

Applications of NMR-based Metabolomics: from cells to human body fluids to food

A Thesis

Submitted in Partial Fulfillment of Requirements for the Award of Degree of

Doctor of Philosophy

By

Saleem Yousf

20153380



Indian Institute of Science, Education, and Research, Pune

2020



Jeetender Chugh, Ph.D.
Assistant Professor,
Department of Chemistry & Department of Biology
Indian Institute of Science Education & Research (IISER),
Homi Bhabha Road, Pashan, Pune 411008, India
E-mail: cjeet@iiserpune.ac.in
Phone: +91-20-25908121, +91-8378979667

October 26, 2020

Certificate

I certify that the thesis entitled “**Applications of NMR-based Metabolomics: from cells to human body fluids to food**” presented by **Mr. Saleem Yousf** represents his/her original work which was carried out by him at IISER, Pune under my guidance and supervision during the period from 01st January, 2015 to 28th October, 2020. The work presented here or any part of it has not been included in any other thesis submitted previously for the award of any degree or diploma from any other University or institutions. I further certify that the above statements made by him in regard to his thesis are correct to the best of my knowledge.

26th Oct 2020
Date:


Jeetender
(Supervisor: Dr. Jeetender Chugh)

DECLARATION

I declare that this written submission represents my idea in my own words and where others' ideas have been included; I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Place: **IISER Pune**

Date: **28/10/2020**


Mr. Saleem Yousf

Acknowledgments

First and foremost, all praises and thanks to the Almighty Allah for his showers of blessing throughout my research and has helped me to complete this thesis. I would like to thank all the people who have helped and inspired me during my Ph.D. journey. So, I will attempt to give them their due here, and I sincerely apologize for any omissions.

I would like to express my gratitude to my supervisor Dr. Jeetender Chugh for his continuous supervision, trust, patience, motivation, enthusiasm and immense knowledge. His guidance helped me out in all the time of research and writing of the thesis. Above all and the most needed, he provided me unflinching encouragement and support in various ways. I am really indebted to him more than he knows. I'm deeply indebted to Dr. Shilpy Sharma (Savitribai Phule Pune University, Pune), thanks a lot for her concern, great help and patience in my life as well as new ideas, fruitful discussion and constructive comment in my study. I would like to thank my Research Advisory Committee (RAC) members, Dr. Pankaj Mandal and Dr. Santosh Kumar Jha (NCL Pune) for tracking the progress of my research over the years and providing valuable insights, suggestions and fruitful discussions during all meetings.

I would also like to extend my deepest gratitude to my collaborators Dr. Shilpy Sharma (Savitribai Phule Pune University, Pune), Dr. Sharmistha Banerjee (University of Hyderabad), Dr. Sagar (Savitribai Phule Pune University, Pune) in designing the study experiments and helping me in various biochemical and cellular assays. I am very grateful for the support of all co-authors in my papers. I would also acknowledge IISER-Pune for all the research facilities. This work would not have been complete without generous support by NMR facility at IISER-Pune (co-funded by DST-FIST and IISER Pune) for access to 600 MHz NMR spectrometer.

I am also thankful to Nitin, Sandeep, and Chinmay for maintaining the NMR facility. I am also grateful to Mr. Sachin Kate and Mr. Raghav (Bruker Biospin) for quick support to resolve all technical issues with the NMR spectrometers.

I take this opportunity to sincerely acknowledge the University Grants Commission (UGC), Government of India, New Delhi, for providing financial assistance in the form of Research Fellowship which buttressed me to perform my work comfortably. I am also obliged to Infosys Foundation for the financial support provided in the form of travel grants to present part of this work at Weizmann Institute of Science, Israel.

Collective and individual acknowledgments are also owed to past and present lab members Santosh, Harshad, Firdousi, Sarita, Arpita, Aazam, Guneet, Samra, Osama and many more for sharing all good and bad moments of the Ph.D. life. Furthermore, I really appreciate my friends for making my life more exciting and funnier. Special thanks to my best buddy Zahid Bhat, who is not only a friend but a brother.

And most of all, I would like to share this moment of happiness with my loving, supportive, encouraging, family where the most basic source of my life energy resides. The support of my parents, brother and sisters has been unconditional all these years and without their encouragement, prayers and understanding it would have been impossible for me to finish this work. I very fondly acknowledge my brother-in laws for their love and support. I sincerely thank all my relatives who were with me for their love and encouragement.....

This thesis is dedicated to my ammi and dadi...

Saleem Yousf

Abbreviations

°C	Degree Celsius
%	Percent
ADP	Adenosine Diphosphate
AMP	Adenosine Monophosphate
ATP	Adenosine Triphosphate
AUC	Area Under the Curve
ANOVA	Analysis of Variance
BMRB	Biological Magnetic Resonance Data Bank
COSY	Correlation Spectroscopy
CPMG	Carr-Purcell-Meiboom-Gill spin-echo sequence
CSV	Comma-separated values file
D2O	Deuterium oxide
DSS	4,4-dimethyl-4-silapentane-1-sulfonic acid
FDR	False Discovery Rate
FT	Fourier Transform
GABA	γ -Aminobutyric Acid
GC	Gas Chromatography
GC-MS	Gas Chromatography-Mass Spectrometry
HMDB	Human Metabolome Database
HSQC	Heteronuclear Single Quantum Coherence
J-RES	J-Resolved
K	Kelvin
KEGG	Kyoto Encyclopedia of Genes and Genomes
LDL	Low density lipoproteins
MS	Mass Spectrometry
MVA	MultiVariate Analysis
NOESY	Nuclear Overhauser Effect Spectroscopy
NMR	Nuclear Magnetic Resonance
NUS	Non-Uniform Sampling
OPLS-DA	Orthogonal Partial Least Squares Discriminant Analysis
OSC	Orthogonal Signal Correction

PC	Principal Component
PCA	Principal Component Analysis
PLS	Partial Least Squares
PLS-DA	Partial Least Squares Discriminant Analysis
ppm	Parts per million
PQN	Probabilistic Quotient Normalization
ROC	Receiver Operating Characteristic
RF	Radiofrequency
S/N	Signal-to-noise ratio
SAM	Significance Analysis of Microarrays
TCA	Tricarboxylic acid
TOCSY	Total Correlation Spectroscopy
TSP	3-(trimethylsilyl)-2,2,3,3-tetradeuteropropionic acid
T2DM	Type 2 Diabetes mellitus
UF	Ultra-Fast
UVA	UniVariate Analysis
VIP	Variable Importance on Projection

Synopsis

Applications of NMR-based Metabolomics: from cells to human body fluids to food

Name: Mr. Saleem Yousf

Registration No.: 20153380

Thesis Supervisor: Dr. Jeetender Chugh.

Department: Department of Chemistry, IISER, Pune.

Living organisms possess the ability to tune themselves in response to various intra- and extracellular cues, which in turn creates an assortment of metabolites in the cellular soup. Metabolomics – the science that deals with the identification and quantification of metabolites in a biological system under a set of conditions – is gaining recognition as a powerful tool that helps comprehend the effect of various stressors and environmental conditions on the metabolome of a biological system. NMR-based metabolomics is being used in different areas, including disease biology, medicine, pharmacy, toxicology, food, environmental sciences, etc., to obtain the metabolic fingerprint associated with different biological systems. The work carried out in this Ph.D. thesis has focussed on the analysis of samples ranging from animal cells, microbial cultures, plants to human body fluids using a metabolomics approach. This thesis is divided into six chapters.

Chapter 1: Introduction

This chapter details on the need of performing metabolomics studies and describes the methods commonly used in NMR-based metabolomics. It provides an overview of metabolomics and covers important topics such as sample preparation and collection, multiple analytical platforms currently in use to study the metabolome, description of majorly used NMR experiments, various spectral pre- and post-processing steps needed to analyze raw spectral data, data processing tools to eliminate spectral artifacts and remove biologically irrelevant variations, statistical analysis for data interpretation and modelling for validation. Recent developments in NMR methodology pertaining to metabolomics has also been discussed.

Chapter 2: General Methodology

The combined use of NMR spectroscopy and chemometrics techniques can provide the metabolic “*fingerprint*” of the different types of biological samples analysed. This chapter provides information about important parameters, e.g., NMR experiments, metabolomics spectral databases, computational tools for metabolomic data and pathway analysis, software packages for metabolite identification, their quantification, and types of analysis that have been performed for data interpretation and feature selection. The selection criteria for different chemometrics techniques used in metabolite biomarker discovery have also been highlighted.

Chapter 3: Metabolic signatures suggest o-phosphocholine to UDP-N-acetylglucosamine ratio as a potential biomarker for high-glucose and/or palmitate exposure in pancreatic β -cells

Chronically elevated glucose (hyperglycemia) and lipid levels (dyslipidemia in T2DM confounded with obesity) are the major phenotypes associated with T2DM and have been known to induce dysfunction and apoptosis of pancreatic β -cell. In particular, in this chapter, we aimed to determine the metabolic signatures associated with high glucose (glucotoxic), high lipid exposure (lipotoxic) alone and in combination (glucolipotoxic) conditions in INS-1E cells (pancreatic β -cells) and have identified metabolic pathways that play significant roles in excess-fuel detoxification in these cells. Here, we employed untargeted ^1H NMR-based metabolomics to identify and quantify the metabolites, and use this information to gain insights into the associated metabolic shifts and key metabolic pathways that are significantly altered in *INS-1E* cells exposed to different toxic conditions used in this study. The perturbed metabolites identified in the course of this work majorly belong to glycolysis, TCA cycle, amino acid metabolism, and hexosamine metabolism pathways. Interestingly, UDP-N-acetylglucosamine and o-phosphocholine were identified as the commonly dysregulated metabolites under all three stress conditions used, and we proposed to use the ratio of these metabolites as a biomarker for these conditions. The results from this study helped us in annotating the critical metabolites associated with diabetes-like conditions and consequently in constructing the metabolic pathways that play crucial roles in excess-fuel detoxification associated with chronic T2DM.

Chapter 4: Identification of potential serum biomarkers linked with Type 2 Diabetes in an Asian Indian population using NMR-based metabolomics

Type 2 diabetes mellitus (T2DM) is a progressive metabolic disorder that has been characterized by chronic hyperglycemia that occurs due to insulin insufficiency from

pancreatic β -cells and has been associated with the development of insulin resistance (IR) in the insulin-target tissues. The prevalence of prediabetes and T2DM is increasing globally, with worrying statistics from the children, adolescents, and young adults worldwide and in developing countries like India. As per the International Diabetes Federation, ~77 million people, which account for one in six people, or 17% of the world diabetic population, belong to India (<https://www.diabetesatlas.org/en/>). Here, we aimed to identify unique circulating metabolic markers associated with pre-diabetes and T2DM in Asian Indians using NMR-based metabolomics that could be used as potential biomarkers for prognosis and disease diagnosis. A total of 284 individuals (Asian Indians) were recruited for the study. Our study identified 36 aqueous metabolites from the methanolic extracts of serum samples collected from each individual after overnight fasting, of which 24 metabolites showed a statistically significant difference between normal individuals (referred to as healthy controls), prediabetic individuals, and subjects with established T2DM. Using ROC curve analysis, 12 metabolites (including glucose, pyroglutamate, serine, proline, glutamate, methionine, isoleucine, alanine, citrate, betaine, glycerol, and o-phosphocholine) in the T2DM subjects; and six metabolites (including glucose, pyroglutamate, o-phosphocholine, serine, snglycero-3-phosphocholine, and methionine) in prediabetic subjects, were identified with high specificity and sensitivity (AUC > 0.7). On performing multivariate ROC curve analysis with the panel of selected 5 metabolites commonly dysregulated in prediabetes and T2DM subjects, AUC values obtained were 0.96 (95% confidence interval (CI) = 0.93, 0.98) for established T2DM; and 0.88 (95% CI = 0.81, 0.93) for prediabetic subjects. Hence, we propose that this panel of 5 metabolic biomarkers (namely, glucose, pyroglutamate, o-phosphocholine, serine, and methionine) can be used in the future for clinical diagnosis, patient surveillance, and for predicting individuals at risk for developing overt diabetes in the future in the South Asian Indians.

Chapter 5: Mapping metabolic perturbations in *Mycobacterium smegmatis* in response to different stress conditions using NMR spectroscopy

Mycobacterium smegmatis, the saprophytic soil mycobacteria besides being extensively applied as a surrogate for the human pathogen *Mycobacterium tuberculosis*, is an important constituent of natural and human-engineered habitats, including environmental inter-surfaces like aerosolized water droplets. Its survival in consistently changing ecological setups and turning into an opportunistic pathogen in immune-compromised hosts suggest high versatility at molecular and cellular levels. Here, we aimed to identify metabolites/metabolic

pathways critical for early adaptive response to abiotic stresses like acidic, oxidative, and nutrient starvation in *Mycobacterium smegmatis*. Using untargeted ¹H NMR based metabolomics, 22, 21, and 47 dysregulated metabolites were identified in acidic, oxidative, and nutrient starvation conditions. Distinct topological shifts were observed in purine-pyrimidine, amino-acid metabolisms, and energy metabolism pathways. The accumulation of organic osmolytes, such as dimethylamine, methylamine, and betaine during nutrient starvation and oxidative stress, was specifically noticeable. Tracing these accumulated osmolytes through computational search tools and gene-expression studies, we deciphered pathways of biosynthesis of betaine, methylamine and dimethylamine (previously undocumented and unreported in *Mycobacterium smegmatis*). We also observed the differential levels of intermediary metabolites involved in the α -glucan biosynthesis pathway. This study documents, for the first time, the metabolic changes that occur in *Mycobacterium smegmatis* as a response to three stresses, namely, acidic stress, oxidative stress, and nutrient starvation. These stresses are also faced by intracellular mycobacteria during infection and, therefore, may be extended to frame therapeutic interventions for pathogenic mycobacteria.

Chapter 6: Cold storage reveals distinct metabolic perturbations in processing and non-processing cultivars of potato (*Solanum tuberosum* L.)

Potato (*Solanum tuberosum*), the largest popular non-grain vegetable food crop worldwide, belongs to the Solanaceae family and ranks third most wanted food crop after wheat and rice. Cold-induced sweetening (CIS) has been known to cause a significant loss to the potato processing industry, wherein the selection of potato genotypes using biochemical information through marker-trait associations has been found to be advantageous. The potato industry faces a massive loss due to CIS as chips and French fries get discolored, a parameter which is primarily determined by RS content in potato tubers. As a result, CIS is known to be one of the critical parameters in potato cultivation; the selection and breeding of CIS-resistant potato tubers have become a priority in potato breeding programs. In this study, an untargeted ¹H nuclear magnetic resonance (NMR)-based approach was conducted to assess the alterations in the metabolic profiles of five different potato cultivars namely, Atlantic, Frito Lay-1533, Kufri Pukhraj, Kufri Jyoti, and PU1, upon cold storage at 4 °C. These potato cultivars differ in their CIS abilities and processing characteristics. The key purpose of this work was to investigate the variations in metabolic profiles of different potato cultivars at fresh harvest and after cold storage to further advance the knowledge of metabolic events associated with the CIS phenomenon. From this study, key metabolites were identified that could potentially be

used in breeding programs for the development of CIS-resistant cultivars with improved processing characteristics and thereby would enhance the quality of potato

Conclusions

Using different biological systems, ranging from animal cells, microbial cultures, plants to human body fluids, we have been able to gain insights into important biological problems. Based on the results obtained in this thesis, we have been able to identify perturbed metabolites (using ^1H NMR spectroscopy) that can be used as biomarkers in future.

List of Publications

1. **Saleem Yousf**¹, Devika M Sardesai¹, Abraham B Mathew, Rashmi Khandelwal, Jhankar D Acharya, Shilpy Sharma*, and Jeetender Chugh* *Metabolic signatures suggest o-phosphocholine to UDP-N-acetylglucosamine ratio as a potential biomarker for high-glucose and/or palmitate exposure in pancreatic beta-cells.* *Metabolomics*; 2019, 15(4).
2. **Saleem Yousf**, Rakesh Mohan, Devika M Sardesai, Shilpy Sharma*, and Jeetender Chugh* *Identification of potential serum biomarkers linked with Type 2 Diabetes in an Asian Indian population using NMR-based metabolomics.* (Manuscript under preparation).
3. Arshad Rizvi¹, **Saleem Yousf**¹, Kannan Balakrishnan, Harish K Dubey, Shekhar C Mande, Jeetender Chugh*, and Sharmistha Banerjee* *Metabolomics studies to decipher stress responses in Mycobacterium smegmatis point to a putative pathway of methylated amines biosynthesis.* *Journal of Bacteriology*; 2019, 201(15), 477-489
4. Sagar S Datir*¹, **Saleem Yousf**¹, Shilpy Sharma, Mohit Koshle, Ameeta Ravikumar, and Jeetender Chugh* *Cold storage reveals distinct metabolic perturbations in processing and non-processing cultivars of potato (Solanum tuberosum L.).* *BioRxiv*; 2019, doi: <https://doi.org/10.1101/661611>; *Nature Scientific Reports*; 2020, 10, Article No. 6268
5. **Saleem Yousf**, Nazia Hussain, Shilpy Sharma*, Jeetender Chugh* *Identification & Characterization of Secondary Metabolites in the Biological Soup by NMR Spectroscopy.* *Applications of NMR Spectroscopy*; Bentham Science 2017.
6. Malla, Javid; Umesh, Rintu; **Saleem Yousf**; Mane, Shrinal; Sharma, Shilpy; Lahiri, Mayurika; Talukdar, Pinaki *. *A Glutathione Activatable Ion Channel Induces Apoptosis in Cancer Cells by Depleting Intracellular Glutathione Levels.* *Angew. Chem. Int. Ed.* 2020, Accepted. DOI: 10.1002/anie.202000961.
7. Madhuri Vangala*, **Saleem Yousf**, Jeetender Chugh, and Srinivas Hotha *Solid-Phase Synthesis of Clickable Psicofuranose Glycocarbamates and Application of Their Self-Assembled Nanovesicles for Curcumin Encapsulation.* *ChemistrySelect*; 2020, 5(9), 2672-2677

(¹Authors contributed equally).

Table of Contents

<i>Abbreviations</i>	<i>VI</i>
<i>Synopsis</i>	<i>VIII</i>
<i>Chapter 1</i>	<i>1</i>
<i>Introduction</i>	<i>1</i>
1.1 <i>Introduction to metabolomics</i>	<i>2</i>
1.2 <i>Targeted and Untargeted Metabolomics</i>	<i>3</i>
1.3 <i>The metabolomics workflow</i>	<i>4</i>
1.4 <i>Sample preparation</i>	<i>5</i>
1.5 <i>Analytical platforms</i>	<i>5</i>
1.6 <i>NMR as a tool in metabolomics</i>	<i>7</i>
1.6.1 <i>¹H NMR Spectroscopy</i>	<i>8</i>
1.6.2 <i>2D NMR Spectroscopy</i>	<i>9</i>
1.6.3 <i>Recent advances in NMR spectroscopy</i>	<i>14</i>
1.7 <i>Post-acquisition processing of NMR spectra</i>	<i>16</i>
1.7.1 <i>Chemical shift referencing</i>	<i>16</i>
1.7.2 <i>Phase correction</i>	<i>17</i>
1.7.3 <i>Baseline correction</i>	<i>17</i>
1.7.4 <i>Zero filling</i>	<i>18</i>
1.7.5 <i>Apodization</i>	<i>18</i>
1.8 <i>NMR data pre-processing for chemometric analysis</i>	<i>18</i>
1.8.1 <i>Binning</i>	<i>18</i>
1.8.2 <i>Spectral alignment</i>	<i>19</i>
1.8.3 <i>Normalisation</i>	<i>19</i>
1.8.4 <i>Scaling</i>	<i>20</i>
1.8.5 <i>Mean centering</i>	<i>20</i>
1.8.6 <i>Autoscaling</i>	<i>21</i>
1.8.7 <i>Pareto scaling</i>	<i>21</i>
1.9 <i>Univariate Analysis</i>	<i>21</i>
1.10 <i>Multivariate Analysis (MVA)</i>	<i>22</i>
1.10.1 <i>Principal Component Analysis (PCA)</i>	<i>22</i>
1.10.2 <i>Partial least squares for discrimination (PLS-DA)</i>	<i>24</i>
1.10.3 <i>Statistical validation of multivariate models</i>	<i>26</i>
1.11 <i>Receiver operating characteristic (ROC) curve</i>	<i>26</i>
1.12 <i>Correlation analysis and heat maps</i>	<i>27</i>
1.13 <i>Significance Analysis for Microarrays (SAM)</i>	<i>29</i>
1.14 <i>Pathway analysis</i>	<i>30</i>
1.15 <i>Summary of the chapter</i>	<i>30</i>

Chapter 2	43
General Methodology	43
2.1 Introduction	44
2.2 NMR Spectroscopy	44
2.3 Spectral processing	45
2.4 Metabolite Identification and Quantification	45
2.5 Statistical Analysis	46
2.6 Metabolic pathway analysis	48
Chapter 3	51
Metabolic signatures suggest o-phosphocholine to UDP-N-acetylglucosamine ratio as a potential biomarker for high-glucose and/or palmitate exposure in pancreatic β-cells	51
3.1 Introduction	52
3.2 Experimental Methods	54
3.2.1 Cell culture and establishment of cytotoxic conditions	54
3.2.2 Measurement of lipid accumulation	57
3.2.3 Reactive Oxygen Species (ROS) measurement	57
3.2.4 Metabolite extraction and sample preparation for NMR	57
3.3 NMR spectroscopy and spectral processing	58
3.4 Spectral assignment and quantitative evaluation	58
3.5 Statistical Analysis	58
3.6 Metabolic pathway analysis	59
3.7 Results	59
3.7.1 HG and/or PA exposure induces cytotoxicity in INS-1E cells	59
3.7.2 Global profiling of metabolites using ^1H NMR spectroscopy	62
3.7.3 Pair-wise analysis of metabolic changes	67
3.8 Metabolic Pathway Analysis	70
3.9 Discussion	72
3.10 Conclusions	78
Chapter 4	91
Identification of potential serum biomarkers linked with Type 2 Diabetes in an Asian Indian population using NMR-based metabolomics	91
4.1 Introduction	92
4.2 Experimental Procedures	94
4.2.1 Study Population, Anthropometric parameters, and biochemical tests	94
4.2.2 NMR sample preparation	95
4.3 NMR spectroscopy and Spectral processing	95
4.4 Metabolite Identification and quantitation	95

4.5 Statistical Analysis	96
4.6 Receiver operating characteristic (ROC) curve.....	97
4.7 Metabolic pathway analysis.....	97
4.8 Results	97
4.8.1 Clinical characteristics of the study population	97
4.8.2 Resonance Assignment	98
4.8.3 Classification and Feature selection.....	100
4.8.4 Identification of Metabolite Biomarkers associated with diabetes.....	103
4.8.5 Metabolic Biomarkers for prediabetes	104
4.8.6 Metabolic Pathway Analysis.....	106
4.9 Discussion	109
4.10 Conclusions	112
Chapter 5	119
<i>Mapping metabolic perturbations in Mycobacterium smegmatis in response to different stress conditions using NMR spectroscopy.....</i>	<i>119</i>
5.1 Introduction.....	120
5.2 Experimental Methods.....	122
5.2.1 Bacterial strains and experimental conditions.	122
5.2.2 Metabolite Extraction.	122
5.3 NMR acquisition and spectral processing	123
5.4 Metabolite identification and quantification	123
5.5 Statistical analysis.....	124
5.6 Metabolic pathway analysis.....	124
5.7 RNA isolation and semi-quantitative RT-PCR.....	125
5.8 Identification of putative pathway using in silico tools.....	125
5.9 Results	126
5.9.1 Global metabolome profiling and metabolic alterations of <i>M. smegmatis</i>	126
5.10 Pathway analysis.....	134
5.11 Discussion	135
5.11.1 Putative pathway of biosynthesis of methylated amines in <i>M. smegmatis</i>	138
5.12 Conclusion.....	140
Chapter 6	149
<i>Cold storage reveals distinct metabolic perturbations in processing and non-processing cultivars of potato (Solanum tuberosum L.).....</i>	<i>149</i>
6.1 Introduction.....	150
6.2 Materials and Methods	153
6.2.1 Plant Material	153
6.2.2 Potato Plantation and Harvesting	153

6.2.3 Metabolite extraction	155
6.3 NMR spectroscopy and spectral processing.....	156
6.4 Quantification assessment and statistical analysis.....	156
6.5 Metabolic pathway analysis.....	157
6.6 Results	157
6.6.1 Global profiling of metabolites in different potato cultivars – Processing versus non-processing cultivars.....	157
6.6.2 Cold treatment influences the metabolome of potato cultivars.....	160
6.6.3 Metabolic perturbations in processing, non-processing, and local cultivars under CS treatments.	163
6.6.4 Metabolic correlation network analysis	167
6.7 Metabolic pathway analysis and potential metabolite insights of CIS	170
6.8 Discussion	172
6.8.1 Sugar metabolism under cold storage condition	172
6.8.2 Amino acid metabolism under cold storage condition	173
6.8.3 TCA cycle metabolites under cold storage condition.....	174
6.9 Conclusions.....	177
Appendix 1.....	188
Supporting Tables	188
Appendix 2.....	212
Questionnaire, Consent form and Approvals	212

List of Figures

Figure 1.1: The central dogma of biology and the omics cascade.	3
Figure 1.2: Workflow followed for conducting a metabolomics study.	4
Figure 1.3: NMR experiments typically used for the analysis of metabolites.	7
Figure 1.4: 2D ^1H - ^1H Correlation Spectroscopy (COSY).	10
Figure 1.5: 2D ^1H - ^1H Total Correlation Spectroscopy (TOCSY).	11
Figure 1.6: 1D TOCSY spectrum of glycoconjugates in MeOH- d_3 .	11
Figure 1.7: 2D J-resolved NMR Spectroscopy (J-RES).	12
Figure 1.8: 2D ^1H - ^{13}C Heteronuclear Single Quantum Coherence (HSQC).	13
Figure 1.9: Ultrafast acquisition strategy.	15
Figure 1.10: (a) The 3D [^1H - ^{13}C] HSQC-TOCSY spectrum.	16
Figure 1.11: Schematic illustration of a PCA model.	23
Figure 1.12: Representative Score and Loading plot of the first two components of a PCA model.	24
Figure 1.13: Heat map of <i>Mycobacterium smegmatis</i> grown under different stress conditions.	28
Figure 1.14: Significance analysis of microarray (SAM) plots for selection of significant metabolites.	29
Figure 3.1: Effect of high glucose (HG), Palmitic acid (PA), and High glucose in combination with Palmitic acid (HG/PA) exposure on cellular viability, triglyceride content, lipid accumulation and ROS generation in <i>INS-IE</i> cells.	61
Figure 3.2: Representative ^1H -NMR spectrum of the methanolic extract of <i>INS-IE</i> cells grown in basal glucose (5 mM) containing media.	63
Figure 3.3: ^1H - ^1H TOCSY correlation spectrum of the methanolic extract of <i>INS-IE</i> cells cultured in media containing basal glucose concentration.	64
Figure 3.4: One-way ANOVA analysis followed by post-hoc analysis showing 24 significantly regulated (q -value ≤ 0.05 , marked in red circles) metabolites. The dotted horizontal line corresponds to threshold of FDR-adjusted p -value (q -value ≤ 0.05) on a \log_{10} scale.	65
Figure 3.5: Box-Whisker plots displaying the comparative variation in concentration of each individual significantly (q -value ≤ 0.05) altered metabolite identified from ANOVA and post-hoc analysis against six experimental conditions; namely, G1 (Basal glucose; red); G2 (Basal Glucose with BSA as vehicle control for palmitate; green); G3 (High PA; blue); G4 (High	

glucose; cyan); G5 (High glucose with BSA; magenta); and G6 (High glucose and PA; yellow).

66

Figure 3.6: Score plots of PLS-DA (A) and variable importance in projection (VIP) (B) obtained from MetaboAnalyst 4.0.

67

Figure 3.7: Pairwise metabolite comparison between *INS-IE* cells cultured in, basal glucose vs high glucose (HG) (A) PLS-DA score plot and (B) volcano plot; basal glucose vs High FFA (PA) (C) PLS-DA score plot and (D) volcano plot; and basal glucose vs high glucose and FFA (HG/PA) (E) PLS-DA score plot and (F) volcano plot.

70

Figure 3.8: Metabolic Pathway Analysis (MetPa) of significantly altered metabolites displaying significant metabolic pathways in *INS-IE* cells when associated with (A) glucotoxic, (B) lipotoxic, and (C) glucolipotoxic conditions.

71

Figure 3.9: Pictorial depiction of the significantly dysregulated metabolites and Metabolomic Pathway Analysis (MetPA) construction of the metabolic pathways identified in HG, PA and HG/PA.

77

Figure 4.1: Representative ^1H NMR spectrum of serum obtained from healthy and diabetic patients.

99

Figure 4.2: Box and whisker plots displaying relative normalized concentrations of significantly altered metabolites identified from One-way ANOVA (q -value < 0.05) in control, prediabetic, and T2DM groups.

101

Figure 4.3: Significance analysis of microarray (SAM) plots for the selection of significant metabolites in (A) T2DM group, and (B) prediabetic group.

102

Figure 4.4: Supervised 2D OPLS-DA analysis of (A) Healthy control versus T2DM patients, (C) Healthy controls versus prediabetic patients. OPLS-DA model validation by permutation tests based on the separation between (B) Healthy control versus T2DM patients (D) Healthy control versus prediabetic patients.

103

Figure 4.5: Univariate ROC curves of the potential biomarkers ($\text{AUC} > 0.7$) and their respective box-cum-whisker plots showing discriminatory ability between T2DM patients and healthy controls.

104

Figure 4.6: Univariate ROC curves of the potential biomarkers ($\text{AUC} > 0.7$) along with their respective box-cum-whisker plots showing discriminatory ability between prediabetic patients and healthy controls.

105

Figure 4.7: Multivariate ROC curve analysis for the evaluation of model prediction using balanced-subsampling based Monte-Carlo cross-validation (MCCV).

106

Figure 4.8: Summary of pathway enrichment analysis in T2DM serum samples.

107

Figure 4.9: Metabolic Pathway Analysis (MetPa) of significantly differential metabolites between healthy controls and (A) T2DM groups, (B) Prediabetic groups.	108
Figure 4.10: Pictorial depiction of the significantly dysregulated metabolites and Metabolomic Pathway Analysis (MetPA) construction of the metabolic pathways identified in T2DM.	109
Figure 4.11: Multivariate ROC curve analysis of serum samples using the combination of five metabolite biomarkers (glucose, pyroglutamate, o-phosphocholine, serine, and methionine) common in T2DM and Prediabetic subjects with the AUC > 0.7.	110
Figure 5.1: Representative ¹ H-NMR spectrum of the methanolic extract of <i>M. smegmatis</i> grown in Sauton media.	127
Figure 5.2: Two-dimensional (2D) ¹ H- ¹ H TOCSY correlation spectrum of the methanolic extract of <i>M. smegmatis</i> grown in Sauton media.	128
Figure 5.3: PLS-DA score plots for <i>M. smegmatis</i> samples grown in different stress conditions showing model discrimination between (A) acidic stress versus control; (B) oxidative stress versus control; (C) nutrient starvation versus control; (D) Acidic, oxidative, and nutrient starvation versus the control.	129
Figure 5.4: Volcano plot, where a dotted horizontal line corresponded to FDR-correct p-value 0.05 on a log ₁₀ scale and dotted vertical lines represent a 1.5-fold change in concentration of metabolites on a log ₂ scale.	131
Figure 5.5: Box and whisker plots of relative concentrations for significantly differential metabolites identified from the VIP score plot (VIP > 1) in <i>M. smegmatis</i> grown under different stress conditions.	133
Figure 5.6: Metabolic Pathway Analysis (MetPa) of significantly altered metabolites under (A) acidic, (B) oxidative, and (C) Starvation stress.	135
Figure 5.7: Pictorial depiction of the significantly dysregulated metabolites and Metabolomic Pathway Analysis (MetPA) construction of the metabolic pathways identified in <i>M. smegmatis</i> .	136
Figure 5.8: Metabolic pathways associated with biosynthesis of capsular α-glucan via GlgE pathway.	138
Figure 5.9: A putative pathway of biosynthesis of methylated amines in <i>M. smegmatis</i> .	140
Figure 6.1: ¹ H-NMR spectrum of the methanolic extract of Kufri Pukhraj potato cultivar (cold storage).	159
Figure 6.2: ¹ H- ¹ H TOCSY correlation spectrum of the methanolic extract of Kufri Pukhraj potato cultivar (cold storage).	160

- Figure 6.3:** Scores plot of PLS-DA generated from MetaboAnalyst software for the different potato cultivars – Atlantic (G1, G2), Frito Lay-1533 (G3, G4), Kufri Pukhraj (G5, G6), Kufri Jyoti (G7, G8), and PU1 (G9, G10) at fresh harvest and one-month cold storage at 4°C. 161
- Figure 6.4:** PLS-DA scores plots for pair-wise comparison analysis of metabolites between potato tubers samples at fresh harvest (Red) and cold storage at 4°C for 1 month (Green) in A) Atlantic, B) Frito Lay-1533, C) Kufri Jyoti, D) Kufri Pukhraj, and E) PU1. 163
- Figure 6.5:** VIP scores obtained after pair-wise PLS-DA analysis for A) Atlantic, B) Frito Lay-1533, C) Kufri Jyoti, D) Kufri Pukhraj, and E) PU1. 163
- Figure 6.6:** Volcano plots pairwise metabolite comparison between freshly harvested (FH) and cold storage (CS) tubers. 165
- Figure 6.7:** Box-Whisker plots displaying the comparative alteration in the concentration of discriminating metabolites (identified from VIP score plot and Volcano plot analysis) for both freshly harvested and cold storage tuber samples across replicates and in different potato cultivars used in the study. FH – fresh harvest and CS – cold storage at 4 °C. 167
- Figure 6.8:** Correlation plots between fresh harvest (FH; upper-right half of the plot marked in white) and cold storage (CS; lower-left half of the plot marked in light blue) metabolites for A) Atlantic, B) Frito Lay-1533, C) Kufri Jyoti, D) Kufri Pukhraj, and E) PU1. 169
- Figure 6.9:** Pictorial depiction of metabolic pathways affected during cold-induced sweetening in the different cultivars of potato. 176

List of Tables

Table 3.1: Experimental conditions used in this study.	56
Table 3.2: Calculation and comparison of ratio of o-phosphocholine to UDP-N-GlcNAc under gluco-, lipo-, and glucolipototoxicity conditions in comparison to basal glucose.	76
Table 5.1: VIP value obtained from PLS-DA models for Control vs Acidic stress, Control vs Oxidative stress, Control vs Starvation stress, and Control vs all stresses (Acidic, Oxidative and Starvation).	132
Table 6.1: information on potato cultivars.	155
Table 6.2: The number of correlation observed for the metabolites in the different cultivars.	170
Table 6.3: Metabolic pathway analysis revealing the significantly impacted metabolic pathways in potato tubers under cold storage	171
Table 3.3: List of chemical shift assignment of abundant water-soluble metabolites as extracted from <i>INS-IE</i> cells	189
Table 3.4: Summary of metabolic studies performed in pancreatic β -cells following high-glucose and FFA exposure	195
Table 4.2: Serum metabolites (as identified from Figure 4.1) were assigned to their respective proton chemical shifts (ppm)	198
Table 4.3: p-values and FDR-values of Control vs. T2DM group obtained from ANOVA.	199
Table 4.4a: d(i) values, p-values and FDR-values of Control vs. T2DM group obtained from SAM analysis	200
Table 4.4b: d(i) values, p-values, and FDR-values of Control vs. Prediabetic group obtained from SAM analysis.	200
Table 5.2: <i>Mycobacterium smegmatis</i> metabolites (as identified from Figure 5.1) were assigned to their respective proton chemical shifts (ppm)	201
Table 5.3: Fold change (FC), p-values and FDR-values of Control vs Acidic, Oxidative and Starvation stresses.	203
Table 5.4: Primers used in the study	206
Table 6.4: List of chemical shift assignment of abundant water-soluble metabolites extracted from potato tuber samples	207

Chapter 1

Introduction

1.1 Introduction to metabolomics

Omics sciences are rapidly evolving disciplines that have been developed for studying and describing complex biological systems in a holistic manner (Berry et al. 2011; Kell 2004). The initial understanding of complexity of living organisms was classically described in the simplest way with the help of central dogma – the directional flow of information (Kaddurah-Daouk, Kristal, and Weinshilboun 2008) from genomic DNA to proteins via mRNA transcripts, all are integrated to generate a desired phenotype. Under this ‘Omics’ umbrella, the analysis of genome (referred to as **Genomics**); the analysis of complete set of RNA transcripts produced by the genome at a particular instant (referred to as **Transcriptomics**); the analysis of expression profile of proteins (referred to as **Proteomics**); and the analysis of small molecules called metabolites generated as a result of protein-protein and protein/nucleic acid interactions (referred to as **Metabolomics**), is increasingly being used to generate a system-wide view of a biological system (Dettmer and Hammock 2004). Metabolomics has been considered as the endpoint of this omics cascade (Kaddurah-Daouk et al. 2008), as metabolome is the terminal downstream product of the genome. It consists of the complete set of low molecular weight (<1500 Da) metabolites in a biological system, including cells, tissues, fluids, and whole organisms (Wishart et al. 2007a). Being the downstream product of the omics cascade (Figure 1.1) and eventual reaction to various pathophysiological stimuli or genetic modifications, the metabolome is highly dynamic and hence, can be considered as a reliable snapshot of the physiological state of an organism. In addition to providing snapshot of the physiological state of an organism, metabolites are also known to take part in various cellular pathways, e.g., glycolysis, pentose phosphate pathway, fatty acid biosynthesis, oxidative phosphorylation, Krebs cycle, etc., (Bruntz et al. 2017; Jalloh et al. 2015; Martínez-Reyes and Chandel 2020). This sometimes makes it difficult to understand if they are indeed cause or the effect of any particular disease (Mi et al. 2020; Nemet et al. 2020).

Metabolomics, an emerging and rapidly evolving technology, has been defined by Fiehn as ‘the qualitative and quantitative study of the metabolome in a biological system’ (Fiehn 2002). As explained before as well, it emphasizes on the study of a diverse range of low molecular weight (<1500 Da) endogenous and exogenous metabolites that include various organic species, such as, amino acids, fatty acids, nucleic acids, carbohydrates, organic acids, vitamins, polyphenols, and lipids (collectively called as metabolome). The information about these low molecular weight compounds present in the biological soup can be used in various aspects of scientific research, such as development of biomarkers for disease diagnosis,

monitoring the effects of medical interventions like medications, identification of food adulteration and understanding of the chemical diversity of a species based on geographical origin. In addition to this, metabolomics may provide not only disease-specific biomarkers but also provide deep insights into the etiology and progression of a variety of complex disorders. For example, biofluid metabolomics has been used to understand various human diseases, to examine biochemical pathways and their perturbations that arise due to mutations, ageing, diet, exercise, lifestyle changes, etc. (Beckonert et al. 2007; Markley et al. 2017; Wishart 2008, 2016). Metabolomics, therefore, seeks to provide detailed insights into the physiological state of an organism by analyzing the global metabolic profile of a biological sample in relation to genetic variations or external stimuli (Monteiro et al. 2013).

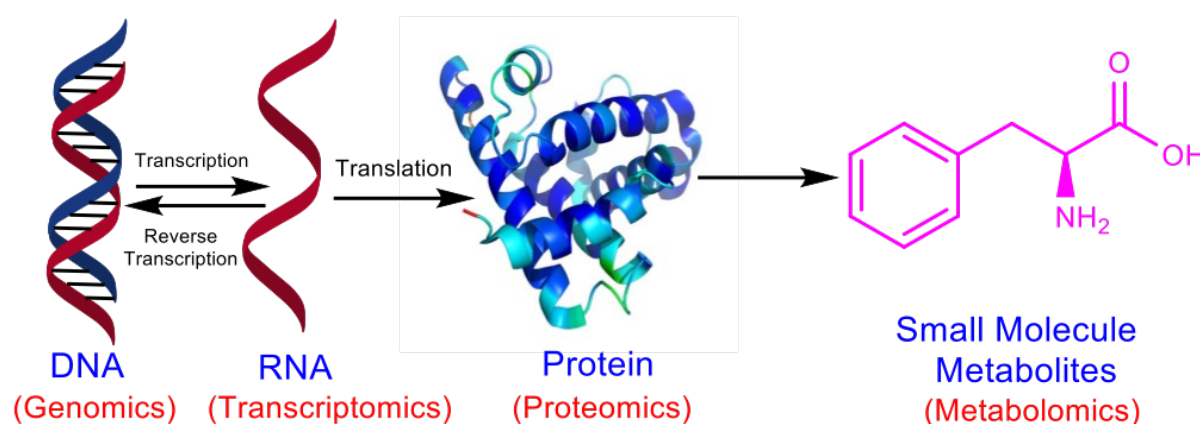


Figure 1.1: The central dogma of biology and the omics cascade.

1.2 Targeted and Untargeted Metabolomics

Two different approaches have been developed for performing metabolomics analysis – targeted and untargeted metabolomics (Fernández-Peralbo and Luque de Castro 2012). The choice of the approach depends on the kind of information that is sought from the metabolomic analysis.

In untargeted approaches or *global profiling approach*, significant metabolites are, by definition, unknown prior to analysis. This approach provides an unbiased overview of the entire metabolome and focuses on simultaneously measurement of as many metabolites as possible from biological samples (without any previous bias) in order to generate a metabolic fingerprint of a sample (Shulaev 2006). On the other hand, targeted metabolomics aims to measure a set of specific and pre-defined metabolites whose physico-chemical characteristics are known. This approach focuses on the absolute quantification of metabolites, that have been identified in advance, and which are highly related to a specific pathway (Bingol 2018). This

method has been characterised as a “hypothesis-driven approach” rather than as a “hypothesis-generating approach”. In this thesis, the untargeted metabolomics approach has been used to address different biological problems.

1.3 The metabolomics workflow

Independent of the field of application, metabolomics investigations need to follow a specific and systematic workflow in order to achieve appropriate and reliable results. The typical workflow involved in obtaining the metabolomic fingerprint of a biological system has been presented in Figure 1.2. The first and most important step is the formulation of a hypothesis or biological problem to be resolved by metabolomics. This is followed by the set-up of experimental design wherein the sample in question is collected and processed to extract the metabolites. The extraction methods need to be standardized to cover the maximum metabolites that can be obtained from a sample. This is followed by an analysis of the extracted metabolites using techniques that separate, identify and quantitate them. This, in turn, leads to the generation of large datasets, which are curated, and then analysed either manually or by pre-designed software packages, followed by rigorous processing of the data and analysis.

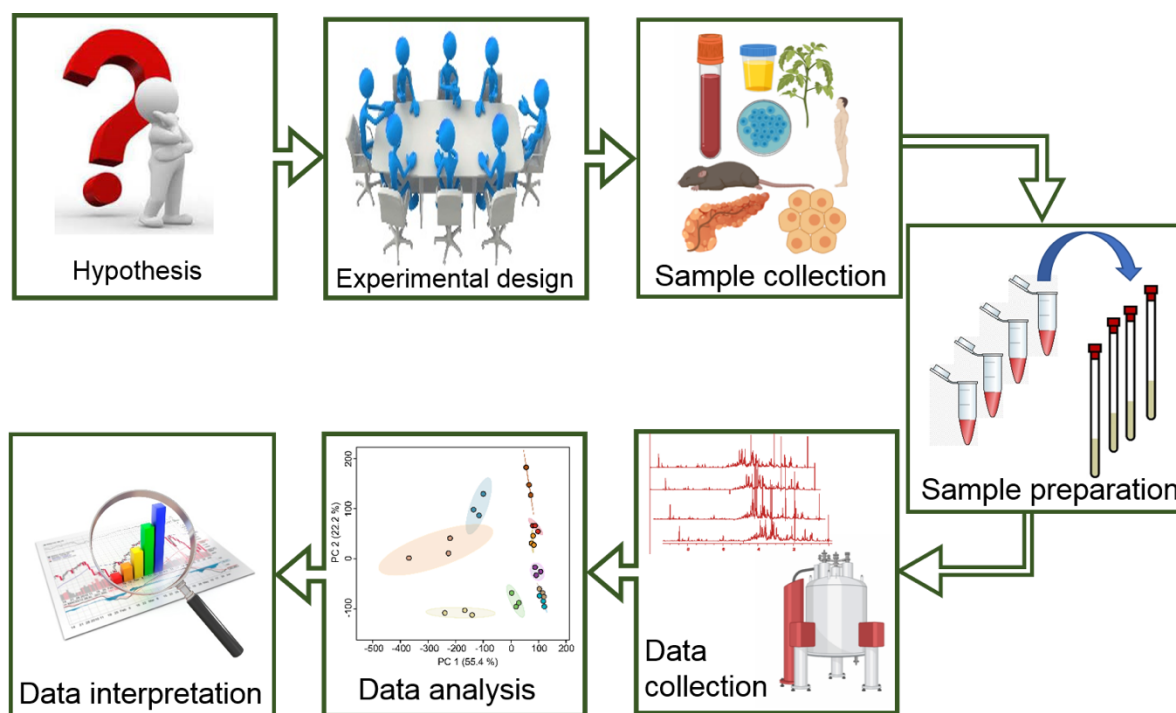


Figure 1.2: Workflow followed for conducting a metabolomics study: The basic steps involved in performing a metabolomics study include: hypothesis generation, experimental design, sample collection, sample preparation, data acquisition, statistical analysis –generally multivariate data analysis, followed by the interpretation of the results (i.e. identification of the major perturbed metabolic pathways).

1.4 Sample preparation

Sample collection, preparation, and storage are crucial steps since only a correct sampling method would provide a real snapshot of the metabolome at a given point in time. Thus, strict protocols both for the collection and the preparation of the samples have to be followed by all the operators involved in the project. For the most common analysed biofluids such as plasma, serum and urine, generally methods that involve minimal sample treatment are preferred, wherein samples are diluted, filtered and centrifuged to perform the extraction (Beckonert et al. 2007). On the other hand, in case of semi-solid or solid biological samples, extraction is performed after weighing and grinding in a pre-cooled mortar-pestle under liquid nitrogen and/or a pre-cooled blender (for harder tissues/samples) (Kim, Choi, and Verpoorte 2010). This is followed by the step-wise addition of reagents and solvents with different polarities to extract the metabolites (Kim and Verpoorte 2010). The performance of various organic solvents such as alcohol, acetonitrile, acetone, acid and their combination have been studied for metabolite extraction from biological fluids or tissues using NMR analysis. It has been shown that protein precipitation and metabolite extraction using methanol gives better results for metabolite concentrations as compared to other solvents. The use of methanol provides high extraction efficiency and minimal metabolite losses, and hence is potentially well suited for routine NMR based metabolomics studies (Nagana Gowda and Raftery 2014; Snytnikova et al. 2019). The entire extraction process must be followed at low temperatures (< -20 °C) in order to inhibit the enzymes responsible for the degradation of metabolites (Dunn and Ellis 2005; Pinu, Villas-Boas, and Aggio 2017). It is important to follow a unique standardized protocol for sample preparation in NMR-based metabolomic studies in order to avoid the identification of spurious biomarkers due to a general lack of reproducibility between laboratories (Emwas et al. 2015).

1.5 Analytical platforms

The most important analytical tools commonly used in metabolomics are NMR spectroscopy and Mass spectrometry (Djukovic, Nagana Gowda, and Raftery 2013; Dunn and Ellis 2005). Both these techniques have advantages and disadvantages, thus there is not a single analytical technique fully suited for metabolomic studies. Indeed, NMR and MS have been demonstrated to be complementary and powerful analytical approaches for the complete characterization of the metabolome (Pan and Raftery 2007).

Mass spectrometry (MS) is an analytical technique that separates the compounds in a biological sample based on their mass to charge ratios by taking into account the degree of deflection of charged particles in an electromagnetic field (Picó 2015). The higher sensitivity and selectivity of MS, with detection limits in the picogram range makes it an important technique for the metabolic profiling of biological samples (Pan and Raftery 2007). MS is often coupled with chromatographic techniques, as Liquid Chromatography-Mass Spectrometry (LC-MS) and Gas Chromatography-Mass Spectrometry (GC-MS) (Scalbert et al. 2009). MS-based metabolomics typically requires complex pre-processing of samples (i.e., via derivatization in GC-MS) that results in loss of many non-derivatised chemical constituents. The overall throughput of this platform is further hampered by many unsolved problems such as a non-uniform detection caused by variable ionization efficiency, lack of standardized protocols or procedures (as it requires optimization of separation conditions each time) and a lack of a universal database due to different ionisation mode, etc. (Aretz and Meierhofer 2016).

NMR spectroscopy is a highly powerful and versatile analytical technique in metabolomics since it allows the qualitative and quantitative analysis of chemical compounds from complex mixtures as well as the structural elucidation of unknown compounds (Huang et al. 2007; Kruk et al. 2017). NMR has an important role in metabolomics owing to the use of easy and rapid sample preparation methods, non-destructiveness of samples, lack of dependence on chromatographic separation and high degree of reproducibility (Clayton et al. 2006; Lindon et al. 2000). It is non-selective (i.e., it is not biased towards the detection of certain metabolites present in a biological sample), and can simultaneously analyse all highly abundant aqueous metabolites present in a biological mixture. NMR experiments generally require minimal sample preparation, often consisting only of pH adjustments and internal standard addition. Further, one of the greatest strength of NMR lies in its utility for the absolute quantification of metabolites as the NMR spectral integral of a peak shows direct proportionality to the molar concentration of the corresponding metabolites (Mahrous and Farag 2015). The metabolite concentrations can be calculated by comparing the area under each peak corresponding to a particular metabolite with that of the internal standard of known concentration, such as 3-trimethylsilylpropionic acid (TPS) or 2,2-dimethyl-2-silapentane-5-sulfonate (DSS). In view of this, a direct comparison of different metabolites is possible without the need for any calibration curve of each individual metabolite. NMR is, however, largely limited by its low sensitivity as it can only detect medium to highly abundant metabolites, typically anything in or above the micromolar range (Pan and Raftery 2007). Increased signal- to noise- (S/N) ratio and spectral resolution can be achieved by the

application of a higher magnetic field instrument and cryoprobe to reduce thermal noise (Kovacs, Moskau, and Spraul 2005). In addition, the use of small volume probes (1 mm and 1.7 mm tube probes), which require less sample volumes (less than 30 μ l), is able to provide several-fold increased sensitivity when compared to the conventionally used probes (3 mm or 5 mm cryoprobes). This is due to the fact that for a given mass of analyte, a reduction in the diameter of the radio-frequency (RF) coil increases the S/N ratio (Dalisay and Molinski 2009; Dona 2018; Hoult and Richards 1976; Schroeder and Gronquist 2006). In addition to this, isotopic enrichment of the metabolites through chemical reactions can be commonly applied to enhance the sensitivity of heteronuclear 2D NMR experiments. Other limitations of NMR experiments include the heavy overlap of spectral resonances as several endogenous metabolites may contribute to a signal; variation in chemical shift of resonances due to changes in pH, concentration, and ionic strength due to alterations in the acid-base equilibrium and solute-solute interactions; poor water suppression in case of dilute samples; baseline distortions; and chemical exchange between metabolites, particularly with water. In this thesis, we have performed metabolomics investigations in different biological systems using NMR spectroscopy.

1.6 NMR as a tool in metabolomics

In the following sections, a brief overview of typical NMR experiments (Figure. 1.3) that are being used extensively to analyze a mixture of metabolites has been provided.

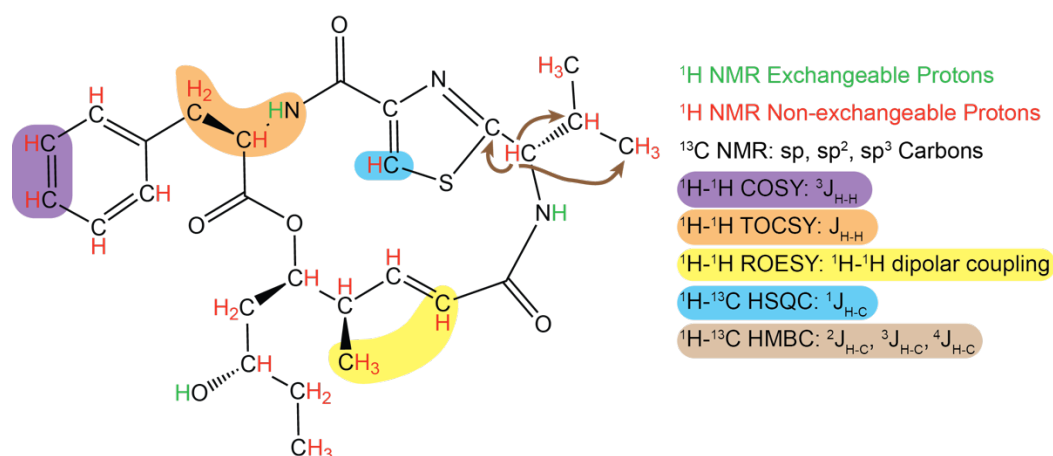


Figure 1.3: NMR experiments typically used for the analysis of metabolites. Using Antalid (a recently isolated secondary metabolite from *Polyangium* species) as an example (Tautz et al. 2016), the NMR experiments employed for the identification of metabolites and the corresponding information content from these experiments (highlighted) have been enlisted. [Adapted from (Yousf et al. 2017)].

1.6.1 ¹H NMR Spectroscopy

One dimensional (1D) ¹H NMR is the most robust and powerful technique used in the analysis of overwhelming majority of metabolomics studies (Larive, Barding, and Dinges 2015) performed to date because it is fully automated, efficient, and very rapid (Beckonert et al. 2007). In fact, the data collection times for ¹H NMR experiments can often be as short as a few minutes per sample. In addition to this, the information generated by the 1D ¹H NMR spectrum is often enough to identify and quantify 40–80 metabolites at a time in a given biological sample. Furthermore, 1D ¹H NMR experiments has the possibility to directly measure the metabolite concentration by integrating the peak using an internal standard of known concentration. In most cases, DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) or TSP (3-(trimethylsilyl)-2,2,3,3-tetradeuteropropionic acid) have been used as internal standards for quantitation of metabolites (Mallol et al. 2013). Besides quantification, they are also used for the calibration of NMR chemical shifts during spectral analysis.

