LC AND LC-FREE MASS SPECTROMETRY APPLICATIONS IN NON-TARGETED METABOLOMICS FOR DISEASE DETECTION AND EARLY PREDICTION

A Dissertation
Presented to
The Academic Faculty

by

Xiaoling Zang

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Chemistry and Biochemistry

Georgia Institute of Technology
August 2018

LC AND LC-FREE MASS SPECTROMETRY APPLICATIONS IN NON-TARGETED METABOLOMICS FOR DISEASE DETECTION AND EARLY PREDICTION

Approved by:

Dr. Facundo M. Fernández, Advisor
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Ronghu Wu
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. María E. Monge, Co-advisor
Centro de Investigaciones en Bionanociencias (CIBION)
Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

Dr. Matthew Torres
School of Biology
Georgia Institute of Technology

Dr. Julia Kubanek
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Mark Styczynski
School of Chemical and Biomolecular Engineering
Georgia Institute of Technology

Date Approved: July 23, 2018

Never give up on what you really want to do. The person with big dreams is more powerful than one with all the facts.

-Albert Einstein, The World as I See It[1]

ACKNOWLEDGEMENTS

There are many people who have helped and supported me a lot during my 6-year Ph.D. life at Georgia Tech. First of all, I want to express my deepest gratitude to my advisor, Dr. Facundo M. Fernández, for his guidance, support, patience and encouragement during my whole Ph.D. study. Under his mentorship over the last 6 years, I have gained so much knowledge and experience in mass spectrometry and have grown to become more independent and capable in handling the research projects. Secondly, I want to thank my co-advisor, Dr. María Eugenia Monge, who took much of her time in training me in metabolomics research during my beginning years in graduate school and have offered much advice and support to me in my Ph.D. research. Thirdly, I want to thank Dr. David Gaul, who have been very kind and patient in helping me to troubleshoot instrument problems and offering me advice and suggestions in my research projects. These people have been wonderful mentors and friend to me during my journey to pursue a Ph.D. degree and I wish to stay in contact with them and keep learning from them throughout my science career. I would also like to thank my committee members Dr. Julia Kubanek, Dr. Ronghu Wu, Dr. Matthew Torres and Dr. Mark Styczynski for their advisement and interest in my graduate research.

I also want to pay thanks to the previous and present Fernández group members for their advice and help in my Ph.D. study, including Dr. Christina Jones, Dr. Manshui Zhou, Dr. Eric Parker, Dr. Jay Forsythe, Stephen Zambrzycki, José Pérez, Dr. Anyin Li, Dr. Li Li, Dr. Molly Hopper, Danning Huang, Dr. Rachel Stryffeler, Dr. Joel Keelor, Dr. Martin Paine, Dr. Matthew Bernier, Scott Hogan, Adam Kaylor, Kristin Mckenna, Dr.

Yafeng Li, Dr. Rosana Alberici, Dr. Julía Culzoni, Dr. Jennifer Pittman, Dr. Marcos Bouza Areces, David Donndelinger, Dr. Prabha Dwivedi, and Dr. Chaminda Gamage. I want to pay special thanks to Dr. Christina Jones, who have passed her knowledge and experience in mass spectrometry to me during our overlapping years in the group and offered great help and encouragement to me during my Ph.D. life. I also want to thank Dr. Manshui Zhou for his guidance in career and life during my first year of graduate school and the remote contact in the following years in my Ph.D. study.

In addition, I would like to thank all my friends who have accompanied me though my Ph.D. journey in many ways, giving me great help and encouragement, including Yu Cao, Ying Sha, Chu Han, Ran Lin, Annie Cheng, Zhishuai Geng, Haopeng Xiao, Zhanjun Guo, Cong Liu, Mengqi Cui, Miao Zhou, etc.

Finally, I want to express my great gratitude to my family for their unconditional love and support during the 6 years of my Ph.D. life. My parents always trust and believe in me, and never doubt that I can achieve my goals. My dad always gives me laughter and comfort when I encounter difficulties, with his witty and humorous mind. They have helped in any way they can in my academic career and life to make me stronger, and I could not appreciate enough.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

A	Ampere
Å	Angstrom
Da	Dalton
Dt	Drift time
g	Relative centrifugal force
g	Gram(s)
h	Hour(s)
L	Liter(s)
m	Meter(s)
M	Molar concentration
min	Minute(s)
m/z	Mass-to-charge ratio
pH	Negative base-10 logarithm of hydrogen ion molar concentration
ppm	Parts per million
rpm	Revolutions per minute
R_t	Retention time
s	Second(s)
v	Volume
V	Volt(s)
Δm	Mass error
°C	Celcius
Ω	Ohm

APCI	Atmospheric pressure chemical ionization
APE	Acute pulmonary exacerbation
AUC	Area under the receiver operating characteristic curve
AuNR	Gold nanorod
BPI	Base peak intensity
CCS	Collision cross-section
CE-MS	Capillary electrophoresis-mass spectrometry
CF	Cystic fibrosis
CFTR	Cystic fibrosis transmembrane conductance regulator
CV	Cross-validation
DART	Direct analysis in real time
DDA	Data-dependent acquisition
DESI	Desorption electrospray ionization
DI	Direct infusion
DIA	Data independent acquisition
DIMS	Direct infusion mass spectrometry
DTIM	Time-dispersive drift tube ion mobility
EBC	Exhaled breath condensate
ESCi	Electrospray chemical ionization
ESI	Electrospray ionization
FEV_1	Forced expiratory volume in 1 second
FI	Flow injection
FIMS	Flow injection mass spectrometry
FTICR	Fourier transform ion cyclotron resonance
FWHM	Full width at half maximum

GC-MS	Gas chromatography mass spectrometry
HDMS	High-definition mass spectrometry
HILIC	Hydrophilic interaction chromatography
HMDB	Human Metabolome Database
HPLC	High performance liquid chromatography
HR	High resolution
HRMS	High resolution mass spectrometry
IM	Ion mobility
IM-MS	Ion mobility-mass spectrometry
IMS	Ion mobility spectrometry
iPLS-DA	Interval partial least-squares-discriminant-analysis
IVDMIA	*In vitro* diagnostic multivariate index assay
KEGG	Kyoto Encyclopedia of Genes and Genomes
LC-MS	Liquid chromatography-mass spectrometry
LOOCV	Leave-one-out cross-validation
LPC	Lysophosphatidylcholine
LPE	Lysophosphatidylethanolamine
LV	Latent variable
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
NMR	Nuclear magnetic resonance
No.	Number
oPLS-DA	Orthogonal partial least-squares-discriminant-analysis
PC	Principal component
PCA	Principal component analysis

PCa	Prostate cancer
PPTT	Plasmonic photothermal therapy
PSA	Prostate-specific Antigen
QC	Quality control
QC-RLSC	Quality control sample based robust locally estimated scatterplot smoothing signal correction
QTOF	Quadrupole-time-of-flight
RSD	Relative standard deviation
SD	Standard deviation
SERS	Surface enhanced Raman spectroscopy
SVM	Support vector machine
TM	Transmission mode
TM-DART	Transmission mode-direct analysis in real time
TOF	Time-of-flight
TWIM	Traveling wave ion mobility
TWIM-MS	Traveling wave ion mobility-mass spectrometry
TWIMS	Traveling wave ion mobility spectrometry
UPLC	Ultraperformance liquid chromatography
UPLC-MS	Ultraperformance liquid chromatography-mass spectrometry
UPLC-MS/MS	Ultraperformance liquid chromatography-tandem mass spectrometry

SUMMARY

Metabolomics is the science of studying small molecule composition of biological systems. Non-targeted metabolomics, as the analytical technology for unbiased simultaneous measurement and analysis of the collection of low molecular weight metabolites within biological samples, has been widely adopted as a novel and powerful approach to study pathophysiological processes and discover potential biomarkers for disease diagnosis and preventive screening. By comparing and analyzing the global metabolome of different classes of samples with different phenotypes, non-targeted metabolomics can serve as a top-down strategy to discover disease related metabolic perturbations, and it has been applied in studies of various diseases.

In this thesis work, mass spectrometry (MS) based non-targeted metabolomics was applied to discover potential biomarkers of two kinds of diseases: prostate cancer (PCa) and cystic fibrosis (CF) acute pulmonary exacerbations (APEs). Current clinical practices for prostate cancer (PCa) diagnosis focus on prostate-specific antigen (PSA) level. Although it exibits fair discriminating power for PCa detection, the PSA test for PCa screening remains controversial due to the risk of over-diagnosis and overtreatment. Another disease we have studied, CF lung disease, has intermittent episodes of acute worsening of symptoms termed acute pulmonary exacerbations (APEs), which is a major cause of morbidity for CF patients. To date, however, there is no consensus diagnostic criteria for CF APEs. Also, there is no preventive screening method for stable CF patients to signal an oncoming APE event, which hinders the initiation of early intervention before the establishment of substantial immune response. These drawbacks, together with a lack

of in-depth information on the pathophysiology of these two diseases may prevent clinicians from making the best possible therapeutic interventions and treatment decisions to improve patient healthcare. Consequently, there has been a constant drive to discover novel biomarkers to improve PCa diagnosis and prediction of APE onset in CF patients via non-targeted metabolomics strategy.

Mass spectrometry (MS) has been increasingly applied in metabolomics studies due to its high sensitivity. MS methods often include chromatography separation prior to ion detection, which helps to increase metabolite coverage and resolution, decrease spectral congestion and ion suppression (or enhancement) effects. As current metabolomics research focuses more on large scale studies with hundreds to thousands of samples, high-throughput metabolic profiling techniques with fast sample analysis speed become a pivotal necessity. Flow injection (FI) and direct infusion (DI) MS are alternative approaches involving direct introduction of biological samples into MS systems without prior chromatography separation, increasing sample analysis speed. The combination of FI or DI methods with ion mobility (IM) MS is generally appealing for its ability to simplify spectra, raise signal to noise ratio by eliminating chemical noise, produce cleaner MS/MS spectra and provide rapid separation of closely related compounds. Therefore this strategy has great potential in non-targeted metabolomics research demanding high sample throughput. In this thesis work, liquid chromatography (LC) MS method and LC-free FI-IM-MS and DI-IM-MS methods were employed for metabolic profiling of biological samples to find potential biomarkers for PCa and CF APEs and study the associated metabolic perturbations. An introduction to MS-based non-targeted metabolic profiling for human disease studies is provided in **Chapter 1**, with recent developments in disease

biomarker discovery reviewed. Sample preparation, MS platforms utilized, metabolite identification, innovations in data analysis and pathway mapping were discussed.

Part I of the dissertation consists of **Chapters 2 and 3**, which present LC-MS based non-targeted metabolomics studies of PCa and CF APE diseases. In **Chapter 2**, a metabolite-based *in vitro* diagnostic multivariate index assay (IVDMIA) was developed to predict PCa in serum samples with a panel of 40 metabolic features, yielding 92.1% sensitivity, 94.3% specificity, and 93.0% accuracy. The performance of the IVDMIA was demonstrated to be higher than the prevalent PSA test. The identification of amino acids, fatty acids, lysophospholipids, and bile acids provided insights into the metabolic alterations associated with the disease. In addition, several metabolites were mapped to the steroid hormone biosynthesis pathway, indicating its association with PCa. **Chapter 3** discusses the feasibility of predicting APE in CF patients using EBC metabolites. In a pilot study, LC-MS was used to profile metabolites in exhaled breath condensate (EBC) samples in negative ion mode from 17 clinically stable CF patients, 9 CF patients with an APE severe enough to require hospitalization (termed APE), 5 CF patients during recovery from a severe APE (termed post-APE), and 4 CF patients who were clinically stable at the time of collection but in the subsequent 1 to 3 months developed a severe APE (termed pre-APE). Using multivariate analysis, a panel containing 2 metabolic discriminant features identified as 4-hydroxycyclohexylcarboxylic acid and pyroglutamic acid differentiated the APE from the stable CF samples with 84.6% accuracy. In addition, the pre-APE samples were distinguished from the stable CF samples with 90.5% accuracy using a panel of two discriminant features including lactic acid and pyroglutamic acid. In a larger EBC sample cohort (n=210) study, negative ion mode data and the combination of negative and positive

ion mode data showed that classification was possible for age and gender-matched samples grouped into adult and pediatric patients. Negative ion mode data yielded acceptable sensitivities (83.3% and 76.2%), specificities (91.7% and 83.7%), and accuracies (88.9% and 81.3%) for discriminating APE from stable CF EBC samples, from pediatric and adult patients, respectively. For the pre-APE vs. stable CF comparison, good sensitivities (85.7% and 89.5%), specificities (88.4% and 84.1%), and accuracies (87.7% and 85.7%) were obtained for EBC samples from pediatric and adult patients, respectively. By combining positive with negative ion mode data, improved classification performance was achieved for most binary comparisons with accuracies enhanced between 3 and 9.6%. The discriminant metabolites identified in the pilot study were also selected in some of the discriminant metabolite panels. Some of the identified discriminant metabolites had microbial relevance, indicating a possible central role of bacterial metabolism in APE development.

Part II of the dissertation includes **Chapters 4 and 5**, describing non-targeted metabolomics studies on PCa and CF APE disease using LC-free FI-IM-MS and DI-IM-MS. **Chapter 4** presents the application of FI-IM-MS to the non-targeted metabolic profiling of serum extracts from 61 PCa patients and 42 controls from the same cohort in Chapter 2. Comprehensive data mining of the mobility-mass domain was used to discriminate compounds with various charges and filter matrix salt cluster ions. Specific criteria were set to ensure correct grouping of adducts, in-source fragments, and impurities in the dataset. Endogenous metabolites were identified with high confidence using tandem MS experiments and collision cross-section (CCS) matching with chemical standards or CCS databases. PCa patient samples were distinguished from control

samples with good accuracies (88.3-89.3%), sensitivities (88.5-90.2%), and specificities (88.1%) using supervised multivariate classification methods. Results from this study show the potential of FI-IM-MS as a high throughput metabolic profiling tool for large scale metabolomics studies. In **Chapter 5**, transmission-mode direct analysis in real time (TM-DART) coupled to IM-MS was tested as a high-throughput alternative to conventional DI electrospray ionization (ESI) and atmospheric pressure chemical ionization (APCI) methods, and the performances of the three ionization methods were compared. When using pooled EBC collected from a healthy control, ESI detected the most metabolites, and TM-DART the least. TM-DART-TWIM-TOF-MS was used to profile metabolites in the EBC samples from 5 healthy individuals and 4 CF patients, and a panel of 3 discriminant EBC metabolites was found to differentiate these two classes with excellent cross-validated accuracy.

Appendix A presents a collaborative work that combined results from surface enhanced Raman spectroscopy (SERS), metabolomics and proteomics experiments, to study the molecular mechanisms of the cellular processes during the plasmonic photothermal therapy (PPTT) process. Our metabolomics results showed increased levels of phenylalanine and metabolites tentatively identified as its derivatives and phenylalanine-containing peptides, aiding in assignments of SERS bands with observed changes during PPTT. To better understand the mechanism of phenylalanine increase upon PPTT, we combined metabolomics and proteomics results using network analysis, which demonstrated that phenylalanine metabolism was perturbed. In addition, several apoptosis pathways were activated *via* key proteins (e.g. HADHA and ACAT1), which

are consistent with the proposed role of altered phenylalanine metabolism in inducing apoptosis.

At last, **Chapter 6** summarizes the conclusions drawn from the thesis work, and also presents the outlook and possible future work.

REFERENCES

1. Einstein, A.; Rouben Mamoulian Collection (Library of Congress), *The world as I see it*. Abridged ed.; Philosophical Library: New York,, 1949; p xiii, 112 p.

CHAPTER 1. INTRODUCTION

Adapted with permission from
Zang X, Monge ME, Fernández FM. Mass Spectrometry-Based Non-targeted Metabolic Profiling for Disease Detection: Recent Developments. (in preparation).

1.1 Abstract

Mass spectrometry (MS) techniques possess great potential in clinical diagnosis and non-targeted metabolic profiling of biological samples plays an important role in seeking biomarkers for disease detection and prediction. High-quality quantitative data is needed for accurate analysis of metabolic perturbations in patients. This introduction chapter describes recent developments of MS-based non-targeted metabolomics research with applications to the detection of a variety of human diseases (diabetes, liver and breast cancer, cardiovascular disease, Parkinson's disease, etc.), focusing on sample preparation, MS platforms utilized, procedures for metabolite identification, innovations in data analysis, and efforts towards pathway mapping. Potential disease biomarkers discovered are summarized, and limitations and future perspectives discussed.

1.2 MS-based Non-targeted Metabolomics for Disease Biomarker Discovery

Most current screening tests for disease diagnosis suffer from low specificity or sensitivity, among other aspects.[1-6] Diagnosis of lung cancer *via* imaging methods such as computed tomography (CT) and magnetic resonance imaging (MRI), for example, cannot be performed at an early-enough stage due to instrumental resolution limitations.[7] Blood

biomarkers for lung cancer, are often not diagnostic, but only limited to prognostic assessment.[8] In the case of colorectal cancer, diagnosis and treatment decisions are typically based on histopathologic inspection, which often has undesirable precision and uniformity.[9] For cancers such as extrahepatic cholangiocarcinomas (ECC), a rare tumor that arises from epithelial cells of the bile duct and has extremely low survival rate [1, 10, 11], imaging methods such as CT, magnetic resonance cholangiopancreatography, and positron emission tomography, generally have insufficient sensitivity for diagnosis.[1] The current gold standard screening method for gestational diabetes mellitus (GDM), the oral glucose tolerance test, is complicated and time-consuming.[12, 13] Cardiovascular disease is commonly diagnosed by CT angiography, which is an invasive method and may cause cancer risk due to the high dose of radiation exposure.[14] Together, these limitations have led to a need to discover new clinical biomarkers to improve diagnosis, enabling more accurate clinical decisions to be made, and ultimately leading to increased patient survival rates and better clinical outcomes.

Metabolites, as end products of cellular regulatory processes, play central roles in energy metabolism, regulation and signaling, which are essential to life.[15, 16] Being close in biological proximity to the phenotype, metabolites can act as fingerprints of biochemical status, and their dynamic changes could reveal perturbations of a given biological system due to a number of factors, including disease development.[16, 17] Metabolomics, the comprehensive analysis of the collection of all small molecules (MW<1500) in a biological system (the metabolome), opens a window to discovering biomarkers and improving disease diagnosis, showing great promise in clinical applications.[17-19] Disease-associated metabolic signatures and pathways provide

important biological insight into the pathophysiology, revealing the affected underlying molecular mechanisms.[20]

Over the last decade, mass spectrometry (MS) has witnessed rapid growth and has become an indispensable tool in clinical metabolomics to discover biomarkers for human disease prediction, diagnosis, prognosis and follow-up care for patients, owing to its inherent high sensitivity, specificity, throughput, mass accuracy and wide metabolite coverage.[2, 6, 16, 18, 21-24] MS-based metabolomics studies are generally divided into targeted and non-targeted approaches. Targeted metabolomics experiments quantitatively measure a predefined set of metabolites based on a prior hypothesis, commonly by the use of isotope-labeled internal standards.[25-27] On the other hand, non-targeted metabolomics provides the opportunity for novel biomarker discovery, through the global measurement of metabolites within a biological sample to identify a specific metabolic fingerprint responsible for perturbations such as a disease, without knowledge of the molecular and biochemical information of the metabolites being studied.[25, 27, 28] However, global metabolite measurement is challenged by the fact that no single analytical technique can achieve full coverage of the whole metabolome, which contains thousands of metabolites with a wide range of physicochemical properties and concentrations in the biological system.[23, 29] Current non-targeted metabolomics studies are still at the discovery phase, and subsequent validation steps are needed before true translation into the clinical settings is accomplished. These include pre-validation studies of candidate biomarkers through non-targeted metabolomics approaches in a different cohort preferably from a geographically distinct area,[30] targeted metabolomics studies[25] for absolute quantification of biomarkers using isotopically labeled compounds as internal standards, and final

validation in a larger patient cohort with thousands of samples.[30, 31] In recent years, stable isotope labeling has been increasingly favored in MS-based metabolomics research to study metabolic fluxes.[32, 33] It is traditionally used in a highly targeted manner to study known metabolites and predicted pathways, and recently has been combined with non-targeted metabolomics to discover unexpected labeled metabolites and novel pathways.[34, 35] However, wide application of non-targeted stable isotope labeling analysis in metabolomics is currently limited to cell studies, and is hindered by a lack of dedicated data analysis software tools.[34]

MS coupled with chromatography is widely adopted for non-targeted metabolic profiling, owing to its wide coverage and efficient separation of a large number of metabolites in complex biological matrices.[16, 36-38] Gas chromatography (GC) MS[39] and liquid chromatography (LC) MS[40] are the most frequently used MS techniques in non-targeted metabolic profiling studies, and to a lesser extent capillary electrophoresis MS (CE-MS)[41]. The introduction of ultraperformance LC (UPLC) to MS-based metabolomics has greatly improved chromatography efficiency, sensitivity, resolution and throughput in complex biological sample matrices when comparing to traditional high performance LC (HPLC).[42] High resolution (HR) MS with electrospray ionization (ESI) has been increasingly embraced by the metabolomics community due to its intrinsic high mass accuracy, with time-of-flight (TOF) MS, Orbitrap MS and Fourier transform ion cyclotron resonance (FTICR) MS being the most used techniques.[43] The coupling of UPLC and HRMS is highly attractive for metabolic profiling studies, and has enjoyed wide popularity in recent years due to the significantly improved data quality arising from the combined advantages of both UPLC and HRMS.

As current metabolomics shifts its focus to larger scale studies with hundreds to thousands of samples, high-throughput metabolic profiling techniques with high sample analysis speed are becoming a necessity. Direct infusion (DI) and flow injection (FI) MS are alternative approaches to LC, involving direct introduction of the sample into the MS systems without prior chromatographic separation,[44-46] therefore maximizing analysis speed with the capability of running more than 200 samples per day.[44] These high-throughput methods can be combined with ion mobility spectrometry (IMS), a gas-phase electrophoretic technique that rapidly separates ions based on their charges, shapes and sizes,[47] to provide simplified spectra, enhanced signal to noise ratio, and cleaner tandem mass spectrometry (MS/MS) spectra to assist metabolite annotation.[48, 49] In addition, the collision cross-section (CCS) values derived from IMS experiments serve as an orthogonal molecular descriptor in addition to mass-to-charge ratio (*m/z*) to improve the confidence of compound identifications.[47, 49] IMS has been routinely hybridized with mass spectrometers in the past decade and increasingly employed in the metabolomics field.[47, 49-53] Currently, the majority of commercial ion mobility-mass spectrometry (IM-MS) instruments employ traveling wave ion mobility (TWIM)[54] or time-dispersive drift tube ion mobility (DTIM)[55] techniques.

Since metabolite coverage in non-targeted metabolomics studies is determined by both sample preparation and instrument performance, methods must be carefully designed to fulfill the experiment purpose. The challenges associated with sample preparation and MS instrumentation in global metabolic profiling include ionization suppression, matrix effects and inability to achieve full metabolome coverage, leading to the development of new ionization methods and chemical derivatization methods to

increase ionization efficiency and application of complementary analytical methods to improve metabolite coverage.[29, 56] In addition, the analysis of sample preparation blanks, quality controls and quality assurance are important to ensure the quality of metabolomics data acquired in non-targeted studies.[57] With the large volume of raw data generated from non-targeted metabolomics experiments, picking out the true potential biomarkers from a complex dataset with both known and unknown metabolites and interpret the result from a biological perspective can be very challenging. To address this issue, efficient, dedicated and accurate bioinformatics tools have been developed for data processing, modeling and pathway mapping. Also, great efforts have been made in developing new algorithms and databases for accurate and comprehensive annotation of metabolites, which is often difficult and time consuming due to their class diversity and structural heterogeneity.[43, 58, 59] In addition, recent advancements in computational tools offer integrative analysis of multi-omics data, providing deeper insight into pathway dysregulations associated with diseases at a system-wide level.[60-63]

1.3 Recent MS-based Non-targeted Metabolomics Studies in Human Disease Biomarker Discovery

1.3.1 Diabetes

Diabetes is the most frequently studied disease using MS-based metabolomics in the past three years. There are 415 million people affected with diabetes in the world, with type 2 diabetes being the most common form.[64, 65] It is characterized by insulin

resistance caused by lifestyle choices and genetic background, and it accounts for 90% of all types of diabetes.[13, 65] Diagnosis of type 2 diabetes is often delayed until complications occur, with the underlying pathophysiology still remaining elusive.[66, 67] Significant efforts have been made towards discovering new biomarkers for type 2 diabetes diagnosis and prediction using metabolomics.[67-72] Peddinti *et al.* performed non-targeted and targeted metabolic profiling of fasting serum samples from a Finnish cohort containing 543 nondiabetic individuals, including 146 who progressed to type 2 diabetes in a follow-up period of 10 years, by using UPLC-linear trap quadrupole (LTQ)-MS and GC-MS.[68] By combining these measurements with machine learning-based feature selection methods, a panel of discriminant metabolites was found capable of predicting type 2 diabetes development with an average area under the receiver operating characteristic curve (AUC) of 0.75.[68] These differentiating metabolites included novel markers such as α-tocopherol and bradykinin hydroxyproline, as well as known markers like glucose, mannose and α-hydroxybutyrate.[68] In a different UPLC-LTQ-MS based non-targeted metabolomics study of fasting plasma or serum samples from three Swedish cohorts (each consisted of *ca.* 1000 subjects), elevated deoxycholic acid and monoacylglyceride (18:2) and decreased cortisol levels were found to be associated with type 2 diabetes risk.[69] Another non-targeted metabolomics study applied LC-quadrupole-TOF-MS (LC-QTOF-MS) and GC-QTOF-MS to collect serum metabolic profiles from 197 type 2 diabetes and 197 age and gender-matched healthy controls in a Chinese population.[73] Metabolic signatures associated with type 2 diabetes risk were identified with increased branched-chain amino acids (BCAAs), non-esterified fatty acids and lysophosphatidylinositol (LPI) compounds.[73] A panel of 6 metabolites containing proline,

glycerol, aminomalonic acid, LPI (16:1), 3-carboxy-4-methyl-5-propyl-2-furanpropionic acid and urea predicted high-risk patients with an AUC of 0.935, and lysophosphatidylglycerol (LPG) (12:0) and LPI (16:1) predicted low-risk patients with an AUC of 0.781.[73]

Type 1 diabetes is characterized by absolute insulin deficiency, with an onset age that continues to decrease.[13, 74] In order to explore the metabolic background of children who developed type 1 diabetes, a urine metabolic profiling study of 56 type 1 diabetes children and 30 controls was conducted by Galderisi *et al.* using UPLC-QTOF-MS. A total of 59 endogenous metabolites were found to have elevated abundance in children with type 1 diabetes, including steroids, fatty acids and glycerolphospholipids, purine derivatives, carbohydrate conjugates and phenylalanine derivatives, amino acids and small peptides, and gut bacterial metabolites.[74]

Gestational diabetes mellitus (GDM) is typically diagnosed during pregnancy. The current gold standard screening method is the oral glucose tolerance test, often seen as being complicated and time-consuming.[12, 13] To find improved ways for GDM diagnosis, Hou *et al.* comprehensively measured serum metabolic profiles of 131 patients with GDM and 138 controls using GC-TOF-MS and UPLC-QTOF-MS and UPLC-TQ-MS. The authors identified a total of 131 metabolites, including compounds like fatty acids, lipids, amino acids and bile acids which had significant differential distributions between samples from GDM patients and controls.[12] GDM detection was achieved with an AUC around 0.75 using multi-marker models with different combinations of clinical variables and metabolites selected based on the statistical significance between the two groups.[12] Another assay of global plasma metabolomes from 27 GDM patients and 34

healthy controls using UPLC-QTOF-MS revealed significantly decreased abundance of polyunsaturated phospholipids in GDM patients.[75]

1.3.2 Hepatocellular carcinoma (HCC)

HCC is the fifth most prevalent cancer in the world. The current blood screening test using α-fetoprotein (AFP) has low sensitivity in HCC diagnosis.[2, 3] Several studies have been conducted using a non-targeted metabolomics approach to address this issue.[2, 3, 76-78] A global metabolic profiling study of plasma samples from 70 HCC patients and 65 age-matched healthy controls was performed using UPLC-TOF-MS.[2] Multivariate and univatiate analyses of the data identified a panel of five metabolites, including deoxycholic acid 3-glucuronide, 6-hydroxymelatonin glucuronide, 4-methoxycinnamic acid, 11β-hydroxyprogesterone, and 4-hydroxyretinoic acid, leading to the detection of HCC with an AUC of 0.996.[2] Gong *et al.* performed a non-targeted serum metabolic profiling analysis of 51 HCC patients, 49 hepatitis B virus (HBV) cirrhosis patients and 39 healthy controls, using GC-QTOF-MS and UPLC-QTOF-MS. Fourteen metabolites were identified to increase or decrease progressively from the healthy group to HBV-cirrhosis and HCC, and they were therefore identified as to be promising biomarkers.[76] Out of the 14 metabolites, malate, citrate, succinate, lysine, carnitine, proline, ornithine, serine, phenylalanine, tyrosine and arachidonic acid increased in abundance from the healthy group to HBV-cirrhosis and HCC, and arabinose, galactose and uric acid showed a decreasing trend.[76] Di Poto *et al.* conducted non-targeted metabolomics analysis of plasma samples from 63 HCC patients and 65 liver cirrhotic controls using GC-quadrupole-MS (GC-Q-MS) and GC-TOF-MS.[77] Eleven metabolites, including amino

acids, sugars, alcohols, and fatty acids, were selected by least absolute shrinkage and selection operator (LASSO) logistic regression, [79] providing a discrimination of HCC from cirrhotic controls with an AUC value of 0.808, greater than the diagnosis performance of AFP (AUC 0.723).[77]

1.3.3 Breast cancer

Breast cancer is the second leading cause of cancer mortality in women in the US.[80] Although current screening methods using imaging techniques could help to diagnose breast cancer, mortality remained high for tumor at advanced stage, with a 5-year relative survival rate of ~22% for metastatic or stage IV breast cancers.[81, 82] Recent metabolomics studies strived to identify new biomarkers to improve breast cancer diagnosis and staging, which could also be beneficial to its prognosis.[81, 83-85] In a GC-TQ-MS-based global serum metabolic profiling study of 152 breast cancer patients and 155 controls, breast cancer was successfully detected with 99% accuracy by multivariate analysis, and stages and grades of breast cancer were predicted with ~70% accuracy by Decision Tree models.[83] Seven metabolites were found to be significantly altered between breast cancer patients and healthy controls, including tetradecane, α-D-glucopyranoside, methyl stearate, dodecane, 1-4-benzene, D-galactose and octadecanoic acid.[83] Do Canto *et al.* performed non-targeted metabolic profiling of breast ductal fluid samples from affected breasts and the unaffected contralateral breasts from 43 women with unilateral breast cancer using UPLC-QTOF-MS.[81] The authors found 66 features with putative identities that had significant changes between cancerous and non-cancerous regions based on paired *t*-tests.[81] These tentatively identified features included

N-acetyl-tryptophan, *N*-linoleoyl taurine, trans-2-dodecenoylcarnitine, lysophosphatidylcholine (LPC(18:2)), glycerophospholipid(18:0) and phosphatidylserine (20:4).[81]

1.3.4 Cardiovascular disease

Cardiovascular disease is the leading cause of mortality in the US,[86] accounting for over 800,000 deaths in 2015.[87] CT angiography is a widely used imaging method for cardiovascular disease diagnosis, however, it is invasive and may lead to cancer risk in patients due to the high dose of radiation exposure.[14] As a result, less invasive diagnostic methods need to be developed to assess cardiovascular disease risk. In the last three years, several global metabolic profiling studies have been performed to provide new biomarkers for disease detection and to reveal disease-related underlying metabolic mechanisms.[88-92] Li *et al.* conducted a non-targeted plasma metabolic profiling study of 49 patients with cardiovascular disease and 50 controls using UPLC-TripleTOF-MS.[88] *N6*,*N6*,*N6*-trimethyl-L-lysine, a nutrient precursor for gut microbes to the atherogenic metabolite trymethylamine *N*-oxide, was highlighted to be associated with cardiovascular disease risk.[88] The result was validated by a study on a larger cohort (n=1,162) and confirmed by a targeted assay on an independent sample cohort (n=2,140) using stable isotope dilution tandem MS.[88] Non-targeted UPLC-QTOF-MS-based serum lipidomics studies on samples from patients with severe calcific coronary artery disease (n=17) and controls with no calcification (n=26) were performed independently in two laboratories from different countries at different times.[89] Six lipids including phosphatidylcholines (PC(16:0/20:4)), PC(18:2/18:0), PC(18:2/18:2) and *sphingomyelins* (SM(d18:1/16:0)),

SM(d18:1/22:0) and SM(d18:1/23:0) were found to change significantly between patients and controls by both laboratories, and were also detected with correlated intensities between the two laboratories, demonstrating the reproducibility of the results.[89]

1.3.5 Parkinson disease

Parkinson's disease (PD) is the most common neurodegenerative movement disorder, with its clinical diagnosis based on motor symptoms often delayed from the actual onset of the disease.[93, 94] PD has been studied using a non-targeted GC-Q-MS approach by comparing cerebrospinal fluid (CSF) profiles from 44 early-stage sporadic PD patients with those from 43 gender- and age-matched healthy controls.[93] By combining with a logistic regression-based machine learning method, a model based on mannose, threonic acid and fructose was built using a training set of samples (PD patients, n=34, controls, n=35), which successfully distinguished the two groups with an AUC value of 0.833.[93] This model was validated by a small test set (PD, n=10, controls, n=8) and an independent external set (PD, n=24, controls, n=12).[93] Another non-targeted study analyzed CSF and plasma metabolic profiles from early PD patients (n=40 and 80, respectively) and gender- and age-matched controls (n=38 and 76, respectively) by HPLC-QTOF-MS.[94] Using partial least square (PLS) and random forest modeling with a 70:30 training/testing split, 14 CSF metabolites and 20 plasma metabolites were found to be discriminative, detecting early PD with an AUC value of 0.9 and 0.8, respectively.[94] The discriminant metabolites were identified to be associated with glycerophospholipids, sphingolipids and amino acid pathways.[94]

1.3.6 Respiratory diseases

Diagnosis of respiratory diseases is challenging due to various confounding factors including environmental exposures,[95, 96] tobacco use,[97, 98] coexistence of other diseases,[97, 99] viral and bacterial infections,[100, 101] etc. For respiratory disease investigation, non-invasive sampling methods using exhaled breath, exhaled breath condensate (EBC), sputum and saliva are ideal. Analysis of the metabolomes of patients with respiratory diseases may reveal details about the disease pathophysiology, as well as links to bacterial infections. Gaisl *et al*. performed real-time breath analysis of 30 cystic fibrosis (CF) patients and 30 healthy subjects by using secondary ESI-HRMS (SESI-HRMS).[102] A support vector machine (SVM) algorithm selected two discriminant features, tentatively identified as oxohexanoic acid and $C_5H_{10}N_2O_3$, which yielded an AUC of 0.771 for CF prediction.[102] Malkar *et al.* performed a non-targeted metabolomics analysis of saliva samples from 9 asthma patients and 21 healthy individuals using UPLC-TOF-MS.[103] A partial least squares discriminant analysis (PLS-DA) model using 10 discriminant features predicted asthma with a cross-validated accuracy of 97%.[103] Global metabolic profiling of serum and EBC samples from 10 idiopathic pulmonary fibrosis (IPF) patients and 10 healthy controls were performed by Rindlisbacher *et al.* using UPLC-QTOF-MS.[104, 105] Fifty-eight EBC features were found to be differential between the two groups, with one feature tentatively assigned with the elemental formula of $C_{21}H_{44}N_2O$ having a 2.5-fold increase in IPF compared to controls.[104] Also, another study identified a serum LPC with a two-fold increase in abundance in IPF compared to controls, although its structure remains to be elucidated.[105]

1.3.7 Prostate cancer (PCa)

PCa is the second leading cause of cancer mortality in men in the US.[80] The current blood test using prostate specific antigen has low specificity,[4, 5] therefore leading to constant drives in discovering novel PCa diagnostic markers. In a urine metabolic profiling study of 236 PCa patients and age-matched 233 healthy controls by UPLC-TOF-MS, a panel of three metabolites including glycocholic acid, hippurate, 5-hydroxy-L-tryptophan was found to detect PCa presence with AUC > 0.95 using multivariate analysis.[106] McDunn *et al.* applied GC-MS and UPLC-MS/MS to analyze metabolic profiles of from 331 prostate tumor tissues and 178 tumor-free tissues, and identified significant decreased levels of metabolites associated with cell energetics, and elevated levels of amino acids, peptides, carnitines, cofactors, lipids, nucleotides and metabolites associated with stress in tumor tissues.[107] Addition of metabolite markers improved the AUC for predictions of organ confinement and 5-year recurrence from of 0.53 to 0.62 and 0.53 to 0.64, respectively.[107] Wang *et al.* performed global metabolomic imaging to analyze three human prostate tissue specimens with PCa and non-PCa regions using matrix-assisted laser desorption/ionization (MALDI) Q-FTICR MS combined with matrix coating assisted by an electric field technique.[108] Metabolites with differential distributions between PCa and non-PCa tissue regions reflected altered metabolism related to PCa, including elevated energy charge and under-expression of neutral acyl glycerides.[108] Lima et al. analyzed volatile organic compounds profiles in the exometabolome of four PCa cell lines and one normal prostate cell line using GC-ion trap-MS (GC-IT-MS), highlighting significantly increased pentadecane-2-one and

decanoic acid, and significantly decreased 4-methylbenzaldehyde, nonanoic acid, cyclohexanone, 4-methylheptan-2-one, 2-methylpentane-1,3-diol, 1-(3,5-dimethylfuran-2-yl) ethanone and methyl benzoate in PCa cells.[109]

1.3.8 Human immunodeficiency virus (HIV)

Non-targeted metabolomics has also been applied to elucidate metabolic signatures of HIV. In a study by Li *et al.*, non-targeted metabolic profiling was performed using GC-Q-MS and UPLC-LTQ-MS on 38 plasma samples from 19 patients with HIV before and after antiretroviral treatment, and 18 plasma samples from healthy controls.[110] Out of the 331 identified features, 67 were found to discriminate patients with HIV from healthy individuals, belonging to the histidine, tryptophan, fatty acids, and acyl carnitine pathways.[110]

1.3.9 Other diseases

MS-based non-targeted metabolomics has been applied in the investigation of a number of other diseases including childhood acute myeloid leukemia,[111] β-thalassemia,[112] colorectal cancer,[113] endometrial cancer,[114] pancreatic ductal adenocarcinoma,[115] nonalcoholic fatty liver disease,[116] nephrotic syndromes,[117] pneumonia,[118] Crohn's disease,[119] Snyder-Robinson syndrome,[120] shock with respiratory failure,[121] ECC,[1] non-small-cell lung cancer,[122] *etc.*

1.4 Advances in Sample Preparation for Non-targeted Metabolomics for Disease Detection

Since non-targeted metabolic profiling aims to collect as much information as possible from a sample containing numerous metabolites with various physicochemical properties, the sample preparation method should be simple, fast, non-specific and reproducible so as to avoid any sort of unwanted bias.[28, 123] Appropriate sample preparation protocols should be carefully designed and optimized according to the sample type and research objectives to ensure comprehensive metabolome coverage. Typical pretreatment procedures for various sample types are detailed in Table 1.1.

Blood and urine samples are most frequently used for metabolomics.[124] The pretreatment of blood samples (*e.g.* serum or plasma) starts with a deproteinization step by adding the extraction solvent (or mixture of solvents) to the sample, followed by vortex-mixing and centrifugation to collect the supernatant containing the extracted metabolites,[2, 46] which can then be directly used for MS analysis,[24, 125] or lyophilized for storage, followed by reconstitution at a later stage. For GC-MS analysis, the supernatant is dried by lyophilization followed by chemical derivatization to increase the thermal stability and volatility of the metabolites.[125-127] Liquid-liquid extraction (LLE) using a mixture of organic solvents such as methanol (MeOH) and chloroform ($CHCl_3$) allows for simultaneous extraction of polar and lipid metabolites into different phases for separate analysis.[46, 126, 127]

Table 1.1: Typical examples of MS-based non-targeted metabolic profiling studies for human disease detection and prediction.

Sample Type	Disease	Sample pretreatment	MS platform	Data analysis	Biomarkers/Discri-minant metabolites	References
Serum	Insulin resistance	Polar metabolites were extracted by MeOH, followed by lyophilization and derivatization for GC×GC-TOF-MS analysis; lipid metabolites were extracted $CHCl_3$/MeOH (2:1) and analyzed by UPLC-QTOF-MS	GC×GC-TOF and UPLC-QTOF	Mann-Whitney *U*-test and Spearman rank correlation test	Branched-chain amino acids	[127]
Serum	Type 2 diabetes	MeOH for protein precipitation and extraction	UPLC-LTQ and GC-MS	Logistic regression, regularized least-squares regression (RLS) and GreedyRLS	α-Tocopherol and bradykinin hydroxyproline, and known markers such as glucose, mannose and α-hydroxybutyrate	[68]
Serum	Breast cancer	MeOH for protein precipitation followed by SPE in a 96 well plate, followed by lyophilization and derivatization	GC-TQ	PLS-DA, decision tree model, ANOVA, student's *t*-test	Tetradecane, α-D-glucopyranoside, methyl stearate, dodecane, 1-4-benzene, D-galactose and octadecanoic acid	[83]
Serum	Pancreatic cancer related diabetic mellitus	MeOH for protein precipitation and extraction	UPLC-QTOF	PCA and oPLS-DA	N-succinyl-L-diaminopimelic acid and phosphatidylethanolamine (PE(18:2))	[24]
Serum	Non-small-cell lung cancer	Samples were plotted on the chip and dried in the air at room temperature for LDI MS analysis	on-chip LDI MS	oPLS-DA and sPlot	Features with *m/z* of 165.04, 139.03. 141.02, 125.03, 197.13, etc.	[128]
Plasma	GDM in pregnant women	Eluting through Ostro Pass-through 96-well Sample Preparation Plate	UPLC-QTOF	PCA and PLS-DA	Polyunsaturated phospholipids	[75]

Table 1.1 (continued).

Plasma	HCC	Cold MeOH for deproteinization, supernatants were dried and re-suspended	UPLC-QTOF	PCA, oPLS-DA, two-tailed Welch's *t*-test	Deoxycholic acid 3-glucuronide, 6-hydroxymelatonin glucuronide, 4-methoxycinnamic acid, 11β-hydroxyprogesterone, and 4-hydroxyretinoic acid	[2]
Urine	Type 1 diabetes in children	Centrifugation and dilution in 1:5 ratio with 0.1% formic acid	UPLC-QTOF	PCA and PLS-DA	Glycerolphospholipids, steroids, amino acids, fatty acids and gut bacterial metabolites	[74]
Urine	ECC	Centrifugation and filtering by syringe filter	UPLC-QTOF	PCA and oPLS-DA	Sarcosine, allantoin, glutaric acid, salicyluric acid, etc.	[1]
Urine	HCC	Centrifugation and filtering by microporous membrane	UPLC-QTOF	PCA, oPLS-DA and random forests	Palmitic acid, α-N-phenylacetyl-L-glutamine, phytoshingosine, indoleacetyl glutamine and glycocholic acid	[3]
Urine	Type 1 diabetes	Urease pre-treatment, followed by ice-cold MeOH extraction, dried and resuspended	SPE-IM-QTOF	*t*-test	Sucrose, lactose, and one of the two structural isomers with *m/z* 312.21	[129]
EBC	IPF	Protein precipitation by -20 °C ACN/MeOH (1:1)	UPLC-QTOF	PCA and oPLS-DA	One feature with formula $C_{21}H_{44}N_2O$	[104]
EBC	CF and CF acute pulmonary exacerbation	Lyophilization and resuspension to obtain 20-fold up-concentration	UPLC-QTOF DI-IM-QTOF TM-DART-IM-QTOF	oPLS-DA	Lactic acid, pyroglutamic acid, 4-hydroxycyclohexanecarboxylic acid	[130, 131]
Breath	CF	N/A	SESI-HRMS	Mann-Witney *U* test and SVM	Oxohexanoic acid and a feature with formula $C_5H_{10}N_2O_3$	[102]

Table 1.1 (continued).

Tissue	Colorectal cancer	Tissue samples were pestled in liquid nitrogen followed by metabolite extraction with MeOH/$CHCl_3$/H_2O (1:1:0.4), filtering by 5-kDa cutoff filters, lyophilization, and resuspension	CE-TOF-MS	Paired-sample *t*-test	152 significantly changed metabolites including leucine, phenylalanine, arginine, serine, proline, cysteine, methionine, tryptophan, lysine, etc.	[113]
Tissue	Ovarian cancer	N/A	REIMS with QTOF	PCA-LDA, and Wilcoxon rank-sum or Kruskal–Wallis test	Phosphatidic acid (PA(34:1)), PE(36:1), PA(36:2)/PE(34:1), PE(36:2), and PA(P-36:1)	[132]
Tissue	PCa	Cyrosectioning and matrix coating assisted by an electric field	MALDI-MSI with Q-FTICR	Student's *t*-test	Cell energy carrier nucleotides such as adenosine triphosphate, adenosine monophosphate, guanosine triphosphate, etc., and neutral acylglycerides	[108]
Tissue	Brain glioma	Cyrosectioning for tissues and physically spreading to make tissue smears	DESI-MSI with LTQ and Q Exactive Orbitrap	PCA and Kruskal–Wallis test	*N*-acetyl-aspartic acid and 2-hydroxyglutaric acid	[133]
Tissue	Oral tongue squamous cell carcinoma	Cryosectioning and thaw mounting	DESI-MSI with LTQ	PCA-LDA, ANOVA	Cholesterol sulfate and oleic acid dimer	[6]
Tissue	PCa	Cryosectioning and thaw mounting	DESI-MSI with LTQ-Orbitrap XL	Lasso	Glucose/citrate ratio	[134]
CSF	PD	Extraction with MeOH/H_2O (8:1)	GC-Q-MS	Logistic regression and two-tailed Welch's *t*-test	Mannose, threonic acid and fructose	[93]
Cell	Glioma stem-like cell self- renewal and differentiation	-80°C MeOH for quenching and extraction	UPLC-Q Exactive Orbitrap	PCA, oPLS-DA, sPlot, independent *t*-test, R-package camera, etc.	Kynurenine, L-formylkynurenine, stearoylcarnitine, xanthosine, xanthine monophosphate, LPC(15:0), etc.	[135]

In a global metabolic profiling study by Pedersen *et al.*, MeOH was used for polar metabolite extraction, and a mixture of $CHCl_3$ and MeOH (2:1) was used for lipid extraction from serum samples of 291 non-diabetic Danish adults and over 100 Danish patients with diabetes. They detected a total of 325 polar metabolites including 94 known and 231 unknown species, and 876 serum lipids including 289 known and 587 unknown species, using two dimensional GC (GC×GC) TOF-MS and UPLC-QTOF-MS, respectively.[127] Hadi *et al.* prepared serum samples for metabolomics experiments by using MeOH for protein precipitation, followed by solid phase extraction (SPE) in a 96 well plate.[83] The eluate was evaporated under N_2, then dried and derivatized for GC-TQ-MS analysis, with a total of 424 features detected in 307 serum sample extracts from 152 breast cancer patients and 155 healthy controls.[83]

For urine samples, sample preparation is relatively simple due to the high water and low protein content.[46, 136] Urine samples are commonly centrifuged to remove solid debris, and the supernatant may be analyzed with or without dilution.[1, 3, 28, 137, 138] In a metabolic profiling study to understand underlying metabolic mechanisms related to type 1 diabetes in children, urine samples were centrifuged and then diluted in 1:5 ratio with 0.1% (v/v) aqueous formic acid solution for UPLC-QTOF-MS analysis, with a total of 2381 and 1435 features detected in positive and negative ionization modes, respectively, from a cohort consisting of 56 children with type 1 diabetes and 30 controls.[74]

Recently there has been an increasing interest in using EBC for probing pathophysiological processes occurring within the lung due the ease and non-invasive nature of the sample collection.[104, 130, 139-141] EBC consists of aerosolized epithelial lining fluid containing volatile and nonvolatile compounds trapped and diluted by water vapor

condensation,[139, 140] resulting in low metabolite concentrations ranging from nM to μM.[142] Therefore, a preconcentration step is typically recommended for EBC metabolomics.[130, 131] Different EBC sample preparation methods for LC-MS-based metabolic profiling have been compared, with the lyophilization method shown to have a higher number of metabolites detected than methods using protein precipitation or SPE.[143] In our own work (Chapter 3 and 5), we processed EBC samples by lyophilization followed by resuspension in water H_2O/MeOH (90:10) to provide a 20-fold up-concentration, resulting in a total of 491 features extracted from UPLC-QTOF-MS data of EBC samples from CF patients.[130]

Another type of valuable biological fluid is CSF, which is produced by the choroid plexus within the central nervous system and it is an excellent source of information for the study of neurological disorders.[93] In a non-targeted metabolomics study of PD, CSF was extracted using MeOH/H_2O (8:1), vortexed and centrifuged.[93] The collected supernatant was dried in a refrigerated rotary vacuum evaporator and derivatized before GC-Q-MS analysis.[93]

Tissue sample preparation for MS analysis generally includes tissue homogenization and metabolite extraction.[46, 113] The extracts are then dried and resuspended in appropriate solvents before MS analysis.[46, 113] In a non-targeted CE-MS-based metabolic profiling experiment, colorectal cancer and paracancerous tissues were pestled in liquid nitrogen. Metabolites were extracted with a MeOH/$CHCl_3$/H_2O (1:1:0.4) solution, followed by filtering using 5-kDa cutoff membranes, lyophilization, and resuspension before CE-MS analysis.[113] For MSI experiments using MALDI or desorption electrospray ionization (DESI), cryosectioning of frozen tissues followed by

thaw-mounting of tissue slices on glass slides are the most common sample pretreatment steps.[108, 133, 144-146] An alternative strategy is the use of tissue smears, produced by physically spreading a tiny amount of tissue on a glass slide to form a uniform and diffuse layer.[133] By eliminating the freezing and sectioning tissue preparation steps, tissue smears enabled fast diagnosis of brain human tumor and intraoperative implementation of DESI-MSI.[133]

Sample preparation for mammalian cell endometabolome analysis generally includes quenching and extraction.[147] Martano *et al.* developed a fast sampling method for LC-MS-based metabolic profiling of intracellular metabolites in adherent mammalian cells, by applying a 2 s fast wash step in water, followed by cell metabolism quenching using a -20 °C cold MeOH/acetonitrile (ACN)/0.5 M formic acid (2:2:1) solvent mixture, followed by freeze drying, and resuspension prior to nanoLC-MS analysis.[148] We have applied this protocol to extract intracellular metabolites from human oral squamous cell carcinoma cells that underwent plasmonic photothermal therapy (PPTT) for UPLC-QTOF-MS profiling, identifying metabolic changes associated with cancer cell death mechanisms induced by PPTT (Appendix).[149] Cuykx *et al.* prepared HepaRG cells by quenching the cell metabolism with liquid nitrogen, followed by addition of 80% MeOH solution, and LLE with a H_2O/MeOH/$CHCl_3$ (2:3:2) solution to separate polar and non-polar phases, which were then separately analyzed by different UPLC-MS platforms.[150] A lipid profiling method for single cells using a novel biocompatible surface-coated probe coupled with nanoESI (nESI) MS was developed by Deng *et al.* [151] In this approach, lipids were extracted using the biocompatible surface-coated solid-phase microextraction probe, which was inserted into a nanospray tip and subject to analysis in

a LTQ Orbitrap or a FTICR mass spectrometer. A total of sixty lipids were identified from lipid profiling of 100 HepG2 cells using this method.[151] For LC or GC-MS-based cell exometabolomics, conditioned medium is collected, followed by protein precipitation and extraction with methanol, or a centrifugation step to collect the supernatant for analysis.[109, 152] The medium without cells, treated in the same way as the samples, serves as a blank comparison to help identify the metabolites secreted by cells in the samples.[109, 153] In MSI-based exometabolomics, direct analysis by spatially defined desorption/ionization techniques is used to provide a map of metabolite distribution in the cell medium.[153]

1.5 MS-based Metabolic Profiling Platforms

When selecting a proper MS platform for a metabolic profiling study, sensitivity, resolution, throughput and metabolite coverage should be considered, since no single analytical method can offer a complete coverage of the metabolome due to the diversity and heterogeneity of metabolites present in the sample.[154-156] This section covers the recent developments in various MS-based platforms for metabolic profiling, including classical GC-MS and LC-MS methods, CE-MS, high-throughput direct infusion and flow injection techniques, ambient and imaging MS, IM-MS, and other new techniques.

1.5.1 GC-MS and LC-MS

Hyphenated chromatographic techniques such as GC-MS and LC-MS are by far the most widely applied in non-targeted metabolic profiling studies due to their high

sensitivity, resolution and reproducibility.[16, 157] GC-MS typically has high reproducibility for the analysis of volatile compounds with low molecular weights.[158, 159] Non-volatile compounds, however, need to be chemically derivatized to achieve increased thermal stability and volatility.[159] Comprehensive GC×GC TOF-MS drastically improves the peak capacity and resolution, with larger number of compounds detected in a single analysis run compared to GC-MS.[159]

LC-MS is typically used to profile non-volatile metabolites, with polar metabolites analyzed by hydrophilic interaction chromatography (HILIC) and non-polar species by reversed phase (RP) chromatography.[127, 160] In recent years, UPLC-HRMS has been increasingly favored for metabolomics studies due to the significant enhancement offered in terms of sensitivity, resolution and mass accuracy. For example, Wang *et al.* developed a novel on-line heart-cutting two-dimensional (2D) UPLC-HRMS method for comprehensive coverage of both polar and lipid metabolites in a single analytical run.[161] In this method, a C_8 pre-column was used to divide the plasma extract into metabolomic and lipidomic fractions, with the former analyzed on a C_{18} column and the later on a T3 column. This new approach covered 99% features detected by conventional metabolomic and lipidomic UPLC-HRMS approaches, with 447 and 289 metabolites from diverse classes identified in the positive and negative mode, respectively.[161] Some studies have combined both GC- and LC-MS platforms to enhance metabolome coverage. A serum metabolomics study using GC×GC-TOF-MS and UPLC-QTOF-MS identified a metabolome signature of insulin-resistant non-diabetic individuals characterized by increased branched chain amino acid biosynthesis, which also correlated with gut microbiome functional modules.[127]

1.5.2 CE-MS

To date, retention and separation of highly charged and polar compounds still remains challenging for LC-MS. CE-MS solves this problem with its capability to analyze polar ionogenic metabolites in small-volume biological samples. [162, 163] Recently, new interface designs for CE-MS have been tested for metabolomics to increase metabolite coverage and lower detection limits.[164, 165] Gulersonmez and co-workers evaluated the use of sheathless CE-MS with a porous sprayer in anionic metabolic profiling and demonstrated its high separation efficiency, acceptable repeatability, and improved detection limits compared to conventional CE-MS.[165] They implemented this approach for profiling of intracellular metabolites from glioblastoma cell line extracts and identified small organic acids, sugar phosphates and nucleotides by accurate mass and migration time matching to chemical standards.[165] CE-MS was also investigated for tissue metabolic profiling to discriminate advanced adenoma from colorectal cancer, with metabolite enrichment analysis revealing altered pathways of protein biosynthesis and metabolisms of several amino acids.[113]

1.5.3 High-throughput MS: direct and flow injection methods

Large scale metabolic profiling studies involving thousands of samples calls for high-throughput analytical platforms, giving rise to the application of DIMS and FIMS in metabolomics. Ultrahigh resolution MS detectors such as Orbitrap and FTICR are ideal for DIMS or FIMS due to the significant improvement in metabolite annotation accuracy

and peak capacity[46]. Habchi *et al.* demonstrated FIMS using FTICR with a dynamically harmonized cell as a robust analytical tool for large-scale high-throughput metabolomics studies.[166] With a resolving power greater than 10^6, mass accuracy less than 1 ppm, and accurate relative isotopic mass defect measurements, the number of possible elemental compositions was considerably reduced, therefore enhancing compound annotation accuracy and efficiency.[166] Compared to ESI, nanoESI (nESI) has higher sensitivity with reduced ion suppression and enhancement effects owing to the low flow rates.[46, 167] Southam *et al.* presented a protocol of high resolution spectral-stitching nESI DIMS for high-throughput non-targeted metabolomic and lipidomic fingerprinting using an automated chip-based ion source coupled to FTICR or Orbitrap MS.[46] By recording data from a series of overlapping *m/z* windows and subsequently stitching them together to produce a complete spectrum, the dynamic range and sensitivity were considerably increased while still maintaining high mass accuracy.[46] Recently, Sun *et al.* developed a new plasmonic gold chip for laser desorption/ionization (LDI) TOF/TOF-MS-based metabolic profiling of biofluids, with high sensitivity and reproducibility and second-scale sample analysis speed.[128] By coupling with microarrays, this technique allowed for automated sampling using only 500 nL of biofluids including serum, cerebrospinal fluid and urine.[128] For the first time, on-chip LDI-MS was applied for global serum metabolic profiling, successfully differentiating patients with non-small-cell lung cancer from healthy controls through the use of orthogonal partial least squares discriminant analysis (oPLS-DA).[128]

1.5.4 Ambient MS and Imaging MS

The development of ambient MS in the last decade has significantly decreased the sample preparation and handling requirements in metabolomics.[23] By combining different desorption and ionization techniques, ambient MS allows for detection of both polar and non-polar, volatile and non-volatile compounds.[168] Direct analysis in real time (DART) is an open air direct sampling plasma ionization technique, in which the sample is exposed to a stream of heated gas flow with excited metastable species that are responsible for ionization through gas-phase mechanisms.[131, 169] Our group coupled transmission mode (TM) DART to IM-MS to perform global metabolic profiling of EBC samples from CF patients and controls, and a panel of three metabolites was found to discriminate the two groups with excellent cross-validated accuracy by oPLS-DA analysis (Chapter 5).[131] Gu *et al.* combined DART-MS, nuclear magnetic resonance (NMR) and multivariate statistical methods for metabolomic analysis of serum samples from 27 breast cancer patients and 30 healthy controls, providing discrimination between the two classes.[170] Rapid evaporative ionization MS (REIMS) was developed for online tissue sample analysis leveraging the detection of compounds in tissue aerosols produced from electrosurgical dissection.[171] REIMS is highly suitable for intra-operative diagnosis by analyzing lipid-based fingerprints obtained in real-time by means of multivariate analysis.[171] This technique was applied to perform lipidomics analysis during electrosurgical dissection of gynecological tissues and successfully discriminated ovarian cancer from normal or borderline tissues with accuracies of 93.5% and 90.0%, respectively.[132] REIMS has also been applied to analyze tissue lipidome from 28 patients with colorectal cancer, distinguishing between cancer and normal adjacent mucosa (NAM) with an AUC of 0.96 and between NAM and adenoma with an AUC of 0.99.[172]

Lipidomic fingerprints of cancer, healthy mucosa and adenomas were characterized by over-expression of long-chain phosphatidylserines and bacterial phosphatidylglycerols, plasmalogens and triacylglycerols, and ceramides, respectively.[172] MS-based imaging capabilities can be achieved with DESI, among other ambient MS techniques. DESI-MSI allows measuring spatial distribution of metabolites and lipids in tissue samples, providing chemical information related to histopathological states that could be utilized for disease diagnosis and resection guidance.[6, 134, 173, 174] In DESI, electrospray charged droplets are directed onto the sample surface to create a desorption event, followed by analyte ionization *via* ESI mechanisms.[175] DESI-MSI has the advantages of fast sample analysis and little sample preparation, and it has been applied for diagnosis of brain tumor,[133] oral tongue squamous cell carcinoma[6] and PCa[134]. Banerjee *et al.* applied DESI-MSI for metabolite and lipid profiling of human cancerous and normal prostate tissues and detected PCa with ~90% accuracy using LASSO.[134] The authors found the abundance ratio of glucose/citrate to be capable of identifying PCa specimens, which showed great promise for real-time diagnosis to guide resection in the surgery.[134]

Secondary electrospray ionization (SESI) MS was developed for rapid and sensitive analysis of trace volatile compounds in breath and vapors.[176, 177] In SESI, neutral vapors are ionized by interacting with a pure electrospray solvent at atmospheric pressure *via* gas phase chemical ionization.[177, 178] Gaisl *et al.* performed real-time exhaled breath profiling of 30 CF patients and 30 controls by SESI-HRMS.[102] From a total of 3273 features detected, 49 were found to be have significantly changes between CF and control samples, including several fatty acids and compounds related to airway bacteria colonization.[102] Martinez-Lozano Sinues *et al.* coupled a laboratory-built SESI ion source

and IT mass spectrometer to analyze breath samples from 14 breast cancer patients and 11 controls.[179] SVM modeling analysis of the breath profiles discriminated breast cancer patients from healthy controls with high classification sensitivity and specificity above 90%.[179]

1.5.5 IM-MS

Although still relatively new in terms of applications to non-targeted metabolomics, a few examples in the literature already explore the use of IM-MS in metabolic profiling of biological samples. Paglia and Astarita, for example, established a protocol for metabolic and lipidomic profiling by using a UPLC-TWIM-MS platform.[52] The addition of IMS to UPLC-MS was shown to provide increased peak capacity and spectral clarity, with various classes of lipids and metabolites mapped to distinct trend lines on the 2D CCS–mass plot.[52] The authors also demonstrated the utility of LC-TWIM-MS for improving metabolite identification confidence, thanks to the use of CCS as an additional orthogonal descriptor and the increased fragmentation specificity obtained following TWIM separations.[52] In another study, the performance of UPLC-IM-MS was evaluated for human plasma and HaCaT cell metabolic profiling, with results showing the benefits of IMS in reducing chemical noise, increasing peak capacity, and improving isomer separation and compound annotation.[180] Zhang *et al.* established a novel metabolic profiling platform by combining RapidFire ultra-fast online SPE with an IM-MS system, reporting that this technique was capable of profiling endogenous metabolites and xenobiotics in human plasma and urine with excellent reproducibility and sensitivity, and a 10-s sample-to-sample duty cycle.[129] The authors applied SPE-IM-

MS for global metabolic profiling of urine samples from type 1 diabetes (T1D) patients and controls, and they identified disaccharides and a previously unreported isomer with significant alterations between T1D patients and controls.[129] Maleki *et al.* investigated the potential of using IM-MS coupled with gas-phase hydrogen/deuterium exchange (HDX) for metabolomics, and showed that the combination of ion mobility with HDX reactivity data facilitated metabolite identification.[181]

1.5.6 Multiplexed activity metabolomics

In a recent study by Earl *et al.*, a multiplexed activity metabolomics (MAM) method was developed by combining cytometry and single-cell biology with metabolomic arrays in non-targeted HPLC-MS analysis, with the aim to identify anticancer metabolites that could target primary human leukemia cells.[182] In this methodology, microbial crude extracts were subject to split-flow polarity-switching HPLC-MS to generate metabolomics data and simultaneously produce metabolomic arrays in a microtiter plate. Bone marrow mononuclear cells from an acute myeloid leukemia patient were added and incubated with the metabolomic arrays in the microtiter plate wells to allow for interaction and biological response production. Fluorescence cytometry cell barcoding and immunoassays were used for quantitation of biological responses for each cell type.[182] Targeting effects of the metabolomic arrays were identified by correlation analysis of the bioassay result and metabolomics data from HPLC-MS analysis. Using this approach, the authors discovered a microbial anthracycline that could target leukemia blasts and a polyene macrolactam that could target either leukemia blasts or nonmalignant cells, depending on the

photoisomerization.[182] MAM therefore proved to be useful for elucidating the specific targeting effects of microbial metabolites to the bioactivities of different types of cells from cancer patients, demonstrating a high potential in drug discovery applications.[182]

1.6 Data Analysis

Development of robust and efficient data processing software tools for metabolomics is challenging due to the complex and large-volume nature of the associated data, and the diverse analytical platforms used by different laboratories.[183, 184] The most widely used free software tools for LC-MS and GC-MS-based metabolomics data preprocessing and analysis have been recently reviewed by Spicer *et al.*, with XCMS[185] and MZmine 2[186] being the most popular ones.[184] However, some of these software packages suffer from problems in feature extraction and integration, leading to incentives to develop new algorithms for improved feature quantification. For example, a new metabolomics data preprocessing method named bakedpi was developed to reduce unnecessary quantification variability in XCMS or MZmine 2, by applying intensity-weighted bivariate kernel density estimation to the 2D *m/z*–retention time space of a metasample created from pooling of all samples in an experiment.[187] Mass spectral feature list optimizer (MS-FLO) was developed by DeFelice *et al.* for non-targeted metabolomics data processing, which could identify the erroneous features including duplicate peaks, isotopic peaks and adducts that failed to be removed or grouped by common software packages based on retention time alignment, accurate mass matching, and peak height correlation.[188] Common software packages for LC-MS metabolomics data processing cannot handle FIMS data directly due to the high variability of peak

shapes in the flowgram, therefore a software tool named proFIA was developed for automatic preprocessing of FI-HRMS data.[183] In proFIA, an innovative approach was adopted for feature detection and quantification using analyte flowgram modeling by taking into account matrix effects, solvent baseline, and experimental noise.[183]

Various normalization methods have been applied to non-targeted metabolomics datasets with extracted spectral features to reduce unwanted biological and experimental variations,[189] including MS total useful signal (MSTUS) normalization,[190] median normalization,[191] probabilistic quotient normalization (PQN),[192] quality control sample based robust LOESS (locally estimated scatterplot smoothing) signal correction (QC-RLSC),[16] etc. Normalization methods should be carefully selected to obtain correct results from non-targeted metabolomics studies. Li *et al.* developed an online tool named NOREVA for comprehensive evaluation of the performance of 24 popular MS-based metabolomics data normalization methods using multiple criteria, allowing users to choose the best normalization method for their dataset.[193] Gagnebin *et al.* proposed a normalization strategy with three sequential steps: normalization by osmolality dilution, QC-RLSC normalization, and MSTUS or PQN normalization.[137] Application of this strategy to UPLC-MS urine metabolomics data collected from patients with kidney failure and healthy controls showed the sequential normalization method improved the predictive ability and reduced the complexity of the oPLS-DA models.[137] An R-based mixture modeling approach named mixnorm was developed for large-scale non-targeted GC-MS metabolomics data normalization.[194] This method accounted for batch order, run order, and threshold of detectability, making it more suitable than other normalization

methods for handling low abundance metabolites with varied detectability thresholds across batches.[194]

Metabolic features with normalized abundances are typically analyzed by univariate and multivariate methods to select discriminant metabolites that are statistically significant and important for sample classification. Univariate analyses consist of parametric tests including Student's *t*-test, [12, 83, 113, 144] a nested analysis of variance (ANOVA) test,[6, 83] and non-parametric tests including the Mann-Witney *U* test,[12, 127] Wilcoxon rank-sum test,[132] and Kruskal–Wallis test[132, 133].[195, 196] When performing univariate analyses on multiple potential biomarkers, Bonferroni correction[197] or Benjamini-Hochberg procedure[198] should be applied to adjust the false discovery rate. Multivariate analyses consist of unsupervised methods such as principal component analysis (PCA),[1, 3, 74, 75, 104, 121, 132, 135] supervised classification methods such as PLS-DA or oPLS-DA,[1, 3, 12, 74, 75, 83, 104, 121, 130, 135] linear discriminant analysis (LDA),[132] SVMs,[102, 179, 199], random forests,[3, 94] *etc*.[200-204] For feature selection, commonly used approaches include univariate tests, PLS-DA model-embedded features including interval PLS-DA (iPLS-DA),[130] variable importance on projection (VIP) and loading weights,[202] recursive feature elimination,[4] greedy feature-selection for regularized least-squares (RLS) (GreedyRLS),[68, 205] and LASSO regression[77, 79].

ominiClassifier was developed as a prediction modeling method for big data analysis, where the prediction model parameters are optimized using cross validation and the final model validated by an external dataset.[206] It has been applied to analyze UPLC-MS serum metabolomics data from rats with traumatic brain injury.[207] From the 120 models built based on various classifiers, the best models were chosen, yielding a panel

of 26 lipid compounds which differentiated injured and uninjured samples with 85.3% accuracy.[207]

Huang *et al*. described a novel strategy to construct pathway-based metabolomics data from the original dataset, which was then subject to feature selection and multivariate modeling for classification.[208] This method was tested on a dataset generated by the metabolic profiling of breast cancer and control blood samples using LC-MS and GC-MS, with the results showing increased classification power compared to metabolite-based data and revealing potential pathways associated with breast cancer.[208]

Recently, deep learning (DL) has gained attention in many aspects in computational biology and genomics research.[209-212] DL is superior to shallow machine learning methods in the sense that it transforms simple features of the data into a high-level hierarchical structure with multiple layers of neurons to maximize the model accuracy and find robust features.[209, 213] Alakwaa *et al.* evaluated the performance of DL in the analysis of data collected from GC-MS based metabolic profiling of breast cancer tissue samples.[213] Compared with other commonly used machine learning methods, DL yielded the highest average AUC of 0.93 for classification of estrogen receptor status in breast cancer, and revealed significantly enriched pathways of protein digestion and absorption, and ATP-binding cassette transporters.[213]

1.7 Metabolite Annotation and Identification

The major challenge in non-targeted metabolomics data analysis is the comprehensive and accurate identification of metabolites. The most common approach is

data-dependent acquisition (DDA), in which a tandem MS scan is triggered for a precursor ion that passes a certain intensity threshold in the MS survey scan.[214, 215] DDA is capable of producing good quality MS/MS spectra, however, it is biased towards high abundant precursor ions and therefore the coverage is not ideal.[216-218] To improve MS/MS efficiency and the coverage for low signal intensity precursor ions, data independent acquisition (DIA) approaches have been developed, including sequential window acquisition of all theoretical mass spectra (SWATH-MS),[219, 220] all ion fragmentation (AIF),[221] and MS^{E}[222], for unbiased and systematic generation of MS/MS data.[215, 223, 224] In DIA, MS/MS is acquired for each predefined *m/z* interval without the need for a survey scan, thus it is independent of the precursor ion intensity.[224, 225] The common strategy for metabolite identification is to compare experimental MS/MS spectra to those in reference libraries including NIST and Wiley for GC-MS, and Metlin, Massbank, mzCloud, the Human Metabolome Database (HMDB), and LipidMaps for LC-MS/MS.[215, 217, 223, 226] A detailed review on databases and software tools for MS/MS identification of small molecule metabolites was reported by Kind *et al*.[215] However, accurate and unique compound identification using LC-MS/MS still remains a bottleneck in non-targeted metabolomics studies due to the limited number of MS/MS spectra in public databases, and the variation of MS/MS relative ion abundances across different instruments and conditions.[217] These limitations have triggered new developments of in-house databases and algorithms to improve metabolite identification coverage and accuracy. Along these lines, a LC-MS/MS metabolomic spectral library was constructed using SWATH acquisition on 532 metabolites from HMDB, at 16 discrete collision energies with a wide range from 5 to 100 eV, aiming to generate more informative MS/MS for DIA metabolite

identification.[223] An in-house LC-HRMS metabolomics library of 408 metabolites was created using all ion fragmentation (ALF) acquisition, with product/precursor ion ratio added to conventional identification criteria based on accurate mass, retention time and MS/MS.[221] Chen *et al.* developed a bioinformatics tool named MetaboDIA to allow users to build customized MS/MS spectral libraries using their own consensus DDA data, maximizing the utility of MS/MS data for metabolite identification.[217] Uppal *et al*. implemented an automated workflow using a multistage scoring algorithm named xMSannotator to assign confidence levels to tentative identities of spectral features from database searches, accounting for intensity correlations, retention time clustering, mass defect, isotope and adduct characteristics, and biological pathway information.[227] Li *et al*. developed MetDIA to analyze metabolomics data acquired using SWATH by targeted extraction of metabolites according to accurate mass match to 786 metabolites from an in-house spectral library.[224] Metabolite identification was based on the correlation score calculated for the extracted ion chromatograms of the precursor and fragments and the similarity score calculated between the extracted experimental MS/MS and the corresponding MS/MS spectra in the library.[224] When applied to biological sample analysis, MetDIA identified 152, 138 and 192 metabolites in human serum, *E. coli* bacteria, and rat liver tissue, respectively.[224] In addition to SWATH library search, LC retention time prediction could further support metabolite identification when authentic standards are not available.[228-231] Bruderer *et al.* evaluated LC retention time prediction for a metabolomics database with 532 compounds in HMDB.[231] A quantitative structure retention relationship model was built with 16 compounds representing the whole library for retention time prediction, using logD2 (consensus logP and classic pKa) and the

molecular volume as molecular descriptors, which was validated on two types of C_{18} columns with comparable prediction accuracy.[231] Combined with SWATH MS/MS spectra, LC retention time prediction aided in assignment of isomeric metabolites in human urine, enhancing metabolite identification confidence and reducing false positives.[231]

Recently, a unified method to annotate metabolites by integrating three cheminformatics tools was proposed by Lai *et al.*,[232] in which BinVestigate was used to query BinBase, a large GC-MS non-targeted metabolomics database containing 1,561 studies with 114,795 samples, to obtain biological metadata of the metabolites. Then, MS-DIAL was applied for spectral deconvolution for unknown compounds from GC-MS or LC-MS/MS data, followed by the use of MS-FINDER for formula prediction and structural annotation of compounds.[232] By adding to MS-FINDER all the Metabolic In silico Network Expansions Database (MINE-DB) virtual epimetabolites, predicted by applying enzymatic transformations on KEGG metabolites, this novel systematic strategy offered new possibilities for improving the chances of correct annotation of unknown metabolites.[232] For example, an unknown compound was identified and confirmed as *N*-methyl-uridine monophosphate (UMP), a predicted metabolite in MINE-DB that has never been detected in biological samples.[232]

With the increasing popularity of IM-MS technology in metabolomics research, compound identification accuracy is improved by the use of CCS as an additional physiochemical descriptor to provide structural information of the ion.[233] To date, many CCS databases have been constructed to support metabolite identification by measurement of a large number of standards using commercial IM-MS instruments or by

CCS prediction using computational tools.[234] Zheng *et al.* created a CCS database of over 500 metabolites and xenobiotics measured by DTIM-MS.[235] Zhou and coworkers developed a support vector regression-based CCS prediction algorithm with 14 molecular descriptors, which was validated to have higher prediction precision than MOBCAL (a software for theoretical CCS calculation) using an external set of metabolites.[233] With this approach, the authors generated a large-scale database of predicted CCS for 35,203 HMDB metabolites named MetCCS.[233] A web server MetCCS predictor was later developed by the same group for rapid prediction of compound CCS values using HMDB ID or other chemical identifiers from user input.[236]

1.8 Pathway Mapping and Multi-omics Analysis

Metabolic pathway mapping and multi-omics analysis by combining metabolomics data with transcriptomics and proteomics data could offer systematic understanding of the dynamic interactions among biomolecules in different layers of the hierarchical biological system.[18] Popular bioinformatics tools for pathway analysis in non-targeted metabolomics studies include MetaboAnalyst,[1, 3, 113, 135, 237] Galaxy-M,[238] mummichog,[239] and Open MS[240]. In 2018, MetaboAnalyst was updated to version 4.0 with a new feature to integrate metabolomics, metagenomics and transcriptomics data for network analysis.[61] Another tool to link metabolomics data to other omics data is xMWAS, which is capable performing integrative analysis of four datasets from different omics platforms, differential network analysis and grouping of related metabolites, genes and proteins into communities.[241] WikiPathways is an open and collaborative platform for the research community to edit and curate pathway information obtained from

integrated transcriptomics, proteomics and metabolomics data, promoting the growth and accuracy of the pathway database.[62, 242] A novel debiased sparse partial correlation algorithm was developed to calculate partial correlation networks among features in large-scale metabolomics data using the CorrelationCalculator program, and the network can be visualized using the Metscape tool.[243] This algorithm aided the annotation of unknown features in non-targeted metabolomics by comparing with related features that were identified in the same network or adjacent subnetwork, providing the chance to discover unexpected interactions between metabolites.[243]

Most of the current pathway analysis software tools require preprocessed data and/or validation of metabolite identities using other programs,[244] therefore increasing the total time required for analysis and the complexity of the workflow. Huan *et al*. provided an online XCMS workflow encompassing all steps in global metabolomics data analysis including raw LC-MS data preprocessing, differential analysis, dysregulated pathway analysis, and combining proteomic and transcriptomic data to provide a deeper insight into metabolic mechanisms on a system-wide scale, which significantly increased analysis efficiency with the whole process taking around 1.5-3.5 h.[63, 244] Chemical similarity enrichment analysis (ChemRICH) was developed by Barupal *et al*. as an alternative approach to pathway mapping for metabolomics studies. In this approach metabolites are clustered into non-overlapping sets based on chemical structure similarity and ontologies, from which statistical significant metabolite sets were identified by Kolmogorov-Smirnov testing.[245] Using this method to analyze published plasma metabolomics data from non-obese diabetic mice and controls, 90% coverage was achieved for the identified metabolites, in contrast to only 40% coverage using classical

pathway mapping.[245] Significant altered metabolite sets including complex lipids, branch-chain amino acids and compounds related to carbohydrate metabolism were found, which were not clearly identified in the original literature.[245]

1.9 Limitations and Outlook

Rapid developments in MS-based non-targeted metabolic profiling strategies in the last decade have greatly promoted disease biomarker discovery in clinical metabolomics science. Nonetheless, significant efforts have yet to be made towards standardization of workflows and stringent validation of the discovered biomarkers using targeted assays, independent patient cohorts with larger numbers of samples, and repetition on different analytical platforms in different laboratories.[246, 247] In addition, there is still space for improvement in metabolite coverage and efficient and accurate compound identification to ensure correct biological interpretation. As more high-throughput large-scale MS-based metabolomics experiments are being performed, the number and size of the datasets continues to grow. Public data repositories such as Metabolomics Workbench [248] and MetaboLights [249] make it possible for researchers from different laboratories around the world to share and re-analyze data,[250] expanding our knowledge gained from such experiments and promoting the healthy growth of the metabolomics community.[250, 251] At present, metabolomics is at the research laboratory-based discovery level. With continuous and collaborative efforts from the metabolomics community, we see hope for a true translation into the clinic in the future. This transition will provide tremendous benefit for improving early disease diagnosis to help guide

clinical intervention, ultimately leading to higher survival rates, better patient care, and lower health care costs.[30]

1.10 References

1. Wang, X. X.; Li, J.; Zhang, A. H., Urine metabolic phenotypes analysis of extrahepatic cholangiocarcinoma disease using ultra-high performance liquid chromatography-mass spectrometry. *Rsc Advances* **2016,** 6, (67), 63049-63057.