In NMR-based metabolomics, solvent suppression is one of the important aspects that has the greatest impact on overall spectral quality. Solvent suppression is required during the measurement due to the large water signal to avoid signal overlapping and baseline distortions. The most popular water suppression schemes commonly used in NMR-based metabolomics include simple presaturation (pre-sat) which uses weak radiofrequency to saturate the water resonance during relaxation delay, 1-D Nuclear Overhauser Effect Spectroscopy (NOESY) presaturation with gradients (noesygppld), water suppression by WATERGATE that uses gradient tailored excitation (Piotto, Saudek, and Sklenář 1992; Sklenář et al. 1993) and excitation sculpting (ES) using pulse field gradient echo sequence (Hwang and Shaka 1995; Nguyen et al. 2007) for water suppression. Amongst these, noesygppld has been widely used to suppress water signal efficiently in case of biological samples, since it is highly robust. This pulse sequence (noesygppld) uses the first increment of a NOESY pulse sequence, with water presaturation and spoiler gradients during relaxation delay and mixing time and is in the form of $-RD-90^\circ-t-90^\circ-t_m-90^\circ-ACQ$, where RD is the relaxation delay, t is a short delay, 90° represents the RF pulse, t_m is the mixing time and ACQ is the data acquisition period. The noesygppld pulse sequence has emerged as the leading choice for NMR-based metabolomic studies as it provides good solvent suppression without rolling baselines.

Another important 1D NMR experiment commonly used in NMR-based metabolomics is CPMG (Carr-Purcell-Meiboom-Gill), that employs T2 filtering to attenuate the broad signals from macromolecules, mainly proteins and lipids, which have short T2 relaxation time

compared to metabolites (Carr and Purcell 1954). CPMG experiment is based on the spin-echo pulse sequence and consists of $-RD-90^{\circ}_x-(t_E-180^{\circ}-t_E)_n-ACQ$, where t_E represents the *echo time* and n is the number of repetition of the block in parenthesis. By carefully optimising these two parameters, the resonances in the spectrum can be separated according to their spin-spin relaxation time (T_2).

1.6.2 2D NMR Spectroscopy

The main drawback of 1D NMR experiments is the extensive signal overlapping that seriously limits the clear identification and quantification of metabolites during metabolite analysis. For this reason, two dimensional (2D) NMR experiments that provide additional dimensions are carried out in order to alleviate the congestion of 1D spectrum. The benefits of using 2D NMR spectroscopy include the reduced overlap of spectral peaks allowing unambiguous identification of the compounds in the mixture, additional information regarding the resonance multiplicity, and coupling patterns. The most commonly used experiments in metabolomic studies include 1H - 1H COSY (COrrrelation SpectroscopY), 1H - 1H TOCSY (TOtal Correlation SpectroscopY), 2D J-resolved spectroscopy (J-Res), 1H - ^{13}C HSQC (Heteronuclear Single-Quantum Correlation spectroscopy), and HMBC (Heteronuclear Multiple-Bond Correlation spectroscopy).

1.6.2.1 1H - 1H Correlation Spectroscopy (COSY)

2D 1H - 1H correlation spectroscopy (COSY) is the simplest experiment of all 2D NMR experiments used to identify homonuclear correlations between coupled 1H nuclei. During the COSY experiment, protons with mutual spin-spin coupling correlations are observed and are displayed in the form of cross-peaks in the COSY spectrum (Figure 1.4B). These cross-peaks in the 2D spectrum indicate pair of nuclei connected through scalar coupling on the same carbon (germinal coupling, $^2J_{H-H}$), or on adjacent carbon (vicinal coupling, $^3J_{H-H}$), or in rare cases protons connected via four/five bonds (long-range COSY) (Jacobsen 2007) (Figure 1.4A). There are several variants of COSY that have been developed to improve the resolution of the cross-peaks and to reduce the acquisition time.

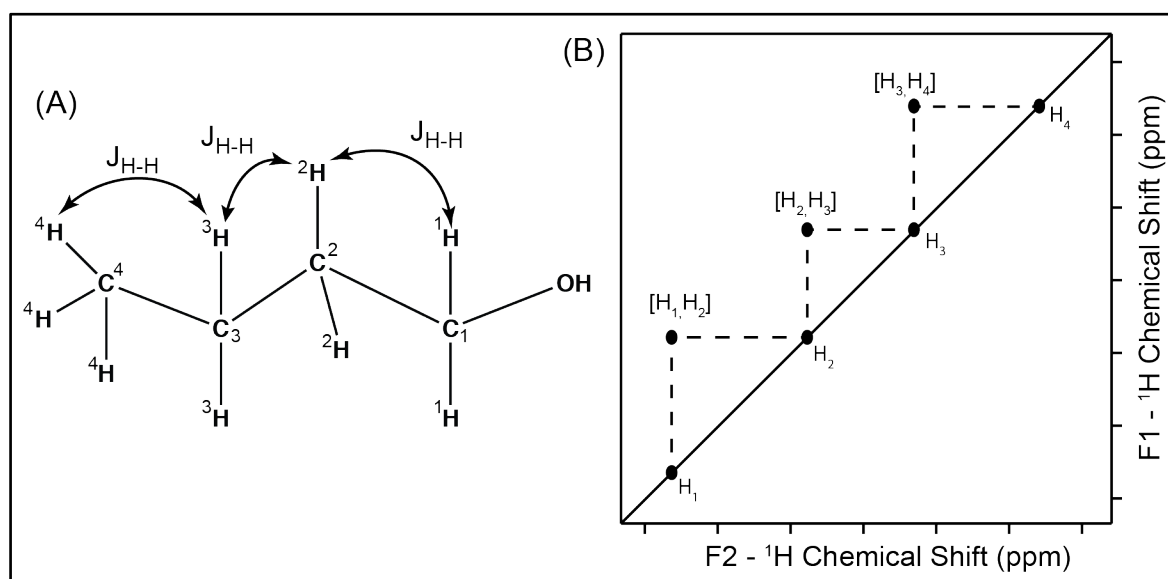


Figure 1.4: 2D ^1H - ^1H Correlation Spectroscopy (COSY). A) ^1H - ^1H correlations that can be observed as a cross peak in COSY spectrum have been connected by arrows in a small molecule taken as an example. B) A typical COSY spectrum: Diagonal peaks are observed at the respective chemical shifts of each proton and can be read in either F1 or F2 dimension. Cross-peaks are observed between the proton pairs marked by arrows in (A) that read ^1H chemical shifts of the respective protons in F1/F2 dimension. For simplicity, cross peaks have been shown only on one side of the diagonal. Adapted from (Yousf et al. 2017).

1.6.2.2 ^1H - ^1H Total Correlation Spectroscopy (TOCSY)

Another powerful ^1H - ^1H correlation spectroscopy is the Total Correlation Spectroscopy (TOCSY), also known as the homonuclear Hartmann–Hahn (HOHAHA) experiment. In a typical TOCSY experiment, all protons involved in the same spin system are detected and are shown either in 1D mode or most commonly in the form of cross peaks in 2D TOCSY spectrum (Jacobsen 2007) (Figure 1.5B). In a metabolomics study, TOCSY experiment is used to provide the complete assignment for proton resonances of sugars and peptides, as all the protons that are correlated in the same sugar residue or in a single amino acid will get detected (Adell et al. 1997). In the 1D version of a TOCSY experiment, a particular signal with a good signal- to- noise ratio is selected in the spectrum and the other corresponding signals that appear correlated with the selected signal can be determined by applying excitation pulse on the selected signal (Figure 1.6). 1D TOCSY is very fast (can be recorded in one or two minutes) providing a relatively simple 1D NMR spectrum which is easier to analyse

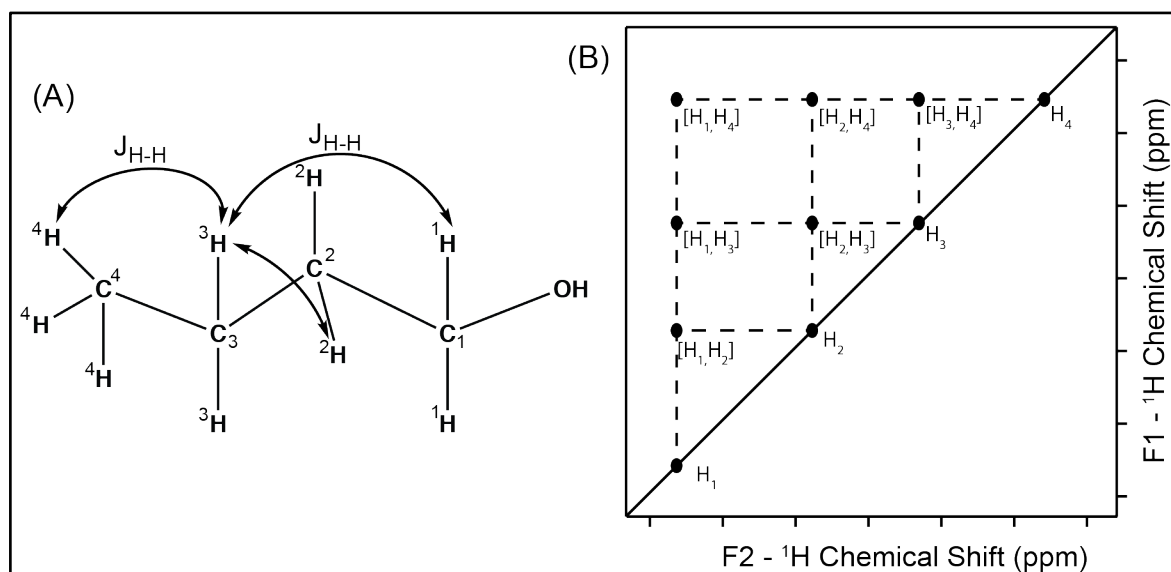


Figure 1.5: 2D ^1H - ^1H Total Correlation Spectroscopy (TOCSY). A) A set of ^1H - ^1H correlations that show a cross-peak in TOCSY are connected by arrows in a small molecule taken as an example. For example, H_3 shows a cross-peak with H_1 , H_2 and H_4 . B) A typical TOCSY spectrum; F1 (corresponding to t_1 period) and F2 (corresponding to t_2 period) dimensions both read ^1H chemical shifts. Diagonal peaks are observed at the respective chemical shifts of each proton and can be read in either F1 or F2 dimension. Cross-peaks are observed between the proton pairs marked by arrows in (A) that read ^1H chemical shifts of the respective protons in F1/F2 dimension. For simplicity, cross peaks have been shown only on one side of the diagonal. Adapted from (Yousf et al. 2017).

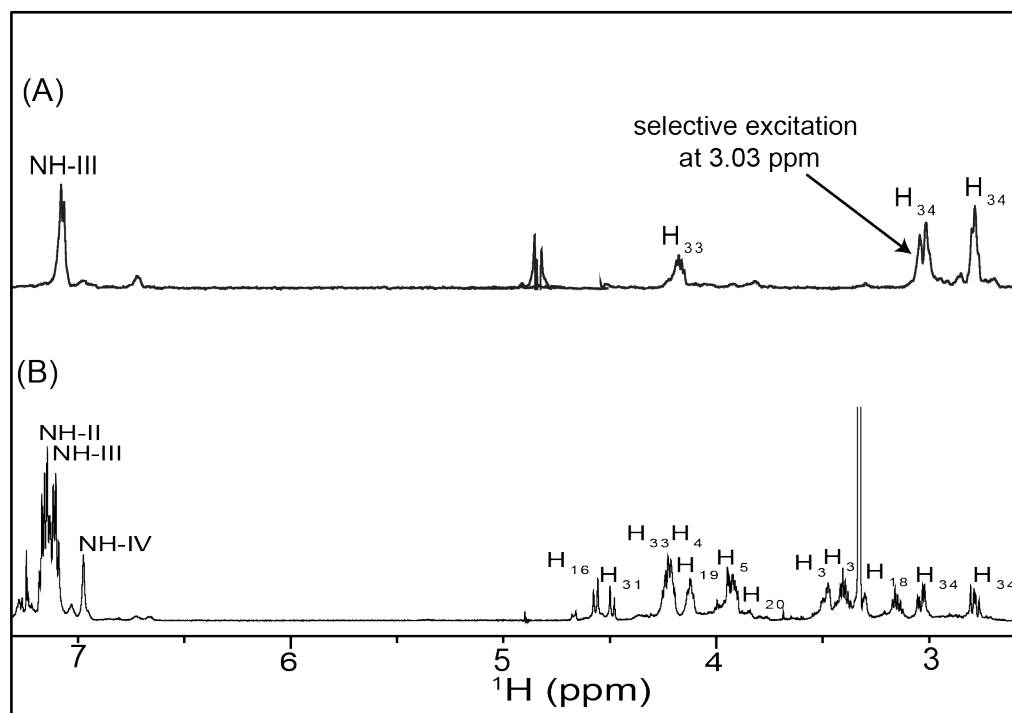


Figure 1.6: 1D TOCSY spectrum of glycoconjugates in MeOH-d_3 . (A) 1D-TOCSY sub spectra obtained by selective excitation of the signals at $\delta \text{H } 3.03 \text{ ppm}$. (B) ^1H -NMR spectrum control.

1.6.2.3 2D J-resolved NMR Spectroscopy (J-RES)

The 2D J-resolved spectroscopy (J-Res) experiment has become a popular method for a wide range of NMR-based metabolomic studies because of its simplicity and short acquisition time (Ludwig and Viant 2010). It is a 2D homonuclear experiment in which proton spectrum (chemical shift) is displayed in one dimension (F2) and the coupling constant (J-couplings) of each peak is presented in a second dimension (F1) (Huang et al. 2015) (Figure 1.7B). A skyline projection of F2 yields a proton-decoupled 1D spectrum that highly reduces the spectral congestion and thus facilitates the assignment and quantifications of metabolites. Unlike chemical shifts, J-couplings are resistant to physiological factors (such as temperature or pH value) and thus simplifies the resonance assignment, especially for low-intensity peaks which may be completely hidden by the overlapping of main peaks. Additionally, details about the J-value can be used to differentiate between certain isomers, such as α and β anomers of sugars and glycosides or cis- and trans- isomers of olefinic compounds (Mahrous and Farag 2015).

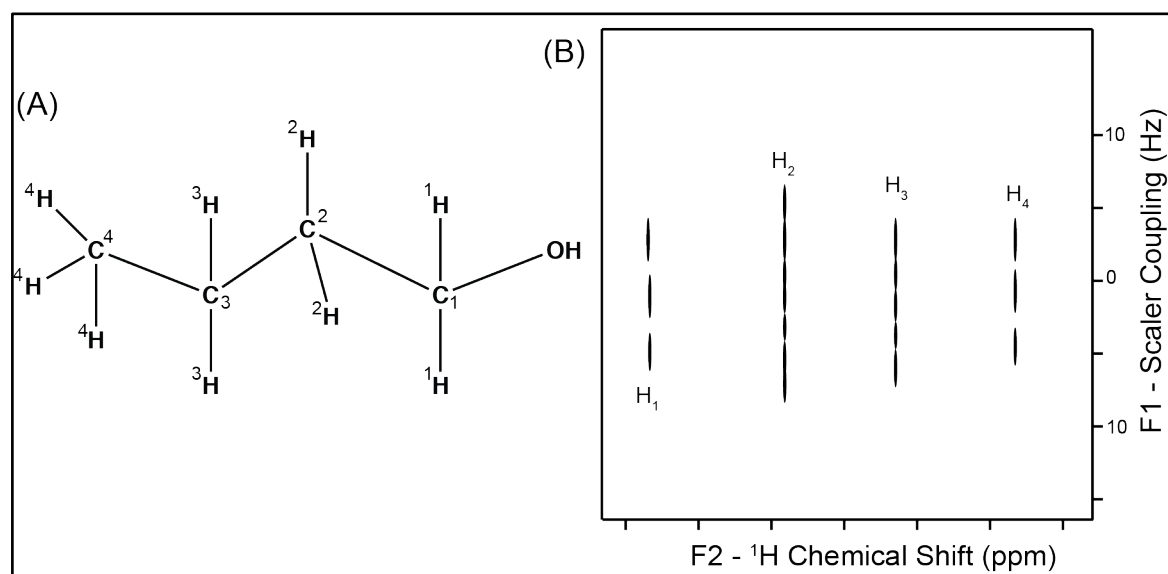


Figure 1.7: 2D J-resolved NMR Spectroscopy (J-RES). A) A small molecule taken as an example. B) A typical J-resolved spectrum of a small molecule (A) taken as an example. F2 dimension reads ¹H chemical shift and F1 dimension reads scalar coupling of fine structure. H₁ proton shows a triplet in the spectrum indicating it is next to a group having two equivalence protons in the molecular structure. Adapted from (Yousf et al. 2017).

1.6.2.4 2D ^1H - ^{13}C Heteronuclear Single Quantum Coherence (HSQC)

The 2D ^1H - ^{13}C heteronuclear NMR spectroscopy in metabolomics represents a valuable tool for resolving and assigning overlapping proton peaks, especially for metabolite signals resulting from complex biological samples (Mahrous and Farag 2015). The 2D correlation spectroscopy such as Heteronuclear Single Quantum Coherence (HSQC), maps a single-bond proton and carbon correlation (Figure 1.8B) is used for identification of molecules through complete resonance assignments. The HSQC experiment provides the highest resolution compared to the other 2D experiments. Nevertheless, the increased resolution is associated with a reduction in sensitivity, especially at the relatively low natural abundance of ^{13}C . To increase the sensitivity, isotopic labelling with ^{13}C is carried out in order to increase the abundance of ^{13}C . This can be done either by using ^{13}C glucose in nutrient media or by performing chemical reaction after the collection of samples (Buescher et al. 2015; Klein and Heinzle 2012).

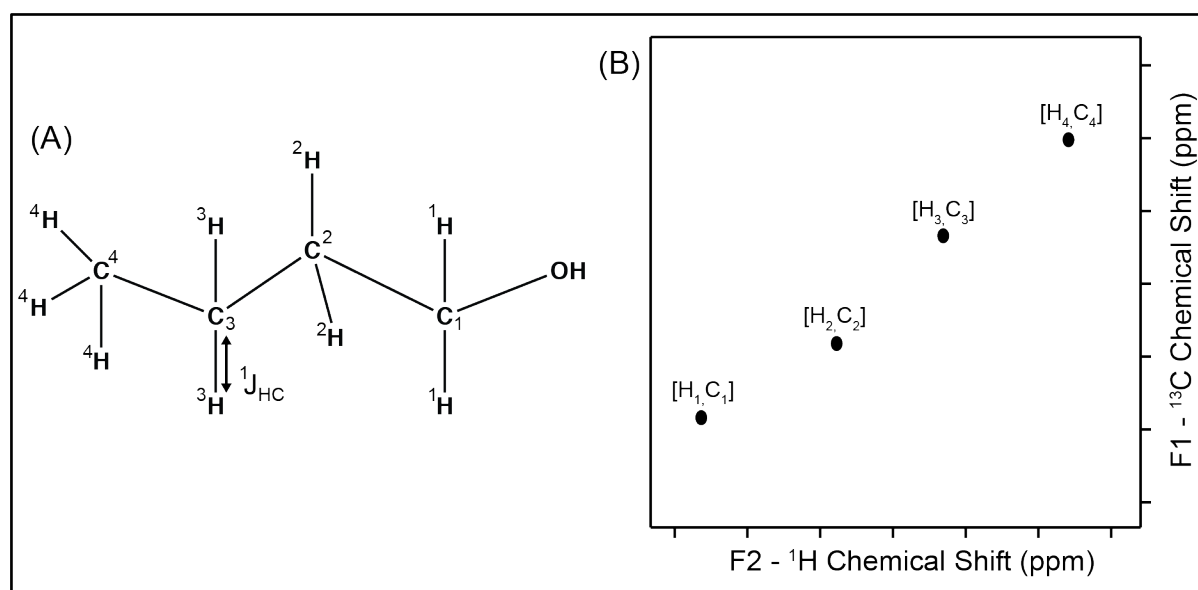


Figure 1.8: 2D ^1H - ^{13}C Heteronuclear Single Quantum Coherence (HSQC). A) One-bond ^1H - ^{13}C correlation that shows a cross-peak in HSQC is connected by an arrow in a small molecule taken as an example. For example, H₃ shows a cross-peak with C₃. B) A typical HSQC spectrum; F1 (corresponding to t₁ period) reads ^{13}C chemical shifts and F2 (corresponding to t₂ period) dimension reads ^1H chemical shifts. Unlike homonuclear experiments one does not observe diagonal peaks in heteronuclear experiments. Cross-peaks are observed between a directly connected ^1H and ^{13}C pair marked by arrows in (A) that reads ^1H chemical shift in F2 dimension and ^{13}C chemical shift in F1 dimension. Adapted from (Yousf et al. 2017).

For metabolite analysis, crucial NMR data acquisition parameters must be optimised in all aforesaid NMR experiments for different biological samples. This includes spectral width, time-domain points, number of scans, receiver gain, inter-scan delay and the acquisition time.

1.6.3 Recent advances in NMR spectroscopy

The main limitation of 2D NMR experiments is the relatively longer acquisition time. To overcome this problem, several approaches have been suggested in recent past, which continue to evolve, for the reduction of NMR experiment measurement time. Among which the two of the methods that show great promises, in a quantitative metabolomics context, are: 1) non-uniform sampling (NUS) which acquires only a subset of the data points in the indirect dimension and remaining points can be predicted using novel reconstruction methods that ultimately allow the extraction of complete sets of chemical shift information (Le Guennec et al. 2015; Mobli and Hoch 2014); and 2) Ultrafast (UF) 2D NMR, the only current approach that makes it possible to record a 2D spectrum in a single scan (Giraudeau and Frydman 2014a). In ultrafast experiments, instead of repeating N successive experiments on a sample with an array of independent time increments, the sample is divided into N virtual slices, where the spins located at different Z-positions undergo different evolution periods, but happens simultaneously for all the spatial slices in the same scan (Figure 1.9) (Giraudeau and Frydman 2014b). In addition, rapid data collection using multiple NMR receivers have been used in metabolomics for acquiring multiple NMR spectra required for resonance assignments simultaneously in a single data set. For instance, in metabolomics the most commonly used 2D NMR experiments are: 2D ^1H - ^1H TOCSY, 2D ^1H - ^{13}C HSQC and 2D ^1H - ^{13}C] HSQC-TOCSY (Dubey, Pudakalakatti, and Atreya 2019). These three spectra can be visualised as three orthogonal projections (2D planes) of 3D [^1H - ^{13}C] HSQC-TOCSY (Figure 1.10). By combining all the three 2D planes together provides complete 3D spectral information. Thus, a 3D spectrum, which requires few days for data recording can be replaced by recording the three 2D projections, which can be acquired rapidly with high resolution. The main problem in implementing these experiments is the requirement of specialized hardware and software technology, and technically involved acquisition schemes (Kupce 2015).

The resolution of ^1H NMR spectra of complex mixtures could be significantly boosted using “Pure Shift“ methods (Adams 2014; Zangger 2015). Pure shift methods also known as broadband homonuclear decoupling are particularly used to enhance the spectral resolution in the direct proton dimension, which is typically crowded due to the limited proton chemical shift range (10-15 ppm) and extensive signal splitting due to proton–proton scalar coupling. Pure shift methods removes all homonuclear scalar couplings by collapsing all multiplets to singlets, simplifying spectra with improved resolution along the proton dimension (Timári et

al. 2019). However, these methods typically come with a significant loss in sensitivity and are not commonly applied to the metabolomic data analysis (Giraudeau 2020).

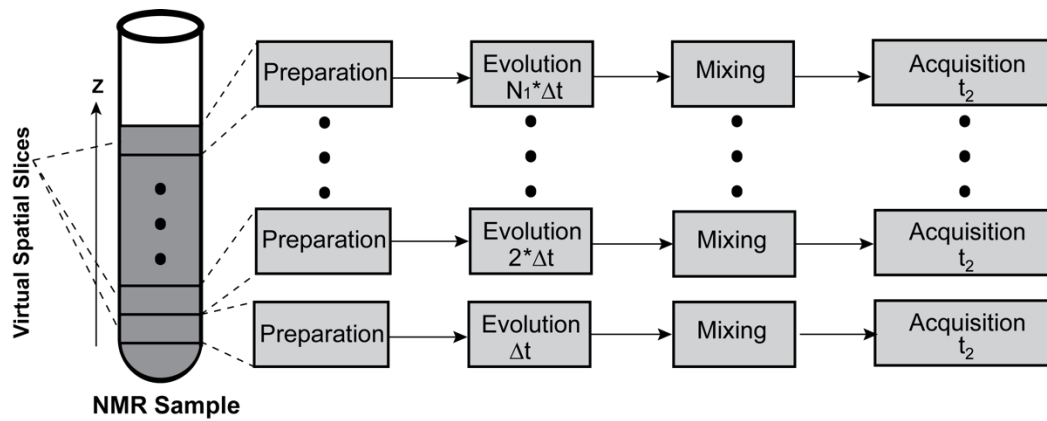


Figure 1.9: Ultrafast acquisition strategy. In ultrafast experiments, the NMR sample is divided into spatial slices along Z-dimension using gradients and each slice is then evolved to a different time during first evolution period. Adapted from (Yousf et al. 2017).

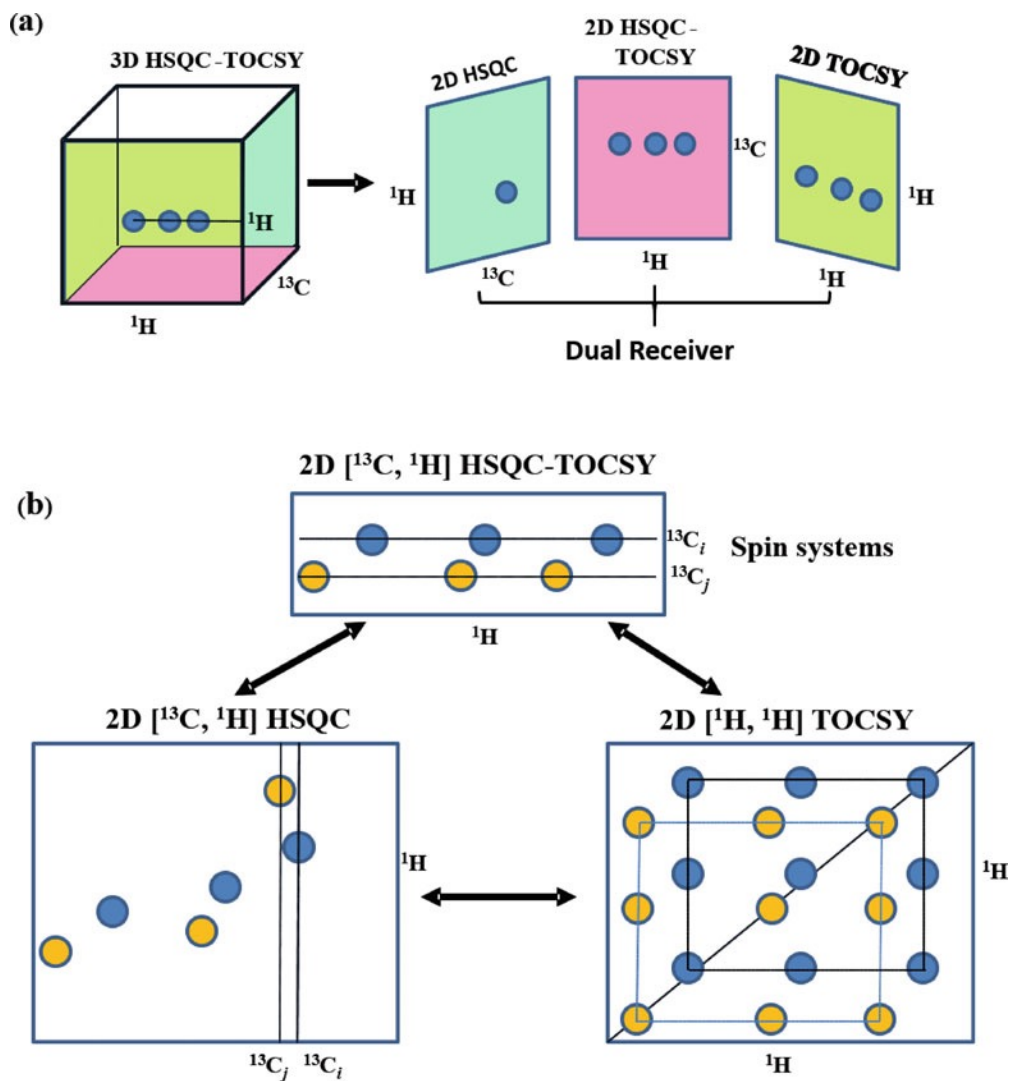


Figure 1.10: (a) The 3D ^1H - ^{13}C HSQC-TOCSY spectrum. The 3D ^1H - ^{13}C HSQC-TOCSY spectrum can be visualized as a combination of three 2D experiments: 2D ^1H - ^{13}C HSQC, 2D ^1H - ^{13}C HSQC-TOCSY and 2D ^1H - ^1H TOCSY. (b) The peak pattern observed in the three 2D spectra is illustrated for two spin systems. Adapted from (Dubey et al. 2019).

1.7 Post-acquisition processing of NMR spectra

Prior to chemometric analysis, a number of processing steps have to be performed before and after the Fourier transformation (FT) of the FID. The key stages of NMR data processing have been briefly summarised in the following sub-sections.

1.7.1 Chemical shift referencing

The acquired NMR spectra needs to be properly referenced with an internal chemical shift standard prior to any qualitative and quantitative analysis. Chemical shift referencing is important for resonance assignment of metabolites, spectral alignment and thus any statistical

analyses. For most metabolomics studies, referencing standards including trimethylsilyl propanoic acid (TSP) and 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) are widely used for chemical shift referencing (Dona et al. 2016). Also, the chemical shift referencing with respect to solvent peaks (i.e., water) must be avoided since its signal position is very sensitive to various sample parameters including pH, salt, exchangeable moieties, and temperature.

1.7.2 Phase correction

Phase correction is an NMR spectral adjustment procedure designed to increase the absorptive character and symmetry of all NMR signals across all regions of an NMR spectrum. It is generally generated by two main phenomena: (i) temporal delays between the RF pulse and the opening of the receiver for the FID acquisition; and (ii) off-resonance effects due to inability of the pulse to excite all the nuclei equally. This is considered as an essential spectral processing step since even a minor error in phasing can result in severe problems that will cascade down through all remaining processing steps, especially quantification. A spectrum that is not phase-corrected has a signal with a dispersive line shape, as well as inverted signals. As a result, a zero-order and first-order phasing adjustments are required to deal with this problem. Zero-order correction is frequency independent and thus, the correction applied in phase factor is same across the spectrum for all the lines. Whereas the first-order correction is frequency dependent and applies a phase change, which increases linearly with the distance from a “**pivot**” point where the first-order phase correction is zero.

1.7.3 Baseline correction

The existence of the baseline artefacts is one of the most ubiquitous problems in FT-NMR spectra that could adversely affect the identification and quantification of NMR signals. It is a crucial feature in a metabolomics experiment. Indeed, a flat baseline is fundamental for an accurate integration and thus, the quantification of the chemical compounds. Furthermore, small peaks could be hidden under a distorted baseline. This issue is usually caused by the corruption of the first few data points in FID that add low frequency modulations in the Fourier-transformed spectrum, and thus forms a distorted baseline. Correcting such distortions is an important step in the processing of NMR spectra. Baseline correction models the baseline by fitting it to polynomial, sine, or exponential functions. Even little baseline distortions lead to large errors in the quantification of metabolites of low abundance.

1.7.4 Zero filling

Zero filling is a data processing technique which is always done prior to Fourier Transformation to increase the digital resolution of the NMR spectra by increasing the FID data points. Moreover, it is essential to point out that zero filling does not increase the true resolution, it only enhances the apparent resolution that might be helpful in visualising the fine coupling. Zero filling is often applied to at least double the original FID data size i.e., twice the number of experimentally collected data points.

1.7.5 Apodization

Apodization is a digital filtering manipulation of the NMR spectra that consists of multiplying the FID with different window functions (Lorentzian, Exponential, Gaussian, or Sine-bell function) prior to Fourier transformation. Window apodization is applied to the FID either to enhance the signal-to-noise ratio (S/N), usually with a Lorentzian function or to improve resolution, using a shifted sine-bell or a Gaussian function.

1.8 NMR data pre-processing for chemometric analysis

NMR data are not readily suitable for the analysis by chemometric methods. Therefore, these data need to be transformed into a format that is suitable for carrying out multivariate analysis. This can be achieved through several steps: binning, alignment, normalisation, and scaling.

1.8.1 Binning

Binning consists of segmenting of an NMR spectrum into a number of small regions or buckets (usually 0.02-0.04 ppm) that are large enough to have one or more NMR peaks in it. The bin intensity is then calculated by means of integrating signal intensities within each bin. One of the important advantage of binning is that it prevents the NMR shift variations from affecting statistical processing, provided such shifts remain within the borders of corresponding bins (Emwas et al. 2018). Binning approach generally tends to decrease the complexity of the data by reducing the number of variables and thus makes it more manageable. After binning, the statistical analysis can then be conducted on the extracted bin intensities, and the peaks are then assigned to metabolites. Binning method requires less effort and less time. The major disadvantage of the equidistant binning is the lack of flexibility in situations where peaks split between two adjacent bins, thereby resulting in a dramatic loss of resolution. To overcome this

problem, several other approaches can be used, such as adaptive-intelligent binning and dynamic adaptive binning, which dynamically determines the size and location of each bin. (Anderson et al. 2011; Davis et al. 2007). Adaptive binning approach identifies bin edges (the beginning and end of each peak) in the existing bins using undecimated wavelet transform.

1.8.2 Spectral alignment

Spectral alignment is a method that iteratively adjusts NMR peak positions in multiple spectra in order to directly overlay or align the peaks corresponding to the same compounds. However, it can happen that a peak belonging to the same analyte, and thus expected at the same chemical shift in all the samples, changes its position across the spectra, even when properly referenced. This is known as chemical shift drift and the most common reasons for this phenomenon arise mainly due to changes in experimental conditions such as pH, temperature, ionic strength, physicochemical interactions, etc. Such chemical shift drifts have been known to create variation in the ppm values across different NMR spectra. Even slight changes in the NMR spectra can deeply influence the statistical analysis and thus complicate the discovery of biomarkers or the pattern of metabolic profiles. Therefore, to correct local signal shifts, it is preferable to perform a spectral alignment that will allow a good chemometric modelling. Some of the examples of alignment algorithms include, interval correlation shifting (icoshift) (Savorani, Tomasi, and Engelsen 2010), Correlation Optimised Warping (COW) (Nielsen et al. 1998) and Recursive Segment-Wise Peak Alignment (RSWPA) (Veselkov et al. 2009).

1.8.3 Normalisation

After the metabolite identification and quantification, the next step in the data post-processing pipeline is to normalize the data in order to remove any variation in signal intensities due to dilutions of each sample or the amount of material analysed rather than changes in metabolic responses. Normalization is a table row operation in the data matrix (containing N observation row vectors of various samples having K variables each in columns where variable can be NMR peak intensity or NMR peak volume) that aims to make spectra comparable with each other (Worley and Powers 2013).

Different normalisation methods mostly used in metabolomics studies include:

- 1) Normalization with respect to an internal standard
- 2) Normalization to a particular reference feature
- 3) Normalization to total intensity or total area

4) Probabilistic Quotient Normalization (*PQN*)

5) Sample-specific Normalisation

Normalization to total or total integral is the most common method of normalisation in NMR-based metabolomic studies, which assumes that the total sum of the intensities of each spectrum is constant (Emwas et al. 2018). This algorithm divides the individual bin integral by the total integrated intensity for each sample. This method is used to normalize deviations between spectra due to sample concentration/dilution. Nonetheless, this approach has been reported to have robustness and accuracy limitations when the samples contain extreme quantities of one or few metabolites. To overcome this issue, Probabilistic Quotient Normalization (*PQN*) approach has been developed, which is based on the assumption that changes in concentration of single metabolites only affect certain portions of the NMR spectra, whereas dilution effects will influence the complete spectrum (Dieterle et al. 2006). This algorithm relies on the determination of most probable dilution factor for all signals of the spectrum compared to a reference spectrum.

When undertaking any normalisation, it is important to consider eliminating NMR peaks considered to be irrelevant to biological effects, such as solvent peaks, the intensities of which can interfere with this procedure.

1.8.4 Scaling

Scaling, a column-wise operation on the variables in the matrix defined above, is a data pre-treatment method that divides each variable by a factor, the scaling factor, that is different for different variables (Hendriks et al. 2011). Generally most abundant metabolites display greater variances among samples that may obscure smaller, but biologically important changes in low-abundant metabolites. Scaling aims to remove the bias towards the most abundant metabolites in the data analysis in order to make data more normally distributed. The most commonly used data scaling methods in multivariate analysis include Autoscaling and Pareto scaling (van den Berg et al. 2006a; Keun et al. 2003).

1.8.5 Mean centering

Prior to scaling, a column-wise operation called mean centering is used to convert all concentrations such that they fluctuate around zero rather than mean value (Smolinska et al. 2012). This is achieved by subtracting the column mean intensity from each individual intensity values, thus allowing the correction for the gap among metabolites present in high and low abundances.

1.8.6 Autoscaling

Autoscaling, also referred as standardization or Unit Variance Scaling (UV), is a method that converts all the variables to have unit variance (van den Berg et al. 2006b). This approach of scaling uses standard deviation as the scaling factor, dividing each value in a column by the standard deviation after mean centering and thus making all variables equally important, including noise. However, this method blows up baseline noise and can be a concern in NMR applications where noise covers major portions of the spectra.

1.8.7 Pareto scaling

Pareto scaling, a highly recommended scaling approach for the metabolomics study, is achieved by dividing each variable by the square root of the standard deviation (van den Berg et al. 2006b). This is quite similar to autoscaling. In this approach, large fold changes in the metabolic data are downscaled more than small fold changes, thus it significantly reduces the former's effect while comparing with the clean data (Yang et al. 2015). This approach is considered as an intermediate between the extremes of no scaling (only mean centering), where medium features (low abundance metabolites) are overwhelmed by big ones (high abundance metabolites), and autoscaling that enlarges baseline noise.

1.9 Univariate Analysis

Univariate analysis is the simplest form of data analysis where only one variable is involved while analysing the data (Sandilands 2014). Because it only includes one variable, therefore it does not deal with causes or relationships (unlike regression). The major aim of univariate analysis is to describe the data and to find patterns that exist within it. The univariate analysis uses one dependent variable – the outcome; and one independent variable – the intervention. To assess the significance of the difference between the means of two independent sets of data, the Student's T-test is commonly used. For data that contains more than two independent groups, one-way analysis of variance (ANOVA) can be applied to assess the difference between group means of each variable. Both t-test and ANOVA are parametric tests that rely on the assumption of normality and equality of variances between groups (Horne 1998). If, in any case, the assumption of normal distribution for the sample population is not met, non-parametric analysis – such as the Wilcoxon rank sum test, can be used to evaluate the difference between two independent samples by comparing their medians (Gauthier and

Hawley 2015). Kruskal-Wallis one-way analysis of variance is the nonparametric equivalent of ANOVA comparing data with multiple groups.

1.10 Multivariate Analysis (MVA)

Multivariate analysis (MVA) methods are crucial in metabolomics studies since all metabolites are considered simultaneously, allowing the detection of trends between samples and metabolites, as well as within samples and metabolites (Worley and Powers 2013). MVA is performed to find the metabolite patterns and relationships in the data. MVA is highly graphical in its approach, provide tools to visualize the relationships between samples and variables. This approach is also used for predicting behaviour and improving forecasting of likely outcomes using predictive tools.

The basic categories of analysis techniques are:

- 1) *Exploratory analysis*: unbiased overview of the data where patterns and outliers can be easily detected.
- 2) *Classification and discrimination*: allows to discriminate between groups and to find biomarker candidates.
- 3) *Regression analysis*: models the quantitative relationship between two blocks of data.

The selection of the type of analysis mainly depends on the scientific goal.

1.10.1 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is one of the most commonly used approaches in metabolomics. It is an unsupervised projection method commonly used in chemometrics to reduce multidimensional data complexity in order to visualize and interpret relations between samples and between studied variables (e.g. levels of metabolites) (Hotelling 1933). This method summarizes the variation within a dataset by a smaller number of (uncorrelated) variables, called principal components (PCs) (Jolliffe 2005). PCs are linearly weighted combinations of the original variables calculated in such a way that each PC consecutively models the maximum variation in the data and at the same time is orthogonal to the other PCs.

PCA always begins with a normalised data often expressed as a data matrix with N rows (*observations*) and K columns (*variables*) (Wold, Esbensen, and Geladi 1987). The observations can be analytical samples, biological samples from different individuals, batch processing time points, and so on. The columns can be spectral or chromatographic variables. Although theoretically a number of PCs can be obtained from the calculation, the truly effective

PCs that generates the most of the relevant information are usually first few of them, while the other ones are often computed considering chance variation and noise. The outcome of this analysis consists of two plots: a ‘*scores plot*’, that displays the distribution trend of data points, in which each point represents a single spectrum (a single sample), and a ‘*loadings plot*’ that shows the variables of the spectrum (e.g., metabolites). The similarities and differences between samples are reflected by the clustering pattern (aggregation or dispersion trend) of samples in the score plots. The aggregation of points in the score plot indicates similarity of observation variables, whereas the well separate clustering of points represents the significant difference of observation variables. Both score and loading plots always need to be inspected together since the directions in the score plot correspond to direction in the loading plot. Thus, in order to understand the reason of a particular grouping observed in the score plot, it is needed to look at the same direction in loadings.

A more appropriate way to explain the PCA is possible using the mathematical interpretation of this method, as illustrated in Figure 1.11.

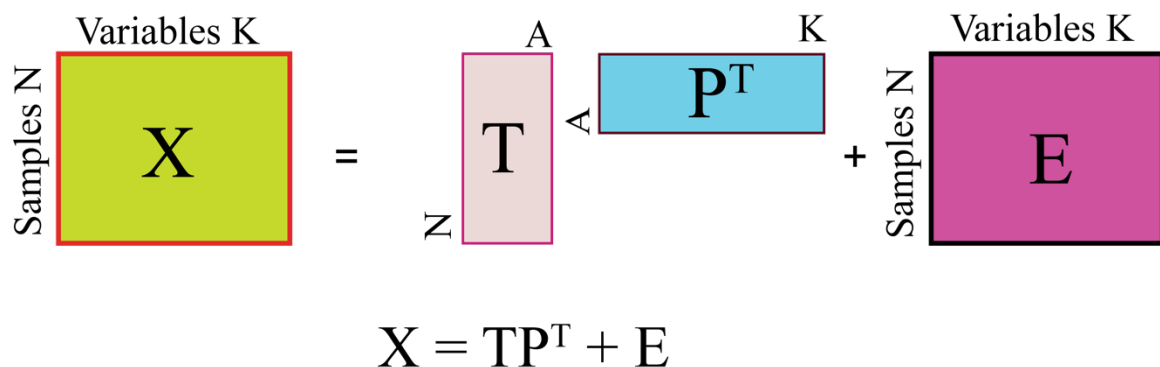


Figure 1.11: Schematic illustration of a PCA model. T and P represent the scores and loading matrices, while E is the residual matrix. N and K are the number of samples and variables respectively. A is the number of Principal Components chosen to build the model. Redrawn and adapted from (Bro and Smilde 2014).

The data matrix is decomposed into a structure part and a noise part:

$$X = TP^T + E = \text{Structure} + \text{Noise} \quad (1.1)$$

Where, T is the scores matrix, P the transposed loadings matrix, and E the residual matrix. Another way to clarify the scores definition is to consider them as element of the T matrix, where each row is an observation while each column represents the value that the observation has along each Principal Component (Bro and Smilde 2014).

Loadings, also called weights, allow one to understand the influence of the original variables on the scores T . The loading plot (Figure 1.12) shows the loadings of a certain PC; also, in this case, it is possible to plot the loadings of a PC against the loading of another PC.

The residuals are not part of the model; thus this part should be “small” in order to not remove too much information from the original dataset. If more PCs are included in the model, higher will be the variance that it is able to explain, thus smaller will be the residuals, following this calculation:

$$\text{Explained Variance} + \text{Residual Variance} = 100\% \quad (2.2)$$

However, retaining too many PCs has the drawback to include not relevant information in the model, thus affecting its goodness.

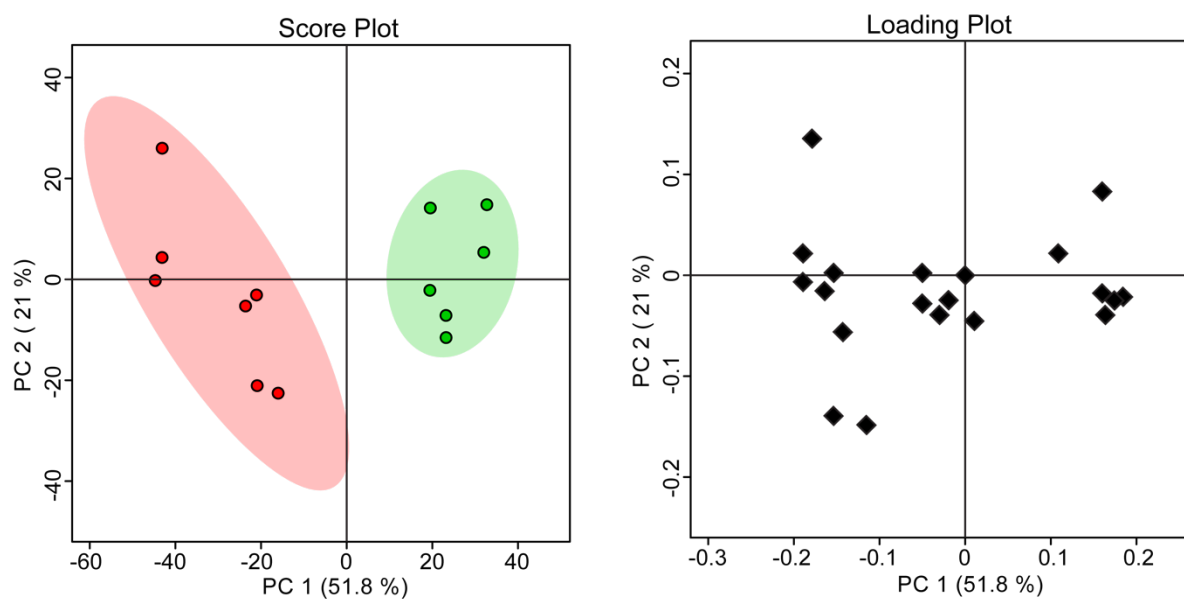


Figure 1.12: Representative Score and Loading plot of the first two components of a PCA model. Each point (Red and green circles) in the score plot represent samples, while each point (black diamonds) in the loading plot represent the variables (loadings) that are responsible for the observed clustering trend observed in the score plot.

1.10.2 Partial least squares for discrimination (PLS-DA)

Partial Least Squares (PLS), also known as projection to latent structures (PLS) regression is a supervised method which can be used when predefined knowledge about each samples, such as class membership and quantitative information is available (Wold, Sjöström, and Eriksson 2001). It is a multivariate technique that assesses the relationship between two blocks of data: a descriptor matrix X (i.e. spectral data) and a response matrix Y (known sample

information) and thus finds components (e.g. latent variables) in the predictor matrix that best predict the response variables. The Y matrix can either contain quantitative (i.e. metabolite concentration) or qualitative information (i.e. class membership). When Y contains sample class membership information, the application of PLS regression is called partial least squares-discriminant analysis (PLS-DA) (Barker and Rayens 2003).

The decomposition form of PLS-DA can be written as:

$$X = TP^T + E \quad (3.3)$$

$$Y = UQ^T + F \quad (4.4)$$

where, X is the matrix of predictors and Y is the matrix of responses; T and P are the scores (projections) and loadings for matrix X , respectively; U and Q are, respectively, the scores and loadings for matrix Y ; and E and F matrices represent the residuals or error terms for X and Y , respectively. The decompositions of X and Y is performed such that T and U share maximum covariance (Trygg, Holmes, and Lundstedt 2007a).

Although, the use of class information in PLS-DA allows for a better discrimination between classes, the variation that is not directly correlated with Y being still present in the scores, complicating the interpretation of the model. Thus, in the cases where class separation is not observed in scores plot of PLS-DA model, orthogonal partial least squares discriminant analysis (OPLS-DA) can be carried out to perform the analysis. The OPLS-DA is a modification of PLS-DA model, in which the systematic variation of data that is not related to the response variable (e.g. sample class labels) is removed. OPLS-DA method uses PLS model in combination with Orthogonal Signal Correction (OSC) filter to exclude irrelevant part of data that are uncorrelated (orthogonal) with the response and thus separates Y -predictive variation from the Y -orthogonal (uncorrelated) variation (Sjöblom et al. 1998; Worley and Powers 2013). OPLS-DA improves the classification performance only in cases where individual classes exhibit divergence in within-class variation. It presents a similar prediction ability to PLS-DA, and demonstrates improved model interpretability. The mathematical relationship is described in equation (1.5) (Trygg, Holmes, and Lundstedt 2007b).

$$X = T_p P_p^T + T_o P_o^T + E \quad (5.5)$$

Where, T_p is the predictive score matrix for X ; P_p^T is the predictive loading matrix for X ; T_o is the corresponding Y -orthogonal score matrix; P_o^T is the corresponding Y -orthogonal loading matrix; and E is the residual matrix.

The main benefit of OPLS-DA over PLS-DA is that a single component serves as a predictor for the class, while the other components describe the variation orthogonal to the first predictive component (Westerhuis et al. 2010).

1.10.3 Statistical validation of multivariate models

To assess the predictive ability of the multivariate models, statistical validation step is needed. Two types of validations are commonly performed: external and internal. In external validation, new data are collected/measured and used, and for internal validation, the data are either divided into two sets: training set for modelling and a test/validation set or a permutation technique is used. The training set is used to build the classification models (e.g. PLS-DA), and the resulting model is used to predict the classes of the test set (Broadhurst and Kell 2006). The purpose of validation is to assess the predictive ability of the model. Preferably, both types of methods should be used in parallel to confirm the reliability of the model. A number of methods such as cross-validation (CV), permutation, or bootstrap can be carried out to assess the performance and reliability of multivariate model.

Cross-validation techniques include leave- n -out and K -fold methods (Shao 1993; Worley and Powers 2013). In K -fold CV, the dataset is partitioned into k sized subsets and then iteratively $k-1$ subsets are combined as a training set, with the remaining subset functioning as a test set. In leave- n -out, the data are divided into N choose- n subsets and each subset serves as a validation set in each iteration. If the sample size is n , in leave-one-out, $n-1$ samples are used as a training set to fit a classification model and the remaining sample is used as test data. Therefore, every sample functions as a test set just once. The model built on $n-1$ samples has the same accuracy as a model built on all (n) samples. Leave-one-out techniques are computationally desirable because they involve fitting the classification model n times. This process is repeated until all samples have been left out and predicted once.

1.11 Receiver operating characteristic (ROC) curve

Receiver operating characteristic (ROC) curve analysis is a validation tool that does not assume a normal distribution of the studied variable and measures a variable's predictive accuracy by showing the relation between the true positive rate (sensitivity) and true negative rate (specificity) (Bünger and Mallet 2016; Moroz 2010). Sensitivity is defined as the fraction of positive observations correctly classified by the model into the positive class. On the other hand, specificity is defined as the fraction of negative observations correctly assigned by the

model to the negative class. In an ROC curve, the sensitivity (in Y-axis) is plotted against the specificity of a variable (in X-axis). The area under the curve (AUC) value generated from ROC curve can be used as a criterion for assessing the success of the classification model (Macias et al. 2019). The closer the area under the ROC curve (AUC) to 1 (maximum value), the more successful the classification model. The classifier is of no practical utility when AUC reaches 0.5, as this indicates that subject classification is random. In addition, the ability of the metabolite to differentiate between two groups of samples can be investigated from the shape of the ROC curve. The most desirable curve has a sharp increase in true positive rate and a slight increase in false positive rate. In this thesis, ROC curves have been generated using “*Biomarker Analysis*” module of the MetaboAnalyst (www.metaboanalyst.ca).

1.12 Correlation analysis and heat maps

The analysis of the correlations is an additional tool useful for retrieving information from the metabolomics data. It is commonly based on pairwise correlation between concentration levels of the metabolites in a sample.

The most common correlation approach in metabolomics is *Pearson's correlation* (linear correlation between variables), however other options are also available (e.g. the non-linear *Spearman correlation*). Pearson correlation works with the raw data of the variables whereas the Spearman correlation coefficient is based on the ranked values for each variable rather than the raw data.

A heatmap is a graphical representation of individual metabolite concentrations, where values contained in a matrix, are usually represented using a colour scale (Figure 1.13) The main advantage of heatmaps is that they display relationships between data values that might be complicated to understand if presented numerically in the datasheet.

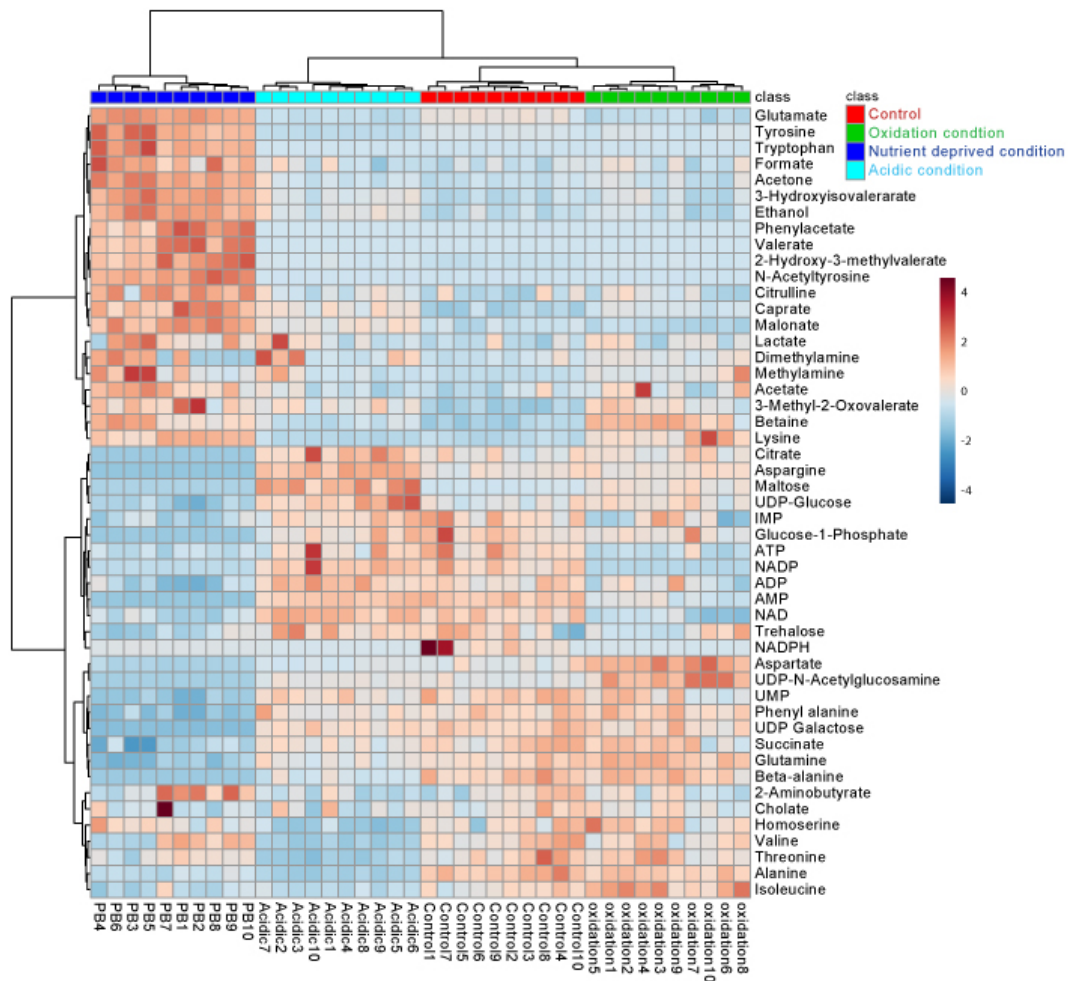


Figure 1.13: Heat map of *Mycobacterium smegmatis* grown under different stress conditions. Normalized signal intensities are visualized as a colour spectrum. The intensity of colour indicates the metabolite concentrations ranging from least abundant (dark blue) to higher ranges (dark red).

1.13 Significance Analysis for Microarrays (SAM)

Significance Analysis for Microarrays (SAM) is a permutation-based (non-parametric) hypothesis testing method which uses a moderated t-test, denoted as d_i , to measure the change in metabolite expression between the two groups. SAM method is used to identify the most discriminant and important metabolites responsible for the intergroup separation. SAM provides an estimation of the FDR, and metabolites with SAM-FDR q-value < 0.05 were considered statistically significant. One of the crucial parameter in SAM is the tunable delta value, which is used to tune the FDR values and selection of important metabolites. The SAM plot has three lines, one centre line crossing the origin corresponds equal observed $d(i)$ and expected $d(i)$ values, and the other two lines parallel to the diagonal line are defined by the Δ cut-off, more precisely going below and above the centre line by Δ (Figure 1.14). Metabolites with observed SAM scores above the upper line are positive significant metabolites (upregulated) and metabolites with observed SAM scores below the lower line are negative significant metabolites (downregulated).

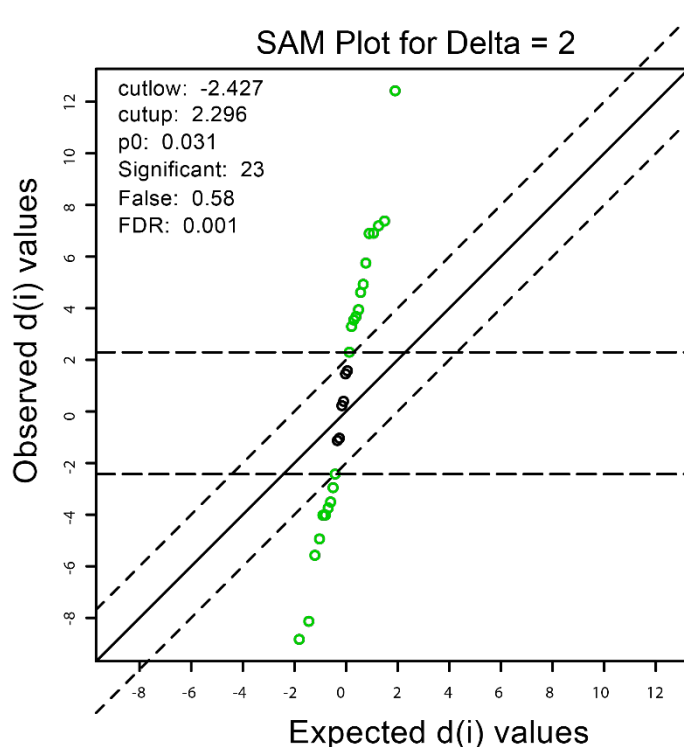


Figure 1.14: Significance analysis of microarray (SAM) plots for selection of significant metabolites. The green circles on the top represents upregulated metabolites and red circles on the bottom represents down-regulated metabolites. $d(i)$ indicates SAM score.

1.14 Pathway analysis

To gain insights into the metabolic mechanism of various cellular processes and diseases, metabolic pathway analysis is essential for understanding the biological significance of the metabolic perturbations and provides insight in the development of treatment methods. After the identification of potential metabolite biomarkers, the pathway analysis is used to identify target pathways involved in the mechanism of different biological problems and is the best route to link metabolites to biological functions or biological consequences. The metabolic pathway analysis in this thesis has been carried out on a free web-based program – Metaboanalyst. MetaboAnalyst encompassed two modules for functional analysis, Metabolic Pathway Analysis (MetPA) and Metabolite Set Enrichment Analysis (MSEA) (Xia and Wishart 2011). The pathway analysis module (MetPa) of MetaboAnalyst 4.0 conducts the metabolic pathway analysis based on the databases of HMDB (Wishart et al. 2007b) and Kyoto Encyclopaedia of Genes and Genomes (KEGG) (Kanehisa and Goto 2000). The most dysregulated metabolic pathways can be identified based on the statistical value (p-value) of the pathway enrichment analysis and pathway impact value calculated from pathway topological analysis (Xia et al., 2011). The pathway topological analysis is based on the centrality measures of a metabolite in a given metabolic network. ‘MetPA’ uses the relative betweenness centrality and out degree centrality measures to calculate the importance of each metabolite in each pathway (Xia, Wishart, and Valencia 2011). The metabolic pathways with impact value > 0.1 and $p < 0.05$ are usually considered as significantly perturbed pathways.

1.15 Summary of the chapter

Over the last decade, the rapid advancement of the Omics technologies has made it feasible for researchers to target a wider area of any given biological sample when looking for disease biomarkers. Metabolomics provides unique snapshots of the interaction between the pathophysiological stimuli and the host response, providing an opportunity to answer questions that have not been addressed with other upstream -omics technologies. Generally, there are three important and crucial steps in any metabolic experiment: sample preparation, data measurement and analysis, and statistical analysis. For each of these steps, a number of issues need to be carefully considered to ensure that the metabolomics experiment can be labelled as both robust and reliable. Nuclear Magnetic Resonance (NMR) and Mass Spectrometry (MS) spectroscopy are two most powerful and information-rich analytical platforms currently employed in metabolomics studies. The unique properties of NMR, including its high degree of reproducibility, relative ease of sample preparation, its ability to handle various types of

samples, its highly quantitative nature, its ability to identify unknown metabolites and its inherently non-destructive nature, have made it an eminent technique useful in several disciplines of metabolomics. NMR-based Metabolomics is being used in different areas including disease biology, medicine, pharmacy, toxicology, food and environmental sciences, etc., to obtain the metabolic fingerprint associated with different biological systems. To ensure robust and accurate quantification of metabolites by NMR, several parameters related to instrumental, acquisition, and post-acquisition including optimal NMR pulse sequences, acceptable acquisition parameter ranges, magnetic field strength, chemical shift referencing, baseline correction, etc. need to be carefully optimised before conducting any quantitative analysis. To analyse the complex and high-dimensional metabolic data, chemometric analysis needs to be required for data interpretation. Prior to multivariate analysis, pre-processing of the metabolome data must be performed to eliminate spectral artifacts and remove biologically irrelevant variations, making the NMR more comparable. Pre-processing of NMR metabolomics data includes several procedures and the optimal method to carry out may vary depending on the type of samples analyzed. Metabolic profiles of plants, animals, or humans, even if healthy, are very complex. In order to validate the classification and prognostic models for the identification of potential biomarkers that could aid the diagnosis, monitoring and the prediction of a disease or the outcome of a stimuli, the metabolomics data are subjected to chemometric analysis. Multivariate analysis (MVA) have been used to explore and discover the overall structure of the data, find trends and groupings, detects outliers in the analysis. MVA is highly graphical in its approach, provide tools to visualize the relationships between samples and variables. Thus, a systematic approach encompassing above-described methods pertaining to metabolomics yields a crafted strategy to understand the phenotypic response to given stimuli.

References

- Adams, Ralph W. 2014. "Pure Shift NMR Spectroscopy." *EMagRes*.
- Adell, Pere, Teodor Parella, Francisco Sánchez-Ferrando, and Albert Virgili. 1997. *Methods for Structure Elucidation by High-Resolution NMR Applications to Organic Molecules of Moderate Molecular Weight*. Vol. 8. Elsevier.
- Anderson, Paul E., Deirdre A. Mahle, Travis E. Doom, Nicholas V. Reo, Nicholas J. DelRaso, and Michael L. Raymer. 2011. "Dynamic Adaptive Binning: An Improved Quantification Technique for NMR Spectroscopic Data." *Metabolomics*.
- Aretz, Ina and David Meierhofer. 2016. "Advantages and Pitfalls of Mass Spectrometry Based Metabolome Profiling in Systems Biology." *International Journal of Molecular Sciences* 17(5).
- Barker, Matthew and William Rayens. 2003. "Partial Least Squares for Discrimination." *Journal of Chemometrics* 17(3):166–73.
- Beckonert, Olaf, Hector C. Keun, Timothy M. D. Ebbels, Jacob Bundy, Elaine Holmes, John C. Lindon, and Jeremy K. Nicholson. 2007. "Metabolic Profiling, Metabolomic and Metabonomic Procedures for NMR Spectroscopy of Urine, Plasma, Serum and Tissue Extracts." *Nature Protocols* 2(11):2692–2703.
- van den Berg, Robert A., Huub C. J. Hoefsloot, Johan A. Westerhuis, Age K. Smilde, and Mariët J. van der Werf. 2006a. "Centering, Scaling, and Transformations: Improving the Biological Information Content of Metabolomics Data." *BMC Genomics*.
- van den Berg, Robert A., Huub C. J. Hoefsloot, Johan A. Westerhuis, Age K. Smilde, and Mariët J. van der Werf. 2006b. "Centering, Scaling, and Transformations: Improving the Biological Information Content of Metabolomics Data." *BMC Genomics* 7:142.
- Berry, D. P., K. G. Meade, M. P. Mullen, S. Butler, M. G. Diskin, D. Morris, and C. J. Creevey. 2011. "The Integration of Omic Disciplines and Systems Biology in Cattle Breeding." *Animal* 5(4):493–505.
- Bingol, Kerem. 2018. "Recent Advances in Targeted and Untargeted Metabolomics by NMR and MS/NMR Methods." *High-Throughput* 7(2).
- Bro, Rasmus and Age K. Smilde. 2014. "Principal Component Analysis." *Analytical Methods* 6(9):2812–31.
- Broadhurst, David I. and Douglas B. Kell. 2006. "Statistical Strategies for Avoiding False

- Discoveries in Metabolomics and Related Experiments.” *Metabolomics* 2(4):171–96.
- Bruntz, Ronald C., Andrew N. Lane, Richard M. Higashi, and Teresa W-M Fan. 2017. “Exploring Cancer Metabolism Using Stable Isotope-Resolved Metabolomics (SIRM) Downloaded From.” *J. Biol. Chem* (28):11601.
- Buescher, Joerg M., Maciek R. Antoniewicz, Laszlo G. Boros, Shawn C. Burgess, Henri Brunengraber, Clary B. Clish, Ralph J. DeBerardinis, Olivier Feron, Christian Frezza, Bart Ghesquiere, Eyal Gottlieb, Karsten Hiller, Russell G. Jones, Jurre J. Kamphorst, Richard G. Kibbey, Alec C. Kimmelman, Jason W. Locasale, Sophia Y. Lunt, Oliver D. K. Maddocks, Craig Malloy, Christian M. Metallo, Emmanuelle J. Meuillet, Joshua Munger, Katharina Nöh, Joshua D. Rabinowitz, Markus Ralser, Uwe Sauer, Gregory Stephanopoulos, Julie St-Pierre, Daniel A. Tennant, Christoph Wittmann, Matthew G. Vander Heiden, Alexei Vazquez, Karen Vousden, Jamey D. Young, Nicola Zamboni, and Sarah Maria Fendt. 2015. “A Roadmap for Interpreting ¹³C Metabolite Labeling Patterns from Cells.” *Current Opinion in Biotechnology* 34:189–201.
- Bünger, Rolf and Robert T. Mallet. 2016. “Metabolomics and Receiver Operating Characteristic Analysis: A Promising Approach for Sepsis Diagnosis*.” *Critical Care Medicine* 44(9):1784–85.
- Carr, H. Y. and E. M. Purcell. 1954. “Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments.” *Physical Review* 94(3):630–38.
- Clayton, T. Andrew, John C. Lindon, Olivier Cloarec, Henrik Antti, Claude Charuel, Gilles Hanton, Jean Pierre Provost, Jean Loïc Le Net, David Baker, Rosalind J. Walley, Jeremy R. Everett, and Jeremy K. Nicholson. 2006. “Pharmaco-Metabonomic Phenotyping and Personalized Drug Treatment.” *Nature* 440(7087):1073–77.
- Dalisay, Doralyn S. and Tadeusz F. Molinski. 2009. “NMR Quantitation of Natural Products at the Nanomole Scale.” *Journal of Natural Products*.
- Davis, Richard A., Adrian J. Charlton, John Godward, Stephen A. Jones, Mark Harrison, and Julie C. Wilson. 2007. “Adaptive Binning: An Improved Binning Method for Metabolomics Data Using the Undecimated Wavelet Transform.” *Chemometrics and Intelligent Laboratory Systems* 85(1):144–54.
- Dettmer, Katja and Bruce D. Hammock. 2004. “Metabolomics - A New Exciting Field within the ‘Omics’ Sciences.” *Environmental Health Perspectives* 112(7):A396.

- Dieterle, Frank, Alfred Ross, Götz Schlotterbeck, and Hans Senn. 2006. "Probabilistic Quotient Normalization as Robust Method to Account for Dilution of Complex Biological Mixtures. Application In ^1H NMR Metabonomics." *Analytical Chemistry* 78(13):4281–90.
- Djukovic, Danijel, G. A. Nagana Gowda, and Daniel Raftery. 2013. "Mass Spectrometry and NMR Spectroscopy-Based Quantitative Metabolomics." Pp. 279–97 in *Proteomic and Metabolomic Approaches to Biomarker Discovery*. Elsevier Inc.
- Dona, Anthony C. 2018. "CHAPTER 1: Instrumental Platforms for NMR-Based Metabolomics." Pp. 1–21 in *New Developments in NMR*. Vols. 2018-January. Royal Society of Chemistry.
- Dona, Anthony C., Michael Kyriakides, Flora Scott, Elizabeth A. Shephard, Dorsa Varshavi, Kirill Veselkov, and Jeremy R. Everett. 2016. "A Guide to the Identification of Metabolites in NMR-Based Metabonomics/Metabolomics Experiments." *Computational and Structural Biotechnology Journal* 14:135–53.
- Dubey, Abhinav, Shivanand M. Pudakalakatti, and Hanudatta S. Atreya. 2019. "Fast NMR Methods for Identification of Resonances and Metabolic Pathways." Pp. 135–47 in *Methods in Molecular Biology*. Vol. 2037. Humana Press Inc.
- Dunn, Warwick B. and David I. Ellis. 2005. "Metabolomics: Current Analytical Platforms and Methodologies." *TrAC - Trends in Analytical Chemistry* 24(4):285–94.
- Emwas, Abdul Hamid, Claudio Luchinat, Paola Turano, Leonardo Tenori, Raja Roy, Reza M. Salek, Danielle Ryan, Jasmine S. Merzaban, Rima Kaddurah-Daouk, Ana Carolina Zeri, G. A. Nagana Gowda, Daniel Raftery, Yulan Wang, Lorraine Brennan, and David S. Wishart. 2015. "Standardizing the Experimental Conditions for Using Urine in NMR-Based Metabolomic Studies with a Particular Focus on Diagnostic Studies: A Review." *Metabolomics* 11(4):872–94.
- Emwas, Abdul Hamid, Edoardo Saccenti, Xin Gao, Ryan T. McKay, Vitor A. P. Martin. dos Santos, Raja Roy, and David S. Wishart. 2018. "Recommended Strategies for Spectral Processing and Post-Processing of 1D ^1H -NMR Data of Biofluids with a Particular Focus on Urine." *Metabolomics* 14(3):1–23.
- Fernández-Peralbo, M. A. and M. D. Luque de Castro. 2012. "Preparation of Urine Samples Prior to Targeted or Untargeted Metabolomics Mass-Spectrometry Analysis." *TrAC - Trends in Analytical Chemistry* 41:75–85.

- Fiehn, Oliver. 2002. "Metabolomics - The Link between Genotypes and Phenotypes." *Plant Molecular Biology* 48(1–2):155–71.
- Gauthier, Thomas D. and Mark E. Hawley. 2015. "Statistical Methods." Pp. 99–148 in *Introduction to Environmental Forensics: Third Edition*. Elsevier Inc.
- Giraudeau, Patrick. 2020. "NMR-Based Metabolomics and Fluxomics: Developments and Future Prospects." *Analyst*.
- Giraudeau, Patrick and Lucio Frydman. 2014a. "Single-Scan 2D NMR : An Emerging Tool in Analytical Spectroscopy." *Annu Rev Anal Chem (Palo Alto Calif)*.
- Giraudeau, Patrick and Lucio Frydman. 2014b. "Ultrafast 2D NMR: An Emerging Tool in Analytical Spectroscopy." *Annual Review of Analytical Chemistry*.
- Le Guennec, Adrien, Jean Nicolas Dumez, Patrick Giraudeau, and Stefano Caldarelli. 2015. "Resolution-Enhanced 2D NMR of Complex Mixtures by Non-Uniform Sampling." *Magnetic Resonance in Chemistry*.
- Hendriks, Margriet M. W. B., Fred A. va. Eeuwijk, Renger H. Jellema, Johan A. Westerhuis, Theo H. Reijmers, Huub C. J. Hoefsloot, and Age K. Smilde. 2011. "Data-Processing Strategies for Metabolomics Studies." *TrAC - Trends in Analytical Chemistry* 30(10):1685–98.
- Horne, Amelia Dale. 1998. "Statistics, Use in Immunology." Pp. 2211–15 in *Encyclopedia of Immunology*. Elsevier.
- Hotelling, H. 1933. "Analysis of a Complex of Statistical Variables into Principal Components." *Journal of Educational Psychology* 24(6):417–41.
- Hoult, D. I. and R. E. Richards. 1976. "The Signal-to-Noise Ratio of the Nuclear Magnetic Resonance Experiment." *Journal of Magnetic Resonance (1969)*.
- Huang, Xiaohua, Robert Powers, Adrienne Tymiak, Robert Espina, and Vikram Roongta. 2007. "Introduction to NMR and Its Application in Metabolite Structure Determination." Pp. 369–409 in *Drug Metabolism in Drug Design and Development: Basic Concepts and Practice*. Hoboken, NJ, USA: John Wiley & Sons, Inc.
- Huang, Yuqing, Zhiyong Zhang, Hao Chen, Jianghua Feng, Shuhui Cai, and Zhong Chen. 2015. "A High-Resolution 2D J-Resolved NMR Detection Technique for Metabolite Analyses of Biological Samples." *Scientific Reports* 5(1):1–9.
- Hwang, T. L. and A. J. Shaka. 1995. "Water Suppression That Works. Excitation Sculpting

- Using Arbitrary Wave-Forms and Pulsed-Field Gradients.” *Journal of Magnetic Resonance - Series A* 112(2):275–79.
- Jacobsen, Neil E. 2007. “Two-Dimensional NMR Spectroscopy: HETCOR, COSY, and TOCSY.” Pp. 353–407 in *NMR Spectroscopy Explained*. Hoboken, NJ, USA: John Wiley & Sons, Inc.
- Jalloh, Ibrahim, Keri L. H. Carpenter, Peter Grice, Duncan J. Howe, Andrew Mason, Clare N. Gallagher, Adel Helmy, Michael P. Murphy, David K. Menon, T. Adrian Carpenter, John D. Pickard, and Peter J. Hutchinson. 2015. “Glycolysis and the Pentose Phosphate Pathway after Human Traumatic Brain Injury: Microdialysis Studies Using 1,2-¹³C₂ Glucose.” *Journal of Cerebral Blood Flow and Metabolism* 35(1):111–20.
- Jolliffe, I. T. 2005. “Principal Component Analysis. Encyclopedia of Statistics in Behavioral Science.” P. 518 in *Encyclopedia of Statistics in Behavioral Science*. Chichester, UK: John Wiley & Sons, Ltd.
- Kaddurah-Daouk, Rima, Bruce S. Kristal, and Richard M. Weinshilboum. 2008. “Metabolomics: A Global Biochemical Approach to Drug Response and Disease.” *Annual Review of Pharmacology and Toxicology* 48(1):653–83.
- Kanehisa, Minoru and Susumu Goto. 2000. “KEGG: Kyoto Encyclopedia of Genes and Genomes.” *Nucleic Acids Research* 28(1):27–30.
- Kell, Douglas B. 2004. “Metabolomics and Systems Biology: Making Sense of the Soup.” *Current Opinion in Microbiology* 7:296–307.
- Keun, Hector C., Timothy M. D. Ebbels, Henrik Antti, Mary E. Bollard, Olaf Beckonert, Elaine Holmes, John C. Lindon, and Jeremy K. Nicholson. 2003. “Improved Analysis of Multivariate Data by Variable Stability Scaling: Application to NMR-Based Metabolic Profiling.” in *Analytica Chimica Acta*.
- Kim, Hye Kyong, Young Hae Choi, and Robert Verpoorte. 2010. “NMR-Based Metabolomic Analysis of Plants.” *Nature Protocols* 5(3):536–49.
- Kim, Hye Kyong and Rob Verpoorte. 2010. “Sample Preparation for Plant Metabolomics.” *Phytochemical Analysis* 21(1):4–13.
- Klein, Sebastian and Elmar Heinzle. 2012. “Isotope Labeling Experiments in Metabolomics and Fluxomics.” *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*.
- Kovacs, Helena, Detlef Moskau, and Manfred Spraul. 2005. “Cryogenically Cooled Probes -

- A Leap in NMR Technology.” *Progress in Nuclear Magnetic Resonance Spectroscopy* 46(2–3):131–55.
- Kruk, Joanna, Marek Doskocz, Elżbieta Jodłowska, Anna Zacharzewska, Joanna Łakomiec, Kornelia Czaja, and Jacek Kujawski. 2017. “NMR Techniques in Metabolomic Studies: A Quick Overview on Examples of Utilization.” *Applied Magnetic Resonance* 48(1):1–21.
- Kupce, Eriks. 2015. “NMR with Multiple Receivers.” *EMagRes*.
- Larive, Cynthia K., Gregory A. Barding, and Meredith M. Dinges. 2015. “NMR Spectroscopy for Metabolomics and Metabolic Profiling.” *Analytical Chemistry* 87(1):133–46.
- Lindon, John C., Jeremy K. Nicholson, Elaine Holmes, and Jeremy R. Everett. 2000. “Metabonomics: Metabolic Processes Studied by NMR Spectroscopy of Biofluids.” *Concepts in Magnetic Resonance* 12(5):289–320.
- Ludwig, Christian and Mark R. Viant. 2010. “Two-Dimensional J-Resolved NMR Spectroscopy: Review of a Key Methodology in the Metabolomics Toolbox.” *Phytochemical Analysis* 21(1):22–32.
- Macias, Shirin, Joseph Kirma, Ali Yilmaz, Sarah E. Moore, Michelle C. McKinley, Pascal P. McKeown, Jayne V. Woodside, Stewart F. Graham, and Brian D. Green. 2019. “Application of ¹H-NMR Metabolomics for the Discovery of Blood Plasma Biomarkers of a Mediterranean Diet.” *Metabolites* 9(10):201.
- Mahrous, Engy A. and Mohamed A. Farag. 2015. “Two Dimensional NMR Spectroscopic Approaches for Exploring Plant Metabolome: A Review.” *Journal of Advanced Research* 6(1):3–15.
- Mallol, Roger, Miguel Angel Rodriguez, Jesus Brezmes, Lluís Masana, and Xavier Correig. 2013. “Human Serum/Plasma Lipoprotein Analysis by NMR: Application to the Study of Diabetic Dyslipidemia.” *Progress in Nuclear Magnetic Resonance Spectroscopy* 70:1–24.
- Markley, John L., Rafael Brüschweiler, Arthur S. Edison, Hamid R. Eghbalnia, Robert Powers, Daniel Raftery, and David S. Wishart. 2017. “The Future of NMR-Based Metabolomics.” *Current Opinion in Biotechnology* 43:34–40.
- Martínez-Reyes, Inmaculada and Navdeep S. Chandel. 2020. “Mitochondrial TCA Cycle Metabolites Control Physiology and Disease.” *Nature Communications* 11(1):1–11.

- Mi, Kai, Yanan Jiang, Jiabin Chen, Dongxu Lv, Zhipeng Qian, Hui Sun, and Desi Shang. 2020. "Construction and Analysis of Human Diseases and Metabolites Network." *Frontiers in Bioengineering and Biotechnology*.
- Mobli, Mehdi and Jeffrey C. Hoch. 2014. "Nonuniform Sampling and Non-Fourier Signal Processing Methods in Multidimensional NMR." *Progress in Nuclear Magnetic Resonance Spectroscopy*.
- Monteiro, M. S., M. Carvalho, M. L. Bastos, and P. Guedes de Pinho. 2013. "Metabolomics Analysis for Biomarker Discovery: Advances and Challenges." *Current Medicinal Chemistry*.
- Moroz, Jennifer. 2010. "Application of ROC Curve Analysis to Metabolomics Data Sets for the Detection of Cancer in a Mouse Model."
- Nagana Gowda, G. A. and Daniel Raftery. 2014. "Quantitating Metabolites in Protein Precipitated Serum Using NMR Spectroscopy." *Analytical Chemistry*.
- Nemet, Ina, Prasenjit Prasad Saha, Nilaksh Gupta, Weifei Zhu, Kymberleigh A. Romano, Sarah M. Skye, Tomas Cajka, Maradumane L. Mohan, Lin Li, Yuping Wu, Masanori Funabashi, Amanda E. Ramer-Tait, Sathyamangla Venkata Naga Prasad, Oliver Fiehn, Federico E. Rey, W. H. Wilso. Tang, Michael A. Fischbach, Joseph A. DiDonato, and Stanley L. Hazen. 2020. "A Cardiovascular Disease-Linked Gut Microbial Metabolite Acts via Adrenergic Receptors." *Cell*.
- Nguyen, Bao D., Xi Meng, Kevin J. Donovan, and A. J. Shaka. 2007. "SOGGY: Solvent-Optimized Double Gradient Spectroscopy for Water Suppression. A Comparison with Some Existing Techniques." *Journal of Magnetic Resonance* 184(2):263–74.
- Nielsen, Niels Peter Vest, Jens Michael Carstensen, and Jørn Smedsgaard. 1998. "Aligning of Single and Multiple Wavelength Chromatographic Profiles for Chemometric Data Analysis Using Correlation Optimised Warping." *Journal of Chromatography A* 805(1–2):17–35.
- Pan, Zhengzheng and Daniel Raftery. 2007. "Comparing and Combining NMR Spectroscopy and Mass Spectrometry in Metabolomics." *Analytical and Bioanalytical Chemistry* 387(2):525–27.
- Picó, Yolanda. 2015. "Mass Spectrometry in Food Quality and Safety: An Overview of the Current Status." Pp. 3–76 in *Comprehensive Analytical Chemistry*. Vol. 68. Elsevier B.V.

- Pinu, Farhana R., Silas G. Villas-Boas, and Raphael Aggio. 2017. "Analysis of Intracellular Metabolites from Microorganisms: Quenching and Extraction Protocols." *Metabolites*.
- Piotto, Martial, Vladimir Saudek, and Vladimir Sklenář. 1992. "Gradient-Tailored Excitation for Single-Quantum NMR Spectroscopy of Aqueous Solutions." *Journal of Biomolecular NMR* 2(6):661–65.
- Sandilands, Debra. 2014. "Univariate Analysis." Pp. 6815–17 in *Encyclopedia of Quality of Life and Well-Being Research*. Dordrecht: Springer Netherlands.
- Savorani, F., G. Tomasi, and S. B. Engelsen. 2010. "Icoshift: A Versatile Tool for the Rapid Alignment of 1D NMR Spectra." *Journal of Magnetic Resonance* 202(2):190–202.
- Scalbert, Augustin, Lorraine Brennan, Oliver Fiehn, Thomas Hankemeier, Bruce S. Kristal, Ben van Ommen, Estelle Pujos-Guillot, Elwin Verheij, David Wishart, and Suzan Wopereis. 2009. "Mass-Spectrometry-Based Metabolomics: Limitations and Recommendations for Future Progress with Particular Focus on Nutrition Research." *Metabolomics* 5(4):435–58.
- Schroeder, Frank C. and Matthew Gronquist. 2006. "Extending the Scope of NMR Spectroscopy with Microcoil Probes." *Angewandte Chemie - International Edition*.
- Shao, Jun. 1993. "Linear Model Selection by Cross-Validation." *Journal of the American Statistical Association* 88(422):486.
- Shulaev, Vladimir. 2006. "Metabolomics Technology and Bioinformatics." *Briefings in Bioinformatics* 7(2):128–39.
- Sjöblom, Jonas, Olof Svensson, Mats Josefson, Hans Kullberg, and Svante Wold. 1998. "An Evaluation of Orthogonal Signal Correction Applied to Calibration Transfer of near Infrared Spectra." Pp. 229–44 in *Chemometrics and Intelligent Laboratory Systems*. Vol. 44. Elsevier.
- Sklenář, Vladimír, Martial Piotto, Raymond Leppik, and Vladimír Saudek. 1993. "Gradient-Tailored Water Suppression for 1H-15N HSQC Experiments Optimized to Retain Full Sensitivity." *Journal of Magnetic Resonance - Series A* 102(2):241–45.
- Smolinska, Agnieszka, Lionel Blanchet, Lutgarde M. C. Buydens, and Sybren S. Wijmenga. 2012. "NMR and Pattern Recognition Methods in Metabolomics: From Data Acquisition to Biomarker Discovery: A Review." *Analytica Chimica Acta* 750:82–97.
- Snytnikova, Olga A., Anastasiya A. Khlichkina, Renad Z. Sagdeev, and Yuri P. Tsentlovich.

2019. "Evaluation of Sample Preparation Protocols for Quantitative NMR-Based Metabolomics." *Metabolomics*.
- Tautz, Thomas, Judith Hoffmann, Thomas Hoffmann, Heinrich Steinmetz, Peter Washausen, Brigitte Kunze, Volker Huch, Andreas Kitsche, Hans Reichenbach, Gerhard Höfle, Rolf Müller, and Markus Kalesse. 2016. "Isolation, Structure Elucidation, Biosynthesis, and Synthesis of Antalid, a Secondary Metabolite from *Polyangium* Species." *Organic Letters* 18(11):2560–63.
- Timári, István, Cheng Wang, Alexandar L. Hansen, Gilson Costa Dos Santos, Sung Ok Yoon, Lei Bruscheiler-Li, and Rafael Brüscheiler. 2019. "Real-Time Pure Shift HSQC NMR for Untargeted Metabolomics." *Analytical Chemistry*.
- Trygg, Johan, Elaine Holmes, and Torbjörn Lundstedt. 2007a. "Chemometrics in Metabonomics." *Journal of Proteome Research* 6(2):469–79.
- Trygg, Johan, Elaine Holmes, and Torbjörn Lundstedt. 2007b. "Chemometrics in Metabonomics." *Journal of Proteome Research* 6(2):469–79.
- Veselkov, Kirill A., John C. Lindon, Timothy M. D. Ebbels, Derek Crockford, Vladimir V. Volynkin, Elaine Holmes, David B. Davies, and Jeremy K. Nicholson. 2009. "Recursive Segment-Wise Peak Alignment of Biological ¹H NMR Spectra for Improved Metabolic Biomarker Recovery." *Analytical Chemistry* 81(1):56–66.
- Westerhuis, Johan A., Ewoud J. J. van Velzen, Huub C. J. Hoefsloot, and Age K. Smilde. 2010. "Multivariate Paired Data Analysis: Multilevel PLSDA versus OPLSDA." *Metabolomics* 6(1):119–28.
- Wishart, David S. 2008. "Metabolomics: Applications to Food Science and Nutrition Research." *Trends in Food Science & Technology* 19(9):482–93.
- Wishart, David S. 2016. "Emerging Applications of Metabolomics in Drug Discovery and Precision Medicine." *Nature Reviews Drug Discovery* 15(7):473–84.
- Wishart, David S., Dan Tzur, Craig Knox, Roman Eisner, An Chi Guo, Nelson Young, Dean Cheng, Kevin Jewell, David Arndt, Summit Sawhney, Chris Fung, Lisa Nikolai, Mike Lewis, Marie Aude Coutouly, Ian Forsythe, Peter Tang, Savita Shrivastava, Kevin Jeroncic, Paul Stothard, Godwin Amegbey, David Block, David D. Hau, James Wagner, Jessica Miniaci, Melisa Clements, Mulu Gebremedhin, Natalie Guo, Ying Zhang, Gavin E. Duggan, Glen D. MacInnis, Alim M. Weljie, Reza Dowlatabadi, Fiona Bamforth,

- Derrick Clive, Russ Greiner, Liang Li, Tom Marrie, Brian D. Sykes, Hans J. Vogel, and Lori Querengesser. 2007a. "HMDB: The Human Metabolome Database." *Nucleic Acids Research*.
- Wishart, David S., Dan Tzur, Craig Knox, Roman Eisner, An Chi Guo, Nelson Young, Dean Cheng, Kevin Jewell, David Arndt, Summit Sawhney, Chris Fung, Lisa Nikolai, Mike Lewis, Marie Aude Coutouly, Ian Forsythe, Peter Tang, Savita Shrivastava, Kevin Jeronic, Paul Stothard, Godwin Amegbey, David Block, David D. Hau, James Wagner, Jessica Miniaci, Melisa Clements, Mulu Gebremedhin, Natalie Guo, Ying Zhang, Gavin E. Duggan, Glen D. MacInnis, Alim M. Weljie, Reza Dowlatabadi, Fiona Bamforth, Derrick Clive, Russ Greiner, Liang Li, Tom Marrie, Brian D. Sykes, Hans J. Vogel, and Lori Querengesser. 2007b. "HMDB: The Human Metabolome Database." *Nucleic Acids Research* 35(SUPPL. 1):D521-526.
- Wold, Svante, Kim Esbensen, and Paul Geladi. 1987. "Principal Component Analysis." *Chemometrics and Intelligent Laboratory Systems* 2(1-3):37-52.
- Wold, Svante, Michael Sjöström, and Lennart Eriksson. 2001. "PLS-Regression: A Basic Tool of Chemometrics." Pp. 109-30 in *Chemometrics and Intelligent Laboratory Systems*. Vol. 58. Elsevier.
- Worley, Bradley and Robert Powers. 2013. "Multivariate Analysis in Metabolomics." *Current Metabolomics* 1(1):92-107.
- Xia, Jianguo and David S. Wishart. 2011. "Metabolomic Data Processing, Analysis, and Interpretation Using MetaboAnalyst." *Current Protocols in Bioinformatics*.
- Xia, Jianguo, David S. Wishart, and Alfonso Valencia. 2011. "MetPA: A Web-Based Metabolomics Tool for Pathway Analysis and Visualization." Pp. 2342-44 in *Bioinformatics*. Vol. 27. Oxford University Press.
- Yang, Jun, Xinjie Zhao, Xin Lu, Xiaohui Lin, and Guowang Xu. 2015. "A Data Preprocessing Strategy for Metabolomics to Reduce the Mask Effect in Data Analysis." *Frontiers in Molecular Biosciences* 2(FEB).
- Yousf, Saleem, Nazia Hussain, Shilpy Sharma, and Jeetender Chugh. 2017. "Identification & Characterization of Secondary Metabolites in the Biological Soup by NMR Spectroscopy." Pp. 47-96 in.
- Zangger, Klaus. 2015. "Pure Shift NMR." *Progress in Nuclear Magnetic Resonance*

Spectroscopy.

Chapter 2

General Methodology

2.1 Introduction

This chapter gives an outline of the research methods that have been followed in the study. It provides information about essential parameters related to NMR experiments, metabolomics spectral databases, computational tools for metabolomic data and pathway analysis, software packages for metabolite identification and quantification, and types of analysis that we performed for data interpretation and feature selection. The other specific methods including other biochemical and cellular experiments, sample preparation and processing for NMR spectroscopy, etc. have been given in the respective chapters.

2.2 NMR Spectroscopy

All the NMR data were measured at VT controlled 298 K using a Bruker AVANCE III HD Ascend NMR spectrometer operating at 14.1 Tesla equipped with pulsed-field gradients in x, y, and z-direction (operating at 54 Gauss/cm); Bruker high-performance shim system with 36 orthogonal shim gradients and integrated real-time shim gradient for 3-axis shimming. A quad-channel (^1H , ^{13}C , ^{15}N , ^{31}P , and one separate channel for lock using ^2H) cryogenic solution-state 5 mm probe with automated tuning-matching was used for recording all the NMR data without sample spinning. Water suppression pulse sequence noesygppld from Bruker library was used to record 1D ^1H NMR spectra, which uses water pre-saturation (using a cw irradiation at $5.56\text{E}-05$ W during the inter-scan delay of 5 s) and spoiler gradients (Smoothed square shape SMSQ10.100, where G1 was with 50% power and G2 was with -10% power for 1 ms duration) during the relaxation delay and is of the form: RD-G1- 90° -t- 90° -tm-G2- 90° -ACQ, where RD is the inter-scan relaxation delay of 5 s, t is a short delay typically of ~ 3 μs , 90° represents a 90° RF pulse (15.47 μs square pulse at 5.8479 W), tm is the mixing time of 100 ms, and ACQ is the data acquisition period (6.95 s). A long relaxation delay of 5 s was chosen since peak intensities obtained from NMR data had to be used for quantification. For a given sample, a total of 64 transients and 16 steady-state transients were collected into 32K data points for each spectrum with a spectral width of 7200 Hz. Pulse width, receiver gain, and water suppression parameters were kept identical among all the ^1H experiments recorded for different samples to rule out intensity variations while recording the NMR data. For the ^1H - ^1H total correlation spectroscopy (TOCSY) experiment, the mlevesgppl pulse program from Bruker library was used with acquisition times of 0.065 s and 0.131 s in F1 and F2 dimension, respectively. TOCSY spin-lock of 80 ms was used for Hartman-Hahn mixing using composite blocks of 90° - 180° - 90° pulses with 90° pulse width of 25 μs at 2.29 W of power. A total of 2048×1024

data points with 64 transients per increment in the indirect dimension were recorded, spanning a spectral width of 6000 Hz in both the dimensions. A total of 4 steady-state transients were used in the TOCSY experiment, and data was collected in States-TPPI mode. Smoothed square-shaped (SMSQ10.100) gradients were used with 31% power (after the spin-lock period) and 11% power (before refocusing) for a duration of 1 ms.

2.3 Spectral processing

The ^1H NMR spectra of all the samples were processed using Bruker's NMR data processing software Topspin (v3.5) (www.bruker.com/bruker/topspin). ^1H NMR raw data was multiplied with exponential function and zero-filled to 64K data points before Fourier transformation. No linear prediction was applied during the processing of the data. Standard pre-processing steps such as phasing, baseline correction, line broadening was carried out for all the individual ^1H NMR spectra. All the ^1H chemical shifts were directly referenced with respect to methyl singlet of DSS internal reference, set to a chemical shift (δ) of 0.00 ppm. For ^1H - ^1H TOCSY processing, the FIDs were weighted in both dimensions by a pure cosine function (SINE with SSB = 2) and zero-filled to 2048 and 1024 data points in F1 and F2 dimensions, respectively, before subjecting the data to Fourier transformation.

2.4 Metabolite Identification and Quantification

After spectral processing, identification and quantification of metabolites were carried out with the Chenomx NMR Suite 8.1 (Edmonton, AB, Canada) software. The profiler module of this software was used to carry out the ^1H resonance assignment of metabolites based on the chemical shift values, coupling values, line shape, line-width, and multiplicity information. All the identified metabolites were further confirmed with biological magnetic resonance data bank (BMRB) (Ulrich et al. 2008) database and human metabolome database (HMDB) (Wishart et al. 2007). In addition, two dimensional ^1H - ^1H TOCSY was used for further metabolite confirmation via a semi-automated software – MetaboMiner (Xia et al. 2008).

All the identified metabolites were then quantified using the profiler module of Chenomx software, which enables metabolites quantification relative to an internal standard of known concentration (400 μM). Quantification is achieved using targeted profiling in which mathematically modelled pure compound NMR resonances from the Chenomx library are fit to the acquired spectra (Weljie et al. 2006). The concentration data obtained after metabolite quantification were converted to comma-separated values (CSV) format using Microsoft excel

format and imported into MetaboAnalyst 4.0 (www.metaboanalyst.ca), a free web-based program for multivariate analysis (Chong et al. 2018). Metabolites with low signal-to-noise ($s/n \leq 15$) in NMR analysis were excluded from the analysis as they may be vulnerable to over- or under-estimation of concentrations.

2.5 Statistical Analysis

Because of the high dimensionality and massive complexity of NMR metabolomics data, multivariate statistical analysis was carried out to discern the inter-group differences and to establish a systematic overview of the discrimination of metabolic patterns of different experimental conditions. For predicting the metabolite variations and identification of important features in different experimental conditions and samples, chemometric analysis – Principal Component Analysis (PCA), Partial Least Squares Discriminant Analysis (PLS-DA), and Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA) were conducted using a free web server program, MetaboAnalyst (Chong et al. 2018). Before subjecting the data for multivariate analysis, row-wise normalization of the data matrix [containing N observation row vectors of various samples (experimental conditions and replicates) having K variables (metabolites) in columns] was performed to normalize the data and reduce the variation within replicates. The normalized metabolite concentration was used as an input parameter to carry out chemometric analysis of the data in the MetaboAnalyst web tool. The metabolomics data sets were then subjected to scaling prior to chemometric analysis. Pareto-scaling has typically been the method of choice in NMR-based metabolomics studies as it minimizes the effect of intense signals while emphasizing weaker ones that could be more biologically important (Worley and Powers 2013a). PCA is an unsupervised method that does not use class label information and is used for clustering, while PLS-DA is a supervised method that uses class label information. and is used for classification (Worley and Powers 2013b). The score plot of the first two principal components was obtained and ellipses showing 95% confidence limits of a normal distribution for each group using PLS-DA/PCA utilities. OPLS-DA is a modification of the PLS-DA model, in which the systematic variation of data that is not related to the response variable (e.g., sample class labels) is removed (Boccard and Rutledge 2013; Westerhuis et al. 2010) The ellipse in the scores of the OPLS-DA model represents Hotelling's T² regions with a 95% confidence interval of the modelled variation. The quality of the OPLS-DA models was assessed by R², which defines the total explained variance (indicating goodness of fit) and Q² values, indicating the predictability of the model. The

permutation statistic was further used with 100 permutations to validate the OPLS-DA models. To identify the key metabolites responsible for the differential clustering of score plots in the PLS-DA model, the variable importance of projection (VIP) score plot was generated from PLS-DA analysis. VIP score measures the contribution of a variable to the PLS-DA model. The VIP score of a variable is calculated as a weighted sum of the squared correlations between the PLS-DA components and the original variable. VIP score identifies the important metabolites that vary significantly between the groups and better describes the intergroup variation. Metabolites with a VIP score of ≥ 1.0 are considered to be statistically significant as these contributed most to group discrimination.

Further one way ANOVA analysis followed by Fishers's least significant difference (LSD) post-hoc tests were used to assess the significance of changes in the levels of metabolite concentrations in different experimental conditions used in different chapters of this thesis. A false discovery rate (FDR)- adjusted p-value (or q value) threshold ($q\text{-value} \leq 0.05$) has been used in ANOVA analysis to determine the significance of differences in metabolite levels. Box and Whisker plots were generated to visualise the comparative variation in the metabolite concentrations of all the metabolites across the replicates and different experimental conditions. Pair-wise analysis was carried out using the Volcano plot utility of MetaboAnalyst that screens important metabolites based on fold-change (FC) and FDR-adjusted p-value (q-value). Metabolites with fold change ≥ 1.5 and $q\text{-value} \leq 0.05$ were considered significant. In addition, metabolite-metabolite correlations were obtained using Pearson's correlation coefficient analysis to identify all the significant correlations ($p\text{-value} \leq 0.05$).

Significance Analysis for Microarrays (SAM) method was employed in chapter 4 to identify the most discriminant and important metabolites responsible for the separation between control and diabetic groups. SAM is a permutation-based (non-parametric) hypothesis testing method that uses a moderated t-test, denoted as d_j , to measure the change in metabolite expression between the two groups (Roxas and Li 2008; Zhang 2007). SAM also estimated the FDR, and metabolites with SAM-FDR $q\text{-value} < 0.05$ were considered statistically significant. Metabolites with observed SAM scores above the upper line are positive significant metabolites (upregulated), and metabolites with observed SAM scores below the lower line are negative significant metabolites (downregulated).

To identify the potential candidate biomarkers, the Receiver operating characteristic (ROC) curve analysis was carried out in chapter 4 to assess the effectiveness of the differential metabolites in discriminating the serum profiles of diabetes patients from healthy patients. In

ROC curve specificity of a variable (in x-axis) is plotted against sensitivity (in the y-axis)(Metz 1978). The area under the curve (AUC) value generated from the ROC curve can be used as a criterion for biomarker discovery (Carter et al. 2016; Hajian-Tilaki 2013). The closer the area under the ROC curve (AUC) to 1 (maximum value), the more successful the classification model is. The classifier is of no practical utility when AUC reaches 0.5, indicating that subject classification is random. The most desirable curve has a sharp increase in the true positive rate and a slight increase in the false-positive rate. ROC curves were generated using the “Biomarker Analysis” module of the MetaboAnalyst (Chong et al., 2018; Xia, Wishart, & Valencia, 2011).

2.6 Metabolic pathway analysis

After the identification of potential metabolite biomarkers, pathway analysis is performed to identify target pathways involved in the mechanism of different biological problems. This indeed is the best route to link metabolites to biological functions or biological consequences. Metabolic pathway analysis was carried out on a free web-based program, Metaboanalyst (www.metaboanalyst.ca) (Chong et al. 2018). MetaboAnalyst encompasses two modules for functional analysis, Metabolic Pathway Analysis (MetPA) and Metabolite Set Enrichment Analysis (MSEA) (Xia and Wishart 2011). The pathway analysis module (MetPa) of MetaboAnalyst 4.0 conducts the metabolic pathway analysis based on the databases of HMDB (Wishart et al. 2007) and the Kyoto Encyclopaedia of Genes and Genomes (KEGG) (Kanehisa and Goto 2000). The most dysregulated metabolic pathways were identified based on the statistical value (p-value) of the pathway enrichment analysis and pathway impact value calculated from pathway topological analysis (Xia et al., 2011). The metabolic pathways with impact values ≥ 0.1 and *p-value* ≤ 0.05 were considered as significantly perturbed pathways.

References

- Boccard, Julien and Douglas N. Rutledge. 2013. "A Consensus Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA) Strategy for Multiblock Omics Data Fusion." *Analytica Chimica Acta*.
- Carter, Jane V., Jianmin Pan, Shesh N. Rai, and Susan Galandiuk. 2016. "ROC-Ing along: Evaluation and Interpretation of Receiver Operating Characteristic Curves." *Surgery (United States)*.
- Chong, Jasmine, Othman Soufan, Carin Li, Iurie Caraus, Shuzhao Li, Guillaume Bourque, David S. Wishart, and Jianguo Xia. 2018. "MetaboAnalyst 4.0: Towards More Transparent and Integrative Metabolomics Analysis." *Nucleic Acids Research* 46(W1):W486–94.
- Hajian-Tilaki, Karimollah. 2013. "Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation." *Caspian Journal of Internal Medicine*.
- Kanehisa, Minoru and Susumu Goto. 2000. "KEGG: Kyoto Encyclopedia of Genes and Genomes." *Nucleic Acids Research* 28(1):27–30.
- Metz, Charles E. 1978. "Basic Principles of ROC Analysis." *Seminars in Nuclear Medicine*.
- Roxas, Bryan A. P. and Qingbo Li. 2008. "Significance Analysis of Microarray for Relative Quantitation of LC/MS Data in Proteomics." *BMC Bioinformatics* 9(1):187.
- Ulrich, Eldon L., Hideo Akutsu, Jurgen F. Doreleijers, Yoko Harano, Yannis E. Ioannidis, Jundong Lin, Miron Livny, Steve Mading, Dimitri Maziuk, Zachary Miller, Eiichi Nakatani, Christopher F. Schulte, David E. Tolmie, R. Kent Wenger, Hongyang Yao, and John L. Markley. 2008. "BioMagResBank." *Nucleic Acids Research* 36(SUPPL. 1):D402-408.
- Weljie, Aalim M., Jack Newton, Pascal Mercier, Erin Carlson, and Carolyn M. Slupsky. 2006. "Targeted Profiling: Quantitative Analysis Of ¹H NMR Metabolomics Data." *Analytical Chemistry* 78(13):4430–42.
- Westerhuis, Johan A., Ewoud J. J. van Velzen, Huub C. J. Hoefsloot, and Age K. Smilde. 2010. "Multivariate Paired Data Analysis: Multilevel PLS-DA versus OPLS-DA." *Metabolomics*.
- Wishart, David S., Dan Tzur, Craig Knox, Roman Eisner, An Chi Guo, Nelson Young, Dean Cheng, Kevin Jewell, David Arndt, Summit Sawhney, Chris Fung, Lisa Nikolai, Mike Lewis, Marie Aude Coutouly, Ian Forsythe, Peter Tang, Savita Shrivastava, Kevin Jeroncic, Paul Stothard, Godwin Amegbey, David Block, David D. Hau, James Wagner, Jessica Miniaci, Melisa Clements, Mulu Gebremedhin, Natalie Guo, Ying Zhang, Gavin

- E. Duggan, Glen D. MacInnis, Alim M. Weljie, Reza Dowlatabadi, Fiona Bamforth, Derrick Clive, Russ Greiner, Liang Li, Tom Marrie, Brian D. Sykes, Hans J. Vogel, and Lori Querengesser. 2007. "HMDB: The Human Metabolome Database." *Nucleic Acids Research* 35(SUPPL. 1):D521-526.
- Worley, Bradley and Robert Powers. 2013a. "Multivariate Analysis in Metabolomics." *Current Metabolomics* 1(1):92-107.
- Worley, Bradley and Robert Powers. 2013b. "Multivariate Analysis in Metabolomics." *Current Metabolomics* 1(1):92-107.
- Xia, Jianguo, Trent C. Bjorndahl, Peter Tang, and David S. Wishart. 2008. "MetaboMiner - Semi-Automated Identification of Metabolites from 2D NMR Spectra of Complex Biofluids." *BMC Bioinformatics* 9:507-507.
- Xia, Jianguo and David S. Wishart. 2011. "Metabolomic Data Processing, Analysis, and Interpretation Using MetaboAnalyst." *Current Protocols in Bioinformatics*.
- Zhang, Shunpu. 2007. "A Comprehensive Evaluation of SAM, the SAM R-Package and a Simple Modification to Improve Its Performance." *BMC Bioinformatics*.

Chapter 3

Metabolic signatures suggest o-phosphocholine to UDP-N-acetylglucosamine ratio as a potential biomarker for high-glucose and/or palmitate exposure in pancreatic β -cells

3.1 Introduction

The increased global prevalence of Diabetes mellitus (DM) – a complex metabolic disorder marked by changes in glucose and lipid metabolism – is alarming and has been associated with great losses at the social, health, and economical fronts (Alberti et al. 2004; Zimmet, Alberti, and Shaw 2001). As per the recent statistics by the International Diabetes Federation (IDF; <https://www.idf.org/>) in 2020, 463 million people are suffering from diabetes mellitus worldwide. The numbers from a developing country like India are alarming, as IDF reports that 77 million people with diabetes are from India, which means that one in six people (17 %) in the world with diabetes belong to India (<https://www.diabetesatlas.org/en/>). Diabetes is generally associated with defects in insulin secretion, insulin action or the signalling of insulin, thereby resulting in chronic hyperglycaemia (Association 2009; Sönksen 1984). This has been primarily associated with defects in carbohydrate, fat, and protein metabolism (Dasgupta and Wahed 2014; Franz 1997). Type 2 Diabetes Mellitus (T2DM, also referred to as non-insulin dependent diabetes mellitus (NIDDM), adult onset diabetes mellitus) is the most prevalent form of diabetes and accounts for >90% of the cases.

T2DM is majorly caused due to the dysfunction and/or death of the pancreatic β -cells (Meier and Bonadonna 2013). Insulin, a hormone produced by the β -cells of pancreas, regulates carbohydrate and fat metabolism by storing excess glucose in the form of glycogen in liver and skeletal muscles, and in the form of triglycerides in fat tissue. In T2DM, insulin resistance develops in peripheral tissues, such as skeletal muscle, liver, and adipose tissues which fail in responding to blood concentrations of glucose. This has been associated with impaired insulin secretion and excessive hepatic glucose production, which in turn leads to the development of hyperglycemia and hyperlipidaemia (Saltiel and Kahn 2001). With the progression of the disease, the levels of advanced glycated end products and reactive oxygen species (ROS) rise, leading to severe macro- and micro-vascular complications, including cardiovascular diseases, hypertension, retinopathy, nephropathy, neuropathy, etc. (Chawla, Chawla, and Jaggi 2016) (Yamagishi and Matsui 2010).