2. Li, Y. F.; Qiu, S.; Gao, L. J.; Zhang, A. H., Metabolomic estimation of the diagnosis of hepatocellular carcinoma based on ultrahigh performance liquid chromatography coupled with time-of-flight mass spectrometry. *Rsc Advances* **2018,** 8, (17), 9375-9382.

3. Liang, Q.; Liu, H.; Wang, C.; Li, B., Phenotypic Characterization Analysis of Human Hepatocarcinoma by Urine Metabolomics Approach. *Sci Rep* **2016,** 6, 19763.

4. Zang, X. L.; Jones, C. M.; Long, T. Q.; Monge, M. E.; Zhou, M. S.; Walker, L. D.; Mezencev, R.; Gray, A.; McDonald, J. F.; Fernandez, F. M., Feasibility of Detecting Prostate Cancer by Ultraperformance Liquid Chromatography-Mass Spectrometry Serum Metabolomics. *Journal of Proteome Research* **2014,** 13, (7), 3444-3454.

5. Tombal, B., Over- and underdiagnosis of prostate cancer: The dangers. *European Urology Supplements* **2006,** 5, (6), 511-513.

6. D'Hue, C.; Moore, M.; Summerlin, D. J.; Jarmusch, A.; Alfaro, C.; Mantravadi, A.; Bewley, A.; Gregory Farwell, D.; Cooks, R. G., Feasibility of desorption electrospray ionization mass spectrometry for diagnosis of oral tongue squamous cell carcinoma. *Rapid Commun Mass Spectrom* **2018,** 32, (2), 133-141.

7. Arya, S. K.; Bhansali, S., Lung Cancer and Its Early Detection Using Biomarker-Based Biosensors. *Chemical Reviews* **2011,** 111, (11), 6783-6809.

8. Liu, B.; Li, Y. L.; Wan, H.; Wang, L.; Xu, W.; Zhu, S. J.; Liang, Y. Y.; Zhang, B.; Lou, J. T.; Dai, H. J.; Qian, K., High Performance, Multiplexed Lung Cancer Biomarker Detection on a Plasmonic Gold Chip. *Advanced Functional Materials* **2016,** 26, (44), 7994-8002.

9. Sanduleanu, S.; Driessen, A.; Gomez-Garcia, E.; Hameeteman, W.; de Bruine, A.; Masclee, A., In vivo diagnosis and classification of colorectal neoplasia by

chromoendoscopy-guided confocal laser endomicroscopy. *Clin Gastroenterol Hepatol* **2010,** 8, (4), 371-8.

10. Altekruse, S. F.; Petrick, J. L.; Rolin, A. I.; Cuccinelli, J. E.; Zou, Z. H.; Tatalovich, Z.; McGlynn, K. A., Geographic Variation of Intrahepatic Cholangiocarcinoma, Extrahepatic Cholangiocarcinoma, and Hepatocellular Carcinoma in the United States. *Plos One* **2015,** 10, (4).

11. Anderson, C. D.; Pinson, C. W.; Berlin, J.; Chari, R. S., Diagnosis and treatment of cholangiocarcinoma. *Oncologist* **2004,** 9, (1), 43-57.

12. Hou, W. L.; Meng, X. Y.; Zhao, A. H.; Zhao, W. J.; Pan, J. M.; Tang, J. L.; Huang, Y. J.; Li, H. P.; Jia, W.; Liu, F.; Jia, W. P., Development of Multimarker Diagnostic Models from Metabolomics Analysis for Gestational Diabetes Mellitus (GDM). *Molecular & Cellular Proteomics* **2018,** 17, (3), 431-441.

13. American Diabetes, A., 2. Classification and Diagnosis of Diabetes. *Diabetes Care* **2017,** 40, (Suppl 1), S11-S24.

14. Alkhorayef, M.; Babikir, E.; Alrushoud, A.; Al-Mohammed, H.; Sulieman, A., Patient radiation biological risk in computed tomography angiography procedure. *Saudi J Biol Sci* **2017,** 24, (2), 235-240.

15. Fiehn, O., Metabolomics - the link between genotypes and phenotypes. *Plant Molecular Biology* **2002,** 48, (1-2), 155-171.

16. Dunn, W. B.; Broadhurst, D.; Begley, P.; Zelena, E.; Francis-McIntyre, S.; Anderson, N.; Brown, M.; Knowles, J. D.; Halsall, A.; Haselden, J. N.; Nicholls, A. W.; Wilson, I. D.; Kell, D. B.; Goodacre, R.; C, H. S. M. H., Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nature Protocols* **2011,** 6, (7), 1060-1083.

17. Patti, G. J.; Yanes, O.; Siuzdak, G., Metabolomics: the apogee of the omics trilogy. *Nature Reviews Molecular Cell Biology* **2012,** 13, (4), 263-269.

18. Tebani, A.; Abily-Donval, L.; Afonso, C.; Marret, S.; Bekri, S., Clinical Metabolomics: The New Metabolic Window for Inborn Errors of Metabolism Investigations in the Post-Genomic Era. *Int J Mol Sci* **2016,** 17, (7).

19. Guma, M.; Tiziani, S.; Firestein, G. S., Metabolomics in rheumatic diseases: desperately seeking biomarkers. *Nature Reviews Rheumatology* **2016,** 12, (5), 269-281.

20. Johnson, C. H.; Ivanisevic, J.; Siuzdak, G., Metabolomics: beyond biomarkers and towards mechanisms. *Nature Reviews Molecular Cell Biology* **2016,** 17, (7), 451-459.

21. Gowda, G. A.; Djukovic, D., Overview of mass spectrometry-based metabolomics: opportunities and challenges. *Methods Mol Biol* **2014,** 1198, 3-12.

22. Wang, Y.; Liu, S. Y.; Hu, Y. J.; Li, P.; Wan, J. B., Current state of the art of mass spectrometry-based metabolomics studies - a review focusing on wide coverage, high throughput and easy identification. *Rsc Advances* **2015,** 5, (96), 78728-78737.

23. Clendinen, C. S.; Monge, M. E.; Fernandez, F. M., Ambient mass spectrometry in metabolomics. *Analyst* **2017,** 142, (17), 3101-3117.

24. He, X. Y.; Zhong, J.; Wang, S. W.; Zhou, Y. F.; Wang, L.; Zhang, Y. P.; Yuan, Y. Z., Serum metabolomics differentiating pancreatic cancer from new-onset diabetes. *Oncotarget* **2017,** 8, (17), 29116-29124.

25. Roberts, L. D.; Souza, A. L.; Gerszten, R. E.; Clish, C. B., Targeted metabolomics. *Curr Protoc Mol Biol* **2012,** Chapter 30, Unit 30 2 1-24.

26. Vinayavekhin, N.; Saghatelian, A., Untargeted metabolomics. *Curr Protoc Mol Biol* **2010,** Chapter 30, Unit 30 1 1-24.

27. Alonso, A.; Marsal, S.; Julia, A., Analytical methods in untargeted metabolomics: state of the art in 2015. *Front Bioeng Biotechnol* **2015,** 3, 23.

28. Khamis, M. M.; Adamko, D. J.; El-Aneed, A., Mass spectrometric based approaches in urine metabolomics and biomarker discovery. *Mass Spectrom Rev* **2017,** 36, (2), 115-134.

29. Vuckovic, D., Current trends and challenges in sample preparation for global metabolomics using liquid chromatography-mass spectrometry. *Anal Bioanal Chem* **2012,** 403, (6), 1523-48.

30. Trivedi, D. K.; Hollywood, K. A.; Goodacre, R., Metabolomics for the masses: The future of metabolomics in a personalized world. *New Horizons in Translational Medicine* **2017,** 3, (6), 294-305.

31. Koulman, A.; Lane, G. A.; Harrison, S. J.; Volmer, D. A., From differentiating metabolites to biomarkers. *Anal Bioanal Chem* **2009,** 394, (3), 663-70.

32. Bueschl, C.; Krska, R.; Kluger, B.; Schuhmacher, R., Isotopic labeling-assisted metabolomics using LC-MS. *Analytical and Bioanalytical Chemistry* **2013,** 405, (1), 27-33.

33. Dai, Z. W.; Locasale, J. W., Understanding metabolism with flux analysis: From theory to application. *Metabolic Engineering* **2017,** 43, 94-102.

34. Weindl, D.; Wegner, A.; Hiller, K., Metabolome-Wide Analysis of Stable Isotope Labeling-Is It Worth the Effort? *Frontiers in Physiology* **2015,** 6.

35. Creek, D. J.; Chokkathukalam, A.; Jankevics, A.; Burgess, K. E.; Breitling, R.; Barrett, M. P., Stable isotope-assisted metabolomics for network-wide metabolic pathway elucidation. *Anal Chem* **2012,** 84, (20), 8442-7.

36. Fu, Y. Q.; Zhao, C. X.; Lu, X.; Xu, G. W., Nontargeted screening of chemical contaminants and illegal additives in food based on liquid chromatography-high resolution mass spectrometry. *Trac-Trends in Analytical Chemistry* **2017,** 96, 89-98.

37. Garcia, A.; Barbas, C., Gas chromatography-mass spectrometry (GC-MS)-based metabolomics. *Methods Mol Biol* **2011,** 708, 191-204.

38. Forcisi, S.; Moritz, F.; Kanawati, B.; Tziotis, D.; Lehmann, R.; Schmitt-Kopplin, P., Liquid chromatography-mass spectrometry in metabolomics research: mass analyzers in ultra high pressure liquid chromatography coupling. *J Chromatogr A* **2013,** 1292, 51-65.

39. Jonsson, P.; Johansson, A. I.; Gullberg, J.; Trygg, J.; A, J.; Grung, B.; Marklund, S.; Sjostrom, M.; Antti, H.; Moritz, T., High-throughput data analysis for detecting and identifying differences between samples in GC/MS-based metabolomic analyses. *Analytical Chemistry* **2005,** 77, (17), 5635-5642.

40. Theodoridis, G.; Gika, H. G.; Wilson, I. D., LC-MS-based methodology for global metabolite profiling in metabonomics/metabolomics. *Trac-Trends in Analytical Chemistry* **2008,** 27, (3), 251-260.

41. Ramautar, R.; Somsen, G. W.; de Jong, G. J., CE-MS for metabolomics: Developments and applications in the period 2010-2012. *Electrophoresis* **2013,** 34, (1), 86-98.

42. Wilson, I. D.; Nicholson, J. K.; Castro-Perez, J.; Granger, J. H.; Johnson, K. A.; Smith, B. W.; Plumb, R. S., High resolution "ultra performance" liquid chromatography coupled to oa-TOF mass spectrometry as a tool for differential metabolic pathway profiling in functional genomic studies. *J Proteome Res* **2005,** 4, (2), 591-8.

43. Rathahao-Paris, E.; Alves, S.; Junot, C.; Tabet, J. C., High resolution mass spectrometry for structural identification of metabolites in metabolomics. *Metabolomics* **2016,** 12, (1).

44. Beckmann, M.; Parker, D.; Enot, D. P.; Duval, E.; Draper, J., High-throughput, nontargeted metabolite fingerprinting using nominal mass flow injection electrospray mass spectrometry. *Nature Protocols* **2008,** 3, (3), 486-504.

45. Castrillo, J. I.; Hayes, A.; Mohammed, S.; Gaskell, S. J.; Oliver, S. G., An optimized protocol for metabolome analysis in yeast using direct infusion electrospray mass spectrometry. *Phytochemistry* **2003,** 62, (6), 929-937.

46. Southam, A. D.; Weber, R. J. M.; Engel, J.; Jones, M. R.; Viant, M. R., A complete workflow for high-resolution spectral-stitching nanoelectrospray direct-infusion mass-spectrometry-based metabolomics and lipidomics. *Nature Protocols* **2017,** 12, (2), 310-328.

47. Paglia, G.; Williams, J. P.; Menikarachchi, L.; Thompson, J. W.; Tyldesley-Worster, R.; Halldorsson, S.; Rolfsson, O.; Moseley, A.; Grant, D.; Langridge, J.; Palsson, B. O.; Astarita, G., Ion Mobility Derived Collision Cross Sections to Support Metabolomics Applications. *Analytical Chemistry* **2014,** 86, (8), 3985-3993.

48. Kurulugama, R. T.; Valentine, S. J.; Sowell, R. A.; Clemmer, D. E., Development of a high-throughput IMS-IMS-MS approach for analyzing mixtures of biomolecules. *Journal of Proteomics* **2008,** 71, (3), 318-331.

49. Paglia, G.; Angel, P.; Williams, J. P.; Richardson, K.; Olivos, H. J.; Thompson, J. W.; Menikarachchi, L.; Lai, S.; Walsh, C.; Moseley, A.; Plumb, R. S.; Grant, D. F.; Palsson, B. O.; Langridge, J.; Geromanos, S.; Astarite, G., Ion Mobility-Derived Collision Cross Section As an Additional Measure for Lipid Fingerprinting and Identification. *Analytical Chemistry* **2015,** 87, (2), 1137-1144.

50. Dwivedi, P.; Wu, P.; Klopsch, S. J.; Puzon, G. J.; Xun, L.; Hill, H. H., Metabolic profiling by ion mobility mass spectrometry (IMMS). *Metabolomics* **2008,** 4, (1), 63-80.

51. Kaplan, K.; Dwivedi, P.; Davidson, S.; Yang, Q.; Tso, P.; Siems, W.; Hill, H. H., Monitoring Dynamic Changes in Lymph Metabolome of Fasting and Fed Rats by Electrospray Ionization-Ion Mobility Mass Spectrometry (ESI-IMMS). *Analytical Chemistry* **2009,** 81, (19), 7944-7953.

52. Paglia, G.; Astarita, G., Metabolomics and lipidomics using traveling-wave ion mobility mass spectrometry. *Nature Protocols* **2017,** 12, (4), 797-813.

53. Zhang, X.; Kew, K.; Reisdorph, R.; Sartain, M.; Powell, R.; Armstrong, M.; Quinn, K.; Cruickshank-Quinn, C.; Walmsley, S.; Bokatzian, S.; Darland, E.; Rain, M.; Imatani, K.; Reisdorph, N., Performance of a High-Pressure Liquid Chromatography-Ion Mobility-Mass Spectrometry System for Metabolic Profiling. *Analytical Chemistry* **2017,** 89, (12), 6384-6391.

54. Shvartsburg, A. A.; Smith, R. D., Fundamentals of Traveling Wave Ion Mobility Spectrometry. *Analytical Chemistry* **2008,** 80, (24), 9689-9699.

55. May, J. C.; McLean, J. A., Ion Mobility-Mass Spectrometry: Time-Dispersive Instrumentation. *Analytical Chemistry* **2015,** 87, (3), 1422-1436.

56. Matsuda, F., Technical Challenges in Mass Spectrometry-Based Metabolomics. *Mass Spectrom (Tokyo)* **2016,** 5, (2), S0052.

57. Broadhurst, D.; Goodacre, R.; Reinke, S. N.; Kuligowski, J.; Wilson, I. D.; Lewis, M. R.; Dunn, W. B., Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. *Metabolomics* **2018,** 14, (6).

58. Schymanski, E. L.; Ruttkies, C.; Krauss, M.; Brouard, C.; Kind, T.; Duhrkop, K.; Allen, F.; Vaniya, A.; Verdegem, D.; Bocker, S.; Rousu, J.; Shen, H.; Tsugawa, H.;

Sajed, T.; Fiehn, O.; Ghesquiere, B.; Neumann, S., Critical Assessment of Small Molecule Identification 2016: automated methods. *J Cheminform* **2017,** 9, (1), 22.

59. Kind, T.; Fiehn, O., Advances in structure elucidation of small molecules using mass spectrometry. *Bioanal Rev* **2010,** 2, (1-4), 23-60.

60. Wanichthanarak, K.; Fan, S.; Grapov, D.; Barupal, D. K.; Fiehn, O., Metabox: A Toolbox for Metabolomic Data Analysis, Interpretation and Integrative Exploration. *PLoS One* **2017,** 12, (1), e0171046.

61. Chong, J.; Soufan, O.; Li, C.; Caraus, I.; Li, S.; Bourque, G.; Wishart, D. S.; Xia, J., MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res* **2018**.

62. Slenter, D. N.; Kutmon, M.; Hanspers, K.; Riutta, A.; Windsor, J.; Nunes, N.; Melius, J.; Cirillo, E.; Coort, S. L.; Digles, D.; Ehrhart, F.; Giesbertz, P.; Kalafati, M.; Martens, M.; Miller, R.; Nishida, K.; Rieswijk, L.; Waagmeester, A.; Eijssen, L. M. T.; Evelo, C. T.; Pico, A. R.; Willighagen, E. L., WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res* **2018,** 46, (D1), D661-D667.

63. Forsberg, E. M.; Huan, T.; Rinehart, D.; Benton, H. P.; Warth, B.; Hilmers, B.; Siuzdak, G., Data processing, multi-omic pathway mapping, and metabolite activity analysis using XCMS Online. *Nat Protoc* **2018,** 13, (4), 633-651.

64. Taylor, S. I., Deconstructing type 2 diabetes. *Cell* **1999,** 97, (1), 9-12.

65. Chatterjee, S.; Khunti, K.; Davies, M. J., Type 2 diabetes. *Lancet* **2017,** 389, (10085), 2239-2251.

66. Knowler, W. C.; Barrett-Connor, E.; Fowler, S. E.; Hamman, R. F.; Lachin, J. M.; Walker, E. A.; Nathan, D. M.; G, D. P. P. R., Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *New England Journal of Medicine* **2002,** 346, (6), 393-403.

67. de Mello, V. D.; Paananen, J.; Lindstrom, J.; Lankinen, M. A.; Shi, L.; Kuusisto, J.; Pihlajamaki, J.; Auriola, S.; Lehtonen, M.; Rolandsson, O.; Bergdahl, I. A.; Nordin, E.; Ilanne-Parikka, P.; Keinanen-Kiukaanniemi, S.; Landberg, R.; Eriksson, J. G.; Tuomilehto, J.; Hanhineva, K.; Uusitupa, M., Indolepropionic acid and novel lipid

metabolites are associated with a lower risk of type 2 diabetes in the Finnish Diabetes Prevention Study. *Sci Rep* **2017,** 7, 46337.

68. Peddinti, G.; Cobb, J.; Yengo, L.; Froguel, P.; Kravic, J.; Balkau, B.; Tuomi, T.; Aittokallio, T.; Groop, L., Early metabolic markers identify potential targets for the prevention of type 2 diabetes. *Diabetologia* **2017,** 60, (9), 1740-1750.

69. Fall, T.; Salihovic, S.; Brandmaier, S.; Nowak, C.; Ganna, A.; Gustafsson, S.; Broeckling, C. D.; Prenni, J. E.; Kastenmuller, G.; Peters, A.; Magnusson, P. K.; Wang-Sattler, R.; Giedraitis, V.; Berne, C.; Gieger, C.; Pedersen, N. L.; Ingelsson, E.; Lind, L., Non-targeted metabolomics combined with genetic analyses identifies bile acid synthesis and phospholipid metabolism as being associated with incident type 2 diabetes. *Diabetologia* **2016,** 59, (10), 2114-24.

70. Andrianou, X. D.; Charisiadis, P.; Makris, K. C., Coupling Urinary Trihalomethanes and Metabolomic Profiles of Type II Diabetes: A Case-Control Study. *Journal of Proteome Research* **2017,** 16, (8), 2743-2751.

71. Tam, Z. Y.; Ng, S. P.; Tan, L. Q.; Lin, C. H.; Rothenbacher, D.; Klenk, J.; Boehm, B. O.; Team, S.; Grp, A. S., Metabolite profiling in identifying metabolic biomarkers in older people with late-onset type 2 diabetes mellitus. *Scientific Reports* **2017,** 7.

72. Knebel, B.; Mack, S.; Lehr, S.; Barsch, A.; Schiller, M.; Haas, J.; Lange, S.; Fuchser, J.; Zurek, G.; Muller-Wieland, D.; Kotzka, J., Untargeted mass spectrometric approach in metabolic healthy offspring of patients with type 2 diabetes reveals medium-chain acylcarnitine as potential biomarker for lipid induced glucose intolerance (LGIT). *Archives of Physiology and Biochemistry* **2016,** 122, (5), 266-280.

73. Lu, Y. H.; Wang, Y. L.; Ong, C. N.; Subramaniam, T.; Choi, H. W.; Yuan, J. M.; Koh, W. P.; Pan, A., Metabolic signatures and risk of type 2 diabetes in a Chinese population: an untargeted metabolomics study using both LC-MS and GC-MS. *Diabetologia* **2016,** 59, (11), 2349-2359.

74. Galderisi, A.; Pirillo, P.; Moret, V.; Stocchero, M.; Gucciardi, A.; Perilongo, G.; Moretti, C.; Monciotti, C.; Giordano, G.; Baraldi, E., Metabolomics reveals new metabolic perturbations in children with type 1 diabetes. *Pediatr Diabetes* **2018,** 19, (1), 59-67.

75. Law, K. P.; Mao, X.; Han, T. L.; Zhang, H., Unsaturated plasma phospholipids are consistently lower in the patients diagnosed with gestational diabetes mellitus

throughout pregnancy: A longitudinal metabolomics study of Chinese pregnant women part 1. *Clinica Chimica Acta* **2017,** 465, 53-71.

76. Gong, Z. G.; Zhao, W.; Zhang, J.; Wu, X.; Hu, J.; Yin, G. C.; Xu, Y. J., Metabolomics and eicosanoid analysis identified serum biomarkers for distinguishing hepatocellular carcinoma from hepatitis B virus-related cirrhosis. *Oncotarget* **2017,** 8, (38), 63890-63900.

77. Di Poto, C.; Ferrarini, A.; Zhao, Y.; Varghese, R. S.; Tu, C.; Zuo, Y.; Wang, M.; Nezami Ranjbar, M. R.; Luo, Y.; Zhang, C.; Desai, C. S.; Shetty, K.; Tadesse, M. G.; Ressom, H. W., Metabolomic Characterization of Hepatocellular Carcinoma in Patients with Liver Cirrhosis for Biomarker Discovery. *Cancer Epidemiol Biomarkers Prev* **2017,** 26, (5), 675-683.

78. Tan, W. F.; He, J. Q.; Deng, J. L.; Yang, X. W.; Cui, L. J.; Ran, R. Z.; Du, G. W.; Jiang, X. Q., Small molecule metabolite biomarkers for hepatocellular carcinoma with bile duct tumor thrombus diagnosis. *Scientific Reports* **2018,** 8.

79. Tibshirani, R., Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* **1996,** 58, (1), 267-288.

80. Siegel, R. L.; Miller, K. D.; Jemal, A., Cancer statistics, 2018. *CA Cancer J Clin* **2018,** 68, (1), 7-30.

81. Do Canto, L. M.; Marian, C.; Varghese, R. S.; Ahn, J.; Da Cunha, P. A.; Willey, S.; Sidawy, M.; Rone, J. D.; Cheema, A. K.; Luta, G.; Ranjbar, M. R. N.; Ressom, H. W.; Haddad, B. R., Metabolomic profiling of breast tumors using ductal fluid. *International Journal of Oncology* **2016,** 49, (6), 2245-2254.

82. Breast Cancer Survival Rates. https://www.cancer.org/cancer/breast-cancer/understanding-a-breast-cancer-diagnosis/breast-cancer-survival-rates.html (July 7th).

83. Hadi, N. I.; Jamal, Q.; Iqbal, A.; Shaikh, F.; Somroo, S.; Musharraf, S. G., Serum Metabolomic Profiles for Breast Cancer Diagnosis, Grading and Staging by Gas Chromatography-Mass Spectrometry. *Scientific Reports* **2017,** 7.

84. Roig, B.; Rodriguez-Balada, M.; Samino, S.; Lam, E. W. F.; Guaita-Esteruelas, S.; Gomes, A. R.; Correig, X.; Borras, J.; Yanes, O.; Guma, J., Metabolomics reveals

novel blood plasma biomarkers associated to the BRCA1-mutated phenotype of human breast cancer. *Scientific Reports* **2017,** 7.

85. Zhong, L. P.; Cheng, F.; Lu, X. Y.; Duan, Y. X.; Wang, X. D., Untargeted saliva metabonomics study of breast cancer based on ultra performance liquid chromatography coupled to mass spectrometry with HILIC and RPLC separations. *Talanta* **2016,** 158, 351-360.

86. Pagidipati, N. J.; Gaziano, T. A., Estimating Deaths From Cardiovascular Disease: A Review of Global Methodologies of Mortality Measurement. *Circulation* **2013,** 127, (6), 749-756.

87. Benjamin, E. J.; Virani, S. S.; Callaway, C. W.; Chamberlain, A. M.; Chang, A. R.; Cheng, S.; Chiuve, S. E.; Cushman, M.; Delling, F. N.; Deo, R.; de Ferranti, S. D.; Ferguson, J. F.; Fornage, M.; Gillespie, C.; Isasi, C. R.; Jimenez, M. C.; Jordan, L. C.; Judd, S. E.; Lackland, D.; Lichtman, J. H.; Lisabeth, L.; Liu, S. M.; Longenecker, C. T.; Lutsey, P. L.; Mackey, J. S.; Matchar, D. B.; Matsushita, K.; Mussolino, M. E.; Nasir, K.; O'Flaherty, M.; Palaniappan, L. P.; Pandey, A.; Pandey, D. K.; Reeves, M. J.; Ritchey, M. D.; Rodriguez, C. J.; Roth, G. A.; Rosamond, W. D.; Sampson, U. K. A.; Satou, G. M.; Shah, S. H.; Spartano, N. L.; Tirschwell, D. L.; Tsao, C. W.; Voeks, J. H.; Willey, J. Z.; Wilkins, J. T.; Wu, J. H. Y.; Alger, H. M.; Wong, S. S.; Muntner, P.; Epidemi, A. H. A. C., Heart Disease and Stroke Statistics-2018 Update A Report From the American Heart Association. *Circulation* **2018,** 137, (12), E67-E492.

88. Li, X. M. S.; Wang, Z. N.; Cajka, T.; Buffa, J. A.; Nemet, I.; Hurd, A. G.; Gu, X. D.; Skye, S. M.; Roberts, A. B.; Wu, Y. P.; Li, L.; Shahen, C. J.; Wagner, M. A.; Hartiala, J. A.; Kerby, R. L.; Romano, K. A.; Han, Y.; Obeid, S.; Luscher, T. F.; Allayee, H.; Rey, F. E.; DiDonato, J. A.; Fiehn, O.; Tang, W. H. W.; Hazen, S. L., Untargeted metabolomics identifies trimethyllysine, a TMAO-producing nutrient precursor, as a predictor of incident cardiovascular disease risk. *Jci Insight* **2018,** 3, (6).

89. Djekic, D.; Pinto, R.; Vorkas, P. A.; Henein, M. Y., Replication of LC-MS untargeted lipidomics results in patients with calcific coronary disease: An interlaboratory reproducibility study. *International Journal of Cardiology* **2016,** 222, 1042-1048.

90. Xu, X. B.; Gao, B. B.; Guan, Q. J.; Zhang, D. D.; Ye, X. H.; Zhou, L.; Tong, G. X.; Li, H.; Zhang, L.; Tian, J. K.; Huang, J. Y., Metabolomic profile for the early detection of coronary artery disease by using UPLC-QTOF/MS. *Journal of Pharmaceutical and Biomedical Analysis* **2016,** 129, 34-42.

91. Li, R. J.; Li, F. Y.; Feng, Q.; Liu, Z. P.; Jie, Z. Y.; Wen, B.; Xu, X.; Zhong, S. L.; Li, G. L.; He, K. L., An LC-MS based untargeted metabolomics study identified novel biomarkers for coronary heart disease. *Molecular Biosystems* **2016,** 12, (11), 3425-3434.

92. Liang, Q.; Liu, H.; Zhang, T. Y.; Jiang, Y.; Zhang, A. H., Untargeted lipidomics study of coronary artery disease by FUPLC-Q-TOF-MS. *Analytical Methods* **2016,** 8, (6), 1229-1234.

93. Trezzi, J. P.; Galozzi, S.; Jaeger, C.; Barkovits, K.; Brockmann, K.; Maetzler, W.; Berg, D.; Marcus, K.; Betsou, F.; Hiller, K.; Mollenhauer, B., Distinct Metabolomic Signature in Cerebrospinal Fluid in Early Parkinson's Disease. *Movement Disorders* **2017,** 32, (10), 1401-1408.

94. Stoessel, D.; Schulte, C.; dos Santos, M. C. T.; Scheller, D.; Rebollo-Mesa, I.; Deuschle, C.; Walther, D.; Schauer, N.; Berg, D.; da Costa, A. N.; Maetzler, W., Promising Metabolite Profiles in the Plasma and CSF of Early Clinical Parkinson's Disease. *Frontiers in Aging Neuroscience* **2018,** 10.

95. Schwartz, J., Particulate air pollution and chronic respiratory disease. *Environ Res* **1993,** 62, (1), 7-13.

96. Kim, W. J.; Lee, C. Y., Environmental exposures and chronic obstructive pulmonary disease. *Molecular & Cellular Toxicology* **2017,** 13, (3), 251-255.

97. Le Jemtel, T. H.; Padeletti, M.; Jelic, S., Diagnostic and therapeutic challenges in patients with coexistent chronic obstructive pulmonary disease and chronic heart failure. *J Am Coll Cardiol* **2007,** 49, (2), 171-80.

98. Shahab, L.; Jarvis, M. J.; Britton, J.; West, R., Prevalence, diagnosis and relation to tobacco dependence of chronic obstructive pulmonary disease in a nationally representative population sample. *Thorax* **2006,** 61, (12), 1043-7.

99. Prentice, B.; Hameed, S.; Verge, C. F.; Ooi, C. Y.; Jaffe, A.; Widger, J., Diagnosing cystic fibrosis-related diabetes: current methods and challenges. *Expert Rev Respir Med* **2016,** 10, (7), 799-811.

100. Henrickson, K. J., Advances in the laboratory diagnosis of viral respiratory disease. *Pediatr Infect Dis J* **2004,** 23, (1 Suppl), S6-10.

101. Ciofu, O.; Hansen, C. R.; Hoiby, N., Respiratory bacterial infections in cystic fibrosis. *Curr Opin Pulm Med* **2013,** 19, (3), 251-8.

102. Gaisl, T.; Bregy, L.; Stebler, N.; Gaugg, M. T.; Bruderer, T.; Garcia-Gomez, D.; Moeller, A.; Singer, F.; Schwarz, E. I.; Benden, C.; P, M. L. S.; Zenobi, R.; Kohler, M., Real-time exhaled breath analysis in patients with cystic fibrosis and controls. *J Breath Res* **2018,** 12, (3), 036013.

103. Malkar, A.; Wilson, E.; Harrrison, T.; Shaw, D.; Creaser, C., Untargeted metabolic profiling of saliva by liquid chromatography-mass spectrometry for the identification of potential diagnostic biomarkers of asthma. *Analytical Methods* **2016,** 8, (27), 5407-5413.

104. Rindlisbacher, B.; Strebel, C.; Guler, S.; Kollar, A.; Geiser, T.; Martin Fiedler, G.; Benedikt Leichtle, A.; Bovet, C.; Funke-Chambour, M., Exhaled breath condensate as a potential biomarker tool for idiopathic pulmonary fibrosis-a pilot study. *J Breath Res* **2017,** 12, (1), 016003.

105. Rindlisbacher, B.; Schmid, C.; Geiser, T.; Bovet, C.; Funke-Chambour, M., Serum metabolic profiling identified a distinct metabolic signature in patients with idiopathic pulmonary fibrosis - a potential biomarker role for LysoPC. *Respiratory Research* **2018,** 19.

106. Liang, Q.; Liu, H.; Xie, L. X.; Li, X.; Zhang, A. H., High-throughput metabolomics enables biomarker discovery in prostate cancer. *Rsc Advances* **2017,** 7, (5), 2587-2593.

107. McDunn, J. E.; Li, Z.; Adam, K. P.; Neri, B. P.; Wolfert, R. L.; Milburn, M. V.; Lotan, Y.; Wheeler, T. M., Metabolomic signatures of aggressive prostate cancer. *Prostate* **2013,** 73, (14), 1547-60.

108. Wang, X. D.; Han, J.; Hardie, D. B.; Yang, J. C.; Pan, J. X.; Borchers, C. H., Metabolomic profiling of prostate cancer by matrix assisted laser desorption/ionization-Fourier transform ion cyclotron resonance mass spectrometry imaging using Matrix Coating Assisted by an Electric Field (MCAEF). *Biochimica Et Biophysica Acta-Proteins and Proteomics* **2017,** 1865, (7), 755-767.

109. Lima, A. R.; Araujo, A. M.; Pinto, J.; Jeronimo, C.; Henrique, R.; Bastos, M. D.; Carvalho, M.; de Pinho, P. G., Discrimination between the human prostate normal and cancer cell exometabolome by GC-MS. *Scientific Reports* **2018,** 8.

110. Li, X.; Wu, T.; Jiang, Y.; Zhang, Z.; Han, X.; Geng, W.; Ding, H.; Kang, J.; Wang, Q.; Shang, H., Plasma metabolic changes in Chinese HIV-infected patients receiving lopinavir/ritonavir based treatment: Implications for HIV precision therapy. *Cytokine* **2018,** 110, 204-212.

111. Stockard, B.; Garrett, T.; Guingab-Cagmat, J.; Meshinchi, S.; Lamba, J., Distinct Metabolic features differentiating FLT3-ITD AML from FLT3-WT childhood Acute Myeloid Leukemia. *Scientific Reports* **2018,** 8.

112. Musharraf, S. G.; Iqbal, A.; Ansari, S. H.; Parveen, S.; Khan, I. A.; Siddiqui, A. J., beta-Thalassemia Patients Revealed a Significant Change of Untargeted Metabolites in Comparison to Healthy Individuals. *Scientific Reports* **2017,** 7.

113. Gao, P.; Zhou, C.; Zhao, L.; Zhang, G.; Zhang, Y., Tissue amino acid profile could be used to differentiate advanced adenoma from colorectal cancer. *J Pharm Biomed Anal* **2016,** 118, 349-355.

114. Audet-Delage, Y.; Villeneuve, L.; Gregoire, J.; Plante, M.; Guillemette, C., Identification of Metabolomic Biomarkers for Endometrial Cancer and Its Recurrence after Surgery in Postmenopausal Women. *Frontiers in Endocrinology* **2018,** 9.

115. Peces, S. R.; Navarro, C. D.; Lopez, C. M.; Caba, O.; Jimenez-Luna, C.; Melguizo, C.; Prados, J. C.; Genilloud, O.; Perez, F. V.; del Palacio, J. P., Untargeted LC-HRMS-Based Metabolomics for Searching New Biomarkers of Pancreatic Ductal Adenocarcinoma: A Pilot Study. *Slas Discovery* **2017,** 22, (4), 348-359.

116. Jin, R.; Banton, S.; Tran, V. T.; Konomi, J. V.; Li, S. Z.; Jones, D. P.; Vos, M. B., Amino Acid Metabolism is Altered in Adolescents with Nonalcoholic Fatty Liver Disease-An Untargeted, High Resolution Metabolomics Study. *Journal of Pediatrics* **2016,** 172, 14-+.

117. Lee, J. E.; Lee, Y. H.; Kim, S. Y.; Kim, Y. G.; Moon, J. Y.; Jeong, K. H.; Lee, T. W.; Ihm, C. G.; Kim, S.; Kim, K. H.; Kim, D. K.; Kim, Y. S.; Kim, C. D.; Park, C. W.; Lee, D. Y.; Lee, S. H., Systematic biomarker discovery and coordinative validation for different primary nephrotic syndromes using gas chromatography-mass spectrometry. *Journal of Chromatography A* **2016,** 1453, 105-115.

118. Ning, P.; Zheng, Y. L.; Luo, Q. Z.; Liu, X. H.; Kang, Y.; Zhang, Y.; Zhang, R. B.; Xu, Y.; Yang, D. H.; Xi, W.; Wang, K. Q.; Chen, Y. S.; An, S. C.; Gao, Z. C., Metabolic profiles in community-acquired pneumonia: developing assessment tools for disease severity. *Critical Care* **2018,** 22.

119. Scoville, E. A.; Allaman, M. M.; Brown, C. T.; Motley, A. K.; Horst, S. N.; Williams, C. S.; Koyama, T.; Zhao, Z. G.; Adams, D. W.; Beaulieu, D. B.; Schwartz, D. A.; Wilson, K. T.; Coburn, L. A., Alterations in lipid, amino acid, and energy metabolism distinguish Crohn's disease from ulcerative colitis and control subjects by serum metabolomic profiling. *Metabolomics* **2018,** 14, (1).

120. Abela, L.; Simmons, L.; Steindl, K.; Schmitt, B.; Mastrangelo, M.; Joset, P.; Papuc, M.; Sticht, H.; Baumer, A.; Crowther, L. M.; Mathis, D.; Rauch, A.; Plecko, B., N-8-acetylspermidine as a potential plasma biomarker for Snyder-Robinson syndrome identified by clinical metabolomics. *Journal of Inherited Metabolic Disease* **2016,** 39, (1), 131-137.

121. Fermier, B.; Blasco, H.; Godat, E.; Bocca, C.; Moenne-Loccoz, J.; Emond, P.; Andres, C. R.; Laffon, M.; Ferrandiere, M., Specific Metabolome Profile of Exhaled Breath Condensate in Patients with Shock and Respiratory Failure: A Pilot Study. *Metabolites* **2016,** 6, (3).

122. Klupczynska, A.; Derezinski, P.; Garrett, T. J.; Rubio, V. Y.; Dyszkiewicz, W.; Kasprzyk, M.; Kokot, Z. J., Study of early stage non-small-cell lung cancer using Orbitrap-based global serum metabolomics. *J Cancer Res Clin Oncol* **2017,** 143, (4), 649-659.

123. Fernandez-Peralbo, M. A.; de Castro, M. D. L., Preparation of urine samples prior to targeted or untargeted metabolomics mass-spectrometry analysis. *Trac-Trends in Analytical Chemistry* **2012,** 41, 75-85.

124. Pasikanti, K. K.; Ho, P. C.; Chan, E. C. Y., Gas chromatography/mass spectrometry in metabolic profiling of biological fluids. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* **2008,** 871, (2), 202-211.

125. Liu, R. X.; Hong, J.; Xu, X. Q.; Feng, Q.; Zhang, D. Y.; Gu, Y. Y.; Shi, J.; Zhao, S. Q.; Liu, W.; Wang, X. K.; Xia, H. H.; Liu, Z. P.; Cui, B.; Liang, P. W.; Xi, L. Q.; Jin, J. B.; Ying, X. Y.; Wang, X. L.; Zhao, X. J.; Li, W. Y.; Jia, H. J.; Lan, Z.; Li, F. Y.; Wang, R.; Sun, Y. K.; Yang, M. L.; Shen, Y. X.; Jie, Z. Y.; Li, J. H.; Chen, X. M.; Zhong, H. Z.; Xie, H. L.; Zhang, Y. F.; Gu, W. Q.; Deng, X. X.; Shen, B. Y.; Xu, X.; Yang, H. M.; Xu, G. W.; Bi, Y. F.; Lai, S. H.; Wang, J.; Qi, L.; Madsen, L.; Wang, J. Q.; Ning, G.; Kristiansen, K.; Wang, W. Q., Gut microbiome and serum metabolome alterations in obesity and after weight-loss intervention. *Nature Medicine* **2017,** 23, (7), 859-+.

126. Park, H. M.; Park, K. T.; Park, E. C.; Kim, S. I.; Choi, M. S.; Liu, K. H.; Lee, C. H., Mass Spectrometry-Based Metabolomic and Lipidomic Analyses of the Effects of Dietary Platycodon grandiflorum on Liver and Serum of Obese Mice under a High-Fat Diet. *Nutrients* **2017,** 9, (1).

127. Pedersen, H. K.; Gudmundsdottir, V.; Nielsen, H. B.; Hyotylainen, T.; Nielsen, T.; Jensen, B. A. H.; Forslund, K.; Hildebrand, F.; Prifti, E.; Falony, G.; Le Chatelier, E.; Levenez, F.; Dore, J.; Mattila, I.; Plichta, D. R.; Poho, P.; Hellgren, L. I.; Arumugam, M.; Sunagawa, S.; Vieira-Silva, S.; Jorgensen, T.; Holm, J. B.; Trost, K.; Kristiansen, K.; Brix, S.; Raes, J.; Wang, J.; Hansen, T.; Bork, P.; Brunak, S.; Oresic, M.; Ehrlich, S. D.; Pedersen, O.; Consortium, M., Human gut microbes impact host serum metabolome and insulin sensitivity. *Nature* **2016,** 535, (7612), 376-+.

128. Sun, X. M.; Huang, L.; Zhang, R.; Xu, W.; Huang, J. Y.; Gurav, D. D.; Vedarethinam, V.; Chen, R. P.; Lou, J. T.; Wang, Q.; Wan, J. J.; Qian, K., Metabolic Fingerprinting on a Plasmonic Gold Chip for Mass Spectrometry Based in Vitro Diagnostics. *Acs Central Science* **2018,** 4, (2), 223-229.

129. Zhang, X.; Romm, M.; Zheng, X.; Zink, E. M.; Kim, Y. M.; Burnum-Johnson, K. E.; Orton, D. J.; Apffel, A.; Ibrahim, Y. M.; Monroe, M. E.; Moore, R. J.; Smith, J. N.; Ma, J.; Renslow, R. S.; Thomas, D. G.; Blackwell, A. E.; Swinford, G.; Sausen, J.; Kurulugama, R. T.; Eno, N.; Darland, E.; Stafford, G.; Fjeldsted, J.; Metz, T. O.; Teeguarden, J. G.; Smith, R. D.; Baker, E. S., SPE-IMS-MS: An automated platform for sub-sixty second surveillance of endogenous metabolites and xenobiotics in biofluids. *Clin Mass Spectrom* **2016,** 2, 1-10.

130. Zang, X.; Monge, M. E.; McCarty, N. A.; Stecenko, A. A.; Fernandez, F. M., Feasibility of Early Detection of Cystic Fibrosis Acute Pulmonary Exacerbations by Exhaled Breath Condensate Metabolomics: A Pilot Study. *J Proteome Res* **2017,** 16, (2), 550-558.

131. Zang, X.; Perez, J. J.; Jones, C. M.; Monge, M. E.; McCarty, N. A.; Stecenko, A. A.; Fernandez, F. M., Comparison of Ambient and Atmospheric Pressure Ion Sources for Cystic Fibrosis Exhaled Breath Condensate Ion Mobility-Mass Spectrometry Metabolomics. *J Am Soc Mass Spectrom* **2017,** 28, (8), 1489-1496.

132. Phelps, D. L.; Balog, J.; Gildea, L. F.; Bodai, Z.; Savage, A.; El-Bahrawy, M. A.; Speller, A. V.; Rosini, F.; Kudo, H.; McKenzie, J. S.; Brown, R.; Takats, Z.; Ghaem-Maghami, S., The surgical intelligent knife distinguishes normal, borderline and malignant gynaecological tissues using rapid evaporative ionisation mass spectrometry (REIMS). *Br J Cancer* **2018,** 118, (10), 1349-1358.

133. Jarmusch, A. K.; Pirro, V.; Baird, Z.; Hattab, E. M.; Cohen-Gadol, A. A.; Cooks, R. G., Lipid and metabolite profiles of human brain tumors by desorption electrospray ionization-MS. *Proc Natl Acad Sci U S A* **2016,** 113, (6), 1486-91.

134. Banerjee, S.; Zare, R. N.; Tibshirani, R. J.; Kunder, C. A.; Nolley, R.; Fan, R.; Brooks, J. D.; Sonn, G. A., Diagnosis of prostate cancer by desorption electrospray ionization mass spectrometric imaging of small metabolites and lipids. *Proceedings of the National Academy of Sciences of the United States of America* **2017,** 114, (13), 3334-3339.

135. Zhang, R.; Hu, P. S.; Zang, Q. C.; Yue, X. F.; Zhou, Z.; Xu, X. Y.; Xu, J.; Li, S. S.; Chen, Y. H.; Qiang, B. Q.; Peng, X. Z.; Han, W.; Zhang, R. P.; Abliz, Z., LC-MS-based metabolomics reveals metabolic signatures related to glioma stem-like cell self-renewal and differentiation. *Rsc Advances* **2017,** 7, (39), 24221-24232.

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

137. Gagnebin, Y.; Tonoli, D.; Lescuyer, P.; Ponte, B.; de Seigneux, S.; Martin, P. Y.; Schappler, J.; Boccard, J.; Rudaz, S., Metabolomic analysis of urine samples by UHPLC-QTOF-MS: Impact of normalization strategies. *Anal Chim Acta* **2017,** 955, 27-35.

138. Rainville, P. D.; Wilson, I. D.; Nicholson, J. K.; Isaac, G.; Mullin, L.; Langridge, J. I.; Plumb, R. S., Ion mobility spectrometry combined with ultra performance liquid chromatography/mass spectrometry for metabolic phenotyping of urine: Effects of column length, gradient duration and ion mobility spectrometry on metabolite detection. *Anal Chim Acta* **2017,** 982, 1-8.

139. Hunt, J., Exhaled breath condensate: an evolving tool for noninvasive evaluation of lung disease. *J Allergy Clin Immunol* **2002,** 110, (1), 28-34.

140. Mutlu, G. M.; Garey, K. W.; Robbins, R. A.; Danziger, L. H.; Rubinstein, I., Collection and analysis of exhaled breath condensate in humans. *Am J Respir Crit Care Med* **2001,** 164, (5), 731-7.

141. Montuschi, P.; Barnes, P. J., Analysis of exhaled breath condensate for monitoring airway inflammation. *Trends Pharmacol Sci* **2002,** 23, (5), 232-7.

142. Goldoni, M.; Corradi, M.; Mozzoni, P.; Folesani, G.; Alinovi, R.; Pinelli, S.; Andreoli, R.; Pigini, D.; Tillo, R.; Filetti, A.; Garavelli, C.; Mutti, A., Concentration of exhaled breath condensate biomarkers after fractionated collection based on exhaled CO2 signal. *J Breath Res* **2013,** 7, (1), 017101.

143. Fernandez-Peralbo, M. A.; Santiago, M. C.; Priego-Capote, F.; de Castro, M. D. L., Study of exhaled breath condensate sample preparation for metabolomics analysis by LC-MS/MS in high resolution mode. *Talanta* **2015,** 144, 1360-1369.

144. Wang, X. D.; Han, J.; Hardie, D. B.; Yang, J. C.; Borchers, C. H., The use of matrix coating assisted by an electric field (MCAEF) to enhance mass spectrometric imaging of human prostate cancer biomarkers. *Journal of Mass Spectrometry* **2016,** 51, (1), 86-95.

145. Guenther, S.; Muirhead, L. J.; Speller, A. V.; Golf, O.; Strittmatter, N.; Ramakrishnan, R.; Goldin, R. D.; Jones, E.; Veselkov, K.; Nicholson, J.; Darzi, A.; Takats, Z., Spatially resolved metabolic phenotyping of breast cancer by desorption electrospray ionization mass spectrometry. *Cancer Res* **2015,** 75, (9), 1828-37.

146. Dong, Y.; Li, B.; Malitsky, S.; Rogachev, I.; Aharoni, A.; Kaftan, F.; Svatos, A.; Franceschi, P., Sample Preparation for Mass Spectrometry Imaging of Plant Tissues: A Review. *Front Plant Sci* **2016,** 7, 60.

147. Cuperlovic-Culf, M.; Barnett, D. A.; Culf, A. S.; Chute, I., Cell culture metabolomics: applications and future directions. *Drug Discovery Today* **2010,** 15, (15-16), 610-621.

148. Martano, G.; Delmotte, N.; Kiefer, P.; Christen, P.; Kentner, D.; Bumann, D.; Vorholt, J. A., Fast sampling method for mammalian cell metabolic analyses using liquid chromatography-mass spectrometry. *Nat Protoc* **2015,** 10, (1), 1-11.

149. Ali, M. R.; Wu, Y.; Han, T.; Zang, X.; Xiao, H.; Tang, Y.; Wu, R.; Fernandez, F. M.; El-Sayed, M. A., Simultaneous Time-Dependent Surface-Enhanced Raman Spectroscopy, Metabolomics, and Proteomics Reveal Cancer Cell Death Mechanisms Associated with Gold Nanorod Photothermal Therapy. *J Am Chem Soc* **2016,** 138, (47), 15434-15442.

150. Cuykx, M.; Negreira, N.; Beirnaert, C.; Van den Eede, N.; Rodrigues, R.; Vanhaecke, T.; Laukens, K.; Covaci, A., Tailored liquid chromatography-mass spectrometry analysis improves the coverage of the intracellular metabolome of HepaRG cells. *J Chromatogr A* **2017,** 1487, 168-178.

151. Deng, J.; Li, W.; Yang, Q.; Liu, Y.; Fang, L.; Guo, Y.; Guo, P.; Lin, L.; Yang, Y.; Luan, T., Biocompatible Surface-Coated Probe for in Vivo, in Situ, and Microscale Lipidomics of Small Biological Organisms and Cells Using Mass Spectrometry. *Anal Chem* **2018**.

152. Marasco, C. C.; Enders, J. R.; Seale, K. T.; McLean, J. A.; Wikswo, J. P., Real-time cellular exometabolome analysis with a microfluidic-mass spectrometry platform. *PLoS One* **2015,** 10, (2), e0117685.

153. Silva, L. P.; Northen, T. R., Exometabolomics and MSI: deconstructing how cells interact to transform their small molecule environment. *Current Opinion in Biotechnology* **2015,** 34, 209-216.

154. Berg, M.; Vanaerschot, M.; Jankevics, A.; Cuypers, B.; Breitling, R.; Dujardin, J. C., LC-MS metabolomics from study design to data-analysis - using a versatile pathogen as a test case. *Comput Struct Biotechnol J* **2013,** 4, e201301002.

155. Boccard, J.; Veuthey, J. L.; Rudaz, S., Knowledge discovery in metabolomics: An overview of MS data handling. *Journal of Separation Science* **2010,** 33, (3), 290-304.

156. Creek, D. J.; Anderson, J.; McConville, M. J.; Barrett, M. P., Metabolomic analysis of trypanosomatid protozoa. *Mol Biochem Parasitol* **2012,** 181, (2), 73-84.

157. Haggarty, J.; Burgess, K. E., Recent advances in liquid and gas chromatography methodology for extending coverage of the metabolome. *Curr Opin Biotechnol* **2017,** 43, 77-85.

158. Fukumoto, T.; Nishiumi, S.; Fujiwara, S.; Yoshida, M.; Nishigori, C., Novel serum metabolomics-based approach by gas chromatography/triple quadrupole mass spectrometry for detection of human skin cancers: Candidate biomarkers. *Journal of Dermatology* **2017,** 44, (11), 1268-1275.

159. Pasikanti, K. K.; Ho, P. C.; Chan, E. C., Gas chromatography/mass spectrometry in metabolic profiling of biological fluids. *J Chromatogr B Analyt Technol Biomed Life Sci* **2008,** 871, (2), 202-11.

160. Barbier Saint Hilaire, P.; Hohenester, U. M.; Colsch, B.; Tabet, J. C.; Junot, C.; Fenaille, F., Evaluation of the High-Field Orbitrap Fusion for Compound Annotation in Metabolomics. *Anal Chem* **2018,** 90, (5), 3030-3035.

161. Wang, S. Y.; Zhou, L. N.; Wang, Z. C.; Shi, X. Z.; Xu, G. W., Simultaneous metabolomics and lipidomics analysis based on novel heart-cutting two-dimensional liquid chromatography-mass spectrometry. *Analytica Chimica Acta* **2017,** 966, 34-40.

162. Zhang, W.; Hankemeier, T.; Ramautar, R., Next-generation capillary electrophoresis-mass spectrometry approaches in metabolomics. *Curr Opin Biotechnol* **2017,** 43, 1-7.

163. Kuehnbaum, N. L.; Kormendi, A.; Britz-McKibbin, P., Multisegment injection-capillary electrophoresis-mass spectrometry: a high-throughput platform for metabolomics with high data fidelity. *Anal Chem* **2013,** 85, (22), 10664-9.

164. Lindenburg, P. W.; Haselberg, R.; Rozing, G.; Ramautar, R., Developments in Interfacing Designs for CE-MS: Towards Enabling Tools for Proteomics and Metabolomics. *Chromatographia* **2015,** 78, (5-6), 367-377.

165. Gulersonmez, M. C.; Lock, S.; Hankemeier, T.; Ramautar, R., Sheathless capillary electrophoresis-mass spectrometry for anionic metabolic profiling. *Electrophoresis* **2016,** 37, (7-8), 1007-14.

166. Habchi, B.; Alves, S.; Jouan-Rimbaud Bouveresse, D.; Appenzeller, B.; Paris, A.; Rutledge, D. N.; Rathahao-Paris, E., Potential of dynamically harmonized Fourier transform ion cyclotron resonance cell for high-throughput metabolomics fingerprinting: control of data quality. *Anal Bioanal Chem* **2018,** 410, (2), 483-490.

167. Schultz, G. A.; Corso, T. N.; Prosser, S. J.; Zhang, S., A fully integrated monolithic microchip electrospray device for mass spectrometry. *Analytical Chemistry* **2000,** 72, (17), 4058-4063.

168. Huang, M. Z.; Cheng, S. C.; Cho, Y. T.; Shiea, J., Ambient ionization mass spectrometry: A tutorial. *Analytica Chimica Acta* **2011,** 702, (1), 1-15.

169. Cody, R. B.; Laramee, J. A.; Durst, H. D., Versatile new ion source for the analysis of materials in open air under ambient conditions. *Anal Chem* **2005,** 77, (8), 2297-302.

170. Gu, H.; Pan, Z.; Xi, B.; Asiago, V.; Musselman, B.; Raftery, D., Principal component directed partial least squares analysis for combining nuclear magnetic resonance and mass spectrometry data in metabolomics: application to the detection of breast cancer. *Anal Chim Acta* **2011,** 686, (1-2), 57-63.

171. Balog, J.; Sasi-Szabo, L.; Kinross, J.; Lewis, M. R.; Muirhead, L. J.; Veselkov, K.; Mirnezami, R.; Dezso, B.; Damjanovich, L.; Darzi, A.; Nicholson, J. K.; Takats, Z., Intraoperative Tissue Identification Using Rapid Evaporative Ionization Mass Spectrometry. *Science Translational Medicine* **2013,** 5, (194).

172. Alexander, J.; Gildea, L.; Balog, J.; Speller, A.; McKenzie, J.; Muirhead, L.; Scott, A.; Kontovounisios, C.; Rasheed, S.; Teare, J.; Hoare, J.; Veselkov, K.; Goldin, R.; Tekkis, P.; Darzi, A.; Nicholson, J.; Kinross, J.; Takats, Z., A novel methodology for in vivo endoscopic phenotyping of colorectal cancer based on real-time analysis of the mucosal lipidome: a prospective observational study of the iKnife. *Surg Endosc* **2017,** 31, (3), 1361-1370.

173. Chaurand, P., Imaging mass spectrometry of thin tissue sections: A decade of collective efforts. *Journal of Proteomics* **2012,** 75, (16), 4883-4892.

174. Guenther, S.; Muirhead, L. J.; Speller, A. V. M.; Golf, O.; Strittmatter, N.; Ramakrishnan, R.; Goldin, R. D.; Jones, E.; Veselkov, K.; Nicholson, J.; Darzi, A.; Takats, Z., Spatially Resolved Metabolic Phenotyping of Breast Cancer by Desorption Electrospray Ionization Mass Spectrometry. *Cancer Research* **2015,** 75, (9), 1828-1837.

175. Wiseman, J. M.; Ifa, D. R.; Zhu, Y. X.; Kissinger, C. B.; Manicke, N. E.; Kissinger, P. T.; Cooks, R. G., Desorption electrospray ionization mass spectrometry: Imaging drugs and metabolites in tissues. *Proceedings of the National Academy of Sciences of the United States of America* **2008,** 105, (47), 18120-18125.

176. Zhu, J.; Bean, H. D.; Jimenez-Diaz, J.; Hill, J. E., Secondary electrospray ionization-mass spectrometry (SESI-MS) breathprinting of multiple bacterial lung pathogens, a mouse model study. *J Appl Physiol (1985)* **2013,** 114, (11), 1544-9.

177. Gaugg, M. T.; Gomez, D. G.; Barrios-Collado, C.; Vidal-de-Miguel, G.; Kohler, M.; Zenobi, R.; Sinues, P. M. L., Expanding metabolite coverage of real-time breath analysis by coupling a universal secondary electrospray ionization source and high resolution mass spectrometry-a pilot study on tobacco smokers. *Journal of Breath Research* **2016,** 10, (1).

178. Rioseras, A. T.; Gaugg, M. T.; Sinues, P. M. L., Secondary electrospray ionization proceeds via gas-phase chemical ionization. *Analytical Methods* **2017,** 9, (34), 5052-5057.

179. Sinues, P. M. L.; Landoni, E.; Miceli, R.; Dibari, V. F.; Dugo, M.; Agresti, R.; Tagliabue, E.; Cristoni, S.; Orlandi, R., Secondary electrospray ionization-mass

spectrometry and a novel statistical bioinformatic approach identifies a cancer-related profile in exhaled breath of breast cancer patients: a pilot study. *Journal of Breath Research* **2015,** 9, (3).

180. Zhang, X.; Kew, K.; Reisdorph, R.; Sartain, M.; Powell, R.; Armstrong, M.; Quinn, K.; Cruickshank-Quinn, C.; Walmsley, S.; Bokatzian, S.; Darland, E.; Rain, M.; Imatani, K.; Reisdorph, N., Performance of a High-Pressure Liquid Chromatography-Ion Mobility-Mass Spectrometry System for Metabolic Profiling. *Anal Chem* **2017,** 89, (12), 6384-6391.

181. Maleki, H.; Karanji, A. K.; Majuta, S.; Maurer, M. M.; Valentine, S. J., Ion Mobility Spectrometry-Mass Spectrometry Coupled with Gas-Phase Hydrogen/Deuterium Exchange for Metabolomics Analyses. *J Am Soc Mass Spectrom* **2018,** 29, (2), 230-241.

182. Earl, D. C.; Ferrell, P. B., Jr.; Leelatian, N.; Froese, J. T.; Reisman, B. J.; Irish, J. M.; Bachmann, B. O., Discovery of human cell selective effector molecules using single cell multiplexed activity metabolomics. *Nat Commun* **2018,** 9, (1), 39.

183. Delabriere, A.; Hohenester, U. M.; Colsch, B.; Junot, C.; Fenaille, F.; Thevenot, E. A., proFIA: a data preprocessing workflow for flow injection analysis coupled to high-resolution mass spectrometry. *Bioinformatics* **2017,** 33, (23), 3767-3775.

184. Spicer, R.; Salek, R. M.; Moreno, P.; Canueto, D.; Steinbeck, C., Navigating freely-available software tools for metabolomics analysis. *Metabolomics* **2017,** 13, (9), 106.

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

186. Pluskal, T.; Castillo, S.; Villar-Briones, A.; Oresic, M., MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **2010,** 11, 395.

187. Myint, L.; Kleensang, A.; Zhao, L.; Hartung, T.; Hansen, K. D., Joint Bounding of Peaks Across Samples Improves Differential Analysis in Mass Spectrometry-Based Metabolomics. *Anal Chem* **2017,** 89, (6), 3517-3523.

188. DeFelice, B. C.; Mehta, S. S.; Samra, S.; Cajka, T.; Wancewicz, B.; Fahrmann, J. F.; Fiehn, O., Mass Spectral Feature List Optimizer (MS-FLO): A Tool To Minimize False Positive Peak Reports in Untargeted Liquid Chromatography-Mass Spectroscopy (LC-MS) Data Processing. *Analytical Chemistry* **2017,** 89, (6), 3250-3255.

189. Mizuno, H.; Ueda, K.; Kobayashi, Y.; Tsuyama, N.; Todoroki, K.; Min, J. Z.; Toyo'oka, T., The great importance of normalization of LC-MS data for highly-accurate non-targeted metabolomics. *Biomedical Chromatography* **2017,** 31, (1).

190. Warrack, B. M.; Hnatyshyn, S.; Ott, K. H.; Reily, M. D.; Sanders, M.; Zhang, H. Y.; Drexler, D. M., Normalization strategies for metabonomic analysis of urine samples. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* **2009,** 877, (5-6), 547-552.

191. De Livera, A. M.; Olshansky, M.; Speed, T. P., Statistical analysis of metabolomics data. *Methods Mol Biol* **2013,** 1055, 291-307.

192. Dieterle, F.; Ross, A.; Schlotterbeck, G.; Senn, H., Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in H-1 NMR metabonomics. *Analytical Chemistry* **2006,** 78, (13), 4281-4290.

193. Li, B.; Tang, J.; Yang, Q. X.; Li, S.; Cui, X. J.; Li, Y. H.; Chen, Y. Z.; Xue, W. W.; Li, X. F.; Zhu, F., NOREVA: normalization and evaluation of MS-based metabolomics data. *Nucleic Acids Research* **2017,** 45, (W1), W162-W170.

194. Reisetter, A. C.; Muehlbauer, M. J.; Bain, J. R.; Nodzenski, M.; Stevens, R. D.; Ilkayeva, O.; Metzger, B. E.; Newgard, C. B.; Lowe, W. L.; Scholtens, D. M., Mixture model normalization for non-targeted gas chromatography/mass spectrometry metabolomics data. *Bmc Bioinformatics* **2017,** 18.

195. Hollander, M.; Wolfe, D. A., *Nonparametric statistical methods*. Wiley: New York,, 1973; p xviii, 503 p.

196. Broadhurst, D. I.; Kell, D. B., Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics* **2006,** 2, (4), 171-196.

197. Armstrong, R. A., When to use the Bonferroni correction. *Ophthalmic and Physiological Optics* **2014,** 34, (5), 502-508.

198. Benjamini, Y.; Hochberg, Y., Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological* **1995,** 57, (1), 289-300.

199. Brereton, R. G.; Lloyd, G. R., Support vector machines for classification and regression. *Analyst* **2010,** 135, (2), 230-67.

200. Worley, B.; Powers, R., Multivariate Analysis in Metabolomics. *Curr Metabolomics* **2013,** 1, (1), 92-107.

201. Beckonert, O.; Bollard, M. E.; Ebbels, T. M. D.; Keun, H. C.; Antti, H.; Holmes, E.; Lindon, J. C.; Nicholson, J. K., NMR-based metabonomic toxicity classification: hierarchical cluster analysis and k-nearest-neighbour approaches. *Analytica Chimica Acta* **2003,** 490, (1-2), 3-15.

202. Gromski, P. S.; Muhamadali, H.; Ellis, D. I.; Xu, Y.; Correa, E.; Turner, M. L.; Goodacre, R., A tutorial review: Metabolomics and partial least squares-discriminant analysis - a marriage of convenience or a shotgun wedding. *Analytica Chimica Acta* **2015,** 879, 10-23.

203. Hammer, B.; Gersmann, K., A note on the universal approximation capability of support vector machines. *Neural Processing Letters* **2003,** 17, (1), 43-53.

204. Breiman, L., Random forests. *Machine Learning* **2001,** 45, (1), 5-32.

205. Pahikkala, T.; Okser, S.; Airola, A.; Salakoski, T.; Aittokallio, T., Wrapper-based selection of genetic features in genome-wide association studies through fast matrix operations. *Algorithms Mol Biol* **2012,** 7, (1), 11.

206. Phan, J. H.; Kothari, S.; Wang, M. D., omniClassifier: a Desktop Grid Computing System for Big Data Prediction Modeling. *ACM BCB* **2014,** 2014, 514-523.

207. Hogan, S. R.; Phan, J. H.; Alvarado-Velez, M.; Wang, M. D.; Bellamkonda, R. V.; Fernandez, F. M.; LaPlaca, M. C., Discovery of Lipidome Alterations Following Traumatic Brain Injury via High-Resolution Metabolomics. *J Proteome Res* **2018**.

208. Huang, S.; Chong, N.; Lewis, N. E.; Jia, W.; Xie, G. X.; Garmire, L. X., Novel personalized pathway-based metabolomics models reveal key metabolic pathways for breast cancer diagnosis. *Genome Medicine* **2016,** 8.

209. Min, S.; Lee, B.; Yoon, S., Deep learning in bioinformatics. *Briefings in Bioinformatics* **2017,** 18, (5), 851-869.

210. Angermueller, C.; Parnamaa, T.; Parts, L.; Stegle, O., Deep learning for computational biology. *Molecular Systems Biology* **2016,** 12, (7).

211. Chen, Y. F.; Li, Y.; Narayan, R.; Subramanian, A.; Xie, X. H., Gene expression inference with deep learning. *Bioinformatics* **2016,** 32, (12), 1832-1839.

212. Kelley, D. R.; Snoek, J.; Rinn, J. L., Basset: learning the regulatory code of the accessible genome with deep convolutional neural networks. *Genome Res* **2016,** 26, (7), 990-9.

213. Alakwaa, F. M.; Chaudhary, K.; Garmire, L. X., Deep Learning Accurately Predicts Estrogen Receptor Status in Breast Cancer Metabolomics Data. *Journal of Proteome Research* **2018,** 17, (1), 337-347.

214. Benton, H. P.; Ivanisevic, J.; Mahieu, N. G.; Kurczy, M. E.; Johnson, C. H.; Franco, L.; Rinehart, D.; Valentine, E.; Gowda, H.; Ubhi, B. K.; Tautenhahn, R.; Gieschen, A.; Fields, M. W.; Patti, G. J.; Siuzdak, G., Autonomous metabolomics for rapid metabolite identification in global profiling. *Anal Chem* **2015,** 87, (2), 884-91.

215. Kind, T.; Tsugawa, H.; Cajka, T.; Ma, Y.; Lai, Z.; Mehta, S. S.; Wohlgemuth, G.; Barupal, D. K.; Showalter, M. R.; Arita, M.; Fiehn, O., Identification of small molecules using accurate mass MS/MS search. *Mass Spectrom Rev* **2017**.

216. Lawson, T. N.; Weber, R. J. M.; Jones, M. R.; Chetwynd, A. J.; Rodriguez-Blanco, G.; Di Guida, R.; Viant, M. R.; Dunn, W. B., msPurity: Automated Evaluation of Precursor Ion Purity for Mass Spectrometry-Based Fragmentation in Metabolomics. *Analytical Chemistry* **2017,** 89, (4), 2432-2439.

217. Chen, G.; Walmsley, S.; Cheung, G. C. M.; Chen, L.; Cheng, C. Y.; Beuerman, R. W.; Wong, T. Y.; Zhou, L.; Choi, H., Customized Consensus Spectral Library Building for Untargeted Quantitative Metabolomics Analysis with Data Independent Acquisition Mass Spectrometry and MetaboDIA Workflow. *Anal Chem* **2017,** 89, (9), 4897-4906.

218. Zhu, X. C.; Chen, Y. P.; Subramanian, R., Comparison of Information-Dependent Acquisition, SWATH, and MSAll Techniques in Metabolite Identification Study Employing Ultrahigh-Performance Liquid Chromatography-Quadrupole Time-of-Flight Mass Spectrometry. *Analytical Chemistry* **2014,** 86, (2), 1202-1209.

219. Anjo, S. I.; Santa, C.; Manadas, B., SWATH-MS as a tool for biomarker discovery: From basic research to clinical applications. *Proteomics* **2017,** 17, (3-4).

220. Gillet, L. C.; Navarro, P.; Tate, S.; Rost, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R., Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Molecular & Cellular Proteomics* **2012,** 11, (6).

221. Naz, S.; Gallart-Ayala, H.; Reinke, S. N.; Mathon, C.; Blankley, R.; Chaleckis, R.; Wheelock, C. E., Development of a Liquid Chromatography-High Resolution Mass Spectrometry Metabolomics Method with High Specificity for Metabolite Identification Using All Ion Fragmentation Acquisition. *Anal Chem* **2017,** 89, (15), 7933-7942.

222. Bateman, K. P.; Castro-Perez, J.; Wrona, M.; Shockcor, J. P.; Yu, K.; Oballa, R.; Nicoll-Griffith, D. A., MSE with mass defect filtering for in vitro and in vivo metabolite identification. *Rapid Communications in Mass Spectrometry* **2007,** 21, (9), 1485-1496.

223. Bruderer, T.; Varesio, E.; Hidasi, A. O.; Duchoslav, E.; Burton, L.; Bonner, R.; Hopfgartner, G., Metabolomic spectral libraries for data-independent SWATH liquid chromatography mass spectrometry acquisition. *Anal Bioanal Chem* **2018,** 410, (7), 1873-1884.

224. Li, H.; Cai, Y.; Guo, Y.; Chen, F.; Zhu, Z. J., MetDIA: Targeted Metabolite Extraction of Multiplexed MS/MS Spectra Generated by Data-Independent Acquisition. *Anal Chem* **2016,** 88, (17), 8757-64.

225. Panchaud, A.; Scherl, A.; Shaffer, S. A.; von Haller, P. D.; Kulasekara, H. D.; Miller, S. I.; Goodlett, D. R., Precursor Acquisition Independent From Ion Count: How to Dive Deeper into the Proteomics Ocean. *Analytical Chemistry* **2009,** 81, (15), 6481-6488.

226. Go, E. P., Database resources in metabolomics: an overview. *J Neuroimmune Pharmacol* **2010,** 5, (1), 18-30.

227. Uppal, K.; Walker, D. I.; Jones, D. P., xMSannotator: An R Package for Network-Based Annotation of High-Resolution Metabolomics Data. *Anal Chem* **2017,** 89, (2), 1063-1067.

228. Creek, D. J.; Jankevics, A.; Breitling, R.; Watson, D. G.; Barrett, M. P.; Burgess, K. E., Toward global metabolomics analysis with hydrophilic interaction liquid

chromatography-mass spectrometry: improved metabolite identification by retention time prediction. *Anal Chem* **2011,** 83, (22), 8703-10.

229. Aicheler, F.; Li, J.; Hoene, M.; Lehmann, R.; Xu, G.; Kohlbacher, O., Retention Time Prediction Improves Identification in Nontargeted Lipidomics Approaches. *Anal Chem* **2015,** 87, (15), 7698-704.

230. Cao, M.; Fraser, K.; Huege, J.; Featonby, T.; Rasmussen, S.; Jones, C., Predicting retention time in hydrophilic interaction liquid chromatography mass spectrometry and its use for peak annotation in metabolomics. *Metabolomics* **2015,** 11, (3), 696-706.

231. Bruderer, T.; Varesio, E.; Hopfgartner, G., The use of LC predicted retention times to extend metabolites identification with SWATH data acquisition. *J Chromatogr B Analyt Technol Biomed Life Sci* **2017,** 1071, 3-10.

232. Lai, Z. J.; Tsugawa, H.; Wohlgemuth, G.; Mehta, S.; Mueller, M.; Zheng, Y. X.; Ogiwara, A.; Meissen, J.; Showalter, M.; Takeuchi, K.; Kind, T.; Beal, P.; Arita, M.; Fiehn, O., Identifying metabolites by integrating metabolome databases with mass spectrometry cheminformatics. *Nature Methods* **2018,** 15, (1), 53-+.

233. Zhou, Z.; Shen, X.; Tu, J.; Zhu, Z. J., Large-Scale Prediction of Collision Cross-Section Values for Metabolites in Ion Mobility-Mass Spectrometry. *Anal Chem* **2016,** 88, (22), 11084-11091.

234. Zhou, Z.; Tu, J.; Zhu, Z. J., Advancing the large-scale CCS database for metabolomics and lipidomics at the machine-learning era. *Curr Opin Chem Biol* **2018,** 42, 34-41.

235. Zheng, X.; Aly, N. A.; Zhou, Y.; Dupuis, K. T.; Bilbao, A.; Paurus, V. L.; Orton, D. J.; Wilson, R.; Payne, S. H.; Smith, R. D.; Baker, E. S., A structural examination and collision cross section database for over 500 metabolites and xenobiotics using drift tube ion mobility spectrometry. *Chem Sci* **2017,** 8, (11), 7724-7736.