Under normoglycemic conditions, after a diet, the plasma glucose levels rise gradually. Concomitant to this, insulin secretion is stimulated by the simultaneous activation of multiple metabolic pathways in response to glucose and other nutrients in the pancreatic β -cell. *In vitro* studies on β -cell lines and isolated islets have suggested that chronic hyperglycaemia (glucotoxicity) and dyslipidaemia (lipotoxicity) alone; or in combination (glucolipotoxicity) contribute to altered expression of genes involved in the insulin signalling pathway, glucose

stimulated insulin secretion (GSIS). This leads to the worsening of β -cell function over time (Erion et al. 2015; Kim and Yoon 2011; Mugabo et al. 2017; Somesh et al. 2013; Ward et al. 2017) and have been known to induce their apoptosis (Cnop et al. 2005). Indeed, increased plasma levels of free fatty acids (FFA) or non-esterified fatty acids (NEFAs) and elevated glucose levels have also been reported in diabetic and obese subjects (Carlson et al. 2007; Ni et al. 2015).

Metabolic disorders, including T2DM, go unnoticed in subjects for years before being clinically observable due to lack of potential prognostic biomarkers (Bhatia et al. 2015; Flowers et al. 2017). It has been found that in most of the cases, the development of T2DM is irreversible; and once the metabolic syndrome is established, the symptoms of T2DM can only be delayed but not entirely removed from the system. Also, the efficacy of the drugs could be greatly increased if they could be utilized at an early stage of the disorder before the actual metabolic syndrome sets in. Therefore, there exists a dire need to characterize and identify early biomarkers and key modulators involved in the initiation and progression of such disorders. In this context, Metabolomics – the systematic identification and quantitation of small molecule metabolites in a biological system (body fluids, cellular extracts, tissue extracts, etc.) at a specific time point – offers an attractive avenue for the identification of novel risk associated biomarkers (Molnos et al. 2018; Mugabo et al. 2017; Tam et al. 2017). Importantly, metabolomics is the final frontier of the omics cascade and therefore, the changes in the metabolome are amplified relative to changes in the transcriptome and the proteome. Thus, it has been used frequently either alone or in combination with other omics data (genomics, transcriptomics and proteomics) to obtain more insights into the pathophysiology of many diseases, including T2DM (Wanichthanarak, Fahrman, and Grapov 2015; Zheng and Hu 2015).

Even though extensive research has been carried out to understand the pathophysiology of T2DM, the exact cause of insulin resistance remains unknown. Several mechanisms including oxidative stress, mitochondrial dysfunction, endoplasmic reticulum stress, DNA damage, etc. have been proposed to play key roles in propagating and worsening T2DM conditions by affecting pancreatic β -cell viability and function (Bhandary et al. 2013; Hafizi Abu Bakar et al. 2015; Yazıcı and Sezer 2017). However, more studies are warranted to gain deeper mechanistic insights. Along these lines, a comprehensive metabolomic analysis of the pancreatic β -cells and the target tissues would provide cues into the mechanisms by which glucotoxicity, lipotoxicity, and glucolipotoxicity exerts its effect. These investigations would

also provide information on the major metabolic pathways affected under these conditions and would help in predicting novel drug targets.

Here, we aimed to employ metabolomics approach to identify the metabolic signatures and metabolic shifts in glucotoxic, lipotoxic, and glucolipotoxic conditions associated with T2DM using untargeted proton nuclear magnetic resonance ($^1\text{H-NMR}$) spectroscopy (Emwas et al. 2019). Using a series of biochemical and cellular experiments, high-glucose and/or palmitate (a saturated FFA) induced cytotoxic conditions were established in pancreatic β -cells. When compared to the control conditions, all three stress conditions were associated with elevated mitochondrial and cellular ROS levels. However, exposure to FFA alone (lipotoxicity) or in combination with high-glucose (glucolipotoxicity) was associated with increased cytotoxicity. A total of 48 abundant aqueous metabolites were identified, quantified, and compared across different experimental conditions using NMR spectroscopy. Pair-wise analysis was performed to identify distinct metabolic signatures for glucotoxic, lipotoxic, and glucolipotoxic conditions. The results from this investigation would aid in the better understanding of the metabolic pathways majorly functioning in the different stress conditions associated with T2DM.

3.2 Experimental Methods

3.2.1 Cell culture and establishment of cytotoxic conditions

The pancreatic rat insulinoma cells - *INS-IE* were used to perform our current study. These cells were obtained as a kind gift from Prof. Claes Wollheim and Prof. Pierre Machler from University Medical Centre, Geneva, Switzerland. *INS-IE* cells have ability to respond to glucose changes and has been recognized as a good surrogate for pancreatic β -cells (Merglen et al. 2004). Cells between passages 62 and 72 were grown in monolayer cultures in a humidified atmosphere containing 5% CO_2 at 37°C in complete medium composed of RPMI-1640 supplemented with 10 mM HEPES, 1 mM pyruvate, 50 μM 2-mercaptoethanol, 10% (v/v) heat inactivated FBS, 100 units/ml penicillin and 100 $\mu\text{g/ml}$ streptomycin. The cellular and biochemical experiments reported in this chapter were performed by Ms. Devika Sardesai in Dr. Shilpy Sharma's lab at the Department of Biotechnology, Savitribai Phule Pune University, Pune.

The viability of the *INS-IE* cells under different experimental conditions was determined using the MTT assay and lactate dehydrogenase (LDH) release assay (Himedia). Briefly, *INS-IE* cells were seeded onto 96-well plates (Corning) at an initial density of 1×10^4

cells/well and were allowed to adhere at 37°C for 24 h. Subsequently, the medium was replaced with fresh culture medium (100 μ l) containing 5 mM glucose (basal glucose media) and the plates were further incubated for 24 hours. Post-incubation, the glucose concentration was increased to 16 mM glucose (HG; mimicking glucotoxic conditions) alone and with 0.5 mM palmitate (HG/PA; mimicking glucolipotoxic conditions). The basal glucose concentration 5 mM used in this study corresponds to the fasting blood glucose levels (\leq 100 mg/dl) in a normal individual. On the other hand, the concentration of 16 mM glucose corresponds to plasma glucose levels \sim 300 mg/dl in a diabetic individual (Borg et al. 2010). While there are previous studies in the literature that have performed glucotoxicity experiments with glucose concentration $>$ 20 mM (Mugabo et al. 2017; Nyblom et al. 2008; Wallace, Whelan, and Brennan 2013), it is too high when compared to the physiological blood sugar levels. Thus, the concentration of 16 mM glucose was selected for HG and HG/PA conditions. For high lipid exposure, the basal media was supplemented with 0.5 mM palmitate (PA; mimicking lipotoxic conditions). It has been reported that palmitic acid — a saturated FFA, induces significantly higher apoptosis; higher oxidative stress; mitochondrial dysfunction, etc. when compared to the unsaturated FFAs (Cunha et al. 2008; El-Assaad et al. 2003; Yuzefovych, Wilson, and Rachek 2010). Therefore, PA was used for mimicking lipotoxic conditions. Untreated wells containing only cells in basal media were evaluated as controls. The different experimental conditions used in the study have been listed in Table 3.1.

Table 3.1: Experimental conditions used in this study.

Experimental Condition	Abbreviated as	Description
Basal Glucose (5 mM)	5G	Normal plasma glucose levels in humans corresponding to 100 mg/dl in a fasted state
Basal Glucose with BSA	5G+BSA	Vehicle control for Palmitate under normal glucose conditions
Basal Glucose with 0.5 mM Palmitate in BSA	5G+PA	Lipotoxic condition
High glucose (16 mM)	16G	Glucotoxic condition wherein 16 mM glucose corresponds to 300 mg/dl in a diabetic individual
High glucose with BSA	16G+BSA	Vehicle control for Palmitate under Glucotoxic condition
High glucose and 0.5 mM Palmitate in BSA	16G+PA	Glucolipotoxic condition

For MTT assay, media was removed after 24 h of treatment and the plates were incubated with 100 μ l of MTT solution (0.5 mg/ml in PBS) for 4 h at 37°C. Post-incubation, formazan crystals were dissolved with 100 μ l SDS/0.1N HCl for 4 h at RT. Absorbance was measured at 540 nm using a plate reader (Thermo Scientific). The reduction in cell viability was expressed as the percentage for each treatment relative to the control wells (set as 100%).

LDH release assay was performed using the LDH release assay kit (Himedia) as per the manufacturer's instructions. Briefly, 25 μ l of lysis buffer was added to untreated (control) wells and incubated at 37°C for 30 min to obtain maximum LDH release. Subsequently, 25 μ l of supernatant was taken from wells corresponding to different experimental conditions and the wells subjected to lysis and 25 μ l LDH reagent was added to it. The plate was incubated at 37°C for 15 min in dark. Post-incubation, 25 μ l of stop solution was added and absorbance was measured at 450 and 620 nm using a plate reader (Thermo Scientific). The % cytotoxicity was calculated as per the formula below:

$$\%Cytotoxicity = \frac{Treated\ LDH\ activity - Spontaneous\ LDH\ activity}{Maximum\ LDH\ activity - Spontaneous\ LDH\ activity} \times 100$$

3.2.2 Measurement of lipid accumulation

For relative quantitation of lipid accumulation inside *INS-1E* cells, staining was performed with Oil red O - a lysochrome, diazo dye used for staining triglycerides and lipids. Briefly, cells (0.5×10^5) were seeded in 24 well plates (Corning). The cells were treated with the six experimental conditions as described above. After incubation, the cells were washed twice with ice cold PBS and fixed with 10% formalin for 30 min. The plates were washed twice with 1X PBS twice and stained with oil red O solution (0.5% in isopropanol) for 1 h at RT. Cells were washed with water (three-four times) to remove excess stain. The stain was then quantified by dissolving in 300 μ l isopropanol and measuring the absorbance at 510 nm.

3.2.3 Reactive Oxygen Species (ROS) measurement

INS-1E cells were plated on 96-well plates (Eppendorf) at an initial density of 1×10^4 cells/well and were allowed to adhere at 37°C for 24 h. Post-incubation, the cells were treated with the six experimental conditions as mentioned before. ROS levels were assessed by incubating the cells with either DCFH-DA (10 μ mol) or MitoSox (5 μ mol) for 30 min at 37°C as described previously (Dubey et al. 2017). Simultaneous detection of cell viability was also carried out using the MTT assay (as described above). The data was normalized to the number of viable cells.

3.2.4 Metabolite extraction and sample preparation for NMR

INS-1E cells were seeded in 100 mm plates (in duplicate) at a density of 3×10^6 and were given different treatments as described above. Post-incubation, the cells were harvested by centrifugation for 5 mins at 100 g at 4°C and washed twice with 1X PBS. The cell pellets, thus obtained, were stored at -20°C until further use. There were seven replicates for each of the six experimental conditions. The cell culture pellets were re-suspended in 300 μ l of pre-chilled 1X PBS and 25 μ l was kept separately for protein isolation and estimation using BCA reagent (ThermoFisher Scientific). The cell suspension was mixed with two volumes of ice-cold methanol, vortexed briefly and incubated at -20°C for 30 minutes. Post-incubation, the samples were centrifuged at 13,000 g for 30 mins at 4°C. The supernatant was frozen into liquid nitrogen and lyophilized to remove residual water and methanol. The final extracts were stored at -80°C until NMR data acquisition. The lyophilized extracts were allowed to thaw on ice

before reconstituting into 580 μ l 100% NMR buffer (20 mM sodium phosphate, pH 7.4 in D₂O containing 0.4 mM DSS (2,2-dimethyl-2-silapentane-5-sulfonic acid)). Briefly, 17.46 ± 0.01 mg of DSS was weighed (mol wt. 218.32 g/mol) and dissolved in $2000 \mu\text{l} \pm 2 \mu\text{l}$ of phosphate buffer. This stock solution was then diluted to 100 fold resulting in a final buffer solution containing 87.30 ± 0.16 mg/L of DSS in solution, which corresponds to $399.9 \pm 0.7 \mu\text{M}$ of DSS in buffer. The samples were vortexed for 2 min at room temperature and centrifuged at 4000 g for 2 min. The supernatants were transferred to 5 mm NMR tubes for NMR measurements. A total of 42 distinct samples (seven replicates for six experimental conditions) were used for NMR data measurement.

3.3 NMR spectroscopy and spectral processing

All the ¹H NMR spectra were acquired at 298 K using the NOESY-presaturation pulse sequence noesygppr1d as described in detail in Chapter 2. For each spectrum, a total of 64 scans were collected into 32K data points with a spectral width of 7200 Hz. ¹H-¹H total correlation spectroscopy (TOCSY) experiment was performed to further assist in the resonance assignment of metabolites. All ¹H NMR spectra were manually phased, baseline-corrected, and referenced with respect to methyl singlet of DSS internal reference using Topspin (v3.5) (www.bruker.com/bruker/topspin).

3.4 Spectral assignment and quantitative evaluation

After spectral processing, annotation of metabolites was carried out with the Chenomx NMR Suite 8.1 (Edmonton, AB, Canada) software, which were further confirmed with biological magnetic resonance data bank (BMRB) (Ulrich et al. 2008) database and human metabolome database (HMDB) (Wishart et al. 2007). In addition, two dimensional ¹H -¹H TOCSY was used for further metabolite confirmation via a semi-automated software – MetaboMiner (Xia et al. 2008). All the identified metabolites were then quantified using the profiler module of Chenomx software, which enables metabolites quantification relative to an internal standard of known concentration (400 μM). The absolute concentrations were normalized with respect to the protein concentration obtained using BCA assay.

3.5 Statistical Analysis

All cellular experiments were conducted in triplicate and the results have been presented as mean \pm SEM (standard error of mean), unless stated otherwise. One-way ANOVA

(analysis of variance) and Tukey HSD post-hoc test methods were used to perform statistical analysis. $p\text{-value} \leq 0.05$ was considered to indicate statistically significant differences in groups.

Supervised Partial Least Squares Discriminant Analysis (PLS-DA) was performed using a free web server, MetaboAnalyst, to explore clustering patterns of observations, trends in the data and outliers. Before subjecting the data for multivariate analysis, row-wise normalization of the data matrix [containing N observation row vectors of various samples (experimental conditions and replicates) having K variables (metabolites) in columns] was performed with respect to the protein concentration of cellular lysate (measured before metabolite extraction) in order to normalize the data and reduce the variation within replicates. The metabolomics data sets were then subjected to Pareto-scaling prior to chemometric analysis. Variable importance of projection (VIP) score plot was generated from PLS-DA analysis to identify the metabolites responsible for the differential clustering of score plots in the PLS-DA model. Further one way ANOVA analysis followed by Fishers's least significant difference (LSD) post-hoc tests were used to assess the significance of changes in the levels of metabolite concentrations in all six experimental conditions used in this study. Pair-wise analysis of *INS-IE* cells cultured in basal glucose with different toxic conditions (HG, PA and HG/PA) were carried out using Volcano plot utility of MetaboAnalyst that screens important metabolites based on fold-change (FC) and FDR-adjusted p-value (q-value). Metabolites with fold change ≥ 1.5 and q-value ≤ 0.05 were considered significant.

3.6 Metabolic pathway analysis

Metabolic pathway analysis of all the discriminating metabolites was conducted to understand the biological significance of the metabolic changes and to identify target pathways that predominantly get affected under excess nutrient conditions (HG and/or FFA in *INS-IE* cells). The Metabolic Pathway Analysis (MetPa) was carried out on a free web-based program MetaboAnalyst (Chong et al. 2018; Xia, Wishart, and Valencia 2011). The metabolic pathways with impact value ≥ 0.1 and $p \leq 0.05$ were considered as significantly perturbed pathways.

3.7 Results

3.7.1 HG and/or PA exposure induces cytotoxicity in INS-IE cells

INS-IE cells were pre-incubated with media containing basal glucose, HG, PA (complexed with BSA), HG/PA and BSA (solvent control for PA) for 24 hours to evaluate the

effect of nutrient overload (high-glucose and FFA alone and/or in combination) on these cells. Cytotoxicity of *INS-1E* cells cultured in different conditions was evaluated using MTT (Figure 3.1A). From the cellular viability assays, we observed ~ 50% reduction in cellular viability of *INS-1E* cells to FFA (0.5 mM PA) exposure either alone or in combination with HG (HG/PA) and minimal effect on cellular viability was observed in *INS-1E* cells cultured in basal glucose (5 mM glucose), HG (16 mM glucose) and in BSA (solvent control for PA). The increased cytotoxicity in *INS-1E* cells exposed to PA and HG/PA was attributed to increased intracellular triglyceride content (Figure 3.1B) and increased lipid accumulation (as visualized by Oil Red O staining; Figure 3.1C and 3.1D). This is in line with earlier studies which report increased total FFA, triglyceride, and cholesterol content in rat islets exposed to increasing concentrations of glucose (4-25 mM) for 1 hour and INS-1(832/13) cells exposed increasing concentrations of glucose (5-20 mM) for 2 h (El-Azzouny et al. 2014; Mugabo et al. 2017).

Results from previous studies have shown that exposure to chronically elevated concentrations of glucose and non-esterified fatty acids (NEFAs) lead to induction of ROS levels that in turn mediate apoptosis in pancreatic β -cells (Poitout and Robertson 2008). Concomitant with these findings, increased intracellular ROS and mitochondrial superoxide levels (as determined by DCFH-DA and MitoSox, respectively; Figure 3.1E) were observed in *INS-1E* cells exposed to glucotoxic, lipotoxic and glucolipotoxic conditions.

Collectively, our cellular and biochemical data indicates that exposure of *INS-1E* cells to FFA and/or high glucose results in decreased cellular viability, increased ROS production, and accumulation of triglycerides and FFA in the pancreatic β -cells. All these results validate the establishment of glucotoxic, lipotoxic, and glucolipotoxic conditions in *INS-1E* cells.

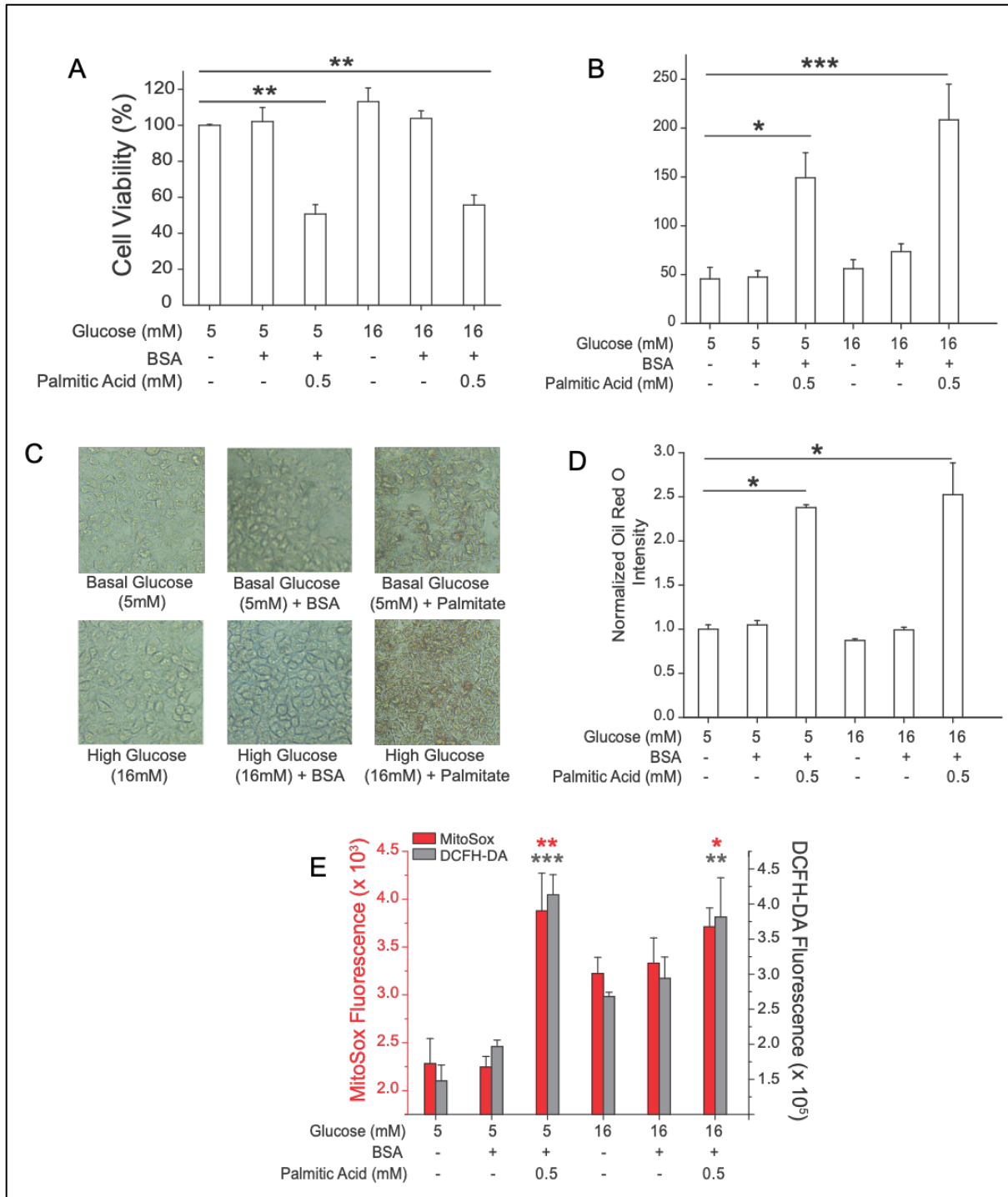


Figure 3.1: Effect of high glucose (HG), Palmitic acid (PA), and High glucose in combination with Palmitic acid (HG/PA) exposure on cellular viability, triglyceride content, lipid accumulation and ROS generation in *INS-1E* cells. (A) *INS-1E* cells were treated with HG, PA, and HG/PA for 24 h and cell viability was determined by MTT assay (N = 4). (B) Increased triglyceride content (in terms of mg% per mg protein) in cells exposed to PA and HG/PA (N=3). (C) Visualization of lipids using oil Red O staining in *INS-1E* cells when subjected to six different experimental conditions. Increased Oil Red O staining was observed in cells exposed to glucolipotoxic conditions, as quantitated in (D). (E) Measurement of Intracellular and mitochondrial ROS levels by DCFH-DA and MitoSox respectively, in *INS-1E* cells exposed to HG, PA and HG/PA conditions (N = 3). Figure provided by Ms. Devika Sardesai (Dr. Shilpy Sharma's lab at Dept. of Biotechnology, Savitribai Phule Pune University).

3.7.2 Global profiling of metabolites using ^1H NMR spectroscopy

Global metabolic profiling of pancreatic rat insulinoma cells - *INS-1E* cultured under basal glucose conditions were carried out and were compared with the metabolic profiles of *INS-1E* cells cultured under excess nutrient conditions (HG and/or FFA). A total of 48 metabolites were identified from the aqueous phase of the metabolic extracts of all the samples using standard one dimensional ^1H NMR (Figure 3.2). All the metabolites were identified using the profiler module of Chenomx NMR Suite 8.1 (Edmonton, AB, Canada) software. The organic phase containing the lipids/steroids/fatty acids and other water insoluble compounds were excluded from our analysis as they gave broad signals in ^1H NMR. Resonance assignment of these metabolites were further confirmed using 2D ^1H - ^1H TOCSY NMR spectroscopy (Figure 3.3). The NMR spectra presents signals mainly from amino acids, sugars, carbohydrates, and membrane metabolites. All the 48 metabolites and their respective ^1H chemical shifts (in reference to DSS) have been tabulated in Appendix Table 3.3.

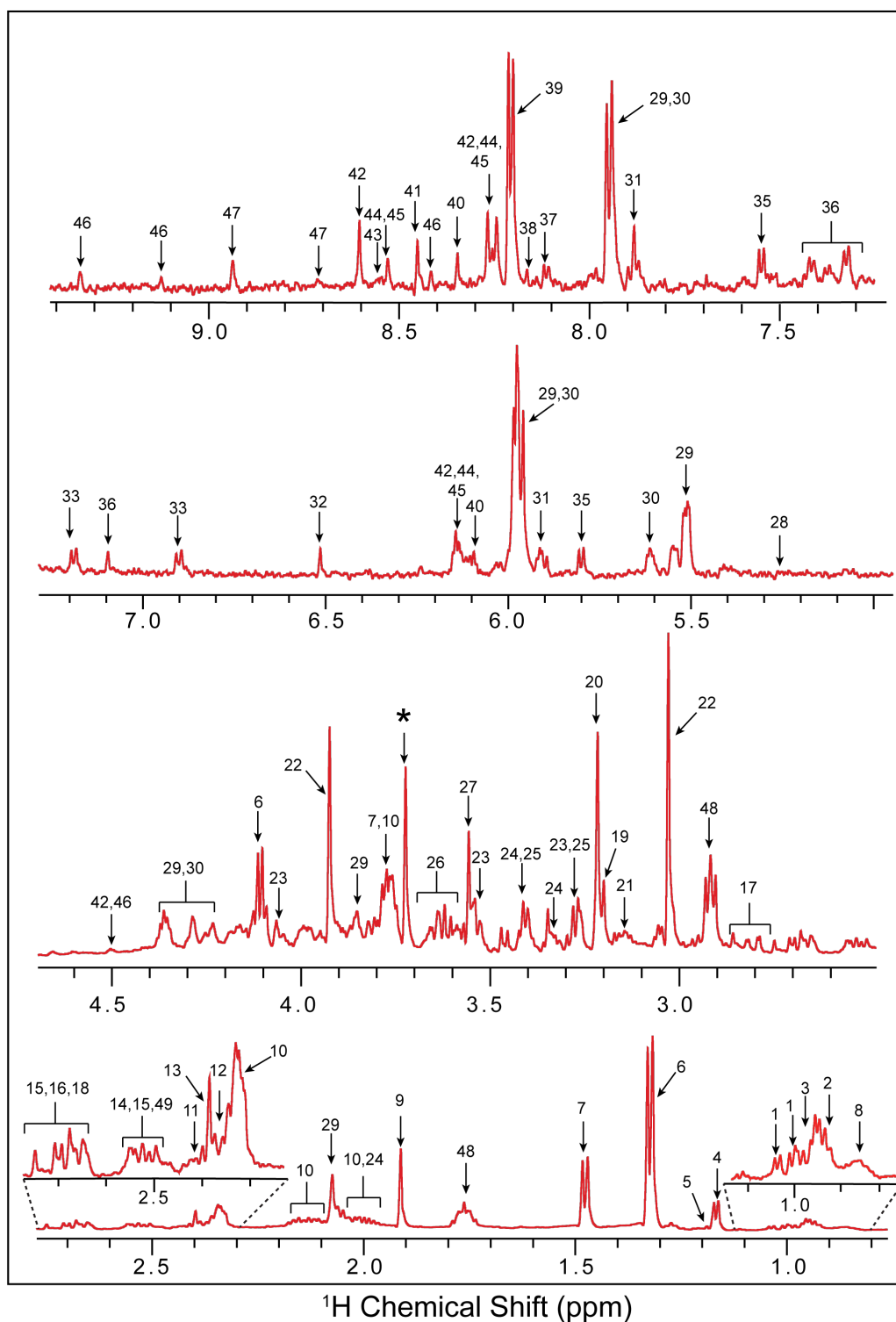


Figure 3.2: Representative ^1H -NMR spectrum of the methanolic extract of *INS-IE* cells grown in basal glucose (5 mM) containing media. 1, Valine; 2, Isoleucine; 3, Leucine; 4, Isopropanol; 5, Ethanol; 6, Lactate; 7, Alanine; 8, LDL; 9, Acetate; 10, Glutamate; 11, Glutamine; 12, Pyruvate; 13, Succinate; 14, Isocitrate; 15, Citrate; 16, Aspartate; 17, Asparagine; 18, Malate; 19, Choline; 20, o-Phosphocholine; 21, Citrulline; 22, Creatine; 23, Myo-inositol; 24, proline; 25, Taurine; 26, Glycerol; 27, Glycine; 28, Glucose ; 29, UDP-N-acetylglucosamine; 30, UDP-glucose; 31, Uridine; 32, Fumarate; 33, Tyrosine; 34, Phenylalanine; 35, Uracil; 36, Histidine; 37, UMP; 38, GTP; 39, Hypoxanthine; 40, Inosine; 41, Formate; 42, AMP; 43, IMP; 44, ADP; 45, ATP; 46, NAD; 47, Niacinamide; 48, DSS. The chemical shift and multiplicity details of the metabolic in the spectrum have been listed in Appendix Table 3.3.

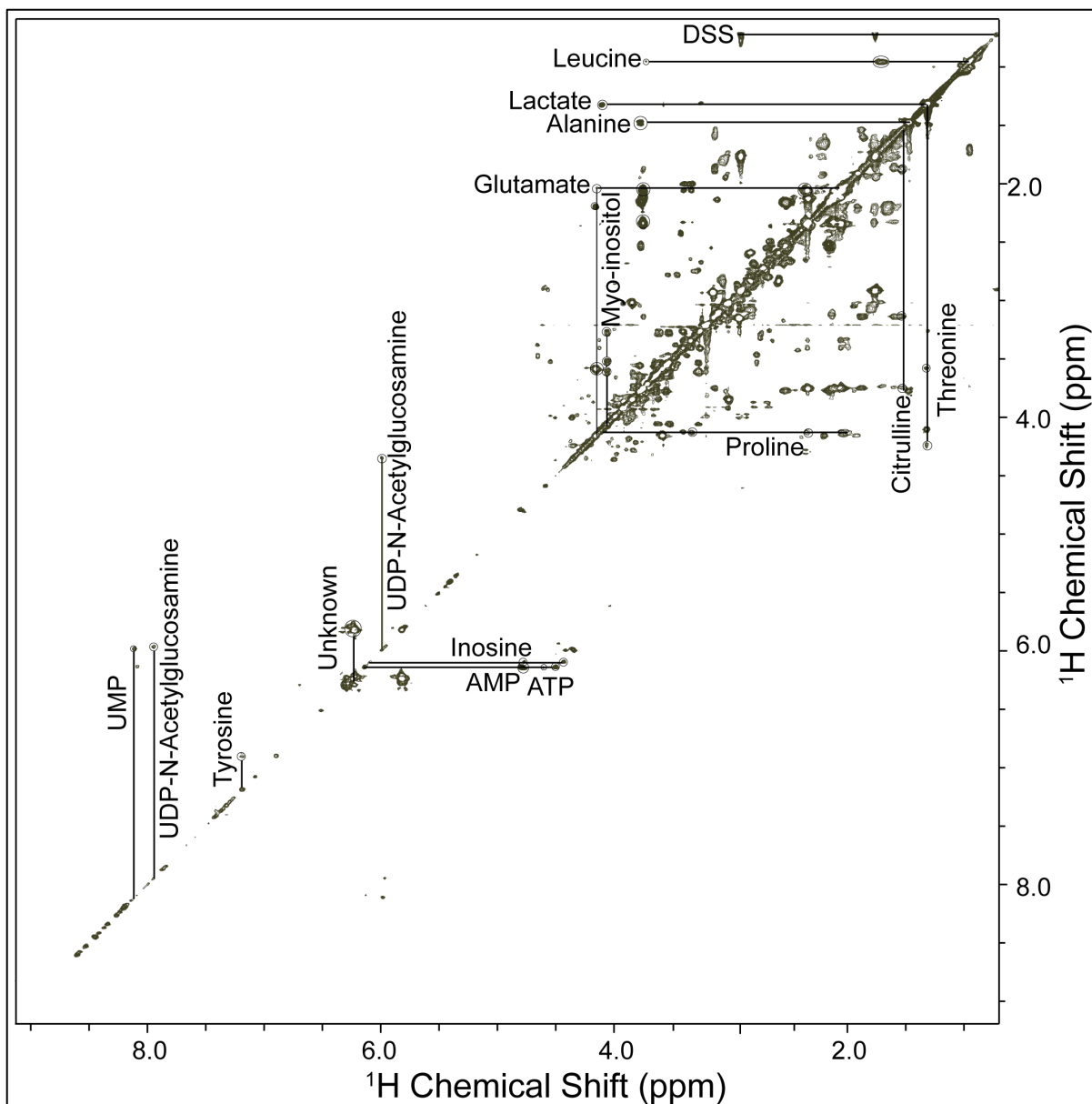


Figure 3.3: ^1H - ^1H TOCSY correlation spectrum of the methanolic extract of *INS-1E* cells cultured in media containing basal glucose concentration. Cross-peaks in the TOCSY spectrum were used to re-confirm the resonance assignments enlisted in Appendix Table 3.3.

All the identified metabolites were then quantified in the profile module of the Chenomx software. The concentration data of all the metabolites were used for feature selection and group discrimination using chemometric analysis. Univariate analysis using one-way ANOVA (followed by post-hoc) identified 24 statistically significant metabolites (Lactate, o-Phosphocholine, Taurine, Creatine, Glutamate, Glutathione, UDP-N-acetylglucosamine, Myoinositol, UDP-glucose, Niacinamide, ATP, Proline, NAD, ADP, Fumarate, Malate, IMP, Choline, Glucose, Glutamine, Isocitrate, Citrate, Asparagine, and Alanine) that display

significant differences (FDR corrected p-value ≤ 0.05) in their concentrations across different experimental conditions (Figure 3.4). Box-Whisker plots for these metabolites (in the order of significance; p-value low to high) have been depicted in Figure 3.5 for a visual interpretation of dysregulated metabolites in HG, PA and HG/PA conditions used in this study.

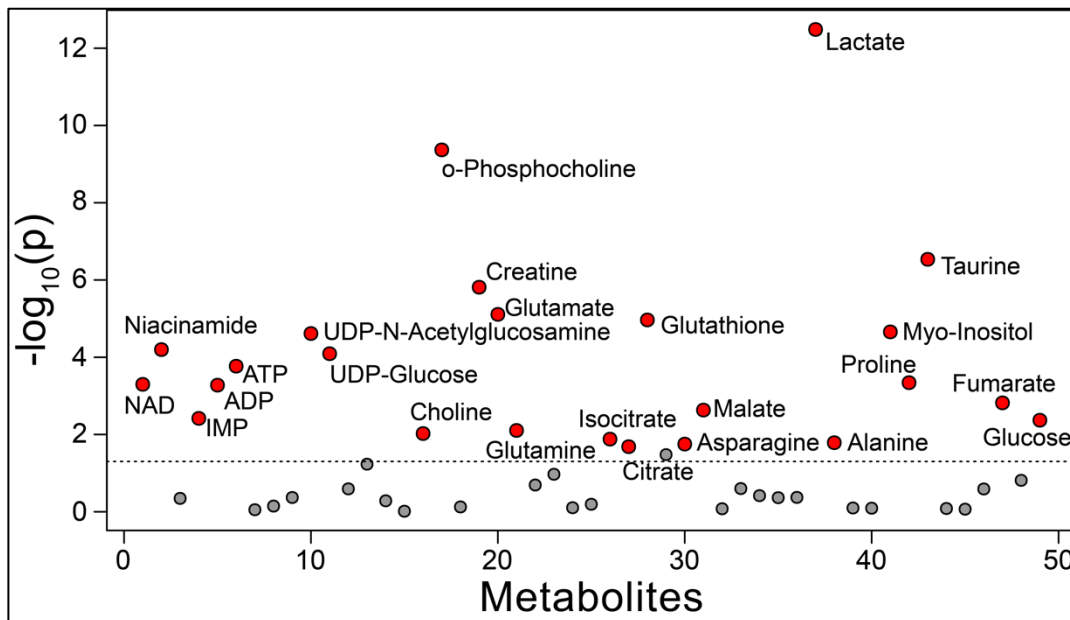


Figure 3.4: One-way ANOVA analysis followed by post-hoc analysis showing 24 significantly regulated (q-value ≤ 0.05 , marked in red circles) metabolites. The dotted horizontal line corresponds to threshold of FDR-adjusted p-value (q-value ≤ 0.05) on a log10 scale.

In the current study, PLS-DA analyses was conducted to investigate the changes in the metabolic profiles of *INS-IE* cells subjected to gluco-, lipo- and glucolipo-toxic conditions and to identify the metabolic signatures associated with different nutrient overload conditions. PLS-DA analysis has widely been used to represent the difference in data across a large set of variables as a smaller set of ‘principal variables’ in the form of scores plot. From the supervised PLS-DA analysis, we observed a differential clustering of *INS-IE* samples in the first and the second principal components. The diverged score plots signified a separation between the basal glucose and various toxic condition along the axes corresponding to PC 1 and 2 with these two principal components accounting for 37.1 % and 22.8% of the variation in the data, respectively (Figure 3.6A). The differential separation of clusters on the score plot of the first two principal components between *INS-IE* cells grown in basal glucose with HG, PA, and HG/PA highlights the differences in metabolome. A high degree of overlap was observed between the score plots

corresponding to BSA control samples, i.e., basal glucose and BSA with basal glucose; and HG/BSA with HG indicating no significant changes in the

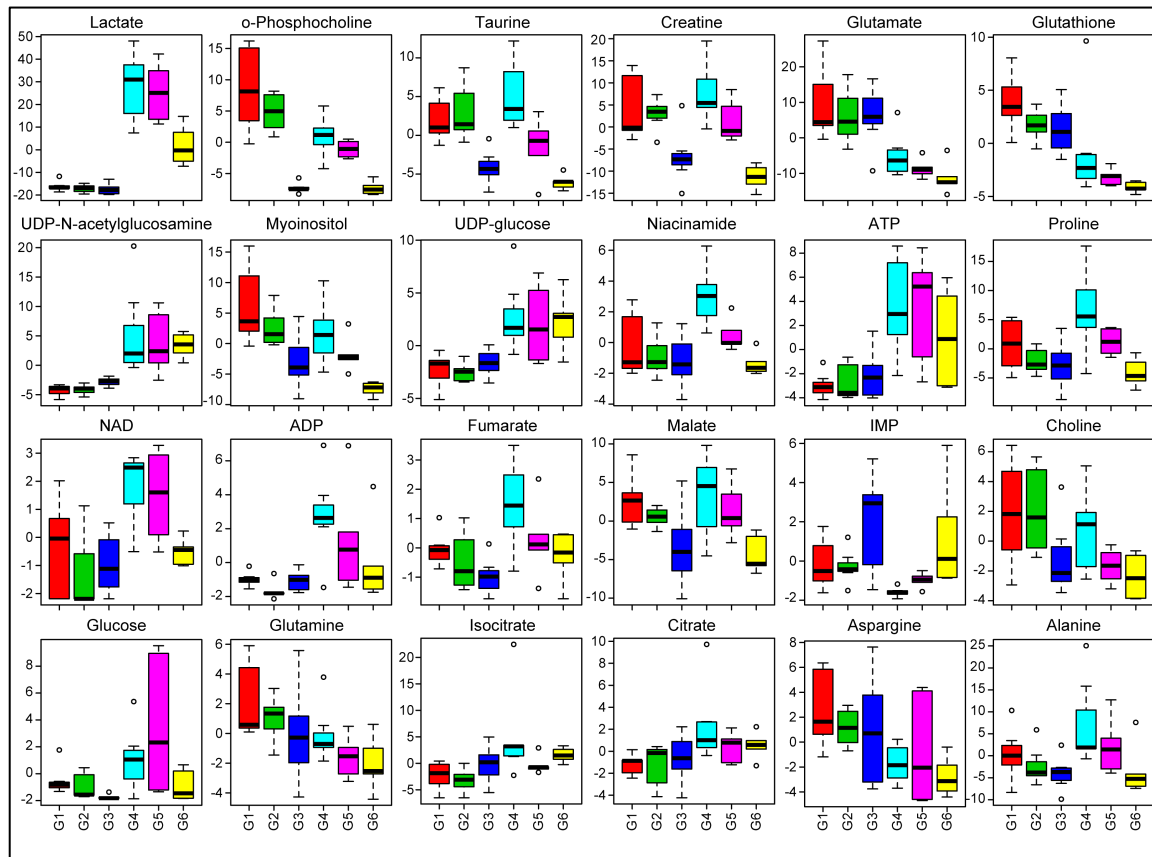


Figure 3.5: Box-Whisker plots displaying the comparative variation in concentration of each individual significantly (q -value ≤ 0.05) altered metabolite identified from ANOVA and post-hoc analysis against six experimental conditions; namely, G1 (Basal glucose; red); G2 (Basal Glucose with BSA as vehicle control for palmitate; green); G3 (High PA; blue); G4 (High glucose; cyan); G5 (High glucose with BSA; magenta); and G6 (High glucose and PA; yellow).

metabolic profiles of *INS-1E* cells grown in presence of BSA (vehicle control for palmitate), thereby negating the effect of BSA on the metabolic perturbations observed amongst the different groups. VIP scores were obtained from PLS-DA model to identify the discriminatory metabolites contributed most to the group separation in the PLS-DA models for each growth condition (Figure 3.6B). The VIP score is based on PLS loadings and reflects the influence of each variable to the overall model, with metabolites having VIP scores ≥ 1 considered important in classification. The discriminatory metabolites were organised in the descending order of the VIP score in component 1. Based on VIP score (VIP ≥ 1), the key metabolites in *INS-1E* samples discriminating among basal glucose, HG, PA, and HG/PA growth conditions were lactate, glutamate, o-phosphocholine, aspartate, myo-inositol, UDP-N-

acetylglucosamine, creatine, glutathione and ATP. These metabolites are largely contributing to the differential clustering of score plots in the PLS-DA analysis.

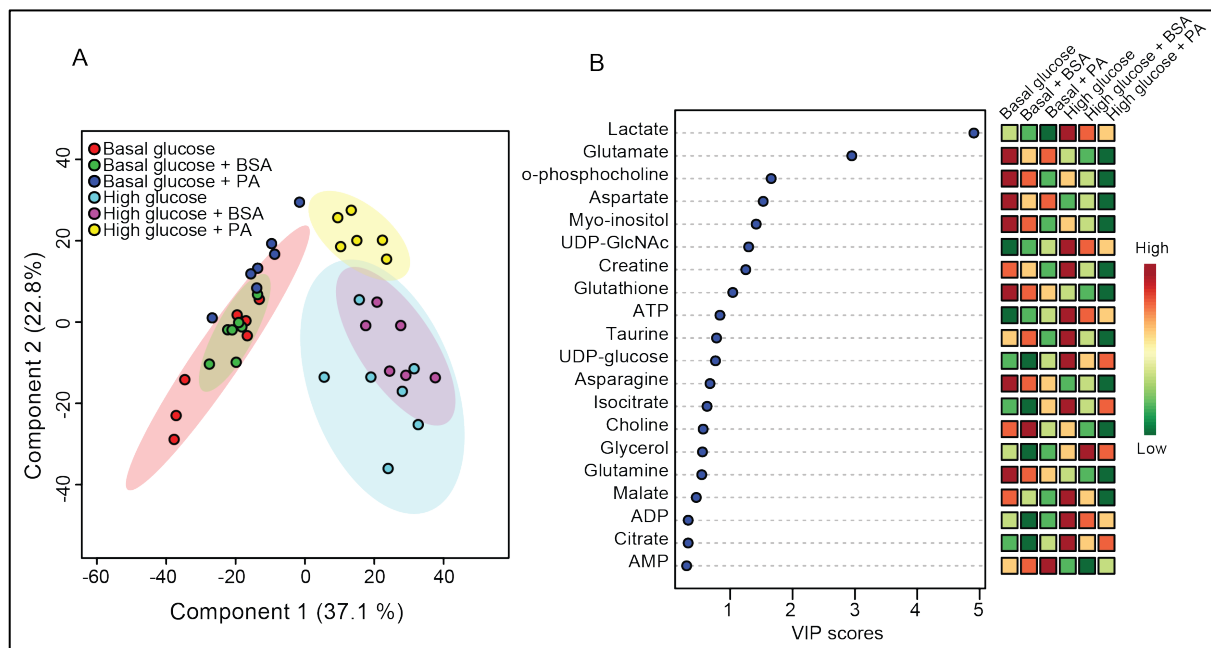


Figure 3.6: Score plots of PLS-DA (A) and variable importance in projection (VIP) (B) obtained from MetaboAnalyst 4.0. PLS-DA score plots of *INS-1E* samples grown under basal glucose (red, normal growth), basal glucose with BSA (vehicle control for palmitate) (green), basal glucose with palmitic acid (blue, lipotoxic), high glucose (cyan, glucotoxic), high glucose with BSA (magenta) and high glucose with palmitate (yellow, glucolipotoxic). For each of the six growth conditions, N=7 replicates were used. Each ellipse on the scores plot represents 95% confidence limits of a normal distribution for each group. VIP plot displays the most significantly altered metabolites detected by PLS-DA model analyses arranged in descending order of VIP score in component 1. The coloured boxes on the right indicate the relative concentrations of the corresponding metabolite in each condition used. The relative concentration of metabolites is represented by a coloured scale from green to red indicating the low and high, respectively.

3.7.3 Pair-wise analysis of metabolic changes

Pair-wise analysis of metabolic extracts obtained from *INS-1E* cells cultured under basal glucose; and HG, PA, and HG/PA conditions was performed using: 1) PLS-DA to discern the differences in the metabolic profiles between basal versus the nutrient overload condition; and 2) Volcano-plot analysis to derive the significantly dysregulated metabolites based on the fold change and p-value (Figure 3.7).

3.7.3.1 High-Glucose induced changes

To investigate the effect of high-glucose, metabolic profiles of *INS-1E* cells cultured under supraphysiological glucose concentration (16 mm) were compared with those cultured in media containing basal glucose (5G) for 24 hours. The PLS-DA analysis revealed differential clustering between basal glucose and glucotoxicity groups in the score plots

accounts for 46% variation in PC1 and 6.2% variation in PC2 (Figure 3.7A). Sixteen metabolites were identified to be significantly dysregulated under glucotoxicity condition as compared to basal condition (Figure 3.7B). The significantly upregulated metabolites upon HG exposure include lactate, UDP-N-acetylglucosamine, ATP, ADP, UDP-glucose, NAD, isocitrate, niacinamide, fumarate and citrate. On the other hand, levels of glutamate, glutathione, asparagine, o-phosphocholine, IMP and aspartate were found to be significantly down-regulated under glucotoxicity condition. Most of the perturbed metabolites under high glucose condition belongs to energy metabolism and amino acid metabolism.

3.7.3.2 High-fat induced changes

The comparison of metabolic profiles of cells grown under basal glucose conditions with those exposed to high saturated fatty acid content (PA) showed a 46.5% variation in PC1 and 19% variation in PC2 in the PLS-DA analysis (Figure 3.7C). Based on the $FC \geq 1.5$ and $FDR\text{-adjusted } p\text{-value} \leq 0.05$ in the volcano-plot analysis (Figure 3.7D), we observed the level of UDP-N-acetylglucosamine was increased under lipotoxic conditions whereas the levels of malate, creatine, myoinositol, taurine, glucose and o-phosphocholine were found to be significantly down-regulated when cells were grown under PA condition. Moreover, it was also observed UDP-N-acetylglucosamine and o-phosphocholine were the only common metabolites that were dysregulated significantly under both glucotoxic and lipotoxic conditions, indicates the possibility of different pathway perturbations under these two conditions.

3.7.3.3 High-Glucose and free fatty acid induced changes

The combination of elevated glucose (hyperglycemia) and FFA (hyperlipidemia) that leads to glucolipotoxicity, is the most detrimental to pancreatic β -cell viability and function (El-Assaad et al. 2003; Kim and Yoon 2011; Yazıcı and Sezer 2017). The combined effect of elevated glucose and elevated lipid levels have been associated with decreased insulin secretion, impaired insulin gene expression, and β -cell death by apoptosis (Cerf 2013). Indeed, INS 832/3 cells and human islets have been shown to undergo apoptosis on exposure to saturated free fatty acids, such as palmitate and stearate along with high glucose (El-Assaad et al. 2003). While there are studies in the literature which have evaluated metabolite alterations in response to high glucose and free fatty acids individually, there are no large-scale assessments of metabolic perturbations in response to long-term exposure of glucolipotoxic

conditions on pancreatic β -cells.

PLS-DA score plot analysis of *INS-1E* cells grown under glucolipotoxic conditions revealed a 51.8% variation in PC1 and 21% variation in PC2 in a pairwise comparison with the basal glucose conditions (Figure 3.7E). Overall 18 metabolites were found significantly dysregulated using volcano-plot analysis ($FC \geq 1.5$; $q\text{-value} \leq 0.05$) under glucolipotoxic condition when compared to basal glucose conditions (Figure 3.7F). Out of these, the levels of UDP-N-acetylglucosamine, lactate, ATP and UDP-glucose were found to be significantly elevated under glucolipotoxicity and the levels of 14 metabolites namely, proline, alanine, malate, glutamine, GTP, asparagine, glutamate, choline, glutathione, creatine, taurine, myoinositol, aspartate, and o-phosphocholine were found to be significantly downregulated under HG condition in comparison with basal glucose condition.

It is interesting to note that even though there were some common dysregulated metabolites identified from HG and FFA conditions with those from glucolipotoxic conditions, but the dysregulation of these overlapping metabolites were much more amplified under glucolipotoxic conditions as compared to glucotoxic and lipotoxic conditions (Figure 3.5). As visualized from the box-whisker plots, the levels of o-phosphocholine, taurine, myoinositol, glutathione, creatine, aspartate, glutamate, asparagine, malate were drastically reduced under glucolipotoxic conditions when compared to glucotoxic and lipotoxic conditions. Similarly, the levels of UDP-N-acetylglucosamine were increased (Figure 3.5) in *INS-1E* cells under glucolipotoxic condition when compared glucotoxic and lipotoxic conditions. In addition, a new group of metabolites including proline, GTP, choline, glutamine, and alanine was observed to be significantly down-regulated under glucolipotoxic conditions, but not significantly altered under glucotoxic and lipotoxic conditions. Similarly, a few metabolites including ADP, niacinamide, NAD, fumarate, citrate, and isocitrate which displayed substantial upregulation under glucotoxic conditions did not display a major difference under glucolipotoxic condition. The presence of differentially altered metabolites under the different stress conditions used in the study thereby suggests the involvement of different metabolic pathway perturbations in *INS-1E* cells.

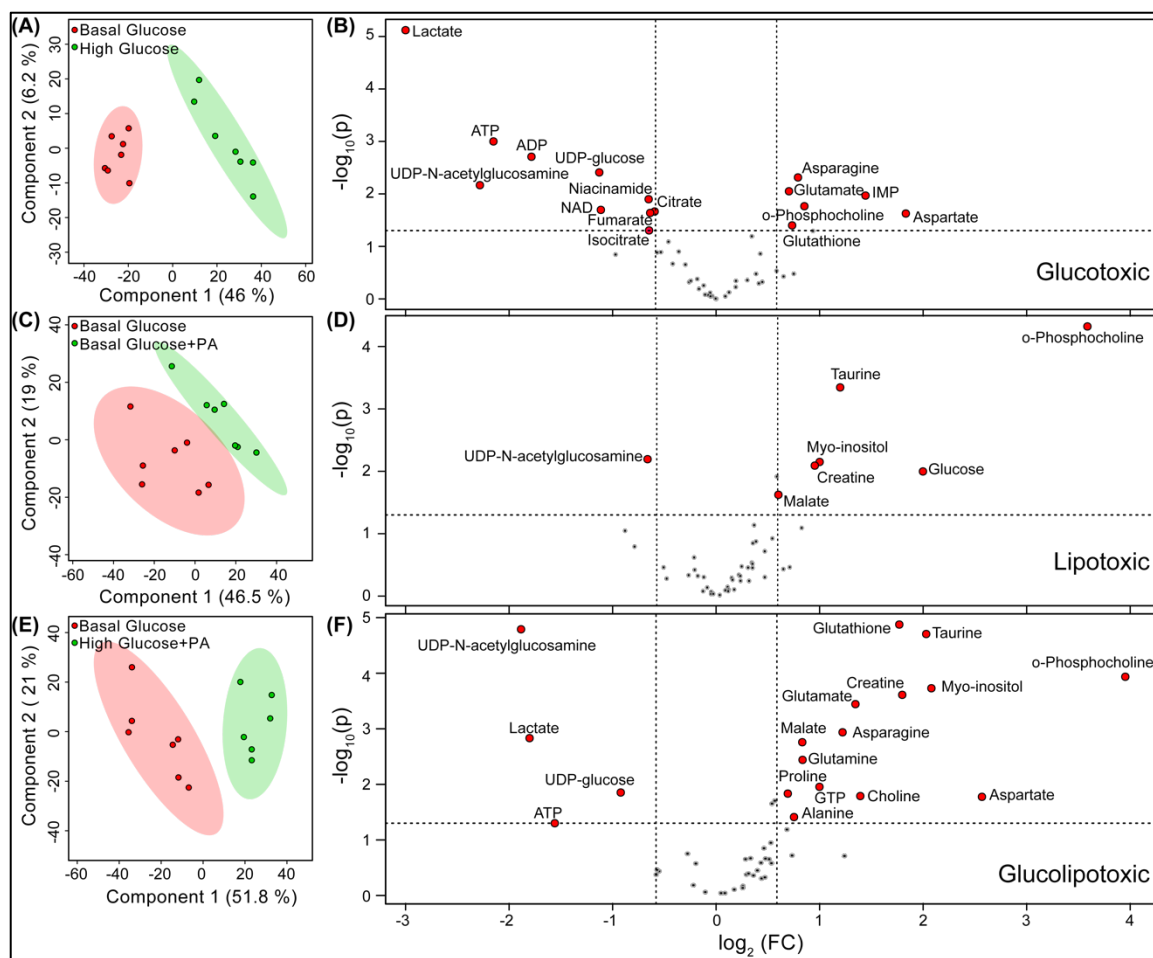


Figure 3.7: Pairwise metabolite comparison between *INS-1E* cells cultured in, basal glucose vs high glucose (HG) (A) PLS-DA score plot and (B) volcano plot; basal glucose vs High FFA (PA) (C) PLS-DA score plot and (D) volcano plot; and basal glucose vs high glucose and FFA (HG/PA) (E) PLS-DA score plot and (F) volcano plot. In all the PLS-DA score plots, basal glucose sample groups were depicted in red colors, whereas toxic-stressed sample groups (HG, PA and HG/PA) has been highlighted with green colors in their respective score plots. Ellipses on the score plots showing 95% confidence limits of a normal distribution for each group. Each dot on the score plot represents an individual sample. In the volcano plot, vertical dotted lines indicate threshold of ± 1.5 -fold changes in concentration of metabolites on a \log_2 scale and the horizontal dotted line corresponds to threshold of p-value significance ($p \leq 0.05$) on a \log_{10} scale. Red dot in the volcano plots represents metabolites that met both thresholds for significant change, with up-regulated metabolites on the left-hand side of each plot and down-regulated metabolites on the right-hand side.