236. Zhou, Z.; Xiong, X.; Zhu, Z. J., MetCCS predictor: a web server for predicting collision cross-section values of metabolites in ion mobility-mass spectrometry based metabolomics. *Bioinformatics* **2017,** 33, (14), 2235-2237.

237. Xia, J. G.; Sinelnikov, I. V.; Han, B.; Wishart, D. S., MetaboAnalyst 3.0-making metabolomics more meaningful. *Nucleic Acids Research* **2015,** 43, (W1), W251-W257.

238. Davidson, R. L.; Weber, R. J.; Liu, H.; Sharma-Oates, A.; Viant, M. R., Galaxy-M: a Galaxy workflow for processing and analyzing direct infusion and liquid chromatography mass spectrometry-based metabolomics data. *Gigascience* **2016,** 5, 10.

239. Li, S.; Park, Y.; Duraisingham, S.; Strobel, F. H.; Khan, N.; Soltow, Q. A.; Jones, D. P.; Pulendran, B., Predicting network activity from high throughput metabolomics. *PLoS Comput Biol* **2013,** 9, (7), e1003123.

240. Aiche, S.; Sachsenberg, T.; Kenar, E.; Walzer, M.; Wiswedel, B.; Kristl, T.; Boyles, M.; Duschl, A.; Huber, C. G.; Berthold, M. R.; Reinert, K.; Kohlbacher, O., Workflows for automated downstream data analysis and visualization in large-scale computational mass spectrometry. *Proteomics* **2015,** 15, (8), 1443-7.

241. Uppal, K.; Ma, C. Y.; Go, Y. M.; Jones, D. P., xMWAS: a data-driven integration and differential network analysis tool. *Bioinformatics* **2018,** 34, (4), 701-702.

242. Pico, A. R.; Kelder, T.; van Iersel, M. P.; Hanspers, K.; Conklin, B. R.; Evelo, C., WikiPathways: Pathway editing for the people. *Plos Biology* **2008,** 6, (7), 1403-1407.

243. Basu, S.; Duren, W.; Evans, C. R.; Burant, C. F.; Michailidis, G.; Karnovsky, A., Sparse network modeling and metscape-based visualization methods for the analysis of large-scale metabolomics data. *Bioinformatics* **2017,** 33, (10), 1545-1553.

244. Huan, T.; Forsberg, E. M.; Rinehart, D.; Johnson, C. H.; Ivanisevic, J.; Benton, H. P.; Fang, M.; Aisporna, A.; Hilmers, B.; Poole, F. L.; Thorgersen, M. P.; Adams, M. W. W.; Krantz, G.; Fields, M. W.; Robbins, P. D.; Niedernhofer, L. J.; Ideker, T.; Majumder, E. L.; Wall, J. D.; Rattray, N. J. W.; Goodacre, R.; Lairson, L. L.; Siuzdak, G., Systems biology guided by XCMS Online metabolomics. *Nat Methods* **2017,** 14, (5), 461-462.

245. Barupal, D. K.; Fiehn, O., Chemical Similarity Enrichment Analysis (ChemRICH) as alternative to biochemical pathway mapping for metabolomic datasets. *Scientific Reports* **2017,** 7.

246. Scalbert, A.; Brennan, L.; Fiehn, O.; Hankemeier, T.; Kristal, B. S.; van Ommen, B.; Pujos-Guillot, E.; Verheij, E.; Wishart, D.; Wopereis, S., Mass-spectrometry-based metabolomics: limitations and recommendations for future progress with particular focus on nutrition research. *Metabolomics* **2009,** 5, (4), 435-458.

247. Dudzik, D.; Barbas-Bernardos, C.; Garcia, A.; Barbas, C., Quality assurance procedures for mass spectrometry untargeted metabolomics. a review. *Journal of Pharmaceutical and Biomedical Analysis* **2018,** 147, 149-173.

248. Sud, M.; Fahy, E.; Cotter, D.; Azam, K.; Vadivelu, I.; Burant, C.; Edison, A.; Fiehn, O.; Higashi, R.; Nair, K. S.; Sumner, S.; Subramaniam, S., Metabolomics Workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. *Nucleic Acids Res* **2016,** 44, (D1), D463-70.

249. Steinbeck, C.; Conesa, P.; Haug, K.; Mahendraker, T.; Williams, M.; Maguire, E.; Rocca-Serra, P.; Sansone, S. A.; Salek, R. M.; Griffin, J. L., MetaboLights: towards a new COSMOS of metabolomics data management. *Metabolomics* **2012,** 8, (5), 757-760.

250. Haug, K.; Salek, R. M.; Steinbeck, C., Global open data management in metabolomics. *Current Opinion in Chemical Biology* **2017,** 36, 58-63.

251. Rocca-Serra, P.; Salek, R. M.; Arita, M.; Correa, E.; Dayalan, S.; Gonzalez-Beltran, A.; Ebbels, T.; Goodacre, R.; Hastings, J.; Haug, K.; Koulman, A.; Nikolski, M.; Oresic, M.; Sansone, S. A.; Schober, D.; Smith, J.; Steinbeck, C.; Viant, M. R.; Neumann, S., Data standards can boost metabolomics research, and if there is a will, there is a way. *Metabolomics* **2016,** 12, (1).

PART I: LIQUID CHROMATOGRAPHY-MASS SPECTROMETRY BASED NON-TARGETED METABOLOMICS FOR DISEASE DETECTION AND EARLY PREDICTION

CHAPTER 2. PROSTATE CANCER DETECTION BY ULTRAPERFORMANCE LIQUID CHROMATOGRAPHY-MASS SPECTROMETRY SERUM METABOLOMICS

Adapted with permission from
Zang X†, Jones CM†, Long TQ, Monge ME, Zhou M, Walker LD, Mezencev R, Gray A, McDonald JF, Fernández FM. Feasibility of detecting prostate cancer by ultraperformance liquid chromatography-mass spectrometry serum metabolomics. *Journal of Proteome Research.* **2014**, *13*, 3444-3454.
†equal contributing author

C. M. Jones and M. Zhou optimized sample preparation and UPLC-MS methods and acquired UPLC-MS data. C. M. Jones and M.E. Monge processed UPLC-MS data and performed data analysis. T.Q. Long performed multivariate statistical analysis. X. Zang generated subpanels of discriminant features based on their prevalence in samples and searched database for tentative annotation of all discriminant metabolites. X. Zang, M.E. Monge and C. M. Jones conducted UPLC-MS/MS experiments. X. Zang analyzed UPLC-MS/MS data to confirm discriminant metabolite identifications, performed chemical standard validation experiments with assistance from M. E. Monge., R. Mezencev, C. M Jones, X. Zang and M.E. Monge determined the biological relevance of discriminant metabolites.

2.1 Abstract

Prostate cancer (PCa) is the second leading cause of cancer-related mortality in American men. The prevalent diagnosis method is based on the serum Prostate-specific antigen (PSA) screening test, which suffers from low specificity, over-diagnosis and over-treatment. In this study, non-targeted metabolomic profiling of age-matched serum samples from PCa patients and control individuals was performed using ultra performance liquid chromatography coupled to high resolution mass spectrometry (UPLC-MS) and machine learning methods. A metabolite-based *in vitro* diagnostic multivariate index assay (IVDMIA) was developed to predict the presence of PCa in

serum samples with high classification sensitivity, specificity and accuracy. A panel of 40 metabolic features was found to be differential with 92.1% sensitivity, 94.3% specificity, and 93.0% accuracy. The performance of the IVDMIA was higher than the prevalent PSA test. Within the discriminant panel, 31 metabolites were identified by MS and MS/MS, with 10 further confirmed chromatographically by standards. The identification of fatty acids, amino acids, lysophospholipids, and bile acids provided insights into the metabolic alterations associated with the disease. In addition, several metabolites were mapped to the steroid hormone biosynthesis pathway, providing further insights into PCa related biological pathway perturbation. With additional work, the results presented here show great potential towards implementation in clinical settings.

2.2 Prostate Cancer Diagnosis

2.2.1 Prostate Cancer Screening

Prostate cancer (PCa) is the second leading cause of cancer-related mortality in American men, with ~29,430 estimated deaths in the United States in 2018.[1] The prevalent diagnosis method is based on the triad of digital rectal examination (DRE), blood Prostate-Specific Antigen (PSA) measurement, and ultrasound-guided prostate biopsy. Although its death rate has dropped 52% from 1993 to 2015,[1] the use of PSA as a diagnostic serum marker still presents several drawbacks. The concentration of this protein in the blood stream increases during the development of cancer, but also can be secreted as a result of benign prostatic hyperplasia, prostatitis, or other traumas to prostate cells.[2] Therefore, this method suffers from low specificity and consequent over-

diagnosis and over-treatment.[3-6] In addition, DRE has low detection rate for non-palpable growth in PCa tumors[7] and transrectal ultrasound-guided biopsy is limited by undersampling significant cases and oversampling insignificant ones[8].

2.2.2 Overview of Biomarker Discovery Approaches for Prostate Cancer Detection

The drawbacks associated to the current PCa diagnostic methods have led to an increased interest in using non-targeted metabolite profiling to discover new potential metabolic biomarkers that could improve the specificity of PCa diagnosis.[9-11] Different metabolites have been suggested in the literature to be considered as potential biomarkers for PCa, but they still need further validation before reaching clinical practice. [9-11]

Different metabolic alterations have been associated with PCa. For example, tissue sarcosine levels have been shown to increase in the aggressive form of the disease during PCa progression to metastasis, but differences in urine were much less marked, using both LC-MS and GC-MS.[12] These results have been very prominent but somewhat controversial as other targeted studies failed in the attempt of differentiating controls from cancer patients based on sarcosine concentration in biological fluids and cancerous tissues.[9, 13-15] The analysis of cancerous tissues by proton high-resolution magic angle spinning NMR spectroscopy has shown a decrease in the concentrations of citrate and polyamines, and increases in cholines, glycerophospholipids, and lactate concentrations during PCa proliferation.[16, 17] Increased levels of cholesterol as well as alterations in amino acid metabolism were detected in metastatic bone samples by GC-MS.[18] However, all of these studies included too few patients to offer strong leads on the metabolic alterations associated with PCa. McDunn *et al.* analyzed metabolic profiles of from 331

prostate tumor tissues and 178 tumor-free tissues by GC-MS and LC-MS/MS and revealed significant decreased levels of metabolites associated with cell energetics, and increased levels of amino acids, peptides, carnitines, cofactors, lipids, nucleotides and metabolites related to stress in tumor tissues.[19] The authors showed that inclusion of metabolite markers improved the AUC for predictions of organ confinement and 5-year recurrence from of 0.53 to 0.62 and 0.53 to 0.64, respectively.[19] A non-targeted metabolic profiling study analyzed prostate tissue samples from 48 PCa patients by NMR and distinguished cancerous from normal adjacent tissue with a sensitivity of 86.9% and a specificity of 85.2% using multivariate analysis.[20] Significantly changed metabolites included choline, lactate, glutamate, glycine, leucine, glucose, and succinate, among others, and concentrations of spermine, citrate and the ratio (choline+creatine+polyamines)/citrate were found to be significantly altered between high and low grade PCa.[20] Interestingly, metabolic profiles were significantly correlated to tissue composition, defined as the percentage of benign glandular tissue, stroma and cancer, but not to serum PSA levels in patients.[20] Zhang *et al.* conducted non-targeted LC-MS based metabolic profiling of urine samples from 60 PCa patients and 30 healthy controls and identified a panel of four metabolites including ureidoisobutyric acid, indolylacryloyglycine, acetylvanilalinine and 2-oxoglutarate that detected PCa with an AUC of 0.896, comparable to the value of 0.94 provided by PSA test.[21] A panel of plasma lipids that included phosphatidylethanolamines, ether-linked phosphatidylethanolamines, and ether-linked phosphatidylcholines was proposed to discriminate PCa patients from controls through direct infusion ESI tandem MS.[22] The authors demonstrated that a combination of multiple biomarkers yielded better predictive power for the diagnosis of

PCa than univariate analysis of single lipid species when analyzing 105 PCa patients and 36 controls. However, the predictive power was not compared with that of PSA, as this information was not available at the time of cohort design.[22]

2.3 Hypothesis

In this study, we hypothesize that a specific signature is associated with PCa serum metabolic profile, which is distinct from that of age-matched non-PCa controls based on a panel of discriminant metabolites. To test this hypothesis, we performed a LC-MS based non-targeted metabolomics study of serum samples collected from PCa patients and controls (without PCa), in an attempt to detect PCa by developing a metabolite-based *in vitro* diagnostic multivariate index assay (IVDMIA)[23] based on machine learning algorithms.

2.4 Materials and Methods

2.4.1 Chemicals

Healthy human blood serum (S7023-50 mL) and acetic acid (≥99.7%) were purchased from Sigma-Aldrich Corp. (St. Louis, MO, USA). Omnisolv LC-MS grade acetonitrile, Omnisolv high purity dichloromethane and HPLC grade acetone were purchased from EMD (Billerica, MA, USA). LC-MS grade methanol and 2-propanol were purchased from J.T. Baker Avantor Performance Materials, Inc. (Center Valley, PA, USA). Ultrapure water with 18.2 MΩ cm resistivity (Barnstead Nanopure UV ultrapure water system, USA) was used to prepare mobile phases. Uric acid (≥99%), azelaic acid

(98%), undecanedioic acid (97%), heptadecanoic acid (≥98%) and decanoic acid (≥98%) were purchased from Sigma-Aldrich Corp. (St. Louis, MO, USA). Hexadecanedioic acid (98%) was purchased from Ark Pharm, Inc. (Libertyville, IL, USA). Phenylalanyl phenylalanine was purchased from MP Biomedicals (Solon, OH, USA). Phenylacetyl glutamine was purchased from Bachem (Hauptstrasse, Bubendorf, Sitzerland). Indoxyl sulfate potassium was purchased from Alfa Aesar (Ward Hill, MA, USA). Lysophosphatidylcholine (LPC(18:0/0:0)) (1-stearoyl-2-hydroxy-*sn*-glycerol-3-phosphocholine was purchased from Avanti Polar Lipids, Inc. (Alabaster, AL, USA).

2.4.2 Cohort Description

Age-matched blood serum samples were obtained from 64 PCa patients (age range 49-65, mean (standard deviation (SD)) age 59 (4) years) and 50 PCa free controls (age range: 45-76, mean age 57 (7) years). At the 0.05 level, the population means of PCa and controls were not significantly different with the two-sample t-test. The cohort ethnicity was as follows: 28 African American (24.6%); 76 Caucasian (66.7%); 5 Hispanic (4.4%); 2 Asian (1.8%); 2 Jewish ancestry (1.8%); and 1 unknown (0.9%). After approval by the Institutional Review Board (IRB), blood samples were collected at Saint Joseph´s Hospital of Atlanta (GA, USA) by venipuncture from each donor into evacuated blood collection tubes that contained no anticoagulant. Serum was obtained by centrifugation at 5000 rpm for 5 min at 4 °C. Immediately after centrifugation, 200 µL aliquots of serum were frozen and stored at -80 °C for further use. The sample collection and storage procedures for PCa patients and controls were identical. Gleason scores were obtained for 61 PCa patients.

2.4.3 Sample Preparation

A stock sample of healthy human blood serum was used to develop the serum sample preparation protocol and UPLC-MS method. Serum samples were thawed on ice, and protein precipitation was performed by the addition of a mixture of acetone, acetonitrile and methanol (1:1:1 v/v) to 100 µL of serum in a 3:1 volume ratio. Samples were vortex-mixed for 20 s, and centrifuged at 16000 × *g* for 5 min. After centrifugation, 800 µL of dichloromethane were added to 350 µL of supernatant, and vortex-mixed. Following the addition of 250 µL of deionized water, samples were vortex-mixed again to extract the non-polar lipid fraction. The aqueous phase was used for metabolite analysis by UPLC-MS. Samples were randomly separated into 7 batches and consecutively analyzed.

2.4.4 Metabolic Profiling by UPLC-MS

UPLC-MS analysis was performed using a Waters ACQUITY Ultra Performance LC (Waters Corporation, Manchester, UK) system, fitted with a Waters ACQUITY UPLC BEH C_{18} column (2.1 × 50 mm, 1.7 µm particle size), and coupled to a high resolution Synapt G2 high-definition mass spectrometry (HDMS) system (Waters Corporation, Manchester, UK). The Synapt G2 HDMS is a hybrid quadrupole-ion mobility-orthogonal acceleration time-of-flight (TOF) instrument with typical resolving power of 20,000 Full width at half maximum (FWHM) and mass accuracy of 9 ppm at *m/z* 554.2615. The instrument was operated in negative ion mode with a probe capillary voltage of 2.3 kV, and a sampling cone voltage of 45 V. The source and desolvation temperatures were 120 °C and 350 °C, respectively; and the nitrogen desolvation flow rate was 650 L h^{-1}.

The mass spectrometer was calibrated across the range of *m/z* 50-1800 using a 0.5 mM sodium formate solution prepared in 90:10 2-propanol:water v/v. Data were mass corrected during acquisition using a leucine enkephalin reference spray infused at 2 µL min^{-1}. Data were acquired in the range of *m/z* 50-1750 and the scan time was set to 1 s. Data acquisition and processing was carried out using MassLynx v4.1. The chromatographic method for sample analysis involved elution with acetonitrile (mobile phase A) and water with 0.1% acetic acid (mobile phase B) using the following gradient program: 0-1 min 0-10% A; 1-2.5 min 10-15% A; 2.5-4 min 15-22% A; 4-6 min 22-38% A; 6-9 min 38-65% A; 9-12 min 65-80% A; 12-16 min 80-100% A; 16-18 min 100% A. The flow rate was constant at 0.25 mL min^{-1} for 12 min. It was increased to 0.30 mL min^{-1} between 12 and 16 min, and from 0.30 to 0.45 mL min^{-1} between 16 and 18 min. The gradient was returned to its initial conditions over a period of 8 minutes after each sample injection. The column temperature was set to 35 °C, the autosampler tray temperature was set to 5 °C, and the injection volume was 10 µL. UPLC-tandem mass spectrometry (MS/MS) experiments were performed by acquiring mass spectra with applied voltages between 5 and 50 V in the trap cell, using ultra purity argon (≥ 99.999%) as the collision gas. The instrument was calibrated before analysis and solvent and sample preparation blanks were jointly analyzed with the samples in a random order.

2.4.5 Data Analysis

After UPLC-MS analysis, metabolic features (retention time (R_t), *m/z* pairs) were extracted from chromatograms using MarkerLynx XS software. This procedure involved chromatogram alignment, peak picking and integration, peak area extraction, and

normalization. The matrix containing sample peak areas for each feature (R_t, *m/z*) was utilized to build a model for sample classification and to find the minimum set of discriminant features by means of linear support vector machines (SVMs).[24] This supervised classification technique is effective at handling high dimensionality data as those produced in the work described in this chapter. For a binary classification problem, linearly-separable samples represented as a row vector $\boldsymbol{x}$, had membership of two classes *g* (= H or D), where H stands for control and D for PCa disease with labels *c* = -1 for class H, and +1 for class D. In order to build the classification model, 70% of the samples were randomly selected as a training set, and 30% as a test set. Within the training set, 10% of samples were used for validation and to find the minimum set of discriminant features that maximized accuracy in the classification through a recursive feature elimination (RFE) method.[25] The decision function that separated the two classes, defined here as "PCa metabolic score" for the IVDMIA[23], was as follows:

$$PCa\ \ metabolic\ \ score = b + \sum_{j=1}^{J} w_j x_{ij} \qquad [1]$$

$$g(\boldsymbol{x}_i) = sgn\left(\boldsymbol{w}\boldsymbol{x}_i^{'}\ \ + b\right) = sgn(PCa\ \ metabolic\ \ score) \qquad [2]$$

where *w* and *b* are the weight and bias parameters that were determined from the training set, i is the sample number, j is the feature number and J is the total number of features. The sign of the PCa metabolic score determined which class a sample was assigned to: class H if negative and class D if positive. In this classification function, the two classes were divided in the dataspace by a hyperplane $\boldsymbol{w}\boldsymbol{x}^{'} + b = 0$ that maximized the margins between samples of different classes. The margin between the two classes was defined such that:

$$\boldsymbol{w}\boldsymbol{x}^{'} + b \geq 1, \qquad c = +1 \qquad [3]$$

$$\boldsymbol{w}\boldsymbol{x}^{'} + b \leq -1, \qquad c = -1 \qquad [4]$$

To estimate the classification and feature selection performance, ten iterative validations were performed to randomly select the training and test sets. The statistical significance of the model was further assessed through hypothesis testing by permutation tests. No assumptions were made in this non-parametric approach to hypothesis testing regarding the data distribution, and the *p* value was computed as the cumulative sum using the empirical distribution. Two permutation tests were performed using 100 permutation samples with the following null hypothesis:

i) Null hypothesis 1: feature and labels (positive/negative) are independent (*i.e*. indifference when class labels are permutated).

ii) Null hypothesis 2: features are independent within each class (*i.e*. indifference when value of each features are permutated within each class).

If the *p* value < α (α=0.05), the null hypothesis H_0 was rejected; otherwise the observed result was not statistically significant.

Additionally, principal component analysis (PCA) was used to evaluate the performance of all extracted metabolic features or subsets of them in an unsupervised manner using MATLAB R2011b (Version 7.13.0, The MathWorks, Inc., Natick, MA, USA) and the PLS Toolbox (v.6.71, Eigenvector Research, Inc., Wenatchee, WA, USA). Data were preprocessed by autoscaling.

2.4.6 Metabolite Identification Procedure

Compound identification was attempted for the 40 discriminant features remaining

after the feature selection processes. Due to the biological complexity of serum samples, adduct ion analysis was first performed to ensure the unambiguous assignment of the signal of interest in each mass spectrum. Adduct ions corresponding to SVM-selected variables that were investigated in the mass spectra included $[M - H]^-$, $[M + Cl]^-$, $[M + CH_3COO]^-$, $[M + HCOO]^-$, $[M + Na - 2H]^-$, $[M + K - 2H]^-$, $[M - H_2O - H]^-$ and $[M + H_2O - H]^-$, which are typically observed in negative ion mode ESI. The expected *m/z* values for common adduct species were calculated and compared with the experimental values from peaks within the spectra. For spectra in which no confirmatory adducts were present, the accurate mass of the candidate neutral molecule was calculated based on the assumption that the peak of interest corresponded to $[M - H]^-$. Elemental formulae were generated based on the mass accuracy of the peak of interest and isotopic patterns with a mass error of 8 mDa, using MassLynx 4.1. The list of elements included in the search was C, H, N, O, P, S, and Cl. The list of generated elemental formulae were searched against the Metlin database[26] and the Human Metabolome Database (HMDB)[27] in order to determine the possible endogenous metabolite candidates. The MS/MS Metlin[26] and MassBank[28] databases and a literature survey were subsequently used to confirm the identity of putative candidates. Fragmentation patterns were also manually analyzed to discriminate between different isobaric species.

2.5 Results and Discussion

2.5.1 Classification Performance

A total of 480 features (R_t, *m/*z pairs) were extracted from UPLC-MS negative ion mode data from serum samples of PCa patients and controls. These features were used to build a discriminant SVM model for sample classification. An optimum set of 51 discriminant features were found to maximize classification accuracy through a recursive feature elimination method,[25] as illustrated in Figure 2.1. Out of the 51 selected features, 7 were found to be only present in less than 2% of the samples; 2 features were identified as acetaminophen and its sulfite adduct, and 2 additional features were identified as adducts or fragments of other features in the subset, and were thus removed from further consideration. The optimum panel that best discriminated PCa patients from controls was thus reduced to 40 features, demonstrating that the feature selection process accomplished a high reduction in problem dimensionality.

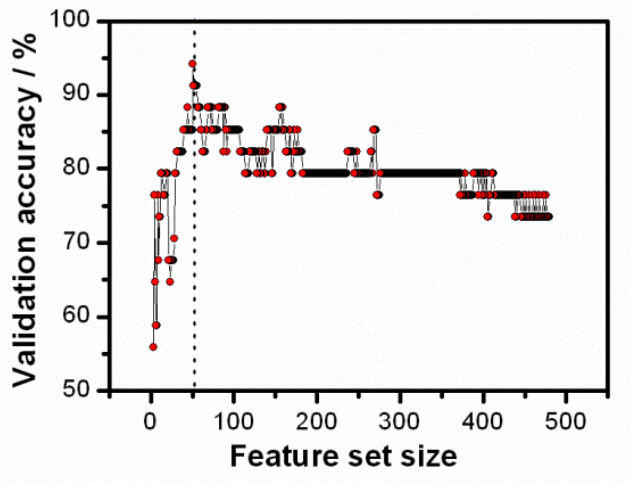

Figure 2.1: Evolution of classification accuracy for a validation sample subset consisting of 10% of the training samples as a function of the number of features retained. The minimum discriminant feature set that maximizes classification accuracy is highlighted with a dashed line.

Figure 2.2 illustrates the “PCa metabolic scores” obtained for the training and the test sets of randomly selected samples that were used to construct and evaluate the classification model, respectively. The separation of the two sample classes (H or D) was determined in the data space by the optimal separating hyperplane for which the margin between the most similar samples in each group was largest, illustrated with a dotted line in the figure. The samples with scores equal to 1 or -1 are the support vectors of the model. For the particular cross-validation iteration illustrated in Figure 2.2 only one sample was misclassified as a false negative. Based on these 40 discriminant features, serum samples were successfully classified with 93.0% accuracy, 92.1% sensitivity, and 94.3% specificity. These values were calculated as the averages from 10 distinct test sets.

In addition, the statistical significance of the model was further evaluated through hypothesis testing and, at 0.05 significance level, the null hypothesis was rejected for all permutations generated (p value = 0.0099). Unambiguously, the classifier did not yield a better leave-one-out-cross validation (LOOCV) accuracy rate than the original data. These results suggest a promising approach that could form the basis for a PCa IVDMIA. In particular, of the 40 differential features, 24 were found to increase in PCa patients, and 16 were found to decrease in PCa, as illustrated in Figure 2.3. It is important to underline, however, that the strength of this IVDMIA resides in the combination of multiple metabolic features using an interpretation function to yield a single, patient-specific result to be used in the disease diagnosis, and not on the average fold change of each differential feature.

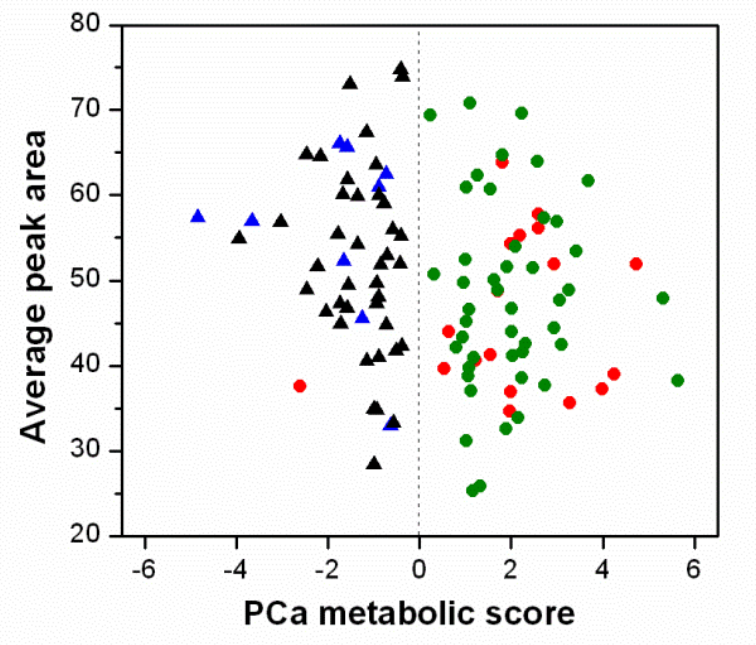

Figure 2.2: Visualization of the PCa metabolic scores obtained by SVMs in one out of 10 iterative model validations based on 40 discriminant features. Green circles correspond to PCa patients in the training set, black triangles correspond to controls in the training set, red circles correspond to PCa patients in the test set built for the iteration shown, and blue triangles correspond to controls in the test set. The dotted line shows the projection of the separating hyperplane: $\boldsymbol{wx}' + b = 0$.

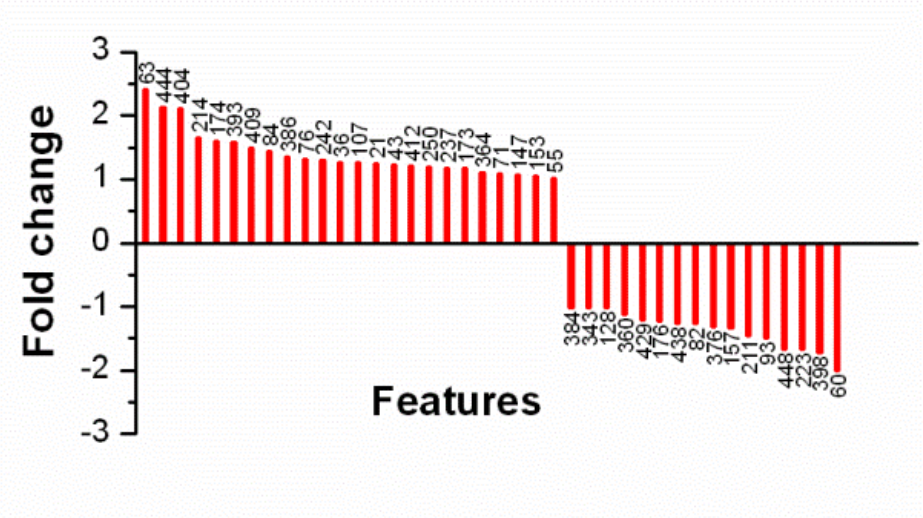

Figure 2.3: Fold change of average peak areas of each discriminant feature. Positive fold changes are calculated as the ratio of average peak areas between PCa patients and controls, and negative fold changes are calculated as the negative value of the ratio of average peak areas between controls and PCa patients. Features are labeled with their codes.

2.5.2 IVDMIA vs. PSA Diagnosis

Table 2.1 summarizes the Gleason scores for the PCa patients, indicating that the most common tumor patterns presented by the patients derived from moderate to aggressive cancers. However, the PSA test performed at surgery did not follow this histological evidence for the entire PCa cohort, as 33% of patients with PCa (n=20) had PSA values lower than the commonly used cutoff point of 4.0 ng mL^{-1}. Figure 2.4 compares PSA and IVDMIA results in terms of true positive and false negative outputs, highlighted in red and black, respectively. The IVDMIA outputs provided by the randomly-selected 10 test sets are visualized as box plots in the figure, and show that the IVDMIA was able to correctly predict 100% of the true positives that were incorrectly diagnosed as negatives by the PSA test. The false negative results provided by the

IVDMIA derived from one sample that was misclassified in all test sets and 4 samples that were misclassified in at least one test set. The classification performance obtained with this cohort shows promise towards detection of PCa that would be missed by the PSA method. The use of multiple discriminant features by this metabolic IVDMIA yields higher predictive power for PCa diagnosis than the univariate analysis of a single marker such as with the PSA method.

Table 2.1: Gleason scores for PCa patients.

Gleason Sum	**# of patients; (%)**
3+3 = 6	13; (20.3)
3+4 = 7	27; (42.2)
3+3 = 6; tert=4	6; (9.4)
4+3 = 7	3; (4.7)
3+4=7; tert=5	2; (3.1)
4+3=7; tert= 5	2; (3.1)
(R) 3+4 = 7; (L) 4+3 = 7	1; (1.6)
4+5 = 9	1; (1.6)
5+4 = 9	1; (1.6)
(R)3+4=7; tert= 5; (L) 3+3=6	1; (1.6)
(R)3+4=7; (L) 4+3 = 7; tert= 5	1; (1.6)
3+5=8; tert= 4	1; (1.6)
(R) 3+3 = 6; (L) 4+5 = 9	1; (1.6)
(R)3+3=6; (L)3+3=6 tert= 4	1; (1.6)

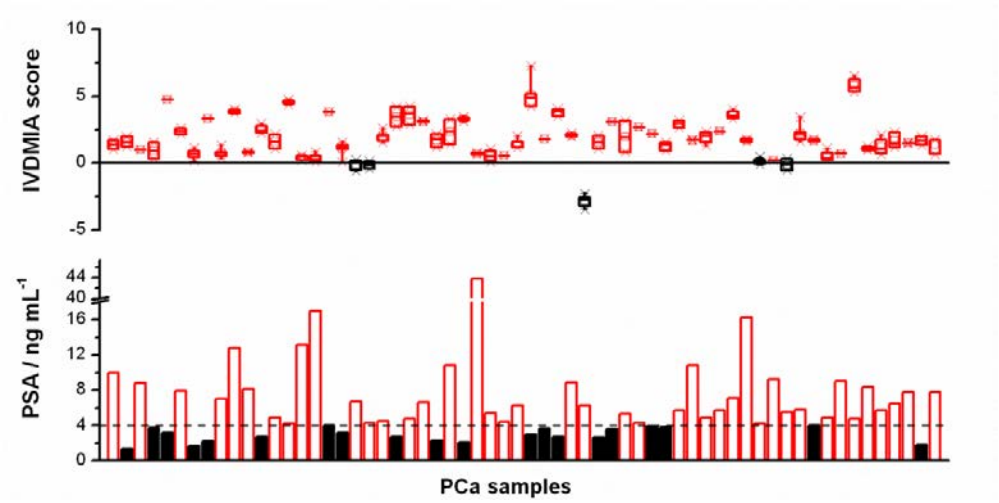

Figure 2.4: Comparison of IVDMIA vs. PSA diagnosis performance for 62 PCa patients. True positive and false negative outputs are highlighted in red and black, respectively. The cutoff point of 4.0 ng mL^{-1} used in PSA-based diagnosis is indicated with a dotted line. The IVDMIA score output is presented as box plots in the figure, each of which is generated by results obtained for each of the 10 test sets where each sample was selected for validation. No comparison is shown for 2 of the 64 PCa samples as they were not randomly selected in any of the 10 cross-validation iterations.

2.5.3 IVDMIA Potential in Clinical Applications

To determine the prevalence of the discriminant features in samples, and to evaluate the feasibility of implementing the PCa IVDMIA in clinical laboratory settings through targeted triple-quadrupole MS-based assays, smaller subgroups of the optimum 40 discriminant features, subsequently referred to as “panel A”, were investigated (Table 2.2). These subpanels were chosen to provide the minimum number of features that collectively captured metabolic PCa patterns with a high level of accuracy, specificity and sensitivity. The selection of these additional subpanels was based on the fraction of features that were present in 50, 70 or 90% of the entire sample cohort, in either PCa

patients or controls. Table 2.2 summarizes the different panels constructed following these criteria, with their corresponding subset of discriminant features. These panels were used to build new SVM models, and cross-validated to provide average values of accuracy, specificity and sensitivity from 10-independent randomly-selected training and testing sets. Thirty eight out of 40 discriminant features were present in more than 50% of controls (Panel B) and 35 out of 40 were present in more than 50% of PCa samples (Panel C), providing similar accuracy, specificity, and sensitivity as panel A. When the criterion for feature prevalence was set to be more stringent, from panel A to panel G; the accuracy, specificity and sensitivity decreased by only ~10%, suggesting the robust biological role that the detected features might have. In other words, the different feature subpanels were not highly sensitive to a reduction in the number of discriminant features, suggesting that the smaller number of metabolites contained in subpanel G could still be potentially useful for building a more focused, simpler IVDMIA for PCa detection in a clinical setting.

To further test this finding, another SVM model was created with only those 13 features that could be confidently assigned to metabolites in subpanel G by high resolution MS and MS/MS (Table 2.3). It was found that this model still provided high classification sensitivity (88.3%), specificity (80.3%), and accuracy (85.0%). The mass spectrometric assay for such model would be much simpler to implement in a targeted fashion due to the reduced number of transitions that a UPLC-MS/MS triple quadrupole method would require, allowing higher analysis throughput and minimizing cost.

The set of 40 SVM weights obtained for panel A from the optimal classification model are shown in Figure 2.5. The figure shows the individual contribution of each of

the 40 discriminant metabolic features in the computed PCa metabolic score, *i.e.*, the weight of each discriminant metabolite in the classification. Figure 2.6 shows a comparison of the different sets of weights for the different panels described in Table 2.2, sorted from the largest to lowest value in panel A and expanded to panels B-G. The figure shows that the sign of the weights generally remained the same across the panels, in agreement with the fact that accuracy, specificity and sensitivity were highly conserved even after restricting the presence of discriminant features to those present in a majority of the patients within the cohort. It was seen that for the most restrictive panels, those features with weights equal to zero, *i.e.*, those that do not contribute to the panels, are those with lower weights in panel A.

Table 2.2: Discriminant feature (sub)panels for PCa detection.

Panel	Accuracy	Specificity	Sensitivity	Discriminant features (#)	Discriminant feature codes	% Controls, % PCa samples
A	93.0	94.3	92.1	40	147, 36, 71, 211, 60, 55, 107, 409, 250, 223, 386, 438, 157, 63, 176, 82, 393, 173, 84, 412, 43, 376, 343, 429, 384, 76, 444, 214, 128, 93, 398, 360, 448, 174, 153, 21, 364, 404, 242, 237	>0%; >0%
B	91.2	90.6	91. 7	38	147, 36, 71, 211, 60, 55, 107, 409, 250, 223, 386, 438, 157, 63, 176, 82, 393, 173, 84, 412, 43, 376, 343, 429, 384, 76, 444, 214, 128, 93, 398, 360, 448, 174, 153, 21, 364, 404	>50%; >0%
C	90.2	87.2	91.8	35	147, 36, 71, 211, 60, 55, 107, 409, 250, 223, 386, 438, 157, 63, 176, 82, 393, 173, 84, 412, 43, 376, 343, 429, 384, 76, 444, 214, 128, 93, 398, 360, 448, 174, 153	>50%; >50% and >0%; >50%
D	86.1	87.2	85.3	28	147, 36, 71, 211, 60, 55, 107, 409, 250, 223, 386, 438, 157, 63, 176, 82, 393, 173, 84, 412, 43, 376, 343, 429, 384, 76, 444, 214	>0%; >70%
E	84.4	80.0	85.8	25	147, 36, 71, 211, 60, 55, 107, 409, 250, 223, 386, 438, 157, 63, 176, 82, 393, 173, 84, 412, 43, 376, 343, 429, 384	>70%; >70% and >70%; >0%
F	85.0	80.0	88.8	22	147, 36, 71, 60, 55, 409, 223, 386, 438, 157, 63, 176, 82, 393, 173, 84, 412, 43, 376, 343, 429, 384	>90%; >0%
G	80.0	81.0	79.3	17	147, 36, 71, 60, 55, 409, 386, 438, 157, 176, 82, 393, 173, 84, 343, 429, 384	>90%; >90% and >0%; >90%

Table 2.3: IVDMIA performance for identified metabolites.

Feature subpanel	Accuracy (%)	Specificity (%)	Sensitivity (%)	Discriminant features (#)	Discriminant feature codes
Identified in Panel G by MS/MS	85.0	80.3	88.3	13	60, 36, 84, 71, 157, 176, 55, 343, 429, 384, 409, 386, 173
Identified in Panel A by MS/MS	91.1	91.3	90.9	31	60, 36, 84, 71, 157, 176, 55, 343, 429, 384, 409, 386, 173, 223, 43, 63, 376, 250, 211, 107, 214, 76, 444, 174, 128, 398, 93, 153, 364, 21, 242
Identified in Panel A by MS/MS and confirmed chromatographically with standards	76.3	70.6	79.9	10	60, 36, 71, 384, 43, 211, 76, 174, 128, 153
Identified in Panel A by MS/MS with xenobiotics and marker 63 excluded	90.2	90.7	89.7	28	60, 36, 84, 71, 157, 176, 55, 343, 429, 384, 409, 386, 173, 223, 43, 376, 250, 211, 107, 214, 76, 444, 174, 128, 398, 93, 153, 242

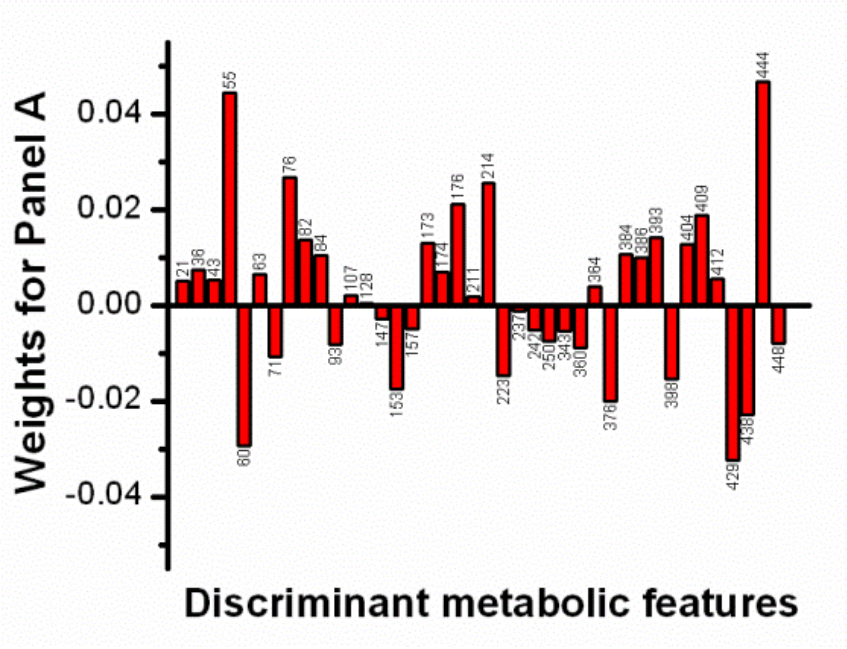

Figure 2.5: Weights for the 40 discriminant metabolic features in panel A. Metabolic features are labeled with their codes.

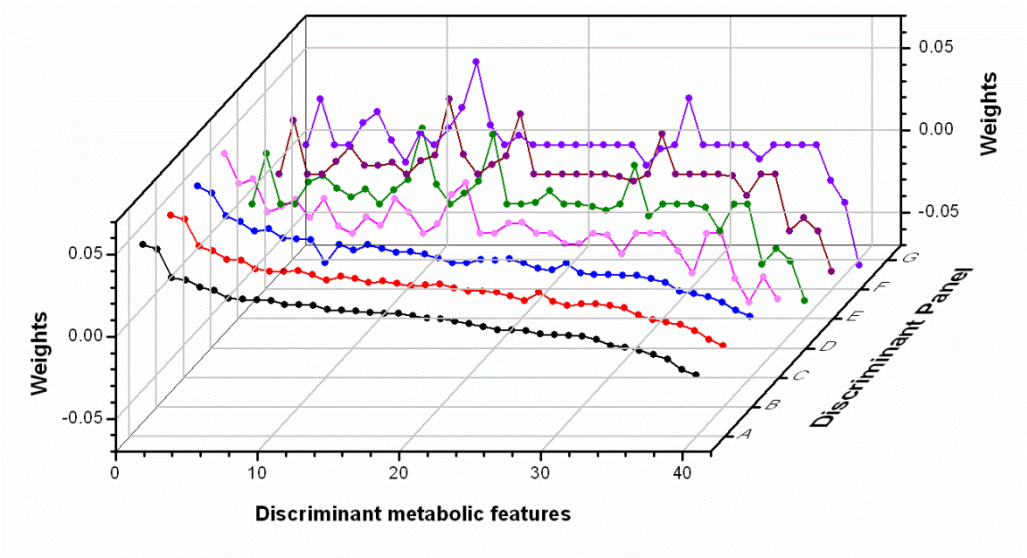

Figure 2.6: Weights for the discriminant metabolic features from panels A-G (indicated in Table 2.2) obtained by the classification model using the total cohort.

2.5.4 Identification of Metabolites Used in the IVDMIA

Once the robustness of the model was established, chemical identification of the 40 discriminant metabolic features was attempted. The high resolving power of the TOF analyzer used allowed generating highly-selective extracted ion chromatograms for each discriminant feature. Adduct ion analysis was used to ensure the unambiguous assignment of the signal of interest in the ESI mass spectrum, and the isotopic pattern and accurate masses were used to generate a list of possible candidate elemental formulae that were searched against databases. Moreover, UPLC-MS/MS experiments were performed to confirm the identities of these candidate metabolites responsible for classification. Tandem MS spectra were compared to those in databases or literature, and fragmentation patterns were manually analyzed as well. Finally, available chemical standards were

subject to UPLC-MS and MS/MS to verify the identity of the candidates by retention time and mass spectral matching.

Of the 40 spectral features in panel A, 31 were identified by high resolution MS and MS/MS, with 10 further confirmed chromatographically by standards. The group of 31 metabolites provided 90.9% sensitivity, 91.3% specificity, and 91.1% accuracy; whereas the 10 differential metabolites confirmed by standards, when considered alone, provided 79.9% sensitivity, 70.6% specificity, and 76.3% accuracy (Table 2.3). It should be noted that among the 31 identified metabolites, 1-α-amino-1H-pyrrole-1-hexanoic acid (feature code 63) had the highest mass error (11.4 mDa), and its identity should be viewed as tentative. However, a classification model built using the set of 30 metabolites excluding feature 63 still provided 92.8% sensitivity, 89.2% specificity and 91.2% accuracy.

2.5.5 Biological Relevance of the IVDMIA Metabolites

Table 2.4 shows the chemical identification for the 40 discriminant features. Metabolites confirmed by retention time and MS/MS matching with standards are shown in bold, and can be viewed as the ones with the higher confidence in the panel. Several discriminant metabolites were identified as fatty acids, amino acids, lysophospholipids, and bile acids, suggesting perturbations in their respective metabolism in PCa. Previous findings have shown abnormality in fatty acid,[29] and amino acid[12, 30, 31] metabolism in PCa patients. Alterations in fatty acid metabolism through an enhanced β-oxidation pathway have been suggested to provide bioenergy for abnormal cell proliferation.[29] Among the different lysophospholipids identified that may play a role in cell signaling,[32] LPC(18:2) and LPC(18:0) have been reported as biomarkers for PCa detection within a

panel of plasma lipids.[22] Uric acid has also been suggested to be a disease risk marker due to its pro-inflammatory properties,[33, 34] and a prospective epidemiological study demonstrated positive association between serum uric acid levels and risk of PCa development.[35] In addition, increased levels of serum uric acid are often found due to tumor lysis syndrome observed as a result of cancer therapy.[36] Interestingly, indoxyl sulfate, a toxic product of dietary tryptophan metabolism that accumulates in the blood of patients with impaired renal function,[37] was also identified among the 40 discriminant features. The reason behind elevated indoxyl sulfate in the sera of PCa patients is not yet fully understood; nonetheless, this nephrotoxic metabolite likely contributes to the disease or its complications *via* multiple mechanisms, including enhanced oxidative stress due to decreased levels of glutathione.[38]

Interestingly, many identified metabolites in the discriminant panel were found to be involved in the steroid hormone biosynthesis pathway. As illustrated in Figure 2.7, the pathway supplies androgens[39-41] such as testosterone and 5α-dihydrotestosterone, to support the growth of androgen-dependent PCa.[42] An average increase of pregnanetriol and androstenedione concentrations in PCa serum suggests that there is a metabolic alteration of the steroid pathway that mimics congenital adrenal hyperplasia (CAH), a metabolic disease that is accompanied by androgen excess due to the diversion of 17-hydroxyprogesterone into the pathway for androgen biosynthesis.[43, 44] In addition, the average decrease of azelaic acid concentration in serum of PCa patients, an inhibitor of 5α-reductase,[45] suggests the disinhibition of 5α-reductase, an enzyme that catalyzes the synthesis of highly active androgen 5α-dihydrotestosterone to support PCa growth. Indeed, azelaic acid, which has a large contribution in the models, has been postulated to

be a potential antitumoral agent.[46] However, the origin of azelaic acid and its monoester identified in the discriminant panel needs further investigation since they have also been reported to originate from corn oil.[47]

Table 2.4: Results for the chemical identification workflow for various discriminant features. Metabolites confirmed by retention time matching with commercially-available standards are highlighted in bold font.

Feature Code	Retention time (min)	*m/z*	Ion type	Elemental Formula	Theoretical *m/z*	Δm (mDa)	Tentative Metabolite Identification	Panel
60	**5.10**	**187.0970**	**$[M-H]^-$**	**$C_9H_{16}O_4$**	**187.0970**	**0.0**	**nonanedioic acid (azelaic acid)**	**G**
36	**0.63**	**167.0206**	**$[M-H]^-$**	**$C_5H_4N_4O_3$**	**167.0205**	**0.1**	**uric acid**	**G**
71	**1.95**	**203.0817**	**$[M-H]^-$**	**$C_{11}H_{12}N_2O_2$**	**203.0821**	**0.4**	**tryptophan**	**G**
384	**11.70**	**508.3403**	**$[M-CH_3]^-$**	**$C_{26}H_{54}NO_7P$**	**508.3403**	**0.0**	**LPC(18:0/0:0)**	**G**
84	8.41	223.1331	$[M-H]^-$	$C_{13}H_{20}O_3$	223.1334	0.3	13-oxo-9,11-tridecadienoic acid	G
157	7.06	273.1703	$[M-H]^-$	$C_{14}H_{26}O_5$	273.1702	0.1	azelaic acid mono-hydroxy-pentyl ester (major) and undecanedioic acid mono-hydroxy-propyl ester	G
176	7.61	287.1854	$[M-H]^-$	$C_{15}H_{28}O_5$	287.1858	0.4	sebacic acid mono-hydroxy-pentyl ester (major) and dodecanedioic acid mono-hydroxy-propyl ester	G
55	5.21	185.0812	$[M-H]^-$	$C_9H_{14}O_4$	185.0814	0.2	5-(2-methylpropyl)-2-oxooxolane-3-carboxylic acid 5-butyl-2-oxooxolane-3-carboxylic acid	G
343	9.77	476.2772	$[M-H]^-$	$C_{23}H_{44}NO_7P$	476.2777	0.5	LPE(0:0/18:2) LPE(18:2/0:0)	G
429	9.80	578.3450	$[M+CH_3COO]^-$	$C_{26}H_{50}NO_7P$	578.3458	0.8	LPC(18:2/0:0)	G
409	5.46	541.2639	$[M-H]^-$	$C_{27}H_{42}O_{11}$	541.2649	1.0	cortolone-3-glucuronide	G
386	6.92	511.2900	$[M-H]^-$	$C_{27}H_{44}O_9$	511.2907	0.7	pregnanetriol glucuronide	G
173	8.19	285.1920	$[M-H]^-$	$C_{19}H_{26}O_2$	285.1855	6.5	androstenedione	G
393	8.12	517.3015		-	-	-	-	G
438	7.04	600.2572		-	-	-	-	G
147	0.55	266.8028		-	-	-	-	G
82	8.12	215.1281		-	-	-	-	G
43	**9.56**	**171.1383**	**$[M-H]^-$**	**$C_{10}H_{20}O_2$**	**171.1385**	**0.2**	**decanoic acid (capric acid)**	**F**
223	6.77	331.1753	$[M-H]^-$	$C_{16}H_{28}O_7$	331.1757	0.4	menthol glucuronide citronellol glucuronide	F
63	7.19	195.1020	$[M-H]^-$	$C_{10}H_{16}N_2O_2$	195.1134	11.4	l-α-amino-1H-pyrrole-1-hexanoic acid	F
376	9.63	504.3081	$[M-CH_3]^-$	$C_{26}H_{50}NO_7P$	504.309	0.9	LPC(0:0/18:2)*	F
412	8.86	545.3323		-	-	-	-	F
211	**4.06**	**311.1387**	**$[M-H]^-$**	**$C_{18}H_{20}N_2O_3$**	**311.1396**	**0.9**	**phenylalanyl phenylalanine**	**E**
250	5.70	383.1521	$[M-H]^-$	$C_{19}H_{28}O_6S$	383.1528	0.7	3β,16α-dihydroxyandrostenone sulfate	E

Table 2.4 (continued).

107	5.40	245.0480	$[M-H]^-$	$C_{10}H_{14}O_5S$	245.0484	0.4	2-tert-butyl-1,4-benzenediol sulfate	E
76	**2.64**	**212.0016**	**$[M-H]^-$**	**$C_8H_7NO_4S$**	**212.0018**	**0.2**	**indoxyl sulfuric acid**	**D**
214	9.87	311.2211	$[M-H]^-$	$C_{18}H_{32}O_4$	311.2222	1.1	9,10-dihydroxy-12Z,15Z-octadecadienoic acid (9,10-DiHODE) 12,13-dihydroxy-9Z,15Z-octadecadienoic acid (12,13-DiHODE) 15,16-dihydroxy-9Z,12Z-octadecadienoic acid (15,16-DiHODE)	D
444	6.82	613.3583	$[M-H]^-$	$C_{32}H_{54}O_{11}$	613.3588	0.5	27-nor-5β-cholestane-3α,7α,12α,24,25-pentol glucuronide	D
174	**9.35**	**285.2059**	**$[M-H]^-$**	**$C_{16}H_{30}O_4$**	**285.2066**	**0.7**	**hexadecanedioic acid**	**C**
128	**2.69**	**263.1023**	**$[M-H]^-$**	**$C_{13}H_{16}N_2O_4$**	**263.1032**	**0.9**	**phenylacetylglutamine**	**C**
153	**14.80**	**269.2475**	**$[M-H]^-$**	**$C_{17}H_{34}O_2$**	**269.2481**	**0.6**	**heptadecanoic acid**	**C**
398	7.06	528.2630	$[M-H]^-$	$C_{26}H_{43}NO_8S$	528.2631	0.1	n-[(3α,5β,7β)-7-hydroxy-24-oxo-3-(sulfooxy)cholan-24-yl]-glycine n-[(3α,5β,7α)-3-hydroxy-24-oxo-7-(sulfooxy)cholan-24-yl]-glycine glycochenodeoxycholate-3-sulfate	C
93	6.36	229.0534	$[M-H]^-$	$C_{10}H_{14}O_4S$	229.0535	0.1	5-isopropyl-2-methylphenol sulfate (carvacrol sulfate)	C
360	8.16	489.2692		-	-	-	-	C
448	8.51	621.3273		-	-	-	-	C
364	5.57	495.2228	$[M-H]^-$	$C_{25}H_{36}O_{10}$	495.2230	0.2	5'-carboxy-α-chromanol glucuronide	B
21	5.16	144.0471	$[M-H]^-$	C_9H_7NO	144.0449	2.2	indole-3-carboxaldehyde	B
404	7.28	537.2501		-	-	-	-	B
242	7.66	369.1740	$[M-H]^-$	$C_{19}H_{30}O_5S$	369.1736	0.4	androsterone sulfate 5α-dihydrotestosterone sulfate etiocholanolone sulfate	A
237	11.34	365.2680		-	-	-	-	A

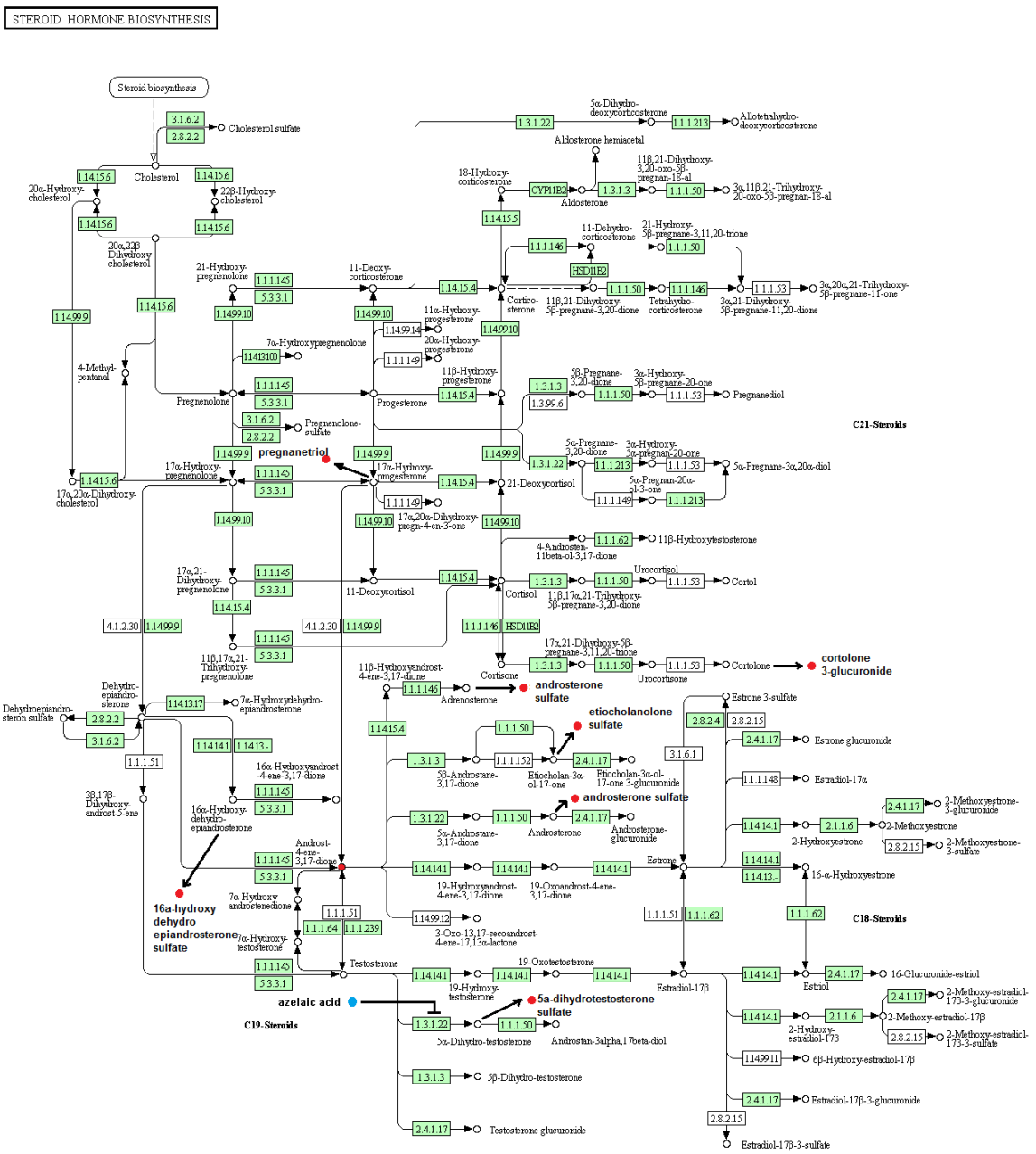

Figure 2.7: Kyoto Encyclopedia of Genes and Genomes (KEGG) steroid hormone biosynthesis pathway (hsa00140). The identified discriminant metabolites are indicated in the pathway. Average increase and decrease of metabolite concentrations in PCa serum is highlighted in red and blue, respectively. Green rectangles: human metabolic enzymes.

Table 2.4 also includes several xenobiotics that can be grouped into two classes based on their origin. Menthol, citronellol, carvacrol, and t-butylhydroquinone are most likely related to food components. Assuming that both PCa patients and controls were equally exposed, on average, to food components/additives, their different metabolism could explain the different levels of these xenometabolites in serum. For instance, the terpenoids menthol, carvacrol and citronellol are metabolized by CYP2A6,[48, 49] which is also involved in steroid metabolism. As a result, average lower concentrations of these terpenoids relative to controls may be suggestive of higher activity of CYP2A6 in PCa patients, supporting inclusion of these xenometabolites in the models. The second group of xenobiotics comprises indole-3-carboxaldehyde and 5'-carboxy-α-chromanol glucuronide, which could possibly result from the consumption of dietary supplements used by cancer patients. Self-medicating with an over-the-counter indole-3-carbinol (I3C) supplement may explain the increased average concentration of indole-3-carboxaldehyde in PCa serum.[50] Indeed, indole-3-carboxaldehyde demonstrated activity against PCa in both *in vitro* and *in vivo* models.[51] Similarly, α-tocopherol, a form of vitamin E and a precursor of 5'-carboxy-α-chromanol glucuronide, have been suggested to influence the development of PCa due to their antioxidant activity.[52] As humans do not normally produce indole-3-carbaldehyde or 5'-carboxy-α-chromanol, and their consideration in the models may reflect dietary supplementation differences rather than endogenous metabolic differences, PCa detection was attempted using 28 of the 31 identified metabolites, excluding from the SVM classification model two metabolites which might result from dietary supplementation and one metabolite with highest mass error (1-α-amino-1H-pyrrole-1-hexanoic acid). This modified classification model yielded 89.7% sensitivity,

90.7% specificity, and 90.2% accuracy (Table 2.3), indicating that the three excluded metabolites had little effect on the overall assay performance, as supported by their low weights in panel A (Figure 2.5 and 2.6).

2.6 Conclusions

This study shows the combined application of UPLC-MS/MS and machine learning methods in the development of a metabolite-based IVDMIA that allows PCa detection in serum samples with high classification sensitivity, specificity and accuracy. A panel of 40 metabolic features was found to be differential with 92.1% sensitivity, 94.3% specificity, and 93.0% accuracy. Of further significance, the detection performance of the IVDMIA was proven to be higher than the prevalent PSA test, highlighting that a combination of multiple discriminant features yields higher predictive power for PCa detection than the univariate analysis. Within the discriminant panel, 31 metabolites were identified by high resolution MS and MS/MS, with 10 further confirmed chromatographically by standards. Fatty acids, amino acids, lysophospholipids, and bile acids have been identified among the discriminant metabolites, suggesting alterations in their metabolism in PCa patients compared to controls. Additionally, several metabolites were mapped to the steroid hormone biosynthesis pathway. These observations demonstrate some of the plausible metabolic alterations in PCa, and provide further insight into the biological pathway changes associated with the disease. The combination of multiple metabolites yielding a single, patient-specific result for disease detection is the strength of the IVDMIA developed in the work presented in this chapter. When the assay is based on the 28 identified disease-related metabolites, PCa can still be detected with 89.7% sensitivity,

90.7% specificity, and 90.2% accuracy. If higher throughput analysis and lower analysis cost and complexity are needed, 13 metabolites that were found to be present in 90% of the entire sample cohort would still offer high classification sensitivity (88.3%), specificity (80.3%), and accuracy (85.0%). Therefore, this assay shows promise towards implementation in the clinical laboratory setting once it is fully validated by the examination of a larger patient cohort.

2.7 References

1. Siegel, R. L.; Miller, K. D.; Jemal, A., Cancer statistics, 2018. *CA Cancer J Clin* **2018,** 68, (1), 7-30.

2. Nadler, R. B.; Humphrey, P. A.; Smith, D. S.; Catalona, W. J.; Ratliff, T. L., Effect of inflammation and benign prostatic hyperplasia on elevated serum prostate-specific antigen levels. *Journal of Urology* **1995,** 154, (2), 407-413.

3. Tombal, B., Over- and underdiagnosis of prostate cancer: The dangers. *European Urology Supplements* **2006,** 5, (6), 511-513.

4. Bickers, B.; Aukim-Hastie, C., New Molecular Biomarkers for the Prognosis and Management of Prostate Cancer - The Post PSA Era. *Anticancer Research* **2009,** 29, (8), 3289-3298.

5. Heijnsdijk, E. A. M.; der Kinderen, A.; Wever, E. M.; Draisma, G.; Roobol, M. J.; de Koning, H. J., Overdetection, overtreatment and costs in prostate-specific antigen screening for prostate cancer. *British Journal of Cancer* **2009,** 101, (11), 1833-1838.

6. Draisma, G.; Etzioni, R.; Tsodikov, A.; Mariotto, A.; Wever, E.; Gulati, R.; Feuer, E.; de Koning, H., Lead Time and Overdiagnosis in Prostate-Specific Antigen Screening: Importance of Methods and Context. *Journal of the National Cancer Institute* **2009,** 101, (6), 374-383.

7. Lee, F.; Littrup, P. J.; Torppedersen, S. T.; Mettlin, C.; Mchugh, T. A.; Gray, J. M.; Kumasaka, G. H.; Mcleary, R. D., Prostate-Cancer - Comparison of Trans-Rectal Us and Digital Rectal Examination for Screening. *Radiology* **1988,** 168, (2), 389-394.

8. Giannarini, G.; Crestani, A.; Rossanese, M.; Ficarra, V., Multiparametric Magnetic Resonance Imaging Targeted Biopsy for Early Detection of Prostate Cancer: All That Glitters Is Not Gold! *Eur Urol* **2017,** 71, (6), 904-906.

9. Trock, B. J., Application of metabolomics to prostate cancer. *Urologic Oncology-Seminars and Original Investigations* **2011,** 29, (5), 572-581.

10. Kelly, R. S.; Vander Heiden, M. G.; Giovannucci, E.; Mucci, L. A., Metabolomic Biomarkers of Prostate Cancer: Prediction, Diagnosis, Progression, Prognosis, and Recurrence. *Cancer Epidemiol Biomarkers Prev* **2016,** 25, (6), 887-906.

11. Lima, A. R.; Bastos Mde, L.; Carvalho, M.; Guedes de Pinho, P., Biomarker Discovery in Human Prostate Cancer: an Update in Metabolomics Studies. *Transl Oncol* **2016,** 9, (4), 357-70.

12. Sreekumar, A.; Poisson, L. M.; Rajendiran, T. M.; Khan, A. P.; Cao, Q.; Yu, J. D.; Laxman, B.; Mehra, R.; Lonigro, R. J.; Li, Y.; Nyati, M. K.; Ahsan, A.; Kalyana-Sundaram, S.; Han, B.; Cao, X. H.; Byun, J.; Omenn, G. S.; Ghosh, D.; Pennathur, S.; Alexander, D. C.; Berger, A.; Shuster, J. R.; Wei, J. T.; Varambally, S.; Beecher, C.; Chinnaiyan, A. M., Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* **2009,** 457, (7231), 910-914.

13. Jentzmik, F.; Stephan, C.; Miller, K.; Schrader, M.; Erbersdobler, A.; Kristiansen, G.; Lein, M.; Jung, K., Sarcosine in Urine after Digital Rectal Examination Fails as a Marker in Prostate Cancer Detection and Identification of Aggressive Tumours. *European Urology* **2010,** 58, (1), 12-18.

14. Schalken, J. A., Is Urinary Sarcosine Useful to Identify Patients With Significant Prostate Cancer? The Trials and Tribulations of Biomarker Development. *European Urology* **2010,** 58, (1), 19-20.

15. Struys, E. A.; Heijboer, A. C.; van Moorselaar, J.; Jakobs, C.; Blankenstein, M. A., Serum sarcosine is not a marker for prostate cancer. *Annals of Clinical Biochemistry* **2010,** 47, 282-282.

16. Swanson, M. G.; Vigneron, D. B.; Tabatabai, Z. L.; Males, R. G.; Schmitt, L.; Carroll, P. R.; James, J. K.; Hurd, R. E.; Kurhanewicz, J., Proton HR-MAS spectroscopy and quantitative pathologic analysis of MRI/3D-MRSI-targeted postsurgical prostate tissues. *Magnetic Resonance in Medicine* **2003,** 50, (5), 944-954.

17. Swanson, M. G.; Zektzer, A. S.; Tabatabai, Z. L.; Simko, J.; Jarso, S.; Keshari, K. R.; Schmitt, L.; Carroll, P. R.; Shinohara, K.; Vigneron, D. B.; Kurhanewicz, J., Quantitative analysis of prostate metabolites using H-1 HR-MAS spectroscopy. *Magnetic Resonance in Medicine* **2006,** 55, (6), 1257-1264.

18. Thysell, E.; Surowiec, I.; Hornberg, E.; Crnalic, S.; Widmark, A.; Johansson, A. I.; Stattin, P.; Bergh, A.; Moritz, T.; Antti, H.; Wikstro, P., Metabolomic Characterization of

Human Prostate Cancer Bone Metastases Reveals Increased Levels of Cholesterol. *PLoS One* **2010,** 5, (12).

19. McDunn, J. E.; Li, Z.; Adam, K. P.; Neri, B. P.; Wolfert, R. L.; Milburn, M. V.; Lotan, Y.; Wheeler, T. M., Metabolomic signatures of aggressive prostate cancer. *Prostate* **2013,** 73, (14), 1547-60.

20. Giskeodegard, G. F.; Bertilsson, H.; Selnaes, K. M.; Wright, A. J.; Bathen, T. F.; Viset, T.; Halgunset, J.; Angelsen, A.; Gribbestad, I. S.; Tessem, M. B., Spermine and Citrate as Metabolic Biomarkers for Assessing Prostate Cancer Aggressiveness. *Plos One* **2013,** 8, (4).

21. Zhang, T.; Watson, D. G.; Wang, L.; Abbas, M.; Murdoch, L.; Bashford, L.; Ahmad, I.; Lam, N. Y.; Ng, A. C.; Leung, H. Y., Application of Holistic Liquid Chromatography-High Resolution Mass Spectrometry Based Urinary Metabolomics for Prostate Cancer Detection and Biomarker Discovery. *PLoS One* **2013,** 8, (6), e65880.

22. Zhou, X. C.; Mao, J. H.; Ai, J. M.; Deng, Y. P.; Roth, M. R.; Pound, C.; Henegar, J.; Welti, R.; Bigler, S. A., Identification of Plasma Lipid Biomarkers for Prostate Cancer by Lipidomics and Bioinformatics. *Plos One* **2012,** 7, (11), e48889.

23. Zhang, Z., An In Vitro Diagnostic Multivariate Index Assay (IVDMIA) for Ovarian Cancer: Harvesting the Power of Multiple Biomarkers. *Reviews in Obstetrics & Gynecology* **2012,** 5, (1), 35-41.

24. Brereton, R. G.; Lloyd, G. R., Support vector machines for classification and regression. *Analyst* **2010,** 135, (2), 230-67.

25. Guan, W.; Zhou, M.; Hampton, C.; Benigno, B.; Walker, L. D.; Gray, A.; McDonald, J.; Fernandez, F., Ovarian cancer detection from metabolomic liquid chromatography/mass spectrometry data by support vector machines. *BMC Bioinformatics* **2009,** 10, (1), 259.

26. Smith, C. A.; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G., METLIN - A metabolite mass spectral database. *Therapeutic Drug Monitoring* **2005,** 27, (6), 747-751.

27. Wishart, D. S.; Jewison, T.; Guo, A. C.; Wilson, M.; Knox, C.; Liu, Y. F.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E.; Bouatra, S.; Sinelnikov, I.; Arndt, D.; Xia, J. G.; Liu, P.; Yallou, F.; Bjorndahl, T.; Perez-Pineiro, R.; Eisner, R.; Allen, F.;

Neveu, V.; Greiner, R.; Scalbert, A., HMDB 3.0-The Human Metabolome Database in 2013. *Nucleic Acids Research* **2013,** 41, (D1), D801-D807.

28. Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K.; Oda, Y.; Kakazu, Y.; Kusano, M.; Tohge, T.; Matsuda, F.; Sawada, Y.; Hirai, M. Y.; Nakanishi, H.; Ikeda, K.; Akimoto, N.; Maoka, T.; Takahashi, H.; Ara, T.; Sakurai, N.; Suzuki, H.; Shibata, D.; Neumann, S.; Iida, T.; Tanaka, K.; Funatsu, K.; Matsuura, F.; Soga, T.; Taguchi, R.; Saito, K.; Nishioka, T., MassBank: a public repository for sharing mass spectral data for life sciences. *Journal of Mass Spectrometry* **2010,** 45, (7), 703-714.

29. Liu, Y., Fatty acid oxidation is a dominant bioenergetic pathway in prostate cancer. *Prostate Cancer and Prostatic Diseases* **2006,** 9, (3), 230-234.

30. Ebenezar, J.; Pu, Y.; Wang, W. B.; Liu, C. H.; Alfano, R. R., Stokes shift spectroscopy pilot study for cancerous and normal prostate tissues. *Applied Optics* **2012,** 51, (16), 3642-3649.

31. Fu, Y. M.; Lin, H.; Liu, X.; Fang, W.; Meadows, G. G., Cell death of prostate cancer cells by specific amino acid restriction depends on alterations of glucose metabolism. *J. Cell Physiol.* **2010,** 224, (2), 491-500.

32. Peyruchaud, O., Novel Implications for Lysophospholipids, Lysophosphatidic Acid and Sphingosine 1-Phosphate, as Drug Targets in Cancer. *Anti-Cancer Agents in Medicinal Chemistry* **2009,** 9, (4), 381-391.

33. Brys, M.; Morel, A.; Forma, E.; Krzeslak, A.; Wilkosz, J.; Rozanski, W.; Olas, B., Relationship of urinary isoprostanes to prostate cancer occurence. *Molecular and Cellular Biochemistry* **2013,** 372, (1-2), 149-153.

34. Fini, M.; Elias, A.; Johnson, R.; Wright, R., Contribution of uric acid to cancer risk, recurrence, and mortality. *Clinical and Translational Medicine* **2012,** 1, (1), 16.

35. Kolonel, L.; Yoshizawa, C.; Nomura, A.; Stemmermann, G., Relationship of serum uric acid to cancer occurrence in a prospective male cohort. *Cancer Epidemiol. Biomarkers Prev.* **1994,** 3, (3), 225 - 228.

36. Tsimberidou, A.; Keating, M., Hyperuricemic syndromes in cancer patients. *Contrib. Nephrol.* **2005,** 147, 47 - 60.

37. Barreto, F. C.; Barreto, D. V.; Liabeuf, S.; Meert, N.; Glorieux, G.; Temmar, M.; Choukroun, G.; Vanholder, R.; Massy, Z. A.; European Uremic Toxin Work Grp, E. U. T., Serum Indoxyl Sulfate Is Associated with Vascular Disease and Mortality in Chronic Kidney Disease Patients. *Clinical Journal of the American Society of Nephrology* **2009,** 4, (10), 1551-1558.

38. Dou, L.; Jourde-Chiche, N.; Faure, V.; Cerini, C.; Berland, Y.; Dignat-George, F.; Brunet, P., The uremic solute indoxyl sulfate induces oxidative stress in endothelial cells. *J. Thromb. Haemost.* **2007,** 5, (6), 1302-8.

39. Kliman, B.; Prout, G. R.; Maclaughlin, R. A.; Daly, J. J.; Griffin, P. P., Altered androgen metabolism in metastatic prostate cancer *Journal of Urology* **1978,** 119, (5), 623-626.

40. Titus, M. A.; Schell, M. J.; Lih, F. B.; Tomer, K. B.; Mohler, J. L., Testosterone and dihydrotestosterone tissue levels in recurrent prostate cancer. *Clinical Cancer Research* **2005,** 11, (13), 4653-7.

41. Zumoff, B.; Levin, J.; Strain, G. W.; Rosenfeld, R. S.; Oconnor, J.; Freed, S. Z.; Kream, J.; Whitmore, W. S.; Fukushima, D. K.; Hellman, L., Abnormal levels of plasma hormones in men with prostate-cancer - evidence toward a 2-disease theory. *Prostate* **1982,** 3, (6), 579-588.

42. Knudsen, K. E.; Scher, H. I., Starving the Addiction: New Opportunities for Durable Suppression of AR Signaling in Prostate Cancer. *Clinical Cancer Research* **2009,** 15, (15), 4792-4798.

43. Horton, R.; Frasier, S. D., Androstenedione and its conversion to plasma testosterone in congenital adrenal hyperplasia *Journal of Clinical Investigation* **1967,** 46, (6), 1003-&.

44. Speiser, P. W.; White, P. C., Congenital adrenal hyperplasia. *New England Journal of Medicine* **2003,** 349, (8), 776-788.

45. Passi, S.; Picardo, M.; De Luca, C.; Nazzaro-Porro, M., Mechanism of azelaic acid action in acne. *G. Ital. Dermatol. Venereol.* **1989,** 124, (10), 455-63.

46. Breathnach, A. S., Azelaic acid: potential as a general antitumoural agent. *Medical Hypotheses* **1999,** 52, (3), 221-226.

47. Matsubara, T.; Tanaka, N.; Krausz, K. W.; Manna, S. K.; Kang, D. W.; Anderson, E. R.; Luecke, H.; Patterson, A. D.; Shah, Y. M.; Gonzalez, F. J., Metabolomics Identifies an Inflammatory Cascade Involved in Dioxin- and Diet-Induced Steatohepatitis. *Cell Metabolism* **2012,** 16, (5), 634-644.

48. Dong, R. H.; Fang, Z. Z.; Zhu, L. L.; Ge, G. B.; Cao, Y. F.; Li, X. B.; Hu, C. M.; Yang, L.; Liu, Z. Y., Identification of CYP isoforms involved in the metabolism of thymol and carvacrol in human liver microsomes (HLMs). *Pharmazie* **2012,** 67, (12), 1002-1006.

49. Miyazawa, M.; Marumoto, S.; Takahashi, T.; Nakahashi, H.; Haigou, R.; Nakanishi, K., Metabolism of (+)- and (-)-Menthols by CYP2A6 in Human Liver Microsomes. *Journal of Oleo Science* **2011,** 60, (3), 127-132.

50. Anderton, M. J.; Manson, M. M.; Verschoyle, R. D.; Gescher, A.; Lamb, J. H.; Farmer, P. B.; Steward, W. P.; Williams, M. L., Pharmacokinetics and tissue disposition of indole-3-carbinol and its acid condensation products after oral administration to mice. *Clinical Cancer Research* **2004,** 10, (15), 5233-5241.

51. Souli, E.; Machluf, M.; Morgenstern, A.; Sabo, E.; Yannai, S., Indole-3-carbinol (I3C) exhibits inhibitory and preventive effects on prostate tumors in mice. *Food and Chemical Toxicology* **2008,** 46, (3), 863-870.

52. Heinonen, O. P.; Albanes, D.; Virtamo, J.; Taylor, P. R.; Huttunen, J. K.; Hartman, A. M.; Haapakoski, J.; Malila, N.; Rautalahti, M.; Ripatti, S.; Maenpaa, H.; Teerenhovi, L.; Koss, L.; Virolainen, M.; Edwards, B. K., Prostate cancer and supplementation with alpha-tocopherol and beta-carotene: incidence and mortality in a controlled trial. *J. Natl. Cancer Inst.* **1998,** 90, (6), 440-6.

CHAPTER 3. EARLY DETECTION OF CYSTIC FIBROSIS ACUTE PULMONARY EXACERBATIONS BY ULTRAPERFORMANCE LIQUID CHROMATOGRAPHY-MASS SPECTROMETRY EXHALED BREATH CONDENSATE METABOLOMICS

Adapted with permission from
Zang X, Monge ME, McCarty NA, Stecenko AA, Fernández FM. Feasibility of Early Detection of Cystic Fibrosis Acute Pulmonary Exacerbations by Exhaled Breath Condensate Metabolomics: A Pilot Study. *Journal of Proteome Research.* **2017**, *16*, 550-558.

Zang X, Monge ME, Gaul D, McCarty NA, Stecenko AA, Fernández FM. Early Detection of Cystic Fibrosis Acute Pulmonary Exacerbations in Adult and Pediatric Patients by Exhaled Breath Condensate Metabolomics. (in preparation).

For the pilot study, M. E. Monge optimized sample preparation and UPLC-MS analysis methods. X. Zang performed sample preparation and acquired UPLC-MS data with assistance from M. E. Monge. X. Zang performed data processing and analysis, and identified all metabolites by using UPLC-MS/MS experiments. For the large cohort study, X. Zang optimized UPLC-MS analysis methods and prepared the samples. UPLC-MS data was acquired by X. Zang with assistance of D. Gaul. X. Zang performed data processing and analysis, discriminant metabolite identification and pathway analysis. N. A. McCarty and A. A. Stecenko provided EBC samples for both studies and clinical expertise.

3.1 Abstract

The most common cause of death in patients with cystic fibrosis (CF) is progressive lung function decline and, ultimately, respiratory failure. This decline is punctuated by acute pulmonary exacerbations (APEs), and, in many cases, there is failure to return to baseline lung function. In a pilot study, ultraperformance liquid chromatography-quadrupole-time-of-flight mass spectrometry (UPLC-QTOF-MS) was used to profile metabolites in exhaled breath condensate (EBC) samples in negative ion

mode from 17 clinically stable CF patients, 9 CF patients with an APE severe enough to require hospitalization (termed APE), 5 CF patients during recovery from a severe APE (termed post-APE), and 4 CF patients who were clinically stable at the time of collection but in the subsequent 1 to 3 months developed a severe APE (termed pre-APE). Using orthogonal partial least-squares-discriminant-analysis (oPLS-DA), a panel containing 2 metabolic discriminant features identified as 4-hydroxycyclohexylcarboxylic acid and pyroglutamic acid differentiated the APE from the stable CF samples with 84.6% accuracy. Furthermore, the pre-APE samples were distinguished from the stable CF samples with 90.5% accuracy using a panel of two discriminant features including lactic acid and pyroglutamic acid.

Metabolic profiling of a larger EBC sample cohort (n=210) was performed using UPLC coupled to ultra-high mass accuracy Orbitrap MS. Negative ion mode data and the combination of negative and positive ion mode data showed that classification was possible for age and gender-matched samples grouped into adult and pediatric patients. APE and pre-APE samples were differentiated from stable CF samples for both patient cohorts using oPLS-DA multivariate classification models. Negative ion mode data, yielded acceptable sensitivities (83.3% and 76.2%), specificities (91.7% and 83.7%), and accuracies (88.9% and 81.3%) for discriminating APE from stable CF EBC samples, from pediatric and adult patients, respectively. For the pre-APE vs. stable CF comparison, good sensitivities (85.7% and 89.5%), specificities (88.4% and 84.1%), and accuracies (87.7% and 85.7%) were obtained for EBC samples from pediatric and adult patients, respectively. By combining positive with negative ion mode data, improved classification performance was achieved for most binary comparisons with accuracies

enhanced between 3 and 9.6%. Interestingly, two of the discriminant metabolites identified in the pilot study, lactic acid and 4-hydroxycyclohexylcarboxylic acid, were also selected by machine learning algorithms in some of the optimized discriminant metabolite panels. Some of the other identified discriminant metabolites had microbial relevance, indicating a possible central role of bacterial metabolism in APE development. Further investigation of these bacterial metabolites may provide insight into the relationship between human-microbial co-metabolism and the trigger for APE, leading to better treatment and patient care to prevent the morbidity and mortality associated with APEs.

3.2 Detection and Prediction of Cystic Fibrosis Acute Pulmonary Exacerbations

3.2.1 Introduction: Cystic Fibrosis Acute Pulmonary Exacerbations

Cystic fibrosis (CF) is a genetic disease caused by mutations in the gene encoding the cystic fibrosis transmembrane conductance regulator (CFTR) protein, leading to abnormal ion and water transport across epithelial cells.[1, 2] Although multiple organs are affected by CF, over 90% of patients die from progressive pulmonary disease and subsequent respiratory failure.[3] CF lung disease is characterized by the triad of impaired mucociliary clearance, chronic poly-microbial bacterial infection, and neutrophil-dominated inflammation. This triad results in progressive decline in lung function that is punctuated by acute episodes of increased respiratory symptoms and often decline in lung function that can be marked. These episodes are termed acute pulmonary exacerbations

(APEs). Therapies are intensified and hospitalization is often required in an attempt to restore lung function to baseline. For example, The CF Foundation Patient Registry showed that 24% of children with CF and 43% of adults with CF had one or more APEs that required intravenous antibiotics treatment in 2017.[4] It is known that the frequency of APEs severe enough to require hospitalization adversely impacts the life quality and life expectancy of patients[5] and associated health care costs,[6] but despite the clinical importance of APEs, there is still a general lack of knowledge regarding their pathophysiology,[7-10] resulting in non-uniform treatment decisions.[11]

Different from asthma, the triggers for APEs in CF are still poorly defined. Viral infections, particularly RSV, rhinovirus, and influenza are thought to be important initiating factors in CF APEs.[12, 13] In addition, exposure to cigarette smoke or other pollutants[14] as well as non-adherence to daily maintenance therapy shown to prevent APEs[15] may also be important. Controversy exists on the role of bacteria with some evidence suggesting that increased bacterial load of resident organisms is associated with APEs versus infection with new bacteria. Regardless of the trigger, a generally held notion is that intensification of bacterial infection and inflammation drives the clinical manifestations of APEs, so the mainstay of therapy is intensive antibiotic treatment and physically clearing the airways of debris.[16] However, there is a flaw with this approach, as 25% of CF patients with APEs severe enough to require hospitalization do not recover to their baseline lung function.[17] As we know, there has yet to be consensus diagnostic criteria for CF APE, as most of the current criteria are based on empirical data which has not been formally validated, and thus may cause problematic treatment decisions.[18] In addition, there is no preventive screening method for stable CF patients to signal an

oncoming APE event, which hinders the initiation of early intervention before the establishment of substantial immune response.[19] These drawbacks trigger the motivation to search for novel biomarkers to improve APE diagnosis, and to predict an oncoming APE event.

3.2.2 Overview of Biomarker Discovery Approaches for Cystic Fibrosis Acute Pulmonary Exacerbation Detection and Prediction

Biomarkers indicating CF APEs have been explored in different studies in various biofluids, however, only a few have been demonstrated to be predictive for an exacerbation. The most extensively studied blood biomarker is the C-reactive protein (CRP), which was found to correlate with CF disease severity with a significant increase in APEs compared to baseline in stable CF patients, and a significant decrease after treatment.[20] Matouk *et al.* measured plasma CRP concentration and collected clinical data from 51 stable CF patients, and they demonstrated that the combination of elevated baseline CRP and clinical disease activity scores could predict future APEs.[21] In another study of 13 CF patients with longitudinal plasma samples collected during APEs, CRP and interleukin (IL)-8 were demonstrated to be promising in predicting re-exacerbations, with significantly elevated levels in CF patients who developed a next APE within 42.5 days.[22] A different promising biomarker for CF APEs is calprotectin, which was proved to perform better than CRP in predicting the time to exacerbation and pulmonary function decline.[23-25] In a study of 15 CF patients by Quon *et al.*, plasma soluble cluster of differentiation 14 (sCD14) was shown to be useful for predicting CF exacerbations requiring intravenous therapy within 4 months.[26] Another targeted proteomics study by

the same group measured 117 peptides from 79 proteins in plasma of 104 CF patients by multiple reaction monitoring MS and identified a six-protein panel that predicted the CF APE onset within 4 months with an area under the curve (AUC) of 0.74, better than predictions using forced expiratory volume in 1 second (FEV_1) that yielded an AUC of 0.55. The top ranking candidate markers of this protein-based panel included CRP and IL-6.[27]

Other biofluids have been investigated as sources of potential APE biomarkers. Promising biomarkers found in nasal lavage fluid include γ-induced protein 10 kDa (IP-10), IL-6 and IL-10.[28] In sputum, a less invasive to collect biofluid, neutrophil elastase and IL-8 have been proposed as tentative biomarkers, although the later was not consistently found to have significant changes in APE, perhaps as a result of differences in the children and adult cohorts involved in the studies.[23, 25, 29, 30] Sputum collection in pediatric CF patients, however, is challenging due to the fact that some children cannot expectorate sputum, and sputum induction is time consuming and expensive in clinical settings.[31, 32] In contrast, exhaled breath condensate (EBC) can be more easily collected from people of all ages.[33, 34] For this reason EBC has been one of the preferred biofluids to study biochemical changes in the lung environment. EBC consists of aerosolized epithelial lining fluid containing volatile and non-volatile compounds trapped and diluted by water vapor condensation.[35, 36] However, EBC components may be present at trace levels (nM to μM concentration range), necessitating very sensitive techniques for analysis. A study by Robrocks *et al.* reported the detection of CF APEs in 6 patients with a 2-feature multivariate logistic regression model using EBC 8-isoprostane and nitrite concentrations, with a sensitivity of 40% and a specificity of 97%.[37] Carpagnano *et al.*

detected significant increased concentrations of EBC leukotriene B4 (LTB_4) and IL-6 in 20 CF APE patients compared to 15 age-matched healthy controls, and the levels of these markers decreasing significantly in 6 patients who returned to a stable CF state after 2 weeks of antibiotic treatment.[38] Another study by van Horck *et al.* measured levels of inflammatory markers including IL-6, IL-8, tumor necrosis factor α and macrophage migration inhibitory factor in EBC from CF pediatric patients. [39] By combining these markers with other clinical or demographic parameters, the best model could only predict APE events with 55% accuracy in the validation set.

Most of the above studies have been focused on finding protein biomarkers for discrimination of APE from stable CF patients based on targeted assays. The difficulty with targeted approaches becomes more evident in studies of inflammatory biomarkers. For example, comparison of sputum from clinically stable CF patients and patients during exacerbations has suggested a correlation between APEs and inflammatory mediators such as IL-1β, IL-8, and myeloperoxidase.[40-43] Other studies, however, failed to differentiate CF patients based on IL-8 alone, finding instead other potential protein biomarkers, such as soluble intercellular adhesion molecule-1, calprotectin, and calgranulin A and B.[44, 45] Overall, the literature evidence so far strongly suggests that new APE biomarkers and better understanding of pathways that are aberrant in CF patients with APEs compared to clinically stable CF patients are needed to develop better mechanistic hypotheses on CF pathophysiology, aiding in APE early detection and development of therapeutic strategies.

Compared to proteins, metabolites are closer in proximity to the phenotype and their changes can better reflect the status of a biological system.[46, 47] A few metabolomics

studies have revealed metabolic alterations associated with CF APEs. Quinn *et al.* identified platelet activating factor and related lipids as potential biomarkers for CF APEs in sputum samples collected from 11 CF patients.[48] This study also demonstrated the personalized nature of the CF sputum metabolome. In a plasma metabolomics study by Laguna *et al.*, metabolic profiles of matched paired samples from 25 CF patients collected during an APE and during outpatient clinic visit were analyzed by LC-MS and gas chromatography-MS.[49] Five out of 398 identified metabolites showed significant alterations between APE and stable CF states, including hypoxanthine, N4-acetylcytidine, N-acetylmethionine, mannose, and cortisol.[49] Alvarez *et al.* applied LC-MS-based plasma metabolic profiling to study the metabolic effects of high-dose vitamin D_3 to CF patients with exacerbations.[50] By comparing the metabolomes of the vitamin D_3-treated group (n = 12) with the placebo group (n = 12), 316 out of 9,258 metabolites showed significant group-by-time interaction and 15 pathways were differential between the two groups, with the amino acid pathway being the dominant one.[50] A possible anti-catabolic mechanism of high-dose vitamin D_3 treatment of CF APE was revealed.[50] To the best of our knowledge, only one study has investigated the discrimination of stable CF patients (n = 29) from unstable CF patients during exacerbations (n = 24), identifying ethanol, acetic acid, 2-propanol and methanol as the best discriminating EBC metabolites using NMR.[51] Although these potential biomarkers offered insight into the possible perturbed biological pathways and pathophysiology associated with CF APEs, the study suffered from insufficient statistical power for predicting an oncoming APE event.

3.3 Hypothesis

In this study, we hypothesize that, regardless of the initiating trigger, CF exacerbations severe enough to require hospitalization are associated with a specific metabolic signature in EBC. We further hypothesize that this metabolic fingerprint precedes any symptoms or signs of an APE, signaling an impending exacerbation that can be treated preemptively. Finally, we hypothesize that this chemical signature returns to the clinically stable EBC signature following treatment for the APE.

We utilized a discovery-based metabolomics approach to investigate EBC samples collected from CF patients who are clinically stable compared to those with APEs severe enough to require hospitalization. Once a discriminant metabolite profile is identified, we investigated its presence in the pre-symptomatic phase of an APE event and also its persistence following treatment for the APE. A pilot study was performed on EBC samples from a small cohort including 4 pre-APE, 9 APE and 17 stable CF samples to evaluate the possibility of detecting and predicting CF APEs. After demonstration of the feasibility, we proceeded to investigate a larger cohort that included 97 stable CF patients, 36 pre-APE patients, 41 APE patients and 36 post-APE patients with two aims: first, to compare the results with those from the pilot study; second, to identify more robust EBC biomarkers for CF APE prediction and detection in adult and pediatric cohorts.

3.4 Materials and Methods

3.4.1 Chemicals

3.4.1.1 Pilot study

LC-MS grade methanol, purchased from J.T. Baker Avantor Performance Materials, Inc. (Center Valley, PA, USA), and ultrapure water with 18.2 MΩ·cm resistivity (Barnstead Nanopure UV ultrapure water system, USA) were used to prepare mobile phases and solutions. DL-Lactic acid lithium salt (~99%) and myristoleic acid (≥99%), were purchased from MP Biomedicals, LLC (Solon, OH, USA), pyroglutamic acid (5-oxoproline) from Anaspec, Inc. (San Jose, CA, USA), hydroxyacetone (96.4%) from TCI America (Portland, OR, USA), 2-methylbutyric acid (98%), 3,3-dimethylglutaric acid (≥98%) and pimelic acid (≥98%) from Alfa Aesar (Ward Hill, MA, USA), 4-methylvaleric acid (98.5%) from Acros Organics (Morris, NJ, USA), 4-hydroxycyclohexanecarboxylic acid, D-lactaldehyde solution (1M in H_2O), 3-hydroxybenzoic acid (99%), 4-hydroxybenzoic acid (99%), propionic acid (99%), isovaleric acid (99%), valeric acid (99%), adenosine (≥99%), *trans*-4-hydroxy-L-proline (≥ 99%), L-proline (≥ 99%), sucrose (≥ 99%), L-glutathione reduced (≥ 98.0%), D-tyrosine (99%), D-(+)-glucose monohydrate (≥ 99%) and D-(-)-fructose (≥ 99%) were purchased from Sigma-Aldrich (St. Louis, MO, USA), 8-isoprostane-d_4 (1050 mg/L), 5(S), 6(R)-lipoxin A_4 (100 mg/L) and 5(S), 6(S)-lipoxin A_4 (100 mg/L) from Cayman Chemical Company (Ann Arbor, MI, USA).

3.4.1.2 Large cohort study

Ultrapure water with 18.2 MΩ·cm resistivity (Barnstead Nanopure UV ultrapure water system, USA), Optima LC-MS grade acetonitrile and methanol (Fisher Scientific,

Suwannee, GA, USA), were used for mobile phase preparation, sample preparation and chemical standard solution preparation. LC-MS grade acetic acid, lactic acid, 4-hydroxycyclohexanecarboxylic acid, N-formylanthranilic acid, butyric acid (≥99%), L-tryptophan (≥98%), DL-malic acid (≥99%), sebacic acid (99%), L-carnitine inner salt (≥98%), (E)-3-methylglutaconic acid (≥97%), (E)-2-methylglutaconic acid, trans-β-hydromuconic acid/3-hexenedioic acid, nonanedioic acid (azelaic acid) (98%), propionic acid, lysine, salicylic acid, leucine, isoleucine (≥98%), levulinic acid (98%), butyric acid, glutamic acid, proline, and arginine were purchased from Sigma-Aldrich Corporation (St. Louis, MO, USA). Pyroglutamic acid (5-oxoproline) was purchased from Anaspec, Inc. (San Jose, CA, USA).

3.4.2 Cohort Description

CF patients are usually seen in CF clinic every three months when stable and more frequently with exacerbations. EBC was collected during these regular clinic visits to the Emory+Children's CF Care Center in Atlanta, Georgia, after obtaining informed consent. The patient's clinical course was then followed over the subsequent months so that they could be grouped according to their APE status. Clinically stable CF was defined as CF subjects whose symptoms were at baseline, physical examination of the lungs was at baseline, FEV_1 was within 10% of the yearly baseline, and no new therapies (particularly antibiotics) were added at that clinic visit, plus the patient was seen at the next clinic visit three months later and was again classified as clinically stable. In the pilot study, EBC was collected on 17 CF subjects meeting this definition of clinically

stable (age range 14-39, mean (SD) age 28 (7) years, 29.4% females). A severe APE was defined as an increase in respiratory symptoms (cough, sputum production) and/or changes in physical examination of the lungs (increase in crackles, decrease in airflow), at least a 10% decrease in FEV_1, and (in the opinion of the clinician) requiring hospitalization for treatment of the APE. For the pilot study, EBC was collected in 9 of such subjects with a severe APE at the time of hospitalization (age range 15-39, mean age 26 (7) years, 55.6% females). In 5 subjects, EBC was collected 1 to 3 months after an APE event requiring hospitalization, labeled as post-APE (age range 19-30, mean age 26 (5) years, 40% females). Finally, EBC was collected in 4 subjects who were clinically stable as defined above but in the subsequent 1 to 3 months developed APEs severe enough to require hospitalization, labeled as pre-APE (age range 15-39, mean age 27 (10) years, 50% females). At the 0.05 level, the age population means were not significantly different with the two-sample t-test for all possible pairs of sample classes. Among the 26 patients from whom all the 35 EBC samples (17 CFs, 9 APEs, 5 post-APEs and 4 pre-APEs) were collected, 6 patients had multiple samples collected in different groups of disease severity. In most cases the samples were not drawn from the same APE episode, so they don't meet the criteria of paired samples. Due to the small number of samples available, we tried to include all the samples to maximize sample size in each group of disease severity.

For the large cohort study, EBC samples were collected from 97 stable CF patients (age range 5-66, 50% female), 36 patients at pre-APE stage (age range 8-61, 61% female), 41 APE patients (age range 8-58, 56% female), and 36 patients at post-APE stage (age range 8-64, 64% female). Due to the wide age distribution in the large cohort,

samples were separated into adult and pediatric groups within each class. Principal component analysis (PCA) was performed for samples in each age group, and the outliers outside 95% confidence intervals were removed. Subsequently, oPLS-DA binary classifications were performed for age-matched samples in each group (Table 3.1).

Table 3.1: Age- and gender-matched samples in the large cohort study used for oPLS-DA analyses.

Data type	Classes Compared	Age group	Class	Number of samples	Mean age (SD)
Negative ion mode	APE vs. stable CF	Pediatric	APE	18	15 (3)
		Pediatric	Stable CF	36	14 (2)
		Adult	APE	21	29 (10)
		Adult	Stable CF	43	33 (12)
	Pre-APE vs. stable CF	Pediatric	APE	14	13 (3)
		Pediatric	Stable CF	43	12 (3)
		Adult	APE	19	32 (15)
		Adult	Stable CF	44	33 (12)
Combined negative and positive ion mode	APE vs. stable CF	Pediatric	APE	16	15 (2)
		Pediatric	Stable CF	39	14 (2)
		Adult	APE	19	28 (11)
		Adult	Stable CF	44	33 (12)
	Pre-APE vs. stable CF	Pediatric	APE	14	13 (3)
		Pediatric	Stable CF	48	13 (3)
		Adult	APE	21	32 (15)
		Adult	Stable CF	44	33 (12)

3.4.3 EBC Sample Collection and Preparation

This study used EBC samples made available by the CF Biospecimen Registry, a part of the Emory+Children's Center for CF and Airways Disease Research. EBC sample collection followed the guidelines approved by the Georgia Institute of Technology and

the Emory University Institutional Review Boards (approval number IRB00000372). An R-Tube collector (Respiratory Research, Inc., Austin, TX, USA) was used to collect EBC samples, which were immediately frozen at -80 °C until processed.

For the pilot study, EBC samples were thawed and lyophilized at -40 °C and ~100 mTorr for 24 hours using a VirTis bench top freeze-dryer (SP Industries, Stone Ridge, NY, USA). Sample residues were reconstituted in water/methanol (90:10 v/v) with a concentration factor of 20, and analyzed by UPLC-MS. Blank samples containing ultrapure water went through the same sample preparation procedure. Prior to UPLC-MS, samples were randomly separated into two batches and analyzed on consecutive days together with solvent and sample preparation blanks. Quality control (QC) samples (5.50 μM L-glutathione (reduced), trans-4-hydroxy-L-proline, adenosine, D-(+)-glucose monohydrate, D-(-)-fructose, sucrose, 5(S), 6(R)-lipoxin A_4 and 5(S), 6(S)-lipoxin A_4, 8.85 μM D-tyrosine, 5.49 μM L-proline and 5.51 μM 8-isoprostane-d_4 solution) were analyzed every 5 hours to verify the stability of retention times, peak shapes and areas during the analysis. Chemical standards for metabolite identity validation were prepared in ultrapure water or methanol (or a mixture of those solvents), depending on their solubility. For the large cohort study, the sample preparation procedure was identical to the pilot study. QC samples were prepared by combining 40 μL of each EBC sample and mixing them together, followed by lyophilization and resuspension to provide a 20 times up-concentration.

3.4.4 Metabolic Profiling by UPLC-MS

For the pilot study, UPLC-MS analyses were performed using a Waters ACQUITY UPLC H Class system fitted with a Waters ACQUITY UPLC BEH C_{18} column (2.1×50 mm, 1.7 μm particle size, Waters Corporation, Milford, MA, USA), coupled to a Xevo G2 QTOF mass spectrometer (Waters Corporation, Manchester, UK) with an electrospray ionization (ESI) source. The typical resolving power and mass accuracy of the Xevo G2 QTOF mass spectrometer were 25,000 FWHM and 1.8 ppm at *m/z* 554.2615, respectively. Gradient elution was employed in the chromatographic separation method using water (mobile phase A) and methanol (mobile phase B), with the following program: 0-1 min, 90%-80% A, 1-3 min 80%-60% A, 3-5 min 60%-50% A, 5-10 min 50%-40% A, 10-15 min 40%-10% A, 15-20 min 10% A. The flow rate was constant at 0.3 mL min^{-1}. After each sample run, the column was re-equilibrated to the initial conditions in 6 min. The injection volume was 5 μL. The column and autosampler tray temperatures were set at 60 and 5 °C, respectively. The mass spectrometer was operated in negative ion mode with a probe capillary voltage of 2.0 kV and a sampling cone voltage of 12.0 V. The source and desolvation gas temperatures were set to 120 and 350 °C, respectively. The nitrogen gas desolvation flow rate was 650 L h^{-1}. The mass spectrometer was calibrated across the range of *m/z* 50-1500 using a 0.5 mM sodium formate solution prepared in 2-propanol/water (90:10 v/v). Data were drift corrected during acquisition using a leucine enkephalin (*m/z* 554.2615) reference spray infused at 4 μL min^{-1}. Data were acquired in the range of *m/z* 50-1500, and the scan time was set to 1 s. Technical duplicates were acquired in all cases, except for 3 samples with too little volume for replicates. For UPLC-MS/MS experiments, the product ion mass spectra were acquired with collision cell voltages between 7 and 35 V, and sampling cone voltages of

12V or 30V, depending on the analyte. Ultra-high-purity argon (≥ 99.999%) was used as the collision gas in UPLC-MS/MS experiments. Data acquisition and processing were carried out using MassLynx version 4.1 (Waters Corp., Milford, MA, USA).

For the large cohort study, UPLC-MS metabolic profiling were also performed on a Dionex Ultimate 3000 UHPLC system (Thermo Scientific, Dionex, Sunnyvale, California, USA) equipped with a Waters ACQUITY UPLC BEH C_{18} column (2.1×50 mm, 1.7 μm particle size, Waters Corporation, Milford, MA, USA), coupled to a Thermo Scientific™ Q Exactive™ HF hybrid quadruple-Orbitrap mass spectrometer. The mass spectrometer parameters were as follows: capillary voltage 3.1 kV, capillary temperature 300 °C, S-lens RF 50.0%, sheath gas, auxiliary gas and sweep gas flow rate 48, 11, and 2 arbitrary units, respectively, and auxiliary gas heater temperature 413 °C. The *m/z* scan range was 50.0 to 750.0 with an AGC target of 3×10^6 and maximum injection time of 512 ms. The mass resolution setting was 240,000. The mass spectrometer was calibrated with Pierce™ LTQ ESI positive ion calibration solution (including caffeine, Met-Arg-Phe-Ala (MRFA) and Ultramark 1621) and negative ion calibration solution (including sodium dodecyl sulfate, sodium taurocholate and Ultramark 1621). Data were acquired in both positive and negative ion modes. LC gradient elution was employed using water (mobile phase A) and methanol (mobile phase B), with the following program: 0-1 min, 90% A, 1-2 min 90%-80% A, 2-6 min 80%-10% A, 6-10 min 10% A. The flow rate was constant at 0.3 mL min^{-1}. After each sample run, the column was re-equilibrated to the initial conditions in 8 min. The injection volume was 5 μL. The column and autosampler tray temperatures were set at 60 and 5 °C, respectively. Each sample was run once. For UPLC-MS/MS experiments, the product ion mass spectra were acquired with a resolution

of 30,000, AGC target of 1×10^5, maximum injection time of 64 ms, isolation window of 0.4 *m/z* and normalized collision cell voltages between 10 and 100, depending on the analyte. Data acquisition and processing were carried out using Thermo Scientific™ XCalibur™ software.

3.4.5 Flow Injection-Traveling Wave Ion Mobility-MS Analysis

For collision cross section (CCS) measurement in the large cohort study, flow injection-traveling wave ion mobility-MS (FI-TWIM-MS) technique was applied to analyze EBC samples and chemical standards. A Waters ACQUITY UPLC I-Class system fitted with a stainless steel union to bypass the chromatographic column, was coupled to a Synapt G2-S high-definition mass spectrometry (HDMS) system (Waters Corporation, Manchester, UK) equipped with an electrospray ionization (ESI) source operated in negative mode. The column compartment was operated at room temperature, and the autosampler temperature was set to 5 °C. Instrument settings were as follows: capillary voltage 3.0 kV for positive mode and 2.0 kV for negative mode, cone voltage 40 V for positive mode and 12 V for negative mode, source offset 30 V, source temperature 120 °C, desolvation gas temperature 350 °C, desolvation gas flow rate 650 L h^{-1}, nebulizer gas flow 4.0 bar, trap cell voltage 4 V (default in MS mode), transfer cell voltage 2 V (default in MS mode), EDC delay coefficient 1.41 V, helium cell gas flow rate 180 mL min^{-1}, IMS gas (N_2) flow rate 90 mL min^{-1}, trap DC entrance 3.0 V, bias 45.0 V, trap DC 0.0 V, exit -6.0 V, IMS DC settings entrance 25.0 V, helium cell DC 50.0 V, helium exit -10.0 V, bias 7.0 V, exit 0.0 V, transfer DC entrance 5.0 V, exit 15.0 V, trap wave velocity 311 m s^{-1}, wave height 4.0 V, IMS wave velocity 650 m s^{-1}, wave

height 40.0 V, transfer wave velocity 175 m s^{-1}, wave height 4.0 V, StepWave 1 wave velocity 300 m s^{-1}, wave height 15.0 V, StepWave 2 wave velocity 300 m s^{-1}, wave height 15.0 V, StepWave 2 offset 25.0 V, StepWave differential aperture1 3.0 V, StepWave differential aperture2 0.0 V, source ion guide wave velocity 300 m s^{-1}, wave height 1.0 V, StepWave RF offset 300.0 V, ion guide RF offset 350.0 V, IMS wave delay 1000 µs. IM resolution was ~40 (FWHM). IM cell pressure was ~3.15 mbar. The instrument was calibrated in the range of *m/z* 50–1200 using a 0.5 mM sodium formate solution prepared in 90:10 2-propanol/water v/v. The injection volume was 5 µL using partial-loop needle overfill mode. The flow rate was set to 0.05 mL min^{-1} with methanol/water (50:50 v/v) as the mobile phase. The duration of each flow injection run was 3 min. Mass spectra were acquired in profile mode over the range of *m/z* 50–1200 in the "mobility-TOF" resolution mode. During data acquisition, spectra were collected with a scan time of 1 s and corrected using a 2 ng μL^{-1} leucine enkephalin reference spray infused at 2 µL min^{-1}. Data acquisition was carried out using MassLynx ver. 4.1 (Waters Corp., Milford, MA, USA). A poly-DL-alanine solution was used as CCS reference in negative ion mode (10 mg L^{-1} in 50:50 v/v acetonitrile/water). Calibration was performed using singly charged oligomers from n = 3 to 14, covering a mass range from 230 to 1012 Da and a CCS range from 150 to 308 $Å^2$. CCS values were derived using previously described procedures.[52]

3.4.6 Data Analysis

The general workflow for the pilot study is shown in Figure 3.1. Spectral features (retention time (R_t), *m/z* pairs) were extracted from UPLC-MS data using Progenesis QI

version 2.0 (Nonlinear Dynamics, Waters Corp.). The procedure included Rt alignment, peak picking, integration, and deconvolution to group together adducts derived from the same compound. Subsequently, *m/z* values of all extracted features were input into the Metlin database[53] to perform a broad search for chemical compound candidates with an error window of 20 ppm, and 16.7% features with no candidates in the database were removed from the list. The remaining features were normalized after blank subtraction. Further, only features that were present in at least 50% of one group class were retained. These were subject to a more stringent search against the Human Metabolome Database (HMDB)[54] using the elemental formula of the compound candidates in Metlin, and only those features that had candidates with endogenous human or microbial origins were retained. The remaining features were further confirmed by MS/MS experiments. The feature matrix obtained after this procedure was utilized to build models for sample discrimination *via* oPLS-DA[55, 56] by comparing the sample classes pairwise (MATLAB, R2015a, The MathWorks, Natick, MA with PLS-Toolbox, version 8.0, Eigenvector Research, Inc., Manson, WA). Reverse interval PLS-DA (iPLS-DA) was applied to autoscaled feature abundances to find the optimum number of latent variables (LVs) and a feature panel that maximized classification accuracy. The iPLS-DA interval size was set to 1 and the maximum number of LVs set to 6. Leave-one-out cross-validation (LOOCV) and contiguous block cross-validation (CV) with 3 data splits were used for oPLS-DA model building. Permutation tests were performed to validate the models and prevent over-fitting.

For the large cohort study, the procedures were modified to improve data analysis efficiency due to the higher number of features extracted (Figure 3.2). In addition,

samples were separated into adult and pediatric groups based on an age cut-off of 18 for binary classifications, to better account for the wider age distribution and possible age-related metabolic differences. After data pre-processing, quality control sample-based robust LOESS (locally estimated scatterplot smoothing) signal correction (QC-RLSC) was applied, followed by quality assurance check, in which features with a relative standard deviation (RSD) > 30% in QC samples were removed.[46] Subsequently, signals from sample blanks, contaminants, chloride salt clusters and high mass defect ions were removed, and the data was normalized by total ion intensity. Then, data matrices were subject to a 50% sample prevalence filter and a putative identity filter search against the HMDB[54] using accurate mass, and only those that had candidates with endogenous human or microbial origins were kept. The samples in each binary class comparison were separated into pediatric and adult groups. After deletion of outliers in each age group for each comparison, samples were age-matched, and the abundances of the remaining features in these samples autoscaled and investigated by reverse iPLS-DA, followed by oPLS-DA classification with three-block CV. Further validation was performed by rigorous permutation testing.[57]

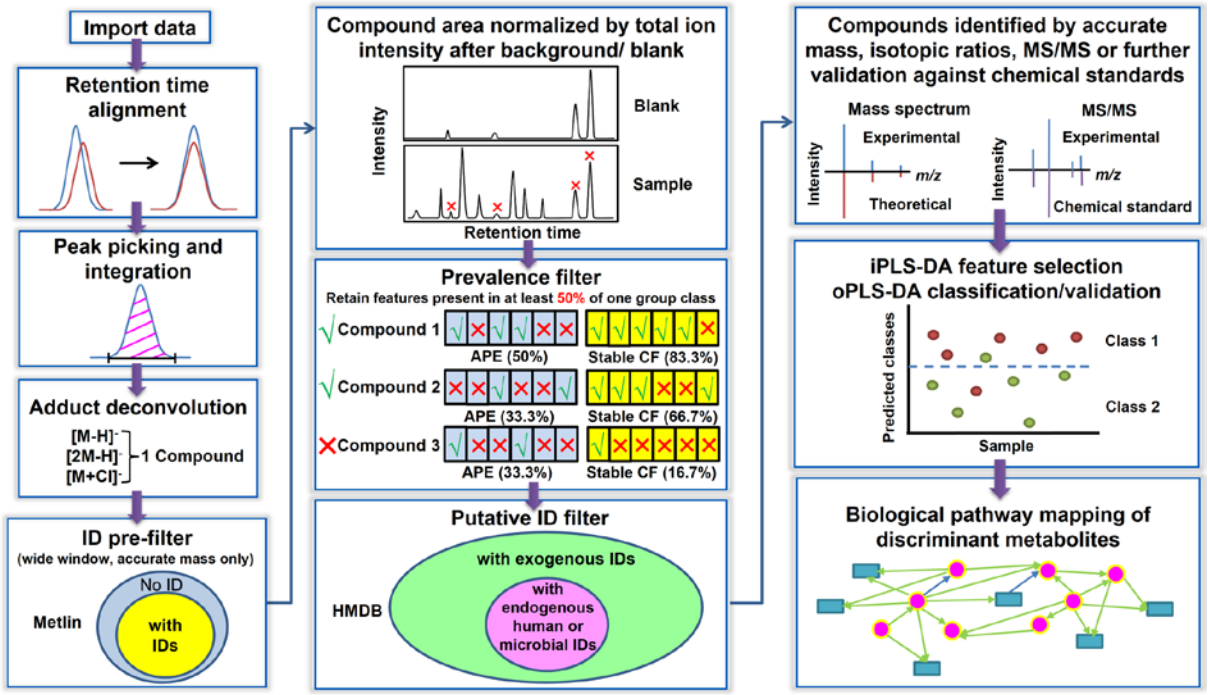

Figure 3.1: Data analysis workflow for the pilot study.

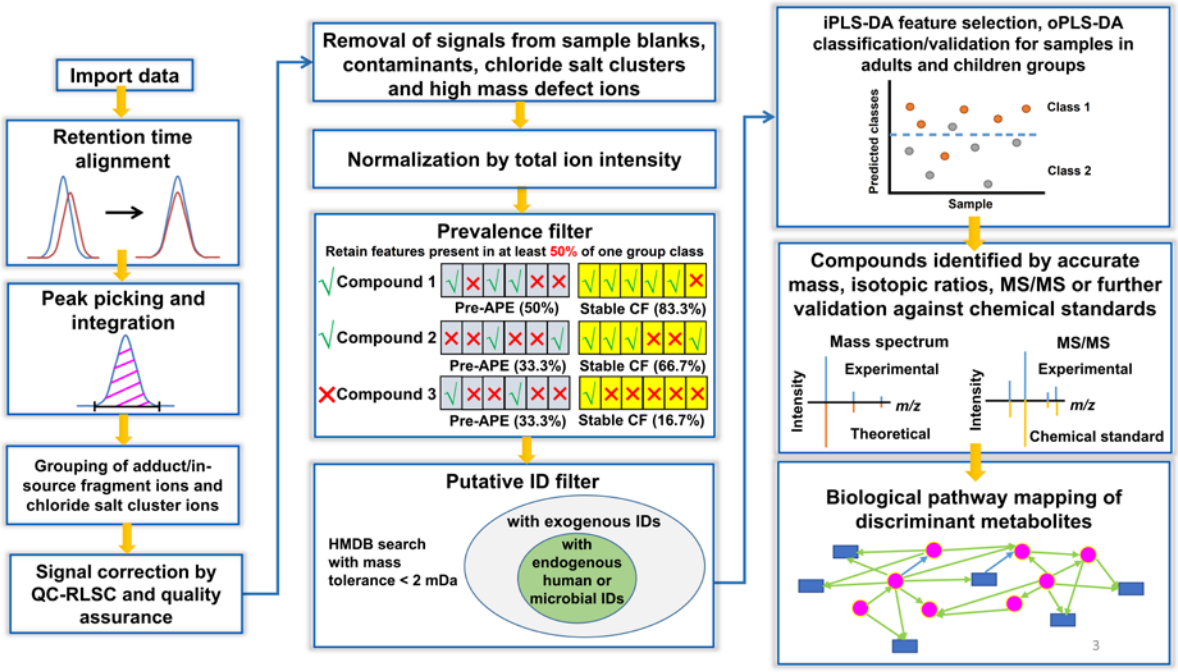

Figure 3.2: Data analysis workflow for the large cohort study. Main differences with the pilot study include application of signal correction by QC-RLSC and quality assurance and identification of discriminant features after feature selection and modeling.

3.4.7 Metabolite Identification and Pathway Analysis

Spectral features with tentative candidates in the HMDB were targeted for identification based on i) the accurate mass and isotopic pattern, ii) tandem MS experiments where the respective precursor ions were quadrupole selected, and iii) further validation against chemical standards (when available). For those cases in which MS/MS spectra were not available in the Metlin database, fragmentation patterns were manually interpreted for metabolite annotation. Commercially available standards were analyzed under identical conditions as EBC samples to validate putative metabolite identities by chromatographic Rt matching and MS/MS fragmentation pattern matching.

In the large cohort study, for metabolites with Rt at the solvent front, CCS values were measured using ion mobility (IM) MS to provide an additional molecular descriptor to increase identification confidence by comparing to those of chemical standards. Pathway analysis was performed by Metaboanalyst v3.0.[58] A total of 23 uniquely identified discriminant metabolites from all panels in the large cohort study were chosen for analysis. Data were autoscaled, with other parameters kept as default.

3.5 Results and Discussion for the Pilot Study

3.5.1 Data Processing

A total of 491 features were extracted from UPLC-MS negative ion mode data of the entire sample cohort by Progenesis software. Following Metlin filtering, 409 spectral features were kept. After deleting contaminant compounds such as known surfactants or plasticizers, background subtraction was applied to remove features in EBC samples that were also present in the sample blanks. If a feature had a maximum peak area in the blank runs which was one-third or more of the peak area of the same feature in EBC samples, it was considered a contaminant[59] and its peak area in the corresponding EBC sample was set to 0. Otherwise, the maximum peak area in the blank samples was deducted from the feature peak areas in the EBC samples. Following background subtraction, features that had zero peak areas in 70% or more of the EBC samples from the studied class pairs were removed, resulting in 176 features that remained in the APE/stable CF class pair and 185 features in the pre-APE/stable CF class pair. This step was aimed at pruning out less significant groups of features. Of these, only features present in at least half of any group

class were selected to increase the robustness of the final marker panel, leaving 144 features for the APE/stable CF classes, and 159 features for the pre-APE/stable CF classes, when considered pairwise. As described in Figure 3.1, the feature datasets were further filtered to keep only those that also had tentative identities based on elemental formula searches in the HMDB. Following this filtering, 20 features remained in the APE/stable CF class pair and 21 features in the pre-APE/stable CF class pair, and of these only 9 and 10, respectively, could be confirmed by MS/MS experiments (Table 3.2) and were then subject to iPLS-DA feature selection process, as described in the next section. This rather stringent filtering approach was chosen to ensure that the features used for multivariate classification had a high certainty in terms of biochemical identity, and therefore improve the chances of understanding their significance in the context of CF APE pathophysiology.

Table 3.2: Chemical identification of features in EBC with tentative identities in the HMDB.

Feature code	Used as part of starting feature set in models	Retention time (R_t) (min)	Experimental *m/z*	Ion type	Elemental formula	Δm (mDa)	Tentative metabolite annotation	Method for tentative annotation	Metabolite ID validation (with standard)
397	(1), (2)	0.82	143.0701	$[M-H]^-$	$C_7H_{12}O_3$	-0.7	4-Hydroxycyclohexylcarboxylic acid	[a]	R_t, MS/MS match
40	(1), (2)	0.48	89.0231	$[M-H]^-$	$C_3H_6O_3$	-0.8	Lactic acid	[a], MS/MS[b]	R_t, MS/MS match
407	(1), (2)	0.48	128.0343	$[M-H]^-$	$C_5H_7NO_3$	-0.5	Pyroglutamic acid (5-oxoproline)	[a], MS/MS[b]	R_t, MS/MS match
186	(1), (2)	0.51	159.0640	$[M-H]^-$	$C_7H_{12}O_4$	-1.7	3,3-Dimethylglutaric acid; Pimelic acid	[a]	R_t[d]
403	(1), (2)	0.54	73.0283	$[M-H]^-$	$C_3H_6O_2$	-0.7	Propionic acid; Hydroxyacetone; Lactaldehyde	[a]	R_t[d]
44	(1), (2)	0.65	137.0232	$[M-H]^-$	$C_7H_6O_3$	-0.7	3-Hydroxybenzoic acid; 4-Hydroxybenzoic acid; Salicylic acid	[a], MS/MS[b/c]	--
388	(1), (2)	1.04	101.0593	$[M-H]^-$	$C_5H_{10}O_2$	-1.0	Ethylmethylacetic acid; Isovaleric acid; Valeric acid	[a]	R_t[d]
52	(1), (2)	1.04	187.0970	$[M-H]^-$	$C_9H_{16}O_4$	0.0	Nonanedioic acid/Azelaic acid	[a], MS/MS[b]	--
385	(1)	1.39	115.0748	$[M-H]^-$	$C_6H_{12}O_2$	-1.1	Isocaproic acid	[a]	R_t
156	(1), (2)	14.65	225.1848	$[M-H]^-$	$C_{14}H_{26}O_2$	-0.7	Myristoleic acid	[a]	R_t

(1) APE/stable CF; (2) pre-APE/stable CF; [a] accurate mass and isotopic pattern matched; [b] Metlin database matched; [c] manual fragmentation analysis; [d] all candidates showed similar R_t.

3.5.2 Classification Performance

Results for the discrimination of 9 APE EBC samples *vs.* 17 stable CF EBC samples are shown in Table 3.3 and Figure 3.3 (A). An optimum panel of 2 discriminant features (panel #1) was selected through the iterative iPLS-DA process. Following a three-block CV approach, the corresponding oPLS-DA model yielded a classification accuracy of 84.6%, a sensitivity of 77.8% and a specificity of 88.2%. One latent variable was used to build the oPLS-DA model that interpreted 44.7 and 36.3% variance from the X- (feature peak areas) and Y- (EBC class membership) blocks, respectively. Two EBC samples from the stable CF patient class and two samples from the APE class were systematically misclassified. Figures 3.3 (B) and (C) show box plots of peak areas for each discriminant feature in panel #1, and denote fold changes obtained between the compared sample classes. Median instead of mean peak area values were used to calculate fold changes to account for sample variability resulting from the relatively small sample size used in this class comparison. Interestingly, when the 4 pre-APE samples were used as an unknown sample set and input into this classification model, 3 out of 4 pre-APE samples were predicted as being similar to APE samples (Figure 3.3 (A)), foreshadowing a metabolic fingerprint of APEs in the pre-APE EBC samples. Conversely, when the 5 post-APE samples were input into the APE vs. stable CF oPLS-DA model, 4 out of 5 post-APE samples were predicted as being like stable CF samples (Figure 3.3 (A)), possibly suggesting that following APE treatment, the EBC metabolic profiles of the discriminant features of most post-APE patients resembled those of the stable CF patients.

Table 3.3: Comparison of discriminant feature panels.

Model /Panel #	Classes compared	Type of CV	No. of features in initial set	No. of discriminant features in oPLS-DA model	Discriminant feature codes	Model accuracy	Model specificity	Model sensitivity
1	APE (9); Stable CF (17)	3-block	9	2	397, 407	84.6	88.2	77.8
2	Pre-APE (4); Stable CF (17)	LOOCV	10	2	40, 407	90.5	94.1	75.0

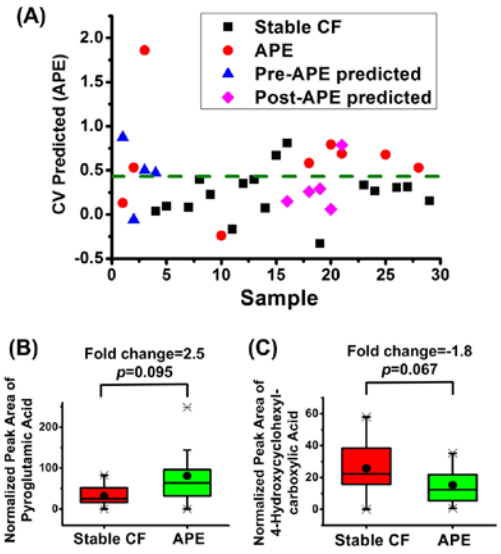

Figure 3.3: (A) oPLS-DA cross-validated classification plot and (B) Box plots of peak areas of each discriminant feature in panel #1. (A) The *x*-axis represents sample number, and *y*-axis represents the cross-validated predicted scores of the oPLS-DA classification model. APE and stable CF samples are represented by red circles and black squares, respectively. The pre- and post-APE samples projected into the model are represented by blue triangles and magenta diamonds, respectively. The threshold for sample classification is represented by the green dashed line. "Pre-APE" samples are collected within 3 months before an APE event; "APE" samples are collected during an APE event; "post-APE" samples are collected within 3 months after an APE event; "stable CF" samples are collected for stable CF patients (3 months before or after collection there were no APE events). (B, C) Box plots for pyroglutamic acid and 4-hydroxycyclohexylcarboxylic acid, respectively, in EBC samples from stable CF and APE patient groups. Mean values are represented by a filled circle in the box; median values are represented by a line in the box; the edges of the box are 25th and 75th percentiles; the whisker extends to the most extreme values in data not including outliers, with a 99.3% coverage; outliers are represented by asterisks. The positive fold change is calculated as the ratio of median peak areas between APE and stable CF samples. The negative fold change is calculated as the negative value of the ratio of median peak areas between stable CF and APE samples. *p* values are calculated from Wilcoxon rank-sum test.

With the purpose of investigating if discriminating the 4 EBC samples collected from patients in a pre-APE state from those 17 EBC samples from stable CF patients was possible, a new oPLS-DA model was developed using a leave-one-out cross-validation (LOOCV) approach. This approach was chosen due to the small number of pre-APE samples. Table 3.3 shows that using a two-feature discriminant metabolite panel (panel #2) selected by iPLS-DA, classification was indeed possible with an accuracy of 90.5%, a sensitivity of 75.0% and a specificity of 94.1% (Figure 3.4 (A)). One EBC sample from a pre-APE patient and one sample from a stable CF patient were misclassified with this model, which used 1 latent variable and interpreted 87.2 and 44.6% variance from the X- and Y-blocks, respectively. Figures 3.4 (B) and (C) show box plots for each discriminant feature in panel #2, with the respective median fold changes obtained for the compared sample classes. Interestingly, feature #407 was common to panels #1 and #2, but feature #40 was only selected in panel #2, suggesting that biomarkers of the asymptomatic phase preceding an APE event may be somewhat different from those associated with biochemical processes occurring during an exacerbation. Overall, these results highlight the feasibility of early indication of an oncoming APE event using these small metabolite panels (Tables 3.3 and 3.4), a possibility that could have significant implications in terms of pre-emptive APE diagnostics, enabling detection and treatment before irreversible damage to lung function occurs. Although CV could prevent over-fitting to some extent, we further performed permutation tests to validate the robustness of the models. A pairwise Wilcoxon signed rank test was chosen for the cross-validated residuals since the population couldn’t be assumed to be normally distributed due to the small sample size. For the model classifying pre-APE/stable CF class pair, the probability that the un-

permuted model was not significantly different from the permuted models was 0.024, indicating that the original model was significant and not over-fitting at the 95% confidence level. For the model classifying APE/stable CF class pair, the probability that the un-permuted and permuted models were indistinguishable was 0.087, indicating that the model was significant at the 90% confidence level.

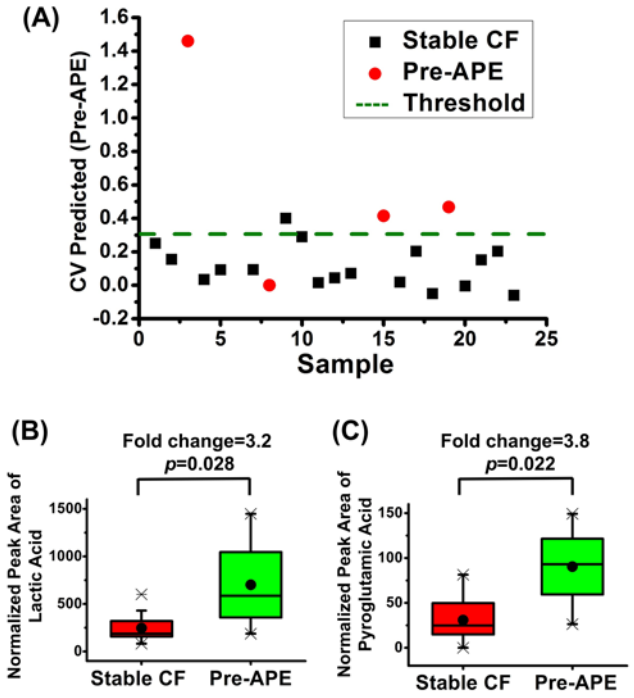

Figure 3.4: (A) oPLS-DA cross-validated classification plot and (B) Box plots of peak areas of each discriminant feature in panel #2. (A) The *x*-axis represents sample number, and *y*-axis represents the cross-validated predicted scores by the oPLS-DA classification model. Pre-APE and stable CF samples are represented by red circles and black squares, respectively. The threshold for sample classification is represented by a green dashed line. (B, C) Box plots for lactic acid and pyroglutamic acid, respectively, in EBC samples from stable CF and pre-APE patient groups. Mean values are represented by a filled circle in the box; median values are represented by a line in the box; the edges of the box are 25th and 75th percentiles; the whisker extends to the most extreme values in data not including outliers, with a 99.3% coverage; outliers are represented by asterisks. Fold changes are calculated as the ratio of median peak areas between pre-APE and stable CF samples. *p* values are calculated from Wilcoxon rank-sum test.

3.5.3 Identification of Discriminant Metabolites and Their Biological Roles

Figure 3.5 displays the procedure used for unambiguous chemical identification of the discriminant features used in the oPLS-DA panels, using feature #40 as an example. Metabolic fingerprints from the same patient at 3 different CF states: pre-APE, stable, and during an APE event are illustrated with the corresponding base peak intensity chromatograms (BPI) displayed in Figures 3.5 (A), (B) and (C), respectively. Extracted ion chromatograms (EICs) (Figure 3.5 (D)) and corresponding mass spectra (Figure 3.5 (E) top) were obtained for features selected by iPLS-DA. According to the experimental monoisotopic mass of each feature, a series of possible candidates were generated from database searches in Metlin database followed by the HMDB, and selected after matching of the observed and theoretical isotopic patterns (Figure 3.5 (E)). Next, fragmentation patterns obtained from MS/MS experiments were compared to MS/MS spectra of the candidates in Metlin, if available, or interpreted manually. Finally, the tentatively identified metabolites were confirmed by matching Rt and fragmentation pattern to those of chemical standards, whenever possible (Figure 3.5 (D, F)). Feature #40 in the example was identified as lactic acid. The identities of the discriminant features in the models are summarized in Table 3.4.

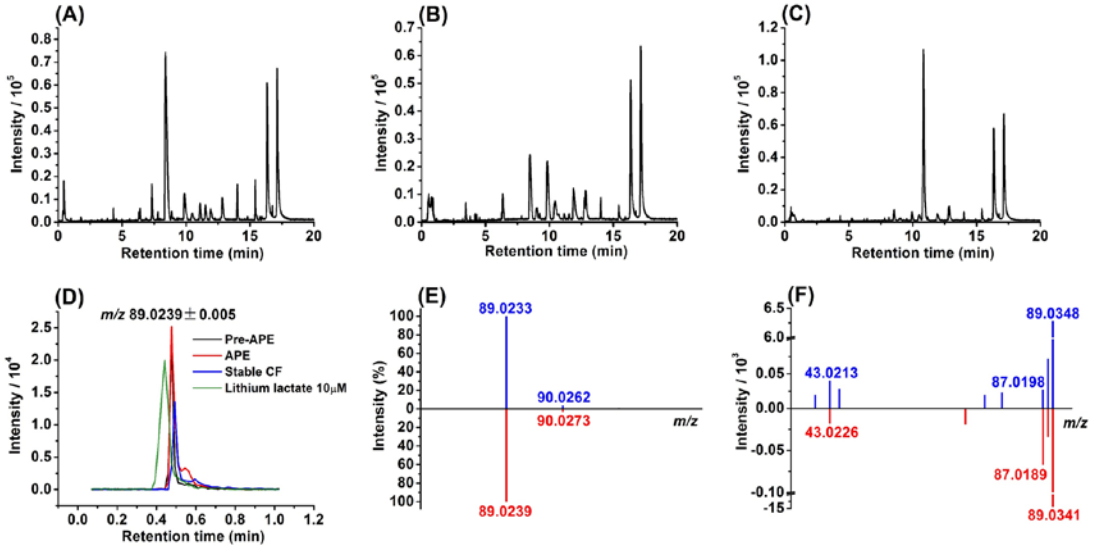

Figure 3.5: Base peak intensity (BPI) chromatograms obtained for EBC samples from the same patient at 3 different CF states: (A) pre-APE, (B) stable CF, and (C) during an APE event. (D) Extracted ion chromatogram for the discriminant feature with *m/z* 89.0239 ± 0.005 (lactic acid) generated from data in (A), (B), (C) and lithium lactate standard. (E) Experimental (top) and theoretical (bottom) mass spectra for the discriminant feature with *m/z* 89.0239 and R_t= 0.48 min. (F) MS/MS spectrum for *m/z* 89.0239 precursor ion using a collision cell voltage of 8V and a sampling cone voltage of 30V. The matching of MS/MS fragmentation between the experimental spectrum (top) and the chemical standard (bottom) is shown.

Table 3.4: Chemical identification of discriminant EBC features in the pilot study.

Feature code	Used in model/ panel	Rt (min)	Experimental *m/z*	Ion type	Elemental formula	Δm (mDa)	Tentative annotation	Method for tentative annotation	Metabolite ID validation (with standard)
40	**2**	0.48	89.0231	$[M-H]^-$	$C_3H_6O_3$	-0.8	Lactic acid	[a], MS/MS[b]	R_t, MS/MS match
407	**1,2**	0.48	128.0343	$[M-H]^-$	$C_5H_7NO_3$	-0.5	Pyroglutamic acid (5-Oxoproline)	[a], MS/MS[b]	R_t, MS/MS match
397	**1**	0.82	143.0701	$[M-H]^-$	$C_7H_{12}O_3$	-0.7	4-Hydroxycyclohexyl -carboxylic acid	[a]	R_t, MS/MS match

[a] Accurate mass and isotopic pattern matched; [b] Metlin database matched.

Feature # 40 was selected by iPLS-DA for the model comparing pre-APE patient samples with stable CF ones, and it was identified as lactic acid. Lactic acid had a significant median fold increase of 3.2 from stable CF to pre-APE patient samples (p<0.05, Figure 3.4 (B)). Interestingly, a previous study using NMR reported increased levels of lactate in bronchoalveolar lavage fluid (BALF) from CF patients with high inflammation compared to those with low inflammation,[60] confirming that increased inflammation both prior to and during an APE event could be also detected in EBC by LC-MS. Increased lactate production has also been reported in patients with respiratory distress syndrome, finding it proportional to lung injury severity.[61] Lactic acid levels in the studied cohort possibly reflect the status of different stages preceding and following an APE event (Figure 3.6 (A)). The higher levels of lactic acid in the pre-APE and APE patients compared to stable CF patients could possibly result from the increasingly hypoxic environment in CF lungs due to poorly cleared thick mucus developing on epithelial surfaces,[62] known to lead to an increased lactate conversion from pyruvate in anaerobic glycolysis. Lactate is also a glucose precursor in gluconeogenesis, and elevated gluconeogenesis has been found in CF-related diabetes, possibly contributing to abnormal glucose tolerance in CF.[63] In addition, lactic acid as a fermentation product of the CF anaerobic bacteria, was suggested as a potential biomarker for CF progression.[64] The decreasing trend of lactic acid from the pre-APE to the APE group in the studied cohort might be understood by considering that APE patients were treated with intravenous antibiotic therapy, so their inflammatory phenotype might be different from pre-APE patients, who had not yet been aggressively treated.

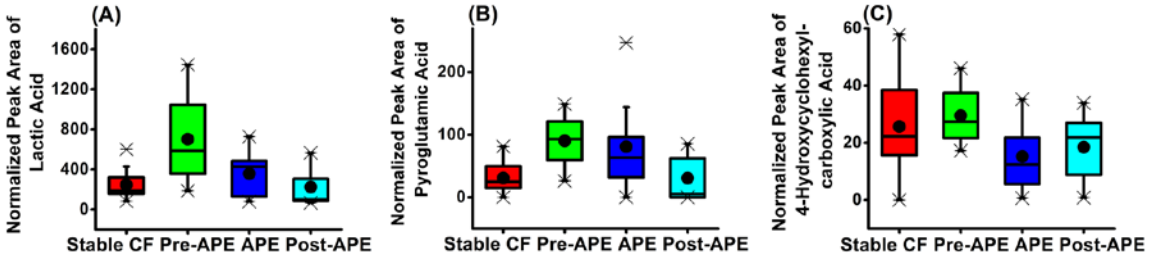

Figure 3.6: Box plots for the three discriminant metabolites: (A) lactic acid, (B) pyroglutamic acid, and (C) 4-hydroxycyclohexylcarboxylic acid in different subgroups of the sample cohort. Mean values are represented by a filled circle in the box; median values are represented by a line in the box; the edges of the box are 25th and 75th percentiles; the whisker extends to the most extreme values in data not including outliers, with a 99.3% coverage; outliers are represented by asterisks.

Feature #407, identified as pyroglutamic acid (5-oxoproline), was present in both marker panels (Table 3.4), suggesting that its relative alterations may reflect processes occurring both during APE as well as during the 3-months-time window preceding the APE episode. Pyroglutamic acid had median fold increases from stable CF to APE and to pre-APE samples of 2.5 and 3.8, respectively (Figure 3.3 (B) and Figure 3.4 (C)). Interestingly, in a recent serum metabolomics study of 31 CF vs. 31 non-CF children reported by Joseloff *et al.*, pyroglutamic acid was also identified as an important metabolite responsible for discrimination between CF and non-CF subjects.[65] This metabolite is a known intermediate in the γ-glutamyl cycle, a pathway for the biosynthesis and degradation of glutathione, and is thus related to redox imbalance. CF mutations cause a primary dysfunction in the glutathione system, leading to a systematic

glutathione deficiency in the respiratory epithelial lining fluid, which is aggravated by oxidative burden.[66, 67] Interestingly, decreased levels of glutathione have also been detected during exacerbations in EBC of children with asthma, hinting at some common mechanisms.[68] Figure 3.6 (B) illustrates the relative concentrations of pyroglutamic acid in the different sample classes.

Feature #397 was selected by iPLS-DA for the model classifying APE from stable CF EBC samples, with a median fold decrease of 1.8 from stable CF to APE samples (Figure 3.3 (C)). This feature was identified by both UPLC-MS and MS/MS, and validated with a standard as 4-hydroxycyclohexylcarboxylic acid, which is a relatively rare organic acid involved in gut microbial mammalian cometabolism. It is a metabolite typically found in urine,[69, 70] but never before reported in EBC to our knowledge. Interestingly, this type of metabolic gut-lung crosstalk has been found to be also associated with inflammatory bowel diseases, in which the pulmonary inflammation is reported to accompany the main inflammatory processes in the bowel.[71] It is yet unclear, however, if these inflammatory processes are manifested through similar alterations in the respective lung and bowel metabolomes, as suggested by this finding. Figure 3.6 (C) illustrates the relative concentrations of 4-hydroxycyclohexylcarboxylic acid in the different sample classes.

3.6 Results and Discussion for the Large Cohort Study

3.6.1 Data Processing Results

A total of 1334 and 923 features were obtained from UPLC-MS data in negative and positive ion modes, respectively. Adducts, in-source fragment ions, and chloride salt cluster ions were grouped according to criteria in Table 3.5. QC-RLSC was applied to the data, followed by removal of features with RSD >30% in quality control (QC) samples.[46] After blank correction, features with zero values in >80% of samples considering all classes (stable CF, pre-APE, APE and post-APE) were removed, leaving 1015 and 635 features in negative and positive ion mode data, respectively. Subsequently, contaminants and ESI artifacts such as chloride salt cluster ions and high mass defect ions defined by McMillan filter[72] were removed, leaving 745 and 538 features in negative and positive ion mode data, respectively. The resulting data matrices were normalized by total ion intensity and then passed through a 50% sample prevalence filter in each class, yielding 281 and 262 features in negative and positive ion mode data for APE/stable CF classes, and 278 and 268 features in negative and positive ion mode data for pre-APE/stable CF classes. The remaining features were then searched in the HMDB and only those with endogenous human or microbial candidate IDs were retained. Following this step, 58 and 46 compounds remained in negative ion mode and positive ion mode data for the APE/stable CF class pair, and 57 and 44 compounds for the pre-APE/stable CF class pair. These remaining features constituted the final datasets used for subsequent multivariate analyses. PCA was performed for each set, and outliers outside 95% confidence intervals were removed. Then, iPLS-DA feature selection and oPLS-DA binary classifications were performed for age-matched samples in each age-based group (Table 3.1).

Table 3.5: Criteria applied for grouping adduct/in-source fragment ions, and chloride salt clusters in the large cohort UPLC-MS dataset.

Criterion	Δm (mDa)	ΔRt (min)	Specific rules	Numbers of compounds or features
Adduct/in-source fragment ion grouping[a]. Adduct/in-source fragment ions considered: Negative mode: adduct ions: $[M-H]^-$, $[M+Na-2H]^-$; in-source fragment ions: $[M-H_2O-H]^-$, $[M-SO_3-H]^-$, $[M-CO_2-H]^-$, $[M-NH_3-H]^-$, $[M-CH_3COOH-H]^-$, $[M-HCOOH-H]^-$ Positive mode: adduct ions: $[M+H]^+$, $[M+Na]^+$; in-source fragment ions: $[M-H_2O+H]^+$, $[M-SO_3+H]^+$, $[M-CO_2+H]^+$, $[M-NH_3+H]^+$, $[M-CH_3COOH+H]^+$, $[M-HCOOH+H]^+$	1.5	0.1	1. $[M-H]^-$ or $[M+H]^+$ ion must be present in group in negative or positive mode, respectively. $[M-H]^-$ or $[M+H]^+$ ion must be most abundant ion for >90% samples. 2. Only singly charged compounds are considered. 3. Either all ion adducts in a group should be present, or all should be absent for >90% of samples. 4. Correlation between features[b] >0.85 to ensure they correspond to different adducts of the same compound.[73]	Twenty-six features were grouped into 11 compounds in negative mode data and 22 peaks were grouped into 11 compounds in positive mode data, based on the presence of multiple adducts or in-source fragments.
Manual grouping or assigning of all chloride salt cluster isotopic signals not correctly grouped by Progenesis. Possible isotopes considered: X, X+1.9971, X+1.9971×2, X+1.9971×3, X+1.9971×4, X+1.9971×5.	4	0.1	1. Ratio between the second largest isotope abundance and the most abundant isotope >0.3. 2. Mass defect >0.3. 3. Satisfy rules 2-4 in adduct/in-source fragment ion grouping criteria.	Ten chloride-containing species were grouped into 3 chloride salt cluster compounds in negative mode, and 2 chloride-containing species were grouped into 1 chloride salt cluster compound in positive mode. All species overlapped with features filtered by the McMillan filter[72].
Identification of cases where Progenesis only identified a single isotopic peak from a chloride salt cluster.	N/A	N/A	1. Mass defect >0.3. 2. *m/z* match theoretical value of known chloride salt cluster ions.	Four sodium chloride cluster ions ($[Na_nCl_{n+1}]^-$ (n=2–5)), and 2 iron chloride cluster ions ($[FeCl_3]^-$ and $[FeCl_4]^-$) in negative mode; 1 sodium chloride cluster ion of $[Na_3Cl_2]^+$ in positive mode.

3.6.2 Classification Performance

Data analysis of negative ion mode data showed that discrimination of APE from stable CF samples was possible for age-matched samples in both adult and pediatric groups (Table 3.6, Figure 3.7). For classification of pediatric patients, 9 discriminant features were selected by iPLS-DA, yielding a sensitivity of 83.3%, specificity of 91.7% and accuracy of 88.9%. For the adult group, the oPLS-DA model provided a sensitivity of 76.2%, specificity of 83.7%, and accuracy of 81.3% in distinguishing 21 APE samples from 43 stable CF samples, using a panel of 10 discriminant features with 6 latent variables. For discrimination of pre-APE samples from stable CF EBC samples, another set of oPLS-DA models was built that provided sensitivities of 85.7% and 89.5%, specificities of 88.4% and 84.1%, and accuracies of 87.7% and 85.7% for differentiating samples from 14 pre-APE pediatric patients or 19 adult patients from 43 stable pediatric patients or 44 adult patients. All models had good AUC values ranging from 0.8 to 0.9, with permutation testing showing no over-fitting.

Table 3.6: Comparison of oPLS-DA models using negative ion mode data.

Classes compared	Age group	No. of features in initial set	No. of discriminant features in oPLS-DA model	Model accuracy	Model specificity	Model sensitivity	AUC (CV)	Percent variance X1, X-block, Y-block	Cross-validated Wilcoxon signed rank test *p* value for permutation test
APE (18); Stable CF (36)	Pediatric	58	9	88.9	91.7	83.3	0.8873	9.2, 59.4, 58.4	0.0010
APE (21); Stable CF (43)	Adult	58	10	81.3	83.7	76.2	0.7940	12.0, 71.6, 42.3	0.0060
Pre-APE (14); Stable CF (43)	Pediatric	57	11	87.7	88.4	85.7	0.8787	10.5, 43.1, 42.1	0.013
Pre-APE (19); Stable CF (44)	Adult	57	5	85.7	84.1	89.5	0.8301	31.2, 31.2, 30.3	0.0080

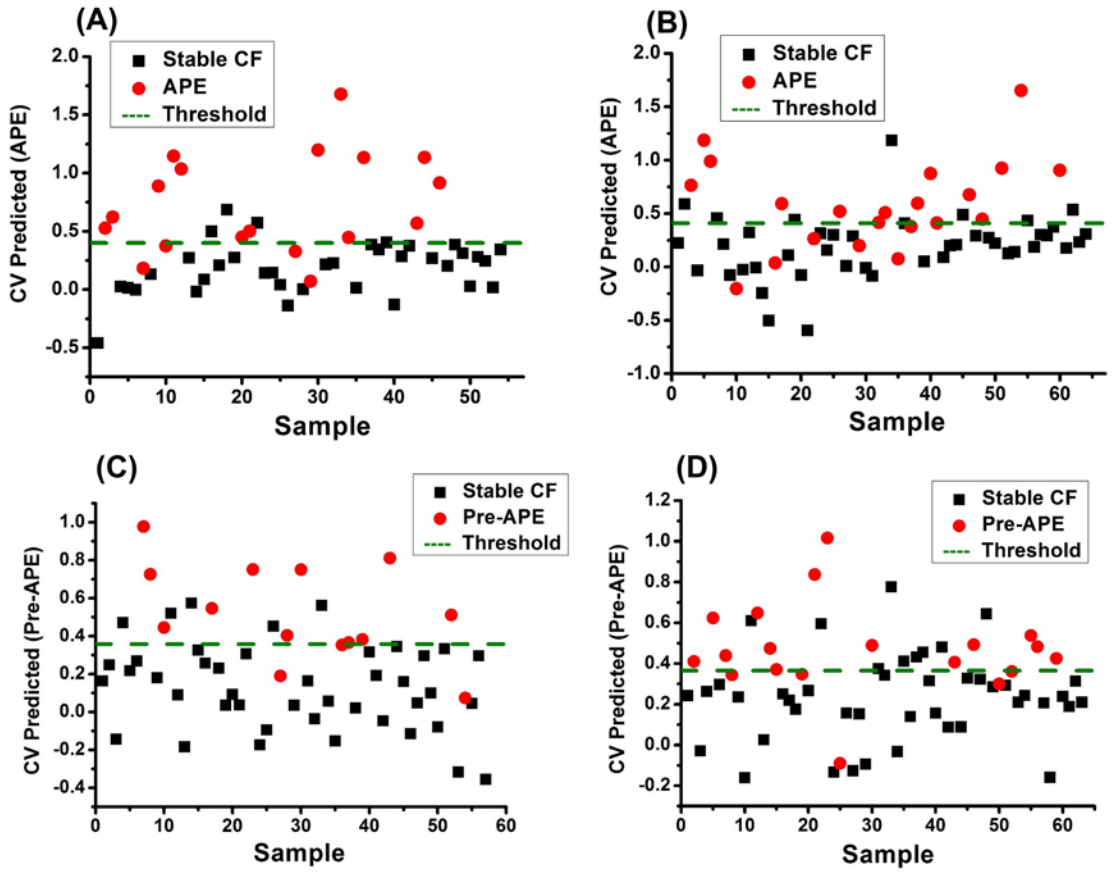

Figure 3.7: oPLS-DA cross-validated classification plots using negative ion mode data, including comparison of APE vs. stable CF samples in age- and gender-matched EBC samples from pediatric patients (A) and those from adult patients (B), and comparison of pre-APE vs. stable CF samples in age- and gender-matched EBC samples from pediatric patients (C) and those from adult patients (D). The *x*-axis represents sample number, and *y*-axis represents the cross-validated predicted scores of the oPLS-DA classification model. APE/pre-APE and stable CF samples are represented by red circles and black squares, respectively.

Data analysis was also performed on the dataset combining positive and negative ion mode features. Results showed classification was also feasible when samples were separated into adult and pediatric groups (Table 3.7, Figure 3.8) and improved classification performance was obtained for most comparisons with this feature dataset. For the comparison of APE and stable CF classes, good classification sensitivities (87.5% and 84.2%), specificities (94.9% and 90.9%) and accuracies (92.7% and 88.9%) were achieved by oPLS-DA in discriminating APE from stable patients for pediatric and adult cohorts, respectively. For the classification between pre-APE and stable CF, oPLS-DA modeling provided good sensitivities (100% and 85.7%), specificities (87.5% and 97.7%) and accuracies (90.3% and 93.9%) for pediatric and adult groups, respectively. All models had high AUC values of approximately 0.9, and no indication of overfitting as suggested by permutation test results.

Table 3.7: Comparison of oPLS-DA models using combined negative and positive ion mode data.