3.8 Metabolic Pathway Analysis

To gain insights into the metabolic mechanism of T2DM and to assess the biological relevance of the changes in metabolite levels, metabolic pathway analysis of substantially differential metabolites was carried under glucotoxic, lipotoxic, and glucolipotoxic conditions using MetaboAnalyst tool and KEGG pathway database. The metabolites identified through Volcano-plot analysis ($q\text{-value} \leq 0.05$, $FC \geq 1.5$) were employed for the detailed analysis of the altered pathways in T2DM. These metabolites were then explored for metabolic pathway analysis using the *Rattus norvegicus* (rat) pathway library with parameters “Hypergeometric

test” for Over-Representation Analysis and “Relative-between-ness Centrality” for pathway topology analysis (Xia et al. 2011). Pathway enrichment analysis was performed to identify all significant pathways ($p\text{-value} \leq 0.05$) based on the fold enrichment ratio (hits/expected) in different nutrient overload conditions used in our study. The significantly perturbed pathways were then identified based on their pathway impact values calculated from pathway topology analysis. All the identified metabolic pathways were plotted with $p\text{-value}$ along $y\text{-axis}$ and pathway impact score along $X\text{-axis}$ (Figure 3.8). The metabolic pathways with a $p\text{-value} \leq 0.05$ and impact score ≥ 0.1 were deemed as significantly perturbed pathway with great relevance to disease mechanism.

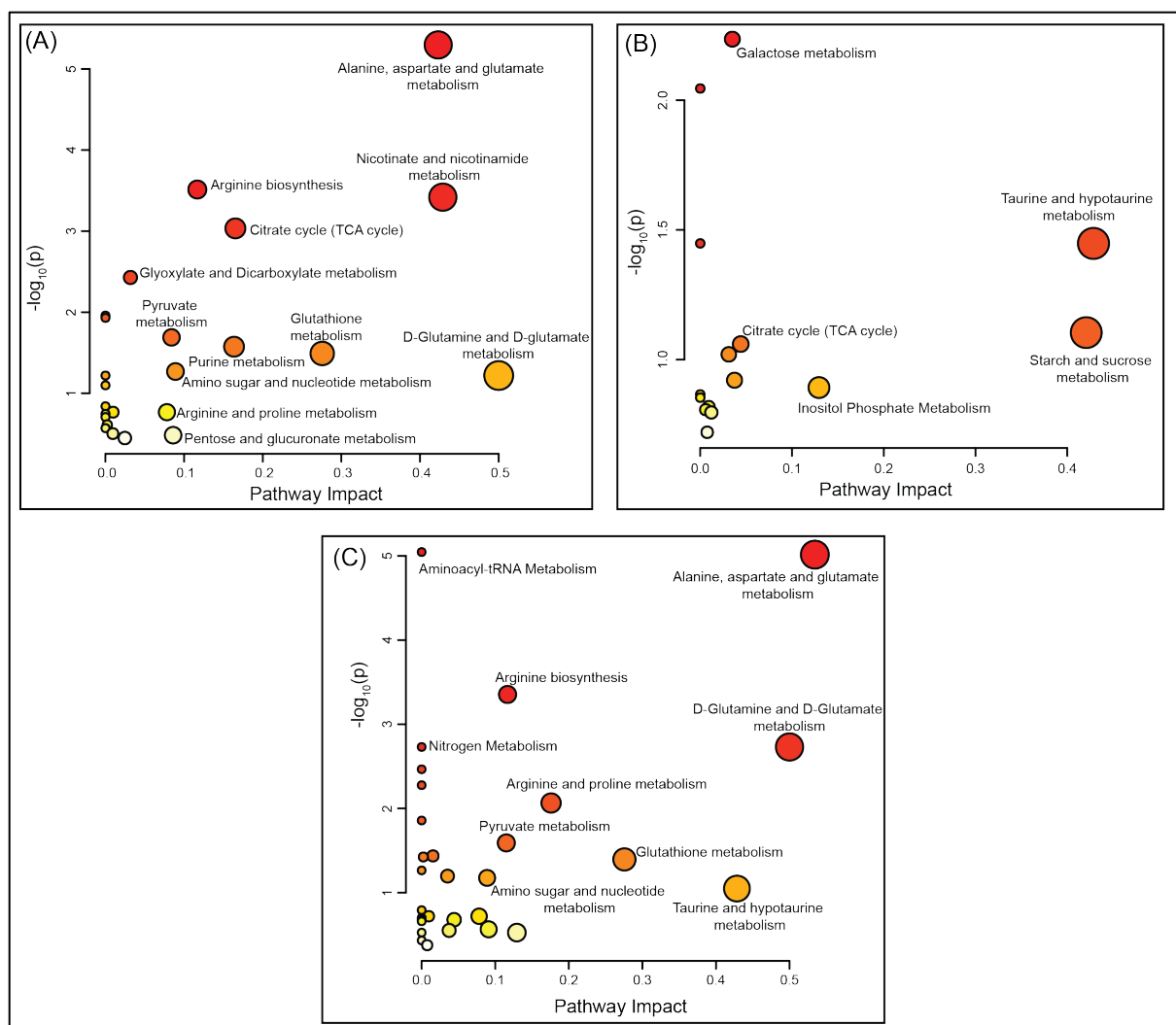


Figure 3.8: Metabolic Pathway Analysis (MetPa) of significantly altered metabolites displaying significant metabolic pathways in *INS-1E* cells when associated with (A) glucotoxic, (B) lipotoxic, and (C) glucolipotoxic conditions. Each node represents a single metabolic pathway with node color corresponding to the $-\log_{10}(P)$ value (red: higher $p\text{-values}$ and yellow: lower $p\text{-values}$) and node size corresponding to the pathway impact score.

The major pathways perturbed significantly in the pairwise comparison under glucotoxic condition were alanine, aspartate and glutamate metabolism, nicotinate and nicotinamide metabolism, citrate cycle, arginine biosynthesis, glutamine-glutamate metabolism, arginine and proline metabolism, pyruvate metabolism, glyoxylate and dicarboxylate metabolism, glutathione metabolism, purine metabolism, amino sugar and nucleotide metabolism and pentose and glucuronate metabolism (Figure 3.8A). Similarly, the pathways perturbed significantly under lipotoxic condition were starch and sucrose metabolism, taurine and hypotaurine metabolism, galactose metabolism, citrate cycle and inositol phosphate metabolism (Figure 3.8B). Significantly pathways perturbed under glucolipotoxic condition included alanine, aspartate and glutamate metabolism, glutamine-glutamate metabolism, arginine-proline metabolism, arginine biosynthesis, pyruvate metabolism, aminoacyl-tRNA biosynthesis, glutathione metabolism, glycerophospholipid metabolism, amino sugar and nucleotide metabolism, taurine and hypotaurine metabolism (Figure 3.8C). Overall the significantly perturbed pathways identified from the current study belonged to amino acid metabolism, TCA cycle, nucleotide metabolism, hexosamine pathway and energy metabolism. All the identified pathways were then constructed using KEGG metabolic network and have been pictorially depicted in Figure 3.9.

3.9 Discussion

Diabetes mellitus, a chronic metabolic disease is associated with profound changes in the energy metabolism in diabetic subjects. The mapping of intracellular metabolites onto metabolic pathways under different nutrient overload conditions revealed that the changes were majorly focused around amino acid metabolism, TCA cycle, nucleotide metabolism, nitrogen metabolism pathways. Chronic exposure of pancreatic β -cells to high glucose results in an elevated metabolic flux into the TCA cycle. The increased levels of TCA cycle intermediates such as citrate, isocitrate, and fumarate under high glucose conditions indicates the upregulation of TCA cycle, a major pathway for the generation of energy and production of metabolic precursors for the biosynthesis of non-essential amino acids. This was associated with an increase in ATP and NAD^+ levels that is required for maintaining glycolytic flux. Previous studies have also reported the increased flux of TCA cycle in INS-1 832/13 cells cultured in increasing concentrations of glucose (5-20 mM) for 2 h (Mugabo et al. 2017). It has been reported that the malate-aspartate shuttle is associated with glucose metabolism in the pancreatic beta cells and plays an important role in linking glucose metabolism to insulin

secretion in these cell (Yokoi et al. 2016). Indeed, the depleted levels of aspartate, asparagine, glutamate, and glutamine under glucotoxic conditions indicate defects in the functioning of the malate-aspartate shuttle. Further, a significant reduction in malate levels was observed during lipotoxic and glucolipotoxic conditions, thereby indicating a reduction in TCA cycle intermediates under these conditions. These observations gain support from previous reports where reduced levels of malate have been observed in high-fat fed obese rats (Eccleston et al. 2011).

Taurine, a sulphonic acid, plays an important role in numerous biological and physiological functions such as maintenance of intracellular osmotic balance, acts as an antioxidant in biological systems, involved in modulation of neurotransmitters, etc. (Ripps and Shen 2012). In our study, decreased levels of taurine were detected in *INS-1E* cells cultured under lipotoxic and glucolipotoxic conditions. DM is known to be associated with a decrease in endogenous taurine levels in a variety of tissues, thereby enhancing the notion that alterations in taurine levels could also lead to the severity of ROS-mediated damage. Insulin-like effects such as accelerating glucose uptake in tissues and glycogen synthesis in the liver have been suggested for taurine (Kim et al. 2009). Several reports have investigated the effects of taurine supplementation in alleviating hyperglycemia, plasma HbA1c and dyslipidemia in animal models associated with diabetes (Kim et al. 2012a; El Mesallamy et al. 2010). It was also shown to improve insulin secretion and insulin sensitivity in acute glucose or lipid infusion models (Haber et al. 2003). In addition to this, significantly lower plasma taurine levels have been reported in streptozotocin- or alloxan-induced diabetic animals as well as in T2DM patients (Ito, Schaffer, and Azuma 2012a; Nakamura et al. 2014; Trachtman et al. 1995). Urinary metabolic profile of high-fat fed rats and Zucker obese rats also showed a significant differences in the taurine levels (Kim et al. 2009; Williams et al. 2006). In Wistar rats, taurine has improved the islet dysfunction induced by free fatty acids (Kim et al. 2012b). These important roles of taurine tend to be primarily based on its antioxidant property, as well as on various protective effects against high glucose toxicity in pancreatic cells (e.g., the modulation of mitochondrial calcium handling and the stabilization of protein folding) (Ito, Schaffer, and Azuma 2012b).

D-myoinositol, a cyclitol present in animal and plant cells has been shown to have insulin mimetic effects in insulin resistant animals (Ortmeyer 1996). In fact, dietary supplementation of D-myoinositol has also been reported to be effective in improving glycemic control in T2DM subjects (Pintaudi, Di Vieste, and Bonomo 2016); and in improving insulin

sensitivity and reducing fat accretion in high-fat fed mice (Croze, Gélöën, and Soulage 2015). In concurrence with these investigations, the current study also reports a significant reduction in D-myoinositol levels in lipo-and glucolipo-toxic conditions.

Glutathione is an important intracellular antioxidant that plays important role in resisting the oxidative stress by scavenging the radicles in the cell. Besides serving a protective role against ROS formation, glutathione has been primarily involved in triggering the signalling pathway to regulate the first phase of GSIS (Liu et al. 2015). Low levels of glutathione was observed in the *INS-IE* cells exposed to glucotoxic and glucolipotoxic conditions in our study. Previous studies indicate that chronic exposure of cells to high glucose results in decrease in glutathione levels, and thus eventually increases the ROS levels which is detrimental for the cells (Kalkan and Suher 2013; Lutchmansingh et al. 2018). This decrease in the level of glutathione concentrations could be attributed both as a results of competition between aldose reductase and glutathione reductase for NADPH, a cofactor, and increased oxidative stress (increased ratio of NADH/NAD) (De Mattia et al. 1994).

Lactate, an indicator of oxidative capacity, predicts the prevalence of diabetes independent of several other risk factors and is closely linked to insulin resistance markers (Juraschek et al. 2013). From our metabolomics data, lactate was observed to be the most significant metabolite with increased levels during glucotoxicity and glucolipotoxicity conditions. Elevated lactate levels have been associated with the development of insulin resistance and have been considered to be the primary biological hallmark of diabetes progression. Indeed, significantly high lactate levels have been reported from serum and plasma of obese and insulin resistant subjects (Doar, Wynn, and Cramp 1968; Lovejoy et al. 1992). In addition, higher level of expression of Lactate dehydrogenase (LDH), an enzyme that facilitates the conversion of pyruvate to lactate, has been reported in diabetic subjects (Ainscow, Zhao, and Rutter 2000).

Creatine, a natural amine, is known to have protective effects against hyperglycaemia and has been shown to have antioxidant activity (Ročić et al. 2011; Stefani et al. 2014). Decreased levels of creatine were observed in *INS-IE* cells cultured under lipotoxic and glucolipotoxicity conditions. Earlier reports indicate that dietary creatine supplementation helps ameliorate hyperglycemia and helps improve glucose metabolism and glycemic control in T2DM subjects (Gualano et al. 2011; Op 'T Eijnde et al. 2001); and influences lipid metabolism to promote lipid secretion and oxidation and prevents lipid accumulation in cultured liver cells and high-fat-fed SD rats (Deminice et al. 2011; da Silva, Leonard, and Jacobs 2017).

Perturbation in amino acid metabolism was observed in all nutrient overload conditions used in the current study. Under suitable conditions, amino acid metabolism stimulates the regulation of insulin secretion in the β -cells (Zhang and Li 2013). Both glutamine and alanine which are the most abundant amino acid in the blood can regulate β -cell function and enhance insulin secretion from the primary islets and pancreatic β -cell line (Newsholme et al. 2005). Glutamate can play a pivotal role in triggering insulin secretion directly, either through metabolism or via its role in enhancing malate/aspartate shuttle activity (Newsholme, Brennan, and Bender 2006). Therefore any chronic changes in the levels of these metabolites could result in dysregulated insulin secretion *in vivo*.

Uridine diphosphate glucose, also known as UDP-glucose is a nucleotide sugar and is a key intermediate in the carbohydrate metabolism. Elevated levels of UDP-glucose were detected in *INS-1E* cells exposed to HG and HG/PA. UDP-glucose is an activated form of glucose and serves as the glycosyl donor for the biosynthesis of glycogen (Zois and Harris 2016). Reduced UDP-glucose flux and glycogen synthesis in the liver has also been reported in diabetic subjects when compared to non-diabetic subjects (Basu et al. 2001). The alterations in the levels of UDP-glucose under these conditions clearly indicates its known tendency to synthesize glycogen in diabetes (Spiro 1984).

Down regulation of o-phosphocholine, an intermediate in the synthesis of phosphatidylcholine, was identified as a common significantly dysregulated metabolite in *INS-1E* cells cultured under HG, PA, and HG/PA conditions. The combined effect of elevated glucose and elevated FFA results in activation of esterification pathway, leading to generation of lipid signalling molecules including phosphatidylcholine, triglyceride, ceramides (Poitout and Robertson 2008). These molecules have been directly shown to be involved in β -cell dysfunction. Indeed, *de novo* synthesis of FFAs and upregulation of fatty acid metabolism enzymes, followed by remodeling of the lipid composition of the plasma membrane have been proposed to be associated with high-glucose exposure associated with glucolipotoxicity in pancreatic β -cells (Couté et al. 2010; Nyblom et al. 2008; Wallace et al. 2013). Also, increased *de novo* synthesis of both saturated and unsaturated fatty acids has been reported in *INS-1E* cells exposed to increased glucose concentrations, thereby supporting our results (Nyblom et al. 2008).

Uridine diphosphate-*N*-acetylglucosamine (UDP-GlcNAc), a sensor molecule and high energy donor substrate for OGT (O-GlcNAc transferase) is the major end product of hexosamine pathway, an important pathway for nutrient sensing. Increased flux of hexosamine

pathway, which plays a key role in the production of insulin, has been reported during hyperglycemia (McClain 2002a). The excess glucose due to the hyperglycemic conditions observed in diabetes is shunted in the hexosamine pathway that leads to increased levels of UDP-GlcNAc. It has been hypothesized that an elevated levels of UDP-GlcNAc under hyperglycemic conditions is required for post-translational protein O-glycosylation events in T2DM and metabolic syndrome (Slawson, Copeland, and Hart 2010). Earlier reports have also shown the increased levels of UDP-GlcNAc in streptozotocin-induced hyperglycemic rats (Robinson et al. 1995); high fat-fed normal rats (McClain 2002b); normal rats where hyperglycemia was induced using glucose (Hawkins et al. 1997; Robinson et al. 1995); in animals with prolonged elevation of serum FFA (Hawkins et al. 1997; Robinson et al. 1995); and human obese and T2DM subjects (Pouwels et al. 2004; Testa et al. 2015). In our study, we detected UDP-GlcNAc as a common dysregulated metabolite that showed upregulation under all three nutrient-overloaded conditions (Figure 3.7). Interestingly, the fold increase for this metabolite was much higher under gluco- (Figure 3.7B) and glucolipo-toxic (Figure 3.7F) conditions when compared to lipotoxic (Fig. 5d) conditions alone, thereby hinting towards a bigger role played by this metabolite under hyperglycemic conditions.

In order to find unique metabolic biomarker fingerprint associated with glucotoxicity, lipotoxicity, and glucolipotoxicity, a ratio of average normalized concentration of the common dysregulated metabolites, namely, o-phosphocholine to UDP-N-acetylglucosamine was calculated and compared. These ratios were $\sim 5.98 (\pm 1.78)$ in HG exposure, $\sim 19.09 (\pm 6.81)$ in the presence of PA and $\sim 51.79 (\pm 24.38)$ in HG/ PA exposure when compared to basal glucose (Table 3.2), thereby suggesting that these ratios can be used to identify the major type of metabolic dysregulation associated with pancreatic β -cell death and dysfunction. Thus, o-phosphocholine to UDP-N-acetylglucosamine ratio can be used as a potential biomarker to identify glucotoxic, lipotoxic, and glucolipotoxic metabolic imbalances associated with T2DM.

Table 3.2: Calculation and comparison of ratio of o-phosphocholine to UDP-N-GlcNAc under gluco-, lipo-, and glucolipotoxicity conditions in comparison to basal glucose.

	Glucotoxic	Lipotoxic	Glucolipotoxic
o-phosphocholine	1.80 ± 0.35	11.97 ± 4.22	12.51 ± 2.91
UDP-N-acetylglucosamine	0.30 ± 0.06	0.62 ± 0.04	0.24 ± 0.09
Ratio	5.98 ± 1.78	19.09 ± 6.81	51.79 ± 24.38

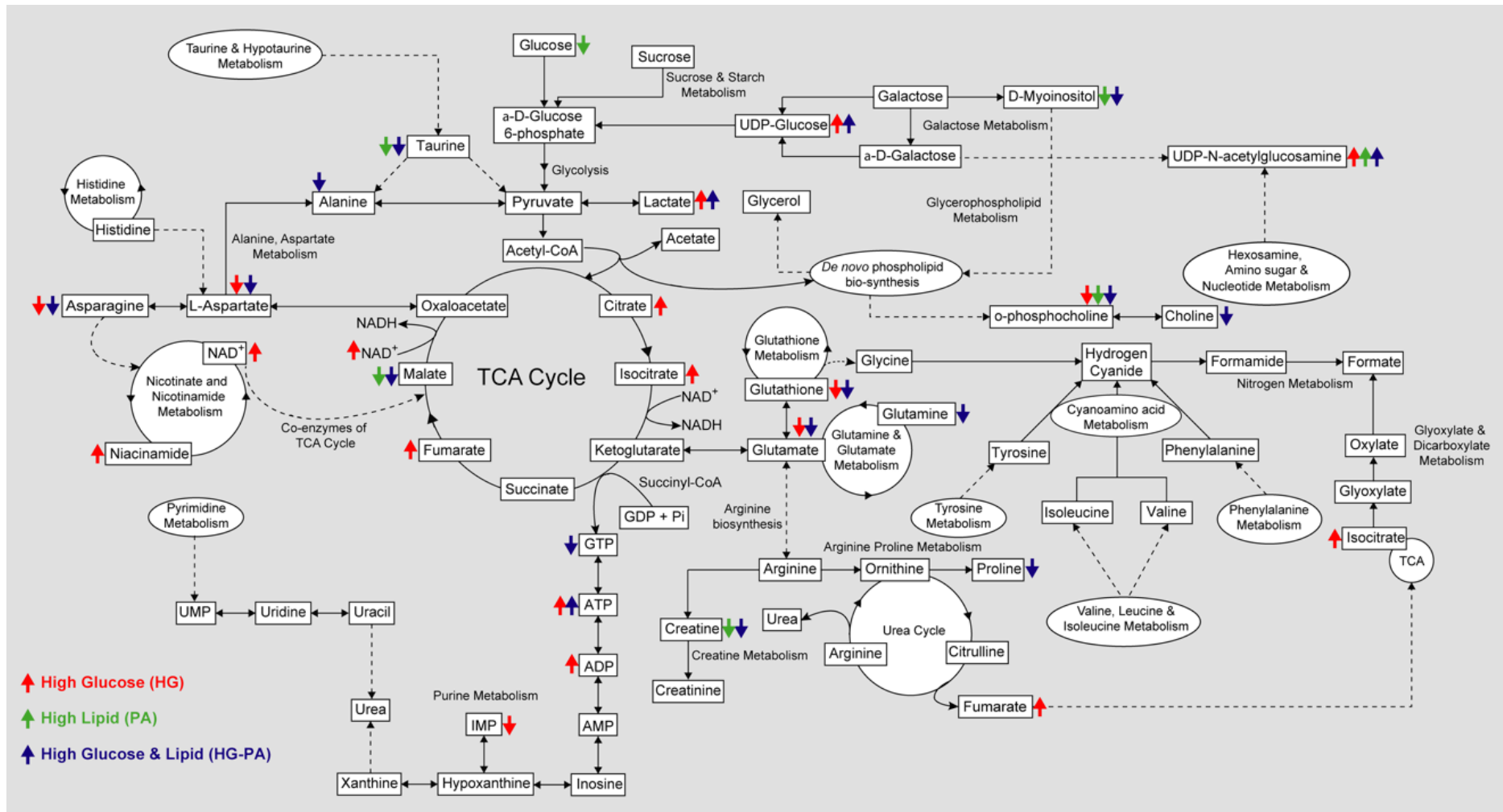


Figure 3.9: Pictorial depiction of the significantly dysregulated metabolites and Metabolomic Pathway Analysis (MetPA) construction of the metabolic pathways identified in HG, PA and HG/PA. The description of the arrows has been provided as a legend in which HG is represented by red arrows, PA is represented by green arrows and HG/PA is represented by blue arrows. The upward and downward direction of arrows indicates the upregulation and down-regulation of metabolites respectively.

The use of primary β -cells for most molecular, biochemical, and metabolomics studies has largely been limited by the availability of the human pancreatic tissue. Furthermore, the isolation and purification of β -cells in large amounts (required for such metabolomics studies) and their maintenance becomes technically challenging. In addition to this, primary β -cells do not proliferate easily in culture and quickly undergo apoptosis. The research conducted in the newly developed human pancreatic insulinomas is still in infancy and need more characterization. Therefore, the current studies were performed on *INS-IE* cells (derived from the rat INS-1 insulinoma cells) that are highly glucose responsive. Even though, the homogeneity of such clonal β -cells has been questioned lately, they have been widely used for mechanistic studies of pancreatic β -cell function. These cell cultures can be expanded in sufficient quantities for metabolomics based experiments in contrast to isolated islets which actually consist of four distinct cell types and would require pooling from multiple animals. However, there are certain drawbacks associated with using these cells that include: (a) concerns about homogeneity of clonal β -cells (Fernandez et al. 2008), (b) a low total insulin content in *INS-IE* cells (only 20% of that of the native cells) (Asfari et al. 1992), and (c) the requirement of β -mercaptoethanol (which is toxic and irritating, and irreversibly denatures the proteins) in the culture media for their propagation and maintenance of important functional characteristics (Asfari et al. 1992; Skelin, Rupnik, and Cencic 2010). Despite these drawbacks, *INS-IE* cells have been vigorously used as a model system since they are highly responsive to small changes in glucose concentrations (Merglen et al. 2004) and can be expanded in sufficient quantities for metabolomics experiments in contrast to isolated islets which actually consist of four distinct cell types and would require pooling from multiple animals. In fact, a study by Spegel et al. compared glucose metabolism in rat islets and INS-1 832/13 clonal β -cells and suggested that the metabolic processes were largely similar in the two systems despite the difference in the expression of genes involved in proliferation (Spégel et al. 2015). Indeed, most previous studies performed till data (September 2020) have reported the effect of HG (ranging from 10 mM to 25 mM) on the metabolic pathways investigated in glucose-responsive INS-1 832/13 cells, BRIN-BD11 cells and isolated pancreatic islets (Appendix Table 3.4).

3.10 Conclusions

In conclusion, our study presented the comprehensive list of the metabolic signatures associated with high-glucose (glucotoxicity), FFA (lipotoxicity), and high-glucose and FFA (glucolipotoxicity) exposure in pancreatic β -cells. Our analysis reveals that the shift in the

metabolic profile was more distinctive for glucotoxic and glucolipotoxic conditions compared to lipotoxic conditions. Cellular and biochemical data indicates that exposure of *INS-IE* cells to FFA and/or high glucose was associated with decreased cellular viability, increased ROS production, and accumulation of triglycerides and FFA in the pancreatic β -cells. The metabolites belonging to the energy metabolism and amino acid metabolism were majorly perturbed; and UDP-*N*-acetylglucosamine and *o*-phosphocholine were identified as the common dysregulated metabolites under all three nutrient excess conditions and their ratio was proposed as a biomarker for the three stress conditions tested in this study. It would also be interesting in future to carry out dose- and time- dependent studies and to examine how *de novo* lipogenesis pathways and water-insoluble metabolites are affected under these conditions in order to get the complete picture of biochemical and metabolic perturbances associated with these conditions. In addition to this, NMR spectroscopy was used as a method of choice for the current study for identification and quantitation of metabolites over mass spectrometry as these measurements are associated with very high analytical reproducibility and ease of molecular characterization in small measurement times. However, due to the inherent limitations associated with NMR spectroscopy, the effect of these stressors on the low-abundant metabolites could have been missed.

References

- Ainscow, Edward K., Chao Zhao, and Guy A. Rutter. 2000. "Acute Overexpression of Lactate Dehydrogenase-A Perturbs β -Cell Mitochondrial Metabolism and Insulin Secretion." *Diabetes*.
- Alberti, George, Paul Zimmet, Jonathan Shaw, Zachary Bloomgarden, Francine Kaufman, and Martin Silink. 2004. "Type 2 Diabetes in the Young: The Evolving Epidemic. The International Diabetes Federation Consensus Workshop." Pp. 1798–1811 in *Diabetes Care*. Vol. 27. American Diabetes Association.
- Asfari, Maryam, Danilo Janjic, Paolo Meda, Guodong Li, Philippe A. Halban, and Claes B. Wollheim. 1992. "Establishment of 2-Mercaptoethanol-Dependent Differentiated Insulin-Secreting Cell Lines." *Endocrinology* 130(1):167–78.
- Association, American Diabetes. 2009. "Diagnosis and Classification of Diabetes Mellitus." *Diabetes Care* 32(SUPPL. 1):S62.
- Basu, Ananda, Rita Basu, Pankaj Shah, Adrian Vella, C. Michael Johnson, Michael Jensen, K. Sreekumaran Nair, W. Frederick Schwenk, and Robert A. Rizza. 2001. "Type 2 Diabetes Impairs Splanchnic Uptake of Glucose but Does Not Alter Intestinal Glucose Absorption during Enteral Glucose Feeding: Additional Evidence for a Defect in Hepatic Glucokinase Activity." *Diabetes* 50(6):1351–62.
- Bhandary, Bidur, Anu Marahatta, Hyung Ryong Kim, and Han Jung Chae. 2013. "An Involvement of Oxidative Stress in Endoplasmic Reticulum Stress and Its Associated Diseases." *International Journal of Molecular Sciences* 14(1):434–56.
- Bhatia, Parnika, Shikha Raina, Jeetender Chugh, and Shilpy Sharma. 2015. "MiRNAs: Early Prognostic Biomarkers for Type 2 Diabetes Mellitus?" *Biomarkers in Medicine* 9(10):1025–40.
- Borg, R., J. C. Kuenen, B. Carstensen, H. Zheng, D. M. Nathan, R. J. Heine, J. Nerup, K. Borch-Johnsen, and D. R. Witte. 2010. "Real-Life Glycaemic Profiles in Non-Diabetic Individuals with Low Fasting Glucose and Normal HbA1c: The A1C-Derived Average Glucose (ADAG) Study." *Diabetologia*.
- Boslem, Ebru, Gemma MacIntosh, Amanda M. Preston, Clarissa Bartley, Anna K. Busch, Maria Fuller, D. Ross Laybutt, Peter J. Meikle, and Trevor J. Biden. 2011. "A Lipidomic Screen of Palmitate-Treated MIN6 β -Cells Links Sphingolipid Metabolites with Endoplasmic Reticulum (ER) Stress and Impaired Protein Trafficking." *Biochemical Journal* 435(1):267–76.
- Carlson, Olga D., Jehan D. David, Jessica M. Schrieder, Dennis C. Muller, Hyeung Jin Jang,

- Byung Joon Kim, and Josephine M. Egan. 2007. "Contribution of Nonesterified Fatty Acids to Insulin Resistance in the Elderly with Normal Fasting but Diabetic 2-Hour Postchallenge Plasma Glucose Levels: The Baltimore Longitudinal Study of Aging." *Metabolism: Clinical and Experimental* 56(10):1444–51.
- Cerf, Marlon E. 2013. "Beta Cell Dysfunction and Insulin Resistance." *Frontiers in Endocrinology*.
- Chawla, Aastha, Rajeev Chawla, and Shalini Jaggi. 2016. "Microvascular and Macrovascular Complications in Diabetes Mellitus: Distinct or Continuum?" *Indian Journal of Endocrinology and Metabolism* 20(4):546–53.
- Chong, Jasmine, Othman Soufan, Carin Li, Iurie Caraus, Shuzhao Li, Guillaume Bourque, David S. Wishart, and Jianguo Xia. 2018. "MetaboAnalyst 4.0: Towards More Transparent and Integrative Metabolomics Analysis." *Nucleic Acids Research* 46(W1):W486–94.
- Cnop, Miriam, Nils Welsh, Jean Christophe Jonas, Anne Jörns, Sigurd Lenzen, and Decio L. Eizirik. 2005. "Mechanisms of Pancreatic β -Cell Death in Type 1 and Type 2 Diabetes: Many Differences, Few Similarities." *Diabetes* 54(SUPPL. 2).
- Couté, Yohann, Yannick Brunner, Domitille Schwartz, Céline Hernandez, Alexandre Masselot, Frédérique Lisacek, Claes B. Wollheim, and Jean Charles Sanchez. 2010. "Early Activation of the Fatty Acid Metabolism Pathway by Chronic High Glucose Exposure in Rat Insulin Secretory β -Cells." *Proteomics* 10(1):59–71.
- Croze, Marine L., Alain Géoën, and Christophe O. Soulage. 2015. "Abnormalities in Myo-Inositol Metabolism Associated with Type 2 Diabetes in Mice Fed a High-Fat Diet: Benefits of a Dietary Myo-Inositol Supplementation." *British Journal of Nutrition* 113(12):1862–75.
- Cunha, Daniel A., Paul Hekerman, Laurence Ladrière, Angie Bazarra-Castro, Fernanda Ortis, Marion C. Wakeham, Fabrice Moore, Joanne Rasschaert, Alessandra K. Cardozo, Elisa Bellomo, Lutgart Overbergh, Chantal Mathieu, Roberto Lupi, Tsonwin Hai, Andre Herchuelz, Piero Marchetti, Guy A. Rutter, Décio L. Eizirik, and Miriam Cnop. 2008. "Initiation and Execution of Lipotoxic ER Stress in Pancreatic β -Cells." *Journal of Cell Science* 121(14):2308–18.
- Dasgupta, Amitava and Amer Wahed. 2014. "Carbohydrate Metabolism, Diabetes, and Hypoglycemia." in *Clinical Chemistry, Immunology and Laboratory Quality Control*.
- Deminice, Rafael, Robin P. da Silva, Simon G. Lamarre, Colin Brown, George N. Furey, Shannon A. McCarter, Alceu Afonso Jordao, Karen B. Kelly, Kirst King-Jones, René L.

- Jacobs, Margaret E. Brosnan, and John T. Brosnan. 2011. "Creatine Supplementation Prevents the Accumulation of Fat in the Livers of Rats Fed a High-Fat Diet." *Journal of Nutrition* 141(10):1799–1804.
- Doar, J. W. H., Victor Wynn, and D. G. Cramp. 1968. "Blood Pyruvate and Plasma Glucose Levels during Oral and Intravenous Glucose Tolerance Tests in Obese and Non-Obese Women." *Metabolism*.
- Dubey, Richa, Pooja Minj, Nikita Malik, Devika M. Sardesai, Shruti H. Kulkarni, Jhankar D. Acharya, Neel Sarovar Bhavesh, Shilpy Sharma, and Ashutosh Kumar. 2017. "Recombinant Human Islet Amyloid Polypeptide Forms Shorter Fibrils and Mediates β -Cell Apoptosis via Generation of Oxidative Stress." *Biochemical Journal* 474(23):3915–34.
- Eccleston, Heather B., Kelly K. Andringa, Angela M. Betancourt, Adrienne L. King, Sudheer K. Mantena, Telisha M. Swain, Heather N. Tinsley, Ryan N. Nolte, Tim R. Nagy, Gary A. Abrams, and Shannon M. Bailey. 2011. "Chronic Exposure to a High-Fat Diet Induces Hepatic Steatosis, Impairs Nitric Oxide Bioavailability, and Modifies the Mitochondrial Proteome in Mice." *Antioxidants and Redox Signaling* 15(2):447–59.
- El-Assaad, Wissal, Jean Buteau, Marie Line Peyot, Christopher Nolan, Raphael Roduit, Serge Hardy, Erik Joly, Ghassan Dbaibo, Lawrence Rosenberg, and Marc Prentki. 2003. "Saturated Fatty Acids Synergize with Elevated Glucose to Cause Pancreatic β -Cell Death." *Endocrinology* 144(9):4154–63.
- El-Azzouny, Mahmoud, Charles R. Evans, Mary K. Treutelaar, Robert T. Kennedy, and Charles F. Burant. 2014. "Increased Glucose Metabolism and Glycerolipid Formation by Fatty Acids and GPR40 Receptor Signaling Underlies the Fatty Acid Potentiation of Insulin Secretion." *Journal of Biological Chemistry* 289(19):13575–88.
- Emwas, Abdul Hamid, Raja Roy, Ryan T. McKay, Leonardo Tenori, Edoardo Saccenti, G. A. Nagana Gowda, Daniel Raftery, Fatimah Alahmari, Lukasz Jaremko, Mariusz Jaremko, and David S. Wishart. 2019. "Nmr Spectroscopy for Metabolomics Research." *Metabolites* 9(7).
- Erion, Karel A., Charles A. Berdan, Nathan E. Burritt, Barbara E. Corkey, and Jude T. Deeney. 2015. "Chronic Exposure to Excess Nutrients Left-Shifts the Concentration Dependence of Glucose-Stimulated Insulin Secretion in Pancreatic β -Cells." *Journal of Biological Chemistry* 290(26):16191–201.
- Fernandez, Céline, Ulrika Fransson, Elna Hallgard, Peter Spégel, Cecilia Holm, Morten Krogh, Kristofer Wårell, Peter James, and Hindrik Mulder. 2008. "Metabolomic and

- Proteomic Analysis of a Clonal Insulin-Producing β -Cell Line (INS-1 832/13).” *Journal of Proteome Research* 7(1):400–411.
- Flowers, E., A. M. Kanaya, Y. Fukuoka, I. E. Allen, B. Cooper, and B. E. Aouizerat. 2017. “Preliminary Evidence Supports Circulating MicroRNAs as Prognostic Biomarkers for Type 2 Diabetes.” *Obesity Science and Practice* 3(4):446–52.
- Franz, Marion J. 1997. “Protein: Metabolism and Effect on Blood Glucose Levels.” *The Diabetes Educator* 23(6):643–51.
- Goehring, I., N. S. Sauter, G. Catchpole, A. Assmann, L. Shu, K. S. Zien, M. Moehlig, A. F. H. Pfeiffer, J. Oberholzer, L. Willmitzer, J. Spranger, and K. Maedler. 2011. “Identification of an Intracellular Metabolic Signature Impairing Beta Cell Function in the Rat Beta Cell Line *INS-1E* and Human Islets.” *Diabetologia* 54(10):2584–94.
- Göhring, Isabel, Vladimir V. Sharoyko, Siri Malmgren, Lotta E. Andersson, Peter Spégel, David G. Nicholls, and Hindrik Mulder. 2014. “Chronic High Glucose and Pyruvate Levels Differentially Affect Mitochondrial Bioenergetics and Fuel-Stimulated Insulin Secretion from Clonal INS-1 832/13 Cells.” *Journal of Biological Chemistry* 289(6):3786–98.
- Gualano, Bruno, Vitor De Salles Paimelli, Hamilton Roschel, Guilherme Giannini Artioli, Manoel Neves, Ana Lúcia De Sá Pinto, Maria Elizabeth Rossi Da Silva, Maria Rosária Cunha, Maria Concepción García Otaduy, Claudia Da Costa Leite, Júlio César Ferreira, Rosa Maria Pereira, Patrícia Chakur Brum, Eloisa Bonfá, and Antonio Herbert Lancha. 2011. “Creatine in Type 2 Diabetes: A Randomized, Double-Blind, Placebo-Controlled Trial.” *Medicine and Science in Sports and Exercise* 43(5):770–78.
- Haber, C. Andrew, Tony K. T. Lam, Zhiwen Yu, Neehar Gupta, Tracy Goh, Elena Bogdanovic, Adria Giacca, and I. George Fantus. 2003. “N-Acetylcysteine and Taurine Prevent Hyperglycemia-Induced Insulin Resistance in Vivo: Possible Role of Oxidative Stress.” *American Journal of Physiology - Endocrinology and Metabolism* 285(4 48-4):E744-753.
- Hafizi Abu Bakar, Mohamad, Cheng Kian Kai, Wan Najihah Wan Hassan, Mohamad Roji Sarmidi, Harisun Yaakob, and Hasniza Zaman Huri. 2015. “Mitochondrial Dysfunction as a Central Event for Mechanisms Underlying Insulin Resistance: The Roles of Long Chain Fatty Acids.” *Diabetes/Metabolism Research and Reviews* 31(5):453–75.
- Hawkins, Meredith, Nir Barzilai, Rong Liu, Meizhu Hu, Wei Chen, and Luciano Rossetti. 1997. “Role of the Glucosamine Pathway in Fat-Induced Insulin Resistance.” *Journal of Clinical Investigation* 99(9):2173–82.

- Huang, Mei and Jamie W. Joseph. 2012. "Metabolomic Analysis of Pancreatic β -Cell Insulin Release in Response to Glucose." *Islets* 4(3):210–22.
- Ito, Takashi, Stephen W. Schaffer, and Junichi Azuma. 2012a. "The Potential Usefulness of Taurine on Diabetes Mellitus and Its Complications." *Amino Acids* 42(5):1529–39.
- Ito, Takashi, Stephen W. Schaffer, and Junichi Azuma. 2012b. "The Potential Usefulness of Taurine on Diabetes Mellitus and Its Complications." *Amino Acids* 42(5):1529–39.
- Juraschek, Stephen P., Ghanshyam Palamaner Subash Shantha, Audrey Y. Chu, Edgar R. Miller, Eliseo Guallar, Ron C. Hoogeveen, Christie M. Ballantyne, Frederick L. Brancati, Maria Inês Schmidt, James S. Pankow, and J. Hunter Young. 2013. "Lactate and Risk of Incident Diabetes in a Case-Cohort of the Atherosclerosis Risk in Communities (ARIC) Study." *PLoS ONE*.
- Kalkan, Ismail Hakki and Murat Suher. 2013. "The Relationship between the Level of Glutathione, Impairment of Glucose Metabolism and Complications of Diabetes Mellitus." *Pakistan Journal of Medical Sciences*.
- Kim, Ji Won and Kun Ho Yoon. 2011. "Glucolipotoxicity in Pancreatic β -Cells." *Diabetes and Metabolism Journal* 35(5):444–50.
- Kim, Kyoung Soo, Da Hee Oh, Jung Yeon Kim, Bong Gn Lee, Jeong Soon You, Kyung Ja Chang, Hyunju Chung, Myung Chul Yoo, Hyung In Yang, Ja Heon Kang, Yoo Chul Hwang, Kue Jeong Ahn, Ho Yeon Chung, and In Kyung Jeong. 2012a. "Taurine Ameliorates Hyperglycemia and Dyslipidemia by Reducing Insulin Resistance and Leptin Level in Otsuka Long-Evans Tokushima Fatty (OLETF) Rats with Long-Term Diabetes." *Experimental and Molecular Medicine* 44(11):665–73.
- Kim, Kyoung Soo, Da Hee Oh, Jung Yeon Kim, Bong Gn Lee, Jeong Soon You, Kyung Ja Chang, Hyunju Chung, Myung Chul Yoo, Hyung In Yang, Ja Heon Kang, Yoo Chul Hwang, Kue Jeong Ahn, Ho Yeon Chung, and In Kyung Jeong. 2012b. "Taurine Ameliorates Hyperglycemia and Dyslipidemia by Reducing Insulin Resistance and Leptin Level in Otsuka Long-Evans Tokushima Fatty (OLETF) Rats with Long-Term Diabetes." *Experimental and Molecular Medicine* 44(11):665–73.
- Kim, So Hyun, Seung Ok Yang, Hee Su Kim, Yujin Kim, Taesun Park, and Hyung Kyoon Choi. 2009. "1H-Nuclear Magnetic Resonance Spectroscopy-Based Metabolic Assessment in a Rat Model of Obesity Induced by a High-Fat Diet." *Analytical and Bioanalytical Chemistry* 395(4):1117–24.
- Liu, Xiaojing, Shuai Han, Ying Yang, Jiahong Kang, and Jiarui Wu. 2015. "Glucose-Induced Glutathione Reduction in Mitochondria Is Involved in the First Phase of Pancreatic β -

- Cell Insulin Secretion.” *Biochemical and Biophysical Research Communications*.
- Lorenz, Matthew A., Mahmoud A. El Azzouny, Robert T. Kennedy, and Charles F. Burant. 2013. “Metabolome Response to Glucose in the β -Cell Line INS-1832/13.” *Journal of Biological Chemistry* 288(15):10923–35.
- Lovejoy, J., F. D. Newby, S. S. P. Gebhart, and M. DiGirolamo. 1992. “Insulin Resistance in Obesity Is Associated with Elevated Basal Lactate Levels and Diminished Lactate Appearance Following Intravenous Glucose and Insulin.” *Metabolism*.
- Lutchmansingh, Fallon K., Jean W. Hsu, Franklyn I. Bennett, Asha V. Badaloo, McFarlane Anderson Norma, M. Gordon Strachan Georgiana, A. Wright Pascoe Rosemarie, Farook Jahoor, and Michael S. Boyne. 2018. “Glutathione Metabolism in Type 2 Diabetes and Its Relationship with Microvascular Complications and Glycemia.” *PLoS ONE*.
- De Mattia, G., O. Laurenti, C. Bravi, A. Ghiselli, L. Iuliano, and F. Balsano. 1994. “Effect of Aldose Reductase Inhibition on Glutathione Redox Status in Erythrocytes of Diabetic Patients.” *Metabolism*.
- McClain, Donald A. 2002a. “Hexosamines as Mediators of Nutrient Sensing and Regulation in Diabetes.” in *Journal of Diabetes and its Complications*.
- McClain, Donald A. 2002b. “Hexosamines as Mediators of Nutrient Sensing and Regulation in Diabetes.” Pp. 72–80 in *Journal of Diabetes and its Complications*. Vol. 16.
- Meier, Juris J. and Riccardo C. Bonadonna. 2013. “Role of Reduced β -Cell Mass versus Impaired β -Cell Function in the Pathogenesis of Type 2 Diabetes.” *Diabetes Care* 36(SUPPL.2).
- Merglen, Arnaud, Sten Theander, Blanca Rubi, Gaelle Chaffard, Claes B. Wollheim, and Pierre Maechler. 2004. “Glucose Sensitivity and Metabolism-Secretion Coupling Studied during Two-Year Continuous Culture in *INS-1E* Insulinoma Cells.” *Endocrinology* 145(2):667–78.
- El Mesallamy, Hala O., Ebtehal El-Demerdash, Lamiaa N. Hammad, and Hekmat M. El Magdoub. 2010. “Effect of Taurine Supplementation on Hyperhomocysteinemia and Markers of Oxidative Stress in High Fructose Diet Induced Insulin Resistance.” *Diabetology and Metabolic Syndrome* 2(1):46.
- Molnos, Sophie, Simone Wahl, Mark Haid, E. Marelise W. Eekhoff, René Pool, Anna Floegel, Joris Deelen, Daniela Much, Cornelia Prehn, Michaela Breier, Harmen H. Draisma, Nienke van Leeuwen, Annemarie M. C. Simonis-Bik, Anna Jonsson, Gonneke Willemsen, Wolfgang Bernigau, Rui Wang-Sattler, Karsten Suhre, Annette Peters, Barbara Thorand, Christian Herder, Wolfgang Rathmann, Michael Roden, Christian

- Gieger, Mark H. H. Kramer, Diana van Heemst, Helle K. Pedersen, Valborg Gudmundsdottir, Matthias B. Schulze, Tobias Pischon, Eco J. C. de Geus, Heiner Boeing, Dorret I. Boomsma, Anette G. Ziegler, P. Eline Slagboom, Sandra Hummel, Marian Beekman, Harald Grallert, Søren Brunak, Mark I. McCarthy, Ramneek Gupta, Ewan R. Pearson, Jerzy Adamski, and Leen M. 't Hart. 2018. "Metabolite Ratios as Potential Biomarkers for Type 2 Diabetes: A DIRECT Study." *Diabetologia* 61(1):117–29.
- Mugabo, Yves, Shangang Zhao, Julien Lamontagne, Anfal Al-Mass, Marie Line Peyot, Barbara E. Corkey, Erik Joly, S. R. Murth. Madiraju, and Marc Prentki. 2017. "Metabolic Fate of Glucose and Candidate Signaling and Excess-Fuel Detoxification Pathways in Pancreatic β -Cells." *Journal of Biological Chemistry* 292(18):7407–22.
- Nakamura, H., H. Jinzu, K. Nagao, Y. Noguchi, N. Shimba, H. Miyano, T. Watanabe, and K. Iseki. 2014. "Plasma Amino Acid Profiles Are Associated with Insulin, C-Peptide and Adiponectin Levels in Type 2 Diabetic Patients." *Nutrition and Diabetes* 4(9):e133.
- Newsholme, Philip, Lorraine Brennan, and Katrin Bender. 2006. "Amino Acid Metabolism, β -Cell Function, and Diabetes." *Diabetes* 55(SUPPL. 2):S39–47.
- Newsholme, Philip, Lorraine Brennan, Blanca Rubi, and Pierre Maechler. 2005. "New Insights into Amino Acid Metabolism, β -Cell Function and Diabetes." *Clinical Science*.
- Ni, Yan, Linjing Zhao, Haoyong Yu, Xiaojing Ma, Yuqian Bao, Cynthia Rajani, Lenora W. M. Loo, Yurii B. Shvetsov, Herbert Yu, Tianlu Chen, Yinan Zhang, Congrong Wang, Cheng Hu, Mingming Su, Guoxiang Xie, Aihua Zhao, Wei Jia, and Weiping Jia. 2015. "Circulating Unsaturated Fatty Acids Delineate the Metabolic Status of Obese Individuals." *EBioMedicine* 2(10):1513–22.
- Nyblom, Hanna K., L. I. Nord, R. Andersson, L. Kenne, and P. Bergsten. 2008. "Glucose-Induced de Novo Synthesis of Fatty Acyls Causes Proportional Increases in *INS-1E* Cellular Lipids." *NMR in Biomedicine* 21(4):357–65.
- Op 't Eijnde, B., B. Ursø, E. A. Richter, P. L. Greenhaff, and P. Hespel. 2001. "Effect of Oral Creatine Supplementation on Human Muscle GLUT4 Protein Content after Immobilization." *Diabetes* 50(1):18–23.
- Ortmeyer, Heidi K. 1996. "Dietary Myoinositol Results in Lower Urine Glucose and in Lower Postprandial Plasma Glucose in Obese Insulin Resistant Rhesus Monkeys." *Obesity Research* 4(6):569–75.
- Pintaudi, Basilio, Giacomina Di Vieste, and Matteo Bonomo. 2016. "The Effectiveness of Myo-Inositol and D-Chiro Inositol Treatment in Type 2 Diabetes." *International*

- Journal of Endocrinology* 2016:9132052.
- Poitout, Vincent and R. Paul Robertson. 2008. "Glucolipototoxicity: Fuel Excess and β -Cell Dysfunction." *Endocrine Reviews* 29(3):351–66.
- Pouwels, Marie Jose J., Cees J. Tack, Paul N. Span, André J. Olthaar, C. G. J. Sweep, Frank C. Huvers, Jos A. Lutterman, and Ad R. M. M. Hermus. 2004. "Role of Hexosamines in Insulin Resistance and Nutrient Sensing in Human Adipose and Muscle Tissue." *Journal of Clinical Endocrinology and Metabolism* 89(10):5132–37.
- Ripps, Harris and Wen Shen. 2012. "Review: Taurine: A 'Very Essential' Amino Acid." *Molecular Vision* 18:2673–86.
- Robinson, Katherine A., Mitchell L. Weinstein, George E. Lindenmayer, and Maria G. Buse. 1995. "Effects of Diabetes and Hyperglycemia on the Hexosamine Synthesis Pathway in Rat Muscle and Liver." *Diabetes* 44(12):1438–46.
- Ročić, Boris, Ariana Znaor, Petra Ročić, David Weber, and Marijana Vučić Lovrenčić. 2011. "Comparison of Antihyperglycemic Effects of Creatine and Glibenclamide in Type II Diabetic Patients." *Wiener Medizinische Wochenschrift*.
- Saltiel, Alan R. and C. Ronald Kahn. 2001. "Insulin Signalling and the Regulation of Glucose and Lipid Metabolism." *Nature* 414(6865):799–806.
- da Silva, Robin P., Kelly Ann Leonard, and René L. Jacobs. 2017. "Dietary Creatine Supplementation Lowers Hepatic Triacylglycerol by Increasing Lipoprotein Secretion in Rats Fed High-Fat Diet." *Journal of Nutritional Biochemistry* 50:46–53.
- Skelin, Masa, Marjan Rupnik, and Avrelija Cencic. 2010. "Pancreatic Beta Cell Lines and Their Applications in Diabetes Mellitus Research." *ALTEX* 27(2):105–13.
- Slawson, C., R. J. Copeland, and G. W. Hart. 2010. "O-GlcNAc Signaling: A Metabolic Link between Diabetes and Cancer?" *Trends in Biochemical Sciences*.
- Somesh, Baggavalli P., Mahesh Kumar Verma, Manoj Kumar Sadasivuni, Anup Mammen-Oommen, Sanghamitra Biswas, Pavagada C. Shilpa, Ashok Kumar Reddy, Aggunda N. Yateesh, Puttrevana M. Pallavi, Siddaraju Nethra, Rachapalli Smitha, Korrapati Neelima, Usha Narayanan, and Madanahalli R. Jagannath. 2013. "Chronic Glucolipotoxic Conditions in Pancreatic Islets Impair Insulin Secretion Due to Dysregulated Calcium Dynamics, Glucose Responsiveness and Mitochondrial Activity." *BMC Cell Biology* 14(1):31.
- Sönksen, Peter H. 1984. "Diabetes Mellitus." Pp. 507–31 in *Endocrine Disorders*. Elsevier.
- Spéigel, Peter, Lotta E. Andersson, Petter Storm, Vladimir Sharoyko, Isabel Göhring, Anders H. Rosengren, and Hindrik Mulder. 2015. "Unique and Shared Metabolic Regulation in

- Clonal β -Cells and Primary Islets Derived from Rat Revealed by Metabolomics Analysis.” *Endocrinology* 156(6):1995–2005.
- Spégel, Peter, Siri Malmgren, Vladimir V. Sharoyko, Philip Newsholme, Thomas Koeck, and Hindrik Mulder. 2011. “Metabolomic Analyses Reveal Profound Differences in Glycolytic and Tricarboxylic Acid Cycle Metabolism in Glucose-Responsive and - Unresponsive Clonal β -Cell Lines.” *Biochemical Journal* 435(1):277–84.
- Spégel, Peter, Vladimir V. Sharoyko, Isabel Goehring, Anders P. H. Danielsson, Siri Malmgren, Cecilia L. F. Nagorny, Lotta E. Andersson, Thomas Koeck, Geoffrey W. G. Sharp, Susanne G. Straub, Claes B. Wollheim, and Hindrik Mulder. 2013. “Time- Resolved Metabolomics Analysis of β -Cells Implicates the Pentose Phosphate Pathway in the Control of Insulin Release.” *Biochemical Journal* 450(3):595–605.
- Spiro, M. J. 1984. “Effect of Diabetes on the Sugar Nucleotides in Several Tissues of the Rat.” *Diabetologia*.
- Stefani, Giuseppe P., Ramiro B. Nunes, André Z. Dornelles, Jadson P. Alves, Marcella O. Piva, Marlise D. Domenico, Cláudia R. Rhoden, and Pedro D. Lago. 2014. “Effects of Creatine Supplementation Associated with Resistance Training on Oxidative Stress in Different Tissues of Rats.” *Journal of the International Society of Sports Nutrition*.
- Tam, Zhi Yang, Sean Pin Ng, Ling Qiao Tan, Chih Hsien Lin, Dietrich Rothenbacher, Jochen Klenk, Bernhard Otto Boehm, Kelvin Goh Kau Kiat, Pipob Suwanchaikasem, Pornpimol Tiphthara, Song Yi Yang, T. Becker, J. Stingl, W. Koenig, M. Riepe, R. Peter, H. Geiger, A. Ludolph, C. V. Arnim, G. Nagel, G. Weinmayr, K. Rapp, M. D. Denking, D. Dallmeier, J. M. Steinacker, and R. Laszlo. 2017. “Metabolite Profiling in Identifying Metabolic Biomarkers in Older People with Late-Onset Type 2 Diabetes Mellitus.” *Scientific Reports* 7(1):4392.
- Testa, Roberto, Valerie Vanhooren, Anna Rita Bonfigli, Massimo Boemi, Fabiola Olivieri, Antonio Ceriello, Stefano Genovese, Liana Spazzafumo, Vincenzo Borelli, Maria Giulia Bacalini, Stefano Salvioli, Paolo Garagnani, Sylviane Dewaele, Claude Libert, and Claudio Franceschi. 2015. “N-Glycomic Changes in Serum Proteins in Type 2 Diabetes Mellitus Correlate with Complications and with Metabolic Syndrome Parameters.” *PLoS ONE* 10(3):e0119983.
- Trachtman, H., S. Futterweit, J. Maesaka, C. Ma, E. Valderrama, A. Fuchs, A. A. Tarectecan, P. S. Rao, J. A. Sturman, T. H. Boles, M. X. Fu, and J. Baynes. 1995. “Taurine Ameliorates Chronic Streptozocin-Induced Diabetic Nephropathy in Rats.” *American Journal of Physiology - Renal Fluid and Electrolyte Physiology* 269(3 38-3):F429-438.

- Ulrich, Eldon L., Hideo Akutsu, Jurgen F. Doreleijers, Yoko Harano, Yannis E. Ioannidis, Jundong Lin, Miron Livny, Steve Mading, Dimitri Maziuk, Zachary Miller, Eiichi Nakatani, Christopher F. Schulte, David E. Tolmie, R. Kent Wenger, Hongyang Yao, and John L. Markley. 2008. "BioMagResBank." *Nucleic Acids Research* 36(SUPPL. 1):D402-408.
- Wallace, Martina, Helena Whelan, and Lorraine Brennan. 2013. "Metabolomic Analysis of Pancreatic Beta Cells Following Exposure to High Glucose." *Biochimica et Biophysica Acta - General Subjects* 1830(3):2583–90.
- Wanichthanarak, Kwanjeera, Johannes F. Fahrman, and Dmitry Grapov. 2015. "Genomic, Proteomic, and Metabolomic Data Integration Strategies." *Biomarker Insights* 10s4(Suppl 4):BMI.S29511.
- Ward, Meliza G., Ge Li, Valéria C. Barbosa-Lorenzi, and Mingming Hao. 2017. "Stigmasterol Prevents Glucolipototoxicity Induced Defects in Glucose-Stimulated Insulin Secretion." *Scientific Reports* 7(1):9536.
- Williams, R. E., E. M. Lenz, M. Rantalainen, and I. D. Wilson. 2006. "The Comparative Metabonomics of Age-Related Changes in the Urinary Composition of Male Wistar-Derived and Zucker (Fa/Fa) Obese Rats." *Molecular BioSystems* 2(3–4):193–202.
- Wishart, David S., Dan Tzur, Craig Knox, Roman Eisner, An Chi Guo, Nelson Young, Dean Cheng, Kevin Jewell, David Arndt, Summit Sawhney, Chris Fung, Lisa Nikolai, Mike Lewis, Marie Aude Coutouly, Ian Forsythe, Peter Tang, Savita Shrivastava, Kevin Jeroncic, Paul Stothard, Godwin Amegbey, David Block, David D. Hau, James Wagner, Jessica Miniaci, Melisa Clements, Mulu Gebremedhin, Natalie Guo, Ying Zhang, Gavin E. Duggan, Glen D. MacInnis, Alim M. Weljie, Reza Dowlatabadi, Fiona Bamforth, Derrick Clive, Russ Greiner, Liang Li, Tom Marrie, Brian D. Sykes, Hans J. Vogel, and Lori Querengesser. 2007. "HMDB: The Human Metabolome Database." *Nucleic Acids Research* 35(SUPPL. 1):D521-526.
- Xia, Jianguo, Trent C. Bjorndahl, Peter Tang, and David S. Wishart. 2008. "MetaboMiner - Semi-Automated Identification of Metabolites from 2D NMR Spectra of Complex Biofluids." *BMC Bioinformatics* 9:507–507.
- Xia, Jianguo, David S. Wishart, and Alfonso Valencia. 2011. "MetPA: A Web-Based Metabolomics Tool for Pathway Analysis and Visualization." Pp. 2342–44 in *Bioinformatics*. Vol. 27. Oxford University Press.
- Yamagishi, Sho-ichi and Takanori Matsui. 2010. "Advanced Glycation End Products, Oxidative Stress and Diabetic Nephropathy." *Oxidative Medicine and Cellular*

Longevity 3(2):101.

- Yazıcı, Dilek and Havva Sezer. 2017. "Insulin Resistance, Obesity and Lipotoxicity." Pp. 277–304 in *Advances in Experimental Medicine and Biology*. Vol. 960. Springer New York LLC.
- Yokoi, Norihide, Ghupurjan Gheni, Harumi Takahashi, and Susumu Seino. 2016. "β-Cell Glutamate Signaling: Its Role in Incretin-Induced Insulin Secretion." *Journal of Diabetes Investigation* 7(Suppl 1):38–43.
- Yousf, Saleem, Devika M. Sardesai, Abraham B. Mathew, Rashi Khandelwal, Jhankar D. Acharya, Shilpy Sharma, and Jeetender Chugh. 2019. "Metabolic Signatures Suggest O-Phosphocholine to UDP-N-Acetylglucosamine Ratio as a Potential Biomarker for High-Glucose and/or Palmitate Exposure in Pancreatic β-Cells." *Metabolomics* 15(4):55.
- Yuzefovych, Larysa, Glenn Wilson, and Lyudmila Rachek. 2010. "Different Effects of Oleate vs. Palmitate on Mitochondrial Function, Apoptosis, and Insulin Signaling in L6 Skeletal Muscle Cells: Role of Oxidative Stress." *American Journal of Physiology - Endocrinology and Metabolism* 299(6):1096–1105.
- Zhang, Tingting and Changhong Li. 2013. "Mechanisms of Amino Acid-Stimulated Insulin Secretion in Congenital Hyperinsulinism." *Acta Biochimica et Biophysica Sinica*.
- Zheng, Yan and Frank B. Hu. 2015. "Comprehensive Metabolomic Profiling of Type 2 Diabetes." *Clinical Chemistry* 61(3):453–55.
- Zimmet, Paul, K. G. M. M. Alberti, and Jonathan Shaw. 2001. "Global and Societal Implications of the Diabetes Epidemic." *Nature* 414(6865):782–87.
- Zois, Christos E. and Adrian L. Harris. 2016. "Glycogen Metabolism Has a Key Role in the Cancer Microenvironment and Provides New Targets for Cancer Therapy." *Journal of Molecular Medicine*.

Chapter 4

Identification of potential serum biomarkers linked with Type 2 Diabetes in an Asian Indian population using NMR-based metabolomics

4.1 Introduction

Type 2 diabetes mellitus (T2DM) is a progressive metabolic disorder that has been characterized by chronic hyperglycemia that occurs due to insulin insufficiency from pancreatic β -cells and has been associated with the development of insulin resistance (IR) in the insulin-target tissues (Zaccardi et al. 2016)(DeFronzo 2009). The prevalence of prediabetes and T2DM has increasing globally, with worrying statistics being reported from the children, adolescents, and young adults worldwide (Rughani, Friedman, and Tryggstad 2020) and in developing countries like India (Shetty 2012; Tabish 2007). As per the International Diabetes Federation, ~77 million people, which accounts for one in six people, or 17% of the world diabetic population, belongs to India (<https://www.diabetesatlas.org/en/>). The increased prevalence and early onset have been linked with a lack of physical activity and a sedentary lifestyle, dietary changes, obesity, and deficiency of micronutrients, including Vitamin B12 in an average Indian. In addition to this, the increased susceptibility to diabetes at a younger age and lesser BMI in Asian Indians, when compared to the Western equals, has been linked to higher body fat and visceral fat; lower skeletal muscle mass; and higher prevalence of IR (Chawla et al. 2020; Ramachandran et al. 1997; Ramachandran, Wan Ma, and Snehalatha 2010). The traditional biomarkers, including fasting and postprandial plasma glucose, glycated hemoglobin (Hb_{A1C}), C-peptide levels, etc. cannot predict early onset of T2DM and have proven to be inefficient, with very little predictive power, for identifying individuals at risk of developing T2DM, with prediabetes and with insulin resistance (Bhatia et al. 2015; Chawla et al. 2020). This raises the need for the development of novel biomarkers of T2DM that would not only help in early diagnosis but will also aid in identifying individuals at risk.

Metabolomics – an emerging high-throughput technique – has the potential of identification and characterization of low molecular weight biochemicals (referred to as metabolites) in biological samples, including tissues, body fluids, etc. (Gowda et al. 2008; Zhang et al. 2015). Since these metabolites are the by-products of the major metabolic pathways, their analysis performed in different physiological conditions have been used to provide a functional readout of a phenotype. Along these lines, several metabolic studies have been conducted that have correlated metabolite changes with the development of insulin resistance, prediabetes, and T2DM (reviewed extensively in (Arneth, Arneth, and Shams 2019; Guasch-Ferré et al. 2016; Klein and Shearer 2016; Satheesh, Ramachandran, and Jaleel 2020a; Zhao et al. 2016)) and have been useful in gaining mechanistic insights into their pathophysiology. Most of these studies have reported an association of branched-chain amino

acids (leucine, isoleucine, and valine) (Del Coco et al. 2019; Qiu et al. 2016; Wang et al. 2011), aromatic amino acids (phenylalanine and tyrosine) (Floegel et al. 2013; Qiu et al. 2016; Wang et al. 2011), acylcarnitines (Mihalik et al. 2010; Sun et al. 2016), and phospholipids (Ahola-Olli et al. 2019; Pérez-Matos et al. 2017), to name a few, with an increased incidence of prediabetes and established T2DM.

Metabolomics studies investigating biomarkers for T2DM from the South-Asian population, especially from India, are few (Gogna et al. 2015; Tai et al. 2010), with most being conducted in European, German, UK, USA, Dutch, Chinese, and Middle Eastern populations (Fikri et al. 2020; Satheesh, Ramachandran, and Jaleel 2020b). A recent study by Gogna et al. investigated the serum from diabetic, obese South Indian Asian subjects using NMR-based metabolomics. This study proposed a set of 19 significantly altered metabolites, including saturated fatty acids, lactate, valine, isoleucine, and phenylalanine, that could be used as potential biomarkers for diabetes onset in subjects with high BMI (Gogna et al. 2015). Another study performed in Asian-Indian and Chinese individuals from Singapore identified alanine, proline, valine, leucine/isoleucine, phenylalanine, tyrosine, glutamate/glutamine, and ornithine, and branched-chain amino acids to be correlated with insulin resistance (Tai et al. 2010). However, none of the studies performed to date (October 2020) aimed to identify markers associated with prediabetes, which could be used for disease prognosis in the Indian population. Hence, more studies are warranted in this direction from the Indian sub-continent.

With this background, in the current study, we aimed to identify unique metabolic markers associated with prediabetes and established T2DM in Asian Indians using NMR-based metabolomics that could be used as potential biomarkers for prognosis and disease diagnosis. Our study identified 36 aqueous metabolites from the methanolic extracts of serum samples, of which 24 metabolites showed a statistically significant difference between normal individuals (referred to as healthy controls), prediabetic individuals, and subjects with established T2DM. Using ROC curve analysis, 12 metabolites in the T2DM subjects (including glucose, pyroglutamate, serine, proline, glutamate, methionine, isoleucine, alanine, citrate, betaine, glycerol, and o-phosphocholine); and 6 metabolites (including glucose, pyroglutamate, o-phosphocholine, serine, snglycero-3-phosphocholine, and methionine) in prediabetic subjects, were identified high specificity and sensitivity ($AUC > 0.7$). On performing multivariate ROC curve analysis with the panel of selected 5 metabolites common between prediabetes and T2DM samples, AUC values obtained were 0.96 (95% confidence interval (CI) = 0.93, 0.98) for established T2DM; and 0.88 (95% CI = 0.81, 0.93) for prediabetes subjects. Hence, we propose that this panel of 5 metabolic biomarkers (namely, glucose, pyroglutamate,

o-phosphocholine, serine, and methionine) can be used in the future for clinical diagnosis, patient surveillance, and for predicting individuals at risk for developing overt diabetes in the future in the South Asian Indians.

4.2 Experimental Procedures

4.2.1 Study Population, Anthropometric parameters, and biochemical tests

A total of 284 individuals (Asian Indians) were recruited for the study from the Out-Patient Department of Armed forces medical college, Khirkee, Pune by participating physicians and paramedical staff. All enrolled subjects provided their informed consent at the time of recruitment. The study protocols were approved by the institutional ethics committees of the participating institutions, centers, and hospitals. After obtaining consent, a comprehensive case history questionnaire was explained to the participants in English or a local language (Marathi/Hindi) in which the subjects felt comfortable. A copy of questionnaire, consent form and approvals has been provided in Appendix 2. Data was recorded for each participant, including information on the age, sex, education, employment details, and occupation; history on the use of alcohol, tobacco chewing, and smoking was also recorded. In addition to this, clinical history with information on other co-morbid medical conditions and surgeries; medication history; family history about T2DM, metabolic syndrome, hypertension (wherever available); physical parameters, such as weight, height, blood pressure, waist-hip measurements, body mass index (BMI); diet (lacto-, ovo, non-vegetarian), and physical activity and exercise regimen followed; were also recorded. Subjects with a history of malignancy or other terminal illness, positive for HIV and/or HBsAg, poor compliance (from whatever cause), with insufficient blood samples, and unable to provide informed consent were excluded from the study.

Subjects between the age of 30 to 75 years were recruited following the guidelines of the American diabetes association in three groups: a) Recently diagnosed prediabetic subjects (fasting blood glucose levels between 100-125 mg/dl and on low doses of Metformin); b) Established diabetic (subjects with established T2D and undergoing treatment with Metformin for at least five years, fasting blood glucose levels \geq 126 mg/dl); and c) age and sex-matched controls (fasting blood glucose levels \leq 100 mg/dl; no history of T2D and metabolic disorders in the parents) with no history of diabetes, cardiovascular disease and/or metabolic syndrome. Blood samples were collected for routine lab assessment for the test, including fasting blood

glucose, postprandial blood glucose, Hb_{A1c}, fasting lipid profile (triglycerides, HDL, LDL, and total cholesterol). Serum vitamin B12 levels were estimated using ELISA (Calbiochem).

4.2.2 NMR sample preparation

Serum samples were separated from blood collected in serum separator vacutainers (BD Bioscience) incubated at room temperature for 30 minutes to allow clot formation. The tubes were spun at 2000 rpm for 5 mins, and the separated serum was collected in fresh 1.5 ml tubes and stored in aliquots at -80 °C until further use. For the preparation of metabolic extracts, an aliquot containing 300 µl serum was thawed on ice, mixed with two volumes of ice-cold methanol, vortexed briefly, and incubated at -20°C for 30 min. This was followed by centrifugation at 13,000 g for 30 min at 4°C. The supernatant, thus obtained, was flash-frozen into liquid nitrogen and lyophilized to remove residual water and methanol. The final extracts were stored at -80°C until NMR acquisition.

For the NMR sample preparation, the lyophilized extracts were allowed to thaw on ice, followed by reconstitution into 550 µl 100% NMR buffer (20 mM sodium phosphate, pH 7.4 in D₂O containing 0.4 mM DSS (2,2-dimethyl-2-silapentane-5-sulfonic acid). The samples were vortexed for 2 min at room temperature and centrifuged at 4000 g for 2 min. The supernatants were transferred to 5 mm NMR tubes for NMR measurements. NMR measurements were recorded on consecutive days with between 15-20 samples (coded) randomly selected samples recorded per day.

4.3 NMR spectroscopy and Spectral processing

All the 1D ¹H NMR spectra were recorded at 298 K using water suppression pulse sequence noesygppr1d from Bruker library as described in detail in Chapter 2. Typical acquisition parameters included a 5 s relaxation delay, 64 transients, 7200 Hz spectral width, 32-K data points. The ¹H NMR spectra of all the serum samples were manually phased, baseline-corrected using Bruker's NMR data processing software Topspin (v3.5) (www.bruker.com/bruker/topspin). All the ¹H chemical shifts were directly referenced with respect to methyl singlet of DSS internal reference, set to a chemical shift (δ) of 0.00 ppm.

4.4 Metabolite Identification and quantitation

After spectral processing, identification of metabolites was carried out with the Chenomx NMR Suite 8.1 (Edmonton, AB, Canada) software and were further confirmed with

biological magnetic resonance data bank (BMRB) (Ulrich et al., 2008) database and human metabolome database (HMDB) (Wishart et al., 2007). All the identified metabolites were then quantified using the profiler module of Chemomx software, which enables metabolites quantification relative to an internal standard of known concentration (400 μ M). The concentration data obtained after metabolite quantification were converted to comma-separated values (CSV) format using Microsoft excel format and imported into MetaboAnalyst 4.0 (Chong et al., 2018), a free web-based program for multivariate analysis.

4.5 Statistical Analysis

All clinical parameters have been represented as Mean \pm SEM unless stated otherwise, where SEM refers to the standard error mean and is expressed as a sample deviation divided by the square root of the sample size. The comparison amongst the groups was performed using one-way ANOVA; Tukey HSD post-doc test. A p-value $<$ 0.05 was considered to be statistically significant when comparing groups.

Multivariate statistical analysis was performed using the Metaboanalyst web tool (www.metabolanalyst.ca). For predicting the metabolite variations among different groups, chemometric analysis – Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA) was carried out. The raw metabolomics data sets were subjected to Pareto-scaling prior to chemometric analysis. The quality of the OPLS-DA models was assessed by R^2 , which defines the total explained variance (indicating goodness of fit) and Q^2 values, indicating the predictability of the model. The permutation statistic was further used with 100 permutations to validate the OPLS-DA models. One-way analysis of variance (ANOVA) followed by post-hoc analyses using Fisher's LSD was performed to assess the metabolites that showed significant variation (p-value $<$ 0.05) in diabetic and prediabetic subjects when compared to control subjects. Further box and whisker plots were generated to visualize the comparative variation in the levels of discriminating metabolites identified in serum profiles of diabetic, prediabetic, and healthy patients.

Significance Analysis for Microarrays (SAM) method was employed to identify the most discriminant and important metabolites responsible for the separation between control and diabetic groups. SAM is a permutation-based (non-parametric) hypothesis testing method that uses a moderated t-test, denoted as d_j , to measure the change in metabolite expression between the two groups. SAM also provides an estimation of the FDR, and metabolites with SAM-FDR q-value $<$ 0.05 were considered statistically significant.

4.6 Receiver operating characteristic (ROC) curve

To identify potential biomarkers for early detection or progression of T2DM, we applied classical univariate ROC curve analysis to evaluate the area under the curve (AUC) and their 95% confidence intervals. ROC curves were generated using the “Biomarker Analysis” module of the MetaboAnalyst 4.0. Multivariate ROC curve analysis was carried out to select a panel of biomarkers (multiple-biomarker test) and to evaluate the classification performance of the generated models. Balanced sub-sampling-based Monte Carlo cross-validation (MCCV) was applied to generate the ROC curves for different models using different metabolite panels. The Linear Support Vector Machine (SVM) for the classification and "Random Forest algorithm" for the feature ranking method was used for the analysis.

4.7 Metabolic pathway analysis

Metabolic Pathway Analysis (MetPa) was carried out on a free web-based program Metaboanalyst 4.0 (Chong et al., 2018; Xia, Wishart, & Valencia, 2011). The most dysregulated metabolic pathways were identified based on the statistical value (p-value) of the pathway enrichment analysis and pathway impact value calculated from pathway topological analysis (Xia et al., 2011). The metabolic pathways with impact value > 0.1 and $p < 0.05$ were considered as significantly perturbed pathways.

4.8 Results

4.8.1 Clinical characteristics of the study population

The subjects recruited for the study were clinically characterized as healthy volunteers, prediabetic subjects, and established T2DM following the recommendations of the American diabetes association and WHO guidelines. The fasting and postprandial plasma glucose levels for the healthy controls (N = 101) were 99 ± 16 mg/ml and 120 ± 40 mg/dl, respectively; for prediabetics were (N = 75) 121 ± 17 mg/dl and 157 ± 27 mg/dl, respectively; and for established T2DM subjects (N = 108) were 178 ± 73 mg/dl and 257 ± 82 mg/dl, respectively. The HbA1c levels were significantly higher in the T2DM subjects (8 ± 1 %) when compared to the healthy controls (5 ± 0.5 %) (Tukey HSD post-doc test $p = 0.009$). The other clinical characteristics of the study population have been listed in Table 4.1.

Table 4.1: Clinical characteristics of the study population

	Healthy individuals	Prediabetic subjects	Established type 2 diabetic subjects
Total Number	101	75	108
Age (in years)	50 ± 16	55 ± 12	56 ± 10
Sex (M:F)	82:19	34:41	60:48
BMI (kg/m ²)	24.8 ± 3.1	26.2 ± 4.5	26.3 ± 3.85
Fasting glucose (mg/dL)	99 ± 16	121 ± 17	178 ± 73
Post-prandial glucose (mg/dL)	120 ± 40	157 ± 27	257 ± 82
HbA1c (%)	5.0 ± 0.5	6.21 ± 0.4	8.0 ± 1.0
Triglyceride levels (mg/dL)	116 ± 69	123 ± 62	184 ± 133
HDL (mg/dL)	47 ± 14	54 ± 14	48 ± 13
Cholesterol (mg/dL)	168 ± 41	186 ± 46	190 ± 57
Vitamin B12 (picogram/mL)	425 ± 476	396 ± 317	408 ± 346

4.8.2 Resonance Assignment

Global metabolic profiling of serum samples of healthy controls was performed and was compared with the metabolic profiles of serum samples from subjects with prediabetes and established T2DM. A total of 36 metabolites, including our internal reference, were identified in the serum samples of different cohorts using standard one-dimensional ¹H-NMR spectra. A single peak at δ 3.71 ppm was left unassigned. All the metabolites were identified using the profiler module of Chenomx NMR Suite 8.1 (Edmonton, AB, Canada) software. The resonance assignment for low-density lipids (LDL) was done using previously reported NMR assignments from the literature (Mallol et al. 2013; Rawat et al. 2016). The NMR spectra present signals mainly from amino acids, sugars, carbohydrates, lipids, and membrane metabolites. All the 36 metabolites and their respective ¹H chemical shifts (in reference to DSS) have been listed in Appendix Table 4.2.

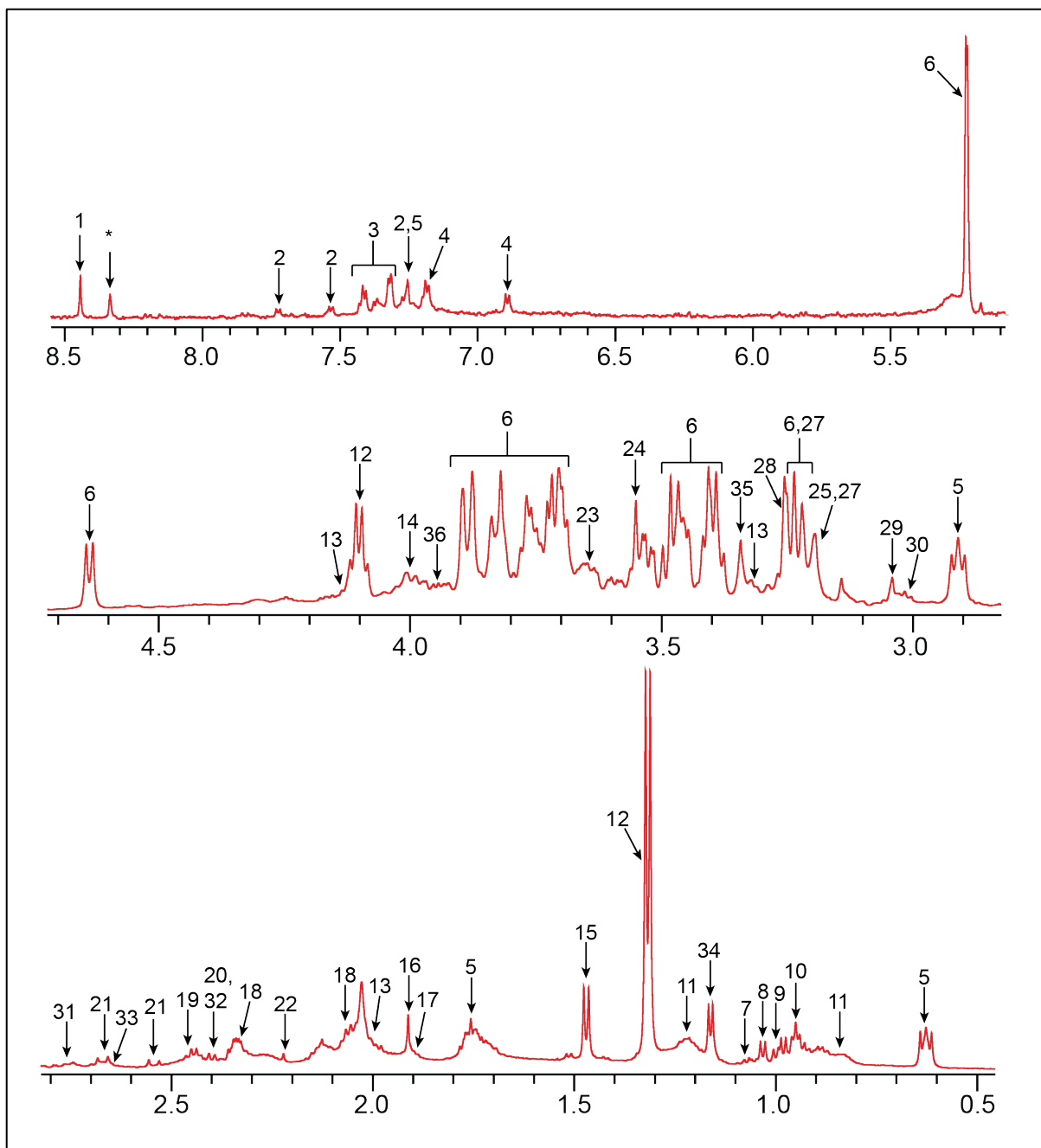


Figure 4.1: Representative ^1H NMR spectrum of serum obtained from healthy and diabetic patients. 1, Formate; 2, Tryptophan; 3, Phenylalanine; 4, Tyrosine; 5, DSS; 6, Glucose; 7, 3-hydroxyisobutyrate; 8, Valine; 9, Isoleucine; 10, Leucine; 11, LDL; 12, Lactate; 13, Proline; 14, Fructose; 15, Alanine; 16, Acetate; 17, Arginine; 18, Glutamate; 19, Glutamine; 20, Succinate; 21, Citrate; 22, Acetone; 23, Glycerol; 24, Glycine; 25, Choline; 26, o-phosphocholine; 27, sn-glycero-3-phosphocholine; 28, Betaine; 29, Creatinine; 30, Lysine; 31, Aspartate; 32, Pyroglutamate; 33, Methionine; 34, Isopropanol; 35, Methanol; 36, Serine.

4.8.3 Classification and Feature selection

All the identified metabolites were then quantified in the profile module of the Chenomx software. Six metabolites, including succinate, methanol, acetone, isopropanol, fructose, and LDL, were excluded from the analysis as their representative peaks were not evident in all spectra and may, thus, be vulnerable to over- or under-estimation of concentrations. The concentration data of all the metabolites were used for feature selection and group discrimination using chemometric analysis. Univariate analysis using one-way ANOVA identified 24 statistically significant metabolites (FDR corrected p -value <0.05) among healthy controls, prediabetic individuals, and subjects with T2DM. The details of these metabolites, along with their p and FDR values, have been listed in Appendix Table 4.3. Box-Whisker plots for these metabolites have been depicted in Figure 4.2 for a visual interpretation of dysregulated metabolites in T2DM and prediabetic subjects. Further, SAM plots were used to identify discriminatory metabolites in prediabetic and diabetic subjects compared to healthy controls (Figure 4.3). To reduce FDR and false positives, the $d(i)$ values were set to 2 for the T2DM group and 1.3 for prediabetic group, respectively. Overall, 23 and 9 metabolites were found to be significantly different (SAM-FDR q -value < 0.05) in T2DM (Figure 4.3A) and prediabetic subjects (Figure 4.3B), respectively, when compared to healthy controls. The details of the $d(i)$ values of these metabolites, along with their p and FDR values, have been listed in Appendix Table 4.4a for the T2DM group and Appendix Table 4.4b for prediabetic group.

The significantly up-regulated and down-regulated metabolites have been represented by green and red circles, respectively, in (Figure 4.3). The significantly up-regulated metabolites observed in T2DM included glucose, glutamate, proline, methionine, isoleucine, citrate, alanine, 3-hydroxyisobutyrate, leucine, valine, lysine, glycerol, and tryptophan. On the other hand, serine, pyroglutamate, betaine, o-phosphocholine, formate, choline, glycine, sn-glycero-3-phosphocholine, aspartate, and arginine were found to be significantly down-regulated in T2DM subjects. Similarly, in prediabetic patients, glucose was the only up-regulated metabolite, while pyroglutamate, o-phosphocholine, serine, choline, SN-glycero-3-phosphocholine, lactate, acetate, and betaine were found significantly down-regulated.

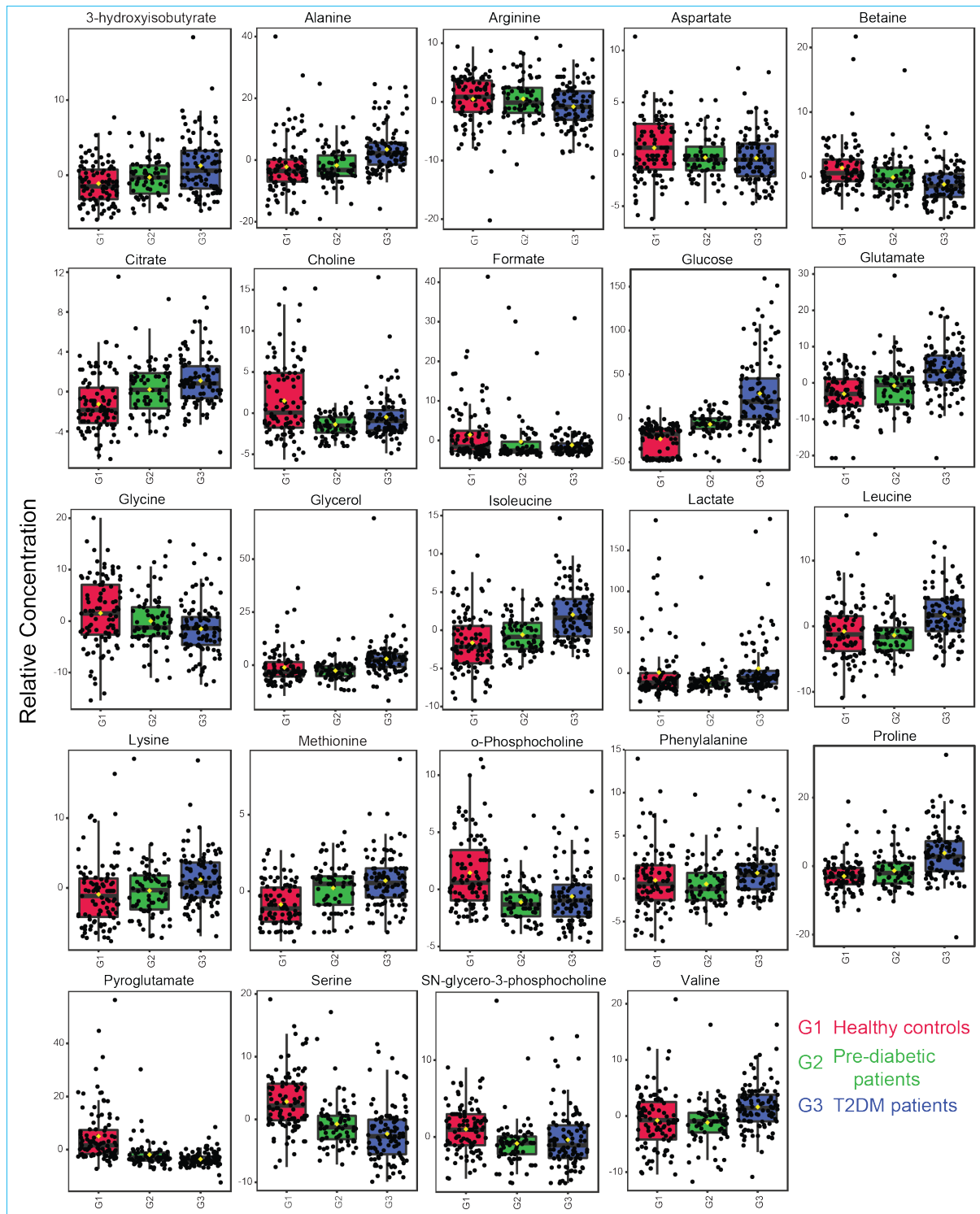


Figure 4.2: Box and whisker plots displaying relative normalized concentrations of significantly altered metabolites identified from One-way ANOVA (q -value < 0.05) in control, prediabetic, and T2DM groups.

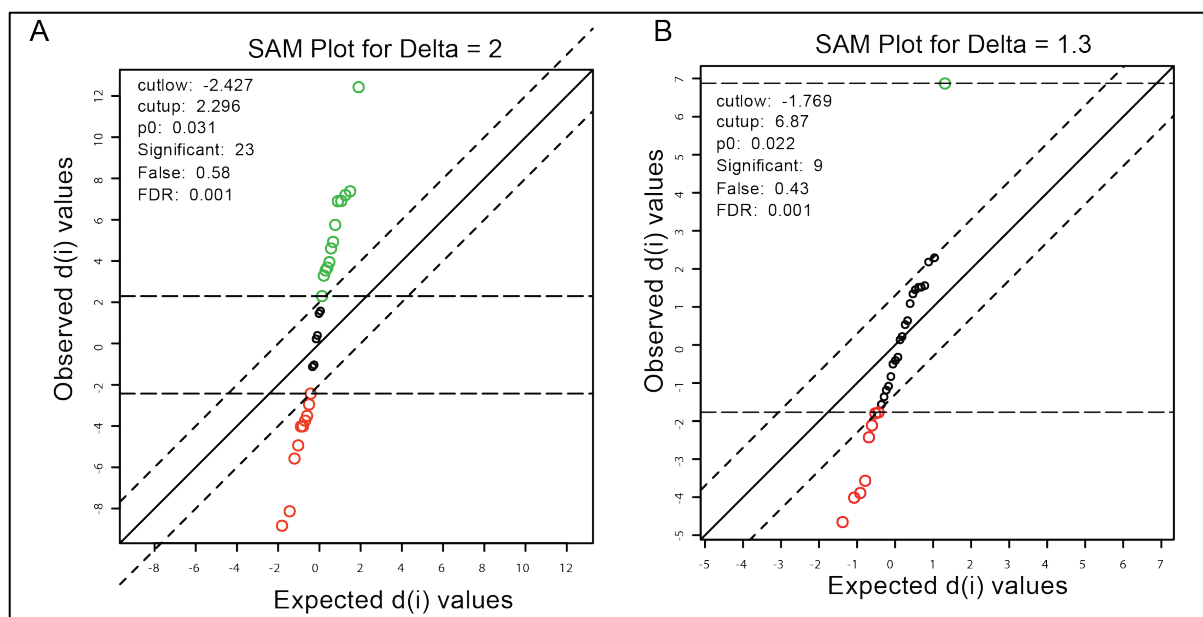


Figure 4.3: Significance analysis of microarray (SAM) plots for the selection of significant metabolites in (A) T2DM group, and (B) prediabetic group. The green circles on the top represent up-regulated metabolites, and the red circles on the bottom represent down-regulated metabolites. $d(i)$ indicates SAM score.

Supervised OPLS-DA analysis was carried out to discern the alterations in metabolic profiles of serum samples between the two groups individually, i.e., T2DM subjects against control samples and prediabetic subjects against healthy controls. The score plot in the OPLS-DA analysis showed a clear separation between healthy controls and T2DM subjects (Figure 4. 1A). The goodness of fit and predictive ability for the OPLS-DA models between T2DM groups and healthy control was revealed in the values of $R^2 = 0.75$, $Q^2 = 0.74$. The permutation test ($n=100$), which was carried out to validate the goodness of these models further, showed robustness and high predictability of the model ($p\text{-value} < 0.05$) (Figure 4. 2B). However, a slight overlap between the ellipses representing the score plot of the healthy controls and prediabetic subjects was observed using OPLS-DA analysis (Figure 4. 3C), with an R^2 value of 0.38 and a Q^2 value of 0.35. Thus, the results from OPLS-DA analysis unveiled good models to discriminate healthy individuals from the diabetic and prediabetic subjects with good predictive ability. Indeed, the OPLS-DA model distinguishing between T2DM and healthy individuals had better predictive performance (R^2 value of 0.75 and Q^2 value of 0.74) than the OPLS-DA model discriminating between healthy and prediabetic patients (R^2 value of 0.38 and Q^2 value of 0.35).

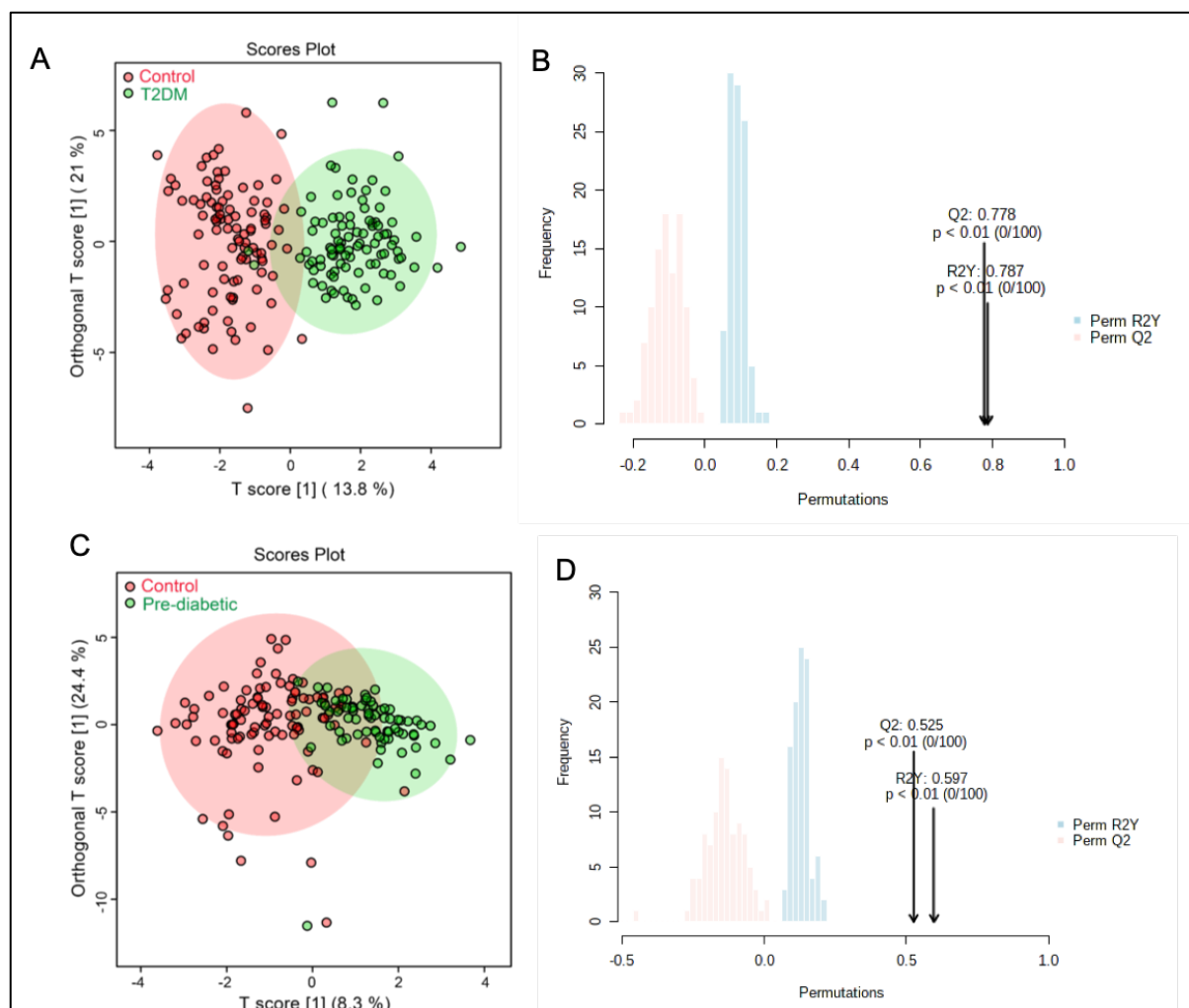


Figure 4.4: Supervised 2D OPLS-DA analysis of (A) Healthy control versus T2DM patients, (C) Healthy controls versus prediabetic patients. OPLS-DA model validation by permutation tests based on the separation between (B) Healthy control versus T2DM patients (D) Healthy control versus prediabetic patients. The p-value based on permutation is $p < 0.01$, indicating the significance of both the models. R2 indicating the goodness of fit, and Q2 indicating the predictability of the model.

4.8.4 Identification of Metabolite Biomarkers associated with diabetes

4.8.4.1 Metabolic Biomarkers associated with T2DM

To identify potential biomarkers for early detection or progression of T2DM, we applied classical univariate ROC curve analysis to produce ROC curves to evaluate the area under the curve (AUC) and their 95% confidence intervals. Metabolites with $AUC > 0.7$ have generally been considered as good biomarkers. As shown in Figure 4.5, twelve metabolites with $AUC > 0.7$ were identified as potential biomarkers for diagnosis, surveillance, and early detection of metabolic changes in T2DM patients. These included glucose, pyroglutamate, serine, proline, glutamate, methionine, isoleucine, alanine, citrate, betaine, glycerol, and o-phosphocholine. Glucose was found to have the highest AUC value, and thereby was the

metabolite associated with the highest sensitivity and specificity (AUC: 0.95; 0.9, 0.8). This was followed by pyroglutamate (AUC: 0.89; 0.9, 0.8) and serine (AUC: 0.83; 0.8, 0.7).

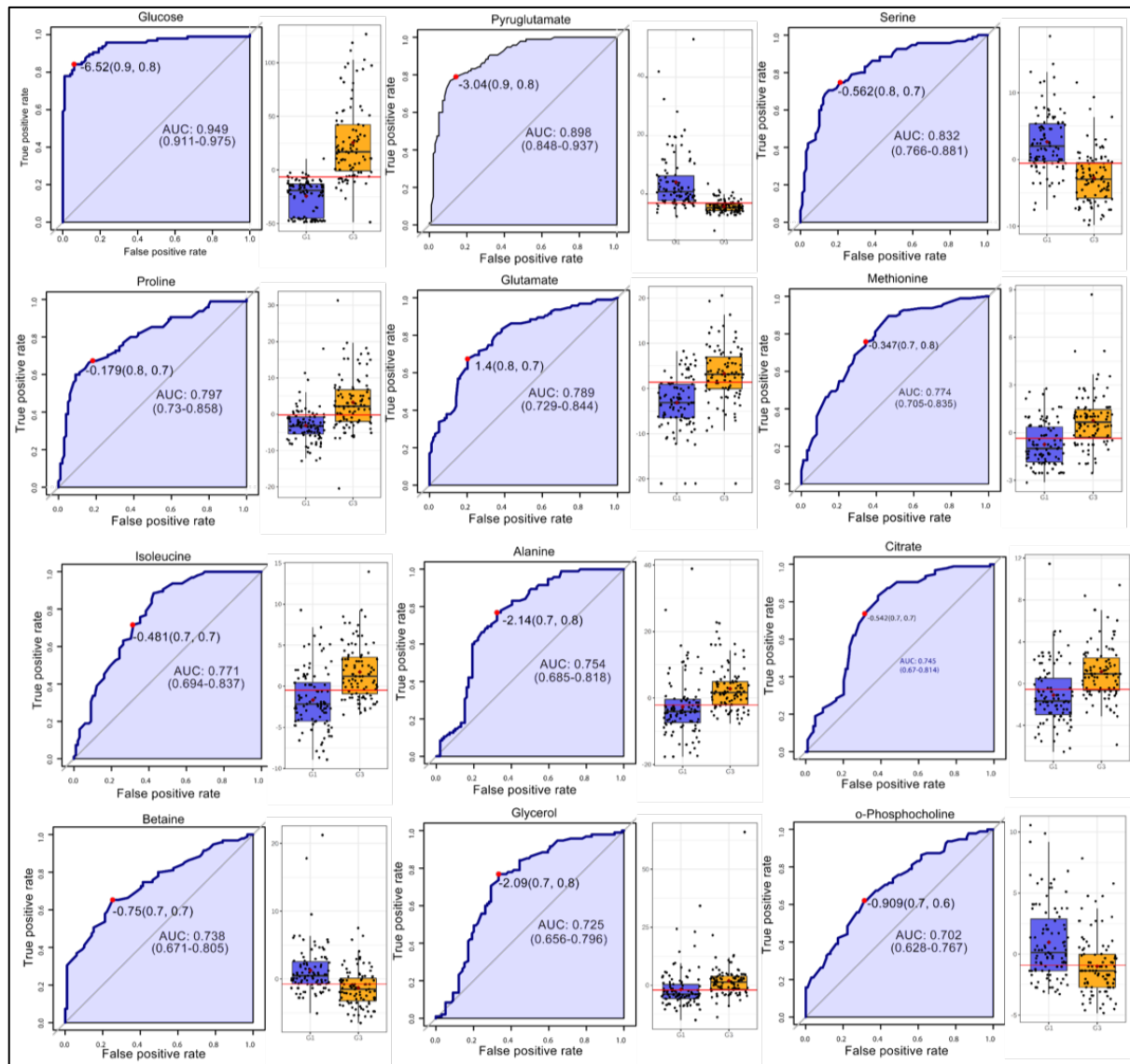


Figure 4.5: Univariate ROC curves of the potential biomarkers (AUC > 0.7) and their respective box-cum-whisker plots showing discriminatory ability between T2DM patients and healthy controls. AUCs and computed confidence intervals have been shown for each biomarker in the figure. AUC: the area under the curve.

4.8.5 Metabolic Biomarkers for prediabetes

ROC curve analysis revealed six metabolite biomarkers associated with the prediabetic groups based on AUC > 0.7 (Figure 4.6). These metabolites have high discriminating power for identifying prediabetic subjects from healthy individuals. These included glucose, pyroglutamate, o-phosphocholine, serine, snlycero-3-phosphocholine, and methionine. As in the T2DM analysis, glucose had the highest AUC value (AUC: 0.81; 0.8, 0.8), followed by

pyroglutamate (AUC: 0.80; 0.7, 0.8) and o-phosphocholine (AUC: 0.75; 0.7, 0.7). The AUC values ranging from 0.81-0.71 suggests that these metabolites could be used as potential biomarkers for individuals at risk for developing overt diabetes in the future. Interestingly, most metabolites were found to follow a similar trend with respect to their concentration in the T2DM and prediabetic groups when compared to the healthy individuals (Figure 4.2).

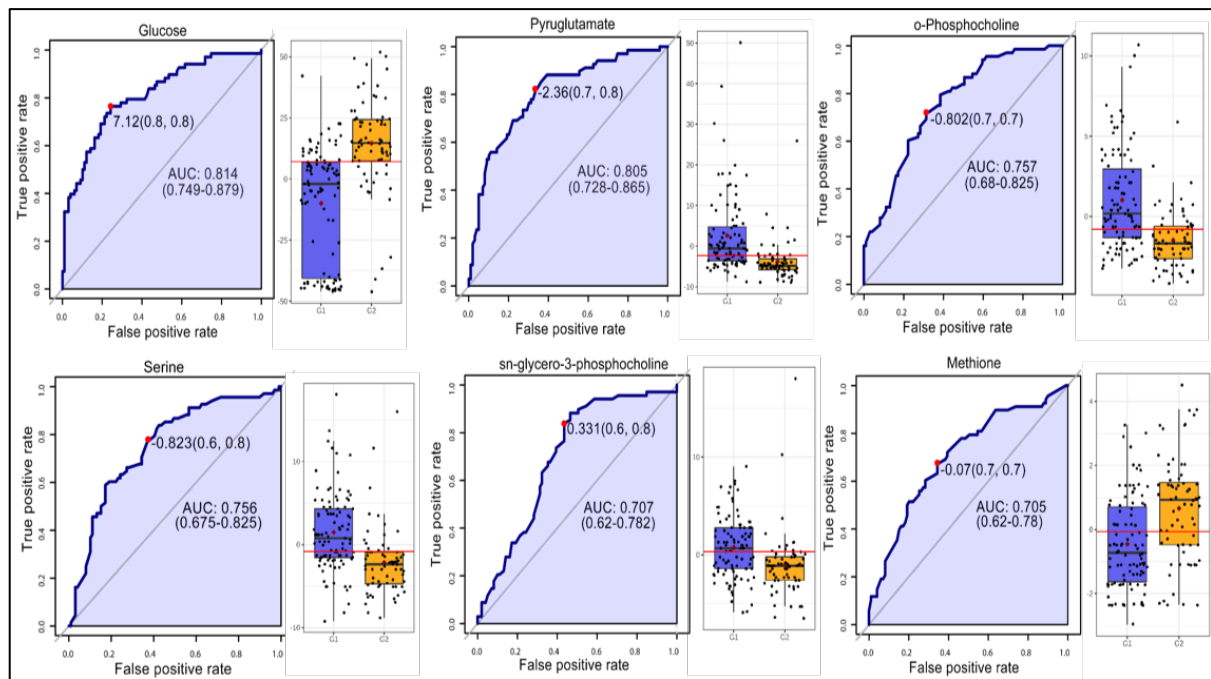


Figure 4.6: Univariate ROC curves of the potential biomarkers (AUC > 0.7) along with their respective box-cum-whisker plots showing discriminatory ability between prediabetic patients and healthy controls. AUCs and computed confidence intervals have been shown for each biomarker in the figure. AUC: the area under the curve.

Further, a multivariate ROC curve analysis was carried out to select a panel of biomarkers (multiple-biomarker test) and to evaluate the classification performance of the generated models. Balanced sub-sampling-based Monte Carlo cross-validation (MCCV) was applied to generate the ROC curves for six different models where each model has a different number of metabolites (2, 3, 5, 10, 20, and 29) included by sub-sampling. The Linear Support Vector Machine (SVM) for the classification and "Random Forest algorithm" for the feature ranking method was used for the analysis. From the results of the multivariate ROC curve analysis, AUC values were increased with an increased number of variables. Figure 4.7 shows that the AUC value ranges from 0.97 for two variables to 0.99 for 29 variables in the T2DM group and from 0.81 for two variables to 0.92 for 29 variables in prediabetic group. Thus, the more the number of variables in the model, the more was the discriminating potential of the

models in distinguishing T2DM and prediabetic subjects from healthy individuals. To further compare different classification models, predictive accuracy was evaluated for each model. The results from predictive accuracy revealed that the model performance was improved with an increasing number of variables, which corresponds to the results obtained from the ROC curve analysis.

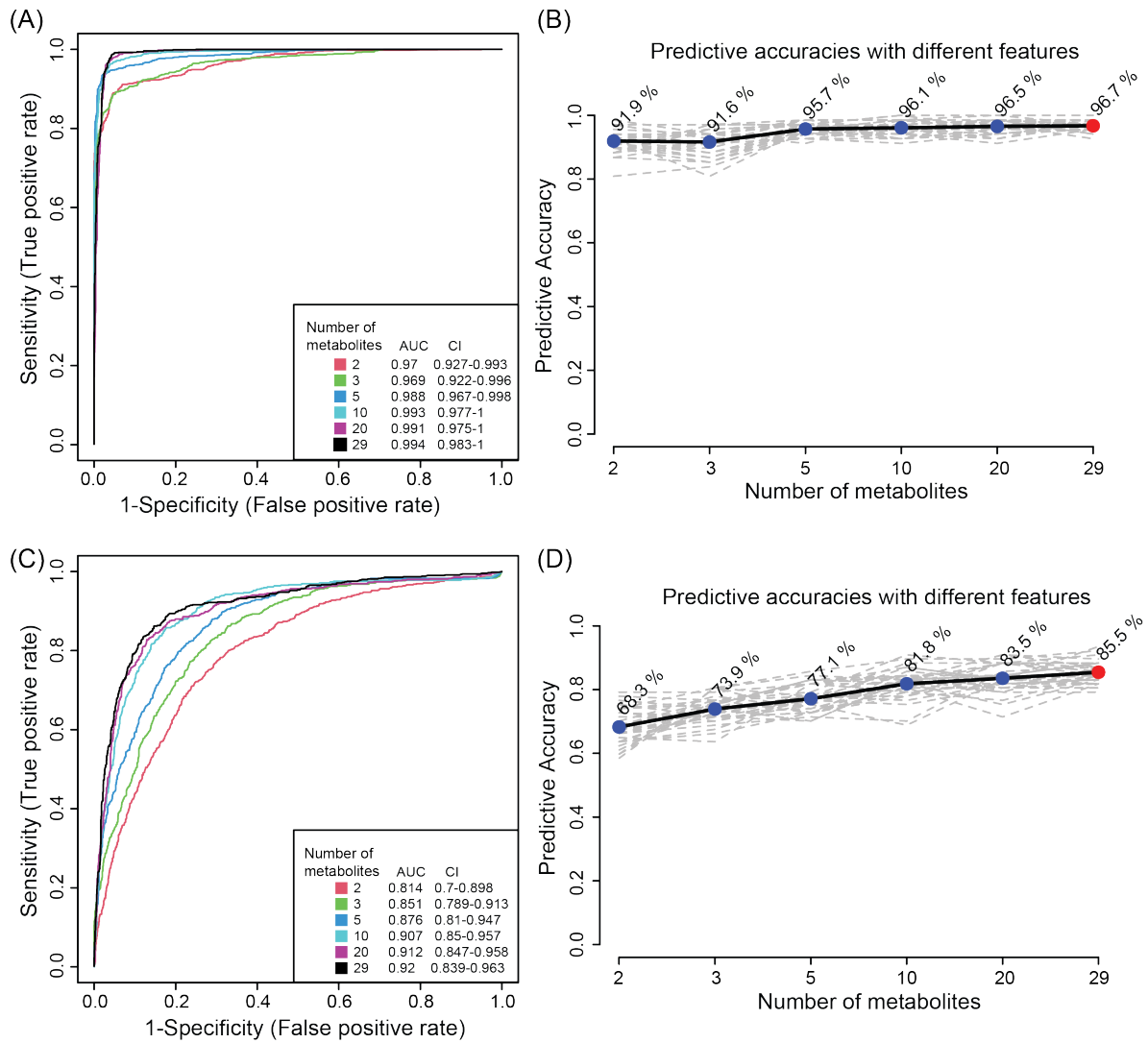


Figure 4.7: Multivariate ROC curve analysis for the evaluation of model prediction using balanced-subsampling based Monte-Carlo cross-validation (MCCV). ROC curve based comparative analysis of 6 different models using different metabolite panels (2, 3, 5, 10, 20, and 29). The corresponding AUCs and confidence intervals were shown for each model in the figure. (A) Metabolite biomarkers detected for T2DM patients, (C) Metabolite biomarkers detected for prediabetic patients. Predictive accuracies for the different number of metabolite panels in (B) T2DM patients, (D) prediabetic patients. The red dot shows the highest predictive accuracy for the model featuring 29 metabolites in both T2DM and prediabetic groups. AUC, the area under the curve; CI, confidence interval.