Classes compared	Age group	No. of features in initial set	No. of discriminant features in oPLS-DA model	Model accuracy	Model specificity	Model sensitivity	AUC (CV)	Percent variance X1, X-block, Y-block	Cross-validated Wilcoxon signed rank test *p* value for permutation test
APE (16); Stable CF (39)	Pediatric	104	6	92.7	94.9	87.5	0.9103	13.6, 36.7, 55.9	0.0010
APE (19); Stable CF (44)	Adult	104	9	88.9	90.9	84.2	0.8911	8.6, 34.1, 49.4	0.0020
Pre-APE (14); Stable CF (48)	Pediatric	101	8	90.3	87.5	100.0	0.9330	8.9, 82.5, 47.5	0.0080
Pre-APE (21); Stable CF (44)	Adult	101	11	93.9	97.7	85.7	0.9075	14.9, 39.2, 59.6	0.0

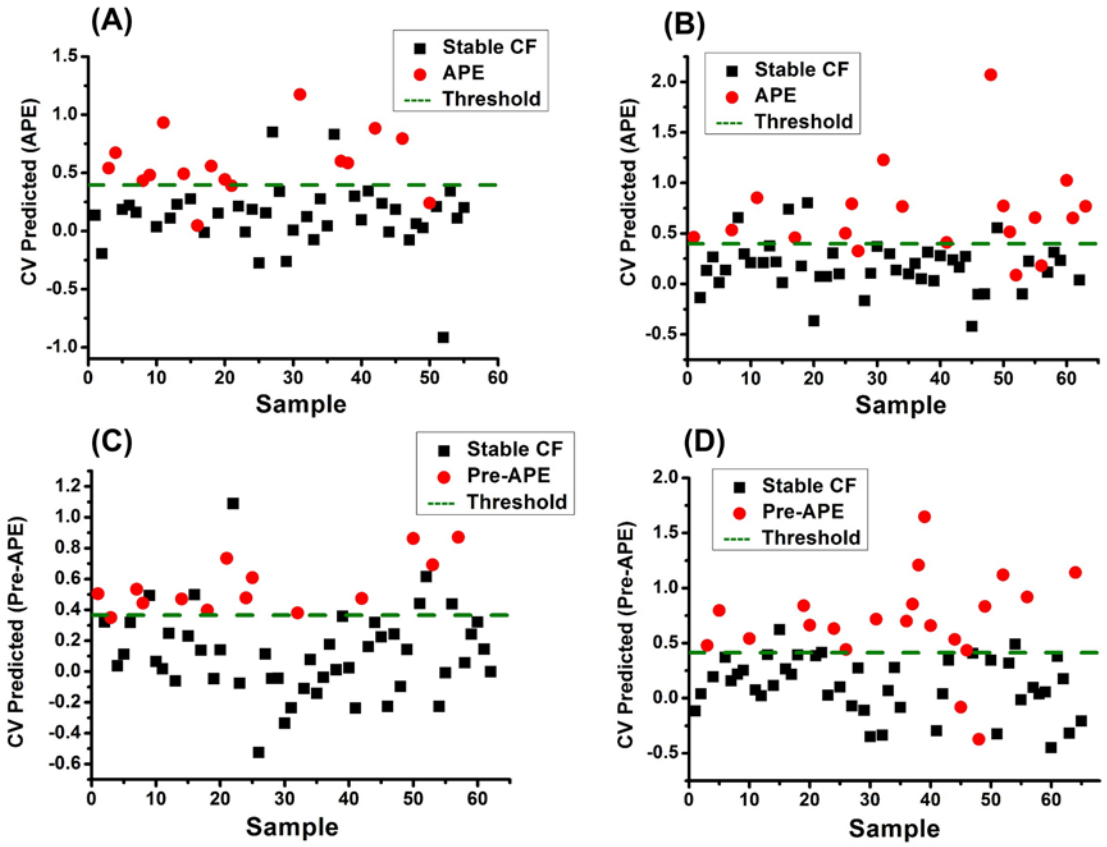

Figure 3.8: oPLS-DA cross-validated classification plots using combined positive and negative ion mode data, including comparison of APE vs. stable CF samples in age- and gender-matched EBC samples from pediatric patients (A) and those from adult patients (B), and comparison of pre-APE vs. stable CF samples in age- and gender-matched EBC samples from pediatric patients (C) and those from adult patients (D). The *x*-axis represents sample number, and *y*-axis represents the cross-validated predicted scores of the oPLS-DA classification model. APE/pre-APE and stable CF samples are represented by red circles and black squares, respectively.

3.6.3 Identification of Discriminant Metabolites and Their Biological Relevance

Features in the discriminant model panels were searched in the HMDB with a mass error of 2 mDa. Experimental MS/MS spectra of discriminant features were then matched to entries in the Metlin database, or manually analyzed if no database spectra were available. In addition, if a chemical standard was available, both Rt and MS/MS of the discriminant feature were compared to it. Since some features co-eluted with the chromatographic solvent front (R_t=0.5-0.6 min), CCS was used as an additional molecular descriptor for identification instead of R_t. A more confident identity is reported if CCS values matched with chemical standards within 2% (Table 3.8.1-3.9.2). Figure 3.9 shows an example of a positive mode identification of a discriminant feature with *m/z* 175.1188 by comparison of its MS/MS spectrum to the arginine chemical standard. In addition, Rt and CCS were also matched (Table 8.2)

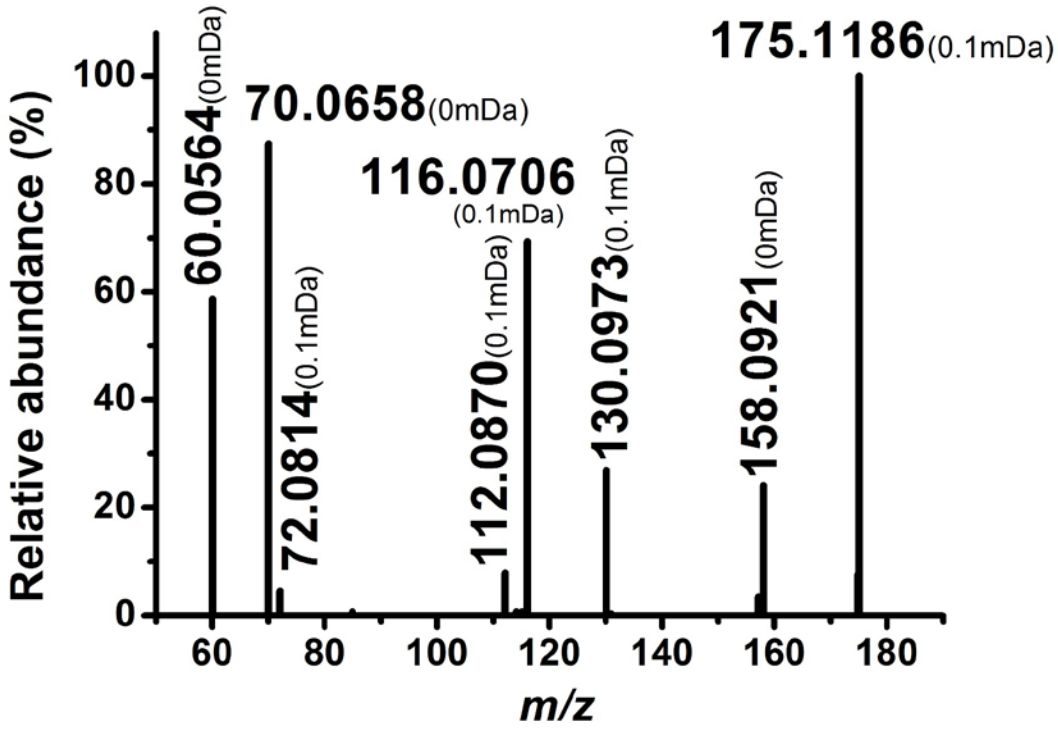

Figure 3.9: Positive ion mode MS/MS spectrum of the feature with *m/z* 175.1188 in a stable CF EBC sample. The experimental fragmentation pattern matches to that of the arginine standard, with mass differences between experimental and theoretical values shown in brackets.

Table 3.8.1: Identification of discriminant features in oPLS-DA classification of APE from stable CF samples using negative ion mode data.

Age Group	*m/z*	R_t (min)	Ion type	Tentative identity	Elemental formula	Median fold change (APE/stable CF)	Identification
Pediatric	89.0232	0.55	$[M-H]^-$	Lactic acid	$C_3H_6O_3$	2.5	CCS[a] and MS/MS[a, b]
	128.0342	0.53	$[M-H]^-$	Pyroglutamic acid	$C_5H_7NO_3$	1.5	CCS[a] and MS/MS[a, b]
	127.0503	0.87	$[M-H]^-$	Dihydrothymine	$C_5H_8N_2O_2$	3.0	MS/MS[c]
	59.0125	0.56	$[M-H]^-$	Acetic acid	$C_2H_4O_2$	-1.7	CCS[a]
	130.0499	0.57	$[M-H]^-$		$C_5H_9NO_3$	1.6	[d]
	143.0340	0.58	$[M-H]^-$	3-Methylglutaconic acid, 2-Methylglutaconic acid, 2-Hexenedioic acid, 3-Hexenedioic acid	$C_6H_8O_4$	-1.4	MS/MS[c] for 2-hexenedioic acid, MS/MS[a] for the other candidates
	227.1036	0.66	$[M-H]^-$	Prolylhydroxyproline	$C_{10}H_{16}N_2O_4$	1.8	MS/MS[c]
	165.0549	1.77	$[M-H]^-$		$C_9H_{10}O_3$	2.4	[d]
	187.0064	0.64	$[M-H]^-$		$C_7H_8O_4S$	*N/A*[1]	[d]
Adult	143.0704	0.67	$[M-H]^-$	4-Hydroxycyclohexylcarboxylic acid	$C_7H_{12}O_3$	1.5	CCS[a] and MS/MS[a, c]
	187.0970	0.63	$[M-H]^-$	Nonanedioic acid/Azelaic acid	$C_9H_{16}O_4$	2.5	CCS[a] and MS/MS[a, b]
	201.1128	1.07	$[M-H]^-$	Sebacic acid	$C_{10}H_{18}O_4$	1.6	Rt[a] and MS/MS[a, b]
	85.0283	0.61	$[M-H]^-$	γ-Butyrolactone, Oxolan-3-one	$C_4H_6O_2$	1.6	MS/MS[c]
	115.0389	0.59	$[M-H]^-$	Levulinic acid	$C_5H_8O_3$	N/A	MS/MS[a, c]
	157.0862	0.97	$[M-H]^-$	4-Hydroxycyclohexylacetic acid	$C_8H_{14}O_3$	1.5	MS/MS[c]
	59.0125	0.56	$[M-H]^-$	Acetic acid	$C_2H_4O_2$	-1.4	CCS[a]
	165.0549	1.77	$[M-H]^-$		$C_9H_{10}O_3$	3.3	[d]
	168.0658	0.73	$[M-H]^-$		$C_8H_{11}NO_3$	-1.3	[d]
	177.0397	0.53	$[M-H]^-$		$C_6H_{10}O_6$	1.6	[d]

[a] match chemical standard; [b] match Metlin MS/MS; [c] match manual analysis; [d] no endogenous ID, not [a], [b] or [c]. Significant median fold change is shown in italics. [1]Unable to calculate median fold change because the denominator is 0.

Table 3.8.2: Identification of discriminant features in oPLS-DA classification of pre-APE from stable CF samples using negative ion mode data.

Age Group	*m/z*	R_t (min)	Ion type	Tentative identity	Elemental formula	Median fold change (pre-APE/stable CF)	Identification
Pediatric	89.0232	0.55	$[M-H]^-$	Lactic acid	$C_3H_6O_3$	1.3	CCS[a] and MS/MS[a, b]
	133.0132	0.86	$[M-H]^-$	Malic acid	$C_4H_6O_5$	*-2.0×10^2*	Rt[a] and MS/MS[a, b]
	189.0762	0.59	$[M-H]^-$	3-Hydroxysuberic acid	$C_8H_{14}O_5$	1.4	MS/MS[c]
	329.2338	6.73	$[M-H]^-$	9,10,13-TriHOME	$C_{18}H_{34}O_5$	1.3	MS/MS[c]
	73.0277	0.64	$[M-H]^-$	Propionic acid	$C_3H_6O_2$	-1.1	CCS[a] and MS/MS[a, c]
	130.0499	0.57	$[M-H]^-$		$C_5H_9NO_3$	1.9	[d]
	145.0973	0.70	$[M-H]^-$	Lysine	$C_6H_{14}N_2O_2$	2.9	Rt[a] and MS/MS[a, b] for positive mode
	101.0596	0.80	$[M-H]^-$		$C_5H_{10}O_2$	N/A[1]	[d]
	122.0237	1.40	$[M-H]^-$		$C_6H_5NO_2$	N/A	[d]
	187.0064	0.64	$[M-H]^-$		$C_7H_8O_4S$	N/A	[d]
	213.1493	5.97	$[M-H]^-$		$C_{12}H_{22}O_3$	1.6	[d]
Adult	143.0704	0.67	$[M-H]^-$	4-Hydroxycyclohexylcarboxylic acid	$C_7H_{12}O_3$	*1.8*	CCS[a] and MS/MS[a, c]
	137.0234	0.74	$[M-H]^-$	Salicylic acid	$C_7H_6O_3$	2.5	Rt[a] and MS/MS[a, b]
	157.0862	0.97	$[M-H]^-$	4-Hydroxycyclohexylacetic acid	$C_8H_{14}O_3$	*2.1*	MS/MS[c]
	171.0655	0.64	$[M-H]^-$	2-Octenedioic acid, 4-Octenedioic acid, 3-Octenedioic acid	$C_8H_{12}O_4$	*2.0*	MS/MS[c]
	130.0499	0.57	$[M-H]^-$		$C_5H_9NO_3$	1.0	[d]

[a] match chemical standard; [b] match Metlin MS/MS; [c] match manual analysis; [d] no endogenous ID, not [a], [b] or [c]. Significant median fold change is shown in italics. [1]Unable to calculate median fold change because the denominator is 0.

Table 3.9.1: Identifications of discriminant features in oPLS-DA classification of APE from stable CF samples using combined positive and negative ion mode data.

Age Group	*m/z*	R_t (min)	Ion type	Tentative identity	Elemental formula	Median fold change (APE/stable CF)	Identification
Pediatric	203.0822	1.21	$[M-H]^-$	Tryptophan	$C_{11}H_{12}N_2O_2$	-1.4	Rt[a] and MS/MS[a, b]
	171.0655	0.64	$[M-H]^-$	2-Octenedioic acid, 4-Octenedioic acid, 3-Octenedioic acid	$C_8H_{12}O_4$	1.2	MS/MS[c]
	134.0599	3.88	$[M+H]^+$	Indoxyl, Oxindole[2]	C_8H_7NO	N/A[1]	MS/MS[c] for indoxyl, [b] for oxindole
	130.0499	0.57	$[M-H]^-$		$C_5H_9NO_3$	1.6	[d]
	89.0602	1.33	$[M+H]^+$		$C_4H_8O_2$	*-16*	[d]
	187.0064	0.64	$[M-H]^-$		$C_7H_8O_4S$	*N/A*	[d]
Adult	143.0704	0.67	$[M-H]^-$	4-Hydroxycyclohexylcarboxylic acid	$C_7H_{12}O_3$	1.5	CCS[a] and MS/MS[a, c]
	130.0863	0.75	$[M-H]^-$	Leucine, Isoleucine	$C_6H_{13}NO_2$	1.2	Rt[a] and MS/MS[a, c]
	162.1122	0.52	$[M+H]^+$	Carnitine	$C_7H_{15}NO_3$	1.5	CCS[a] and MS/MS[a, b]
	187.0970	0.63	$[M-H]^-$	Azelaic acid	$C_9H_{16}O_4$	2.5	CCS[a] and MS/MS[a, b]
	85.0283	0.61	$[M-H]^-$	γ-Butyrolactone, Oxolan-3-one	$C_4H_6O_2$	2.0	MS/MS[c]
	87.0439	0.64	$[M-H]^-$	Butyric acid	$C_4H_8O_2$	-1.9	
	134.0599	3.88	$[M+H]^+$	Indoxyl, Oxindole	C_8H_7NO	-1.4	MS/MS[c] for indoxyl, [b] for oxindole
	137.0683	0.98	$[M+H]^+$			2.7	[e]; Experimental isotopic pattern doesn't match to theoretical value of candidate
	179.0377	0.63	$[M-H]^-$		$C_6H_{12}O_4S$	N/A	[d]

[a] match chemical standard; [b] match Metlin MS/MS; [c] match manual analysis; [d] no endogenous ID, not [a], [b] or [c]; [e] no good MS/MS. Significant median fold change is shown in italics. [1]Unable to calculate median fold change because the denominator is 0. [2]A bacterial metabolite not in HMDB.

Table 3.9.2: Identification of discriminant features in oPLS-DA classification of pre-APE from stable CF samples using combined positive and negative ion mode data.

Age Group	*m/z*	R_t (min)	Ion type	Tentative identity	Elemental formula	Median fold change (APE/stable CF)	Identification
Pediatric	89.0232	0.55	$[M-H]^-$	Lactic acid	$C_3H_6O_3$	-1.0	CCS[a] and MS/MS[a, b]
	146.0449	0.53	$[M-H]^-$	Glutamic acid	$C_5H_9NO_4$	1.3	CCS[a] and MS/MS[a, b]
	116.0707	0.64	$[M+H]^+$	Proline	$C_5H_9NO_2$	*-2.4*	MS/MS[a, b]
	130.0499	0.57	$[M-H]^-$		$C_5H_9NO_3$	1.4	[d]
	161.0320	0.61	$[M+Na]^+$		$C_6H_6N_2O_2$	4.9	[e]
	169.0833	1.16	$[M+Na]^+$		$C_7H_{14}O_3$	-1.4	[e]
	101.0596	0.80	$[M-H]^-$		$C_5H_{10}O_2$	0	[d]
	113.0575	0.88	$[M+Na]^+$		$C_4H_{10}O_2$	N/A[1]	[d]
Adult	143.0704	0.67	$[M-H]^-$	4-Hydroxycyclohexylcarboxylic acid	$C_7H_{12}O_3$	1.7	CCS[a] and MS/MS[a, c]
	164.0345	0.74	$[M-H]^-$	Formylanthranilic acid	$C_8H_7NO_3$	1.3	Rt[a] and MS/MS[a, b]
	175.1188	0.65	$[M+H]^+$	Arginine	$C_6H_{14}N_4O_2$	3.1	CCS[a] and MS/MS[a, b]
	157.0862	0.97	$[M-H]^-$	4-Hydroxycyclohexylacetic acid	$C_8H_{14}O_3$	*2.1*	MS/MS[c]
	171.0655	0.64	$[M-H]^-$	2-Octenedioic acid, 4-Octenedioic acid, 3-Octenedioic acid	$C_8H_{12}O_4$	*2.1*	MS/MS[c]
	59.0125	0.56	$[M-H]^-$	Acetic acid	$C_2H_4O_2$	1.2	CCS[a]
	139.0514	2.24	$[M+H]^+$		$C_6H_6N_2O_2$	*17*	[d]
	139.0839	1.86	$[M+Na]^+$		$C_5H_{12}N_2O$	2.2	[e]
	161.0320	0.61	$[M+Na]^+$		$C_6H_6N_2O_2$	N/A	[e]
	169.0833	1.16	$[M+Na]^+$		$C_7H_{14}O_3$	*2.6*	[e]
	181.0232	0.77	$[M+Na]^+$		$C_5H_6N_2O_4$	N/A	[d]

[a] match chemical standard; [b] match Metlin MS/MS; [c] match manual analysis; [d] no endogenous ID, not [a], [b] or [c]; [e] no good MS/MS. Significant median fold change is shown in italics. [1]Unable to calculate median fold change because the denominator is 0.

Of all discriminant features, several showed interesting trends in their abundances in EBC samples from patients at different stages preceding and following an APE event, in either adult or pediatric patients, including lactic acid, leucine/isoleucine, indoxyl/oxindole, butyric acid, azelaic acid and formylanthranilic acid (Figure 3.10). In EBC samples from pediatric patients, lactic acid was significantly elevated in APE samples compared to stable and pre-APE ones (Figure 3.10 (A)). It was also selected by iPLS-DA in the model comparing EBC samples from pre-APE patients with those from stable CF patients in the pilot study.[74] Lactic acid is a fermentation metabolite of the anaerobic microbial community in CF and a potential biomarker associated with CF progression.[64] In a sputum metabolomics study, lactic acid was detected with significant increased level in samples from APE patients compared to those from stable CF patients using LC-MS/MS.[75] Butyric acid has been found to be secreted by anaerobic bacteria found in CF airways, and it was also detected in BALF samples from CF patients.[76, 77] In this study, butyric acid had an increasing trend in EBC samples from stable CF to pre-APE pediatric patients, and a decreasing trend from pre-APE to APE, and from APE to post-APE (Figure 3.10 (B)). The decrease in abundance of butyric acid from the pre-APE to the APE group in the cohort might be due to the fact that over 50% of the APE patients had already started intravenous antibiotic treatment at the time of EBC collection, resulting in a different inflammatory phenotype from that in pre-APE patients, who had not yet been aggressively treated. Butyric acid, together with acetic acid and propionic acid, which also appeared in the discriminant feature panels (Table 3.8.1-3.9.2), are short-chain fatty acids associated with immune and inflammatory processes in CF airways.[76] Interestingly, acetic acid was identified as one of the potential EBC biomarkers for

discriminating stable CF patients from unstable CF patients during exacerbations in a metabolomics study using NMR.[51] Leucine/isoleucine levels exhibited a similar trend as butyric acid in EBC samples from pediatric patients of this cohort (Figure 3.10 (C)). The rate of appearance of leucine, a measure of protein breakdown using stable isotope and MS analysis, was reported to be significantly higher in CF children compared with age- and gender- matched healthy controls[78] and leucine synthesis has been found to be significantly higher in malnourished CF patients compared to well-nourished ones[79]. In addition, leucine was reported as an important discriminant metabolite for classification of CF patients with high and low inflammation.[60] An overall higher amino acid content, including leucine and isoleucine, was detected in sputum from CF patients compared to non-CF controls, suggesting that it may provide nutrition supplement to auxotrophic *P. aeruginosa* in CF.[80] In the adult group, formylanthranilic acid had an elevated mean value in pre-APE patients compared to CF patients at other stages (Figure 3.10 (D)). Formylanthranilic acid is a bacterial metabolite from indole,[81] which is known to promote *P. aeruginosa* biofilm formation, causing infections in CF patients.[82] In addition, formylanthranilic acid, indoxyl and oxindole (Figure 3.10 (E)) are involved in tryptophan metabolism,[77, 83] and tryptophan was also selected as a discriminant feature in the classification model of APE vs. stable CF pediatric samples (Table 3.9.1). Tryptophan metabolism plays a crucial role in gut mucosal homeostasis and microbiome regulation,[84] and it was found to be significantly altered in primary human airway epithelial cells from CF patients compared to those from non-CF individuals in a metabolic profiling study by Wetmore *et al.*[85]. The remaining metabolite, azelaic acid (Figure 3.10 (F)) has not yet been reported in EBC and its association with CF needs further investigation.

Interestingly, 4-hydroxycyclohexanecarboxylic acid, which is a metabolite associated with gut microbial mammalian cometabolism and one of the two discriminant features in the model for APE vs. stable CF sample classification in the pilot study,[74] appeared in multiple discriminant panels in the large cohort study (Table 3.8.1-3.9.2). In addition, pyroglutamic acid, which is involved in the γ-glutamyl cycle and present in both APE vs. stable CF and pre-APE vs. stable CF discriminant panels in the pilot study,[74] also appeared in the 9-feature negative mode data model differentiating APE from stable CF pediatric samples of the large cohort (Table 3.8.1). These metabolites selected for discrimination in both the pilot and the large cohort studies in a non-supervised fashion, corroborated the biological significance and implications for future diagnostic approaches of the panels described here.

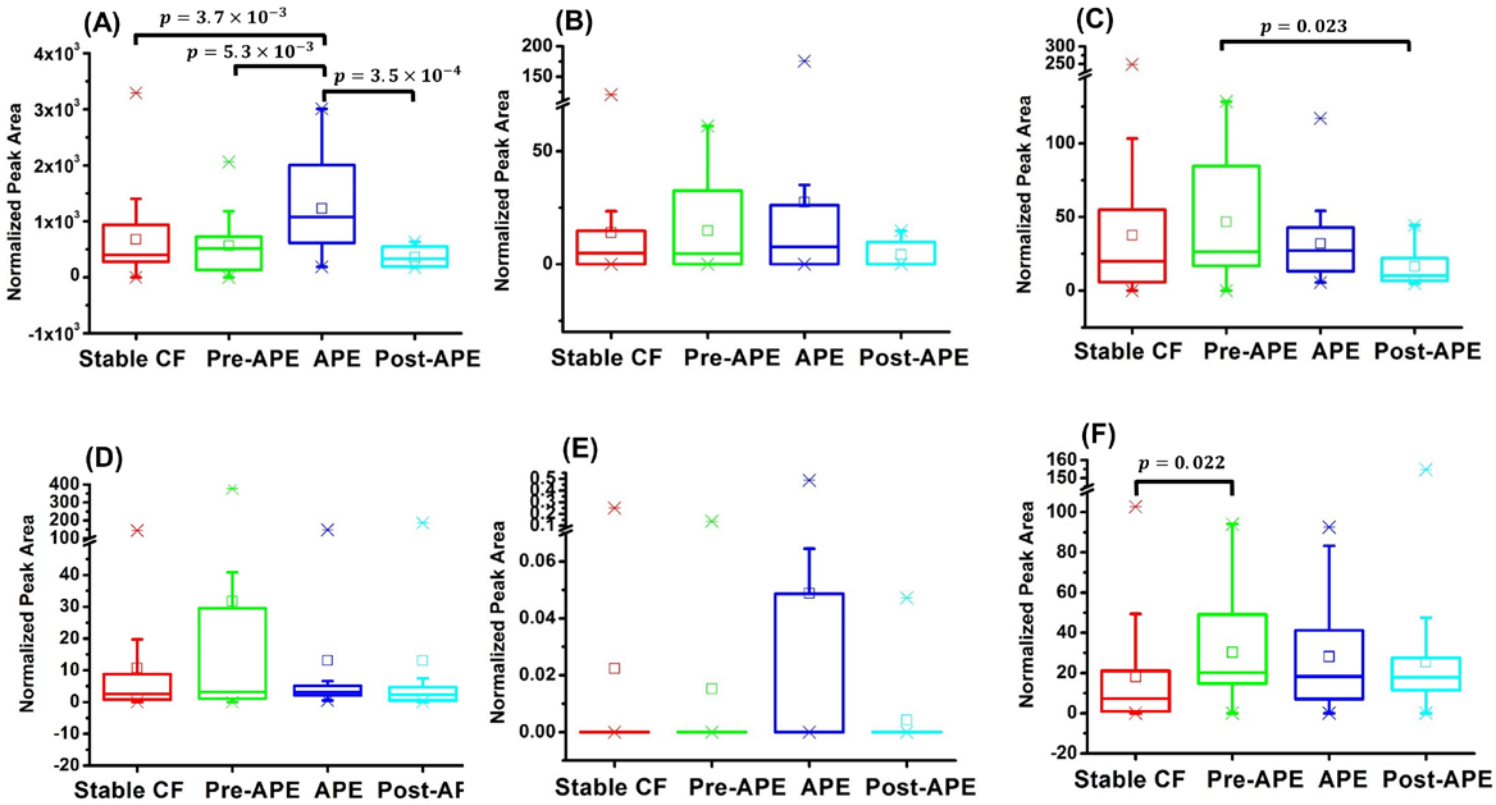

Figure 3.10: Box plots for discriminant metabolites with changing trends in different cohort subgroups: (A) lactic acid, (B) butyric acid and (C) leucine in pediatric patients, (D) formylanthranilic acid in adult patients, (E) indoxyl or oxindole in pediatric patients, and (F) azelaic acid in adult patients. Mean values are represented by a filled circle in the box; median values are represented by a line in the box; the edges of the box are 25th and 75th percentiles; the whisker extends to the most extreme values in data not including outliers, with a 99.3% coverage; outliers are represented by asterisks.

Malic acid, which was present in the 11-feature negative ion mode oPLS-DA model for classification of pre-APE and stable CF pediatric patient samples, was significantly decreased in pre-APE , with a large median fold change of 2.0×10^2 (Table 3.8.2). Malic acid has been recently detected in exhaled breath for the first time by secondary electrospray ionization high resolution MS,[86] and it was reported to have cardioprotective effects possibly due to its anti-inflammatory properties[87]. In addition, it is involved in pyruvate metabolism, expected to be significantly altered between APE and stable CF pediatric samples (Figure 3.11). Free carnitine, which was identified in the adult APE vs. stable CF discriminant panel of the model using combined negative and positive ion mode data, had a 1.5 fold up-regulation in APE samples (Table 3.9.1). Carnitine can be either synthesized in the human body or obtained from dietary red meat, and metabolized by intestinal bacteria.[77, 88] It is involved in fatty acid transportation to mitochondria, and abnormal fatty acid metabolism has been associated with CF.[88-91] Interestingly, free carnitine levels in plasma have been reported to be significantly higher in CF patients compared to healthy controls.[92] To the best of our knowledge, it has not been previously detected in EBC.

A number of identified discriminant metabolites have been reported as being connected with microbial metabolism, re-emphasizing the important role that the CF microbiome has in terms of interacting with the host and participating in inflammatory mechanisms during and before the onset of an APE. Further investigation of the microbial metabolites and integration of CF microbiome studies with metabolomics should improve our current understanding of host-pathogen interactions in APE

development and progression, guiding clinical decisions for early intervention and personalized APE treatment.[93]

3.6.4 Pathway Analysis

Pathway analysis of the 23 uniquely identified discriminant metabolites from all panels revealed pathways with significant changes in APE and pre-APE compared to stable CF patients. These included arginine and proline metabolism, which was found to be significantly altered in pre-APE adult and pediatric samples, and pyruvate metabolism, which was significantly changed between APE and stable pediatric patients (Figure 3.11). The discriminant metabolites involved in arginine and proline metabolism were arginine, proline and glutamic acid. Arginine was one of the 11 discriminant features responsible for classification of adult pre-APE from stable CF samples, with a 3.1-fold increase in the pre-APE group (Table 3.9.2). Arginine serves as the precursor for nitric oxide production and could be *de novo* synthesized by citrulline.[94, 95] Its concentration in EBC has been shown to be associated with lung function, with a significant negative correlation with percent predicted forced vital capacity and a nearly significant negative correlation with percent predicted FEV_1.[94] Another study reported decreased plasma arginine concentration during CF APE, with its level recovering after antibiotic therapy.[96] Grasemann *et al.* showed disturbed arginine metabolism in CF patients.[96] Proline and glutamic acid are downstream products of arginase,[96] and proline had a 2.4-fold decrease in abundance between pre-APE and stable CF pediatric patients (Table 3.9.2). Significantly increased proline and glutamic acid levels were found after antibiotic

treatment of APE.[96] Further research is needed to elucidate the mechanisms involved in the observed alterations of arginine and proline metabolism during CF APE progression.

Pyruvate metabolism, including lactic acid, malic acid, and acetic acid, was found to be significantly altered in EBC samples collected from pediatric patients during an APE compared to those from stable pediatric patients. Pyruvate metabolism is at a key intersection of many pathways in human biological systems including glycolysis and the citric acid cycle.[97] Abnormalities in pyruvate metabolism have been strongly associated with various diseases.[97] The flux of pyruvate metabolism is highly diverse in different strains of *P. aeruginosa*,[98] and its relationship to CF also warrants further investigation.

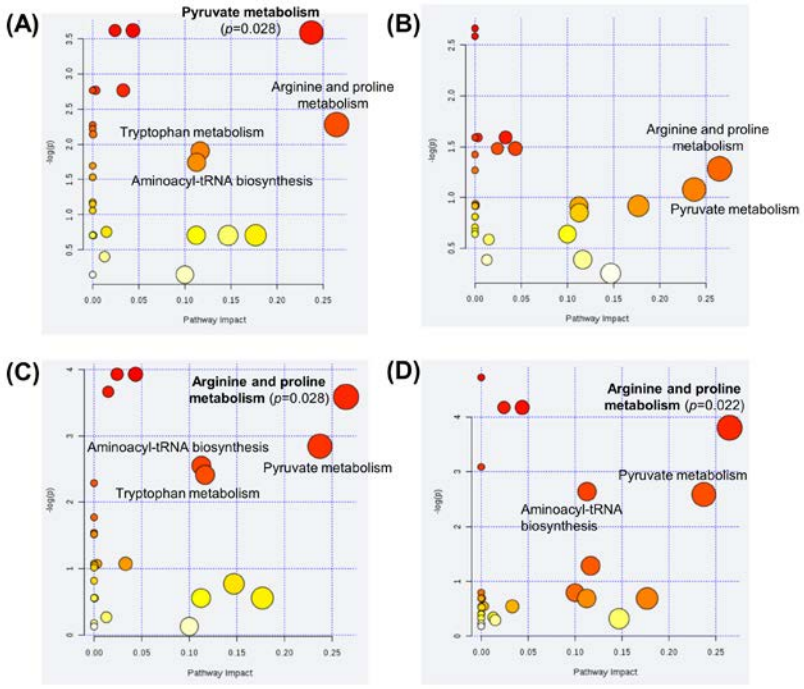

Figure 3.11: Pathway analysis of the 23 uniquely-identified discriminant metabolites from all panels, including EBC samples from APE vs. stable CF (A) pediatric and (B) adult patients, and pre-APE vs. stable CF in (C) pediatric and (D) adult patients. Each circle on the map represents a pathway, and the size and color of the circle are based on the *p* value, indicating the significance of the changes in the matched metabolites in the pathway, and the pathway impact score, which is correlated with the centrality of the metabolites involved.[99, 100]

3.7 Conclusion

Feasibility of detecting EBC metabolites related to APE events in CF patients was demonstrated with a pilot study, by means of an non-targeted UPLC-MS-based metabolomics method coupled to multivariate statistical analysis. oPLS-DA multivariate classification on negative ion mode data yielded acceptable accuracies (84.6%, 90.5%), sensitivities (77.8%, 75.0%) and specificities (88.2%, 94.1%) in distinguishing 9 APE or

4 pre-APE EBC samples from 17 stable CF samples, respectively. A larger cohort study (n= 210) was conducted to validate these findings and to discover novel potential biomarkers for APE detection and prediction. Both negative ion mode data and combined positive and negative ion mode data showed classification of APE and pre-APE vs. stable CF patients in adult and pediatric cohorts. Supervised multivariate models yielded good classification accuracies ranging between 81.3 and 93.9%, depending on the binary comparison, with AUC values of ~0.8-0.9; and provided different discriminant EBC metabolite panels for APE and pre-APE detection in adult and pediatric patients. Metabolites exhibiting significant changes at different stages of an APE event were discussed in relation to the altered metabolic pathways and their microbial relevance. All three discriminant metabolites in the pilot study were also selected in the larger cohort in a non-supervised fashion, validating the biological significance of the panels identified for APE detection and prediction. Limitations of the studies include a lack of sufficient number of longitudinal EBC samples collected from the same patient at different stages of APE progression for time series analysis. Further investigation of the CF microbiome should aid to improve our current understanding of the link between host and pathogen interactions in APE development and progression, and to obtain chemical information on pathogen-specific metabolites that may assist in personalized clinical decisions for early intervention and better APE treatment. The results presented here show promise for detecting APEs and even predicting an oncoming APE event using EBC metabolites, and provide insight into the molecular mechanisms of CF APE development.

3.8 References

1. Kreindler, J. L., Cystic fibrosis: Exploiting its genetic basis in the hunt for new therapies. *Pharmacology & Therapeutics* **2010,** 125, (2), 219-229.

2. Davis, P. B.; Drumm, M.; Konstan, M. W., Cystic fibrosis. *American Journal Of Respiratory And Critical Care Medicine* **1996,** 154, (5), 1229-1256.

3. Ramsey, B. W., Management of pulmonary disease in patients with cystic fibrosis. *N Engl J Med* **1996,** 335, (3), 179-88.

4. 2017 Cystic Fibrosis Foundation Patient Registry Highlights. https://www.cff.org/Research/Researcher-Resources/Patient-Registry/2017-Cystic-Fibrosis-Foundation-Patient-Registry-Highlights.pdf (July 1st).

5. Britto, M. T.; Kotagal, U. R.; Hornung, R. W.; Atherton, H. D.; Tsevat, J.; Wilmott, R. W., Impact of recent pulmonary exacerbations on quality of life in patients with cystic fibrosis. *Chest* **2002,** 121, (1), 64-72.

6. Lieu, T. A.; Ray, G. T.; Farmer, G.; Shay, G. F., The cost of medical care for patients with cystic fibrosis in a health maintenance organization. *Pediatrics* **1999,** 103, (6), e72.

7. Rosenfeld, M.; Gibson, R. L.; McNamara, S.; Emerson, J.; Burns, J. L.; Castile, R.; Hiatt, P.; McCoy, K.; Wilson, C. B.; Inglis, A.; Smith, A.; Martin, T. R.; Ramsey, B. W., Early pulmonary infection, inflammation, and clinical outcomes in infants with cystic fibrosis. *Pediatric Pulmonology* **2001,** 32, (5), 356-66.

8. Fuchs, H. J.; Borowitz, D. S.; Christiansen, D. H.; Morris, E. M.; Nash, M. L.; Ramsey, B. W.; Rosenstein, B. J.; Smith, A. L.; Wohl, M. E., Effect of aerosolized recombinant human DNase on exacerbations of respiratory symptoms and on pulmonary function in patients with cystic fibrosis. The Pulmozyme Study Group. *New England Journal of Medicine* **1994,** 331, (10), 637-42.

9. Saiman, L.; Marshall, B. C.; Mayer-Hamblett, N.; Burns, J. L.; Quittner, A. L.; Cibene, D. A.; Coquillette, S.; Fieberg, A. Y.; Accurso, F. J.; Campbell, P. W., 3rd; Macrolide Study, G., Azithromycin in patients with cystic fibrosis chronically infected

with Pseudomonas aeruginosa: a randomized controlled trial. *JAMA* **2003,** 290, (13), 1749-56.

10. Ramsey, B. W.; Boat, T. F., Outcome measures for clinical trials in cystic fibrosis. Summary of a Cystic Fibrosis Foundation consensus conference. *Journal of pediatrics* **1994,** 124, (2), 177-92.

11. Goss, C. H.; Burns, J. L., Exacerbations in cystic fibrosis. 1: Epidemiology and pathogenesis. *Thorax* **2007,** 62, (4), 360-7.

12. Wark, P. A.; Tooze, M.; Cheese, L.; Whitehead, B.; Gibson, P. G.; Wark, K. F.; McDonald, V. M., Viral infections trigger exacerbations of cystic fibrosis in adults and children. *Eur Respir J* **2012,** 40, (2), 510-2.

13. van Ewijk, B. E.; van der Zalm, M. M.; Wolfs, T. F.; van der Ent, C. K., Viral respiratory infections in cystic fibrosis. *J Cyst Fibros* **2005,** 4 Suppl 2, 31-6.

14. Goss, C. H.; Newsom, S. A.; Schildcrout, J. S.; Sheppard, L.; Kaufman, J. D., Effect of ambient air pollution on pulmonary exacerbations and lung function in cystic fibrosis. *Am J Respir Crit Care Med* **2004,** 169, (7), 816-21.

15. Smyth, A.; Elborn, J. S., Exacerbations in cystic fibrosis: 3--Management. *Thorax* **2008,** 63, (2), 180-4.

16. Stenbit, A. E.; Flume, P. A., Pulmonary exacerbations in cystic fibrosis. *Curr Opin Pulm Med* **2011,** 17, (6), 442-7.

17. Sanders, D. B.; Bittner, R. C. L.; Rosenfeld, M.; Hoffman, L. R.; Redding, G. J.; Goss, C. H., Failure to Recover to Baseline Pulmonary Function after Cystic Fibrosis Pulmonary Exacerbation. *American Journal of Respiratory and Critical Care Medicine* **2010,** 182, (5), 627-632.

18. Goss, C. H.; Burns, J. L., Exacerbations in cystic fibrosis center dot 1: Epidemiology and pathogenesis. *Thorax* **2007,** 62, (4), 360-367.

19. Rogers, G. B.; Hoffman, L. R.; Johnson, M. W.; Mayer-Hamblett, N.; Schwarze, J.; Carroll, M. P.; Bruce, K. D., Using bacterial biomarkers to identify early indicators of cystic fibrosis pulmonary exacerbation onset. *Expert Review of Molecular Diagnostics* **2011,** 11, (2), 197-206.

20. Shoki, A. H.; Mayer-Hamblett, N.; Wilcox, P. G.; Sin, D. D.; Quon, B. S., Systematic review of blood biomarkers in cystic fibrosis pulmonary exacerbations. *Chest* **2013,** 144, (5), 1659-1670.

21. Matouk, E.; Nguyen, D.; Benedetti, A.; Bernier, J.; Gruber, J.; Landry, J.; Rousseau, S.; Ahlgren, H. G.; Lands, L. C.; Wojewodka, G.; Radzioch, D., C-Reactive Protein in Stable Cystic Fibrosis: An Additional Indicator of Clinical Disease Activity and Risk of Future Pulmonary Exacerbations. *J Pulm Respir Med* **2016,** 6, (5), 1000375.

22. Wojewodka, G.; De Sanctis, J. B.; Bernier, J.; Berube, J.; Ahlgren, H. G.; Gruber, J.; Landry, J.; Lands, L. C.; Nguyen, D.; Rousseau, S.; Benedetti, A.; Matouk, E.; Radzioch, D., Candidate markers associated with the probability of future pulmonary exacerbations in cystic fibrosis patients. *PLoS One* **2014,** 9, (2), e88567.

23. Gray, R. D.; Downey, D.; Taggart, C. C., Biomarkers to monitor exacerbations in cystic fibrosis. *Expert Rev Respir Med* **2017,** 11, (4), 255-257.

24. Reid, P. A.; McAllister, D. A.; Boyd, A. C.; Innes, J. A.; Porteous, D.; Greening, A. P.; Gray, R. D., Measurement of serum calprotectin in stable patients predicts exacerbation and lung function decline in cystic fibrosis. *Am J Respir Crit Care Med* **2015,** 191, (2), 233-6.

25. Gray, R. D.; Imrie, M.; Boyd, A. C.; Porteous, D.; Innes, J. A.; Greening, A. P., Sputum and serum calprotectin are useful biomarkers during CF exacerbation. *J Cyst Fibros* **2010,** 9, (3), 193-8.

26. Quon, B. S.; Ngan, D. A.; Wilcox, P. G.; Man, S. F.; Sin, D. D., Plasma sCD14 as a biomarker to predict pulmonary exacerbations in cystic fibrosis. *PLoS One* **2014,** 9, (2), e89341.

27. Quon, B. S.; Dai, D. L. Y.; Hollander, Z.; Ng, R. T.; Tebbutt, S. J.; Man, S. F. P.; Wilcox, P. G.; Sin, D. D., Discovery of novel plasma protein biomarkers to predict imminent cystic fibrosis pulmonary exacerbations using multiple reaction monitoring mass spectrometry. *Thorax* **2016,** 71, (3), 216-222.

28. Solomon, G. M.; Frederick, C.; Zhang, S.; Gaggar, A.; Harris, T.; Woodworth, B. A.; Steele, C.; Rowe, S. M., IP-10 is a potential biomarker of cystic fibrosis acute pulmonary exacerbations. *PLoS One* **2013,** 8, (8), e72398.

29. Ordonez, C. L.; Henig, N. R.; Mayer-Hamblett, N.; Accurso, F. J.; Burns, J. L.; Chmiel, J. F.; Daines, C. L.; Gibson, R. L.; McNamara, S.; Retsch-Bogart, G. Z.; Zeitlin, P. L.; Aitken, M. L., Inflammatory and microbiologic markers in induced sputum after intravenous antibiotics in cystic fibrosis. *Am J Respir Crit Care Med* **2003,** 168, (12), 1471-5.

30. Horsley, A. R.; Davies, J. C.; Gray, R. D.; Macleod, K. A.; Donovan, J.; Aziz, Z. A.; Bell, N. J.; Rainer, M.; Mt-Isa, S.; Voase, N.; Dewar, M. H.; Saunders, C.; Gibson, J. S.; Parra-Leiton, J.; Larsen, M. D.; Jeswiet, S.; Soussi, S.; Bakar, Y.; Meister, M. G.; Tyler, P.; Doherty, A.; Hansell, D. M.; Ashby, D.; Hyde, S. C.; Gill, D. R.; Greening, A. P.; Porteous, D. J.; Innes, J. A.; Boyd, A. C.; Griesenbach, U.; Cunningham, S.; Alton, E. W., Changes in physiological, functional and structural markers of cystic fibrosis lung disease with treatment of a pulmonary exacerbation. *Thorax* **2013,** 68, (6), 532-9.

31. Radhakrishnan, D. K.; Corey, M.; Dell, S. D., Realities of expectorated sputum collection in the pediatric cystic fibrosis clinic. *Arch Pediatr Adolesc Med* **2007,** 161, (6), 603-6.

32. Al-Saleh, S.; Dell, S. D.; Grasemann, H.; Yau, Y. C.; Waters, V.; Martin, S.; Ratjen, F., Sputum induction in routine clinical care of children with cystic fibrosis. *J Pediatr* **2010,** 157, (6), 1006-1011 e1.

33. Rosias, P. P.; Dompeling, E.; Hendriks, H. J.; Heijnens, J. W.; Donckerwolcke, R. A.; Jobsis, Q., Exhaled breath condensate in children: pearls and pitfalls. *Pediatr Allergy Immunol* **2004,** 15, (1), 4-19.

34. Horvath, I.; Hunt, J.; Barnes, P. J.; Alving, K.; Antczak, A.; Baraldi, E.; Becher, G.; van Beurden, W. J.; Corradi, M.; Dekhuijzen, R.; Dweik, R. A.; Dwyer, T.; Effros, R.; Erzurum, S.; Gaston, B.; Gessner, C.; Greening, A.; Ho, L. P.; Hohlfeld, J.; Jobsis, Q.; Laskowski, D.; Loukides, S.; Marlin, D.; Montuschi, P.; Olin, A. C.; Redington, A. E.; Reinhold, P.; van Rensen, E. L.; Rubinstein, I.; Silkoff, P.; Toren, K.; Vass, G.; Vogelberg, C.; Wirtz, H.; Condensate, A. E. T. F. o. E. B., Exhaled breath condensate: methodological recommendations and unresolved questions. *Eur Respir J* **2005,** 26, (3), 523-48.

35. Mutlu, G. M.; Garey, K. W.; Robbins, R. A.; Danziger, L. H.; Rubinstein, I., Collection and analysis of exhaled breath condensate in humans. *American Journal of Respiratory and Critical Care Medicine* **2001,** 164, (5), 731-737.

36. Hunt, J., Exhaled breath condensate: An evolving tool for noninvasive evaluation of lung disease. *Journal of Allergy and Clinical Immunology* **2002,** 110, (1), 28-34.

37. Robroeks, C. M.; Rosias, P. P.; van Vliet, D.; Jobsis, Q.; Yntema, J. B.; Brackel, H. J.; Damoiseaux, J. G.; den Hartog, G. M.; Wodzig, W. K.; Dompeling, E., Biomarkers in exhaled breath condensate indicate presence and severity of cystic fibrosis in children. *Pediatr Allergy Immunol* **2008,** 19, (7), 652-9.

38. Carpagnano, G. E.; Barnes, P. J.; Geddes, D. M.; Hodson, M. E.; Kharitonov, S. A., Increased leukotriene B4 and interleukin-6 in exhaled breath condensate in cystic fibrosis. *Am J Respir Crit Care Med* **2003,** 167, (8), 1109-12.

39. van Horck, M.; Alonso, A.; Wesseling, G.; de Winter-de Groot, K.; van Aalderen, W.; Hendriks, H.; Winkens, B.; Rijkers, G.; Jobsis, Q.; Dompeling, E., Biomarkers in Exhaled Breath Condensate Are Not Predictive for Pulmonary Exacerbations in Children with Cystic Fibrosis: Results of a One-Year Observational Study. *PLoS One* **2016,** 11, (4), e0152156.

40. Osika, E.; Cavaillon, J. M.; Chadelat, K.; Boule, M.; Fitting, C.; Tournier, G.; Clement, A., Distinct sputum cytokine profiles in cystic fibrosis and other chronic inflammatory airway disease. *European Respiratory Journal* **1999,** 14, (2), 339-46.

41. Ordonez, C. L.; Henig, N. R.; Mayer-Hamblett, N.; Accurso, F. J.; Burns, J. L.; Chmiel, J. F.; Daines, C. L.; Gibson, R. L.; McNamara, S.; Retsch-Bogart, G. Z.; Zeitlin, P. L.; Aitken, M. L., Inflammatory and microbiologic markers in induced sputum after intravenous antibiotics in cystic fibrosis. *American Journal of Respiratory and Critical Care Medicine* **2003,** 168, (12), 1471-5.

42. Francoeur, C.; Denis, M., Nitric oxide and interleukin-8 as inflammatory components of cystic fibrosis. *Inflammation* **1995,** 19, (5), 587-98.

43. Sloane, A. J.; Lindner, R. A.; Prasad, S. S.; Sebastian, L. T.; Pedersen, S. K.; Robinson, M.; Bye, P. T.; Nielson, D. W.; Harry, J. L., Proteomic analysis of sputum from adults and children with cystic fibrosis and from control subjects. *American Journal of Respiratory and Critical Care Medicine* **2005,** 172, (11), 1416-26.

44. Wolter, J. M.; Rodwell, R. L.; Bowler, S. D.; McCormack, J. G., Cytokines and inflammatory mediators do not indicate acute infection in cystic fibrosis. *Clinical and Diagnostic Laboratory Immunology* **1999,** 6, (2), 260-5.

45. Salva, P. S.; Doyle, N. A.; Graham, L.; Eigen, H.; Doerschuk, C. M., TNF-alpha, IL-8, soluble ICAM-1, and neutrophils in sputum of cystic fibrosis patients. *Pediatric Pulmonology* **1996,** 21, (1), 11-9.

46. Dunn, W. B.; Broadhurst, D.; Begley, P.; Zelena, E.; Francis-McIntyre, S.; Anderson, N.; Brown, M.; Knowles, J. D.; Halsall, A.; Haselden, J. N.; Nicholls, A. W.; Wilson, I. D.; Kell, D. B.; Goodacre, R.; C, H. S. M. H., Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nature Protocols* **2011,** 6, (7), 1060-1083.

47. Patti, G. J.; Yanes, O.; Siuzdak, G., Metabolomics: the apogee of the omics trilogy. *Nature Reviews Molecular Cell Biology* **2012,** 13, (4), 263-269.

48. Quinn, R. A.; Lim, Y. W.; Mak, T. D.; Whiteson, K.; Furlan, M.; Conrad, D.; Rohwer, F.; Dorrestein, P., Metabolomics of pulmonary exacerbations reveals the personalized nature of cystic fibrosis disease. *PeerJ* **2016,** 4, e2174.

49. Laguna, T. A.; Reilly, C. S.; Williams, C. B.; Welchlin, C.; Wendt, C. H., Metabolomics analysis identifies novel plasma biomarkers of cystic fibrosis pulmonary exacerbation. *Pediatric Pulmonology* **2015,** 50, (9), 869-877.

50. Alvarez, J. A.; Chong, E. Y.; Walker, D. I.; Chandler, J. D.; Michalski, E. S.; Grossmann, R. E.; Uppal, K.; Li, S.; Frediani, J. K.; Tirouvanziam, R.; Tran, V. T.; Tangpricha, V.; Jones, D. P.; Ziegler, T. R., Plasma metabolomics in adults with cystic fibrosis during a pulmonary exacerbation: A pilot randomized study of high-dose vitamin D3 administration. *Metabolism* **2017,** 70, 31-41.

51. Montuschi, P.; Paris, D.; Melck, D.; Lucidi, V.; Ciabattoni, G.; Raia, V.; Calabrese, C.; Bush, A.; Barnes, P. J.; Motta, A., NMR spectroscopy metabolomic profiling of exhaled breath condensate in patients with stable and unstable cystic fibrosis. *Thorax* **2012,** 67, (3), 222-8.

52. Bush, M. F.; Campuzano, I. D. G.; Robinson, C. V., Ion Mobility Mass Spectrometry of Peptide Ions: Effects of Drift Gas and Calibration Strategies. *Analytical Chemistry* **2012,** 84, (16), 7124-7130.

53. Smith, C. A.; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G., METLIN: a metabolite mass spectral database. *Therapeutic Drug Monitoring* **2005,** 27, (6), 747-51.

54. Wishart, D. S.; Jewison, T.; Guo, A. C.; Wilson, M.; Knox, C.; Liu, Y.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E., HMDB 3.0—the human metabolome database in 2013. *Nucleic Acids Research* **2013,** 41, (Supp 1), D801-D807.

55. Worley, B.; Powers, R., Multivariate analysis in metabolomics. *Current Metabolomics* **2013,** 1, (1), 92-107.

56. Boccard, J.; Rutledge, D. N., A consensus orthogonal partial least squares discriminant analysis (OPLS-DA) strategy for multiblock Omics data fusion. *Analytica Chimica Acta* **2013,** 769, 30-9.

57. Szymanska, E.; Saccenti, E.; Smilde, A. K.; Westerhuis, J. A., Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics* **2012,** 8, (Suppl 1), 3-16.

58. Xia, J.; Sinelnikov, I. V.; Han, B.; Wishart, D. S., MetaboAnalyst 3.0--making metabolomics more meaningful. *Nucleic Acids Res* **2015,** 43, (W1), W251-7.

59. Kirwan, J. A.; Weber, R. J.; Broadhurst, D. I.; Viant, M. R., Direct infusion mass spectrometry metabolomics dataset: a benchmark for data processing and quality control. *Scientific Data* **2014,** 1, 140012.

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

61. De Backer, D.; Creteur, J.; Zhang, H.; Norrenberg, M.; Vincent, J. L., Lactate production by the lungs in acute lung injury. *American Journal of Respiratory and Critical Care Medicine* **1997,** 156, (4 Pt 1), 1099-104.

62. Worlitzsch, D.; Tarran, R.; Ulrich, M.; Schwab, U.; Cekici, A.; Meyer, K. C.; Birrer, P.; Bellon, G.; Berger, J.; Weiss, T.; Botzenhart, K.; Yankaskas, J. R.; Randell, S.; Boucher, R. C.; Doring, G., Effects of reduced mucus oxygen concentration in airway Pseudomonas infections of cystic fibrosis patients. *Journal of Clinical Investigation* **2002,** 109, (3), 317-25.

63. Hardin, D. S.; Ahn, C.; Rice, J.; Rice, M.; Rosenblatt, R., Elevated gluconeogenesis and lack of suppression by insulin contribute to cystic fibrosis-related diabetes. *Journal of Investigative Medicine* **2008,** 56, (3), 567-573.

64. Phan, J.; Gallagher, T.; Oliver, A.; England, W. E.; Whiteson, K., Fermentation products in the cystic fibrosis airways induce aggregation and dormancy-associated expression profiles in a CF clinical isolate of Pseudomonas aeruginosa. *FEMS Microbiol Lett* **2018,** 365, (10).

65. Joseloff, E.; Sha, W.; Bell, S. C.; Wetmore, D. R.; Lawton, K. A.; Milburn, M. V.; Ryals, J. A.; Guo, L.; Muhlebach, M. S., Serum metabolomics indicate altered cellular energy metabolism in children with cystic fibrosis. *Pediatric Pulmonology* **2014,** 49, (5), 463-72.

66. Roum, J. H.; Buhl, R.; McElvaney, N. G.; Borok, Z.; Crystal, R. G., Systemic deficiency of glutathione in cystic fibrosis. *Journal of Applied Physiology* **1993,** 75, (6), 2419-24.

67. Hudson, V. M., New insights into the pathogenesis of cystic fibrosis: pivotal role of glutathione system dysfunction and implications for therapy. *Treat. Respir. Med.* **2004,** 3, (6), 353-63.

68. Corradi, M.; Folesani, G.; Andreoli, R.; Manini, P.; Bodini, A.; Piacentini, G.; Carraro, S.; Zanconato, S.; Baraldi, E., Aldehydes and glutathione in exhaled breath condensate of children with asthma exacerbation. *American Journal of Respiratory and Critical Care Medicine* **2003,** 167, (3), 395-399.

69. Sewell, A. C.; Bohles, H. J., 4-Hydroxycyclohexanecarboxylic acid: a rare compound in urinary organic acid analysis. *Clinical Chemistry* **1991,** 37, (7), 1301-2.

70. Zheng, X.; Xie, G.; Zhao, A.; Zhao, L.; Yao, C.; Chiu, N. H.; Zhou, Z.; Bao, Y.; Jia, W.; Nicholson, J. K.; Jia, W., The footprints of gut microbial-mammalian co-metabolism. *Journal of Proteome Research* **2011,** 10, (12), 5512-22.

71. Wang, H.; Liu, J. S.; Peng, S. H.; Deng, X. Y.; Zhu, D. M.; Javidiparsijani, S.; Wang, G. R.; Li, D. Q.; Li, L. X.; Wang, Y. C.; Luo, J. M., Gut-lung crosstalk in pulmonary involvement with inflammatory bowel diseases. *World Journal of Gastroenterology* **2013,** 19, (40), 6794-6804.

72. McMillan, A.; Renaud, J. B.; Gloor, G. B.; Reid, G.; Sumarah, M. W., Post-acquisition filtering of salt cluster artefacts for LC-MS based human metabolomic studies. *Journal of Cheminformatics* **2016,** 8.

73. DeHaven, C. D.; Evans, A. M.; Dai, H. P.; Lawton, K. A., Organization of GC/MS and LC/MS metabolomics data into chemical libraries. *Journal of Cheminformatics* **2010,** 2.

74. Zang, X.; Monge, M. E.; McCarty, N. A.; Stecenko, A. A.; Fernandez, F. M., Feasibility of Early Detection of Cystic Fibrosis Acute Pulmonary Exacerbations by

Exhaled Breath Condensate Metabolomics: A Pilot Study. *J Proteome Res* **2017,** 16, (2), 550-558.

75. Twomey, K. B.; Alston, M.; An, S. Q.; O'Connell, O. J.; McCarthy, Y.; Swarbreck, D.; Febrer, M.; Dow, J. M.; Plant, B. J.; Ryan, R. P., Microbiota and metabolite profiling reveal specific alterations in bacterial community structure and environment in the cystic fibrosis airway during exacerbation. *PLoS One* **2013,** 8, (12), e82432.

76. Mirkovic, B.; Murray, M. A.; Lavelle, G. M.; Molloy, K.; Azim, A. A.; Gunaratnam, C.; Healy, F.; Slattery, D.; McNally, P.; Hatch, J.; Wolfgang, M.; Tunney, M. M.; Muhlebach, M. S.; Devery, R.; Greene, C. M.; McElvaney, N. G., The Role of Short-Chain Fatty Acids, Produced by Anaerobic Bacteria, in the Cystic Fibrosis Airway. *Am J Respir Crit Care Med* **2015,** 192, (11), 1314-24.

77. 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., HMDB 3.0--The Human Metabolome Database in 2013. *Nucleic Acids Res* **2013,** 41, (Database issue), D801-7.

78. Hardin, D. S.; Ellis, K. J.; Dyson, M.; Rice, J.; McConnell, R.; Seilheimer, D. K., Growth hormone decreases protein catabolism in children with cystic fibrosis. *Journal of Clinical Endocrinology & Metabolism* **2001,** 86, (9), 4424-4428.

79. Shepherd, R. W.; Holt, T. L.; Johnson, L. P.; Quirk, P.; Thomas, B. J., Leucine metabolism and body cell mass in cystic fibrosis. *Nutrition* **1995,** 11, (2), 138-41.

80. Barth, A. L.; Pitt, T. L., The high amino-acid content of sputum from cystic fibrosis patients promotes growth of auxotrophic Pseudomonas aeruginosa. *J Med Microbiol* **1996,** 45, (2), 110-9.

81. Arora, P. K.; Sharma, A.; Bae, H., Microbial Degradation of Indole and Its Derivatives. *Journal of Chemistry* **2015**.

82. Kim, S. K.; Park, H. Y.; Lee, J. H., Anthranilate deteriorates the structure of Pseudomonas aeruginosa biofilms and antagonizes the biofilm-enhancing indole effect. *Appl Environ Microbiol* **2015,** 81, (7), 2328-38.

83. Carpenedo, R.; Mannaioni, G.; Moroni, F., Oxindole, a sedative tryptophan metabolite, accumulates in blood and brain of rats with acute hepatic failure. *Journal of Neurochemistry* **1998,** 70, (5), 1998-2003.

84. Puccetti, M.; Paolicelli, G.; Oikonomou, V.; De Luca, A.; Renga, G.; Borghi, M.; Pariano, M.; Stincardini, C.; Scaringi, L.; Giovagnoli, S.; Ricci, M.; Romani, L.; Zelante, T., Towards Targeting the Aryl Hydrocarbon Receptor in Cystic Fibrosis. *Mediators Inflamm* **2018,** 2018, 1601486.

85. Wetmore, D. R.; Joseloff, E.; Pilewski, J.; Lee, D. P.; Lawton, K. A.; Mitchell, M. W.; Milburn, M. V.; Ryals, J. A.; Guo, L., Metabolomic profiling reveals biochemical pathways and biomarkers associated with pathogenesis in cystic fibrosis cells. *J Biol Chem* **2010,** 285, (40), 30516-22.

86. Tejero Rioseras, A.; Singh, K. D.; Nowak, N.; Gaugg, M. T.; Bruderer, T.; Zenobi, R.; Sinues, P. M., Real-Time Monitoring of Tricarboxylic Acid Metabolites in Exhaled Breath. *Anal Chem* **2018,** 90, (11), 6453-6460.

87. Tang, X. L.; Liu, J. X.; Dong, W.; Li, P.; Li, L.; Lin, C. R.; Zheng, Y. Q.; Hou, J. C.; Li, D., The Cardioprotective Effects of Citric Acid and L-Malic Acid on Myocardial Ischemia/Reperfusion Injury. *Evidence-Based Complementary and Alternative Medicine* **2013**.

88. Koeth, R. A.; Wang, Z.; Levison, B. S.; Buffa, J. A.; Org, E.; Sheehy, B. T.; Britt, E. B.; Fu, X.; Wu, Y.; Li, L.; Smith, J. D.; DiDonato, J. A.; Chen, J.; Li, H.; Wu, G. D.; Lewis, J. D.; Warrier, M.; Brown, J. M.; Krauss, R. M.; Tang, W. H.; Bushman, F. D.; Lusis, A. J.; Hazen, S. L., Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat Med* **2013,** 19, (5), 576-85.

89. Meadows, J. A.; Wargo, M. J., Carnitine in bacterial physiology and metabolism. *Microbiology* **2015,** 161, (6), 1161-74.

90. Strandvik, B., Fatty acid metabolism in cystic fibrosis. *Prostaglandins Leukot Essent Fatty Acids* **2010,** 83, (3), 121-9.

91. Freedman, S. D.; Blanco, P. G.; Zaman, M. M.; Shea, J. C.; Ollero, M.; Hopper, I. K.; Weed, D. A.; Gelrud, A.; Regan, M. M.; Laposata, M.; Alvarez, J. G.; O'Sullivan, B. P., Association of cystic fibrosis with abnormalities in fatty acid metabolism. *N Engl J Med* **2004,** 350, (6), 560-9.

92. Kovesi, T. A.; Lehotay, D. C.; Levison, H., Plasma carnitine levels in cystic fibrosis. *J Pediatr Gastroenterol Nutr* **1994,** 19, (4), 421-4.

93. Quinn, R. A.; Phelan, V. V.; Whiteson, K. L.; Garg, N.; Bailey, B. A.; Lim, Y. W.; Conrad, D. J.; Dorrestein, P. C.; Rohwer, F. L., Microbial, host and xenobiotic diversity in the cystic fibrosis sputum metabolome. *ISME J* **2016,** 10, (6), 1483-98.

94. Ayyar, L.; Hsu, J.; Jahoor, F.; Kao, C., The Relationship Of Arginine And Its Metaboites In Exhaled Breath Condensate To Lung Function. *American Journal of Respiratory and Critical Care Medicine* **2014,** 189.

95. Grasemann, H.; Pencharz, P. B., Arginine metabolism in patients with cystic fibrosis. *J Pediatr* **2013,** 163, (2), 317-9.

96. Grasemann, H.; Schwiertz, R.; Grasemann, C.; Vester, U.; Racke, K.; Ratjen, F., Decreased systemic bioavailability of L-arginine in patients with cystic fibrosis. *Respir Res* **2006,** 7, 87.

97. Gray, L. R.; Tompkins, S. C.; Taylor, E. B., Regulation of pyruvate metabolism and human disease. *Cell Mol Life Sci* **2014,** 71, (14), 2577-604.

98. Berger, A.; Dohnt, K.; Tielen, P.; Jahn, D.; Becker, J.; Wittmann, C., Robustness and Plasticity of Metabolic Pathway Flux among Uropathogenic Isolates of Pseudomonas aeruginosa. *Plos One* **2014,** 9, (4).

99. Xia, J. G.; Sinelnikov, I. V.; Han, B.; Wishart, D. S., MetaboAnalyst 3.0-making metabolomics more meaningful. *Nucleic Acids Research* **2015,** 43, (W1), W251-W257.

100. Li, J.; Greenwood, P. L.; Cockett, N. E.; Hadfield, T. S.; Vuocolo, T.; Byrne, K.; White, J. D.; Tellam, R. L.; Schirra, H. J., Impacts of the Callipyge mutation on ovine plasma metabolites and muscle fibre type. *PLoS One* **2014,** 9, (6), e99726.

PART II: FLOW INJECTION-ION MOBILITY-MS AND DIRECT INFUSION-ION MOBILITY-MS BASED NON-TARGETED METABOLOMICS FOR DISEASE DETECTION AND EARLY PREDICTION

CHAPTER 4. FLOW INJECTION-TRAVELING WAVE ION MOBILITY-MASS SPECTROMETRY FOR HIGH THROUGHPUT PROSTATE CANCER METABOLOMICS

Adapted with permission from
Zang X, Monge ME, David A. Gaul, Fernández FM. Flow Injection-Traveling Wave Ion mobility-Mass Spectrometry for Rapid Serum Metabolic Profiling. (to be submitted to *Analytical Chemistry*).

4.1 Abstract

Flow injection-traveling wave ion mobility-mass spectrometry (FI-TWIM-MS) was applied to the non-targeted metabolic profiling of serum extracts from 61 prostate cancer (PCa) patients and 42 controls with an analysis speed of 6 minutes per sample, including a wash run. Comprehensive data mining of the mobility-mass domain was used to discriminate species with various charge states and filter matrix salt cluster ions. Specific criteria were developed to ensure correct grouping of adducts, in-source fragments, and impurities in the dataset. Endogenous metabolites were identified with high confidence using FI-TWIM-MS/MS and collision cross-section (CCS) matching with chemical standards or CCS databases. PCa patient samples were distinguished from control samples with good accuracies (88.3-89.3%), sensitivities (88.5-90.2%), and specificities (88.1%) using supervised multivariate classification methods. Although largely underutilized in metabolomics studies, FI-TWIM-MS proved advantageous in terms of analysis speed, separation of ions in complex mixtures, improved signal-to-noise ratio, and reduction of spectral congestion. Results from this study showcase the potential

of FI-TWIM-MS as a high throughput metabolic profiling tool for large scale metabolomics studies.

4.2 MS-based Metabolic Profiling Strategies for High Sample Throughput

4.2.1 Direct Infusion and Flow Injection Mass Spectrometry in Metabolomics

The most commonly used MS-based metabolomics platform is the hyphenated technique of LC-MS. As mentioned in Chapter 1, LC-MS is widely adopted for non-targeted metabolic profiling, owing to its wide coverage and efficient separation of a large number of metabolites in complex biological matrices.[1-3] Coupling of LC to MS adds an orthogonal axis to mass-to-charge ratio (*m/z*) separations, therefore helping with metabolite identification in non-targeted studies, and lessening the extent of ion suppression in complex matrices.

Evaluation of biological alterations in human metabolomes is often confounded by the inherent biological variance within patient cohorts. One strategy to overcome this challenge is to increase the sample cohort size to hundreds or even thousands of patient samples, leading to the discovery of more robust metabolic signatures.[4, 5] Increasing cohort size, however, necessarily comes at the expense of higher study cost and larger instrument time demands. Most of that instrument time is typically consumed by the front-end LC separation.

Flow injection (FI) or direct infusion (DI) MS, without LC, maximizes analytical throughput and eliminate problems of retention time shifts due to column aging,[6, 7] therefore being well suited for larger patient cohorts. Compared to DI-MS, FI-MS can be more easily automated, with the added advantage of reducing the amount of sample required for analysis. Furthermore, addition of a post-ionization IMS dimension to FI-MS provides rapid separation of gas-phase ions based on CCS differences.[8, 9] Compared to FI-MS alone, the inclusion of an IM dimension i) reduces spectral congestion by separating compounds with different charge states and structural motifs that are distributed into distinct regions on the mobility–mass plot, ii) improves signal-to-noise ratios, iii) increases peak capacity, iv) incorporates CCS values that provide an additional molecular descriptor useful in assigning chemical structures,[10] and v) provides cleaner MS/MS spectra by avoiding precursor ion co-selection, while still maintaining high analysis speed.[11, 12] DI-IM-MS and FI-IM-MS, however, still remain largely unexplored in metabolomics, with the former applied only in a handful of studies without fully exploiting MS/MS or CCS information for metabolite identification.[13-15]

4.2.2 Biomarker Discovery for Prostate Cancer Detection

As described in Chapter 2, there has been a constant drive to discover new biomarkers for Prostate Cancer (PCa) detection to complement or replace existing ones.[16] Metabolomics studies have reported potential PCa biomarkers that include amino acids,[17-19] organic acids,[19-21] polyamines,[20, 21] lipids,[19, 22, 23] and carbohydrates,[18, 24] with the majority of these studies employing the more time-consuming LC-MS approach.[16, 25]

4.3 Hypothesis

In this study, we hypothesize that FI-IM-MS produces data that is comparable in terms of PCa detection performance to LC-MS, with the added advantage of faster sample analysis speed. To test this hypothesis, we performed a FI-IM-MS based non-targeted metabolomics study of serum samples from the same cohort described in Chapter 2 and analyzed the data, by combining with univariate and multivariate methods Based on this hypothesis, we expect to detect discriminant metabolites found in the LC-MS study described in Chapter 2, and discover new metabolites that contribute to sample classification, providing deeper insight into the perturbed biological pathways associated with PCa.

4.4 Materials and Methods

4.4.1 Chemicals

Ultrapure water with 18.2 MΩ·cm resistivity (Barnstead Nanopure UV ultrapure water system, USA), Optima LC-MS grade acetonitrile and methanol (Fisher Scientific, Suwannee, GA, USA), Omnisolv high-purity dichloromethane and HPLC grade acetone (EMD, Billerica, MA, USA) were used for mobile phase preparation, sample preparation and chemical standard solution preparation. LC-MS grade acetic acid, uric acid (≥99%), nonanedioic acid (azelaic acid) (98%), tryptophan, inosine (≥99%), glutamine, histidine,

leucine, isoleucine, L-allo-isoleucine, L-lysine (≥98%), uridine (≥99%), guanosine (≥98%), taurine (≥99%), indole (≥99%), phenylalanine, *m*-cresol (99%), *p*-cresol (99%), *o*-cresol and sodium cholesterol sulfate were purchased from Sigma-Aldrich Corporation (St. Louis, MO, USA). Phenylalanyl phenylalanine was purchased from MP Biomedicals (Solon, OH, USA). Phenylacetylglutamine was purchased from Bachem (Hauptstrasse, Bubendorf, Sitzerland). Lysophosphatidylcholine (LPC(18:0/0:0)) (1-stearoyl-2-hydroxy-*sn*-glycerol-3-phosphocholine) was purchased from Avanti Polar Lipids Inc. (Alabaster, AL, USA).

4.4.2 Cohort Description

Age-matched blood serum samples were obtained from 61 PCa patients (age range 49–65, mean age 59(4) years) and 42 controls (age range 45–76, mean age 58(7) years). At the 0.05 level, the population age means were not significantly different with the two-sample t test. The cohort ethnicity was as follows: 21 African American (20.4%); 72 Caucasian (69.9%); 5 Hispanic (4.9%); 2 Asian (1.9%); 2 Jewish ancestry (1.9%); and 1 unknown (1.0%). Serum samples correspond to a subgroup of the patient cohort described in Chapter 2. Thus, sample collection at Saint Joseph's Hospital of Atlanta after approval by the IRB was performed as previously described.

4.4.3 Sample Preparation

Sample preparation was performed as described in Chapter 2 to compare results. Briefly, frozen serum samples were thawed on ice, and 300 µL of extraction solution

(acetone: acetonitrile: methanol 1:1:1, cooled to -20 ˚C) was added to 100 µL of serum. Samples were vortex-mixed for 20 s and centrifuged at 16,000 × *g* for 5 min to pellet proteins. To remove most lipids and other non-polar metabolites, 800 µL of dichloromethane was added to ~350 µL of supernatant and vortex-mixed. Following the addition of 250 µL deionized water, samples were vortex-mixed and kept on ice for 10 min. The aqueous phase was subject to FI traveling wave IM (TWIM)-MS analysis. Sample blanks were prepared using deionized water instead of serum following the same procedure.

4.4.4 Metabolic Profiling by FI-TWIM-MS

Metabolomic analysis was performed on a Waters ACQUITY UPLC I-Class System fitted with a stainless steel union to bypass the chromatographic column, coupled to a Synapt G2-S HDMS system (Waters Corporation, Manchester, UK) equipped with an ESI source operated in negative mode. The column compartment was operated at room temperature. Instrument settings were as follows: capillary voltage 2.2 kV, cone voltage 45 V, source offset 50 V, source temperature 120 °C, desolvation gas temperature 300 °C, desolvation gas flow rate 600 L h^{-1}, nebulizer gas flow 7.0 bar, trap cell voltage 4 V (default in MS mode), transfer cell voltage 2 V (default in MS mode), EDC delay coefficient 1.41 V, helium cell gas flow rate 180 mL min^{-1}, IMS gas (N_2) flow rate 90 mL min^{-1}, trap DC entrance 3.0 V, bias 45 V, trap DC 0.0 V, exit -6.0 V, IMS DC settings entrance 25.0 V, helium cell DC 50.0 V, helium exit -10.0 V, bias 7.0 V, exit 0.0 V, transfer DC entrance 5.0 V, exit 15.0 V, trap wave velocity 311 m s^{-1}, wave height 4.0 V, IMS wave velocity 650 m s^{-1}, wave height 40.0 V, transfer wave velocity 175 m s^{-1}, wave

height 4.0 V, StepWave 1 wave velocity 300 m s^{-1}, wave height 15.0 V, StepWave 2 wave velocity 300 m s^{-1}, wave height 15.0 V, StepWave 2 offset 25.0 V, StepWave differential aperture1 3.0 V, StepWave differential aperture2 0.0 V, source ion guide wave velocity 300 m s^{-1}, wave height 1.0 V, StepWave Radio frequency (RF) offset 300.0 V, ion guide RF offset 350.0 V, IMS wave delay 1000 µs. IM resolution was ~40 (FWHM). IM cell pressure was ~3.13 mbar. The instrument was calibrated in the range of *m/z* 50–1200 using a 0.5 mM sodium formate solution prepared in 90:10 2-propanol/water v/v. The injection volume was 5 µL using partial-loop needle overfill mode. The flow rate was set to 0.05 mL min^{-1} with a mobile phase consisting of acetonitrile/water (50:50 v/v) with 0.1% acetic acid. Mass spectra were acquired in profile mode over the range of *m/z* 50–1200 in the "mobility-TOF" resolution mode. Scan time was 1 s and run time per sample was 3 min, followed by a 3 min wash run. During data acquisition, spectra were drift corrected using a 2 ng μL^{-1} leucine enkephalin reference spray infused at 2 µL min^{-1}. Data acquisition was carried out using MassLynx ver. 4.1 (Waters Corp., Milford, MA, USA). Sample run order was randomized and sample preparation blanks were included. A wash run followed each sample, and was used to prevent carryover by injecting 5 µL of deionized water using the same analytical method as for samples. After each injection, 200 µL of the strong wash solvent acetonitrile and 600 µL of the weak wash solvent (10:90 acetonitrile:water) were used to wash the needle.