4.8.6 Metabolic Pathway Analysis

The metabolites identified through SAM analysis were employed to thoroughly analyze the altered pathways in T2DM and prediabetes compared to healthy controls using MetaboAnalyst, in order to assess the biological relevance of the changes in metabolite levels.

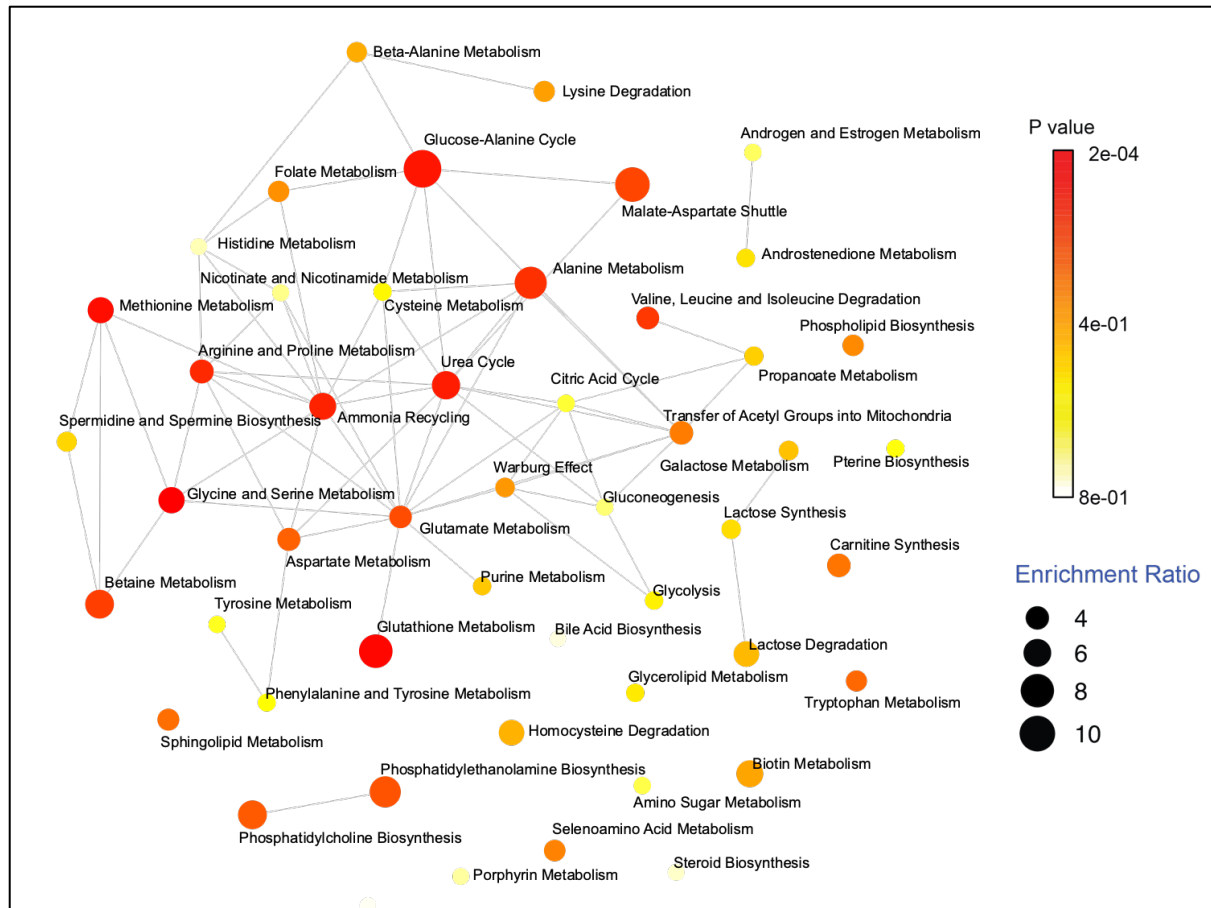


Figure 4.8: Summary of pathway enrichment analysis in T2DM serum samples. Each node represents a specific pathway with node color corresponds to its p-value, and the size of the node corresponds to fold enrichment. Two pathways are connected with edges if they shared more than 25 % metabolite of the total number of combined metabolites.

Metabolic pathway analysis was performed using the *Homo sapiens* pathway library with parameters "Hypergeometric test" for Over-Representation Analysis and "Relative-betweenness Centrality" for pathway topology analysis. Pathway enrichment analysis was performed to identify all significant pathways ($p < 0.05$) based on the fold enrichment ratio (hits/expected) in T2DM and prediabetic groups when compared to the healthy individuals. Each node in the pathway enrichment analysis (Figure 4.8) corresponds to the metabolic pathway with its color based on its p-value, and the size of the node corresponds to its fold enrichment. The significantly perturbed pathways were then identified based on their pathway impact values calculated from pathway topology analysis. All the identified metabolic

pathways were plotted with p-value along the y-axis and pathway impact score along the x-axis (Figure 4.9). The metabolic pathways with a p-value < 0.05 and impact score > 0.1 were deemed as a significantly perturbed pathway with great relevance to disease mechanism.

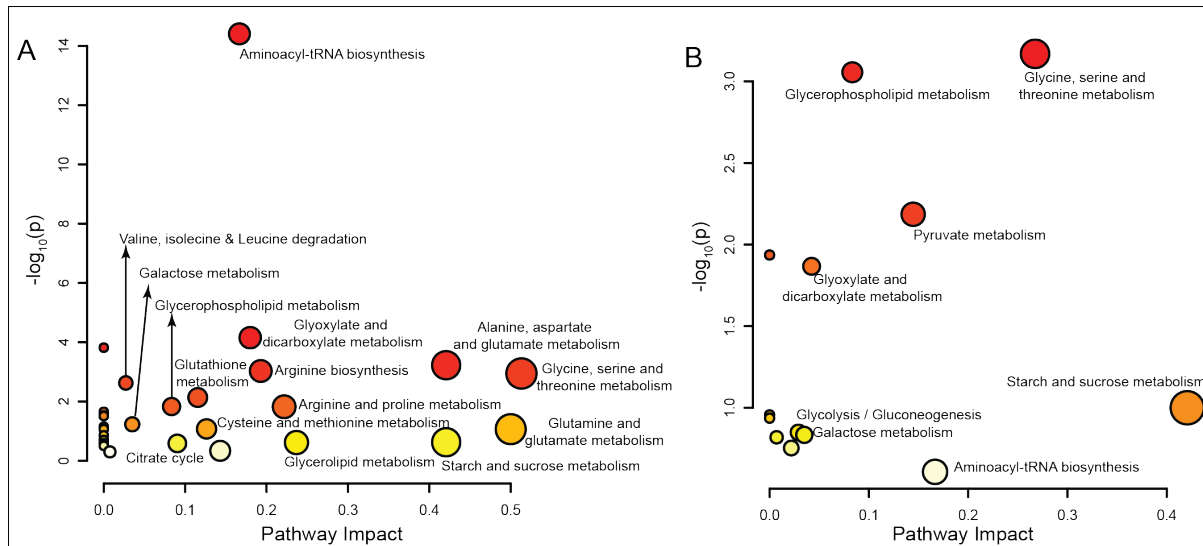


Figure 4.9: Metabolic Pathway Analysis (MetPa) of significantly differential metabolites between healthy controls and (A) T2DM groups, (B) Prediabetic groups. Each node represents a single with node color corresponding to the $-\log_{10}(P)$ value (red: higher p-values and red: lower p-values) and node size corresponding to the pathway impact score.

The major pathways perturbed significantly in the pairwise comparison between healthy individuals and T2DM subjects included glycine-serine-threonine metabolism; alanine, aspartate, and glutamate metabolism; aminoacyl-tRNA biosynthesis; arginine and proline metabolism; glyoxylate and dicarboxylate metabolism; glutathione, glycerophospholipid metabolism; cysteine and methionine metabolism; galactose metabolism; glutamine-glutamate metabolism; and starch and sucrose metabolism (Figure 4.9A). Likewise, the significantly perturbed pathways in the prediabetic individuals were glycine-serine-threonine metabolism; glycerophospholipid metabolism; pyruvate metabolism; starch, and sucrose metabolism; and glyoxylate and dicarboxylate metabolism (Figure 4.9B). Overall the significantly perturbed pathways in our study belong to amino-acid metabolism and energy metabolism. All the identified pathways were then constructed using the KEGG metabolic network and depicted pictorially in Figure 4.10.

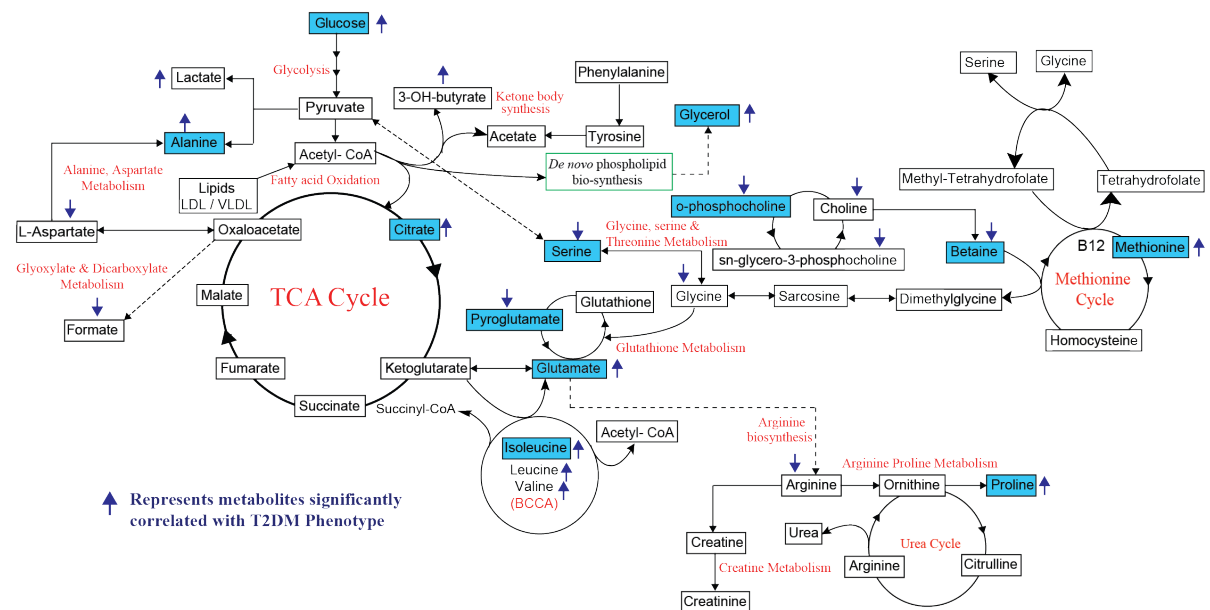


Figure 4.10: Pictorial depiction of the significantly dysregulated metabolites and Metabolomic Pathway Analysis (MetPA) construction of the metabolic pathways identified in T2DM. The cyan colored boxes represent potential metabolites with AUC > 0.07. The upward and downward direction of arrows indicates that the metabolites are up-regulated or down-regulated in T2DM patients.

4.9 Discussion

Using a clinically well-defined cohort of subjects with prediabetes, established T2DM, and healthy volunteers from South Asia, ¹H NMR-based metabolomics was performed, which revealed a clear separation in the serum metabolomes between these groups. Univariate ROC curve analysis identified 6 and 12 metabolites from the prediabetic and diabetic groups, respectively, with AUC values > 0.7. The five common significantly dysregulated metabolites between these groups, namely glucose, pyroglutamate, o-phosphocholine, serine, and methionine, were reanalyzed using multivariate ROC curve analysis, and highly significant AUC values were obtained. The AUC values obtained were 0.96 (95% CI = 0.93, 0.98) for established T2DM (Figure 4. 11A); and 0.88 (95% CI = 0.81, 0.93) for prediabetes subjects (Figure 4. 11B), which are highly significant. Hence, we propose that this panel of 5 metabolic biomarkers can be used in future for predicting individuals at risk for developing overt diabetes in the future in the South Asian Indians.

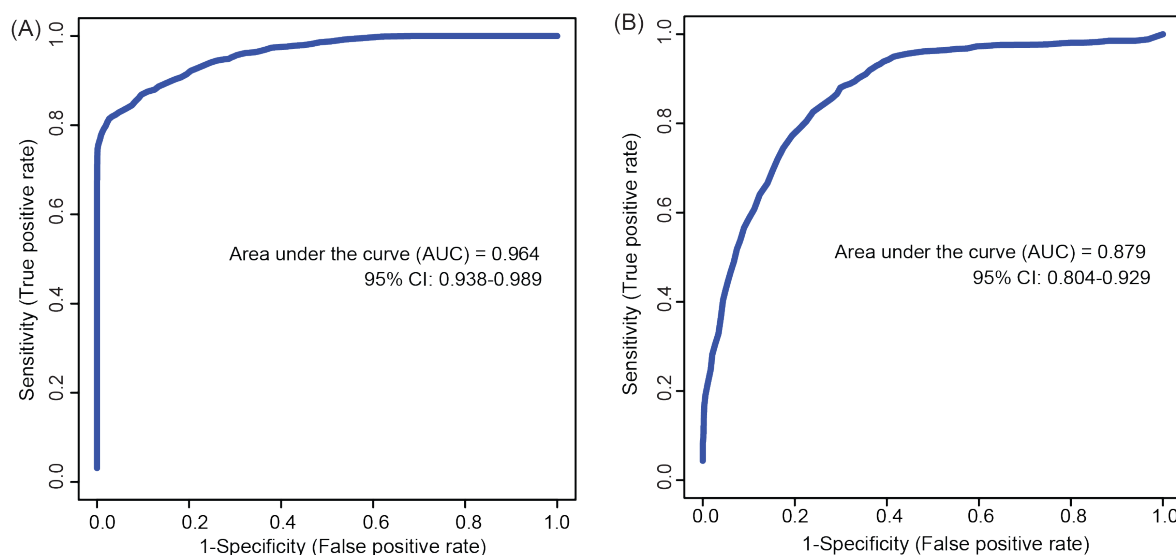


Figure 4.11: Multivariate ROC curve analysis of serum samples using the combination of five metabolite biomarkers (glucose, pyroglutamate, o-phosphocholine, serine, and methionine) common in T2DM and Prediabetic subjects with the AUC >0.7. The Multivariate ROC curve analysis displays a good diagnostic power in discriminating healthy controls from T2DM subjects (AUC = 0.964) (A); and healthy controls from Prediabetic subjects (AUC = 0.879) (B). AUC, the area under the curve; CI, confidence interval.

Among the five metabolites used for multivariate ROC curve analysis, pyroglutamate and methionine levels were significantly increased in the prediabetic and diabetic subjects; and the levels of pyroglutamate, o-phosphocholine, and serine showed a significant decrease in these individuals when compared to the healthy normal controls.

Glucose was the most significantly different metabolite between the prediabetic and diabetic subjects compared to the healthy volunteers. This result is not surprising (and worked as a positive control for our analysis) as fasting plasma glucose levels and glucose levels during an oral glucose tolerance test have routinely been used to diagnose T2DM. However, these levels are often detected after T2DM has progressed and metabolic syndrome has already set in. In addition to this, increased plasma glucose levels are not specific for T2DM alone and can be increased in other metabolic disorders as well (Akter et al. 2017; Jeanes and Reeves 2017). As a result, measuring glucose levels alone does not provide sufficient power for predicting individuals at risk of developing T2DM in the future and needs to be combined with more specific biomarkers. Increased glucose levels play a crucial role in increasing the metabolic flux into the glycolytic pathway, thereby leading to elevated lactate levels, an independent risk factor identified for the development of T2DM (Crawford et al. 2010; Guo et al. 2012). In addition to lactate, pyruvate is also produced during glycolysis, which in the form of acetyl-CoA can either enter the TCA cycle; or used for lipogenesis; or can be used for the production

of dihydroxyacetone phosphate and glycerol-3-phosphate – substrates required for triglyceride synthesis (Guo et al. 2012). We observed increased glycerol levels in prediabetic and diabetic subjects recruited in the study compared to the healthy controls (Figure 4.2). Increased lipolysis that, in turn, leads to the overproduction of glycerol and free fatty acids from triglycerides occur during insulin-resistant states in diabetes. Increased glycerol stimulates gluconeogenesis and further contributes to hyperglycemia (Anon 1963; Mahendran et al. 2013). Indeed, previous studies have proposed the use of glycerol and free fatty acids as a predictor for hyperglycemia and T2DM (Ahola-Olli et al. 2019; Mahendran et al. 2013).

After glucose, the next significant predictor identified from our study was pyroglutamate, a vital metabolite that plays a crucial role in the intracellular transport of free amino acids. Studies in diabetic Goto-Kakizaki rats and KK-Ay mice have shown that dietary inclusion of pyroglutamic acid modifies glucose and lipid metabolism in these animals and potentially contributes to T2DM mitigation (Yoshinari and Igarashi 2011). In addition to this, reduced pyroglutamate levels have been reported in patients with isolated post-challenge diabetes (IPD) (Chou et al. 2018); in subjects with impaired fasting glucose and diabetes (Xu et al. 2013); and T2DM patients with nephropathy (Shao et al. 2020).

Levels of o-phosphocholine, an intermediate in the synthesis of phosphatidylcholine – a major phospholipid in cellular membranes, were also significantly reduced in patients with prediabetes and T2DM. Our previous investigations in pancreatic *INS-1E* cells exposed to gluco-, lipo- and glucolipo-toxic conditions (chapter 3 (Yousf et al. 2019)) and in studies conducted in urine samples of diabetic rats (Guan et al. 2013), reduced o-phosphocholine levels have been reported. Reduced phosphatidylcholine levels have also been demonstrated in type 1 diabetes subjects (Artykbaeva and Saatov 2020); while phosphatidylcholine levels in skeletal muscles have been linked to altered insulin sensitivity (Lee et al. 2018).

Serine, a non-essential amino acid, plays a central role in several crucial biological processes. Several studies have suggested a role of serine metabolism in type 1, type 2, and gestational diabetes; and supplementation of serine has been associated with increased insulin secretion from the pancreatic β -cells, improved insulin sensitivity by the insulin target tissues, thereby leading to improved glucose homeostasis; improved mitochondrial function and relief of endoplasmic reticulum stress (Bertea et al. 2010; Bervoets et al. 2017; Drábková et al. 2015; Holm and Buschard 2019; Vangipurapu et al. 2019). In line with these observations, our study also identified reduced serine levels in subjects with prediabetes and established T2DM.

An essential amino acid – methionine – showed significantly increased levels in prediabetic and T2DM subjects recruited for our study. Dietary restriction (energy restriction) methionine in rodents and humans has been associated with improved glucose homeostasis, improved insulin sensitivity, and reduction in weight gain (Brown-Borg and Buffenstein 2017; Castaño-Martinez et al. 2019; Miller et al. 2005; Stone et al. 2014; Ying et al. 2017).

4.10 Conclusions

To conclude, we performed ¹H NMR-based metabolomics on serum samples collected from age and sex-matched cohort of 284 individuals consisting of sub-groups of healthy controls, prediabetic patients, and type 2 diabetic patients. Out of 36 identified and quantified metabolites, univariate ROC curve analysis yielded a total of six significantly perturbed metabolites in prediabetic patients and twelve in T2DM patients. Further, a set of common five metabolites were chosen from prediabetic and T2DM groups having AUC > 0.7 and subjected to multivariate ROC curve analysis. Multivariate ROC curve analysis highlighted that the selected panel of five metabolites (glucose, pyroglutamate, o-phosphocholine, serine, and methionine) could be used as potential biomarkers for more confident detection of prediabetes and T2DM, and possibly as early biomarkers for prediction of T2DM in individuals where the disease has not set in already. It would also be interesting to analyse the effect of comorbidities, including hypertension and obesity on these metabolites in future.

References

- Ahola-Olli, Ari V., Linda Mustelin, Maria Kalimeri, Johannes Kettunen, Jari Jokelainen, Juha Auvinen, Katri Puukka, Aki S. Havulinna, Terho Lehtimäki, Mika Kähönen, Markus Juonala, Sirkka Keinänen-Kiukaanniemi, Veikko Salomaa, Markus Perola, Marjo Riitta Järvelin, Mika Ala-Korpela, Olli Raitakari, and Peter Würtz. 2019. "Circulating Metabolites and the Risk of Type 2 Diabetes: A Prospective Study of 11,896 Young Adults from Four Finnish Cohorts." *Diabetologia*.
- Akter, R., A. Nessa, M. F. Husain, F. Wahed, N. Khatun, M. Yesmin, S. Nasreen, and T. Tajkia. 2017. "Effect of Obesity on Fasting Blood Sugar." *Mymensingh Medical Journal : MMJ* 26(1):7–11.
- Anon. 1963. "Effect of Insulin on Plasma Glycerol in Man." *JAMA: The Journal of the American Medical Association* 184(3):225.
- Arneth, Borros, Rebekka Arneth, and Mohamed Shams. 2019. "Metabolomics of Type 1 and Type 2 Diabetes." *International Journal of Molecular Sciences* 20(10).
- Artykbaeva, Gulnora and Talat Saatov. 2020. "Phospholipid Metabolism in Patients with Type I and 2 Diabetes Mellitus." *Endocrine Abstracts*.
- Berteau, Mariana, Markus F. Rützi, Alaa Othman, Jaqueline Marti-Jaun, Martin Hersberger, Arnold Von Eckardstein, and Thorsten Hornemann. 2010. "Deoxysphingoid Bases as Plasma Markers in Diabetes Mellitus." *Lipids in Health and Disease*.
- Bervoets, Liene, Guy Massa, Wanda Guedens, Evelyne Louis, Jean Paul Noben, and Peter Adriaenssens. 2017. "Metabolic Profiling of Type 1 Diabetes Mellitus in Children and Adolescents: A Case-Control Study NCT03014908 NCT." *Diabetology and Metabolic Syndrome*.
- Bhatia, Parnika, Shikha Raina, Jeetender Chugh, and Shilpy Sharma. 2015. "MiRNAs: Early Prognostic Biomarkers for Type 2 Diabetes Mellitus?" *Biomarkers in Medicine* 9(10):1025–40.
- Brown-Borg, Holly M. and Rochelle Buffenstein. 2017. "Cutting Back on the Essentials: Can Manipulating Intake of Specific Amino Acids Modulate Health and Lifespan?" *Ageing Research Reviews*.
- Castaño-Martinez, Teresa, Fabian Schumacher, Silke Schumacher, Bastian Kochlik, Daniela Weber, Tilman Grune, Ronald Biemann, Adrian McCann, Klaus Abraham, Cornelia Weikert, Burkhard Kleuser, Annette Schürmann, and Thomas Laeger. 2019. "Methionine Restriction Prevents Onset of Type 2 Diabetes in NZO Mice." *FASEB Journal : Official Publication of the Federation of American Societies for Experimental*

Biology.

- Chawla, Rajeev, SV Madhu, BM Makkar, Sujoy Ghosh, Banshi Saboo, and Sanjay Kalra. 2020. "RSSDI-ESI Clinical Practice Recommendations for the Management of Type 2 Diabetes Mellitus 2020." *Indian Journal of Endocrinology and Metabolism* 24(1):1.
- Chou, Jing, Rui Liu, Jiaying Yu, Xiaowei Liu, Xinshu Zhao, Ying Li, Liyan Liu, and Changhao Sun. 2018. "Fasting Serum A-hydroxybutyrate and Pyroglutamic Acid as Important Metabolites for Detecting Isolated Post-Challenge Diabetes Based on Organic Acid Profiles." *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences.*
- Del Coco, Laura, Daniele Vergara, Serena De Matteis, Emanuela Mensà, Jacopo Sabbatinelli, Francesco Prattichizzo, Anna Rita Bonfigli, Gianluca Storci, Sara Bravaccini, Francesca Pirini, Andrea Ragusa, Andrea Casadei-Gardini, Massimiliano Bonafè, Michele Maffia, Francesco Paolo Fanizzi, Fabiola Olivieri, and Anna Maria Giudetti. 2019. "NMR-Based Metabolomic Approach Tracks Potential Serum Biomarkers of Disease Progression in Patients with Type 2 Diabetes Mellitus." *Journal of Clinical Medicine.*
- Crawford, Stephen O., Ron C. Hoogeveen, Frederick L. Brancati, Brad C. Astor, Christie M. Ballantyne, Maria Inês Schmidt, and Jeffery Hunter Young. 2010. "Association of Blood Lactate with Type 2 Diabetes: The Atherosclerosis Risk in Communities Carotid MRI Study." *International Journal of Epidemiology.*
- Defronzo, Ralph A. 2009. "From the Triumvirate to the Ominous Octet: A New Paradigm for the Treatment of Type 2 Diabetes Mellitus." in *Diabetes.*
- Drábková, Petra, Jana Šanderová, Jakub Kovařík, and Roman Kanmár. 2015. "An Assay of Selected Serum Amino Acids in Patients with Type 2 Diabetes Mellitus." *Advances in Clinical and Experimental Medicine.*
- Fikri, Asma M., Rosemary Smyth, Vijay Kumar, Zainab Al-Abadla, Salahdeen Abusnana, and Michael R. Munday. 2020. "Pre-Diagnostic Biomarkers of Type 2 Diabetes Identified in the UAE's Obese National Population Using Targeted Metabolomics." *Scientific Reports* 10(1):1–10.
- Floegel, Anna, Norbert Stefan, Zhonghao Yu, Kristin Mühlenbruch, Dagmar Drogan, Hans Georg Joost, Andreas Fritsche, Hans Ulrich Häring, Martin Hrabě De Angelis, Annette Peters, Michael Roden, Cornelia Prehn, Rui Wang-Sattler, Thomas Illig, Matthias B. Schulze, Jerzy Adamski, Heiner Boeing, and Tobias Pischon. 2013. "Identification of Serum Metabolites Associated with Risk of Type 2 Diabetes Using a Targeted Metabolomic Approach." *Diabetes.*

- Gogna, Navdeep, Murahari Krishna, Anup Mammen Oommen, and Kavita Dorai. 2015. "Investigating Correlations in the Altered Metabolic Profiles of Obese and Diabetic Subjects in a South Indian Asian Population Using an NMR-Based Metabolomic Approach." *Molecular BioSystems*.
- Gowda, G. A. Nagan., Shucha Zhang, Haiwei Gu, Vincent Asiago, Narasimhamurthy Shanaiah, and Daniel Raftery. 2008. "Metabolomics-Based Methods for Early Disease Diagnostics." *Expert Review of Molecular Diagnostics* 8(5):617–33.
- Guan, Mimi, Liyun Xie, Chengfeng Diao, Na Wang, Wenyi Hu, Yongquan Zheng, Litai Jin, Zhihan Yan, and Hongchang Gao. 2013. "Systemic Perturbations of Key Metabolites in Diabetic Rats During the Evolution of Diabetes Studied by Urine Metabonomics." *PLoS ONE*.
- Guasch-Ferré, Marta, Adela Hruby, Estefanía Toledo, Clary B. Clish, Miguel A. Martínez-González, Jordi Salas-Salvadó, and Frank B. Hu. 2016. "Metabolomics in Prediabetes and Diabetes: A Systematic Review and Meta-Analysis." *Diabetes Care* 39(5):833–46.
- Guo, Xin, Honggui Li, Hang Xu, Shihlung Woo, Hui Dong, Fuer Lu, Alex J. Lange, and Chaodong Wu. 2012. "Glycolysis in the Control of Blood Glucose Homeostasis." *Acta Pharmaceutica Sinica B*.
- Holm, Laurits J. and Karsten Buschard. 2019. "L-Serine: A Neglected Amino Acid with a Potential Therapeutic Role in Diabetes." *APMIS*.
- Jeanes, Yvonne M. and Sue Reeves. 2017. "Metabolic Consequences of Obesity and Insulin Resistance in Polycystic Ovary Syndrome: Diagnostic and Methodological Challenges." *Nutrition Research Reviews*.
- Klein, Matthias S. and Jane Shearer. 2016. "Metabolomics and Type 2 Diabetes: Translating Basic Research into Clinical Application." *Journal of Diabetes Research* 2016.
- Lee, Sindre, Frode Norheim, Hanne L. Gulseth, Torgrim M. Langleite, Andreas Aker, Thomas E. Gundersen, Torgeir Holen, Kåre I. Birkeland, and Christian A. Drevon. 2018. "Skeletal Muscle Phosphatidylcholine and Phosphatidylethanolamine Respond to Exercise and Influence Insulin Sensitivity in Men." *Scientific Reports*.
- Mahendran, Yuvaraj, Henna Cederberg, Jagadish Vangipurapu, Antti J. Kangas, Pasi Soininen, Johanna Kuusisto, Matti Uusitupa, Mika Ala-Korpela, and Markku Laakso. 2013. "Glycerol and Fatty Acids in Serum Predict the Development of Hyperglycemia and Type 2 Diabetes in Finnish Men." *Diabetes Care*.
- Mallol, Roger, Miguel Angel Rodriguez, Jesus Brezmes, Lluís Masana, and Xavier Correig. 2013. "Human Serum/Plasma Lipoprotein Analysis by NMR: Application to the Study

- of Diabetic Dyslipidemia.” *Progress in Nuclear Magnetic Resonance Spectroscopy* 70:1–24.
- Mihalik, Stephanie J., Bret H. Goodpaster, David E. Kelley, Donald H. Chace, Jerry Vockley, Frederico G. S. Toledo, and James P. Delany. 2010. “Increased Levels of Plasma Acylcarnitines in Obesity and Type 2 Diabetes and Identification of a Marker of Glucolipototoxicity.” *Obesity*.
- Miller, Richard A., Gretchen Buehner, Yayi Chang, James M. Harper, Robert Sigler, and Michael Smith-Wheelock. 2005. “Methionine-Deficient Diet Extends Mouse Lifespan, Slows Immune and Lens Aging, Alters Glucose, T4, IGF-I and Insulin Levels, and Increases Hepatocyte MIF Levels and Stress Resistance.” *Aging Cell*.
- Pérez-Matos, Maria Camila, Martha Catalina Morales-álvarez, Freddy Jean Karlo Toloza, Maria Laura Ricardo-Silgado, Jose Oscar Mantilla-Rivas, Jairo Arturo Pinzón-Cortes, Maritza Perez-Mayorga, Elizabeth Jiménez, Edwin Guevara, and Carlos O. Mendivil. 2017. “The Phospholipid Linoleoylglycerophosphocholine as a Biomarker of Directly Measured Insulin Resistance.” *Diabetes and Metabolism Journal*.
- Qiu, Gaokun, Yan Zheng, Hao Wang, Jie Sun, Hongxia Ma, Yang Xiao, Yizhun Li, Yu Yuan, Handong Yang, Xiulou Li, Xinwen Min, Ce Zhang, Chengwei Xu, Yue Jiang, Xiaomin Zhang, Meian He, Ming Yang, Zhibin Hu, Huiru Tang, Hongbing Shen, Frank B. Hu, An Pan, and Tangchun Wu. 2016. “Plasma Metabolomics Identified Novel Metabolites Associated with Risk of Type 2 Diabetes in Two Prospective Cohorts of Chinese Adults.” *International Journal of Epidemiology* 45(5):1507–16.
- Ramachandran, A., C. Snehalatha, Vijay Viswanathan, M. Viswanathan, and S. M. Haffner. 1997. “Risk of Noninsulin Dependent Diabetes Mellitus Conferred by Obesity and Central Adiposity in Different Ethnic Groups: A Comparative Analysis between Asian Indians, Mexican Americans and Whites.” *Diabetes Research and Clinical Practice* 36(2):121–25.
- Ramachandran, Ambady, Ronald Ching Wan Ma, and Chamukuttan Snehalatha. 2010. “Diabetes in Asia.” *The Lancet* 375(9712):408–18.
- Rawat, Atul, Durgesh Dubey, Anupam Guleria, Umesh Kumar, Amit Keshari, Swati Chaturvedi, Anand Prakash, Sudipta Saha, and Dinesh Kumar. 2016. “¹H NMR-Based Serum Metabolomics Reveals Erythromycin-Induced Liver Toxicity in Albino Wistar Rats.” *Journal of Pharmacy and Bioallied Sciences*.
- Rughani, Ankur, Jacob E. Friedman, and Jeanie B. Tryggestad. 2020. “Type 2 Diabetes in Youth: The Role of Early Life Exposures.” *Current Diabetes Reports* 20(9).

- Satheesh, Gopika, Surya Ramachandran, and Abdul Jaleel. 2020a. “Metabolomics-Based Prospective Studies and Prediction of Type 2 Diabetes Mellitus Risks.” *Metabolic Syndrome and Related Disorders* 18(1):1–9.
- Satheesh, Gopika, Surya Ramachandran, and Abdul Jaleel. 2020b. “Metabolomics-Based Prospective Studies and Prediction of Type 2 Diabetes Mellitus Risks.” *Metabolic Syndrome and Related Disorders* 18(1):1–9.
- Shao, Mingmei, Hao Lu, Ming Yang, Yang Liu, Peihao Yin, Guowen Li, Yunman Wang, Lin Chen, Qingguang Chen, Cheng Zhao, Qun Lu, Tao Wu, and Guang Ji. 2020. “Serum and Urine Metabolomics Reveal Potential Biomarkers of T2DM Patients with Nephropathy.” *Annals of Translational Medicine*.
- Shetty, Priya. 2012. “Public Health: India’s Diabetes Time Bomb.” *Nature* 485(7398):S14–16.
- Stone, Kirsten P., Desiree Wanders, Manda Orgeron, Cory C. Cortez, and Thomas W. Gettys. 2014. “Mechanisms of Increased in Vivo Insulin Sensitivity by Dietary Methionine Restriction in Mice.” *Diabetes*.
- Sun, Liang, Liming Liang, Xianfu Gao, Huiping Zhang, Pang Yao, Yao Hu, Yiwei Ma, Feijie Wang, Qianlu Jin, Huaixing Li, Rongxia Li, Yong Liu, Frank B. Hu, Rong Zeng, Xu Lin, and Jiarui Wu. 2016. “Early Prediction of Developing Type 2 Diabetes by Plasma Acylcarnitines: A Population-Based Study.” *Diabetes Care*.
- Tabish, Syed Amin. 2007. “Is Diabetes Becoming the Biggest Epidemic of the Twenty-First Century?” *International Journal of Health Sciences*.
- Tai, E. S., M. L. S. Tan, R. D. Stevens, Y. L. Low, M. J. Muehlbauer, D. L. M. Goh, O. R. Ilkayeva, B. R. Wenner, J. R. Bain, J. J. M. Lee, S. C. Lim, C. M. Khoo, S. H. Shah, and C. B. Newgard. 2010. “Insulin Resistance Is Associated with a Metabolic Profile of Altered Protein Metabolism in Chinese and Asian-Indian Men.” *Diabetologia*.
- Vangipurapu, Jagadish, Alena Stancáková, Ulf Smith, Johanna Kuusisto, and Markku Laakso. 2019. “Nine Amino Acids Are Associated with Decreased Insulin Secretion and Elevated Glucose Levels in a 7.4-Year Follow-up Study of 5,181 Finnish Men.” *Diabetes*.
- Wang, Thomas J., Martin G. Larson, Ramachandran S. Vasan, Susan Cheng, Eugene P. Rhee, Elizabeth McCabe, Gregory D. Lewis, Caroline S. Fox, Paul F. Jacques, Céline Fernandez, Christopher J. O’Donnell, Stephen A. Carr, Vamsi K. Mootha, Jose C. Florez, Amanda Souza, Olle Melander, Clary B. Clish, and Robert E. Gerszten. 2011. “Metabolite Profiles and the Risk of Developing Diabetes.” *Nature Medicine*.

- Xu, Fengguo, Subramaniam Tavintharan, Chee Fang Sum, Kaing Woon, Su Chi Lim, and Choon Nam Ong. 2013. "Metabolic Signature Shift in Type 2 Diabetes Mellitus Revealed by Mass Spectrometry-Based Metabolomics." *Journal of Clinical Endocrinology and Metabolism*.
- Ying, Zhixiong, Hao Zhang, Weipeng Su, Le Zhou, Fei Wang, Yue Li, Lili Zhang, and Tian Wang. 2017. "Dietary Methionine Restriction Alleviates Hyperglycemia in Pigs with Intrauterine Growth Restriction by Enhancing Hepatic Protein Kinase B Signaling and Glycogen Synthesis." *Journal of Nutrition*.
- Yoshinari, Orié and Kiharu Igarashi. 2011. "Anti-Diabetic Effect of Pyroglutamic Acid in Type 2 Diabetic Goto-Kakizaki Rats and KK-A y Mice." *British Journal of Nutrition*.
- Yousf, Saleem, Devika M. Sardesai, Abraham B. Mathew, Rashmi Khandelwal, Jhankar D. Acharya, Shilpy Sharma, and Jeetender Chugh. 2019. "Metabolic Signatures Suggest O-Phosphocholine to UDP-N-Acetylglucosamine Ratio as a Potential Biomarker for High-Glucose and/or Palmitate Exposure in Pancreatic β -Cells." *Metabolomics* 15(4):55.
- Zaccardi, Francesco, David R. Webb, Thomas Yates, and Melanie J. Davies. 2016. "Pathophysiology of Type 1 and Type 2 Diabetes Mellitus: A 90-Year Perspective." *Postgraduate Medical Journal* 92(1084):63–69.
- Zhang, Aihua, Hui Sun, Guangli Yan, Ping Wang, and Xijun Wang. 2015. "Metabolomics for Biomarker Discovery: Moving to the Clinic." *BioMed Research International*.
- Zhao, Xue, Xiaokun Gang, Yujia Liu, Chenglin Sun, Qing Han, and Guixia Wang. 2016. "Using Metabolomic Profiles as Biomarkers for Insulin Resistance in Childhood Obesity: A Systematic Review." *Journal of Diabetes Research* 2016.

Chapter 5

*Mapping metabolic perturbations in *Mycobacterium smegmatis* in response to different stress conditions using NMR spectroscopy*

5.1 Introduction

Tuberculosis – an infectious disease caused by *Mycobacterium tuberculosis* – is responsible for ~2 million total deaths worldwide (Barberis, Bragazzi, Galluzzo, & Martini, 2017). *M. tuberculosis* belongs to the genus *Mycobacterium*, which consists of many non-pathogenic mycobacteria, including *M. smegmatis* (Falkinham, 2009; Yamada et al., 2018). These mycobacteria actively engage in maintaining carbon and nutrient balance in several ecosystems, including endangered coastal swamps and peat-rich wetlands (Cordero et al., 2019). The hydrophobicity of the mycobacterial cell surface permits the accumulation at air-water interfaces and enhances their persistence under various environmental conditions and interfaces, including aerosolized droplets (Falkinham, 2009). These diverse habitats constantly pose a spectrum of challenges and growth-limiting conditions to the inhabitant mycobacterial species, such as pH change, changes in oxygen exposure, and extreme nutrient deprivation. Therefore, for survival in such natural habitats, *M. smegmatis* should possess the ability to circumvent hostile fluctuations posed by the different environmental conditions. For example, secretory antioxidant enzymes, including superoxide dismutase, catalase, and peroxidase, in *M. smegmatis*, play a vital role in detoxifying exogenous oxidants. This clearly indicates that even in its natural habitat, *M. smegmatis* is ready to combat the vulnerabilities induced by reactive oxygen species (Ehrt & Schnappinger, 2009; Tyagi, Dharmaraja, Bhaskar, Chakrapani, & Singh, 2015).

Lately, *M. smegmatis* – though resides in a totally distinct habitat – has been recognized as a good surrogate for the pathogenic *M. tuberculosis* (Altaf, Miller, Bellows, & O’Toole, 2010; Barry, 2001), largely because of the following reasons: i) it is non-pathogenic in nature; ii) has a relatively faster growth rate; iii) shares many features with pathogenic mycobacteria; and iv) is easier to handle in laboratory set-ups. Both *M. smegmatis* and *M. tuberculosis* possess a thick mycolic acid-rich cell wall that helps them in modulating various host immune responses – a property which is very crucial for their survival. In addition, they both respond to anaerobiosis and substrate deprivation; possess an inducible pH-homeostasis system (to tackle the mild acidic conditions); and thus are capable of adaptation and hence, survival at otherwise lethal pH conditions (Anes et al., 2006; Drapal, Wheeler, & Fraser, 2016; O’Toole et al., 2003; M. Rao, Streur, Aldwell, & Cook, 2001; Vandal, Nathan, & Ehrt, 2009). In addition to its role in maintaining environmental homeostasis and as a model system to study pathogenic mycobacteria, *M. smegmatis* lately has been related to opportunistic infections in immune-compromised hosts (Pierre-Audigier et al., 1997).

Most of the studies that have been performed to date have employed omics approaches like genomics, transcriptomics, and proteomics to understand the adaptive physiology displayed by mycobacteria for their survival under different hostile conditions, with a particular focus on mapping regulatory changes at transcriptional or protein levels (Ganji, Dhali, Rizvi, Rapole, & Banerjee, 2016; Xinfeng Li et al., 2017; P. K. Rao & Li, 2009; Roxas & Li, 2009; Wang, Prince, & Marcotte, 2005). Metabolomics has been identified as the final frontier of the omics cascade. It can broadly be defined as the study of the metabolome in a given biological system at a specific point in time, provides extensive information about the metabolic alterations that occur during the process, and thus, provides insights into a particular biological manipulation (Zhao et al., 2017). Notably, the changes in the metabolome are amplified relative to changes in the transcriptome and the proteome, and thus, metabolomic biomarkers can serve as biomarkers for an early response to any environmental change. Therefore, metabolomics data coupled with proteomics and transcriptomics would help generate a system-wide view of the organism or the cell's response (Wanichthanarak, Fahrman, & Grapov, 2015).

Previous studies have revealed perturbation in metabolomic profiles of *M. smegmatis* under different conditions. For example, distinct metabolic signatures have been observed in mycobacteria at various stages of growth (Drupal, Perez-Fons, Wheeler, & Fraser, 2014); during hypoxia (Drupal et al., 2016); and in *M. smegmatis* treated with anti-mycobacterial agents, namely, rifampin, capreomycin (Man et al., 2018), and pretomanid (Baptista, Fazakerley, Beckmann, Baillie, & Mur, 2018). From these studies, one may infer that metabolic adaptation may be the key to survival and adaptation of mycobacteria when subjected to different stress conditions.

Taking leads from these observations, we aimed to employ a metabolomic approach to decipher the distinctive metabolic shifts in *M. smegmatis* (a surrogate for *M. tuberculosis*) associated with different stresses that have not been reported earlier. Among the various techniques used for metabolomics, nuclear magnetic resonance (NMR) spectroscopy has emerged as a versatile and powerful approach due to its non-invasiveness, robustness, fast, reproducibility, and high precision in metabolic profiling (Fan & Lane, 2016). To this end, an untargeted proton nuclear magnetic resonance (¹H-NMR) spectroscopy-based metabolomics approach was implemented to unravel the unique metabolic signatures in *M. smegmatis* subjected to acidic (pH 5.5) stress (lowering the pH of the medium); oxidative stress (addition of hydrogen peroxide); and nutrient starvation (growing mycobacteria in phosphate-buffered saline (PBS)) conditions to identify critical metabolites responsible for immediate adaptive responses. These three stresses are faced by *M. smegmatis*, both as environmental mycobacteria

and as an opportunistic agent inside the human phagocytic cells (Xiaojing Li, Wu, Han, Hu, & Mi, 2015; Pelosi et al., 2012; You, Xu, Yin, & Ye, 2019; Zahrt & Deretic, 2002). In the present investigation, a total of 22, 21, and 47 differential metabolites were identified in response to acidic, oxidative, and nutrient starvation conditions, respectively. Besides significant perturbation in energy metabolism and amino-acid metabolism pathways during different abiotic stress conditions, we also detected, for the first time, accumulation of organic osmolytes dimethylamine, methylamine, and betaine as early adaptive markers to stress associated with nutrient starvation and oxidative stress. These metabolic cues led us to identify possible novel pathways of biosynthesis of these secondary metabolites in *M. smegmatis* similar to those of obligate methylotrophs. The expression of the ORFs of these undocumented pathways was studied at transcript levels using RT-PCR.

5.2 Experimental Methods

5.2.1 Bacterial strains and experimental conditions.

The *M. smegmatis* mc²155 strain, used in the current study, was cultured using already established protocols (24) with few changes in the laboratory of Prof. Dr. Sharmistha Banerjee (University of Hyderabad). Briefly, *M. smegmatis* was grown at 37 °C with shaking at 180 rpm in 7H9 media (consisting 0.4% (vol/vol) glycerol and 0.05% (vol/vol) tyloxapol, and supplemented with 10% (vol/vol) oleic acid-albumin-dextrose-catalase (OADC)) till its optical density at 600 nm (OD₆₀₀) reached 0.5 - 0.7. The cultures were harvested and then washed with 1X phosphate-buffered saline (PBS). The cell pellet was then resuspended in media mimicking microbicidal stress conditions such as acidic stress (pH 5.5) in Sauton's minimal medium (Piddington, Kashkouli, & Buchmeier, 2000); oxidative stress (10 mM H₂O₂) in Sauton's minimal medium (Voskuil, Bartek, Visconti, & Schoolnik, 2011); and nutrient starvation stress in 1X PBS (Loebel, Shorr, & Richardson, 1933) for 4 hours. Each of the experimental conditions was replicated ten times to obtain a statistically significant number of samples. Each sample was checked for contamination using Ziehl-Neelsen (ZN) staining before and after four hours of stress, intracellular metabolites were extracted and subjected to untargeted metabolomics using nuclear magnetic resonance (NMR) spectroscopy.

5.2.2 Metabolite Extraction.

Intracellular metabolites from *M. smegmatis* cultures were extracted using a modified version of extraction previously described by Nicolas P. Tambellini et al (Tambellini,

Zaremborg, Turner, & Weljie, 2013). After 4 h of stress, OD₆₀₀ was measured, and *M. smegmatis* cultures were harvested at 5000 rpm (4°C) for 10 mins. The cell pellet was rapidly frozen in liquid nitrogen to quench any enzymatic or chemical reaction. Subsequently, the cells pellets were allowed to thaw on ice for 5 min and then were resuspended in 500 µl of pre-chilled methanol:chloroform (2:1), and homogenized with bead beating (using 0.1 mm zirconia beads for ten cycles with an interval of 1 min on ice) to lyse the cells. The supernatant was collected after centrifuging at 1000 rpm for 45 sec and then vortexed for 30 sec after the addition of 500 µl of distilled water and 500 µl of chloroform. The suspension was then centrifuged at 12,000 rpm for 30 min at 4°C to separate the phases. The upper aqueous phase and lower organic phase were carefully transferred into separate 1.5 ml microcentrifuge tubes, followed by centrifugation at 12,000 rpm for 2 mins at 4°C to remove any residual solvent from other phase. The upper aqueous phase samples were lyophilized and then transported to IISER-Pune on dry ice. At IISER Pune, the samples were reconstituted in 600 µl of 20 mM sodium phosphate NMR buffer, pH 7.4 in D₂O containing 400 µM DSS (2,2-dimethyl-2-silapentane-5-sulfonic acid). The samples were vortexed for 2 min at room temperature and centrifuged at 4000 g for 5 min. The supernatants were transferred to 5 mm NMR tubes for NMR analysis.

5.3 NMR acquisition and spectral processing

Water-suppression pulse sequence from Bruker library (noesygppr1d) was used to record all the ¹H NMR data at 298 K as described in detail in Chapter 2. For a given sample, a total of 64 transients were collected into 32K data points for each spectrum with a spectral width of 7200 Hz. The acquired spectra were manually corrected for phase and baseline distortions using Topspin (v3.5) software (www.bruker.com/bruker/topspin) and were referenced to the DSS resonance at 0.0 ppm. Additionally, ¹H-¹H total correlation spectroscopy (TOCSY) was performed for chemical shift assignments and verification.

5.4 Metabolite identification and quantification

All ¹H NMR spectra were imported into Chenomx NMR Suite 8.1 (Edmonton, AB, Canada) for identification and quantification of metabolites as described in detail in Chapter 2. All the identified metabolites were further cross checked from the biological magnetic resonance data bank (BMRB) (Ulrich et al., 2008) and human metabolome database (HMDB) (Wishart et al., 2007). In addition, two-dimensional ¹H-¹H TOCSY was used for further

metabolite confirmation via a semi-automated software – MetaboMiner (Xia, Bjorndahl, Tang, & Wishart, 2008).

The identified metabolites were then quantified using the profiler module of Chenomx software, which enables metabolites quantification relative to an internal standard of known concentration (400 μ M). The concentration data obtained after metabolite quantification were converted to comma-separated values (CSV) format using Microsoft excel format and imported into MetaboAnalyst 4.0 (Chong et al., 2018), a free web-based program for multivariate analysis.

5.5 Statistical analysis

Due to the high dimensionality and enormous complexity of NMR data (N = 10 for four experimental conditions), multivariate and univariate statistical analysis was carried out to analyze the effect of acidic, oxidative, and nutrient starvation stresses on *M. smegmatis*. PLS-DA were conducted using normalized concentration data of metabolites as input in the MetaboAnalyst 4.0 web tool (Chong et al., 2018) to demonstrate the metabolite patterns and selection of important features. The raw metabolomics data sets were subjected to Pareto-scaling prior to chemometric analysis. VIP scores were generated from PLS-DA analysis to identify the discriminatory metabolites responsible for the differential clustering of score plots in the PLS-DA model. VIP score measures the contribution of a variable to the PLS-DA model and identifies the important metabolites that vary significantly between the groups and better describes the intergroup variation. Further, pairwise analysis between control and different stresses were carried out to identify significantly altered compounds. Thus, volcano plot analysis was performed to identify important metabolites based on fold change analysis and t-test. Metabolites with fold change >1.5 and FDR adjusted p-value < 0.05 were considered to be significantly different (Yousf et al., 2019).

5.6 Metabolic pathway analysis

Metabolic pathway analysis of all the discriminating metabolites was conducted to understand the biological significance of the metabolic changes and to identify the most important pathways involved in *M. smegmatis* when subjected to various stresses conditions studied in this work. Metabolic Pathway Analysis (MetPa) was carried out on a free web-based program Metaboanalyst 4.0 (Chong et al., 2018; Xia, Wishart, & Valencia, 2011). The

metabolic pathways with impact value > 0.1 and $p < 0.05$ were considered as significantly perturbed pathways.

5.7 RNA isolation and semi-quantitative RT-PCR

RNA isolation was carried out using Trizol method with certain modification. After growing the cells under respective stress conditions, the samples were harvested. The collected pellet was snap frozen in liquid nitrogen and stored at -80°C until RNA extraction. The bacterial pellet was resuspended in Trizol reagent (Invitrogen, CA, USA) along with 0.1 mm glass beads and lysed by bead beating - pulse on: 1 min and pulse off: 2 min on ice. After lysis, glycogen was added to a final concentration of $200\ \mu\text{g}/\text{mL}$ and incubated at room temperature (RT) for 10 min. The samples were vortexed vigorously after adding chloroform and incubate at RT for a further 10 min. The samples were centrifuged at 10,000 rpm and 4°C for 20 min and the upper aqueous layer was collected and transferred to a new tube. The RNA was precipitated using isopropanol in presence of glycogen ($200\ \mu\text{g}/\text{mL}$). The obtained pellet was washed with 75% ethanol, air-dried at RT and resuspended in RNase-free water (Qiagen, Hilden, Germany). Prior to reverse transcription the RNA was subjected to DNase treatment to eliminate any residual DNA contamination. The DNase treated RNA was reverse transcribed with random hexamers as a primer to synthesize cDNA using Superscript III Reverse Transcriptase (Invitrogen). The reverse transcribed RNA was used for Real-time PCR. RT-PCR was carried out with 1:10 diluted cDNA for 30 cycles with respective primers. The details of the primers are provided in Table 5.4. The PCR cycling parameters were (i) initial denaturation 95°C for 3 min, (ii) 30 amplification cycles (95°C for 15 s, annealing temperature [as given in Appendix Table 5.4] for 20 s, and 72°C for 20 s), and (iii) final extension at 72°C for 10 min. A 1.5% agarose gel was used to visualize the RT-PCR product.