4.4.5 Compound Identification

For FI-TWIM-MS/MS experiments, precursor ions were fragmented in the transfer cell with ultrahigh-purity argon (≥99.999%) as the collision gas. Instrument settings were the same as those in MS mode and the trap cell voltage was kept at 4 V. LM resolution was set to 15.0. Two methods were used for tandem MS experiments: automated data dependent acquisition (DDA) and individual MS/MS. For the DDA method, a list of precursor *m/z* values was entered. The scan range was *m/z* 50–600, covering leucine enkephalin at *m/z* 554.262 in negative mode to ensure mass correction could be applied. Survey scan time was set from 0.3 to 0.9 min with 0.2 s scan time. The intensity threshold was set to 5000 so the survey scan switched to MS/MS when the individual ion intensity exceeded this value. The criteria for MS/MS switching to MS was set to be either when accumulated total ion intensity was above 10^5, or when acquisition time reached 0.25 s. Tandem MS scan time was set to 0.1 s. Interscan time was 0.01 s for both survey and MS/MS scans. A voltage of 25 V was applied to the transfer cell. For individual MS/MS methods, different voltage values between 20 to 50 V were applied to the transfer cell, depending on the specific precursor ion under investigation.

For CCS measurement, a poly-DL-alanine solution was used as CCS reference in negative ion mode (10 mg L^{-1} in 50:50 v/v acetonitrile/water). Calibration was performed using singly charged oligomers from n = 3 to 14, covering a mass range from 230 to 1012 Da and a CCS range from 150 to 308 $Å^2$. CCS values were derived using previously described procedures.[26] In MS/MS mode, Dt values were shorter due to elevated collision voltage applied in the transfer cell, leading to increased speed of the ions traversing that chamber. In order to correct for this shift, CCS calibration was performed at each elevated collision voltage applied to the transfer cell in MS/MS experiments. CCS values

for precursor ions in both MS and MS/MS modes were derived, and compared well within 2% tolerance.

4.4.6 Data Processing and Analysis

A general data acquisition and analysis workflow is shown in Figure 4.1. FI-TWIM-MS data was loaded into Progenesis QI ver. 2.0 (Nonlinear Dynamics, Waters Corp.) for processing. Adduct ions such as $[M-H]^-$, $[M-2H]^{2-}$ and $[M-3H]^{3-}$ were selected and no alignment was chosen since there was no chromatography. For the peak picking step, the "retention time" range was set between 0.3 and 0.8 min to remove ghost peaks with apexes outside this range. Sensitivity was set to automatic (default value of 3), and no peak width was set in order to maximize the number of detected features. Each feature was associated with specific *m/z*, Dt, and run time (FI time) value. Manual investigation of the detected spectral features on the mobility–mass plot indicated a few cases where the charge state assignment by the Progenesis software was incorrect, due to either low ion abundance and poor peak shape, or false positive isotope cluster grouping based on adjacent noisy peaks. These wrongly-assigned charge states were manually corrected (Table 4.1).

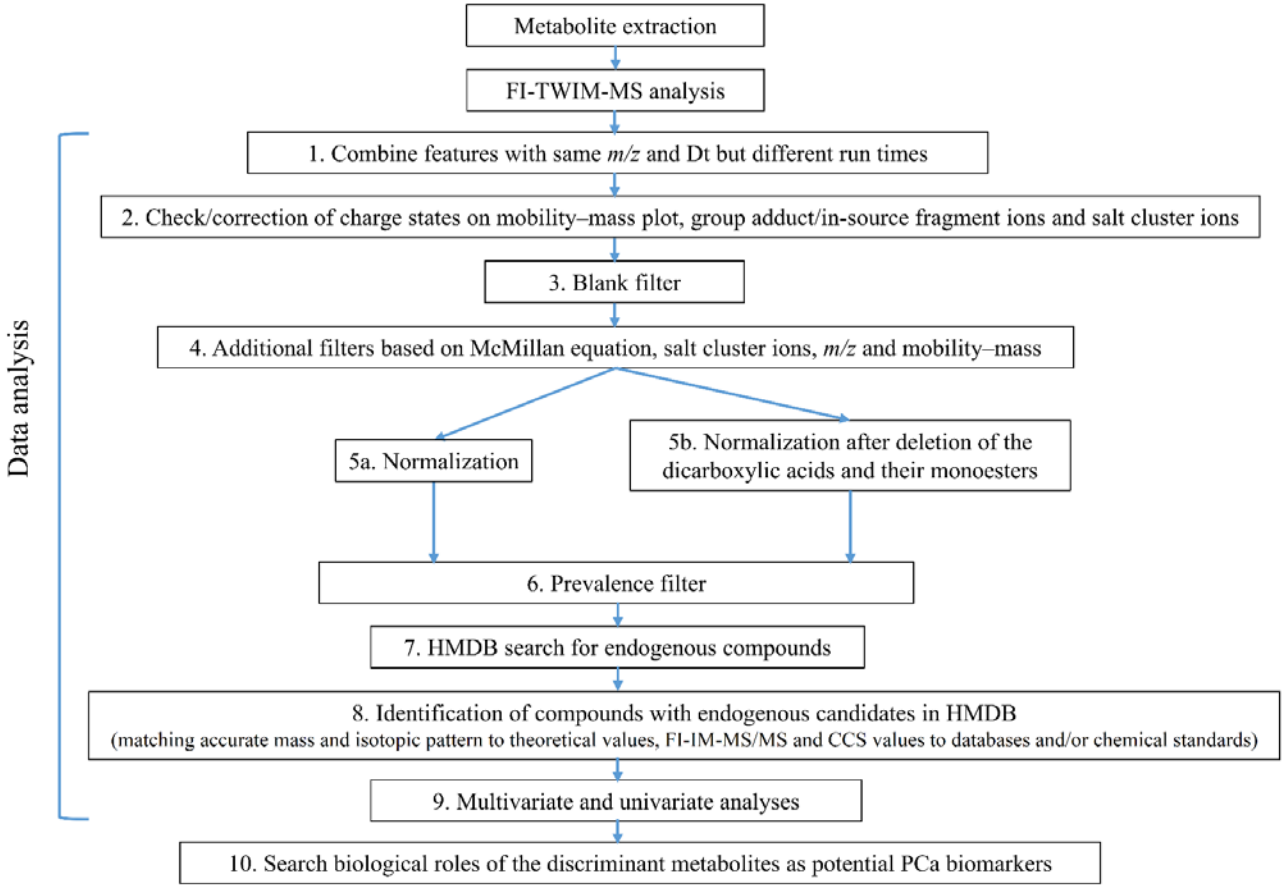

Figure 4.1: FI-TWIM-MS metabolomics workflow used for data collection and analysis.

Table 4.1: Compound charge states found in various areas of mobility–mass plots.

mobility–mass plot area	charge state assigned by Progenesis software		
	-1	-2	-3
1st	Singly charged species, including all identified endogenous metabolites	Only a few signals, mostly doubly charged compounds that were manually checked	Only a few triply charged species. Many noisy peaks with low intensity and poor peak shape, which were manually re-assigned.
2nd	Mostly singly charged chloride salt cluster ions	Doubly charged compounds	Only a few signals, including singly charged chloride salt cluster ions, triply charged compounds and low intensity peaks, which were manual re-assigned.
3rd	Unidentified low intensity singly charged features		

Features with the similar *m/z* (4 mDa tolerance) and Dt (0.06 ms tolerance), but different run times were combined into a single compound (Table 4.2). Following this grouping, features corresponding to adduct ions and in-source generated fragment ions were grouped with the corresponding $[M-H]^-$ ion. Stringent criteria were adopted to avoid false positives, including specific *m/z* differences, relative abundances, and ion species correlations (Table 4.2).

Table 4.2: Criteria applied for combining features, adduct ions, in-source fragment ions, and salt clusters in FI-TWIM-MS data.

Criterion	Δm (mDa)	ΔDt (ms)	Specific Rules	Numbers of Compounds or Features
Spectral feature grouping by *m/z* and Dt	4	0.06	N/A	361 features with different FI run times, but similar *m/z* and Dt were found.
Adduct/in-source fragment ion grouping[a]. Adduct ions considered included: $[M-H]^-$, $[M+Na-2H]^-$, $[M+CH_3COO]^-$,$[M-CH_3]^-$ $[M+CH_3COO]^-$ and $[M+HCOO]^-$(for lysophosphatidylcholine adducts). In-source fragment ions considered included: $[M-H_2O-H]^-$, $[M-SO_3-H]^-$, $[M-CO_2-H]^-$, $[M-NH_3-H]^-$, $[M-CH_3COOH-H]^-$, $[M-HCOOH-H]^-$.	1.5	N/A	1. $[M-H]^-$ ion must be present in group, $[M-H]^-$ ion must be most abundant ion for >90% samples, except for LPC adducts (no $[M-H]^-$). 2. Only singly charged compounds are considered. 3. After blank filtering, either all ion adducts in a group should be present, or all should be absent for >90% of samples. 4. Correlation between features[b] >0.85 to ensure they correspond to different adducts of the same compound[27] 5. Compounds in the group should be in the same mobility–mass plot region.	83 features were grouped into 37 compounds based on the presence of multiple adducts or in-source fragments.
Manual grouping of all chloride salt cluster isotopic signals not correctly grouped by Progenesis. Possible isotopes considered: X, X+1.9971, X+1.9971×2, X+1.9971×3, X+1.9971×4, X+1.9971×5.	4	0.06	1. Ratio between the second largest isotope abundance and the most abundant isotope >0.3. 2. Mass defect >0.3. 3. Satisfy rules 2-5 in adduct/in-source fragment ion grouping criteria.	A total of 248 chloride-containing species were found, these were grouped into 104 NaCl cluster compounds including $[Na_nCl_{n+1}]^-$ (n=3–9). Of the 248 species, 139 overlapped with compounds filtered by the McMillan filter[28].
Identification of cases where Progenesis only identified a single isotopic peak from a chloride salt cluster. Features belonging to salt clusters that matched these specific criteria were manually removed.	N/A	N/A	1. Feature should be in the second mobility–mass plot region. 2. Feature should be singly charged with even nominal mass and mass defect >0.3.	A total of 492 features matched. Of these, 316 overlapped with compounds filtered by McMillan filter[28].
Identification of sodium acetate clusters $[CH_3COO(CH_3COONa)_n]^-$ (n=3–11)	1.5	N/A	N/A	A total of 9 cluster ion species were detected.

[a]Adduct ions, in-source fragment ions, and cluster ions often detected under the described experimental conditions are shown. Adduct/in-source fragment ion grouping was determined by comparing the mass differences between each adduct/in-source fragment ion and the $[M-H]^-$ ion.

[b]$\mathrm{Corr}_{AB} = \frac{\sum_{i=1}^{n}(x_{A,i}-\bar{x}_A)(x_{B,i}-\bar{x}_B)}{\sqrt{\sum_{i=1}^{n}(x_{A,i}-\bar{x}_A)^2}\sqrt{\sum_{i=1}^{n}(x_{B,i}-\bar{x}_B)^2}}$ (A and B represent two features; *n* is the number of samples).

Salt cluster ions, including sodium chloride clusters ($[Na_nCl_{n+1}]^-$, (n=3–9)) and sodium acetate clusters ($[CH_3COO(CH_3COONa)_n]^-$, (n=3–11)), which are known artifacts observed in ESI of complex samples[28], were also detected and grouped according to criteria outlined in Table 4.2. Many of these cluster ion signals were incorrectly grouped or were incorrectly assigned by Progenesis, and required manual curation of the feature list. An example of a sodium chloride cluster ion, annotated as $[Na_4Cl_5]^-$, is shown in Figure 4.2. The distinct isotopic pattern due to the presence of multiple chlorine atoms is clearly observed in Figure 4.2A (top), with an excellent match to the expected abundances (Figure 4.2A bottom, 4.2B). Its identity was confirmed by FI-TWIM-MS/MS experiments (Figure 4.2C). The average of the isotopic distributions of a variety of other sodium chloride cluster ions across all samples was calculated (Figure 4.2D), which allowed to develop rules for the grouping of these species in the dataset (Table 4.2).

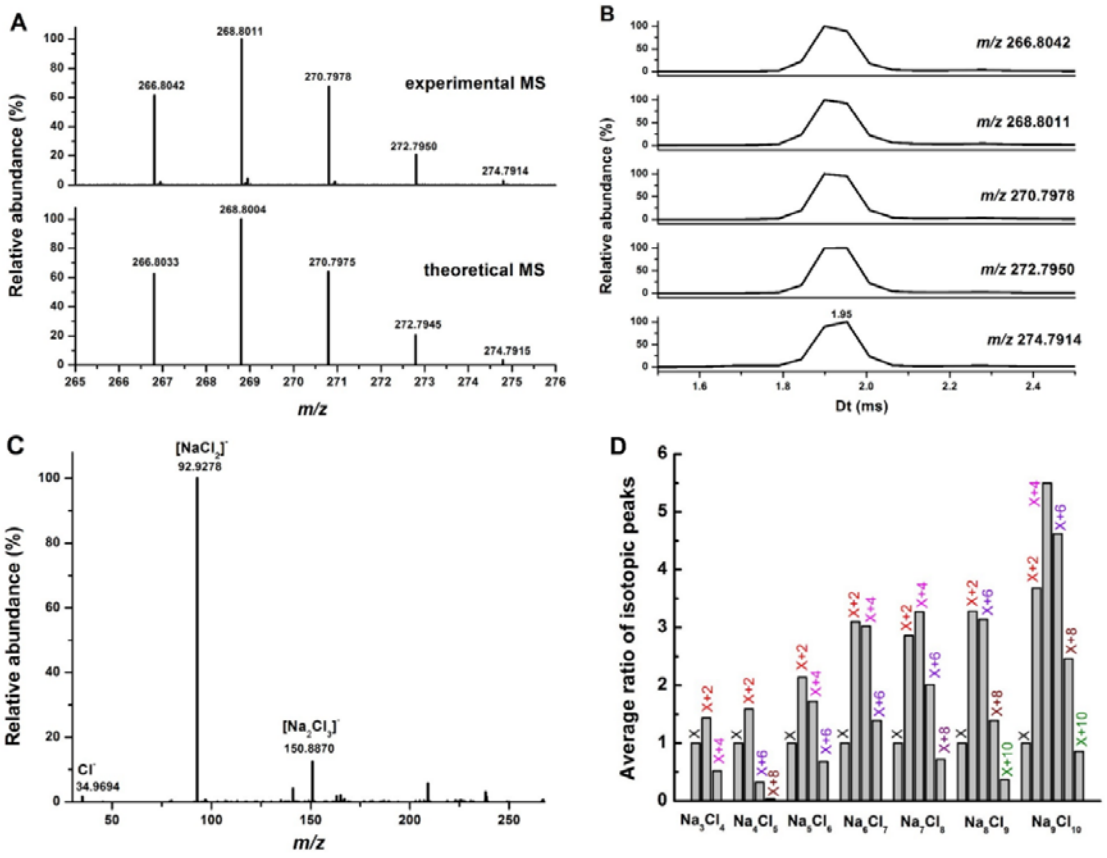

Figure 4.2: (A) Negative ion mode isotopic cluster for the species observed with *m/z* 266.8039 (top) and the theoretical mass spectrum for $[Na_4Cl_5]^-$ (bottom). (B) TWIM extracted ion chronograms for each isotopic peak observed in the top spectrum of (A). (C) FI-TWIM-MS/MS for the precursor with *m/z* 266.8039 for a 25 V transfer cell voltage. (D) Average ratios of isotopic to monoisotopic peak of $[Na_nCl_{n+1}]^-$ (n=3-9) in the raw data matrix.

Following grouping of adduct/in-source fragment ions and salt cluster ions, the dataset was processed using four filters applied sequentially. First, a blank filter was applied where if the compound abundance in the sample was smaller than 5 times the average abundance in the sample blanks, it was set to zero, and compounds with zero values in more than 80% of the samples were removed. Next, a McMillan filter[28] was applied to screen for salt cluster ions with high mass defects. The equation suggested by McMillan *et al.* for positive mode was modified for its application in the negative ion mode; i.e., $y = 0.00112([M-H]^- + 2H) + 0.01953 – 2H +2$. Next, any remaining salt cluster ions that were not removed by the McMilllan filter[28] were manually eliminated using the grouping criteria described in Table 4.2. Lastly, only compounds with m/z <610 were retained since most of the species of interest in the LC-MS PCa study[19] described in Chapter 2 were in this range. The resulting dataset was normalized by total ion intensity, yielding a dataset named “dataset #1”. A different dataset (dataset #2) was generated by total ion intensity normalization after the deletion of species identified as dicarboxylic acids and their monoesters (see discussion below) which might be of either endogenous or exogenous origin. This second dataset was useful in investigating the effect of deletion of these compounds with uncertain origin on the following univariate and multivariate analyses results.

Following normalization, a prevalence filter was applied to the data so that only compounds present in at least 50% of any class were kept, leaving 237 compounds in dataset #1 and 226 in dataset #2. The features in these datasets were searched in the HMDB[29] with a mass tolerance of 4 mDa, and only those matching metabolites of endogenous origin were kept. These features were identified based on i) matching

accurate mass and isotopic pattern to theoretical values, ii) matching MS/MS spectra within 8 mDa mass accuracy, and iii) matching experimental CCS values within ~ ±3% of chemical standard and/or database values.

The identified compounds in dataset #1 matching discriminant metabolites from the LC-MS PCa metabolomics study[19] described in Chapter 2 were grouped into a new dataset named "dataset A", which was used to build a model for sample classification *via* oPLS-DA,[30, 31] using Matlab R2015a (The Mathworks, Natick, MA with PLS-Toolbox, ver. 8.0, Eigenvector Research, Inc., Manson, WA). All identified endogenous metabolites in dataset #2 were grouped into a new dataset named "dataset B". Reverse interval PLS-DA (iPLS-DA) was applied to this dataset after autoscale pre-processing to find the optimum number of oPLS-DA latent variables (LVs) and optimum set of discriminant metabolites that maximized model classification accuracy. The maximum number of LV was set to 6 and three-block cross-validation was applied.

4.5 Results and Discussion

4.5.1 FI-TWIM-MS serum profiling

As discussed in Chapter 2, PCa detection was achieved using UPLC-MS-based serum metabolic profiling,[19] with each sample and wash run taking 18 and 8 min, respectively. In this study, FI-TWIM-MS was investigated as a new approach for metabolic fingerprinting with an analysis speed of 3 min for each run followed by a 3 min wash run. A typical FI-TWIM-MS profile of a serum extract from a PCa patient is shown in Figure 4.3. The total ion chronogram reached maximum intensity after ~0.5 min with

analytes eluting within ~1 min after injection, and beginning to tail off slowly as compounds dispersed in the flow system (Figure 4.3A). Figure 4.3B shows the corresponding TWIM total ion chronogram that, as expected, suggests the presence of ionic species with a variety of shapes and charge states exhibiting drift times in the 1-10 ms range. The combined mass spectrum corresponding to the data in Figure 4.3A and 4.3B is shown in Figure 4.3C. These raw data were cleaned up by stringent data processing methods to filter out unwanted features and group redundant ionic species. The number of species remaining after each major step in the data processing and analysis workflow is shown in Table 4.3.

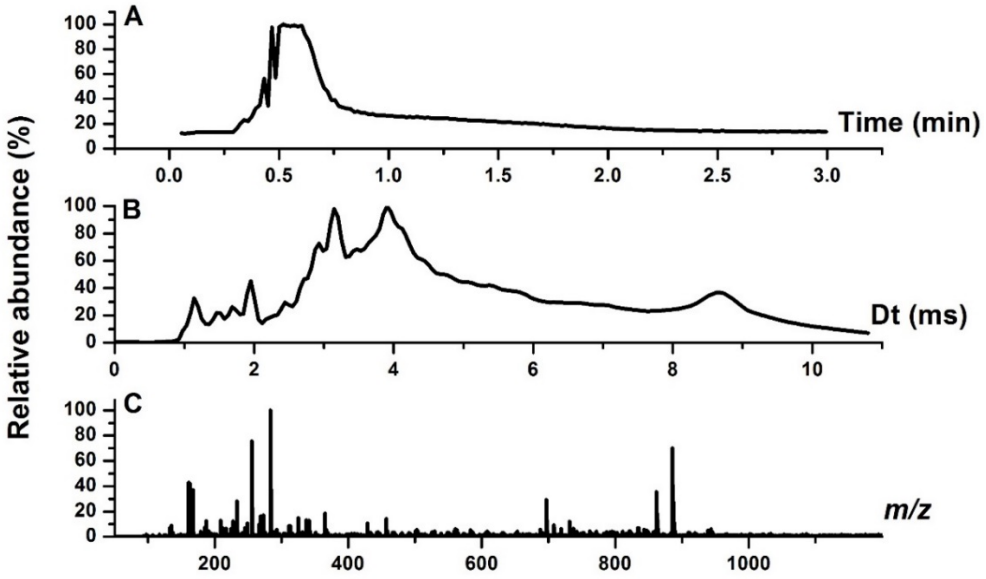

Figure 4.3: Typical FI total ion chronogram (A), TWIM total ion chronogram (B), and combined mass spectrum (C) of a serum extract from a PCa patient.

Table 4.3: Number of ionic compounds retained after each major data processing step.

	mobility–mass plot area									total
	1st			2nd			3rd			
charge	-1	-2	-3	-1	-2	-3	-1	-2	-3	
total compounds[a]	1065	11	6	670	211	10	413	195	26	2607
after blank filter	691	10	2	407	205	10	381	148	24	1878
after McMillan filter[24] and manual salt cluster ion removal	535	10	1	18	202	3	367	148	21	1305
after *m/z* range & prevalence filters	229	8	0							237

[a]After combining features and grouping adducts/in-source fragment ions.

4.5.2 Compound Identification and Validation

Following application of different filters and dataset normalization, features were searched in the HMDB[29] and only those with tentative endogenous identities based on accurate mass matching were retained. The remaining species were identified by matching CCS values and tandem MS spectra to databases, and, whenever possible, their identity was validated by comparing Dt, CCS and FI-TWIM-MS/MS spectra with chemical standards. An example of such identification workflow is illustrated in Figure 4.4 for a feature detected in MS mode with *m/z* 267.0734 and Dt 2.39 ms. Figure 4.4A shows the FI total ion chronogram of a PCa serum sample when performing MS/MS experiments (top panel), the associated TWIM total ion chronogram (middle panel) and the corresponding combined MS/MS spectrum (bottom panel) extracted across Dt=1.5-3 ms. Each of the three peaks shown in the TWIM total ion chronogram with drift times of 1.88, 2.28 and 2.63 ms provided different MS/MS spectra (Figure 4.4B). Following CCS

calibration, only the precursor ion extracted from the MS/MS mobility species with Dt = 2.28 ms matched the CCS of the feature of interest (*m/z* 267.0733) observed in the MS run (Figure 4.4A, middle panel inset) with an error of 0.66%. This species yielded the MS/MS spectrum shown in Figure 4.4B (middle panel). To further confirm that the species detected in Figure 4.4B (middle) correspond to the precursor of interest, fragment ions drift times were matched to that of the precursor ion. Figure 4.4C shows that only 3 out of 5 fragment ions observed in the MS/MS spectrum of Figure 4.4B (middle panel) were actually product ions that aligned with the Dt of the precursor ion of interest. Further analysis of the spectral data indicated that the species at *m/z* 92.9272, identified as $[NaCl_2]^-$, was a fragment ion of the cluster $[Na_4Cl_5]^-$ with Dt = 1.88 ms and *m/z* 266.8039 (Figure 4.2), which partially overlapped with the species at Dt = 2.28 ms in the TWIM chronogram (Figure 4.4A, middle). Accurate mass-based search in the HMDB[29] suggested inosine as the most likely candidate for this compound. Its tentative identity was confirmed by matching the experimental MS/MS spectrum with that of an inosine chemical standard (Figure 4.4D), and by comparing the MS/MS spectrum in the Metlin database[32] (Figure 4.4E). Further validation of this metabolite's identity was achieved by matching the MS-mode Dt and CCS with those of an inosine chemical standard and literature CCS values[8]. Similar procedures were applied to the identification of all compounds retained following application of data filters (Table 4.4 and 4.5).

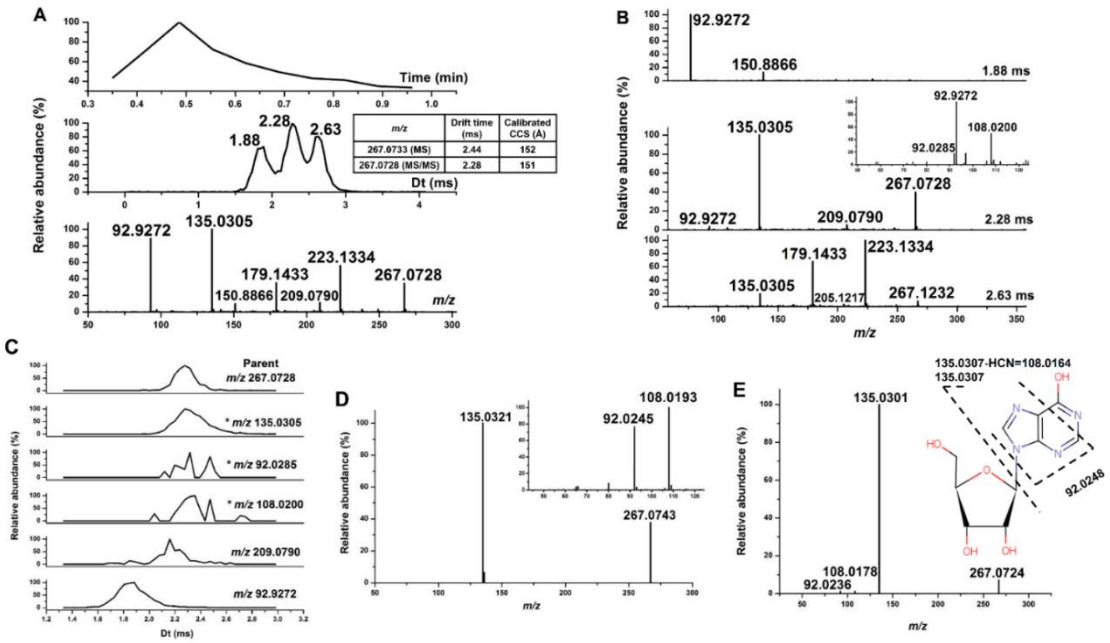

Figure 4.4: (A) FI-TWIM-MS/MS results for a PCa serum extract sample feature detected in MS mode with *m/z* 267.0734 and Dt 2.39 ms. Tandem MS data was acquired by applying 25 V applied to the transfer cell. Typical FI chronogram (top plot), TWIM total ion chronogram (middle plot, with inset showing calibrated CCS values of precursor ions detected in both MS and MS/MS modes), and the corresponding total MS/MS spectrum (bottom plot). Small drift time differences were observed between MS and MS/MS modes due to elevated bias voltage in the transfer cell when performing fragmentation experiments (B) Extracted MS/MS spectra derived at 1.88, 2.28, and 2.63 ms. (C) TWIM extracted ion chronograms for fragment ions with a mass tolerance of 10 mDa. Product ion peaks aligned with the precursor ion with *m/z* 267.0734 are labeled with asterisks. (D) Inosine standard MS/MS spectrum obtained in negative ion mode using 25 V in the transfer cell. The observed Dt and CCS for this standard were 2.44 ms and 152 A^2, respectively. (E) Metlin MS/MS spectrum for inosine obtained in negative ion mode with a collision cell voltage of 20 V. All product ions matched within a 4 mDa error.

Table 4.4: Chemical identities of compounds involved in the oPLS-DA models created.

model	code	*m/z*	Dt (ms)	most abundant ion type	elemental formula	Δm (mDa)	tentative metabolite identity	mean fold change (PCa vs. control)		*p*-value*		metabolite ID validation	CCS (Å²) (% error from database, % error from standard)
								panel A	panel B	panel A	panel B		
A, B	1	311.1401	3.15	$[M-H]^-$	$C_{18}H_{20}N_2O_3$	0.0	phenylalanylphenylalanine (phe-phe)	-1.9	-2.1	**9.1e-5**	**4.4e-5**	Dt, MS/MS[b, d]	175 (N/A, 0)
A, B	2	203.0848	2.17	$[M-H]^-$	$C_{11}H_{12}N_2O_2$	2.2	tryptophan	-1.1	-1.1	0.24	0.026	Dt, MS/MS[a, b, d]	145 (-0.68, 0)
B	3	116.0500	1.55	$[M-H]^-$	C_8H_7N	-0.6	indole	1.2	1.1		0.15	Dt[d]	126 (N/A, 0.81)
A, B	4	167.0214	1.52	$[M-H]^-$	$C_5H_4N_4O_3$	0.3	uric acid	1.3	1.2	**5.2e-9**	**8.4e-6**	Dt, MS/MS[a, b, d]	120 (-2.4, 0)
B	5	187.0075	1.79	$[M-H]^-$	$C_7H_8O_4S$	0.5	*p*-cresol sulfate	-1.0	-1.1		0.86	Dt, MS/MS[a, b, d] for in-source fragment *p*-cresol	131 for 187.0075, 122 for 107.0495 (in-source fragment of *p*-cresol) (N/A, 1.66)
A, B	6	476.2769	4.56	$[M-H]^-$	$C_{23}H_{44}NO_7P$	-1.3	lysophosphatidylethanolamine (LPE(0:0/18:2)), LPE(18:2/0:0)	1.2	1.2	0.23	0.48	MS/MS[b]	213
A, B	7	263.1052	2.60	$[M-H]^-$	$C_{13}H_{16}N_2O_4$	1.5	phenylacetylglutamine	-2.0	-2.1	0.0094	**0.0029**	Dt, MS/MS[b, d]	158 (N/A, -1.3)
B	8	145.0975	1.74	$[M-H]^-$	$C_6H_{14}N_2O_2$	-0.8	lysine		-1.0		0.44	Dt[d]	131 (-1.5, -1.5)
A, B	9	504.3095	4.94	$[M-CH_3]^-$	$C_{26}H_{50}NO_7P$	0.0	LPC(18:2/0:0), LPC(0:0/18:2)	-1.0	-1.1	0.67	0.30	MS/MS[b]	222

Table 4.4 (continued).

B	10	130.0857	1.68	$[M-H]^-$	$C_6H_{13}NO_2$	-1.7	leucine, isoleucine, allo-isoleucine	1.2		0.099	Dt[d]	130 (0.78, 0)
A	11	508.3423	5.13	$[M-CH_3]^-$	$C_{26}H_{54}NO_7P$	1.5	LPC(18:0/0:0), LPC(0:0/18:0)	1.1		0.042	Dt, MS/MS[b, d] for LPC (18:0/0:0)	227 (N/A, 0 for LPC(18:0/0:0))
A	12	369.1729	3.69	$[M-H]^-$	$C_{19}H_{30}O_5S$	-1.2	androsterone sulfate, 5α-dihydrotestosterone sulfate	1.2		0.11	MS/MS[a, b]	190
A	13	187.0971	1.90	$[M-H]^-$	$C_9H_{16}O_4$	-0.5	nonanedioic acid (azelaic acid)	-2.0		**2.2e-12**	Dt, MS/MS[a, b, d]	135 (N/A, 0)
A	14	273.1711	2.71	$[M-H]^-$	$C_{14}H_{26}O_5$	0.4	azelaic acid mono-hydroxy-pentyl ester (major) and undecanedioic acid mono-hydroxy-propyl ester	-1.3		**5.7e-9**	MS/MS[b] for both, [c] for azelaic acid mono-hydroxy-pentyl ester	162
A	15	287.1863	2.88	$[M-H]^-$	$C_{15}H_{28}O_5$	0.0	sebacic acid mono-hydroxy-pentyl ester (major) and dodecanedioic acid mono-hydroxy-propyl ester	-1.3		**2.5e-4**	MS/MS[b]	167

*p-values were calculated by t-test if the data was normally distributed in both PCa and control subject groups, otherwise Wilcoxon rank-sum test was used. Statistically significant fold changes with Bonferroni correction are shown in bold font**.*** All identified compounds were accurate mass and isotopic pattern-matched. [a] MS/MS matched the Metlin database; [b] MS/MS matched manual fragmentation analysis; [c] MS/MS matched literature evidence; [d] Dt and MS/MS matched chemical standard.

Table 4.5: Chemical identities of the compounds not selected by iPLS-DA to create model B.

m/z	Dt (ms)	most abundant ion type	elemental formula	Δm (mDa)	tentative metabolite annotation	metabolite ID validation	CCS (Å²) (% error from database, % error from standard)
274.1038	2.44	$[M-H]^-$	$C_{10}H_{17}N_3O_6$	-0.7	γ-glutamyl glutamine	MS/MS[b]	152
145.0611	1.57	$[M-H]^-$	$C_5H_{10}N_2O_3$	-0.7	glutamine	Dt, MS/MS[a, b, c]	124 (-2.4, 0)
282.0841	2.50	$[M-H]^-$	$C_{10}H_{13}N_5O_5$	-0.2	guanosine	Dt, MS/MS[a, b, c]	154 (-1.9, -2.5)
154.0616	1.57	$[M-H]^-$	$C_6H_9N_3O_2$	-0.7	histidine	Dt, MS/MS[a, b, c]	123 (-3.1, 0)
267.0734	2.39	$[M-H]^-$	$C_{10}H_{12}N_4O_5$	0.0	inosine	Dt, MS/MS[a, b, c]	151 (-1.3, -0.66)
164.0714	1.84	$[M-H]^-$	$C_9H_{11}NO_2$	-0.3	phenylalanine	Dt, MS/MS[a, b, c]	134 (-1.5, 0)
243.0609	2.17	$[M-H]^-$	$C_9H_{12}N_2O_6$	-1.4	pseudouridine and uridine	Dt, MS/MS[a, b, c] (uridine); MS/MS[b] for pseudouridine	143 (-2.1, -1.4 for uridine)
124.0089	1.46	$[M-H]^-$	$C_2H_7NO_3S$	1.5	taurine	Dt, MS/MS[a, b, c]	120 (1.7, 0)
480.3115	4.83	$[M-CH_3]^-$ for LPC, $[M-H]^-$ for LPE	$C_{24}H_{50}NO_7P$ (LPC) and $C_{23}H_{48}NO_7P$ (LPE)	2.0	LPC(16:0) and LPE(18:0/0:0), LPE(0:0/18:0)	MS/MS[b]	220 (0.46 for LPE(18:0/0:0), N/A)
452.2782	4.50	$[M-H]^-$	$C_{21}H_{44}NO_7P$	0.0	LPE(0:0/16:0), LPE(16:0/0:0)	MS/MS[b]	211 (-0.47 for LPE(16:0/0:0), N/A)
512.2682	4.50	$[M-H]^-$	$C_{26}H_{43}NO_7S$	-0.5	sulfolithocholylglycine	MS/MS[b]	211
478.2930	4.69	$[M-H]^-$	$C_{23}H_{46}NO_7P$	-0.9	LPE(0:0/18:1), LPE(18:1/0:0)	MS/MS[b]	217 (-0.46 for LPE(18:1/0:0), N/A)
514.2826	4.29	$[M-H]^-$	$C_{26}H_{45}NO_7S$	-1.8	taurohyocholate, taurocholic acid	MS/MS[a, b] for taurocholic acid; [b] for taurohyocholate	205

Table 4.5 (continued).

498.2893	4.23	$[M-H]^-$	$C_{26}H_{45}NO_6S$	-0.1	tauroursodeoxycholic acid, taurodeoxycholic acid, taurochenodesoxycholic acid	MS/MS[a, b] for tauroursodeoxycholic acid and taurodeoxycholic acid; [b] for taurochenodesoxycholic acid	203
465.3042	5.43	$[M-H]^-$	$C_{27}H_{46}O_4S$	-0.2	cholesterol sulfate	Dt, MS/MS[a, b, c]	235 (N/A, -0.84)
367.1582	3.63	$[M-H]^-$	$C_{19}H_{28}O_5S$	-0.2	dehydroepiandrosterone sulfate, testosterone sulfate	MS/MS[a, b] for dehydroepiandrosterone sulfate; [b] for testosterone sulfate	188

4.5.3 Distribution of Compounds in the Mobility–Mass Plot

The distribution of different classes of ionic species in the FI-TWIM-MS dataset is shown in a mobility–mass plot (Figure 4.5), with each symbol representing an ionic compound with a specific pair of Dt and *m/z* values. Ionic species were distributed across three distinct areas, separated by linear boundaries. The first area consisted mostly of singly charged compounds, including identified polar and lipid metabolites, dicarboxylic acids and their corresponding monoesters, and sodium acetate clusters. Features with high mass defect as defined by the McMillan filter[28] laid between the first and second plot regions. The second plot area mostly consisted of doubly charged compounds, together with some singly charged sodium chloride cluster ions, and very few triply charged compounds. The third plot area contained unidentified, low intensity, singly charged features with $m/z > 600$. Due to their relatively poor signal-to-noise ratios, species in this region were removed from the dataset.

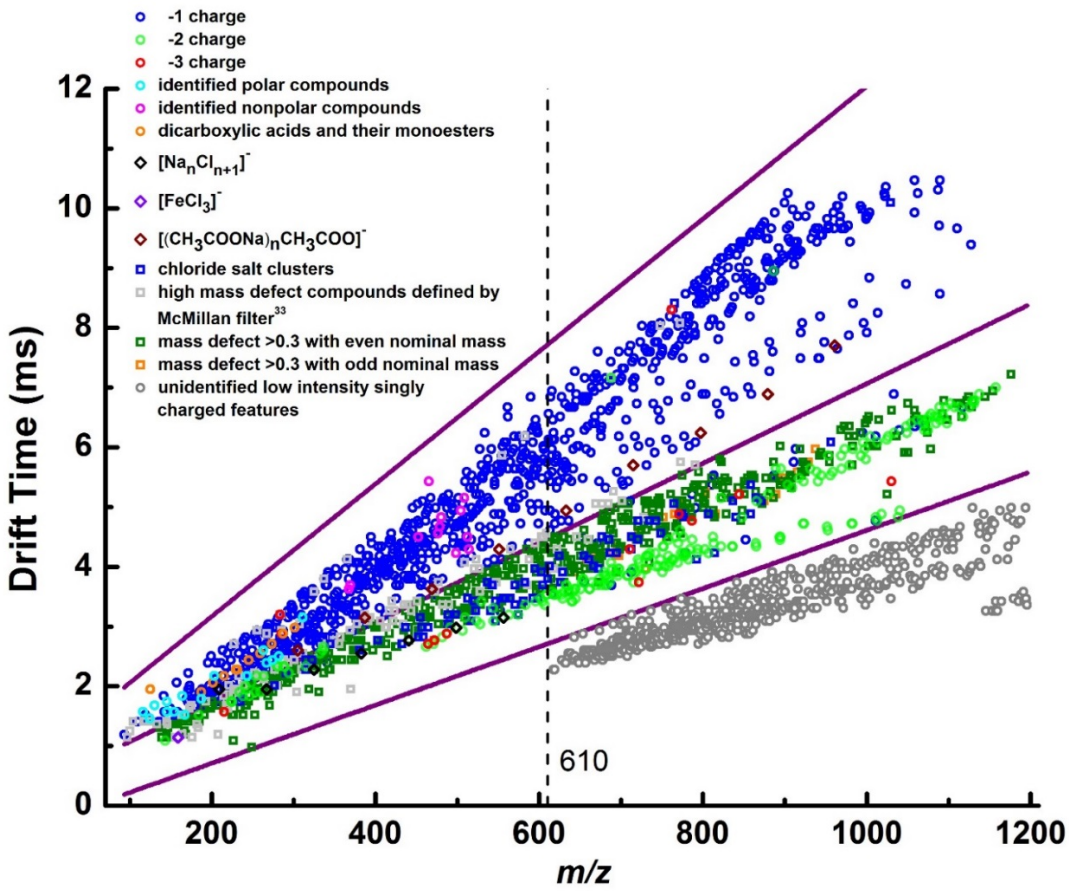

Figure 4.5: Mobility–mass plot for all ionic species detected in the FI-TWIM-MS dataset.

The subset of features detected in the sample preparation blanks included 9 different sodium acetate clusters, 229 chloride salt cluster ions, and 297 high mass defect features, of which 217 were cluster ions (Figure 4.6). These comprised 12.2% of the total detected features shown in Figure 4.5. Following blank filtering, a total of 1878 features were retained in the dataset (Figure 4.7). A number of salt clusters still remained even after filtering for signals present in the blank, likely originating from the sample matrix[28]. Most of these salt clusters were removed by mass defect filtering due to their high mass defect *m/z* values.[28] Features retained following the application of all remaining filters are displayed in Figure 4.8, with all endogenous metabolites with identities confirmed by MS/MS and/or chemical standards shown in Figure 4.9. Less stringent filtering approaches could also be applied to the dataset, with the caveat that identification of the involved compounds may be difficult or even impossible, therefore limiting the biological interpretation of the role of the metabolites involved in PCa pathogenesis.

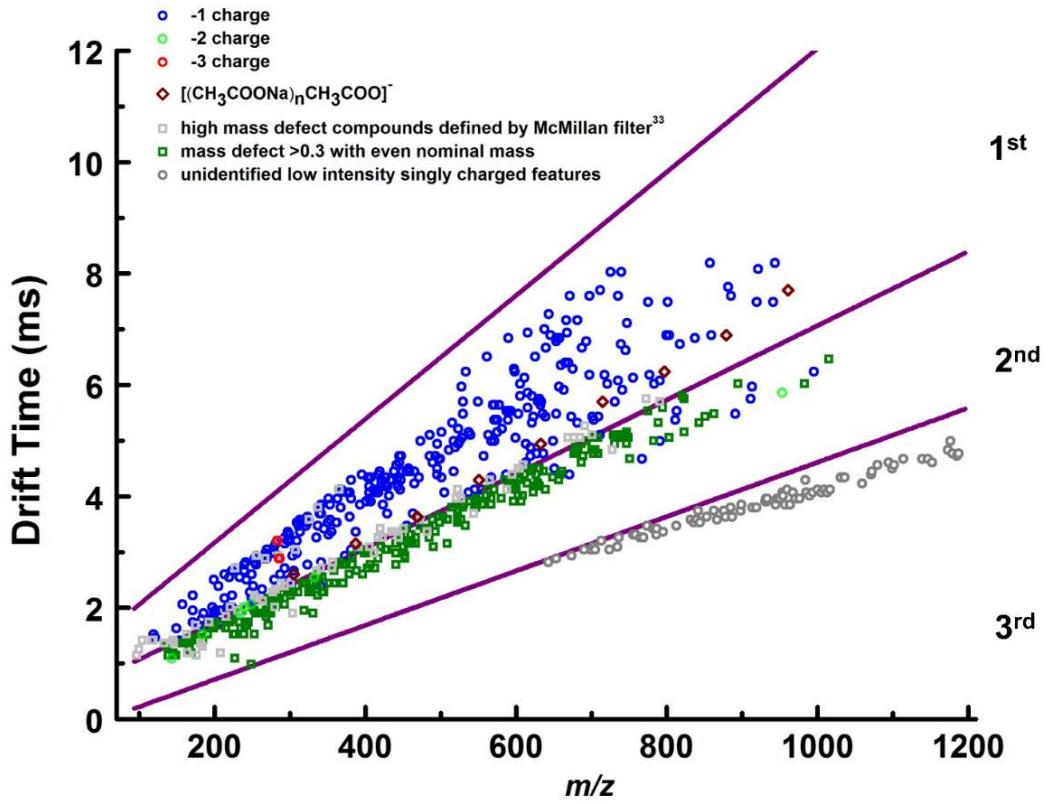

Figure 4.6: Mobility–mass plot of features detected in sample preparation blanks.

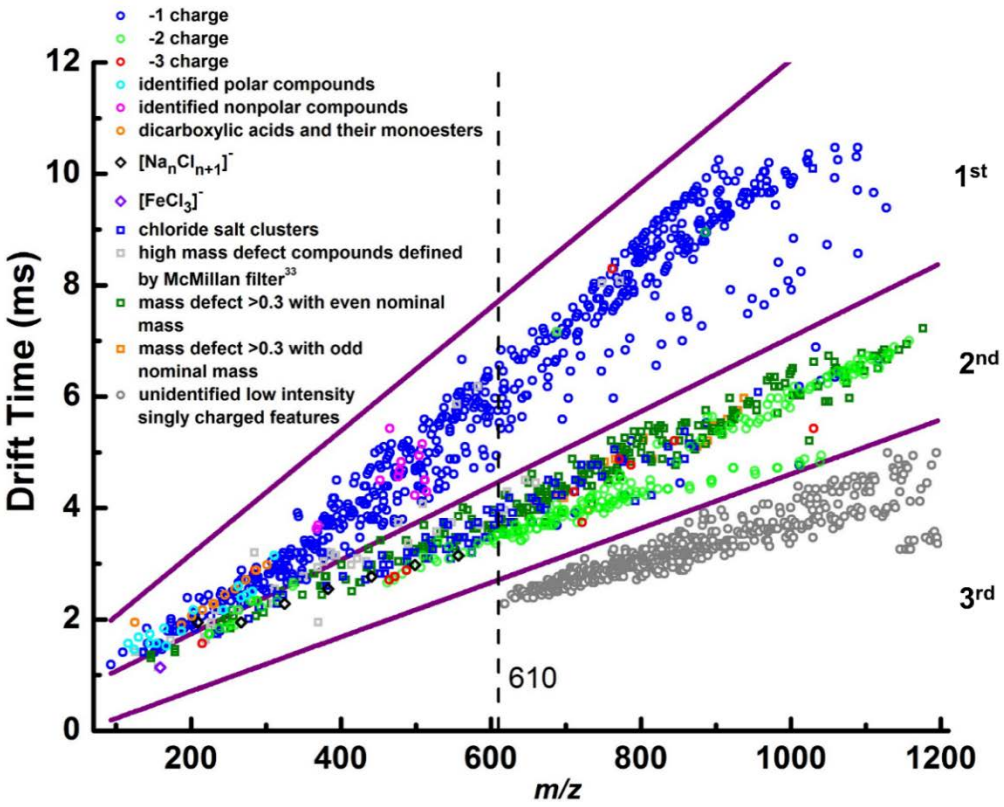

Figure 4.7: Mobility–mass plot of ionic species after blank filtering.

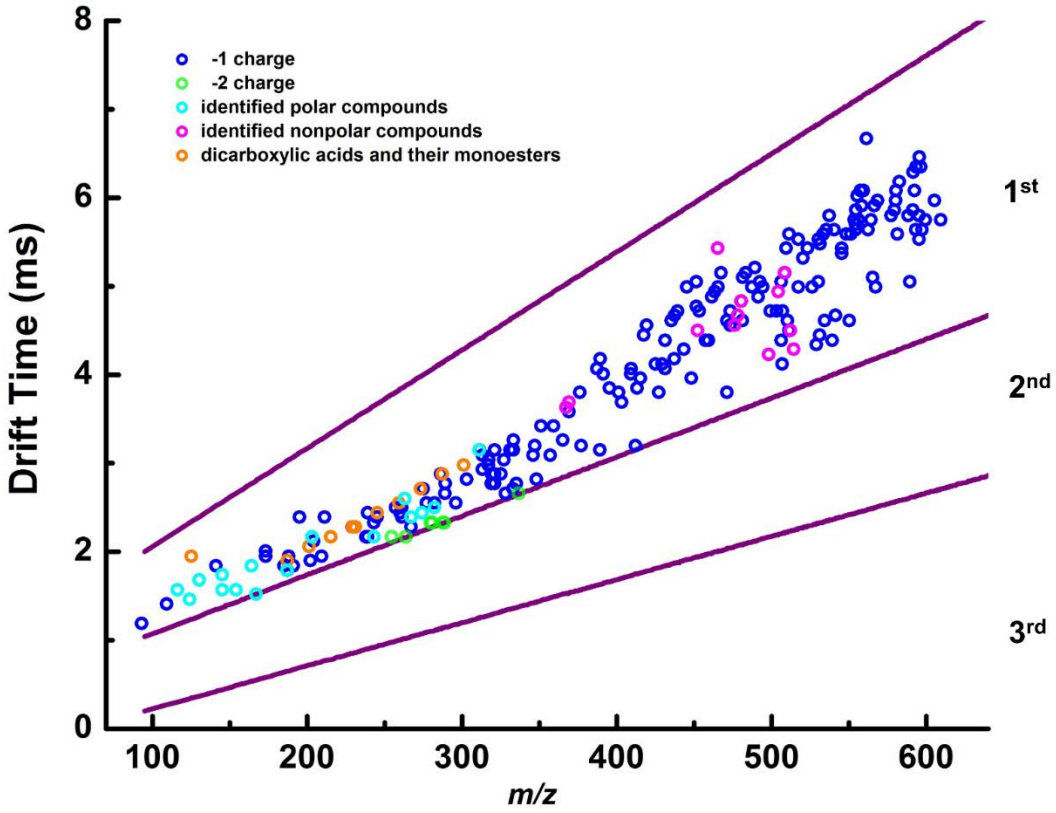

Figure 4.8: Mobility–mass plot after all filters were applied.

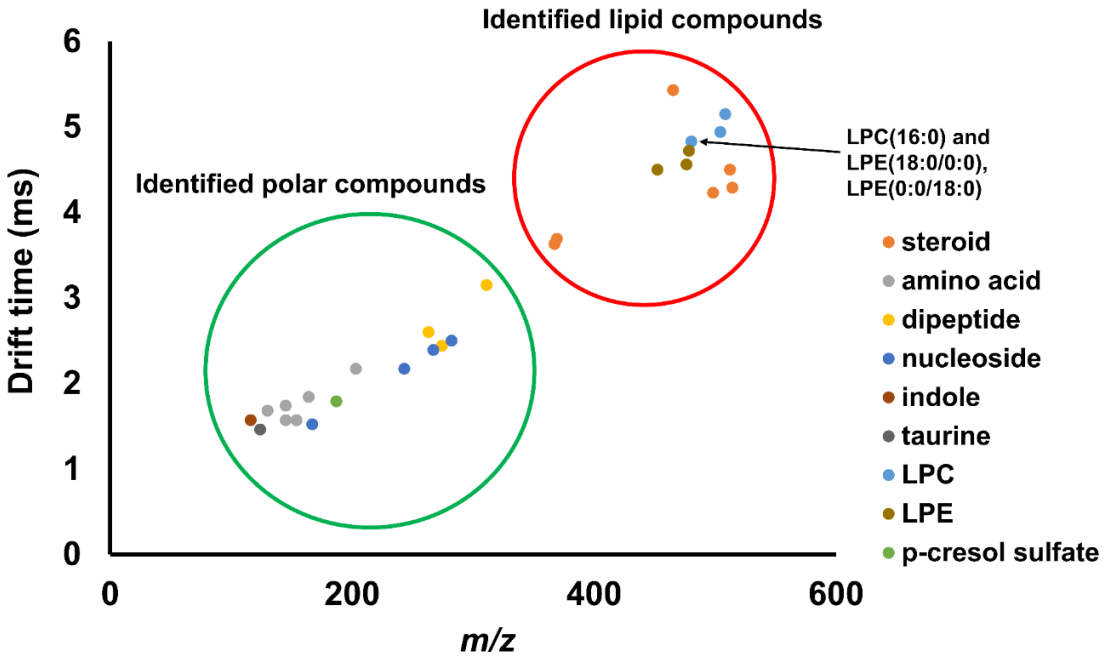

Figure 4.9: Mobility–mass plot of identified polar and lipid compounds grouped by classes in the final dataset after all filters applied.

4.5.4 Multivariate Analysis

After data processing, the final datasets consisting of 11 ("dataset A") and 28 ("dataset B") features were subjected to supervised multivariate analysis. Two oPLS-DA models were built with these datasets using three-block cross-validation. The performance of these two models is shown in Table 4.6. Model A was built using only the 11 identified compounds that were part of the discriminant panel obtained in the LC-MS PCa metabolomics study described in Chapter 2.[19] This model yielded an overall classification accuracy of 89.3%, sensitivity of 90.2% and specificity of 88.1% (Table 4.6 and Figure 4.10). When PCA was performed on this dataset, acceptable unsupervised sample clustering was achieved according to class membership (Figure 4.11).

Azelaic acid was identified as a key differentiating metabolite for PCa detection in the LC-MS study[19] described in Chapter 2. The omission of an LC separation step in FI-TWIM-MS experiments confounds the origin of azelaic acid signals which can originate both from the free acid itself, and from in-source cleavage of its monoesters (Figure 4.12). Since azelaic acid and its esters might be of exogenous origin[33] and their biological role is still unclear, we tested removing the corresponding signals, as well as signals from other dicarboxylic acids and their corresponding esters (Table 4.7) to see if a discriminant metabolite panel could be obtained without their contribution.

Table 4.6: oPLS-DA model classification performance.

Model/ dataset	No. of compounds in initial set	No. of compounds in oPLS-DA model	Compound codes	Model accuracy	Model sensitivity	Model specificity	R^2	Q^2	AUC	Cross-validated Wilcoxon signed rank test *p* value for permutation test
A	11 (matching identified compounds in previous study[19])	11	1,2,4,6,7,9, 11-15	89.3	90.2	88.1	0.68	0.60	0.96	0.0
B	28 (after deleting compounds in Table 4.7)	10 (iPLSDA-selected)	1-10	88.3	88.5	88.1	0.56	0.51	0.93	0.0

A second oPLS-DA model ("model B") was generated using dataset B. Out of the initial set of 28 metabolites, 10 were selected by iPLS-DA variable selection as being the optimum, yielding a discriminant metabolite panel with maximum classification accuracy of 88.3%, sensitivity of 88.5%, and specificity of 88.1% (Table 4.6 and Figure 4.10B). Both models yielded Q^2 values (model predictive ability) larger than 0.5, and an area under the receiver operating characteristic curve (AUC) larger than 0.9. In addition, permutation test results suggested low probability of data overfitting, demonstrating the robustness of the oPLS-DA models and the reliability of both discriminant compound panels.

Table 4.4 details the tentative identifications of the discriminant compounds involved in each of the oPLS-DA models, together with the methods used for validation of their identities. All compounds were identified within 2.5 mDa error for precursor ions, and 8 mDa for fragment ions in their MS/MS spectra. Experimental CCS values were matched within ~ ±3% of database and/or chemical standard values analyzed under

identical conditions. Table 4.5 details the identities of the compounds not selected by iPLS-DA to create model B.

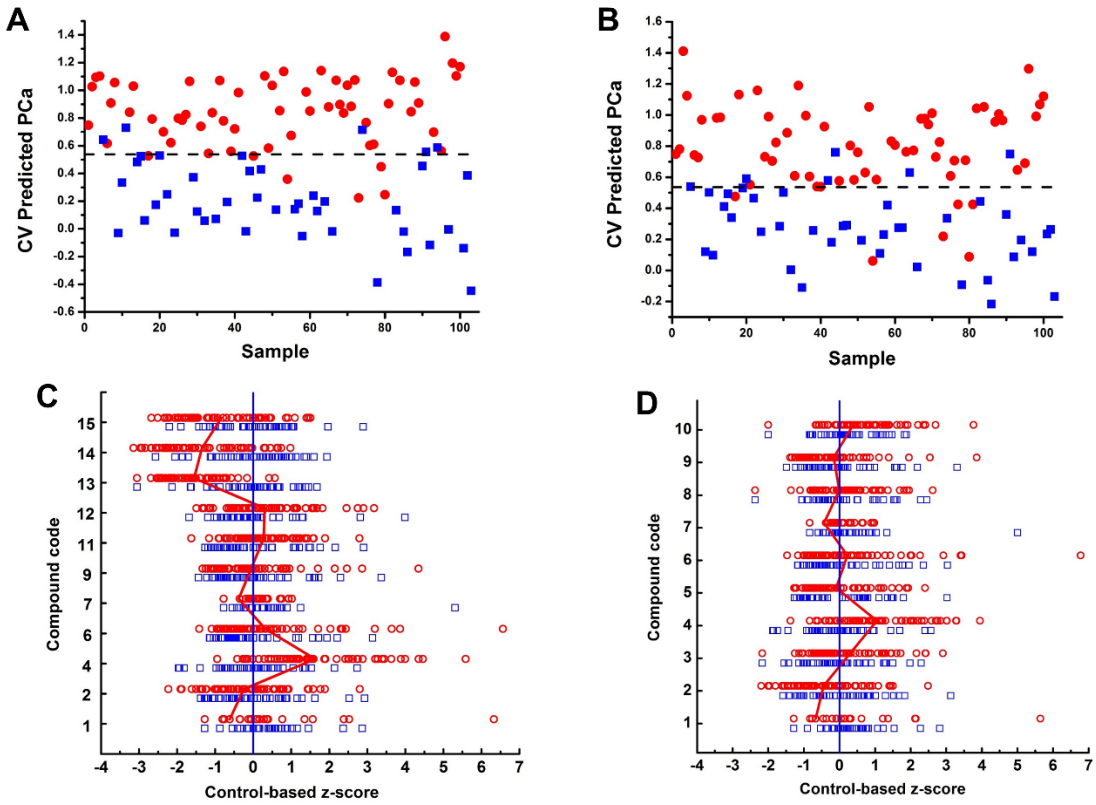

Figure 4.10: (A, B) oPLS-DA three-block cross-validated classification plots for models A and B, respectively. The *x*-axis represents sample number, and the *y*-axis represents the cross-validated predicted scores by the oPLS-DA classification model. PCa and control samples are represented by filled red circles and blue squares, respectively. The threshold for sample classification is represented by a black dashed line. (C, D) Control-based z-score plot of the 11 compounds in panel A and 10 discriminant compounds in panel B, respectively. PCa and control samples are represented by open red and blue squares, respectively. The z-scores are calculated as $(x-\mu)/\sigma$, where x is the normalized peak abundance of the compound in each sample, μ is the mean normalized peak abundance of the compound in the control samples, and σ is the standard deviation of the normalized peak abundance of the compound in control samples. The red and blue line at z-score =0 connect the average z-scores of each compound in PCa and control samples, respectively.

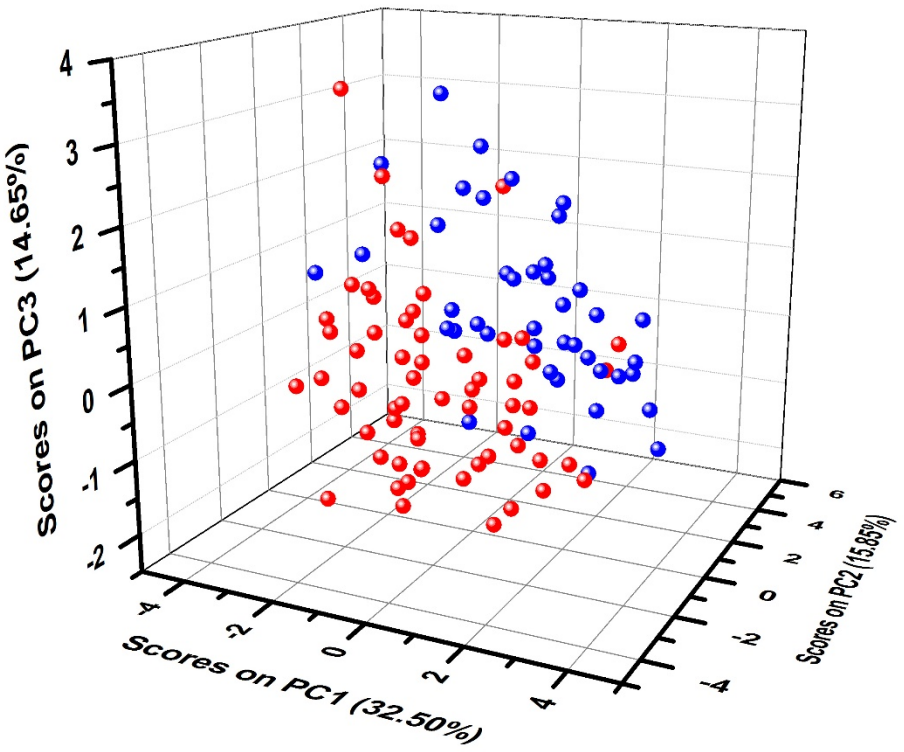

Figure 4.11: PCA plot of all samples using normalized abundances of compounds in panel A. The *x*-, *y*- and *z*- axes represent scores on PC1, PC2 and PC3, respectively. PCa and control samples are represented by red and blue dots, respectively.

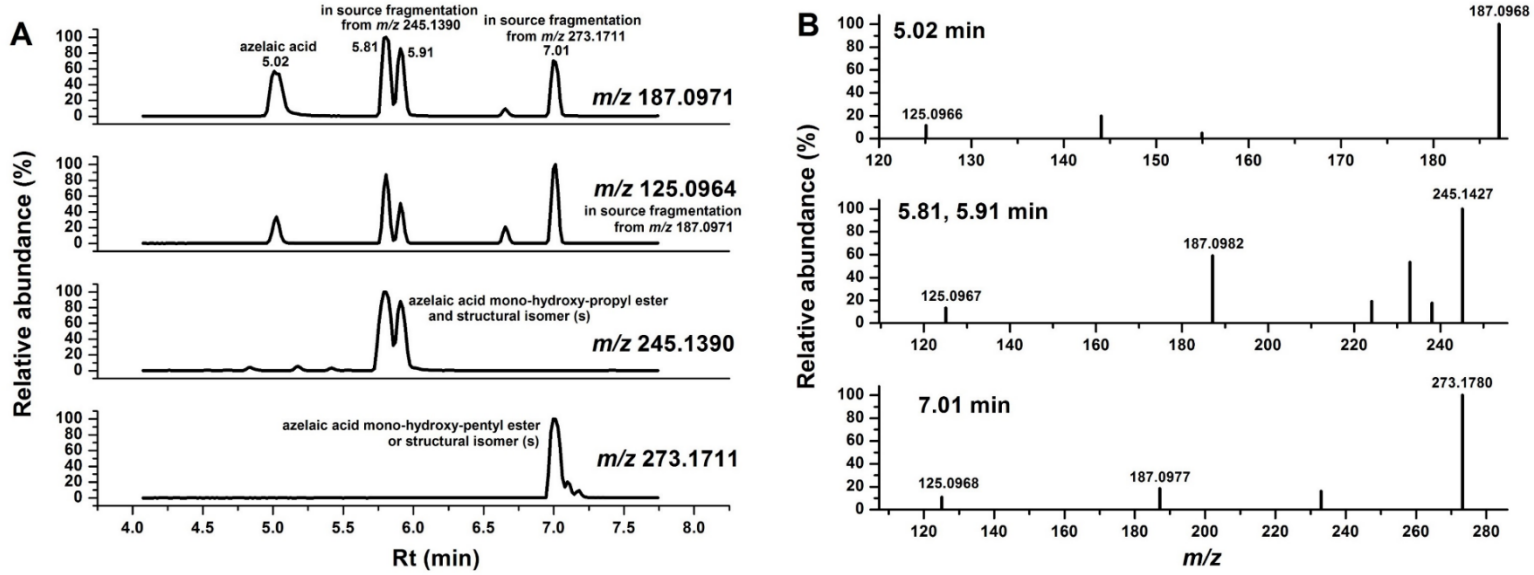

Figure 4.12: (A) Extracted ion chromatograms for azelaic acid $[M-H]^-$ and its monoesters acquired by LC-TWIM-MS. (B) Mass spectra extracted from (A) at various retention times.

Table 4.7: Dicarboxylic acids and corresponding monoesters identified by negative mode FI-TWIM-MS/MS and removed from the dataset.

dicarboxylic acids	elemental formula	$[M-H]^-$	mono-hydroxy-propyl ester $[M-H+C_3H_6O]^-$	mono-hydroxy-pentyl ester $[M-H+C_5H_{10}O]^-$
suberic acid	$C_8H_{14}O_4$		231.1229	259.1545
azelaic acid	$C_9H_{16}O_4$	187.0973	245.1390	273.1711
sebacic acid	$C_{10}H_{18}O_4$	201.1123	259.1545	287.1863
undecanedioic acid	$C_{11}H_{20}O_4$	215.1279	273.1711	301.2016
dodecanedioic acid	$C_{12}H_{22}O_4$	229.1425	287.1863	
1,11-undecanedicarboxylic acid	$C_{13}H_{24}O_4$		301.2016	

Univariate analysis was also performed for the discriminant compounds in panels A and B. Compounds that presented statistically significant fold changes with Bonferroni correction between PCa and control samples are indicated in Table 4.4. The z-score plots for discriminant compounds from panels A and B normalized to the mean abundances of the control samples are shown in Figures 4.10C and 4.10D, respectively. The discriminant compounds with significant changes between PCa and control sample classes were univocally identified as uric acid (significant changes in both panels), phenylalanyl phenylalanine (significant changes in both panels), phenylacetyl glutamine (panel B), azelaic acid and its corresponding monoesters (panel A), and sebacic acid monoester (panel A). The fact that all compounds that were observed as being significantly altered in this study coincided with the discriminant metabolite panel described in Chapter 2, with identical fold change direction between PCa and controls, corroborated the feasibility and accuracy of the FI-TWIM-MS method proposed here.

4.5.5 Biological Roles of Discriminant Metabolites

Uric acid, with a significant positive fold change between PCa and control serum samples, has pro-inflammatory properties, and elevated serum uric acid (hyperuricemia) has been reported to be associated with increased cancer (including PCa) risk, recurrence and mortality.[34, 35] Tryptophan levels were significantly decreased in PCa patient samples compared to controls in panel B. Consumption of tryptophan has been revealed to be a crucial factor in cancer progression;[36] inhibition of tryptophan 2,3-dioxygenase which degrades tryptophan in the kynurenine pathway has been reported to reverse tumor immune resistance in mice.[37] Interestingly, indole, which is a bacterial degradation product of tryptophan, was also identified among the discriminant compounds in panel B with an average abundance elevation in serum of PCa patients compared to that of controls. LPC(18:0) has been reported to be one of the discriminant plasma lipids for PCa, with a significant increase in PCa patients compared to controls,[23] in agreement with our findings. Azelaic acid, reported to be a potential antitumoral agent,[38] had a significant increase in controls compared to PCa patients in panel A, with the same trend observed for its corresponding monoesters. However, the origin of azelaic acid and related monoesters warrants further investigation, as these species have also been reported as originating from corn oil.[33]

Uniquely identified metabolites in dataset #2 were input into Metaboanalyst[39] for pathway analysis, with several metabolic pathways with more than 1 metabolite hit indicated as significantly altered between PCa patients and controls (p <0.05). These included purine metabolism, phenylalanine metabolism, aminoacyl-tRNA metabolism, and tryptophan metabolism (Figure 4.13). Elevated purine nucleotide levels have been

reported to be sufficient to induce major histocompatibility class I chain-related protein A (MICA) expression on abnormal or stressed cells including cancer cells.[40] *De novo* purine biosynthesis has been suggested to support elevated transcription and cell division levels in PCa cells and may provide as a target for PCa treatment.[41] Phenylalanine and tyrosine restriction has been found to induce PCa cell death *via* glucose metabolism modulation.[42] In addition, phenylalanine has also been reported to be significantly increased in PCa metastatic bone tissue compared to normal bone in a non-targeted metabolomics study using GC-MS.[22] Aminoacyl-tRNA synthetases (AARSs) function as enzymes to catalyze the covalent linkage of amino acids to their corresponding tRNAs, and they play a crucial role in translation and cell signaling that are vital for cell function and viability.[43, 44] AARSs have been suggested as potential therapeutic targets for cancer due to their cancer-related genetic profiles, mutations, and biological pathway deregulations.[43, 44] Metabolites involved in tryptophan metabolic pathways have been found to be significantly changed in urine samples of PCa patients compared to healthy controls.[45]

4.6 Limitations of the Proposed Approach

Despite the advantages of the FI-TWIM-MS approach for rapid metabolomics fingerprinting, this technique is not without limitations. Table 4.8 summarizes the strengths and limitations of FI-MS, FI-IM-MS, LC-MS and LC-IM-MS according to various performance parameters. Clearly, a compromise between sample throughput and peak capacity is achieved in FI-IM-MS, with the lack of front-end LC separation being beneficial in terms of speed, but detrimental in terms of differentiating intact ionic species from molecular compounds resulting from in-source fragmentation. Nevertheless,

the power of FI-TWIM-MS resides in its speed and economy, rather than in its comprehensive separation power, as already extensively discussed.

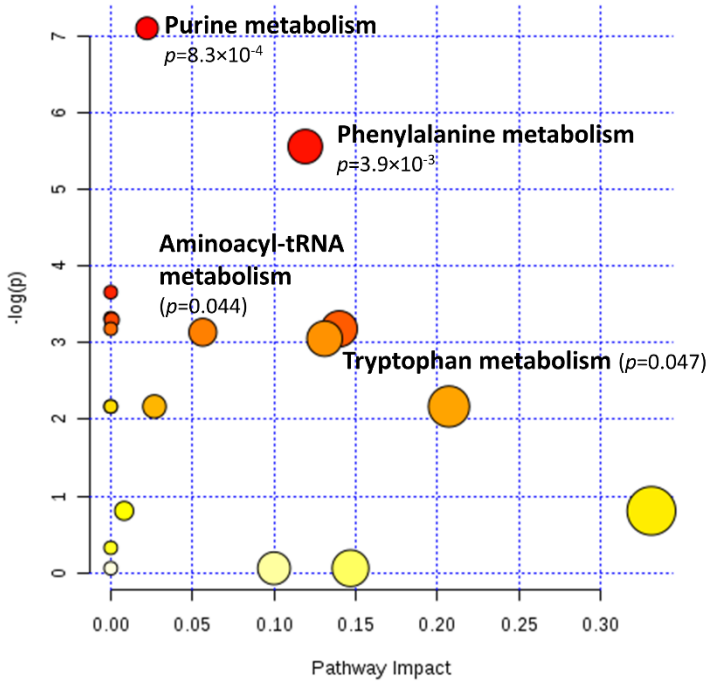

Figure 4.13: Metaboanalyst pathway analysis of uniquely identified metabolites in dataset #2. Each circle on the map represents a pathway, and the size and color of the circle are based on the *p* value, indicating the significance of the changes in the matched metabolites in the pathway, and the pathway impact score, which is correlated with the centrality of the metabolites involved.[39, 46]

Table 4.8: Comparison of four popular MS-based techniques implemented in non-targeted metabolomics applications.

properties	**FI-MS**	**FI-IM-MS**	**LC-MS**	**LC-IM-MS**
novelty	relatively new	largely unexplored	traditional	relatively new
analysis time per sample	<5 min	<5 min	~20-30 min	~20-30 min
maximum number of samples analyzed per day	~200	~200	~30	~30
instrument setup	simple	simple	normal	complex
CCS information	no	yes	no	yes
data dimension	2	3	3	4
ion suppression	yes	yes	reduced	reduced
signal to noise	OK	good	good	excellent
metabolite annotation confidence	OK	good	good	excellent
separation of different classes of compounds	no	yes	yes	yes
separation of compounds with different charge states	no	yes	no	yes
separation of structural isomers	poor	medium	high	highest
discriminate peaks of intact compounds from those resulting from in-source fragmentation of other compounds	no	no	yes	yes

4.7 Conclusion

In this study, a fast FI-TWIM-MS serum metabolic profiling method was applied for PCa detection, with an analysis speed of 3 min per sample, followed by a 3 min wash run. These times can likely be shortened by increasing the mobile phase flow rate or replacing the current front-end FI system with a faster injection platform such as the RapidFire platform.[47] PCa patient and control samples were distinguished with 88.3-89.3% accuracies, 88.5-90.2% sensitivities and 88.1% specificity by using oPLS-DA classification. Discriminant metabolites were identified by matching accurate mass, CCS, and fragmentation patterns in FI-TWIM-MS/MS to those in databases and/or to authentic chemical standards. CCS calibration was utilized to account for ion drift time shifts in TWIM-MS/MS experiments compared to those in TWIM-MS, greatly aiding in assigning the correct precursor ions. Stringent criteria were utilized for combining spectral features, grouping adduct ions, in-source fragments, and salt cluster ions. Overall, results indicated that FI-TWIM-MS is a promising analytical tool that could be successfully applied to metabolic fingerprinting of large scale cohorts, with its fast analysis speed and ion separation capabilities being useful for interrogation of complex biological mixtures. Further improvements in sample throughput and development of an automated metabolite identification pipeline will further increase the efficiency of the proposed metabolomics workflow.

4.8 References

1. Dunn, W. B.; Broadhurst, D.; Begley, P.; Zelena, E.; Francis-McIntyre, S.; Anderson, N.; Brown, M.; Knowles, J. D.; Halsall, A.; Haselden, J. N.; Nicholls, A. W.; Wilson, I. D.; Kell, D. B.; Goodacre, R.; C, H. S. M. H., Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nature Protocols* **2011,** 6, (7), 1060-1083.

2. Fu, Y. Q.; Zhao, C. X.; Lu, X.; Xu, G. W., Nontargeted screening of chemical contaminants and illegal additives in food based on liquid chromatography-high resolution mass spectrometry. *Trac-Trends in Analytical Chemistry* **2017,** 96, 89-98.

3. Forcisi, S.; Moritz, F.; Kanawati, B.; Tziotis, D.; Lehmann, R.; Schmitt-Kopplin, P., Liquid chromatography-mass spectrometry in metabolomics research: mass analyzers in ultra high pressure liquid chromatography coupling. *J Chromatogr A* **2013,** 1292, 51-65.

4. Blaise, B. J.; Correia, G.; Tin, A.; Young, J. H.; Vergnaud, A. C.; Lewis, M.; Pearce, J. T.; Elliott, P.; Nicholson, J. K.; Holmes, E.; Ebbels, T. M., Power Analysis and Sample Size Determination in Metabolic Phenotyping. *Anal Chem* **2016,** 88, (10), 5179-88.

5. Krzywinski, M.; Altman, N., POINTS OF SIGNIFICANCE Power and sample size. *Nature Methods* **2013,** 10, (12), 1139-1140.

6. Beckmann, M.; Parker, D.; Enot, D. P.; Duval, E.; Draper, J., High-throughput, nontargeted metabolite fingerprinting using nominal mass flow injection electrospray mass spectrometry. *Nature Protocols* **2008,** 3, (3), 486-504.

7. Gonzalez-Dominguez, R.; Sayago, A.; Fernandez-Recamales, A., Direct infusion mass spectrometry for metabolomic phenotyping of diseases. *Bioanalysis* **2017,** 9, (1), 131-148.

8. Paglia, G.; Williams, J. P.; Menikarachchi, L.; Thompson, J. W.; Tyldesley-Worster, R.; Halldorsson, S.; Rolfsson, O.; Moseley, A.; Grant, D.; Langridge, J.; Palsson, B. O.; Astarita, G., Ion Mobility Derived Collision Cross Sections to Support Metabolomics Applications. *Analytical Chemistry* **2014,** 86, (8), 3985-3993.

9. Paglia, G.; Angel, P.; Williams, J. P.; Richardson, K.; Olivos, H. J.; Thompson, J. W.; Menikarachchi, L.; Lai, S.; Walsh, C.; Moseley, A.; Plumb, R. S.; Grant, D. F.; Palsson, B. O.; Langridge, J.; Geromanos, S.; Astarite, G., Ion Mobility-Derived Collision Cross Section As an Additional Measure for Lipid Fingerprinting and Identification. *Analytical Chemistry* **2015,** 87, (2), 1137-1144.

10. Soper-Hopper, M. T.; Petrov, A. S.; Howard, J. N.; Yu, S. S.; Forsythe, J. G.; Grover, M. A.; Fernandez, F. M., Collision cross section predictions using 2-dimensional molecular descriptors. *Chem Commun (Camb)* **2017,** 53, (54), 7624-7627.

11. Kurulugama, R. T.; Valentine, S. J.; Sowell, R. A.; Clemmer, D. E., Development of a high-throughput IMS-IMS-MS approach for analyzing mixtures of biomolecules. *Journal of Proteomics* **2008,** 71, (3), 318-331.

12. Paglia, G.; Astarita, G., Metabolomics and lipidomics using traveling-wave ion mobility mass spectrometry. *Nature Protocols* **2017,** 12, (4), 797-813.

13. Dwivedi, P.; Wu, P.; Klopsch, S. J.; Puzon, G. J.; Xun, L.; Hill, H. H., Metabolic profiling by ion mobility mass spectrometry (IMMS). *Metabolomics* **2008,** 4, (1), 63-80.

14. Kaplan, K.; Dwivedi, P.; Davidson, S.; Yang, Q.; Tso, P.; Siems, W.; Hill, H. H., Monitoring Dynamic Changes in Lymph Metabolome of Fasting and Fed Rats by Electrospray Ionization-Ion Mobility Mass Spectrometry (ESI-IMMS). *Analytical Chemistry* **2009,** 81, (19), 7944-7953.

15. Isailovic, D.; Plasencia, M. D.; Gaye, M. M.; Stokes, S. T.; Kurulugama, R. T.; Pungpapong, V.; Zhang, M.; Kyselova, Z.; Goldman, R.; Mechref, Y.; Novotny, M. V.; Clemmer, D. E., Delineating Diseases by IMS-MS Profiling of Serum N-linked Glycans. *Journal of Proteome Research* **2012,** 11, (2), 576-585.

16. Lima, A. R.; Bastos Mde, L.; Carvalho, M.; Guedes de Pinho, P., Biomarker Discovery in Human Prostate Cancer: an Update in Metabolomics Studies. *Transl Oncol* **2016,** 9, (4), 357-70.

17. Miyagi, Y.; Higashiyama, M.; Gochi, A.; Akaike, M.; Ishikawa, T.; Miura, T.; Saruki, N.; Bando, E.; Kimura, H.; Imamura, F.; Moriyama, M.; Ikeda, I.; Chiba, A.; Oshita, F.; Imaizumi, A.; Yamamoto, H.; Miyano, H.; Horimoto, K.; Tochikubo, O.; Mitsushima, T.; Yamakado, M.; Okamoto, N., Plasma Free Amino Acid Profiling of Five Types of Cancer Patients and Its Application for Early Detection. *Plos One* **2011,** 6, (9).

18. Struck-Lewicka, W.; Kordalewska, M.; Bujak, R.; Mpanga, A. Y.; Markuszewski, M.; Jacyna, J.; Matuszewski, M.; Kaliszan, R.; Markuszewski, M. J., Urine metabolic fingerprinting using LC-MS and GC-MS reveals metabolite changes in prostate cancer: A pilot study. *Journal of Pharmaceutical and Biomedical Analysis* **2015,** 111, 351-361.

19. Zang, X. L.; Jones, C. M.; Long, T. Q.; Monge, M. E.; Zhou, M. S.; Walker, L. D.; Mezencev, R.; Gray, A.; McDonald, J. F.; Fernandez, F. M., Feasibility of Detecting Prostate Cancer by Ultraperformance Liquid Chromatography-Mass Spectrometry Serum Metabolomics. *Journal of Proteome Research* **2014,** 13, (7), 3444-3454.

20. Swanson, M. G.; Vigneron, D. B.; Tabatabai, Z. L.; Males, R. G.; Schmitt, L.; Carroll, P. R.; James, J. K.; Hurd, R. E.; Kurhanewicz, J., Proton HR-MAS spectroscopy and quantitative pathologic analysis of MRI/3D-MRSI-targeted postsurgical prostate tissues. *Magnetic Resonance in Medicine* **2003,** 50, (5), 944-954.

21. Swanson, M. G.; Zektzer, A. S.; Tabatabai, Z. L.; Simko, J.; Jarso, S.; Keshari, K. R.; Schmitt, L.; Carroll, P. R.; Shinohara, K.; Vigneron, D. B.; Kurhanewicz, J., Quantitative analysis of prostate metabolites using H-1 HR-MAS spectroscopy. *Magnetic Resonance in Medicine* **2006,** 55, (6), 1257-1264.

22. Thysell, E.; Surowiec, I.; Hornberg, E.; Crnalic, S.; Widmark, A.; Johansson, A. I.; Stattin, P.; Bergh, A.; Moritz, T.; Antti, H.; Wikstro, P., Metabolomic Characterization of Human Prostate Cancer Bone Metastases Reveals Increased Levels of Cholesterol. *Plos One* **2010,** 5, (12).

23. Zhou, X. C.; Mao, J. H.; Ai, J. M.; Deng, Y. P.; Roth, M. R.; Pound, C.; Henegar, J.; Welti, R.; Bigler, S. A., Identification of Plasma Lipid Biomarkers for Prostate Cancer by Lipidomics and Bioinformatics. *Plos One* **2012,** 7, (11), e48889.

24. Wu, H.; Liu, T. T.; Ma, C. G.; Xue, R. Y.; Deng, C. H.; Zeng, H. Z.; Shen, X. Z., GC/MS-based metabolomic approach to validate the role of urinary sarcosine and target biomarkers for human prostate cancer by microwave-assisted derivatization. *Analytical and Bioanalytical Chemistry* **2011,** 401, (2), 635-646.

25. Kelly, R. S.; Vander Heiden, M. G.; Giovannucci, E.; Mucci, L. A., Metabolomic Biomarkers of Prostate Cancer: Prediction, Diagnosis, Progression, Prognosis, and Recurrence. *Cancer Epidemiol Biomarkers Prev* **2016,** 25, (6), 887-906.

26. Bush, M. F.; Campuzano, I. D. G.; Robinson, C. V., Ion Mobility Mass Spectrometry of Peptide Ions: Effects of Drift Gas and Calibration Strategies. *Analytical Chemistry* **2012,** 84, (16), 7124-7130.

27. DeHaven, C. D.; Evans, A. M.; Dai, H. P.; Lawton, K. A., Organization of GC/MS and LC/MS metabolomics data into chemical libraries. *Journal of Cheminformatics* **2010,** 2.

28. McMillan, A.; Renaud, J. B.; Gloor, G. B.; Reid, G.; Sumarah, M. W., Post-acquisition filtering of salt cluster artefacts for LC-MS based human metabolomic studies. *Journal of Cheminformatics* **2016,** 8.

29. 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., HMDB 3.0--The Human Metabolome Database in 2013. *Nucleic Acids Res* **2013,** 41, (Database issue), D801-7.

30. Boccard, J.; Rutledge, D. N., A consensus orthogonal partial least squares discriminant analysis (OPLS-DA) strategy for multiblock Omics data fusion. *Anal Chim Acta* **2013,** 769, 30-9.

31. Worley, B.; Powers, R., Multivariate Analysis in Metabolomics. *Curr Metabolomics* **2013,** 1, (1), 92-107.

32. Smith, C. A.; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G., METLIN: a metabolite mass spectral database. *Ther Drug Monit* **2005,** 27, (6), 747-51.

33. Matsubara, T.; Tanaka, N.; Krausz, K. W.; Manna, S. K.; Kang, D. W.; Anderson, E. R.; Luecke, H.; Patterson, A. D.; Shah, Y. M.; Gonzalez, F. J., Metabolomics Identifies an Inflammatory Cascade Involved in Dioxin- and Diet-Induced Steatohepatitis. *Cell Metabolism* **2012,** 16, (5), 634-644.

34. Fini, M. A.; Elias, A.; Johnson, R. J.; Wright, R. M., Contribution of uric acid to cancer risk, recurrence, and mortality. *Clin Transl Med* **2012,** 1, (1), 16.

35. Kolonel, L. N.; Yoshizawa, C.; Nomura, A. M.; Stemmermann, G. N., Relationship of serum uric acid to cancer occurrence in a prospective male cohort. *Cancer Epidemiol Biomarkers Prev* **1994,** 3, (3), 225-8.

36. Prendergast, G. C., CANCER Why tumours eat tryptophan. *Nature* **2011,** 478, (7368), 192-194.

37. Pilotte, L.; Larrieu, P.; Stroobant, V.; Colau, D.; Dolusic, E.; Frederick, R.; De Plaen, E.; Uyttenhove, C.; Wouters, J.; Masereel, B.; Van den Eynde, B. J., Reversal of tumoral immune resistance by inhibition of tryptophan 2,3-dioxygenase. *Proceedings of the National Academy of Sciences of the United States of America* **2012,** 109, (7), 2497-2502.

38. Breathnach, A. S., Azelaic acid: potential as a general antitumoural agent. *Medical Hypotheses* **1999,** 52, (3), 221-226.

39. Xia, J. G.; Sinelnikov, I. V.; Han, B.; Wishart, D. S., MetaboAnalyst 3.0-making metabolomics more meaningful. *Nucleic Acids Research* **2015,** 43, (W1), W251-W257.

40. McCarthy, M. T.; Moncayo, G.; Hiron, T. K.; Jakobsen, N. A.; Valli, A.; Soga, T.; Adam, J.; O'Callaghan, C. A., Purine nucleotide metabolism regulates expression of the human immune ligand MICA. *Journal of Biological Chemistry* **2018,** 293, (11), 3913-3924.

41. Barfeld, S. J.; Fazli, L.; Persson, M.; Marjavaara, L.; Urbanucci, A.; Kaukoniemi, K. M.; Rennie, P. S.; Ceder, Y.; Chabes, A.; Visakorpi, T.; Mills, I. G., Myc-dependent purine biosynthesis affects nucleolar stress and therapy response in prostate cancer. *Oncotarget* **2015,** 6, (14), 12587-12602.

42. Fu, Y. M.; Lin, H. M.; Fang, W. G.; Meadows, G. G., Cell Death of Prostate Cancer Cells by Specific Amino Acid Restriction Depends on Alterations of Glucose Metabolism. *Journal of Cellular Physiology* **2010,** 224, (2), 491-500.

43. Kim, S.; You, S.; Hwang, D., Aminoacyl-tRNA synthetases and tumorigenesis: more than housekeeping. *Nature Reviews Cancer* **2011,** 11, (10), 708-718.

44. Rajendran, V.; Kalita, P.; Shukla, H.; Kumar, A.; Tripathi, T., Aminoacyl-tRNA synthetases: Structure, function, and drug discovery. *International Journal of Biological Macromolecules* **2018,** 111, 400-414.

45. Fernandez-Peralbo, M. A.; Gomez-Gomez, E.; Calderon-Santiago, M.; Carrasco-Valiente, J.; Ruiz-Garcia, J.; Requena-Tapia, M. J.; de Castro, M. D. L.; Priego-Capote, F., Prostate Cancer Patients-Negative Biopsy Controls Discrimination by Untargeted Metabolomics Analysis of Urine by LC-QTOF: Upstream Information on Other Omics. *Scientific Reports* **2016,** 6.

46. Li, J.; Greenwood, P. L.; Cockett, N. E.; Hadfield, T. S.; Vuocolo, T.; Byrne, K.; White, J. D.; Tellam, R. L.; Schirra, H. J., Impacts of the Callipyge mutation on ovine plasma metabolites and muscle fibre type. *PLoS One* **2014,** 9, (6), e99726.

47. Zhang, X.; Romm, M.; Zheng, X.; Zink, E. M.; Kim, Y. M.; Burnum-Johnson, K. E.; Orton, D. J.; Apffel, A.; Ibrahim, Y. M.; Monroe, M. E.; Moore, R. J.; Smith, J. N.; Ma, J.; Renslow, R. S.; Thomas, D. G.; Blackwell, A. E.; Swinford, G.; Sausen, J.; Kurulugama, R. T.; Eno, N.; Darland, E.; Stafford, G.; Fjeldsted, J.; Metz, T. O.; Teeguarden, J. G.; Smith, R. D.; Baker, E. S., SPE-IMS-MS: An automated platform for sub-sixty second surveillance of endogenous metabolites and xenobiotics in biofluids. *Clin Mass Spectrom* **2016,** 2, 1-10.

CHAPTER 5. EXHALED BREATH CONDENSATE METABOLIC FINGERPRINTING FOR CYSTIC FIBROSIS STUDIES BY TRAVELING WAVE ION MOBILITY-MASS SPECTROMETRY

Adapted with permission from
Zang X†, Pérez JJ†, Jones CM, Monge ME, McCarty NA, Stecenko AA, Fernández FM. Comparison of Ambient and Atmospheric Pressure Ion Sources for Cystic Fibrosis Exhaled Breath Condensate Ion Mobility-Mass Spectrometry Metabolomics. *Journal of the American Society for Mass Spectrometry*. **2017**, *28*, 1489-1496.
†equal contributing author

J. J. Pérez, C. M. Jones and M. E. Monge optimized the DI-TWIM-MS and TM-DART-TWIM-MS methods for EBC analysis. J. J. Pérez conducted sample analysis with assistance from M. E. Monge. X. Zang performed data processing and analysis, discriminant metabolite annotation and biological relevance search.

5.1 Abstract

Cystic fibrosis (CF) is an autosomal recessive disorder caused by mutations in the gene that encodes the cystic fibrosis transmembrane conductance regulator (CFTR) protein. The vast majority of the CF mortality results from progressive lung disease. Targeted and non-targeted CF breath metabolomics investigations *via* exhaled breath condensate (EBC) analyses offer the chance to reveal metabolic alterations associated with CF pathology and aid in assessing the effectiveness of CF treatment. Here, transmission-mode direct analysis in real time traveling wave ion mobility time-of-flight mass spectrometry (TM-DART-TWIM-TOF-MS) was tested as a high-throughput alternative to conventional direct infusion (DI) electrospray ionization (ESI) and atmospheric pressure chemical ionization (APCI) methods, and the performances of the

three ionization methods were critically compared. EBC was chosen as the noninvasive surrogate for airway sampling over expectorated sputum as EBC can be collected in all CF subjects regardless of age and lung disease severity. When using pooled EBC collected from a healthy volunteer, ESI detected the most metabolites, APCI a log order less, and TM-DART the least. TM-DART-TWIM-TOF-MS was used to profile metabolites in EBC samples from 4 CF patients and 5 controls, finding that a panel of 3 discriminant EBC metabolites, some of which had been previously detected by other methods, differentiating these two classes with excellent cross-validated accuracy.

5.2 Mass Spectrometry-Based Approaches for Rapid Exhaled Breath Condensate Metabolomics

5.2.1 Exhaled Breath Condensate Metabolomics to Study Respiratory Diseases

As discussed in the Chapter 1, EBC is being increasingly favored as a non-invasive means for investigating pathophysiological processes occurring within the lung[1-3]. For pulmonary diseases such as asthma, chronic obstructive pulmonary disease (COPD), and cystic fibrosis (CF), the chemical composition of EBC is systematically altered.[3, 4] Targeted and non-targeted breath metabolomics approaches are therefore useful for EBC profiling in an effort to identify markers of airway inflammation, characterize pulmonary disease states, and yield a better understanding of disease pathophysiology.[5-12]

5.2.2 Ambient and Atmospheric Pressure Ion Sources for EBC Metabolomics

MS and NMR spectroscopy have been utilized in EBC metabolomics experiments.[10, 13-16] The higher sensitivity offered by MS[13], makes it highly suitable for the detection of epithelial lining fluid (ELF) metabolites that are 10^3 to 10^4-fold diluted in EBC[17]. As discussed in Chapter 4, MS-based platforms often rely on GC or LC separations, requiring run times in the tens of minutes. DI ESI or APCI methods can be alternatively used to speed up analysis,[18] but often suffer from limited peak capacity, the inability to distinguish overlapping compounds in crowded spectra, and ionization suppression. As demonstrated in the previous chapter, coupling DI methods with IMS can reduce spectra complexity, improve signal to noise ratio by eliminating chemical noise, generate cleaner MS/MS data even when precursor ion co-selection occurs, and separate closely-related compounds such as isobaric species on a millisecond timescale. Various IMS techniques are currently available, including drift tube IMS (DTIMS), differential mobility spectrometry (DMS), and traveling wave ion mobility spectrometry (TWIMS), each offering varying degrees of separation power and ion focusing capabilities [19]. In combination with MS, these mobility techniques have found applications in proteomics,[20] glycomics,[21] clinical analysis,[22] and metabolomics[23].

Direct analysis in real time (DART), first reported by Cody *et al.*[24], is an open air, direct sampling, plasma ionization technique capable of fast analysis of solids, liquids, or gases. It has been combined with stand-alone DTIMS[25, 26], and TWIM-MS[27], but not to stand alone DMS or DMS-MS. In DART-MS, a heated gas stream of metastable atomic or molecular species, typically He or N_2, is directed at a sample placed within the ionization region. Thermally-desorbed analytes are then ionized *via* a number of gas-phase mechanisms,[28-30] and subsequently suctioned into the mass spectrometer. Fluid

dynamics play a critical role in DART-MS, determining not only the extent of ion transmission efficiency[31], but also its reproducibility[32]. To stabilize detrimental fluid dynamic effects, the transmission mode (TM)-DART geometry, in which a sample is directly deposited on a stainless steel mesh in a flow-through fashion, has been proposed[33]. This approach has been shown to be superior to probe DART approaches in terms of precision[34].

Building on previous results from our group where two LC-MS methods were compared with DI-TWIM-MS to quantify expected EBC glucose levels in controls and patients with CF,[35] this chapter describes the implementation of TM-DART-TWIM-TOF-MS as a feasible approach for rapid, high-throughput non-targeted EBC metabolomics studies in CF, with the goal of expanding the toolbox available for exposing metabolic alterations associated with CF disease pathology and aiding in assessing CF therapy effectiveness[36].

5.3 Hypothesis

In this study, we hypothesize that APCI, ESI and TM-DART-TWIM-TOF-MS are feasible high-throughput alternatives to conventional LC-MS-based methods for EBC metabolomics with different metabolome coverages. To test that, we compared DART with ESI and APCI approaches in DI mode to evaluate the degree of overlap between metabolic features produced by each ionization technique. Furthermore, we explored the capability of TM-DART-TWIM-TOF-MS combined with multivariate analysis to discriminate between CF and control EBC samples.

5.4 Materials and Methods

5.4.1 Chemicals

Omnisolv LC-MS grade acetonitrile was purchased from EMD (Billerica, MA, USA). Ultrapure water with 18.2 MΩ cm resistivity (Thermo Scientific Barnstead Nanopure UV ultrapure water system, Marietta, OH, USA) was used in all sample preparation protocols.

5.4.2 Sample Collection and Preparation

5.4.2.1 EBC for Method Development and Comparison

EBC from a healthy volunteer was collected using an R-Tube collector (Respiratory Research, Inc., Austin, TX, USA), pooled to obtain a single, homogenous sample for method development, and kept frozen at −80 °C until processed. Prior to analysis, EBC was thawed, and 2-mL aliquots of the sample were placed into separate vials, stored at −80 °C for a minimum of 2 h, and lyophilized overnight using a VirTis Genesis 25EL lyophilizer (SP Industries, Stone Ridge, NY, USA) according to the program detailed in Table 5.1. The lyophilized residues were resuspended without derivatization in 100 μL of acetonitrile:water 80:20 v/v (concentration factor = 20) and recombined to make a single, homogenous EBC concentrate. This concentrate was then comparatively analyzed by both TM-DART-TWIM-MS and DI electrospray chemical ionization (ESCi) TWIM-MS. Three technical replicates were performed for each method.

Table 5.1: Lyophilization program for EBC sample processing.

Step	Temp (°C)	Time (min)	Pressure (mTorr)	Ramp/Hold
1	-34	360	200	H
2	-20	15	200	R
3	-20	120	200	H
4	0	20	200	H
5	0	120	200	R
6	20	20	200	H
7	20	180	200	H
8	24	20	200	R
9	24	600[a]	200	H

[a]Run time of this step is not critical, and samples can be removed at any point within this time.

5.4.2.2 EBC of Controls and Cystic Fibrosis Patients

Samples were collected by the CF Discovery Core, part of the Emory+Children's Center for CF and Airways Disease Research, under the guidelines approved by the Emory University Institutional Review Boards (approval number IRB00000372) and the Georgia Institute of Technology.

EBC samples collected from 5 controls and 4 patients diagnosed with CF were processed individually using same method described in the previous section. These samples were analyzed by TM-DART-TWIM-MS in the negative ion mode to test the applicability of this technique to CF metabolomics investigations. Two technical replicates were performed for each sample.

5.4.3. Transmission-Mode Direct Analysis in Real-Time (TM-DART)

A DART SVP 100 ion source and transmission module (IonSense, Saugus, MA) were used to conduct TM-DART experiments at a 1 L min^{-1} He gas flow rate. The discharge gas was heated to either 250 °C (positive mode) or 300 °C (negative mode). A custom-built flange and gas-ion separator tube (GIST), connected to a Vacuubrand 2C diaphragm pump (Vacuubrand, Wertheim, Germany), were used to couple the DART ion source to the mass spectrometer in order to reduce the amount of gas flowing into the atmospheric pressure inlet. The exit grid voltage was set at 300 V for both ion polarities. Individual 1-cm stainless steel mesh discs were placed within the 10-position transmission module, and 4 µL of sample (e.g., solvent blank or EBC concentrate) were deposited in the center of the exposed mesh area and allowed to dry for approximately 5 min. A blank mesh was used between sample runs to minimize cross-contamination. An automated software method was used to introduce the transmission module into the DART ionization region at 10 mm s^{-1}, and each position was held within the DART ionization region for 2 min. To avoid disturbing the DART ionizing gas stream, a 1 s hold time was used when the transmission module was advanced to the next position. These experiments generated a transient peak-shaped chronogram with a typical FWHM of approximately 6 s for positive ion mode and between 1–2 s for negative ion mode.

5.4.4 Direct Infusion Electrospray Chemical Ionization (DI-ESCi)

An ESCi multi-mode ion source (Waters Corporation, Manchester, UK) was used for high-speed switching between ESI and APCI within the same analytical run. Polarity specific ion source parameters were as follows: capillary voltage: 3 kV(+)/2.2 kV(-); corona current: 15 µA(+)/20 µA(-). All other source parameters were as follows:

sampling cone voltage 35 V, source temperature 120 °C, desolvation temperature 250 °C, and nitrogen desolvation gas flow rate 650 L h^{-1}. Samples were introduced into the ESCi source using direct infusion at a flow rate of 2 µL min^{-1}, and each run was acquired for 2 minutes.

5.4.5 Traveling Wave Ion Mobility Time-of-Flight Mass Spectrometry (TWIM-TOF-MS)

TWIM-TOF-MS analysis was performed on a Synapt G2 High Definition Mass Spectrometry system (Waters Corporation, Manchester, UK), a hybrid quadrupole-ion mobility-TOF mass spectrometer with a typical resolving power of 20,000 m/Δ*m* (FWHM) and mass accuracy of 9 ppm at *m/z* 554.2615. Initial TWIM-TOF-MS experiments were performed using a control EBC concentrate sample to optimize ion source parameters for maximum signal-to-noise ratio, and TWIMS parameters for ion separation, in both positive and negative ion modes. Optimized TWIMS parameters were as follows: wave height ramped between 10–40 V, wave velocity ramped between 200 and 800 m s^{-1} in positive ion mode and between 400 and 1000 m s^{-1} in negative ion mode, IMS gas flow rate 95 mL min^{-1} in positive ion mode and 40 mL min^{-1} in negative ion mode, and a helium gas flow rate 180 mL min^{-1}. The mass spectrometer was calibrated across the *m/z* 50–1000 range using a 0.5 mM sodium formate solution prepared in 90:10 (v/v) 2-propanol:water. An option within the MS acquisition method was selected to add a drift time function which would contain mobility total ion current chronogram data for each data file. Raw data were examined using MassLynx v4.1 (Waters Corp., Milford, MA, USA).

5.4.6 Data Analysis

Spectral features were extracted from TWIM-TOF-MS data as (drift time, *m/z*) "compounds" using Progenesis QI version 2.0 (Nonlinear Dynamics, Waters Corp). Although originally designed for mining chromatographic data, the acquired TWIM-TOF-MS data could be similarly processed by importing the drift time function of each raw data file generated by the TWIM-MS system. The feature extraction workflow included mass detection followed by drift time alignment, peak picking, integration, and deconvolution to group together adducts derived from the same compound. In all cases, EBC sample data were corrected by corresponding solvent blank data to determine EBC-specific features, filtering out signals with peak areas less than or equal to 2 times of those present in solvent blanks.

After blank correction, the resulting matrix containing TM-DART-TWIM-MS spectral features in 4 CF patients and 5 controls was normalized by total ion intensity. Subsequently, *m/z* values of all extracted features were searched in the HMDB [37] with a mass error window of 10 mDa, and only those that had candidates with endogenous human and/or microbial origins were retained. The remaining features were utilized to build a model for class discrimination *via* oPLS-DA[38] (MATLAB, R2015a, The MathWorks, Natick, MA with PLS-Toolbox, version 8.0, Eigenvector Research, Inc., Manson). Reverse interval PLS-DA (iPLS-DA) was applied to autoscaled data to select an optimum panel of discriminant features and number of latent variables (LVs) that maximized the classification accuracy. The iPLS-DA interval size was set to 1 and the maximum number of LVs set to 6. Leave-one-out cross-validation (LOOCV) as was used for oPLS-DA model building due to the small sample size.

5.5 Results and Discussion

5.5.1 TM-DART-TWIM-TOF-MS Optimization

Parameters that need to be optimized in TM-DART typically include source-to-sample distance and DART gas temperature[33]. Previous work had shown that the greatest sensitivity was obtained when the DART gas outlet and the sample were in close proximity to one another.[34] A similar effect was observed in these experiments and, therefore, this parameter was not investigated further, with all TM-DART experiments using a DART-to-sample distance of ≤ 1 mm, the minimum possible. The plasma gas temperature was optimized in negative ion mode, having an important effect on the number of detected EBC spectral features, as illustrated in Figure 5.1. A compromise between sensitivity and number of detected features was achieved by using a high enough temperature for fast analyte desorption while still minimizing thermal ion fragmentation. The average number of detected features (after blank correction) increased when the gas temperature was augmented from 250 to 300°C, with more reproducible results at 300 °C; beyond 300 °C the number of detected features decreased, and therefore, 300 °C was used in all subsequent experiments.

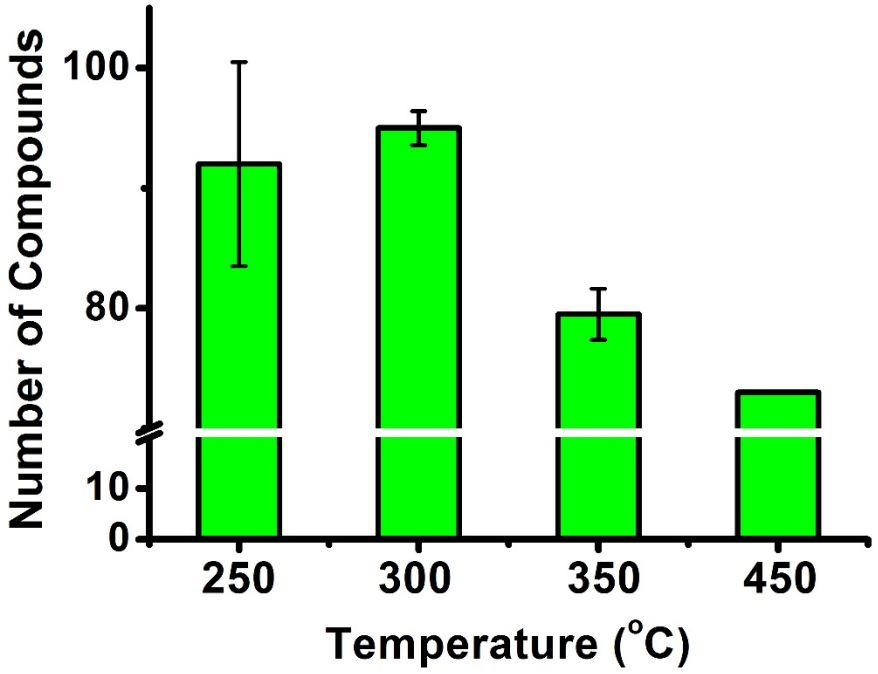

Figure 5.1: Number of compounds detected as a function of plasma gas temperature during TM-DART-TWIMS-TOF-MS optimization in negative ion mode for pooled EBC from a healthy volunteer. Error bars represent standard deviations between duplicate runs.

5.5.2 Comparison of TM-DART and DI- ESCi MS for EBC Analysis

TM-DART and DI-ESCi were compared to investigate if unique and/or complementary EBC metabolome coverage was produced from these different methods. Both positive and negative ion mass spectra of EBC from a healthy volunteer were acquired with both techniques, yielding different mass spectral patterns.

Figure 5.2 shows 2D drift time vs. *m/z* plots in positive and negative ion modes corresponding to the compounds detected after data mining and blank correction. TM-DART (Figures 5.2 (a) and (b)) resulted in the detection of only singly-charged EBC metabolites for both ion modes, most of which were grouped in compact trend lines. DI-

ESI produced a comparatively higher number of compounds (Figures 5.2 (c) and (d)) than DART (~23- and 83-fold higher for positive and negative ion modes, respectively), most likely associated with this technique's characteristic production of both singly and multiply charged ions, as well as the ability to ionize very polar and non-volatile species. The ESCi APCI mode (Figures 5.2 (e) and (f)) yielded ~3- and 4-fold more EBC-specific features than DART, but ~7- and 21-fold less than ESI, in positive and negative ion modes, respectively.

Despite the fact that DART makes use of ionization mechanisms that predominantly follow APCI-like pathways[28, 30], the 2D drift time vs. *m/z* plots of TM-DART appeared to be different from those of DI-APCI for both ion polarities. As indicated above, the ESCi APCI function produced more ions than DART in both ion modes, producing a few multiply charged ions in positive ion mode (Figure 5.2 (e)), and predominantly singly charged ions in negative ion mode (Figure 5.2 (f)). DART only produced singly charged species in both positive and negative ionization modes (Figures 5.2 (a) and (b)). These differences may arise from the way in which the specific dual ESCi ion source operates, i.e., the desolvation parameters (temperature and nitrogen gas flow rate) used in ESCi were unchanged during the alternating switching between ESI and APCI modes. Therefore, analytes may not be exposed to true APCI-like conditions where they are fully desolvated and vaporized at a higher temperature than that utilized in the dual ion source.

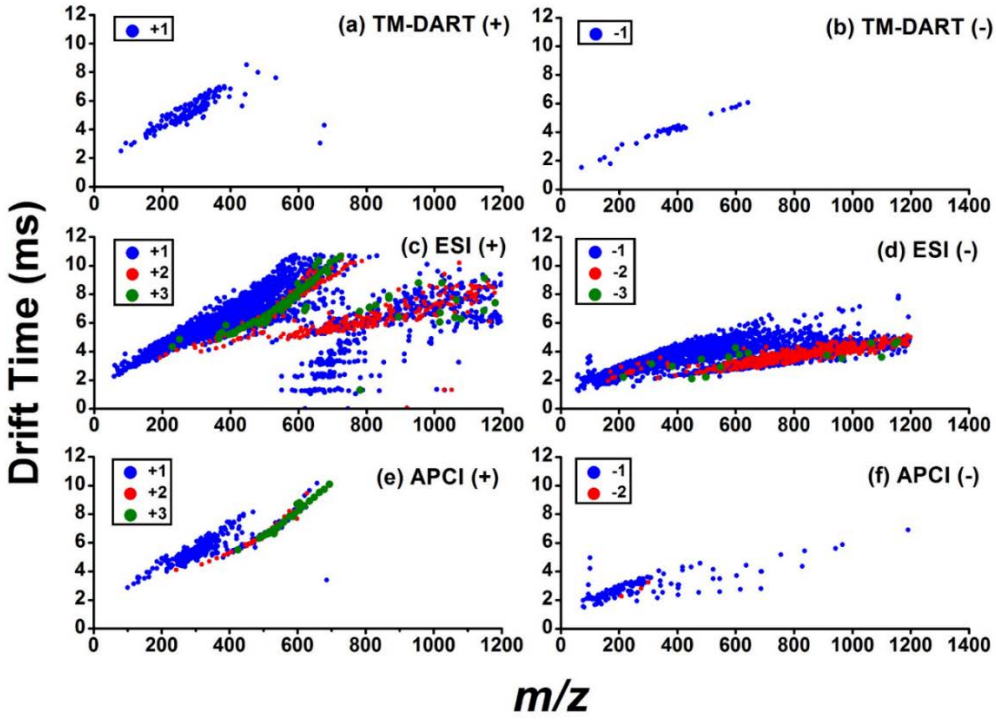

Figure 5.2: Drift time vs. *m/z* plots for datasets obtained with different techniques from a healthy volunteer's EBC. Data were blank-corrected, de-isotoped, and corrected for multiple adducts ((a) TM-DART (+): 106 compounds; (b) TM-DART (-): 31 compounds; (c) ESI (+): 2449 compounds; (d) ESI (-): 2559 compounds; (e) APCI (+): 357 compounds; and (f) APCI (-) 122 compounds).

The data shown in Figure 5.2 were further analyzed to quantitatively determine the number of unique and overlapping features between DART, ESI, and APCI, as shown in Figure 5.3. For the three ionization methods, a total of 2686 and 2619 EBC-specific compounds were detected for positive and negative ion modes, respectively, evidencing the capability of EBC metabolomics approaches to profile airway secretions. Not surprisingly, ESI accounted for a large number of species, detecting 91% of the total compounds in positive ion mode, and 98% in negative ion mode. In contrast, APCI produced about 13% and 5%, and DART detected about 4% and 1% of the total number of compounds in positive and negative ion modes, respectively. Many of the EBC signals produced using DART were removed after blank correction, and therefore the number of DART-detected compounds was much lower, overall.

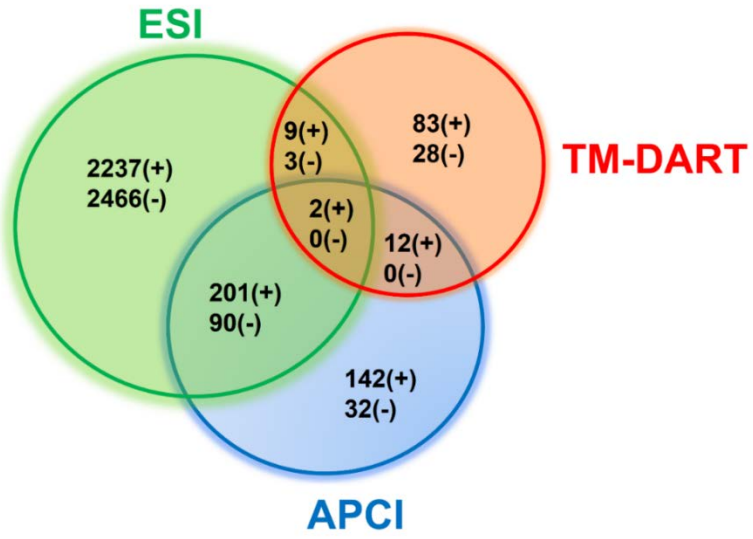

Figure 5.3: Venn diagrams illustrating EBC metabolome overlap in coverage for the investigated ionization techniques, in terms of the number of compounds detected after blank correction.

It is noted that there are numerous unique compounds produced by only one ionization method (Figure 5.3). The proportions of unique compounds to total compounds detected by any of the three ionization methods are 78% and 90% for DART positive and negative ion modes, 91% and 96% for ESI positive and negative ion modes, and 40% and 26% for APCI positive and negative ion modes, respectively. The proportions of compounds detected by APCI overlapping with those detected by ESI reached 57% and 74% for positive and negative ion modes, respectively. However, the compounds detected by DART had a much smaller overlap with the other two ionization methods, with 10% for both positive and negative ESI, and 13% and 0% with positive and negative mode APCI, respectively. These results indicate that although DART produced a lower number of features, these were rather unique and may be an important complement to EBC metabolomics experiments performed by either ESI or APCI.

To explore the signal abundances obtained from the three different ionization methods, scatter plots of average peak areas of overlapping compounds were created (Figures 5.4 to 5.13). In general, for compounds overlapping between ESI and APCI, signal intensities were higher in ESI than in APCI for both positive and negative ion modes (Figures 5.4 to 5.8, 5.11 and 5.12) with a median increase from APCI to ESI of 26-fold for positive ion mode and 64-fold for negative ion mode. For compounds overlapping between ESI and DART, there was no obvious trend in the signal intensities (Figures 5.9 and 5.13). For overlapping compounds between APCI and DART in positive ion mode, signal intensities were higher in DART than in APCI, with a median increase of 6-fold from APCI to DART (Figure 5.10).

Relative standard deviations (RSDs) of overlapping compounds in Figure 5.3 were calculated to compare the precision of each method. The results are shown as box plots in Figure 5.14. In general, ESI had a higher precision than APCI, which had a higher precision than DART. For overlapping compounds between ESI and APCI in positive ion mode, the median RSDs were 7% and 12%, respectively; and 6% and 15% in negative ion mode, respectively. For overlapping compounds between ESI and DART, the median RSD was 8% for both ionization methods in positive ion mode; and 9% and 17% in negative ion mode, respectively. The median RSDs of overlapping compounds between APCI and DART in positive ion mode were 7% and 10%, respectively.

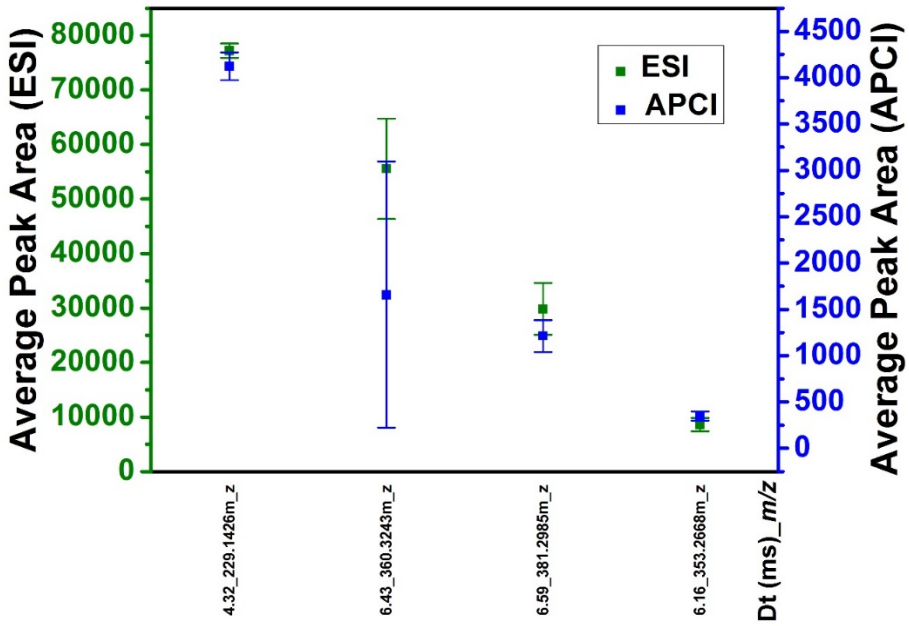

Figure 5.4: Scatter plot of average peak areas (>5000 for ESI) of overlapping compounds between ESI and APCI (+). Error bars represent standard deviations among triplicate runs.

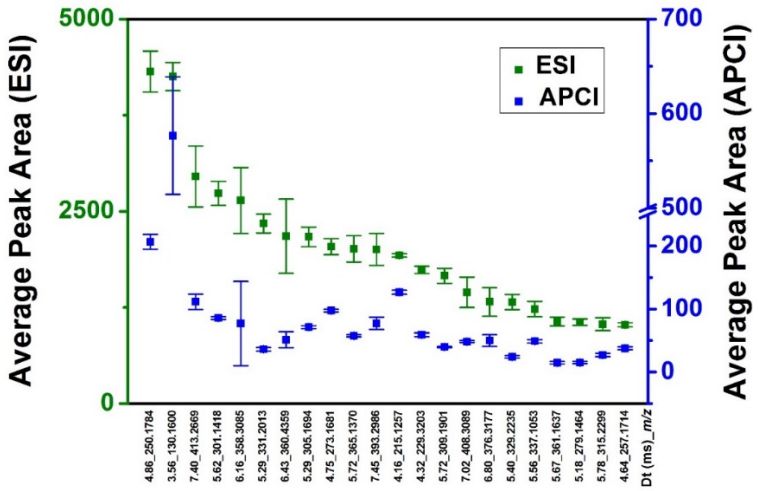

Figure 5.5: Scatter plot of average peak areas (1000–5000 for ESI) of overlapping compounds between ESI and APCI (+). Error bars represent standard deviations among triplicate runs.

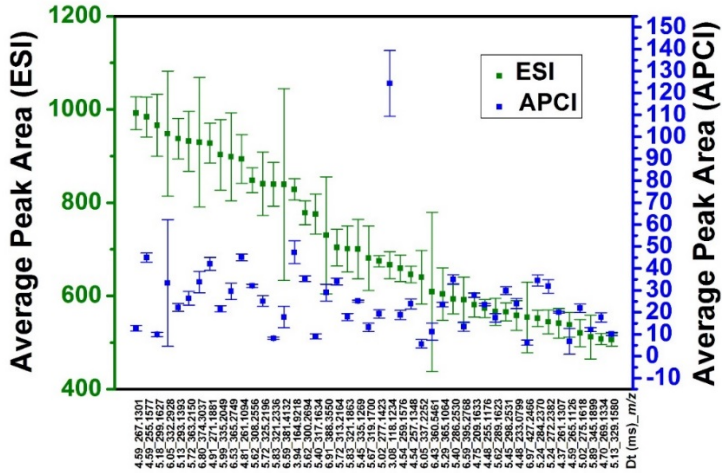

Figure 5.6: Scatter plot of average peak areas (500–1000 for ESI) of overlapping compounds between ESI and APCI (+). Error bars represent standard deviations among triplicate runs.

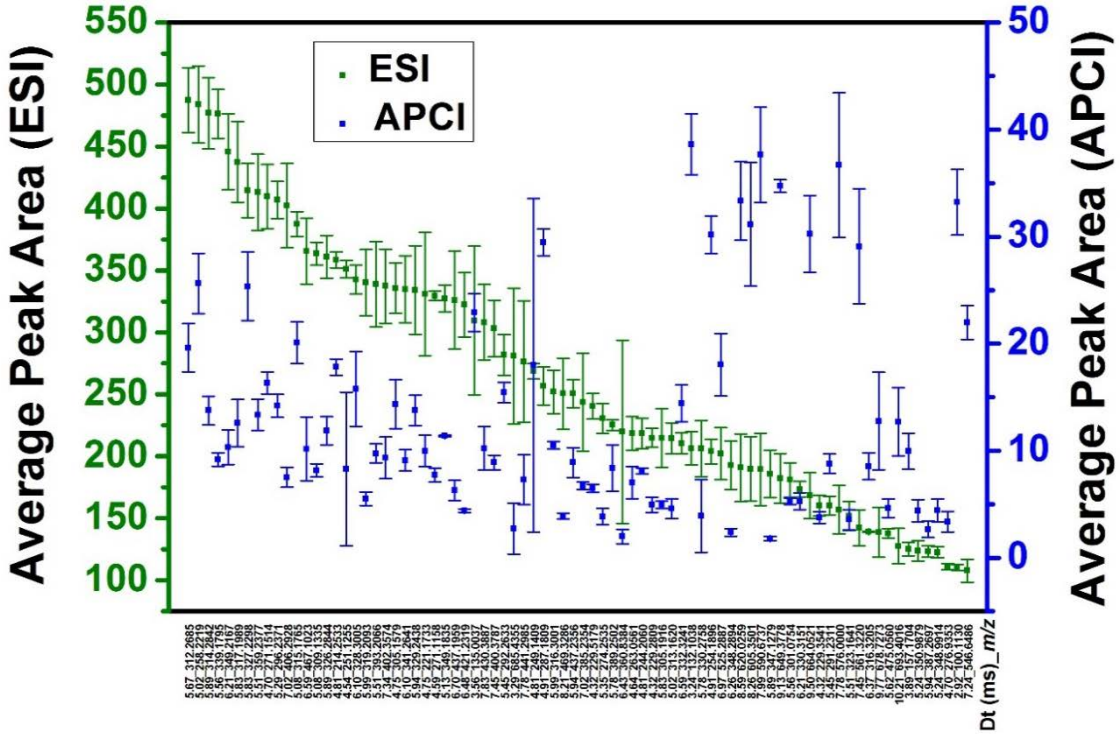

Figure 5.7: Scatter plot of average peak areas (100–500 for ESI) of overlapping compounds between ESI and APCI (+). Error bars represent standard deviations among triplicate runs.

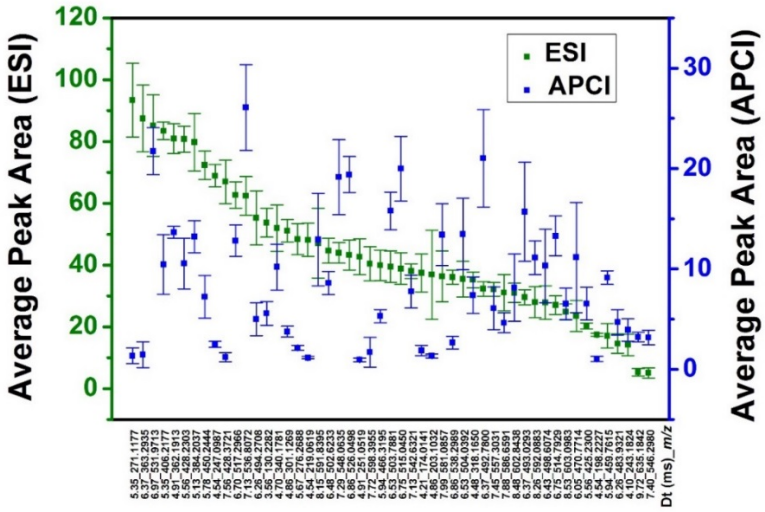

Figure 5.8: Scatter plot of average peak areas (0–100 for ESI) of overlapping compounds between ESI and APCI (+). Error bars represent standard deviations among triplicate runs.

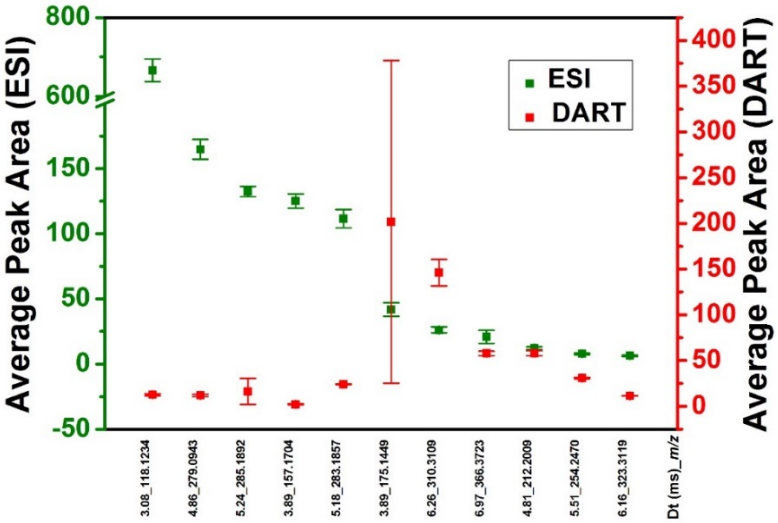

Figure 5.9: Scatter plot of average peak areas of overlapping compounds between ESI and DART (+). Error bars represent standard deviations among triplicate runs.

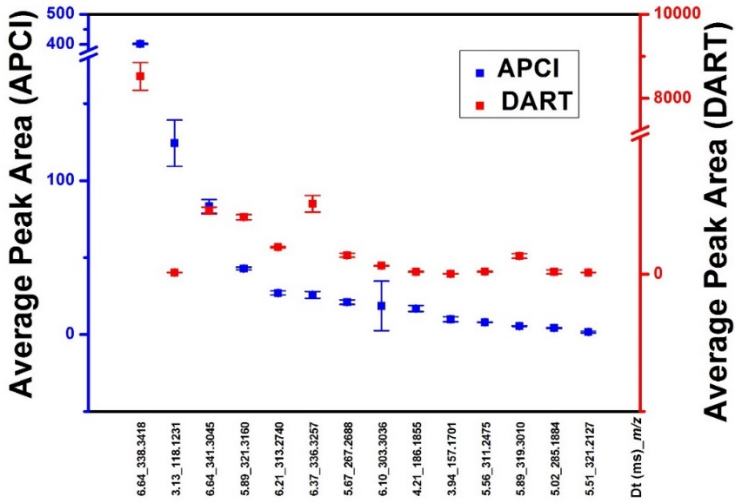

Figure 5.10: Scatter plot of average peak areas of overlapping compounds between APCI and DART (+). Error bars represent standard deviations among triplicate runs.

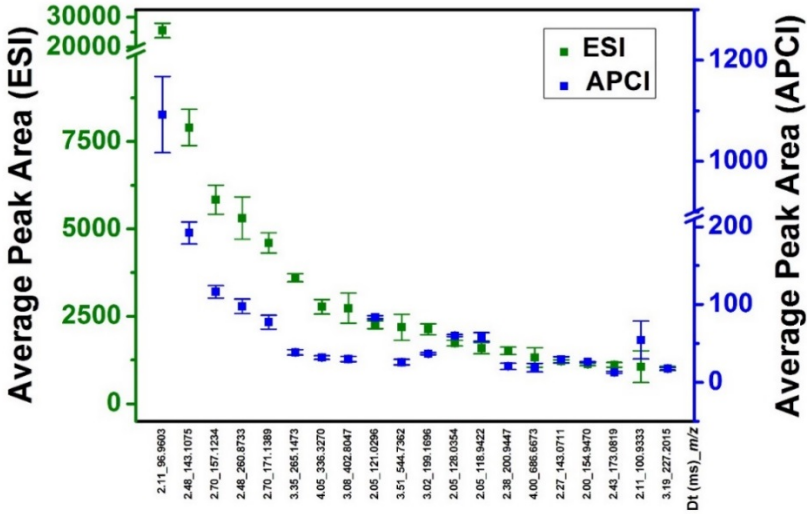

Figure 5.11: Scatter plot of average peak areas (>1000 for ESI) of overlapping compounds between ESI and APCI (-). Error bars represent standard deviations among triplicate runs.

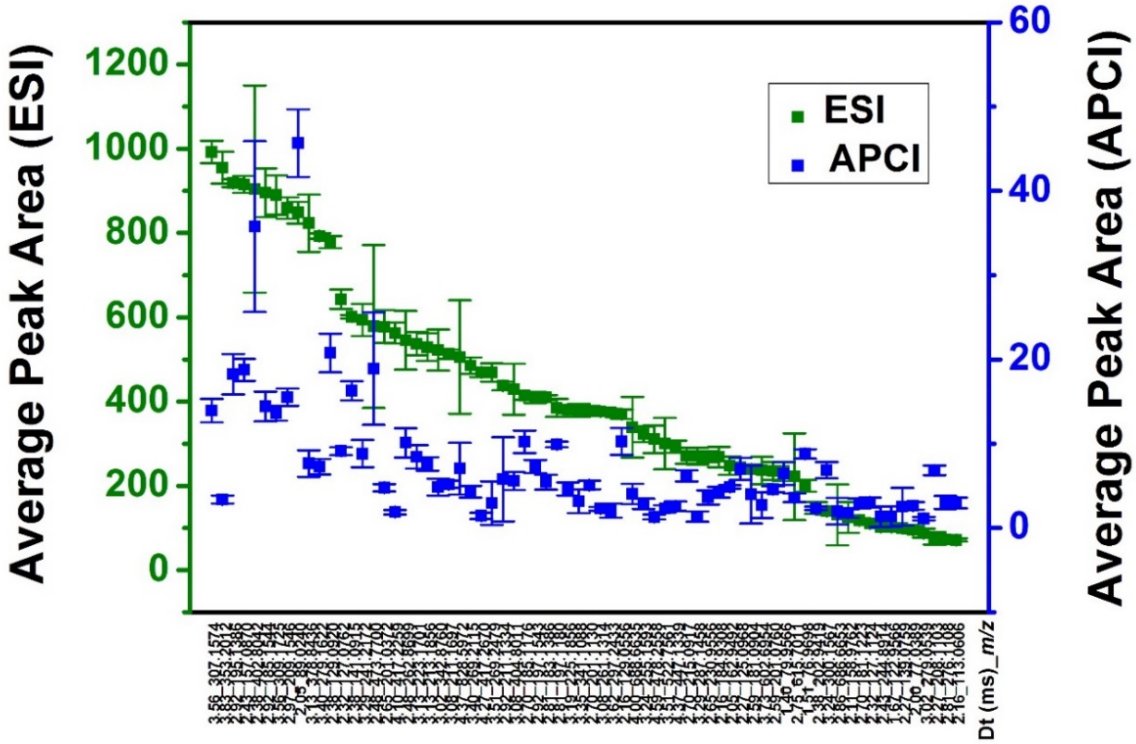

Figure 5.12: Scatter plot of average peak areas (0–1000 for ESI) of overlapping compounds between ESI and APCI (-). Error bars represent standard deviations among triplicate runs.

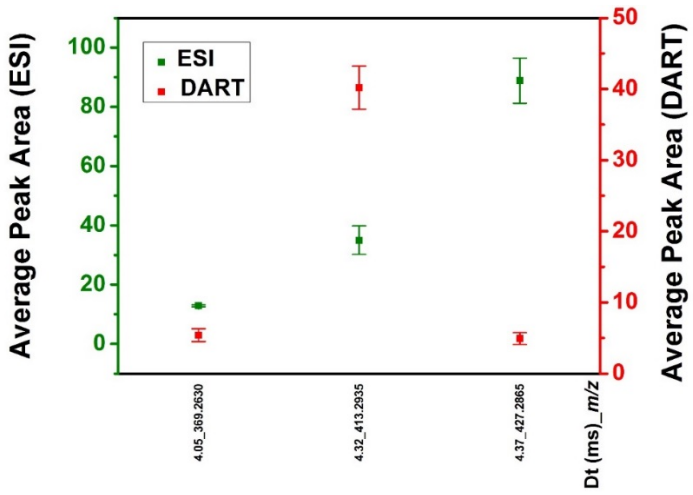

Figure 5.13: Scatter plot of average peak areas of overlapping compounds between ESI and DART (-). Error bars represent standard deviations among triplicate runs.

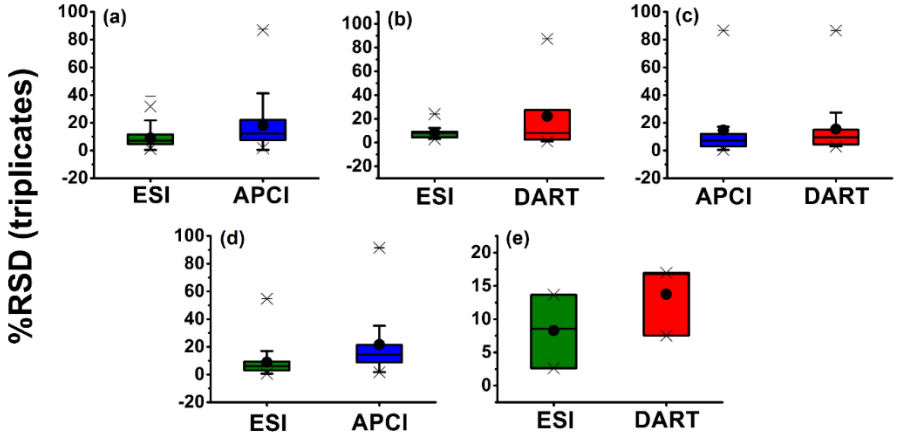

Figure 5.14: Box plots of RSDs of overlapping compounds in Figure 5.3 ((a) ESI and APCI (+): 203 compounds; (b) ESI and DART (+): 11 compounds; (c) APCI and DART (+): 14 compounds; (d) ESI and APCI (-): 90 compounds; (e) ESI and DART (-): 3 compounds). Mean values are represented by a filled circle in the box; median values are represented by a line in the box; the edges of the box are 25th and 75th percentiles; the whisker extends to the most extreme values in the data not including outliers, with a 99.3% coverage; outliers are represented by asterisks.

Regarding analysis time, DART experiments presented the advantage that they did not require rinsing tubing which connected the pump propelling the liquid samples to the ion source as in DI-ESCi, which was translated into higher-throughput sample analysis. In addition, the fact that no wetted tubing is involved in DART, also made this technique much more resistant to carry-over effects.

5.5.3 CF Sample Analysis and Multivariate Classification

The capability of TM-DART-TWIM-TOF-MS to rapidly acquire metabolic profiles from EBC was tested to investigate if the metabolic differences between 4 CF patients and 5 controls could be rapidly established using this approach. Negative ion mode was chosen because the EBC pH in CF patients is known to be lower than in controls[39], suggesting that acidic metabolites may be important discriminant species. Overall, metabolic profile acquisition required less than 20 min per sample following lyophilization, considering ~5 min for sample reconstitution, 5–10 min for sample deposition and drying, and ~4 min per sample for replicate DART analysis. This last step is ~5–15 times faster than a typical 10–30 min LC-MS run. Representative data for two EBC samples from a selected CF patient and a control subject are shown in Figures 5.15 (a) and (b), respectively, depicting mass spectra in the range of *m/z* 50–400. Some differences could be readily observed. A feature with *m/z* 128.0382 (marked with an asterisk), for example, showed much higher abundance in the CF patient sample than in the control sample. A total of 29 metabolic features were extracted from the TM-DART metabolic profiles. Of these features, 11 had candidates with endogenous human and/or microbial origins in the HMDB, which were subjected to multivariate classification.

Figures 5.16 (a) and (b) display the multivariate classification result of EBC samples from the 4 CF patients and those from the 5 controls. A set of 3 metabolic features and 2 LVs was selected by the reverse iPLS-DA feature selection process. The resulting oPLS-DA model (Figure 5.16 (a)) yielded 100% cross-validated accuracy, sensitivity, and specificity. This model captured 70.7% and 96.9% of the X- and Y-block variances, respectively.

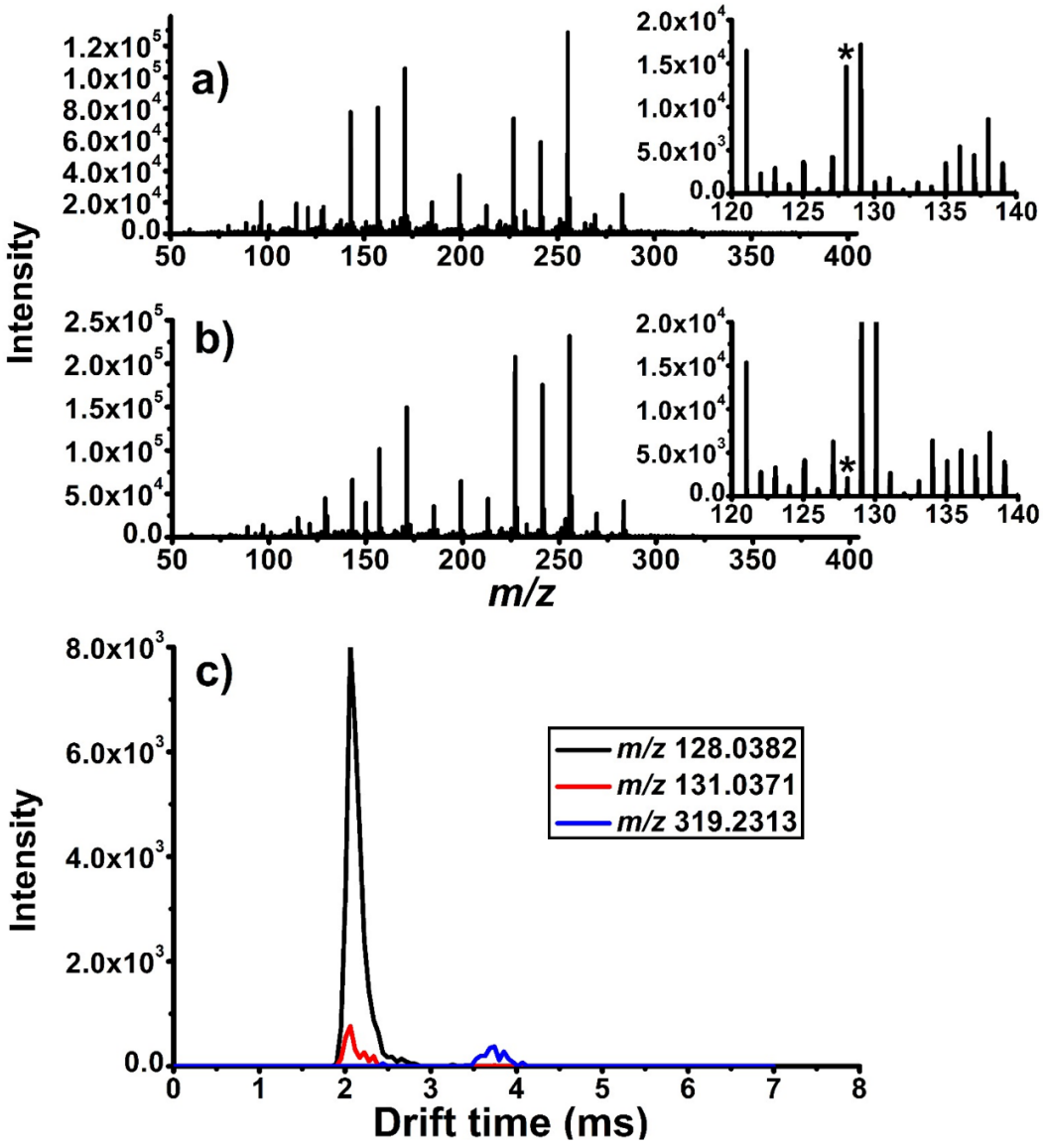

Figure 5.15: Negative ion TM-DART-TWIM-TOF mass spectrum from (a) a sample from a CF patient, and (b) a sample from a control subject (inset shows a zoomed in view of the *m/z* 120–140 range). The asterisk indicates the spectral peak at *m/z* 128.0382. (c) Extracted ion mobility chronograms for the best 3 discriminant features from the CF patient sample illustrated in (a).

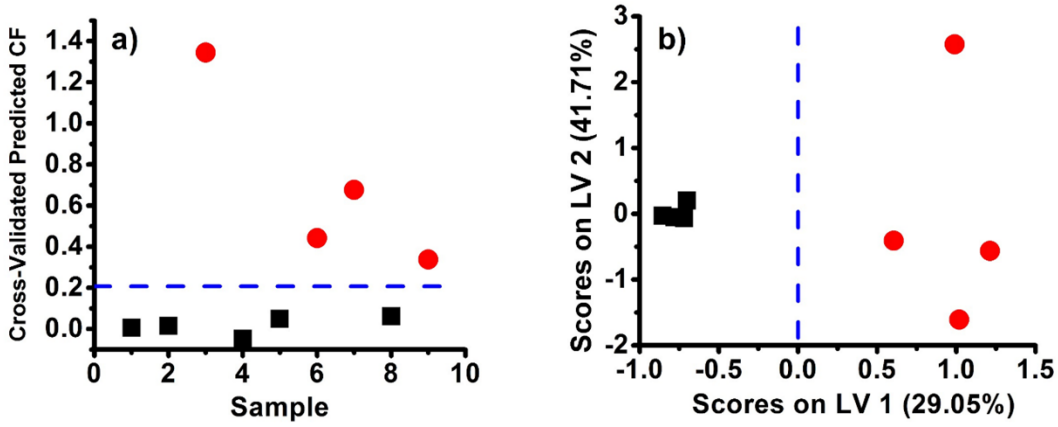

Figure 5.16: oPLS-DA model for discrimination of CF patient samples (red circles) from control samples (black squares). (a) Cross-validated prediction plot using the 3 discriminant metabolic feature panel obtained from iPLS-DA variable selection. (b) oPLS-DA calibration scores plot for (a). The model consisted of 2 LVs with 70.7% and 96.9% total captured X- and Y-block variances, respectively. The accuracy, sensitivity, and specificity were all 100%.

Figure 5.15 (c) shows the TM-DART extracted ion mobility chronograms of the three discriminant features used in the oPLS-DA model, which vary in relative abundances. Metabolites in the 3-feature panel were tentatively identified (Table 5.2) based on accurate mass measurements and database searches. Different metabolites were separated in drift time in TWIMS based on their *m/z* and structure types. Feature #1 had a mean fold increase of 17.4 from control to CF samples, and was tentatively identified as pyroglutamic acid or its structural isomers. Interestingly, in a metabolomics study by Joseloff *et al.*,[40] pyroglutamic acid was found as an important metabolite responsible for distinguishing 31 CF from 31 non-CF serum samples from children, with a CF to non-CF fold change of 1.2. As well, this metabolite, which is involved in the γ-glutamyl cycle,

was identified as increased in CF patients before an acute pulmonary exacerbation in the pilot study described in Chapter 3,[41] and also selected in the 9-feature negative mode data model differentiating APE from stable CF pediatric samples of the larger cohort study (Table 3.8.1). Feature #2 had tentative identifications matching short chain carboxylic acids (dimethylmalonic acid, 2-acetolactate, glutaric acid, monoethyl malonic acid, ethylmalonic acid and methylsuccinic acid, amino acids/peptides (glycyl-glycine and asparagine), and intermediates in amino acid (N-carbamoylsarcosine) or uracil (ureidopropionic acid) metabolic pathways. Feature #3 had a mean fold increase of 14.2 from control to CF samples. It was tentatively assigned to several hydroxyeicosatetraenoic acids (HETEs) (5-, 8-, 9-, 11-, 12-, 15-, 16-, 17-, 18-, 19- and 20-HETE) and epoxyeicosatrienoic acids (EETs) (5,6-, 8,9-, 11,12- and 14,15-EET) involved in arachidonic acid metabolism. Interestingly, an increased arachidonic acid ratio has been reported in bronchial phospholipids in CF patients compared to normal controls, suggesting abnormal arachidonic acid metabolism in CF patients.[42]

Table 5.2: Metabolites tentatively identified as discriminatory between cystic fibrosis patients and controls.

Feature Code	Drift time (ms)	*m/z*	Ion type	Elemental formula	Δm (mDa)	Mean fold change (CF patients to controls)	Tentative metabolite identification
1	2.06	128.0382	$[M-H]^-$	$C_5H_7NO_3$	3.4	17.4	Pyroglutamic acid; 1-Pyrroline-4-hydroxy-2-carboxylate; N-Acryloylglycine; Pyrroline hydroxycarboxylic acid
2	2.06	131.0371	$[M-H]^-$	$C_5H_8O_4$	2.7	N/A[a]	Dimethylmalonic acid; 2-Acetolactate; Glutaric acid; Monoethyl malonic acid; Ethylmalonic acid; Methylsuccinic acid
				$C_4H_8N_2O_3$	-8.6		Glycyl-glycine; Asparagine; Ureidopropionic acid; N-Carbamoylsarcosine
3	3.80	319.2313	$[M-H]^-$	$C_{20}H_{32}O_3$	4.0	14.2	5-HETE[b]; 8-HETE; 9-HETE; 11-HETE; 12-HETE; 15-HETE; 16-HETE; 17-HETE; 18-HETE; 19-HETE; 20-HETE; 5,6-EET[c]; 8,9-EET; 11,12-EET; 14,15-EET

[a]N/A: the fold change cannot be calculated since the abundances are all 0 in controls.

[b]HETE: hydroxyeicosatetraenoic acid. [c]EET: epoxyeicosatrienoic acid.

5.6 Conclusions

TWIM-TOF-MS for CF non-targeted EBC metabolomics studies is demonstrated as a potential high-throughput alternative to conventional LC-MS-based methods typically used in present investigations. The EBC metabolome coverage provided by DART ionization is found to be an important complement to ESI and APCI-based methods. The metabolites detected by DART can provide biochemical information pertinent in metabolomics applications for studying pathophysiological processes occurring within the lung. The analysis of a small set of samples from a cohort of CF patients and controls shows the initial application of DART in EBC metabolomics. Multivariate analyses of the resulting TM-DART-TWIM-TOF-MS datasets successfully discriminated between EBC samples from CF patients and controls.

5.7 References

1. Hunt, J., Exhaled breath condensate: An evolving tool for noninvasive evaluation of lung disease. *Journal of Allergy and Clinical Immunology* **2002,** 110, (1), 28-34.

2. Mutlu, G. M.; Garey, K. W.; Robbins, R. A.; Danziger, L. H.; Rubinstein, I., Collection and analysis of exhaled breath condensate in humans. *American Journal of Respiratory and Critical Care Medicine* **2001,** 164, (5), 731-737.

3. Montuschi, P.; Barnes, P. J., Analysis of exhaled breath condensate for monitoring airway inflammation. *Trends in Pharmacological Sciences* **2002,** 23, (5), 232-237.

4. Grob, N. M.; Aytekin, M.; Dweik, R. A., Biomarkers in exhaled breath condensate: a review of collection, processing and analysis. *Journal of Breath Research* **2008,** 2, (3), 18.

5. Carraro, S.; Giordano, G.; Reniero, F.; Carpi, D.; Stocchero, M.; Sterk, P. J.; Baraldi, E., Asthma severity in childhood and metabolomic profiling of breath condensate. *Allergy* **2013,** 68, (1), 110-117.

6. Carraro, S.; Rezzi, S.; Reniero, F.; Heberger, K.; Giordano, G.; Zanconato, S.; Guillou, C.; Baraldi, E., Metabolomics applied to exhaled breath condensate in childhood asthma. *American Journal of Respiratory and Critical Care Medicine* **2007,** 175, (10), 986-990.

7. Esther, C. R.; Jasin, H. M.; Collins, L. B.; Swenberg, J. A.; Boysen, G., A mass spectrometric method to simultaneously measure a biomarker and dilution marker in exhaled breath condensate. *Rapid Communications in Mass Spectrometry* **2008,** 22, (5), 701-705.

8. Montuschi, P., Exhaled breath condensate analysis in patients with COPD. *Clinica Chimica Acta* **2005,** 356, (1–2), 22-34.

9. Montuschi, P.; Paris, D.; Melck, D.; Lucidi, V.; Ciabattoni, G.; Raia, V.; Calabrese, C.; Bush, A.; Barnes, P. J.; Motta, A., NMR spectroscopy metabolomic profiling of exhaled breath condensate in patients with stable and unstable cystic fibrosis. *Thorax* **2012,** 67, (3), 222-228.

10. Nobakht, M. G. B. F.; Aliannejad, R.; Rezaei-Tavirani, M.; Taheri, S.; Oskouie, A. A., The metabolomics of airway diseases, including COPD, asthma and cystic fibrosis. *Biomarkers* **2015,** 20, (1), 5-16.

11. Snowden, S.; Dahlen, S. E.; Wheelock, C. E., Application of metabolomics approaches to the study of respiratory diseases. *Bioanalysis* **2012,** 4, (18), 2265-2290.

12. Wheelock, C. E.; Goss, V. M.; Balgoma, D.; Nicholas, B.; Brandsma, J.; Skipp, P. J.; Snowden, S.; Burg, D.; D'Amico, A.; Horvath, I.; Chaiboonchoe, A.; Ahmed, H.; Ballereau, S.; Rossios, C.; Chung, K. F.; Montuschi, P.; Fowler, S. J.; Adcock, I. M.; Postle, A. D.; Dahlen, S. E.; Rowe, A.; Sterk, P. J.; Auffray, C.; Djukanovic, R.; Grp, U. B. S., Application of 'omics technologies to biomarker discovery in inflammatory lung diseases. *European Respiratory Journal* **2013,** 42, (3), 802-825.

13. Dettmer, K.; Aronov, P. A.; Hammock, B. D., Mass spectrometry-based metabolomics. *Mass Spectrometry Reviews* **2007,** 26, (1), 51-78.

14. Lenz, E. M.; Wilson, I. D., Analytical strategies in metabonomics. *Journal of Proteome Research* **2007,** 6, (2), 443-458.

15. Sofia, M.; Maniscalco, M.; de Laurentiis, G.; Paris, D.; Melck, D.; Motta, A., Exploring Airway Diseases by NMR-Based Metabonomics: A Review of Application to Exhaled Breath Condensate. *Journal of Biomedicine and Biotechnology* **2011**, 7.

16. Wishart, D. S., Quantitative metabolomics using NMR. *TRAC Trends in Analytical Chemistry* **2008,** 27, (3), 228-237.

17. Effros, R. M.; Hoagland, K. W.; Bosbous, M.; Castillo, D.; Foss, B.; Dunning, M.; Gare, M.; Lin, W.; Sun, F., Dilution of respiratory solutes in exhaled condensates. *American Journal of Respiratory and Critical Care Medicine* **2002,** 165, (5), 663-669.

18. Draper, J.; Lloyd, A. J.; Goodacre, R.; Beckmann, M., Flow infusion electrospray ionisation mass spectrometry for high throughput, non-targeted metabolite fingerprinting: a review. *Metabolomics* **2013,** 9, (1), S4-S29.

19. May, J. C.; McLean, J. A., Ion Mobility-Mass Spectrometry: Time-Dispersive Instrumentation. *Analytical Chemistry* **2015,** 87, (3), 1422-1436.

20. McLean, J. A.; Ruotolo, B. T.; Gillig, K. J.; Russell, D. H., Ion mobility-mass spectrometry: a new paradigm for proteomics. *International Journal of Mass Spectrometry* **2005,** 240, (3), 301-315.

21. Taraszka, J. A.; Counterman, A. E.; Clemmer, D. E., Gas-phase separations of complex tryptic peptide mixtures. *FRESENIUS JOURNAL OF ANALYTICAL CHEMISTRY* **2001,** 369, (3-4), 234-245.

22. Chouinard, C. D.; Wei, M. S.; Beekman, C. R.; Kemperman, R. H. J.; Yost, R. A., Ion Mobility in Clinical Analysis: Current Progress and Future Perspectives. *Clinical Chemistry* **2016,** 62, (1), 124-133.

23. Paglia, G.; Williams, J. P.; Menikarachchi, L.; Thompson, J. W.; Tyldesley-Worster, R.; Halldorsson, S.; Rolfsson, O.; Moseley, A.; Grant, D.; Langridge, J.; Palsson, B. O.; Astarita, G., Ion Mobility Derived Collision Cross Sections to Support Metabolomics Applications. *Analytical Chemistry* **2014,** 86, (8), 3985-3993.

24. Cody, R. B.; Laramee, J. A.; Durst, H. D., Versatile new ion source for the analysis of materials in open air under ambient conditions. *Analytical Chemistry* **2005,** 77, (8), 2297-2302.

25. Harris, G. A.; Kwasnik, M.; Fernandez, F. M., Direct analysis in real time coupled to multiplexed drift tube ion mobility spectrometry for detecting toxic chemicals. *Analytical Chemistry* **2011,** 83, (6), 1908-15.

26. Keelor, J. D.; Dwivedi, P.; Fernandez, F. M., An effective approach for coupling direct analysis in real time with atmospheric pressure drift tube ion mobility spectrometry. *Journal of the American Society for Mass Spectrometry* **2014,** 25, (9), 1538-48.

27. Rasanen, R. M.; Dwivedi, P.; Fernandez, F. M.; Kauppila, T. J., Desorption atmospheric pressure photoionization and direct analysis in real time coupled with travelling wave ion mobility mass spectrometry. *Rapid Communications in Mass Spectrometry* **2014,** 28, (21), 2325-2336.

28. McEwen, C. N.; Larsen, B. S., Ionization Mechanisms Related to Negative Ion APPI, APCI, and DART. *Journal of the American Society for Mass Spectrometry* **2009,** 20, (8), 1518-1521.

29. Song, L. G.; Dykstra, A. B.; Yao, H. F.; Bartmess, J. E., Ionization Mechanism of Negative Ion-Direct Analysis in Real Time: A Comparative Study with Negative Ion-

Atmospheric Pressure Photoionization. *Journal of the American Society for Mass Spectrometry* **2009,** 20, (1), 42-50.

30. Song, L. G.; Gibson, S. C.; Bhandari, D.; Cook, K. D.; Bartmess, J. E., Ionization Mechanism of Positive-Ion Direct Analysis in Real Time: A Transient Microenvironment Concept. *Analytical Chemistry* **2009,** 81, (24), 10080-10088.

31. Harris, G. A.; Fernandez, F. M., Simulations and experimental investigation of atmospheric transport in an ambient metastable-induced chemical ionization source. *Analytical Chemistry* **2009,** 81, (1), 322-9.

32. Harris, G. A.; Falcone, C. E.; Fernandez, F. M., Sensitivity "hot spots" in the direct analysis in real time mass spectrometry of nerve agent simulants. *Journal of the American Society for Mass Spectrometry* **2012,** 23, (1), 153-61.

33. Perez, J. J.; Harris, G. A.; Chipuk, J. E.; Brodbelt, J. S.; Green, M. D.; Hampton, C. Y.; Fernandez, F. M., Transmission-mode direct analysis in real time and desorption electrospray ionization mass spectrometry of insecticide-treated bednets for malaria control. *Analyst* **2010,** 135, (4), 712-719.

34. Jones, C. M.; Fernandez, F. M., Transmission mode direct analysis in real time mass spectrometry for fast untargeted metabolic fingerprinting. *Rapid Communications in Mass Spectrometry* **2013,** 27, (12), 1311-8.

35. Monge, M. E.; Perez, J. J.; Dwivedi, P.; Zhou, M.; McCarty, N. A.; Stecenko, A. A.; Fernandez, F. M., Ion mobility and liquid chromatography/mass spectrometry strategies for exhaled breath condensate glucose quantitation in cystic fibrosis studies. *Rapid Commun Mass Spectrom* **2013,** 27, (20), 2263-71.

36. Wetmore, D. R.; Joseloff, E.; Pilewski, J.; Lee, D. P.; Lawton, K. A.; Mitchell, M. W.; Milburn, M. V.; Ryals, J. A.; Guo, L., Metabolomic profiling reveals biochemical pathways and biomarkers associated with pathogenesis in cystic fibrosis cells. *J Biol Chem* **2010,** 285, (40), 30516-22.

37. 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., HMDB 3.0--The Human Metabolome Database in 2013. *Nucleic Acids Res* **2013,** 41, (Database issue), D801-7.

38. Trygg, J.; Wold, S., Orthogonal projections to latent structures (O-PLS). *Journal of Chemometrics* **2002,** 16, (3), 119-128.

39. Ojoo, J. C.; Mulrennan, S. A.; Kastelik, J. A.; Morice, A. H.; Redington, A. E., Exhaled breath condensate pH and exhaled nitric oxide in allergic asthma and in cystic fibrosis. *Thorax* **2005,** 60, (1), 22-26.

40. Joseloff, E.; Sha, W.; Bell, S. C.; Wetmore, D. R.; Lawton, K. A.; Milburn, M. V.; Ryals, J. A.; Guo, L.; Muhlebach, M. S., Serum metabolomics indicate altered cellular energy metabolism in children with cystic fibrosis. *Pediatr. Pulmonol.* **2014,** 49, (5), 463-72.

41. Zang, X.; Monge, M. E.; McCarty, N. A.; Stecenko, A. A.; Fernandez, F. M., Feasibility of Early Detection of Cystic Fibrosis Acute Pulmonary Exacerbations by Exhaled Breath Condensate Metabolomics: A Pilot Study. *Journal of Proteome Research* **2017,** 16, (2), 550-558.

42. Gilljam, H.; Strandvik, B.; Ellin, A.; Wiman, L. G., Increased mole fraction of arachidonic acid in bronchial phospholipids in patients with cystic fibrosis. *Scand. J. Clin. Lab. Invest.* **1986,** 46, (6), 511-8.

CHAPTER 6. CONCLUSIONS AND OUTLOOK

6.1 Abstract

This chapter summarizes conclusions drawn from this thesis work focused on non-targeted metabolomics for disease detection and prediction using ultraperformance liquid chromatography-mass spectrometry (UPLC-MS) as well as MS-based methods for rapid metabolomics workflows, including flow injection (FI), direct infusion (DI) and transmission-mode direct analysis in real time (TM-DART) ion mobility MS (IM-MS). Impact, outlook and possible future work are also discussed.

6.2 Major Accomplishments

6.2.1 Ultraperformance Liquid Chromatography-Mass Spectrometry Based Non-targeted Metabolomics for Disease Detection

Chapter 2 described a non-targeted metabolomics study for prostate cancer (PCa) detection, in which metabolic profiling of age-matched serum samples from 64 PCa patients and 50 controls was performed using ultraperformance liquid chromatography coupled to high resolution mass spectrometry (UPLC-HRMS). A metabolite-based in vitro diagnostic multivariate index assay (IVDMIA) was developed to detect the presence of PCa in serum samples. A panel of 40 metabolic features was found to be differential with 92.1% sensitivity, 94.3% specificity, and 93.0% accuracy, higher than the prevalent

PSA test.[1, 2] Within this discriminant panel, 31 metabolites were identified by MS/MS, in which 10 were further validated against chemical standards by retention time and MS/MS matching. The identification of amino acids, fatty acids, lysophospholipids and bile acids among the discriminant metabolites suggests that alterations in their metabolism were potentially associated with PCa. In addition, several metabolites were mapped to the steroid hormone biosynthesis pathway, providing further insights into PCa related biological pathway perturbation. When the assay was based on 28 identified disease-related metabolites, PCa was detected with 89.7% sensitivity, 90.7% specificity, and 90.2% accuracy. For higher throughput analysis and lower analysis cost and complexity are needed, 13 metabolites that were found to be present in 90% of the entire sample cohort still provided good classification performance of 88.3% sensitivity, 80.3% specificity and 85.0% accuracy.

Chapter 3 presented two non-targeted metabolomics studies that aimed at early detection of cystic fibrosis (CF) acute pulmonary exacerbations (APEs) by means of UPLC-HRMS coupled to multivariate statistical analysis. In a first pilot study, we demonstrated the feasibility of differentiating exhaled breath condensate (EBC) samples from 9 patients with an APE severe enough to require hospitalization from 17 clinically stable patients at the time of EBC collection, with 77.8% sensitivity, 88.2 specificity and 84.6% accuracy, based on the relative levels of 4-hydroxycyclohexylcarboxylic acids and pyroglutamic acid. Moreover, EBC samples from 4 patients clinically stable at the time of EBC collection but in the subsequent 1-3 months developed a severe APE, were discriminated from 17 stable patients with 75.0% sensitivity, 94.1% specificity, and 90.5% accuracy, based on lactic acid and pyroglutamic acid.

In a second study, metabolic profiling of a larger EBC sample cohort (n=210) was performed using UPLC coupled to ultra-high mass accuracy Orbitrap MS. Both negative ion mode data and combined positive and negative ion mode data showed classification of APE vs. stable CF and pre-APE vs. stable CF patients in adult and pediatric cohorts with good accuracies ranging between 81.3 and 93.9%. Discriminant metabolites found in the pilot study were also selected in the larger cohort study, corroborating the biological significance of the metabolite panels identified for APE detection and prediction. Metabolites exhibiting changes at different stages of an APE event were discussed in terms of their associated metabolic pathways and microbial relevance. The results from these two studies show promise for detecting APEs and even predicting an oncoming APE event using EBC metabolites, as well as providing insight into the molecular mechanisms of CF APE development.

6.2.2 Flow Injection-Ion Mobility-MS and Direct Infusion- Ion Mobility-MS Based Non-targeted Metabolomics for Disease Detection and Early Prediction

Chapter 4 presented the application of flow injection traveling wave ion mobility-time-of-flight-MS (FI-TWIM-TOF-MS) for PCa detection. In this study, non-targeted metabolic profiling was conducted on subgroup of the same sample cohort (n= 103) analyzed in Chapter 2. The high resolution and mass accuracy of FI-TWIM-TOF MS data was extensively exploited and thoroughly analyzed to provide a comprehensive compound annotation on the mobility–mass plane for the purpose of discerning between compounds with different charges and filtering out salt cluster ions, aiding in compound discrimination and dataset cleanup. For FI-TWIM-MS experiments, the software

currently available was designed for LC-MS data processing and efforts were put towards mining FI-TWIM-MS metabolomics data to enable correct data pre-processing. We also developed important criteria to ensure correct grouping of adducts, in source fragments and salt cluster peaks in FI-TWIM-MS data. Endogenous metabolites were identified with high confidence by matching FI-TWIM-MS/MS fragmentation patterns and collision cross-sections (CCSs) to those of authentic chemical standards or in databases. By combining FI-TWIM-MS and supervised classification methods, PCa patient samples were distinguished from control samples with good sensitivities (88.5-90.2%), specificities (88.1%) and accuracies (88.3-89.3%). Results from this study showed the potential of FI-TWIM-TOF-MS for high-throughput metabolic profiling in large scale non-targeted metabolomics studies, based on high analysis speed, effective compound separation and screening for endogenous metabolites in complex biological mixtures, and capability of identifying compounds with high confidence.

In Chapter 4, a transmission-mode (TM) Direct analysis in real time (DART) coupled to TWIM-TOF MS method was tested as a high-throughput alternative to conventional direct infusion (DI) electrospray ionization (ESI) and atmospheric pressure chemical ionization (APCI) methods, and a critical comparison of the three ionization methods was conducted, with the goal of expanding the toolbox available for analyzing metabolic alterations associated with CF disease pathology in EBC samples. When using pooled EBC collected from a healthy volunteer, ESI detected the most metabolites and TM-DART the least. It was found that the EBC metabolome coverage provided by DART ionization is an important complement to ESI and APCI methods, which may provide biochemical information pertinent in metabolomics applications for studying

pathophysiologic processes occurring within the lung. We described the first application of TM-DART-TWIM-TOF MS in EBC metabolomics, showing feasibility of differentiating samples from four CF patients and five controls based on a panel of three discriminant EBC metabolites with excellent cross-validated accuracy. Pyroglutamic acid was identified within this panel, in agreement with the findings reported in Chapter 3 for CF acute pulmonary exacerbation detection. The TWIM-TOF MS method applied for CF non-targeted EBC metabolomics studies was demonstrated to be a feasible high-throughput alternative to conventional LC-MS based methods typically used for these investigations.

6.3 Impact and Future Direction

This thesis work has contributed meaningfully to the metabolomics field by discovering potential biomarkers that might be useful for PCa diagnosis and early detection of APEs for CF patients, as well as by expanding our knowledge of the metabolic perturbations associated with both diseases. One of the great challenges in the metabolomics field is to find and identify useful biomarkers with high confidence. Since the majority of the metabolite composition in biological samples remains poorly understood, a comprehensive coverage and annotation of the metabolome using MS is very challenging. In order to make contribution to this scientific endeavor, significant efforts have been put into *de novo* identification of compounds not present in existing databases by manual analysis of fragmentation patterns. In addition to conventional LC-MS methods, high-throughput LC-free FI-TWIM-MS, DI-TWIM-MS and TM-DART-TWIM-MS metabolomics strategies were employed, which provided a mobility−mass

description of compounds and offered great assistance in dataset cleanup, hence enabling accurate data analysis and metabolite identification in subsequent steps. In addition, FI-TWIM-MS/MS was applied for metabolite identification, in which endogenous metabolites were identified by matching MS/MS fragmentation patterns and CCSs to those of authentic chemical standards or in databases. We performed in-depth analysis on TWIM-TOF-MS data to increase the confidence in compound identification, by developing important criteria for correct grouping of adducts, in-source fragments and salt cluster ions, reducing the number of false-positive or negative compound assignments.

Although the present non-targeted metabolomics studies show promise in clinical laboratory implementation for PCa detection and CF APE early detection, they are still at the discovery stage and future efforts related to this thesis work may include further validation of the candidate biomarkers through non-targeted metabolomics strategies in a different cohort, preferably from a geographically distinct area, and then with targeted metabolomics for absolute quantification of the biomarkers using isotopically labeled compounds as internal standards.[3-5] Finally, the potential biomarkers may be validated in a larger patient cohort with thousands of samples to assess their robustness and reliability, reaching towards the ultimate goal of translation of metabolomics research into the clinic.[5] For the PCa study specifically, María Eugenia Monge, a former research scientist in the Fernández research group and a current Research Staff member from CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina), is currently involved in a PCa study of a large Argentine cohort (n =800), including samples from healthy individuals, PCa patients and patients with benign prostatitis with their associated

clinical metadata and PSA values recorded for all subjects, to validate the results described in Chapter 2. Regarding the prediction of a CF APE onset, it would be useful to perform time-series analysis on EBC samples collected from CF patients at different stages during disease progression, which could provide insightful knowledge on the dynamics of APE development by offering snapshots of the metabolic status of the patient.[6, 7] More importantly, time-series metabolomics analysis can also take into account the personalized nature of EBC,[8] ensuring reliable phenotype biomarkers for anticipating an oncoming APE event.

Additionally, Yafeng Li, a postdoctoral fellow in the Fernández research group, is currently conducting research to analyze a subset of EBC samples from the large CF cohort study described in Chapter 3, using direct nanoESI (nESI) MS with high electric potential generated by triboelectric nanogenerators (TENGs)[9]. Compared to UPLC-MS, TENG-nESI has larger sample analysis speed with low sample consumption, highly suitable for analyzing limited EBC sample volumes. TENG-nESI also has higher sensitivity and lower detection limits compared to standard nESI,[9] offering the possibility of detecting low concentration metabolites in highly diluted EBC samples. This study aims to evaluate the feasibility of using TENG-nESI for detecting a CF APE onset and also to compare the results obtained to those from our UPLC-MS study.

Regarding high-throughput LC-free FI-TWIM-MS, DI-TWIM-MS and TM-DART-TWIM-MS metabolic profiling strategies, the methods developed for conducting experiments and data analyses can be applied to various metabolomics studies using different biological samples. Regarding method improvements, sample analysis speed in FI-TWIM-MS, can be shortened by increasing the mobile phase flow rate or replacing the

current front-end FI system with a faster injection platform such as the RapidFire platform.[10] In addition, further development of an automated metabolite identification pipeline would increase the efficiency of the proposed metabolomics workflow.

Metabolomics could provide insight into system biology, revealing the interactions between different components in cellular networks, relationships among molecular participants in the cell, including genes, proteins and metabolites, and perturbations due to disease or environment stimuli.[11] If additional omics data could be collected in the future, integrative analysis of metabolomics, transcriptomics and proteomics data could help to understand the perturbed biological pathways in a more systematic and comprehensive manner, providing deeper insight into the underlying molecular mechanisms associated with the diseases. For early detection of CF APE, investigation of the CF microbiome would improve our current understanding of the host-pathogen interactions during APE development and progression, and provide chemical information on pathogen-specific metabolites that may assist in personalized clinical decisions for early intervention for APE.

In the next 5 to 10 years, non-targeted metabolomics will shift to large scale studies with thousands of samples analyzed by high-throughput MS techniques, and targeted metabolomics will be the focus for validation of the novel biomarkers identified through non-targeted studies. The advancement in MS techniques and optimization of data analysis methods will enhance the diagnostic and prediction accuracy of the studied diseases. Collaborative efforts of the metabolomics community towards expanding the metabolite databases will greatly improve the coverage and efficiency in metabolite identification. The potential biomarkers identified will also be validated in different

laboratories, in different geographic regions, using different instruments, to affirm the robustness and reproducibility of the biomarkers,[12] which could be facilitated by the web-based data sharing across different laboratories over the world. Standardization of metabolomics workflow is also a pivotal necessity to ensure the quality and validity of the data. Sample collection from patients will be more non-invasive, easier and faster. Improvement in sensitivity and resolution in IMS and MS will improve the coverage for trace-level metabolites in biological matrices. Real-time MS analysis will become popular, and fast and automated data analysis strategy will enable rapid diagnosis and prompt treatment decision making in the hospital. Collaboration between researchers in fields of MS and physical chemistry can improve the prediction accuracy of the fragmentation pattern in tandem MS experiments, improving the metabolite identification coverage and confidence. In addition to MS techniques, combination with other techniques such as NMR will increase the power of metabolomics and confidence of metabolite identification. To realize the futuristic goal of successfully translation of metabolomics into clinics, collaboration among hospitals, pharmacy agencies and academic institutes is required to tightly link the skills, knowledge and practices among these areas.[13]

6.4 References

1. Hoffman, R. M.; Gilliland, F. D.; Adams-Cameron, M.; Hunt, W. C.; Key, C. R., Prostate-specific antigen testing accuracy in community practice. *BMC Fam Pract* **2002,** 3, 19.

2. Hoffman, R. M., Screening for Prostate Cancer. *New England Journal of Medicine* **2011,** 365, (21), 2013-2019.

3. Roberts, L. D.; Souza, A. L.; Gerszten, R. E.; Clish, C. B., Targeted metabolomics. *Curr Protoc Mol Biol* **2012,** Chapter 30, Unit 30 2 1-24.

4. Kitteringham, N. R.; Jenkins, R. E.; Lane, C. S.; Elliott, V. L.; Park, B. K., Multiple reaction monitoring for quantitative biomarker analysis in proteomics and metabolomics. *J Chromatogr B Analyt Technol Biomed Life Sci* **2009,** 877, (13), 1229-39.

5. Trivedi, D. K.; Hollywood, K. A.; Goodacre, R., Metabolomics for the masses: The future of metabolomics in a personalized world. *New Horizons in Translational Medicine* **2017,** 3, (6), 294-305.

6. Huang, X.; Zeng, J.; Zhou, L. N.; Hu, C. X.; Yin, P. Y.; Lin, X. H., A New Strategy for Analyzing Time-Series Data Using Dynamic Networks: Identifying Prospective Biomarkers of Hepatocellular Carcinoma. *Scientific Reports* **2016,** 6.

7. Guo, H. M.; Chen, J. Q.; Huang, Y.; Zhang, W.; Xu, F. G.; Zhang, Z. J., A pseudo-kinetics approach for time-series metabolomics investigations: more reliable and sensitive biomarkers revealed in vincristine-induced paralytic ileus rats. *Rsc Advances* **2016,** 6, (59), 54471-54478.

8. Do, R.; Bartlett, K. H.; Chu, W.; Dimich-Ward, H.; Kennedy, S. M., Within- and between-person variability of exhaled breath condensate pH and NH4+ in never and current smokers. *Respiratory Medicine* **2008,** 102, (3), 457-463.

9. Li, A.; Zi, Y.; Guo, H.; Wang, Z. L.; Fernandez, F. M., Triboelectric nanogenerators for sensitive nano-coulomb molecular mass spectrometry. *Nat Nanotechnol* **2017,** 12, (5), 481-487.

10. Zhang, X.; Romm, M.; Zheng, X.; Zink, E. M.; Kim, Y. M.; Burnum-Johnson, K. E.; Orton, D. J.; Apffel, A.; Ibrahim, Y. M.; Monroe, M. E.; Moore, R. J.; Smith, J. N.; Ma, J.; Renslow, R. S.; Thomas, D. G.; Blackwell, A. E.; Swinford, G.; Sausen, J.; Kurulugama, R. T.; Eno, N.; Darland, E.; Stafford, G.; Fjeldsted, J.; Metz, T. O.; Teeguarden, J. G.; Smith, R. D.; Baker, E. S., SPE-IMS-MS: An automated platform for sub-sixty second surveillance of endogenous metabolites and xenobiotics in biofluids. *Clin Mass Spectrom* **2016,** 2, 1-10.

11. Nielsen, J. J., M. C., Metabolomics: A Powerful Tool in Systems Biology: Topics in Current Genetics. In Springer: Heidelberg, Germany, 2007; Vol. 18, p 284.

12. Djekic, D.; Pinto, R.; Vorkas, P. A.; Henein, M. Y., Replication of LC-MS untargeted lipidomics results in patients with calcific coronary disease: An interlaboratory reproducibility study. *Int J Cardiol* **2016,** 222, 1042-1048.

13. Collino, S.; Martin, F. P.; Rezzi, S., Clinical metabolomics paves the way towards future healthcare strategies. *Br J Clin Pharmacol* **2013,** 75, (3), 619-29.

APPENDIX A: LIQUID CHROMATOGRAPHY-MASS SPECTROMETRY BASED CELL CULTURE METABOLOMICS TO UNDERSTAND MECHANISMS OF CANCER CELL DEATH INDUCED BY GOLD NANOROD PHOTOTHERMAL THERAPY

Adapted with permission from
Ali MR, Wu Y, Han T, Zang X, Xiao H, Tang Y, Wu R, Fernández FM, El-Sayed MA. Simultaneous Time-Dependent Surface-Enhanced Raman Spectroscopy, Metabolomics, and Proteomics Reveal Cancer Cell Death Mechanisms Associated with Gold Nanorod Photothermal Therapy. *Journal of the American Chemical Society*. **2016**, *138*, 15434–15442.

M. R. Ali and Y. Wu synthesized AuNRs, cultured the cells and performed PPTT and SERS experiments. X. Zang performed metabolomics experiments. H. Xiao performed proteomics experiments. T. Han conducted bioinformatics analysis.

A.1 Abstract

In cancer plasmonic photothermal therapy (PPTT), plasmonic nanoparticles are used to convert light into localized heat leading to cancer cell death. Among plasmonic nanoparticles, gold nanorods (AuNRs) with specific dimensions that allow them to absorb the near-infrared (NIR) laser light have been widely used. However, the detailed mechanism of PPTT therapy still remains elusive. Typically, surface enhanced Raman spectroscopy (SERS) has been used to detect time-dependent changes in the intensity of the vibration frequencies of molecules that appear or disappear during different cellular processes. A complete proven assignment of the molecular identity of these vibrations and their biological importance have not yet been achieved. In this work, we comple-

mented the study of changes in SERS spectra with MS-based metabolomics and proteomics to identify the chemical species responsible for the observed changes in SERS band intensities. Using PPTT, the bands at around 1000, 1207 and 1580 cm^{-1} were observed to increase in intensity, which were assigned in the literature to phenylalanine, although with dispute. Our metabolomics results showed increased levels of phenylalanine and metabolites tentatively identified as its derivatives and phenylalanine-containing peptides, providing evidence for more confidence in SERS peak assignments. To better understand the mechanism of phenylalanine increase upon PPTT, we combined metabolomics and proteomics results using network analysis, which demonstrated that phenylalanine metabolism was perturbed. Furthermore, several apoptosis pathways were activated *via* key proteins (e.g. HADHA and ACAT1), which are consistent with the proposed role of altered phenylalanine metabolism in inducing apoptosis. This study shows that the integration of the SERS with MS-based metabolomics and proteomics can aid the assignment of signals in SERS spectra and further characterize the related molecular mechanisms of the cellular processes involved in PPTT.

A.2 Introduction

Plasmonic nanoparticles offer a powerful means to follow dynamic changes associated with intracellular molecular events in real-time.[1-3] Their localized surface plasmon resonance (LSPR) confers these particles unique optical properties. For example, the electromagnetic fields on the surface of plasmonic nanoparticles are greatly increased, and exhibit exponential decay patterns following non-radiative (heat) or radiative (e. g. light scattering) processes.[4, 5] Raman scattering from the molecules localized near the

plasmonic nanoparticles' surface is therefore enhanced by orders of magnitude, resulting in the well-known surface enhanced resonance (SERS) phenomenon.[6, 7] SERS has been successfully applied to single-cell analysis, where plasmonic gold nanoparticles are placed inside the cell and the resulting SERS spectrum collected in order to record the intracellular microenvironment changes occurring in time near the nanoparticles. Molecular event dynamics during cancer cell apoptosis has been observed in real-time by using SERS.[5, 8] However, the molecular species associated with the observed SERS bands could not be confidently assigned, preventing from elucidating the molecular mechanisms involved in these critical cellular processes.[4, 6-8]

Photothermal therapy has its foundation in the targeted destruction of cancerous cells *via* the heat released by gold nanorods (AuNRs) following near-infrared (NIR) radiation absorption. The so called "water wavelength window" between 700–1200 nm is widely considered to be the optimal spectral region for conducting plasmonic photothermal therapy (PPTT),[9, 10] as tissue and water absorption are minimized in this range. AuNRs, on the other hand, readily absorb near-infrared laser light, resulting in effective photothermal generators, for both *in vitro* and *in vivo* applications. AuNRs-based PPTT has been successful at inducing cancer cell apoptosis,[11] resulting in *in vivo* tumor removal.[12-14]

Despite the operational success of PPTT, the molecular mechanisms associated with PPTT-induced apoptosis remain largely unknown or under dispute. Mostafa El-Sayed's group observed PPTT induce apoptosis initiate through heat shock proteins previously,[15] while several reports indicate it is mediated by the mitochondrial apoptotic pathway *via* Bid activation and caspase 3 activity.[16, 17] Although SERS reports on the

real-time biomolecular dynamics in the microenvironment associated with the PPTT process, SERS spectra from cells are incredibly complex, reflecting overlapping signals from a variety of proteins and metabolites that are difficult to assign to individual species. It has been reported, for example, that the 1000 and 1580 cm^{-1} peaks showed significant increase during cell apoptosis,[8, 18] with great debate in their assignment.[19] According to one published literature, the 1000 cm^{-1} signal has been assigned to phenylalanine,[20] while a different report assigned it to tryptophan.[19] Furthermore, it has been argued[21] that these SERS signals actually reflect changes in protein structure, a topic still being intensively debated. One hypothesis states that the 1000 cm^{-1} signal is indicative of the exposure of protein hydrophobic rings following conformational changes,[21] while others report that the protein conformation change induced by adding methanol or SDS[22] or increasing temperature[23] does not alter the intensity of the 1000 cm^{-1} peak, contradicting the hypothesis that this signal is associated with alterations in protein conformation.

In this study, we monitored the SERS spectral signature *in vitro* during apoptosis as a function of PPTT exposure time. We also performed metabolomics and proteomic studies on cell lysates under the same exact PPTT conditions. Integrative multi-omics network analysis revealed specific alterations that explain the underlying changes in SERS spectral data, demonstrating the power of combining SERS with MS for studying cellular processes following PPTT.

A.3 Materials and Methods

A.3.1 Sample Preparation for Metabolomics Experiments

Cells were cultured in 60 mm petri dishes. The culture media was removed and cells were washed 3 times with PBS, followed by a wash with deionized water for 2 s and immediate removal of the wash solution. Seven mL of metabolite extraction solvents (HPLC grade methanol: acetonitrile: 0.5 M formic acid, 2:2:1 v/v, -20 ^{o}C) was added immediately for quenching and lysing the cells.[24] Cells were then scraped down, and the cell suspension was transferred to centrifuge tubes, followed by vortexing and sonication in ice-water bath and incubation on ice for 15 min for metabolite extraction. The cell suspension was then centrifuged at 20,400 × *g* at 4 ^{o}C for 15 min. Solvent in the sample was evaporated using a CentriVap Vacuum Concentrator until dryness. The dried samples were kept at -80 °C until analysis.[24]

A.3.2 Sample Preparation for Proteomics Experiments

Cells were cultured in 60 mm Petri dishes. Ice-cold lysis buffer (50 mM HEPES, pH 7.8, 150 mM NaCl 0.1% SDS (optional) 0.5% sodium deoxycholate 1% Triton X 100 or NP-40, phosphatase inhibitors) was added directly to the cells after washing with PBS twice. The cells were subsequently scraped down and the obtained mixtures homogenized with sonication and vortexing. Cell debris was then removed by centrifugation at 18,000 × *g* for 20 min at 4 °C. Four volumes of ice-cold acetone/ethanol/acetic acid (v/v/v = 50/50/0.1) was added to the supernatant to precipitate the proteins at -20 °C overnight. After centrifugation, the protein pellet was re-dissolved in denaturing buffer (pH 8.0) containing 8 M urea and 50 mM HEPES, and the protein concentration was tested using a

Bradford assay. The disulfide bonds in the protein solution were reduced by 2 mM dithiothreitol (DTT) at 37 °C for 2 h and subsequently alkylated by adding 6 mM iodoacetamide (IAA) and kept in darkness at room temperature for 40 min.[25]

A.3.3 UPLC-MS Based Metabolomics Analysis

Before analysis, ultrapure water was added to each dried sample to obtain a final biomass concentration of ~50,000 cells μL^{-1}. Samples were further vortexed and then centrifuged at 15,000 rpm for 10 min at 4 °C. The supernatant of each biological sample was transferred to autosampler vials for UPLC-MS analysis using a Waters ACQUITY UPLC H Class system fitted with a Waters ACQUITY UPLC BEH C18 column (2.1×50 mm, 1.7 μm particle size, Waters Corporation, Milford, MA, USA), coupled to a Xevo G2 QTOF mass spectrometer (Waters Corporation, Manchester, UK) with an ESI source. Gradient elution was employed in the chromatographic separation method using 0.1% acetic acid in water (mobile phase A) and acetonitrile (mobile phase B), with the following program: 0-1 min, 98% A, 1-3 min 98%-70% A, 3-8 min 70%-50% A, 8-10 min 50%-5% A, 10-15 min 5% A. The flow rate was constant at 0.3 mL min^{-1}. After each sample run, the column was re-equilibrated to the initial conditions in 6 min. The injection volume was 5 μL. The column and auto sampler tray temperatures were set at 35 and 5 °C, respectively. The mass spectrometer was operated in negative ion polarity and resolution mode with a probe capillary voltage of 2.2 kV and a sampling cone voltage of 45.0 V. The source and desolvation gas temperatures were set to 120 and 350 °C, respectively.

The nitrogen gas desolvation flow rate was 650 Lh^{-1}. The mass spectrometer was calibrated across the range of *m/z* 50-1200 using a 0.5 mM sodium formate solution prepared in 2-propanol/water (90:10 v/v). Data were drift corrected during acquisition using a leucine encephalin (*m/z* 554.2615) reference spray infused at 3 μL min^{-1}. Data were acquired in the range of *m/z* 50-1200, and the scan time was set to 1 s. Technical duplicates were acquired in all cases. Tandem MS experiments were performed using a Waters ACQUITY UPLC I Class system fitted with the same column used for UPLC-MS analysis, coupled to a Synapt G2-S HDMS system (Waters Corporation, Manchester, UK) equipped with an ESI source operated in negative polarity and resolution mode. For the fast data dependent acquisition method, targeted ions for MS/MS were entered in an include list. A 0.2 s continuum MS survey scan was collected from 50 to 650 Da until the intensity of an individual precursor ion raised above 5000, then switched to MS/MS acquisition, in which a 0.1s continuum scan was collected from 30 to 650 Da. The MS/MS scan switched off once the accumulated total ion current reached 100,000 or after 0.25 s. A collision energy profile of 15, 25 and 35 V was applied to the trap cell for ion fragmentation. For the MS/MS method, a scan time was 1s and collision voltages between 8 to 30 V were applied to the trap cell. Data acquisition and processing were performed with Masslynx ver. 4.1 (Waters Corp., Milford, MA, USA).

A.3.4 LC-MS/MS Analysis for Proteomic Experiments

Purified and dried peptide samples were dissolved in a 10 μL solution containing 5% ACN and 4% formic acid (FA), and 3 μL was loaded onto a micro capillary column

packed with C18 beads (Magic C18AQ, 3 μm, 200 Å, 100 μm x 16 cm, Michrom Bioresources) by a Dionex WPS-3000T PLUS auto sampler (UltiMate 3000 thermostatted Rapid Separation Pulled Loop Well Plate Sampler). Peptides were separated by reverse-phase chromatography using an UltraMate 3000 binary pump with a 110 min gradient of 8-38% ACN (with 0.125% FA) for the triplicates. Peptides were detected with a data-dependent Top 20 method (the 20 most abundant ions were selected for MS^2),[26] in a hybrid dual-cell quadrupole linear ion trap – Orbitrap mass spectrometer (LTQ Orbitrap Elite, Thermo Fisher, with Xcalibur 3.0.63 software). For each cycle, each full MS scan (resolution: 60,000) in the Orbitrap at 10^6 AGC target was followed by up to 20 MS/MS for the most intense ions in the LTQ. The selected ions were excluded from further analysis for 90 seconds. Ions with singly or unassigned charge were not sequenced. For each full MS scan, the maximum ion accumulation time was 1000 ms and the one for MS/MS scans was 50 ms. Mass spectra Raw files were converted into mzXML format, then searched using the SEQUEST algorithm (version 28).[27] Spectra were matched against a database containing sequences of all proteins in the UniProt Human (Homo sapiens) database (downloaded in February 2014). The search was performed using following parameters: fully digested with trypsin; up to 3 missed cleavages; fixed modifications: carbamidomethylation of cysteine (+57.0214); variable modifications: oxidation of methionine (+15.9949). False discovery rates (FDR) of peptide and protein identifications were controlled by the target-decoy method.[28, 29] Linear discriminant analysis (LDA) was used to control the quality of peptide identifications using parameters such as Xcorr, precursor mass error, and charge state.[30, 31] Peptides fewer than seven amino acid residues in length were deleted. Furthermore, peptide spectral matches were filtered to <1% FDR.

A.3.5 Data Analysis

For metabolomics, spectral features (retention time (Rt), *m/z* pairs) were extracted from UPLC-MS data using Progenesis QI version 2.0 (Nonlinear Dynamics, Waters Corp.). The data preprocessing procedures included retention time alignment, peak picking, integration, and de-convolution to group the adducts derived from the same compound.

Raw data from metabolomics were normalized using supervised normalization of microarray (SNM).[32] In the SNM procedure, variances due to biological and technical replicates were adjusted by setting them as variables in the model. Variance explained by different experimental treatments (control, AuNRs functionalized with nuclear localization signal (AuNRs@NLS), and AuNRs@NLS/PPTT) was fitted as a biological variable in the model. Metabolomics data were log2 transformed before analysis of variance (ANOVA) which was used to detect differential levels of metabolites between control and treatment groups. We fitted models with treatment conditions as fixed effects. A Benjamini-Hochberg 5% false discovery rate (FDR) correction was used to select differential metabolites.[33] For identified differential metabolites perturbed by PPTT, we used the Mummichog program for network-level metabolites annotation.[34] The MS mode considered in Mummichog was negative ion in order to compute isotopic and adduct species. The metabolites identified as being affected by PPTT were subjected to pathway analysis using the MetaCore pathway analysis software (“MetaCore from Thomson Reuters”).

For proteomics, raw data were also normalized using SNM. The identified proteins were subjected to pathway analysis using the MetaCore pathway analysis software ("MetaCore from Thomson Reuters") to study the effect of PPTT.

A.4 Results and Discussion

A.4.1 Metabolomics Results

Metabolic profiling of cell extracts was performed using LC-MS, with two biological replicates and two technical replicates. A total of 1122 spectral features (retention time, *m/z* pairs) were detected in metabolite extracts. Specifically, many of these features corresponded to an increase in the relative amount of phenylalanine (Figure A.1a) and related species after PPTT (Figure A.1b-A.1d, Figure A.2). Tentatively identified phenylalanine derivatives and phenylalanine-containing short peptides, such as glutamylphenylalanine (Figure A.1b), asparaginyl-phenylalanine (Figure A.1c), histidinyl-phenylalanine (Figure A.1d) were amongst those altered, explaining the trends observed in the SERS data. Tandem MS experiments confirmed the identity of the phenylalanine detected by MS with excellent accuracy (Figure A.1e).

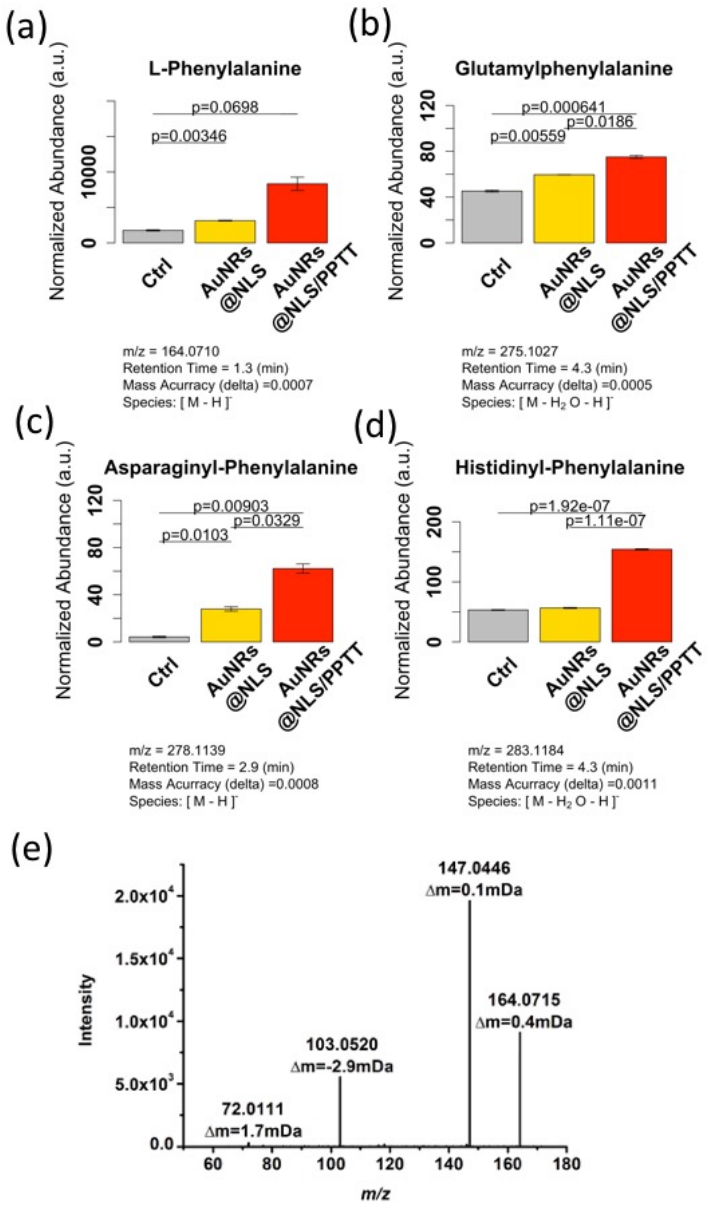

Figure A.1: Metabolite perturbations observed in HSC-3 cells treated with AuNRs-PPTT (NLS conjugated particles). (a-d) Bar graphs showing the normalized abundance of phenylalanine-related metabolites altered following PPTT. Normalized abundances of metabolites following AuNRs@NLS without PPTT are also given for comparison. (a) L-phenylalanine. The result was confirmed by MS/MS (shown in e). (b) Glutamylphenylalanine. (c) Asparaginyl-phenylalanine. (d) Histidinyl-phenylalanine. (e) Product ion spectrum obtained under data dependent acquisition (DDA) conditions for the precursor ion at *m/z* 164.0710. Matching of this mass spectrum to the Metlin database MS/MS reference spectrum of phenylalanine (10 V collision energy) is shown with mass accuracies indicated for each ionic species detected.

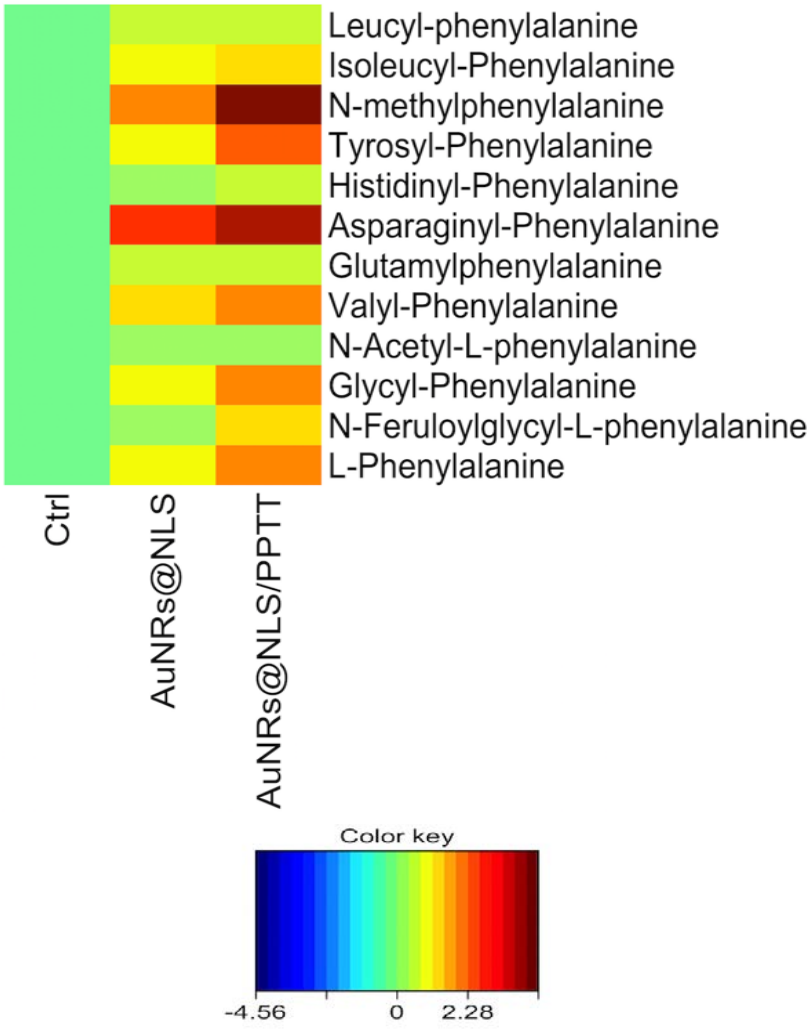

Figure A.2: Heat map showing fold change (log_2) of key metabolites related to phenylalanine metabolism in treatment experiments (AuNRs@NLS, AuNRs@NLS/PPTT) compared to control group.

A.4.2 Proteomics Results

Furthermore, we also conducted a label-free quantitative proteomics experiment for studying alterations in protein abundances and seeking possible evidence for, and understanding of the mechanisms responsible for the phenylalanine concentration increase. A test experiment was done to measure the accuracy of our proteomics workflow using the reported method,[25] where 99% of the proteins have shown accurate quantification (Figure A.3). In our proteomics experiment, two biological replicates and three technical replicates were conducted. In total, 1341 proteins were identified.

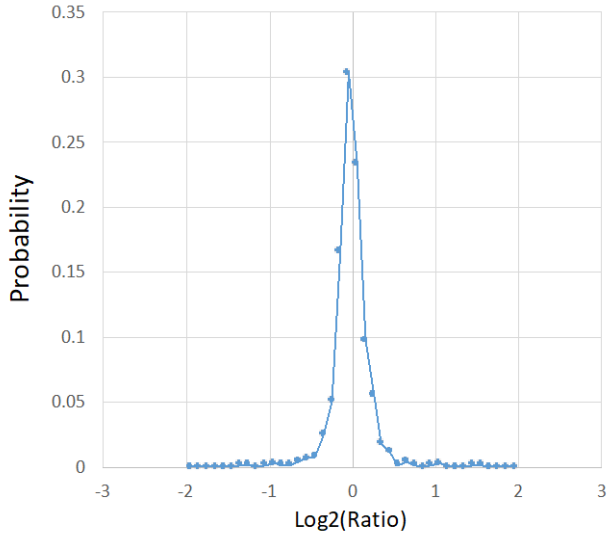

Figure A.3: Quantification accuracy examination of proteomics workflow: Log_2 ratio distributions of quantified peptides from 2 identical test samples (yeast whole proteome sample), each sample having 3 technical replicates.

A.4.3 Integrative Analysis of Metabolomics and Proteomics Results

We conducted an integrative analysis of metabolomics and proteomics results to provide a more holistic understanding of the biological processes involved. Integrated pathway analysis showed that the phenylalanine metabolism pathway was significantly perturbed by PPTT (Figure A.4 and A.5). Approximately half of the metabolites in the phenylalanine metabolism pathway were identified as changed, including 2-phenyl-acetamide (increase), phenylpiruvate (decrease), 2-hydroxy-3-phenyl-acrylic acid (decrease), 3-oxobutanoate (increase), fumarate (increase), phenylacetaldehyde (increase), L-tyrosine (decrease), 4-fumarylacetoacetate (decrease) and 4-maleylacetoacetate (decrease). These alterations were accompanied by perturbations in several key proteins in the phenylalanine metabolism pathway.

Phenylalanine is known to induce apoptosis,[35-37] which is consistent with the apoptotic phenotype observed in PPTT-treated cells. An overview of the pathways identified to be associated with phenylalanine-induced apoptosis is schematically shown in Figure A.6a. Two proteins in the phenylalanine metabolism pathway have been previously associated with apoptosis. Mitochondrial acetyl-CoA acetyltransferase (ACAT1) has been shown to be involved in the development of doxorubicin resistance to decrease cell apoptosis.[38] Another mitochondrial protein, hydroxyl-coenzyme A dehydrogenase/3-ketoacyl-coenzyme A thiolase/enoyl-coenzyme (HADHA), has been shown to prevent chemically-induced apoptosis in cancer treatment.[39, 40] In our experiments, both proteins were observed to be down-regulated following PPTT treatment, suggesting that the anti-apoptotic protection was turned off resulting in enhanced vulnerability to apoptosis (Figure A.6b and c).

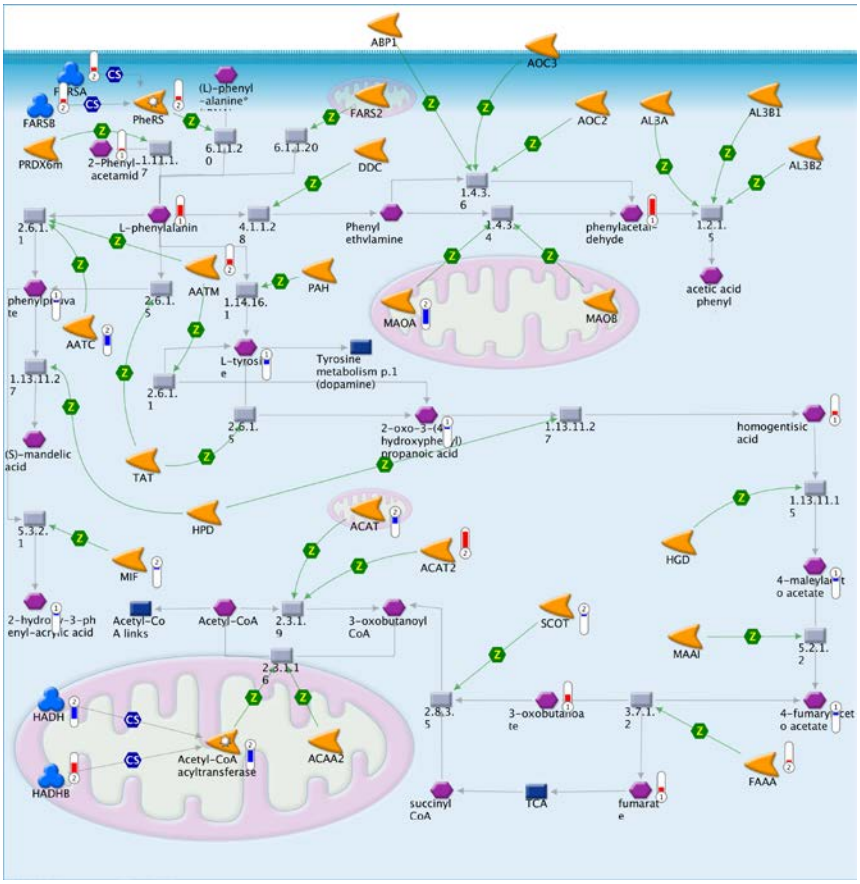

Figure A.4: Pathway map showing that the phenylalanine metabolism pathway was perturbed after PPTT and key proteins (HADHA, ACAT1) were down-regulated, which triggers apoptosis. (Red) means up-regulation after PPTT, (blue) means down-regulation after PPTT. In the thermometer sign, 1 refers to metabolomics results, 2 refers to proteomics results. The thermometers are filled to various degrees, corresponding to the amount by which the markers were up-regulated or down-regulated.

Note: in the figure, "Acetyl-CoA acyltransferase" represents "hydroxyacyl-CoA dehydrogenase/3-ketoacyl-CoA thiolase/enoyl-CoA hydratase (trifunctional protein), alpha subunit", which is a protein complex catalyzing the 3-hydroxyacyl-CoA dehydrogenase and enoyl-CoA hydratase activities. HADHA and HADHB are both subunits of this protein complex. HADHA is down-regulated, while the HADHB is up-regulated. However, the down-regulation of HADHA is the rate limiting step forming the effective protein complex. Therefore, the activity of protein complex Acetyl-CoA acyltransferase is down-regulated, contributing to the mitochondria mediated apoptosis process.

ACAT1 is acetyl-CoA acetyltransferase 1, which is a mitochondrially localized enzyme that catalyzes the reversible formation of acetoacetyl-CoA from two molecules of acetyl-CoA. ACAT2 is cytosolic localized acetyl-CoA acetyltransferase 2, which involved in lipid metabolism. The mitochondrial isoform ACAT1 is down-regulated hints its perturbation contributes to mitochondria mediated apoptosis processes, while the cytosolic isoform ACAT2 is not related to this process.

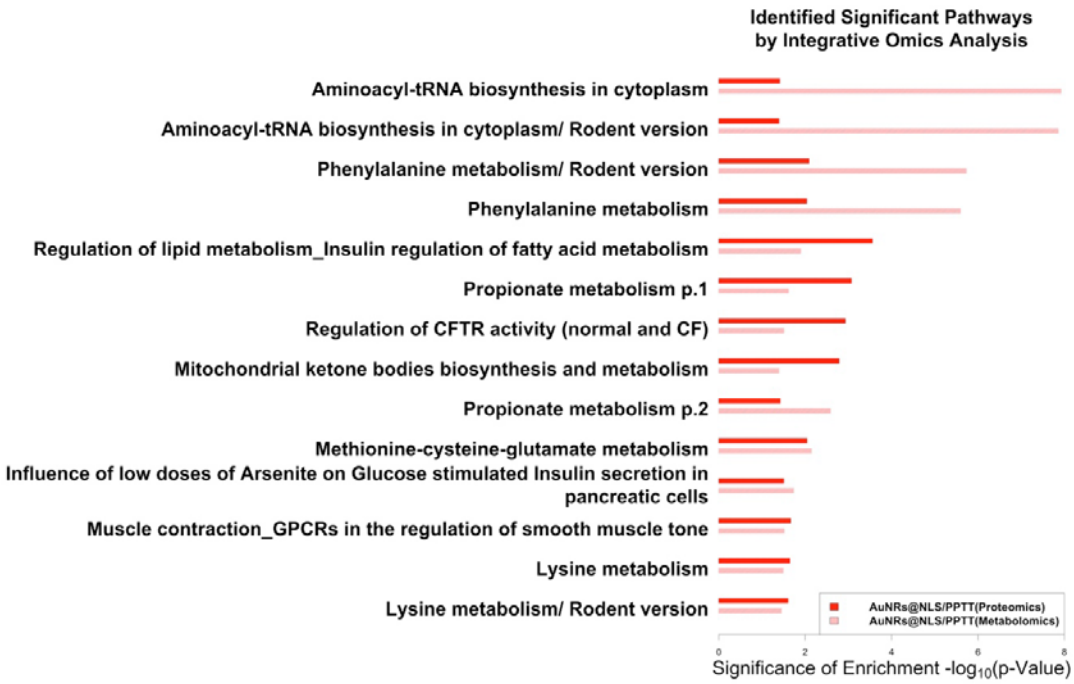

Figure A.5: Significant pathways identified from proteomics (red bars) and metabolomics (light pink bars) that perturbed by photothermal therapy.

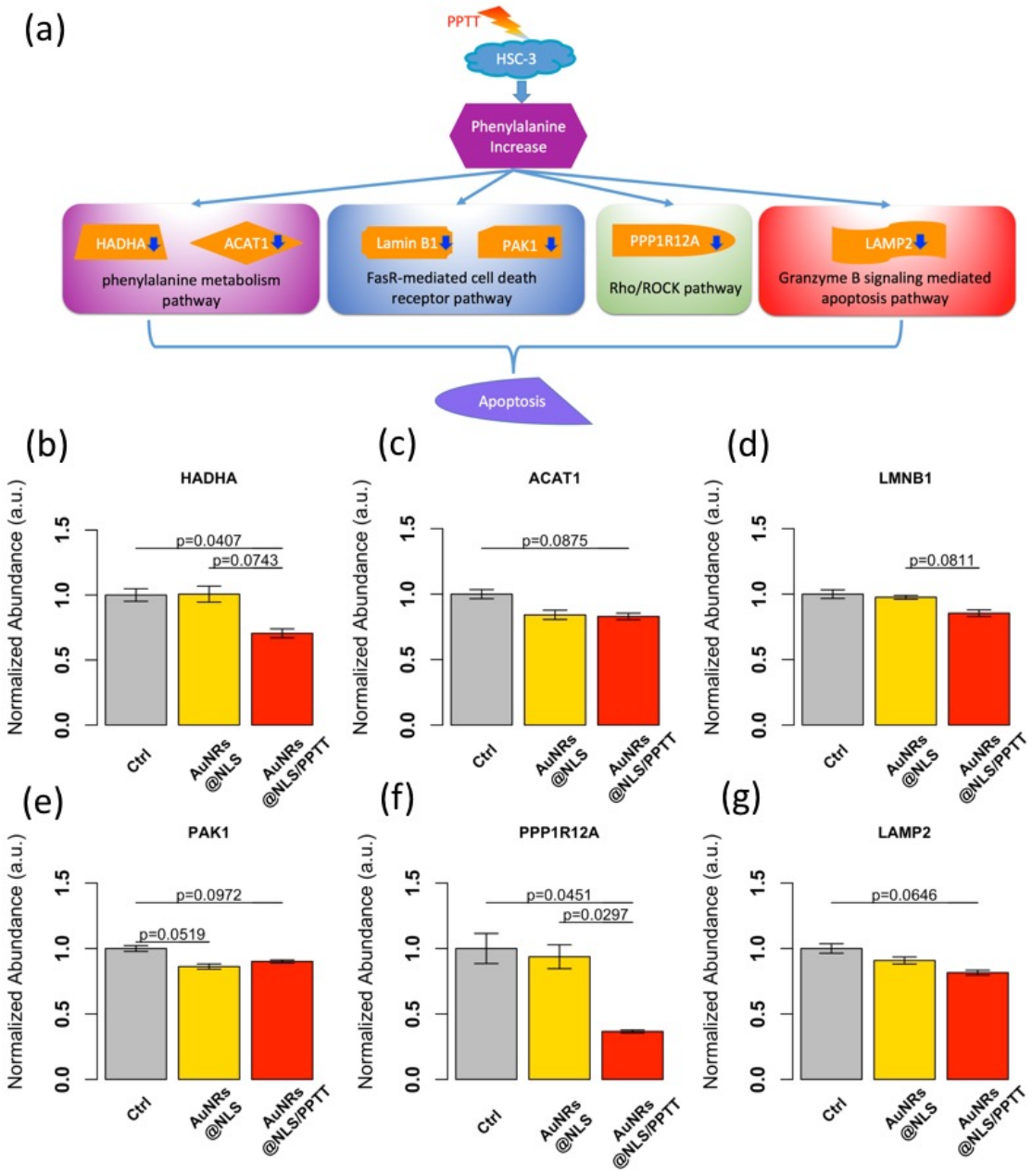

Figure A.6: (a) Schematic diagram explaining the molecular apoptosis mechanisms involved in altering phenylalanine metabolism as induced by PPTT. (b-g) Bar graphs showing the normalized abundance of key proteins contributing to apoptosis involved in altering phenylalanine metabolism following PPTT. Normalized abundances of key proteins following AuNRs@NLS without PPTT are also given for comparison. (b) HADHA. (c) ACAT1. (d) Lamin B1 (LMNB1). (e) PAK1. (f) PPP1R12A. (g) LAMP2.

In addition to the phenylalanine metabolism pathway, three possible mechanisms of phenylalanine-induced apoptosis were suggested by the results. Firstly, increased phenylalanine was shown to induce apoptosis by involvement of the Fas receptor (FasR)-mediated cell death receptor pathway.[35] In this study, two proteins (Lamin B1 and PAK1) in the Fas/Fas ligand death receptor pathway were identified. These two proteins have been previously demonstrated to be associated with apoptosis. Lamin B1, as the major component of the nuclear lamina underlying the nuclear membrane, plays an important role in maintaining nuclear membrane integrity. Destruction of nuclear membrane integrity being a hallmark of apoptosis. During apoptosis, Lamin B1 mRNA level have been shown to decrease,[41] which could result from induction of either p53 or pRB tumor suppressor pathways.[42, 43] Literature results also show that the Fas/Fas ligand complex downstream effector PAK1 is required to prevent apoptosis by limiting the expression of pro-apoptotic proteins or modulating post-translational modifications on effectors.[44-46] In this study, both of these proteins were down-regulated, suggesting a cellular shift towards apoptosis, and reduced anti-apoptotic protection (Figure A.6d and e).

Phenylalanine has also been shown to activate mitochondria-mediated apoptosis through the Rho/ROCK pathway.[47, 48] In this study, we identified down-regulation of the myosin phosphatase targeting subunit 1 (PPP1R12A) in PPTT-treated cells, this being a downstream effector of ROCK (Figure A.6f) that would contribute to the apoptotic phenotype following PPTT. In apoptotic cells, PPP1R12A is cleaved, with the cleaved PPP1R12A inhibiting myosin II binding, which results in membrane blebbing and apoptosis.[49]

A third mechanism of phenylalanine-induced cell death involves a component in Granzyme B signaling-mediated apoptosis, known as lysosome-associated membrane protein 2 (LAMP2). This protein was down-regulated following PPTT treatment (Figure A.6g). LAMP2 is critical to maintain lysosome integrity and normal cellular function, and lower levels of LAMP proteins have been positively associated with apoptosis.[45] It is not yet conclusively established, however, whether decreased LAMP2 levels are also directly associated with phenylalanine-induced apoptosis.

Further investigations into the mechanism of how PPTT treatment increases phenylalanine levels in cells focused on the fact that phenylalanine can be converted to L-tyrosine.[50-52] Metabolomics data indicated that after PPTT treatment, the level of L-tyrosine was actually decreased (Figure A.4). Based on our results, the channel allowing for the conversion from phenylalanine to L-tyrosine could have contributed to the accumulation of phenylalanine, which further induced mitochondria-mediated apoptosis. Besides phenylalanine dependent process, we further identified significantly perturbed pathways by integrative analysis of proteomics and metabolomics (Figure A.5). Other amino acids metabolism pathways are enriched, including methionine-cysteine-glutamate, and lysine metabolism, which are very essential for cell survival. Interestingly, we also found clues on the perturbation of pathways related to lipid metabolism and ketone body metabolism.[15, 53, 54]

A.5 Conclusion

In this work, we combined results from SERS, metabolomics and proteomics experiments, aiming to study the change of the subcellular microenvironment around

AuNRs during the PPTT process. The SERS data showed that the 1000, 1207 and 1580 cm^{-1} bands increased during PPTT, which suggested an increase of phenylalanine and its derivatives. These findings were confirmed by integrative analysis of Raman spectroscopy, metabolomics and proteomics MS data, showing that free phenylalanine and tentatively identified phenylalanine associated metabolites are significantly perturbed by PPTT, leading to cell apoptosis. We therefore propose that phenylalanine measurements by SERS can be developed as a sensitive and convenient readout for non-invasive direct apoptosis characterization.

A.6 References

1. Qian, X.; Peng, X. H.; Ansari, D. O.; Yin-Goen, Q.; Chen, G. Z.; Shin, D. M.; Yang, L.; Young, A. N.; Wang, M. D.; Nie, S., In vivo tumor targeting and spectroscopic detection with surface-enhanced Raman nanoparticle tags. *Nat Biotechnol* **2008,** 26, (1), 83-90.

2. Kneipp, J.; Kneipp, H.; Kneipp, K., SERS--a single-molecule and nanoscale tool for bioanalytics. *Chem Soc Rev* **2008,** 37, (5), 1052-60.

3. Ma, C.; Harris, J. M., Surface-enhanced Raman spectroscopy detection of ionic solutes by surfactant-mediated adsorption to a hydrophobic surface. *Appl Spectrosc* **2013,** 67, (7), 801-7.

4. Jain, P. K.; Huang, X.; El-Sayed, I. H.; El-Sayed, M. A., Noble metals on the nanoscale: optical and photothermal properties and some applications in imaging, sensing, biology, and medicine. *Acc Chem Res* **2008,** 41, (12), 1578-86.

5. Austin, L. A.; Kang, B.; El-Sayed, M. A., Probing molecular cell event dynamics at the single-cell level with targeted plasmonic gold nanoparticles: A review. *Nano Today* **2015,** 10, (5), 542-558.

6. Willets, K. A.; Van Duyne, R. P., Localized surface plasmon resonance spectroscopy and sensing. *Annual Review of Physical Chemistry* **2007,** 58, 267-297.

7. Stiles, P. L.; Dieringer, J. A.; Shah, N. C.; Van Duyne, R. P., Surface-enhanced Raman spectroscopy. *Annu Rev Anal Chem (Palo Alto Calif)* **2008,** 1, 601-26.

8. Kang, B.; Austin, L. A.; El-Sayed, M. A., Observing Real-Time Molecular Event Dynamics of Apoptosis in Living Cancer Cells using Nuclear-Targeted Plasmonically Enhanced Raman Nanoprobes. *Acs Nano* **2014,** 8, (5), 4883-4892.

9. Huang, X.; Jain, P. K.; El-Sayed, I. H.; El-Sayed, M. A., Plasmonic photothermal therapy (PPTT) using gold nanoparticles. *Lasers Med Sci* **2008,** 23, (3), 217-28.

10. Alkilany, A. M.; Thompson, L. B.; Boulos, S. P.; Sisco, P. N.; Murphy, C. J., Gold nanorods: their potential for photothermal therapeutics and drug delivery, tempered

by the complexity of their biological interactions. *Adv Drug Deliv Rev* **2012,** 64, (2), 190-9.

11. Huang, X.; El-Sayed, I. H.; Qian, W.; El-Sayed, M. A., Cancer cell imaging and photothermal therapy in the near-infrared region by using gold nanorods. *J Am Chem Soc* **2006,** 128, (6), 2115-20.

12. Li, Z.; Huang, P.; Zhang, X.; Lin, J.; Yang, S.; Liu, B.; Gao, F.; Xi, P.; Ren, Q.; Cui, D., RGD-conjugated dendrimer-modified gold nanorods for in vivo tumor targeting and photothermal therapy. *Mol Pharm* **2010,** 7, (1), 94-104.

13. Niidome, T.; Akiyama, Y.; Yamagata, M.; Kawano, T.; Mori, T.; Niidome, Y.; Katayama, Y., Poly(ethylene glycol)-modified gold nanorods as a photothermal nanodevice for hyperthermia. *J Biomater Sci Polym Ed* **2009,** 20, (9), 1203-15.

14. Dickerson, E. B.; Dreaden, E. C.; Huang, X.; El-Sayed, I. H.; Chu, H.; Pushpanketh, S.; McDonald, J. F.; El-Sayed, M. A., Gold nanorod assisted near-infrared plasmonic photothermal therapy (PPTT) of squamous cell carcinoma in mice. *Cancer Lett* **2008,** 269, (1), 57-66.

15. Mukherjee, P.; El-Abbadi, M. M.; Kasperzyk, J. L.; Ranes, M. K.; Seyfried, T. N., Dietary restriction reduces angiogenesis and growth in an orthotopic mouse brain tumour model. *Br J Cancer* **2002,** 86, (10), 1615-21.

16. Melamed, J. R.; Edelstein, R. S.; Day, E. S., Elucidating the fundamental mechanisms of cell death triggered by photothermal therapy. *ACS Nano* **2015,** 9, (1), 6-11.

17. Perez-Hernandez, M.; Del Pino, P.; Mitchell, S. G.; Moros, M.; Stepien, G.; Pelaz, B.; Parak, W. J.; Galvez, E. M.; Pardo, J.; de la Fuente, J. M., Dissecting the molecular mechanism of apoptosis during photothermal therapy using gold nanoprisms. *ACS Nano* **2015,** 9, (1), 52-61.

18. Narayanan, N.; Nair, L. V.; Karunakaran, V.; Joseph, M. M.; Nair, J. B.; N, R. A.; Jayasree, R. S.; Maiti, K. K., Investigation of apoptotic events at molecular level induced by SERS guided targeted theranostic nanoprobe. *Nanoscale* **2016,** 8, (22), 11392-7.

19. Aliaga, A. E.; Osorio-Roman, I.; Leyton, P.; Garrido, C.; Carcamo, J.; Caniulef, C.; Celis, F.; Diaz, G.; Clavijo, E.; Gomez-Jeria, J. S.; Campos-Vallette, M. M., Surface-

enhanced Raman scattering study of L-tryptophan. *Journal of Raman Spectroscopy* **2009,** 40, (2), 164-169.

20. Zhu, G.; Zhu, X.; Fan, Q.; Wan, X., Raman spectra of amino acids and their aqueous solutions. *Spectrochim Acta A Mol Biomol Spectrosc* **2011,** 78, (3), 1187-95.

21. Kang, B.; Austin, L. A.; El-Sayed, M. A., Observing real-time molecular event dynamics of apoptosis in living cancer cells using nuclear-targeted plasmonically enhanced Raman nanoprobes. *ACS Nano* **2014,** 8, (5), 4883-92.

22. Maiti, N. C.; Apetri, M. M.; Zagorski, M. G.; Carey, P. R.; Anderson, V. E., Raman spectroscopic characterization of secondary structure in natively unfolded proteins: alpha-synuclein. *J Am Chem Soc* **2004,** 126, (8), 2399-408.

23. Keskin, S.; Efeoglu, E.; Kececi, K.; Culha, M., Label-free detection of proteins in ternary mixtures using surface-enhanced Raman scattering and protein melting profiles. *J Biomed Opt* **2013,** 18, (3), 037007.

24. Martano, G.; Delmotte, N.; Kiefer, P.; Christen, P.; Kentner, D.; Bumann, D.; Vorholt, J. A., Fast sampling method for mammalian cell metabolic analyses using liquid chromatography-mass spectrometry. *Nat Protoc* **2015,** 10, (1), 1-11.

25. Wu, Y.; Wang, F.; Liu, Z.; Qin, H.; Song, C.; Huang, J.; Bian, Y.; Wei, X.; Dong, J.; Zou, H., Five-plex isotope dimethyl labeling for quantitative proteomics. *Chem Commun (Camb)* **2014,** 50, (14), 1708-10.

26. Xiao, H.; Tang, G. X.; Wu, R., Site-Specific Quantification of Surface N-Glycoproteins in Statin-Treated Liver Cells. *Anal Chem* **2016,** 88, (6), 3324-32.

27. Eng, J. K.; McCormack, A. L.; Yates, J. R., An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* **1994,** 5, (11), 976-89.

28. Elias, J. E.; Gygi, S. P., Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* **2007,** 4, (3), 207-14.

29. Peng, J.; Elias, J. E.; Thoreen, C. C.; Licklider, L. J.; Gygi, S. P., Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-

MS/MS) for large-scale protein analysis: the yeast proteome. *J Proteome Res* **2003,** 2, (1), 43-50.

30. Kall, L.; Canterbury, J. D.; Weston, J.; Noble, W. S.; MacCoss, M. J., Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods* **2007,** 4, (11), 923-5.

31. Huttlin, E. L.; Jedrychowski, M. P.; Elias, J. E.; Goswami, T.; Rad, R.; Beausoleil, S. A.; Villen, J.; Haas, W.; Sowa, M. E.; Gygi, S. P., A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* **2010,** 143, (7), 1174-89.

32. Mecham, B. H.; Nelson, P. S.; Storey, J. D., Supervised normalization of microarrays. *Bioinformatics* **2010,** 26, (10), 1308-15.

33. Benjamini, Y.; Hochberg, Y., Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological* **1995,** 57, (1), 289-300.

34. Li, S.; Park, Y.; Duraisingham, S.; Strobel, F. H.; Khan, N.; Soltow, Q. A.; Jones, D. P.; Pulendran, B., Predicting network activity from high throughput metabolomics. *PLoS Comput Biol* **2013,** 9, (7), e1003123.

35. Huang, X.; Lu, Z.; Lv, Z.; Yu, T.; Yang, P.; Shen, Y.; Ding, Y.; Fu, D.; Zhang, X.; Fu, Q.; Yu, Y., The Fas/Fas ligand death receptor pathway contributes to phenylalanine-induced apoptosis in cortical neurons. *PLoS One* **2013,** 8, (8), e71553.

36. Zhang, J. J.; Huang, Q.; Yao, G. H.; Ke, Z. G.; Zhang, H.; Lu, Y. L., SERS study of transformation of phenylalanine to tyrosine under particle irradiation. *Journal of Molecular Structure* **2014,** 1072, 195-202.

37. Kim, S. K.; Kim, M. S.; Suh, S. W., Surface-Enhanced Raman-Scattering (Sers) of Aromatic-Amino-Acids and Their Glycyl Dipeptides in Silver Sol. *Journal of Raman Spectroscopy* **1987,** 18, (3), 171-175.

38. Lo, Y. W.; Lin, S. T.; Chang, S. J.; Chan, C. H.; Lyu, K. W.; Chang, J. F.; May, E. W.; Lin, D. Y.; Chou, H. C.; Chan, H. L., Mitochondrial proteomics with siRNA knockdown to reveal ACAT1 and MDH2 in the development of doxorubicin-resistant uterine cancer. *J Cell Mol Med* **2015,** 19, (4), 744-59.

39. Gonzalez, V. M.; Fuertes, M. A.; Alonso, C.; Perez, J. M., Is cisplatin-induced cell death always produced by apoptosis? *Mol Pharmacol* **2001,** 59, (4), 657-63.

40. Kageyama, T.; Nagashio, R.; Ryuge, S.; Matsumoto, T.; Iyoda, A.; Satoh, Y.; Masuda, N.; Jiang, S. X.; Saegusa, M.; Sato, Y., HADHA is a potential predictor of response to platinum-based chemotherapy for lung cancer. *Asian Pac J Cancer Prev* **2011,** 12, (12), 3457-63.

41. Sato, A.; Hiramoto, A.; Satake, A.; Miyazaki, E.; Naito, T.; Wataya, Y.; Kim, H. S., Association of nuclear membrane protein lamin B1 with necrosis and apoptosis in cell death induced by 5-fluoro-2'-deoxyuridine. *Nucleosides Nucleotides Nucleic Acids* **2008,** 27, (5), 433-8.

42. Freund, A.; Laberge, R. M.; Demaria, M.; Campisi, J., Lamin B1 loss is a senescence-associated biomarker. *Mol Biol Cell* **2012,** 23, (11), 2066-75.

43. Rao, L.; Perez, D.; White, E., Lamin proteolysis facilitates nuclear events during apoptosis. *J Cell Biol* **1996,** 135, (6 Pt 1), 1441-55.

44. Zha, J.; Harada, H.; Yang, E.; Jockel, J.; Korsmeyer, S. J., Serine phosphorylation of death agonist BAD in response to survival factor results in binding to 14-3-3 not BCL-X(L). *Cell* **1996,** 87, (4), 619-28.

45. Schurmann, A.; Mooney, A. F.; Sanders, L. C.; Sells, M. A.; Wang, H. G.; Reed, J. C.; Bokoch, G. M., p21-activated kinase 1 phosphorylates the death agonist bad and protects cells from apoptosis. *Mol Cell Biol* **2000,** 20, (2), 453-61.

46. Jin, S.; Zhuo, Y.; Guo, W.; Field, J., p21-activated Kinase 1 (Pak1)-dependent phosphorylation of Raf-1 regulates its mitochondrial localization, phosphorylation of BAD, and Bcl-2 association. *J Biol Chem* **2005,** 280, (26), 24698-705.

47. Zhang, Y.; Gu, X.; Yuan, X., Phenylalanine activates the mitochondria-mediated apoptosis through the RhoA/Rho-associated kinase pathway in cortical neurons. *Eur J Neurosci* **2007,** 25, (5), 1341-8.

48. Koyanagi, M.; Takahashi, J.; Arakawa, Y.; Doi, D.; Fukuda, H.; Hayashi, H.; Narumiya, S.; Hashimoto, N., Inhibition of the Rho/ROCK pathway reduces apoptosis during transplantation of embryonic stem cell-derived neural precursors. *J Neurosci Res* **2008,** 86, (2), 270-80.

49. Iwasaki, T.; Katayama, T.; Kohama, K.; Endo, Y.; Sawasaki, T., Myosin phosphatase is inactivated by caspase-3 cleavage and phosphorylation of myosin phosphatase targeting subunit 1 during apoptosis. *Mol Biol Cell* **2013,** 24, (6), 748-56.

50. Kaufman, S., The phenylalanine hydroxylating system. *Adv Enzymol Relat Areas Mol Biol* **1993,** 67, 77-264.

51. Martinez, A.; Knappskog, P. M.; Olafsdottir, S.; Doskeland, A. P.; Eiken, H. G.; Svebak, R. M.; Bozzini, M.; Apold, J.; Flatmark, T., Expression of recombinant human phenylalanine hydroxylase as fusion protein in Escherichia coli circumvents proteolytic degradation by host cell proteases. Isolation and characterization of the wild-type enzyme. *Biochem J* **1995,** 306 (Pt 2), 589-97.

52. Moller, N.; Meek, S.; Bigelow, M.; Andrews, J.; Nair, K. S., The kidney is an important site for in vivo phenylalanine-to-tyrosine conversion in adult humans: A metabolic role of the kidney. *Proc Natl Acad Sci U S A* **2000,** 97, (3), 1242-6.

53. Seyfried, T. N.; Mukherjee, P., Targeting energy metabolism in brain cancer: review and hypothesis. *Nutr Metab (Lond)* **2005,** 2, 30.

54. Mukherjee, P.; Abate, L. E.; Seyfried, T. N., Antiangiogenic and proapoptotic effects of dietary restriction on experimental mouse and human brain tumors. *Clin Cancer Res* **2004,** 10, (16), 5622-9.

VITA

Xiaoling Zang was born in Qingdao, Shandong, China. She attended elementary schools in Qingdao and Changchun, and middle school and high school in Changchun. In fall 2008 she enrolled in the pilot class of biology and chemistry at Jilin University, Changchun, Jilin, China. In fall 2009, she transferred to University of North Carolina at Chapel Hill (UNC-CH), NC, USA to continue college study. She received her B.A. in Chemistry (biochemistry track) at UNC-CH in 2012. In fall 2012 she attended graduate school at Georgia Institute of Technology, GA, USA to pursue her doctorate in analytical chemistry. When she is not working, Xiaoling enjoys playing table tennis, running, singing, and drawing.