5.8 Identification of putative pathway using in silico tools

In silico analysis of the *M. smegmatis* mc²155 strain genome was carried out to identify putative pathway in Mycobacterium smegmatis. The primary amino acid sequence of 2,4-dienoyl-coA reductase [*M. smegmatis* mc²155 strain] MSMEG_5124 was obtained from Mycobrowser (Kapopoulou, Lew, & Cole, 2011; Lew, Kapopoulou, Jones, & Cole, 2011) and subjected to domain analysis tools to predict conserved domains in the 2,4-dienoyl-coA reductase [*M. smegmatis* mc²155 strain] MSMEG_5124. These include MOTIF (<http://www.genome.jp/tools/motif/>) and SMART (<http://smart.embl-heidelberg.de/>). Protein

sequence similarity search was carried out using PSI-BLAST (<https://www.ebi.ac.uk/Tools/sss/psiblast/>). In silico analysis and transcriptomic experiments reported in this chapter were performed by Dr. Arshad in Dr. Sharmistha Banerjee's lab at University of Hyderabad.

5.9 Results

5.9.1 Global metabolome profiling and metabolic alterations of *M. smegmatis*.

Global metabolic profiling of *M. smegmatis* (MC²155) grown in Sauton minimal media was performed and compared with *M. smegmatis* subjected to three stress conditions used in the study - acidic stress (pH 5.5), oxidative stress (10 mM H₂O₂) in Sauton minimal media, and nutrient starvation stress in PBS. A total of 56 abundant metabolites were identified from the methanolic extracts using 1D ¹H-NMR (Figure 5.1). Six NMR peaks remain unassigned and have been annotated by an asterisk (*) in Figure 5.1. The resonance assignment of these metabolites were further confirmed using 2D ¹H-¹H TOCSY NMR spectroscopy (Figure 5.2). Five metabolites (GTP, CDP, tryptophan, fructose-1-6 biphosphate, fumarate) were excluded as their representative peaks were not evident in all spectra. The organic phase containing the lipids/steroids/fatty acids and other water-insoluble compounds gave very broad signals in ¹H NMR and, thus, were excluded from the analysis. All the 56 metabolites and their respective ¹H chemical shifts (in reference to DSS) have been listed in Appendix Table 5.2.

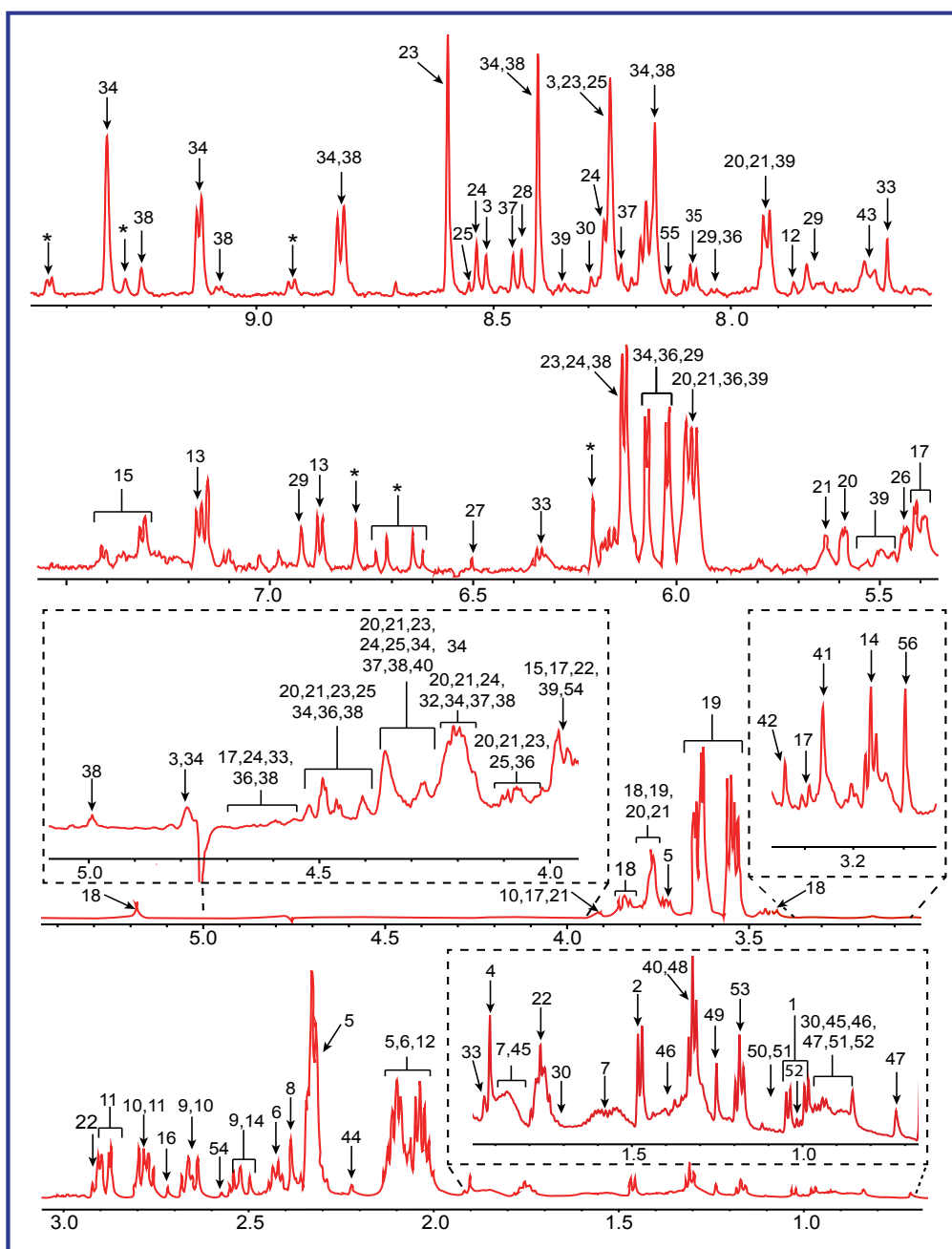


Figure 5.1: Representative $^1\text{H-NMR}$ spectrum of the methanolic extract of *M. smegmatis* grown in Sauton media. The representation of the different metabolites identified from the $^1\text{H NMR}$ spectrum of *M. smegmatis*. Key: 1, Valine; 2, Alanine; 3, ATP; 4, Acetate; 5, Glutamate; 6, Glutamine; 7, Citrulline; 8, Succinate; 9, Citrate; 10, Aspartate; 11, Asparagine; 12, Homoserine; 13, Tyrosine; 14, beta-alanine; 15, Phenylalanine; 16, Dimethylamine; 17, Maltose; 18, Trehalose ; 19, Glycerol; 20, UDP-glucose; 21, UDP-galactose; 22, DSS; 23, AMP; 24, ADP; 25, IMP; 26, Glucose-1- phosphate; 27, fumarate ; 28, Formate; 29, CDP; 30, Leucine; 31, Lysine; 32, Fructose 1-6, biphosphate; 33, dTTP; 34, NAD^+ ; 35, N-acetyl glucosamine; 36, UMP; 37, NADPH; 38, NADP^+ ; 39, UDP-N- acetylglucosamine; 40, Threonine; 41, Betaine; 42, Methanol; 43, Tryptophan; 44, Acetone; 45, 2- aminobutyrate; 46, Caprate; 47, Cholate; 48, Lactate 49, 3-hydroxyisovalerate; 50, 3-methyl-2- oxovalerate; 51, 2-hydroxy-3-methylvalerate; 52, Isoleucine; 53, Ethanol; 54, Methylamine; 55, GTP; and 56, Malonate.

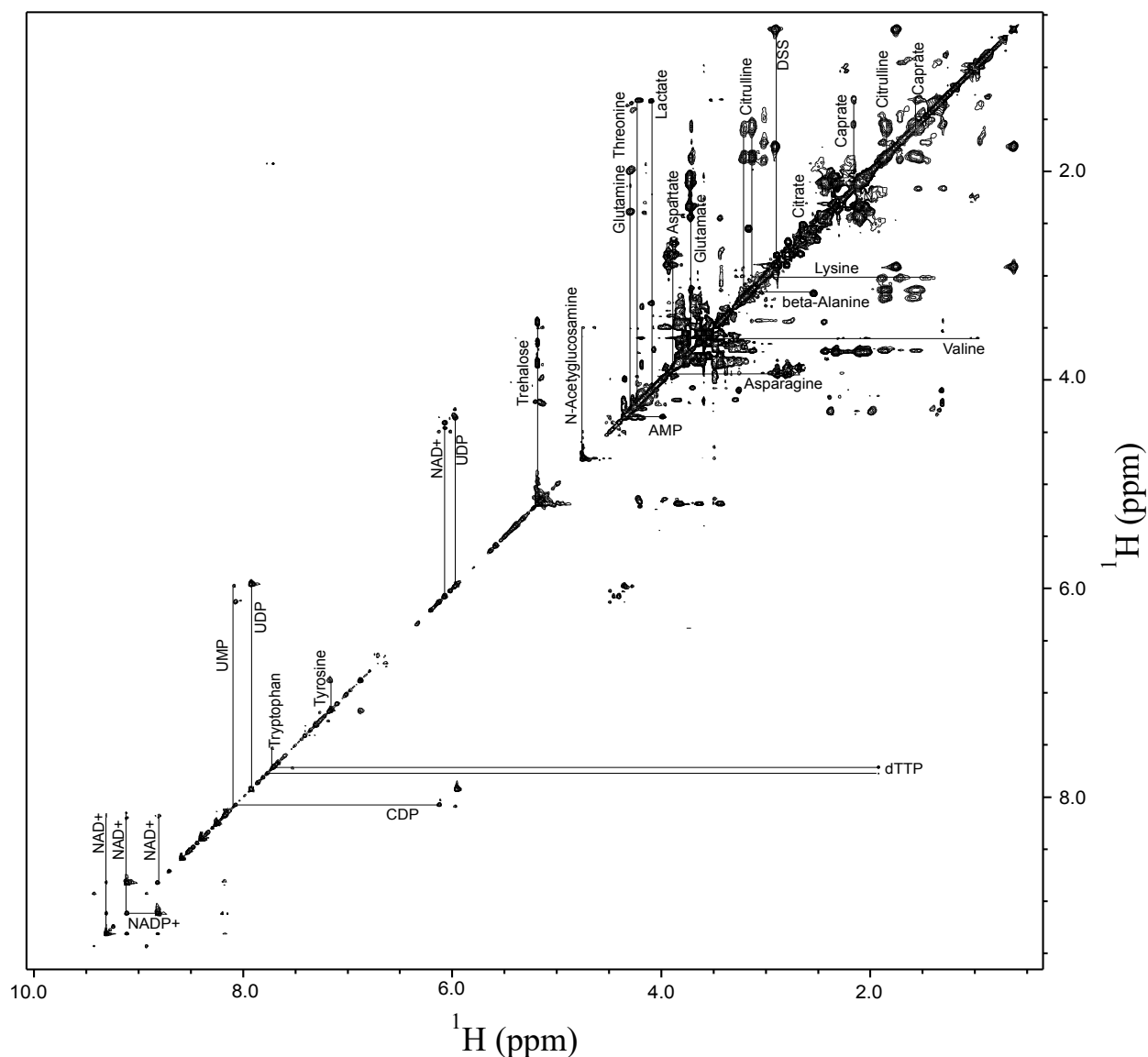


Figure 5.2: Two-dimensional (2D) ^1H - ^1H TOCSY correlation spectrum of the methanolic extract of *M. smegmatis* grown in Sauton media. Cross-peaks in the TOCSY spectrum were used to re-confirm the resonance assignments shown in Figure 5.1 and listed in Appendix Table 5.2.

PLS-DA analysis was performed to discern the differences in metabolic profiles in *M. smegmatis* when exposed to various stresses. The PLS-DA analysis between control and the three stress conditions used in the study showed distinct clustering of groups on their respective score plots with 59.4 percent variance in PC1 and 16.9 percent variance in PC2 for acidic stress (Figure 5.3A), 43.1 percent variance in PC1 and 6.7 percent variance in PC2 for oxidative stress (Figure 5.3B), and 70.6 percent variance in PC1 and 12.8 percent variance in PC2 for starvation stress (Figure 5.3C), respectively.

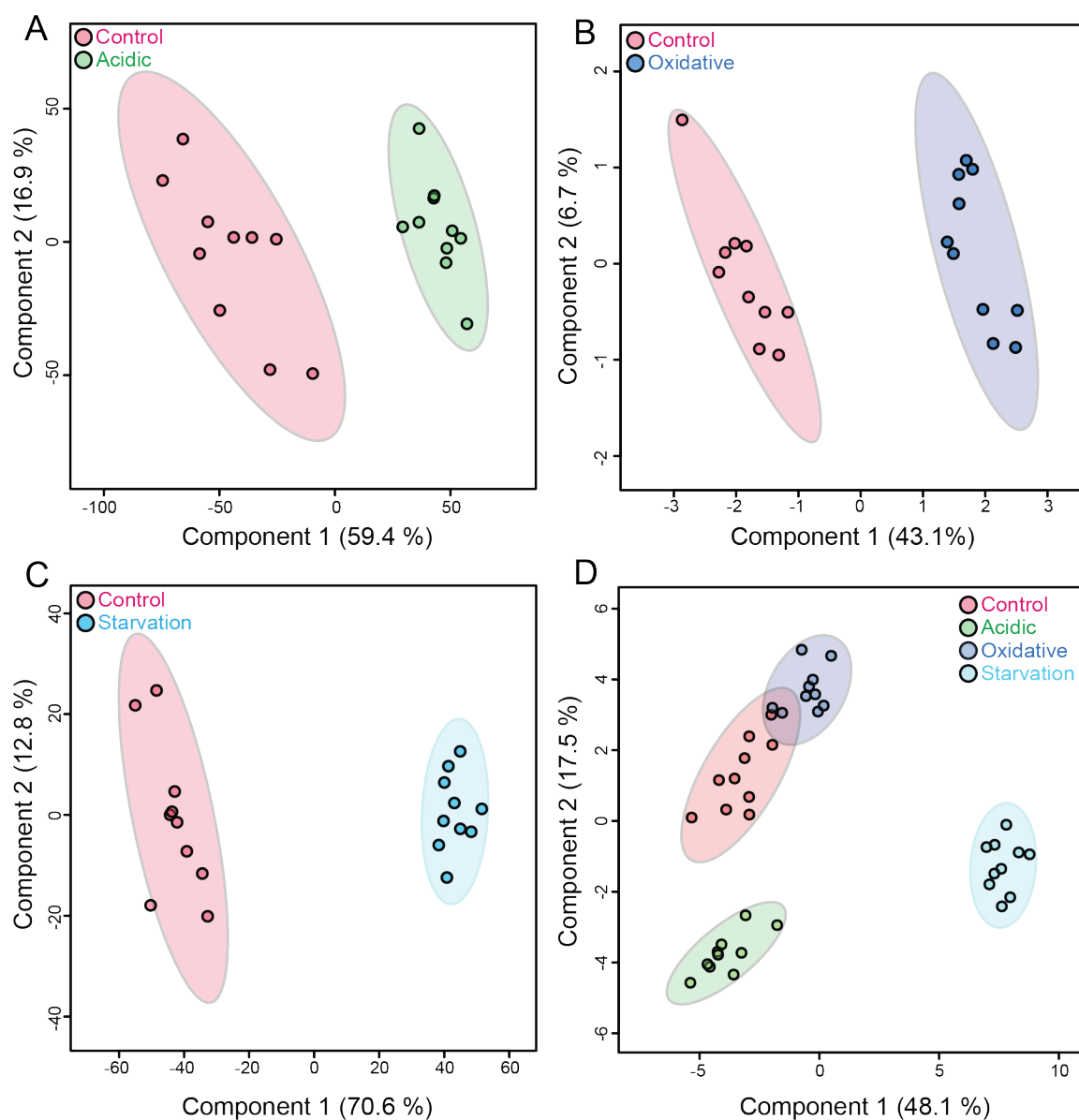


Figure 5.3: PLS-DA score plots for *M. smegmatis* samples grown in different stress conditions showing model discrimination between (A) acidic stress versus control; (B) oxidative stress versus control; (C) nutrient starvation versus control; (D) Acidic, oxidative, and nutrient starvation versus the control. Ten replicates were used for each growth condition. Ellipses showing 95% confidence limits of a normal distribution for each group of the samples have been marked in respective colors, as mentioned above. The dots inside all the plots correspond to biological replicates under each category.

PLS-DA model analysis demonstrates the maximum segregation between control and nutrient-deprived condition, while oxidative stress showed the least segregation from the control. This group separation was based on first two components, which displays a distinct demarcation between the control and stress groups. VIP scores were obtained from PLS-DA model to identify the discriminatory metabolites that contributed most to the group separation in the PLS-DA models for each stress condition: acidic vs control (Table 5.1), oxidative vs control (Table 5.1), and nutrient starvation vs control (Table 5.1). The VIP score which is based

on PLS loadings reflects the influence of each variable to the overall model, with metabolites VIP scores > 1 considered important in classification. The discriminatory metabolites were organised in the descending order of the VIP score in component 1. Based on VIP score (VIP > 1), the key metabolites discriminating control from acidic stress were glutamic acid, homoserine, asparagine, aspartic acid, glutamine, β -alanine, threonine, alanine, and trehalose. While aspartic acid, glutamic acid, lysine, NADP, betaine, AMP, methylamine, NAD, UDP-N-acetylglucosamine (UDP-GlcNAc), 1-methylnicotinamide, capric acid, maltose, ATP, 2-hydroxy-3-methylvalerate (HMVA), and betaine were the most important metabolites that altered significantly during oxidative stress. Similarly, during the nutrient-deprived condition, the discriminatory metabolites (VIP >1) identified were asparagine, glutamic acid, aspartic acid, trehalose, glutamine, citric acid, malonic acid, β -alanine, ethanol, homoserine, and HMVA. We observed most of the discriminating metabolites majorly fall in amino acid metabolism and energy metabolism pathways. Common discriminatory metabolites distinguishing control, acidic stress, oxidative stress, and nutrient starvation included AMP, asparagine, NADP, NAD, lysine, tyrosine, citric acid, ADP, maltose, UDP, acetone, UDP-galactose, 2-Hydroxy-3-methylpentanoic acid (HMVA), sucrose, betaine, leucine, glutamate, UMP, formate, glucose-1-phosphate, 3-hydroxyisovaleric acid, ATP, acetone, and inosinic acid (Table 5.1).

Univariate analysis was applied to identify differential metabolites significantly altered in *M. smegmatis* grown under different experimental conditions. Volcano plot analysis (Figure 5.4) showed differential regulation of 22, 21, and 47 metabolites in case of acidic stress (Figure 5.4A), oxidative stress (Figure 5.4B), and starvation stress (Figure 5.4C) in a pairwise comparison with the control experiment with a cut-off of fold change (FC >1.5), p-value ($P < 0.05$) and false discovery rate (FDR < 0.05). Red and green dots indicate metabolites that are significantly upregulated and downregulated, respectively, in control compared to other stresses used in our study. The fold change, p-value and FDR value for all these differential metabolites under different stress conditions have been shown in Appendix Table 5.3 (a, b, and c for acidic, oxidative, and nutrient starvation stresses, respectively).

Further, box and whisker plots were generated for visual interpretation of discriminating metabolites identified from the PLS-DA and the VIP score plot for all the stress conditions used in our study. After data normalization and scaling, the discriminatory metabolites were selected individually, and the relative concentrations of each of these were plotted along y-axis against different experimental growth condition; namely, Control (green);

Acidic (red), Oxidative (blue), and Nutrient deprivation (represented as Starvation, cyan), have been depicted in Figure 5.5.

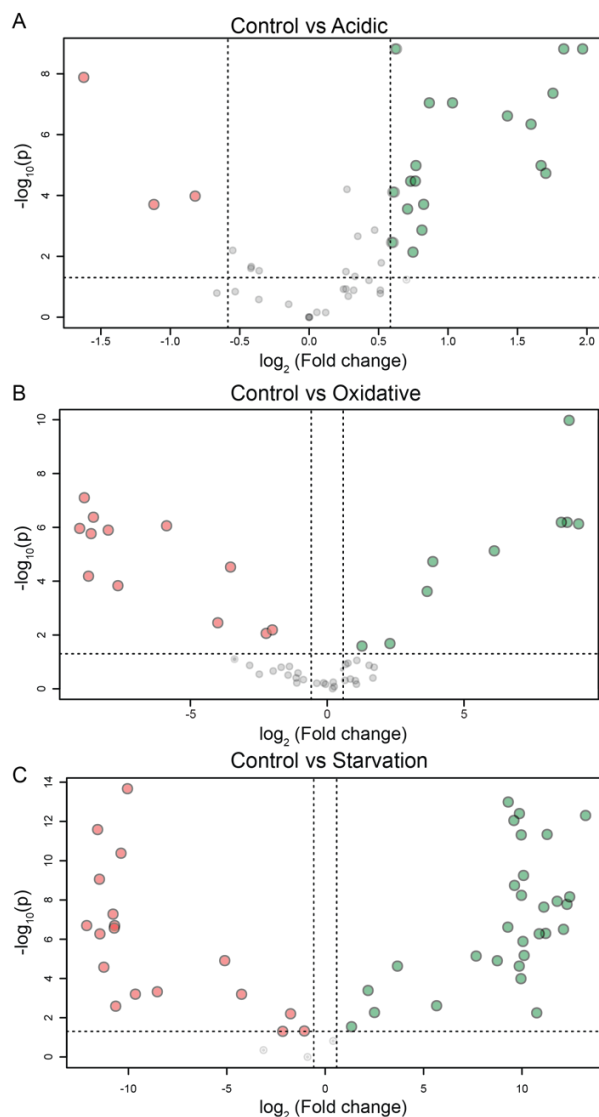


Figure 5.4: Volcano plot, where a dotted horizontal line corresponded to FDR-correct p-value 0.05 on a log₁₀ scale and dotted vertical lines represent a 1.5-fold change in concentration of metabolites on a log₂ scale. Red and green dots indicate metabolites that are significantly upregulated and downregulated, respectively, in the control experiment when compared to A) acidic, B) oxidative, and C) nutrient starvation stresses, respectively. Overall differentially regulated metabolites were 22, 21, and 47 for acidic, oxidative, and nutrient starvation stresses, respectively, when as compared to normal growth conditions with a cut-off of fold change (FC > 1.5), p-value (P < 0.05), and false discovery rate (FDR < 0.05). These metabolites have also been listed in Appendix Table 5.3 (a, b, and c for acidic, oxidative, and nutrient starvation stresses, respectively).

Table 5.1: VIP value obtained from PLS-DA models for Control vs Acidic stress, Control vs Oxidative stress, Control vs Starvation stress, and Control vs all stresses (Acidic, Oxidative and Starvation).

Metabolites	VIP (Control/Acidic)	VIP (Control/Oxidative)	VIP (Control/Starvation)	VIP (Control/all)
1-Methylnicotinamide	0.42	1.38	0.40	0.93
HMVA	0.25	1.21	1.05	1.20
3-Hydroxyisovaleric acid	0.00	0.34	0.40	1.04
Acetic acid	0.41	0.80	0.48	1.01
Acetone	0.25	0.30	0.50	1.21
AMP	0.55	1.61	0.93	1.52
ATP	0.20	1.34	0.35	1.04
ADP	0.15	0.51	0.31	1.30
Alanine	1.19	0.42	0.83	0.23
β -Alanine	1.34	0.09	1.32	0.54
Betaine	0.04	1.68	0.40	1.17
Capric acid	0.47	1.38	0.62	0.85
Citric acid	0.63	0.30	1.65	1.34
Glutamic acid	4.71	2.13	2.27	1.16
D-Maltose	0.65	1.36	0.35	1.30
Ethanol	0.01	0.20	1.14	0.98
Formic acid	0.28	0.29	0.29	1.08
Glucose 1-phosphate	0.25	0.55	0.41	1.07
Inosinic acid	0.12	0.60	0.15	1.01
Asparagine	1.78	0.60	3.60	1.49
Aspartic acid	1.73	2.67	2.06	0.08
Glutamine	1.73	0.94	1.87	0.68
Homoserine	2.61	0.47	1.06	0.94
Leucine	0.28	0.24	0.84	1.17
Lysine	0.29	1.94	0.62	1.35
Threonine	1.25	0.01	0.61	0.69
Tyrosine	0.54	0.40	0.66	1.35
Malonic acid	0.73	1.07	1.65	0.74
Methylamine	0.22	1.55	0.48	0.71
NAD	0.39	1.46	0.70	1.39
NADP	0.03	1.69	0.23	1.39
Sucrose	0.17	0.34	0.36	1.18
Trehalose	1.15	0.28	2.02	0.99
UMP	0.30	0.50	0.38	1.09
UDP glucose	0.11	0.41	0.40	1.27
UDP-GlcNAc	0.13	1.39	0.29	0.34
UDP galactose	0.20	0.29	0.29	1.21

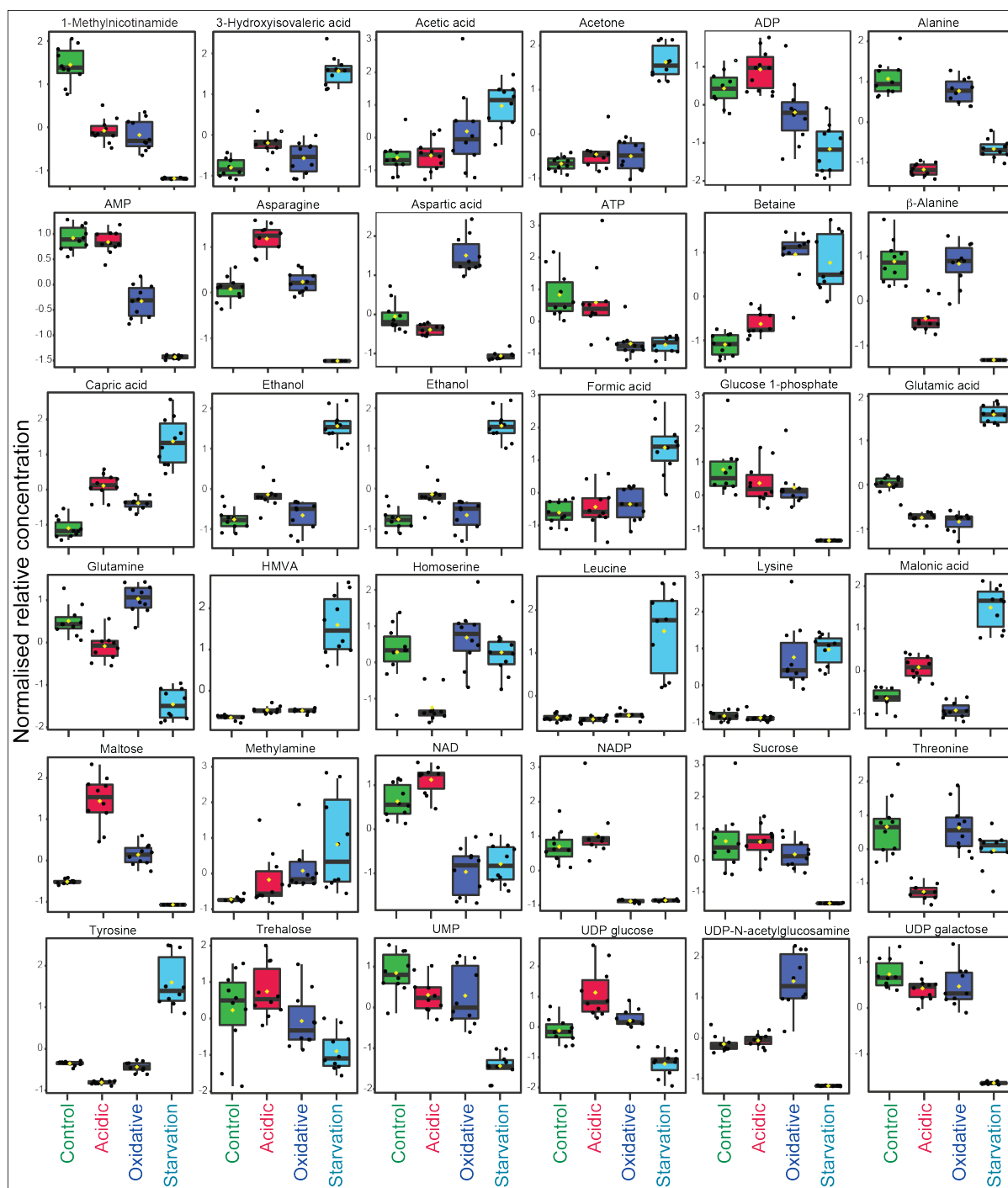


Figure 5.5: Box and whisker plots of relative concentrations for significantly differential metabolites identified from the VIP score plot (VIP >1) in *M. smegmatis* grown under different stress conditions. Data were normalized and scaled in the Y-axes. The concentrations have been represented as relative units. Dysregulated metabolites were selected individually, and the relative concentrations of each of these were plotted against different experimental growth conditions; namely, Control (green); Acidic (red), Oxidative (blue), and nutrient deprivation (represented as Starvation, cyan), respectively.

5.10 Pathway analysis

To identify metabolic pathways that showed the most significant perturbation in *M. smegmatis* cultured in the different experimental conditions used in the study and to assess the biological relevance of the changes in metabolite levels, metabolic pathway analysis of significantly altered metabolites (VIP >1) was carried out for pairwise comparison using MetaboAnalyst. Of all the perturbed pathways, those with an impact value > 0.1 and $p < 0.05$ were identified as significantly perturbed pathways under respective stresses. Results from the pathway analysis have been depicted in Figure 5.6, where each node represents a unique metabolic pathway. The color and radius of the node are based on its p-value and pathway impact values, respectively.

The major pathways perturbed significantly in the pairwise comparison of control versus acidic stress were alanine, aspartate and glutamate metabolism, beta-alanine metabolism, glutamine-glutamate metabolism and glycine-serine-threonine metabolism. These pathways have an important role in nitrogen assimilation and storage. It is reported in the literature that nitrogen assimilation helps the mycobacteria in abating acidic stress by releasing ammonia (Borah et al., 2019; Gouzy et al., 2014). In the control versus oxidative stress, pathways perturbed with significant scores were nicotinate-nicotinamide metabolism, glutamine-glutamate metabolism and lysine biosynthesis. The pathway nicotinate-nicotinamide metabolism have an important role in maintaining homeostasis for redox cofactor in oxidative stress (Nambi et al., 2015).

Similarly, under the nutrient starvation condition, the major pathways perturbed were β -alanine metabolism, alanine, aspartate and glutamate metabolism and glutamine and glutamate metabolism. All the pathways were constructed using KEGG metabolic network and the pictorial representation of these metabolic pathways affecting during various stress conditions have been shown in Figure 5.7. The significantly differential metabolites perturbed under respective stresses against control have been marked with arrows wherein an upward arrow indicates upregulated metabolites, and the downward arrow indicates down-regulated metabolites. Overall the significantly perturbed pathways in our study belongs to amino acid metabolism, nucleotide metabolism and central carbon metabolism.

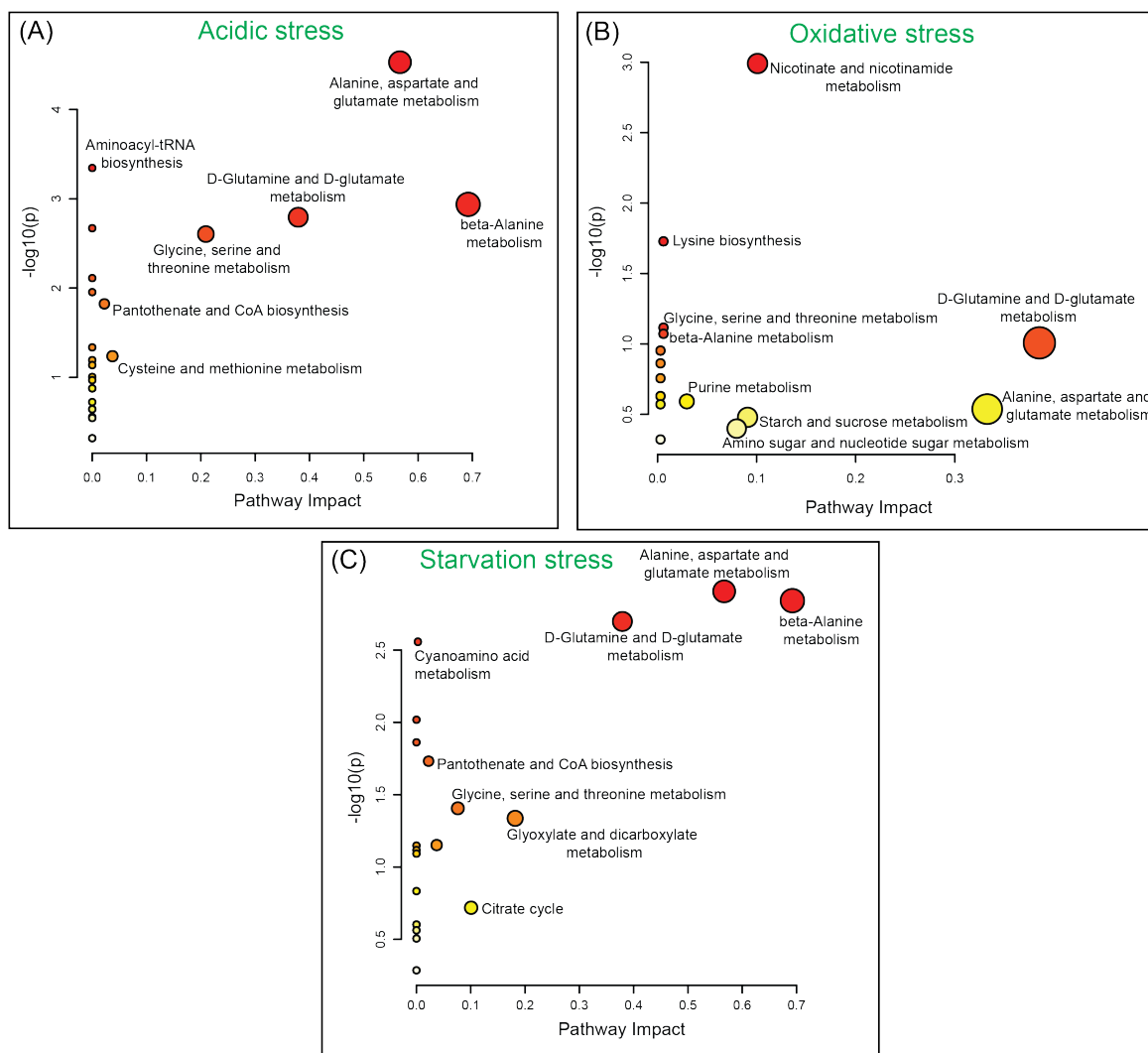


Figure 5.6: Metabolic Pathway Analysis (MetPa) of significantly altered metabolites under (A) acidic, (B) oxidative, and (C) Starvation stress. Each node represents a single metabolic pathway, with node color corresponding to the $-\log_{10}(P)$ value (yellow: higher p-values and red: lower p-values) and node size corresponding to the pathway impact score.

5.11 Discussion

Nitrogen metabolism is an essential process in all mycobacteria for their survival under different hostile conditions. *M. tuberculosis* utilizes multiple amino acids as nitrogen sources in human macrophages (Borah et al., 2019). In our study, the amino acids that are significantly dysregulated under the different stress conditions include asparagine, aspartate, glutamine, glutamate, citrate, alanine, lysine, and threonine. Asparagine and also aspartate are an important nitrogen source for mycobacteria inside the host cells. Besides playing an important role in nitrogen assimilation and determining *M. tuberculosis* virulence, asparagine also mediates resistance to acid stress during infection (Gouzy et al., 2014). From our NMR data, asparagine was found to be the only detected amino acid whose concentration was elevated

during early adaptation to acidic stress (Figure 5.5). One of the suggested mechanisms is that it gets assimilated by enzyme asparaginase, which hydrolyzes this amino acid into aspartate and ammonia, thereby alkalizing the acidic environment. Aspartate was noticed to be significantly dysregulated amino acid in all stress conditions. It is known that aspartate is the primary nitrogen source for *M. tuberculosis* for host colonization and is the precursor for many amino acids, including β -alanine, lysine, threonine, and isoleucine (Gouzy et al., 2013).

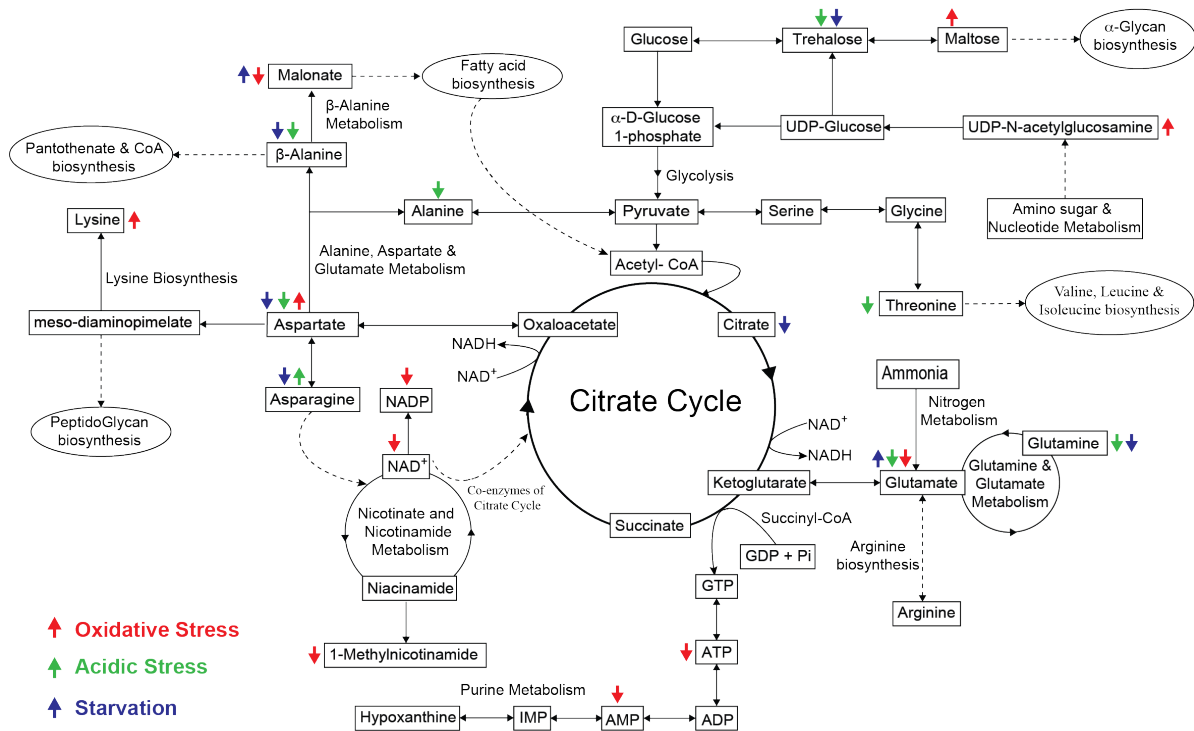


Figure 5.7: Pictorial depiction of the significantly dysregulated metabolites and Metabolomic Pathway Analysis (MetPA) construction of the metabolic pathways identified in *M. smegmatis*. The description of the arrows has been provided as a legend wherein Red arrows represent oxidative stress, Green arrows represent acidic and blue arrows represent starvation, respectively. The upward and downward direction of arrows indicates that the metabolites are upregulated or down-regulated during respective stresses compared to the experimental control condition.

It is important to mention here that the glutamine and glutamate metabolism pathway was perturbed significantly under all three abiotic stresses. Both glutamate and glutamine are central molecules in nitrogen metabolism. Glutamine, a predominant source of nitrogen for the synthesis of nitrogen bases, number of amino acids, amino sugars, is responsible for 85% of nitrogenous compounds in a cell. In addition to nitrogen donor, glutamine can also be used as an additional source of carbon to fuel the Krebs cycle, by a process called glutaminolysis (Koeken et al., 2019). In this process, glutamine is metabolized to glutamate via glutaminase, and the enzyme glutamate dehydrogenase (GDH) facilitates the conversion of glutamate to α -

ketoglutarate, which is a substrate for the Krebs cycle. Catriona et al. reported the transcription regulation of genes encoding for the glutamate dehydrogenase enzyme in *M. smegmatis* (Harper, Hayward, Kidd, Wiid, & van Helden, 2010). The production of ammonia, ATP, and NADH during glutaminolysis may serve as an early metabolic adaptation during stresses and thereby helping mycobacteria to adapt to acidic, oxidative, and nutrient starvation stresses.

β -alanine, a key precursor for the biosynthesis of pantothenate, was identified as a common dysregulated metabolite in acidic and starvation stresses with significantly reduced levels. Pantothenate (or vitamin B5) is the primary precursor for the biosynthesis of Coenzyme A, which is vital for the production of fatty acids and peptidoglycan (Leonardi & Jackowski, 2007). It was also seen that the levels of malonate were upregulated during nutrient starvation, thereby suggesting the utilization of β -alanine to malonate, which can be converted to malonyl-CoA and enter fatty acid metabolism. Altered levels of lysine, a precursor for peptidoglycan was also detected during oxidative stress (Pavelka & Jacobs, 1996). We also observed the drastic reduction of citrate levels in starvation condition, that infers the slowed down of the TCA cycle in mycobacteria under nutrient starvation stress. Furthermore, accumulation of acetone was seen during starvation stress, which depicts *M. smegmatis* are utilizing acetone as a source of carbon and energy, as previously reported by Furuya et al. (Furuya, Nakao, & Kino, 2015). Low levels of ATP production during nutrient starvation were apparent and expected, resulting in low levels of ADP and AMP accumulation.

A unique set of the metabolites associated with the GlgE pathway, such as s, D-maltose, UDP-glucose, Glucose-1-phosphate, etc. to be remarkably differential under different stress conditions. The GlgE pathway, presented schematically in Figure 5.8A, is linked to the biosynthesis of capsular α -glycan and is hypothesized to play a key role in *M. tuberculosis* virulence and immune responses (Koliwer-Brandl et al., 2016). The expression of genes involved in the GlgE pathway at transcription levels was evaluated using RT-PCR. It was observed that the expression level for maltose transferase (GlgE) was upregulated, and trehalase (TreH) was downregulated under oxidative stress when compared with others (Figure 5.8B). It may be concluded that this pathway is directed towards biosynthesis of capsular α -glycan for early adaptation against various stresses. Furthermore, the levels of disaccharide trehalose, which is abundantly present in *M. smegmatis*, sucrose, and maltose, were very low during nutrient deprivation conditions, thereby implying the utilization of these metabolites during starvation stress. It was also apparent from the TreH expression level (Figure 5.8C) that

was slightly higher during starvation stress, suggesting the utilization of trehalose to glucose as bacteria are deprived of nutrients.

Moreover, we also observed significantly altered levels of organic osmolytes, such as methylamine and betaine, in *M. smegmatis* cultures exposed to various stresses. However, the metabolic pathway analysis did not reflect any pathway associated with these osmolytes. Therefore we assumed that this might have been due to the absence of an annotated pathway of methylated amines in *M. smegmatis*.

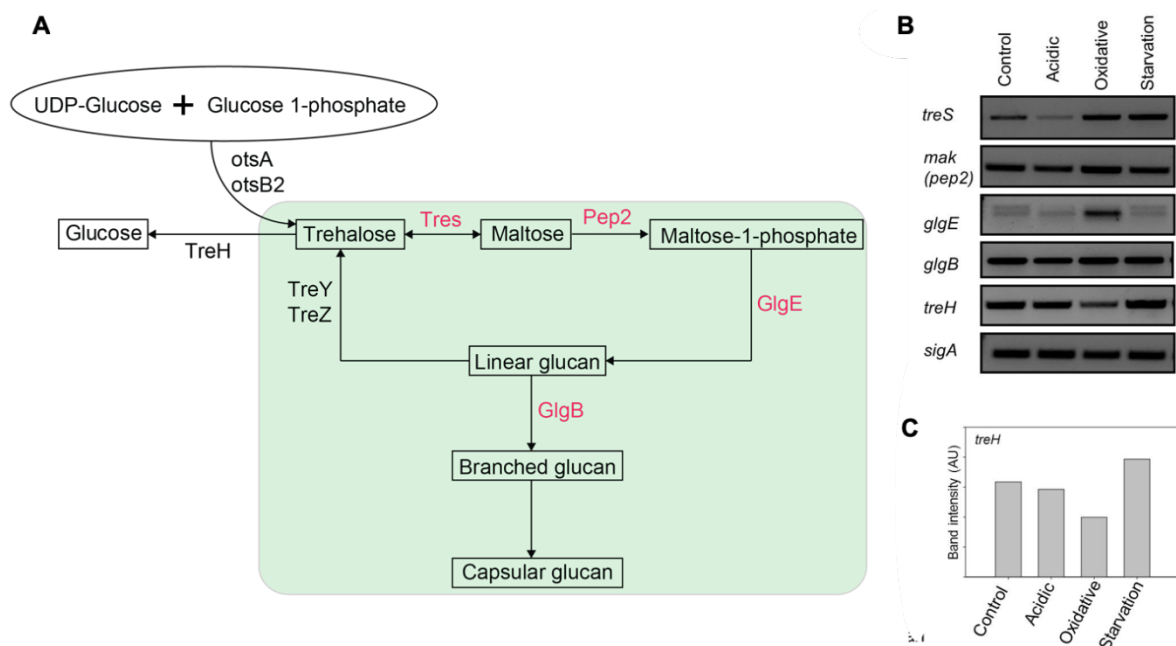


Figure 5.8: Metabolic pathways associated with biosynthesis of capsular α -glucan via GlgE pathway: A) GlgE pathway for the biosynthesis of α -glucan. B) RT-PCR of genes taking part in α -glucan pathway and their possible regulation at transcription levels. C) Densitometric analysis of TreH. Note: Picture B and C were provided by Dr. Sharmistha Banerjee' Lab (University of Hyderabad).

5.11.1 Putative pathway of biosynthesis of methylated amines in *M. smegmatis*.

As described above, our metabolomics data showed altered levels of the osmolytes such as betaine, methylamine, and dimethylamine that are known to play an important role in adaptation to various stresses, besides protecting osmotic imbalances (Burg & Ferraris, 2008; Holmström et al., 1994; Whatmore, Chudek, & Reed, 1990; Paul H. Yancey, Blake, & Conley, 2002). Price et al. reported that *M. tuberculosis* acquires betaine from host macrophages for maintaining osmotic balance (Price, Bukka, Cynamon, & Graham, 2008). Further, tracing the differential levels of these osmolytes through computational search tools, gene expression studies (using reverse transcription-PCR [RT-PCR]) of all the hypothetical ORFs from *M. smegmatis* and the presence of intermediate metabolites, we propose the existence of a putative

pathway of biosynthesis of betaine, methylamine, and dimethylamine previously unreported in *Mycobacterium smegmatis* (Figure 5.9A).

In methylotrophic bacteria, methylamine is formed from dimethylamine using dimethylamine dehydrogenase (DMD), and trimethylamine produces dimethylamine in the presence of trimethylamine dehydrogenase (TMD). In some bacteria, both processes are regulated by single dehydrogenase, which shares similar physical, chemical, and kinetic properties (McIntire, 1990). However, an open reading frame (ORF) or a gene by this name could not be located within the annotated *M. smegmatis* genome (<https://mycobrowser.epfl.ch/>). Using the sequence of trimethylamine dehydrogenase (TMD) of methylotroph *Methylophilus methylotrophs*, a protein sequence similarity search was conducted, and it was observed MSMEG_5124, annotated as 2,4-dienoyl-coA reductase (DCR) shows sequence identity of 26% with the TMD of *Methylophilus methylotrophs*, and hence suggests that there is a possibility that MSMEG_5124 may also function as TMD for *M. smegmatis*. With the confirmation of expression of MSMEG_5124, a possible TMD in *M. smegmatis* using RT-PCR (Figure 5.9B), and the presence of intermediate metabolites, dimethylamine, and methylamine from the metabolomics data, we were able to propose the company of the pathway converting trimethylamine to dimethylamine and methylamine.

Reported in many bacteria, also those associated with gut microbiota, carnitine monooxygenase reductase subunit (YeaX) functions as a complex with an oxygenase component [Rieske (2Fe-2S) (YeaW) region] of *E. coli* K-12 and can use carnitine, γ -butyrobetain, choline, and betaine as substrates to produce trimethylamine (Koeth et al., 2014). Using protein sequence similarity search with these *E. coli* genes using PSI-BLAST, we could identify two orthologs, MSMEG_4371, and MSMEG_0657, with similarity to YeaX (38% identity) and YeaW (37% identity), respectively. From our data, we have seen the accumulation of betaine during oxidative and nutrient-deprived conditions. With the confirmation of expression of MSMEG_4371 and MSMEG_0657 (Figure 5.9B) in *M. smegmatis* during different growth conditions and the presence of betaine from the metabolomics data, it could be inferred that betaine is the possible intracellular source for the synthesis of trimethylamine, catalyzed by MSMEG_4371 and MSMEG_0657.

With the evidence of expression of all these hypothetical ORFs from the *M. smegmatis* genome and the presence of intermediate metabolites that are catalyzed by these ORFs, we propose the existence of a putative pathway of biosynthesis of methylamines in *M. smegmatis*. It is reported that methylamines are used to synthesize formaldehyde and ammonia, which can be an advantage to pathogenic mycobacteria, particularly during acidic stress (Kim, Bae, &

Lee, 2001). Several *in vitro* biochemical studies have shown that methylamines can provide stability to proteins, a property by which they may function as osmolytes, giving protection from osmotic stress (P. H. Yancey & Somero, 1979). One would expect that a similar feature of methylamine may provide an adaptive advantage during oxidative stress.

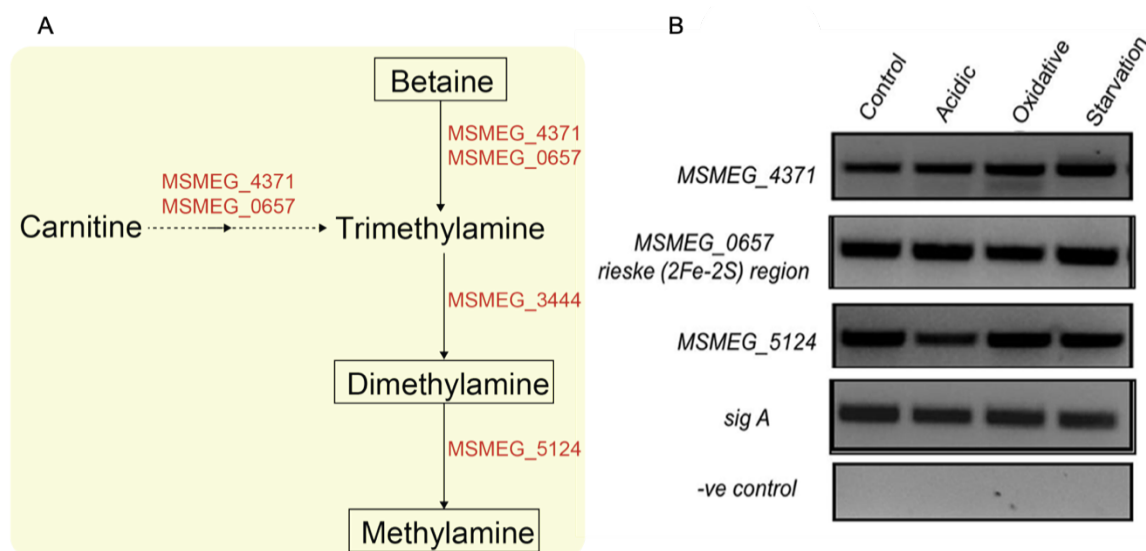


Figure 5.9: A putative pathway of biosynthesis of methylated amines in *M. smegmatis*. (A) Putative methylamine pathway showing the presence of intermediate metabolites (inside boxes) identified in our study. The numbers indicate the orthologues of *M. smegmatis* genes identified in this study. (B) RT-PCR of orthologue ORFs and their existence at the transcript level in *M. smegmatis*; a negative (-ive) control containing RNA instead of cDNA was used to rule out genomic DNA contamination. Note: Picture B was provided by Dr. Sharmistha Banerjee from University of Hyderabad.

5.12 Conclusion

In this chapter, we have presented the first comprehensive list of differential levels of metabolites in response to three microbicidal stresses in *M. smegmatis*. The comparative metabolomics profiles resulted from differential levels of biochemical products validated that these stresses induce a substantial metabolic shift in *M. smegmatis*. Our analysis reveals that the shift in the metabolic profile was more distinctive for nutrient starvation compared to other stresses. We have also listed all the stress-specific metabolites using the VIP score that plays a key role during various cellular stresses. It was noted that while the glutamine-glutamate metabolism pathway was perturbed significantly under all stress conditions, perturbation of the nicotinate-nicotinamide pathway was more significant in oxidative stress; and alanine-aspartate-glutamate metabolism and beta-alanine metabolism were more significant for the nutrient starvation response.

Additionally, we have found the accumulation of osmolytes such as methylamine and betaine that led us to unveil the presence of an unreported putative methylamine biosynthesis pathway in *M. smegmatis*. Capsular α -Glucan biosynthesis was observed to be affected by differential transcriptional regulation of selected enzymes of the pathway. The orthologs of these new pathways can be traced in pathogenic mycobacteria, and their significance may be studied to understand mycobacterial strategies for host invasion.

References

- Altaf, M., Miller, C. H., Bellows, D. S., & O'Toole, R. (2010). Evaluation of the Mycobacterium smegmatis and BCG models for the discovery of Mycobacterium tuberculosis inhibitors. *Tuberculosis*, *90*(6), 333–337.
<https://doi.org/10.1016/j.tube.2010.09.002>
- Anes, E., Peyron, P., Staali, L., Jordao, L., Gutierrez, M. G., Kress, H., ... Griffiths, G. (2006). Dynamic life and death interactions between mycobacterium smegmatis and J774 macrophages. *Cellular Microbiology*, *8*(6), 939–960.
<https://doi.org/10.1111/j.1462-5822.2005.00675.x>
- Baptista, R., Fazakerley, D. M., Beckmann, M., Baillie, L., & Mur, L. A. J. (2018). Untargeted metabolomics reveals a new mode of action of pretomanid (PA-824). *Scientific Reports*, *8*(1), 5084. <https://doi.org/10.1038/s41598-018-23110-1>
- Barberis, I., Bragazzi, N. L., Galluzzo, L., & Martini, M. (2017). The history of tuberculosis: From the first historical records to the isolation of Koch's bacillus. *Journal of Preventive Medicine and Hygiene*. Pacini Editore S.p.A. <https://doi.org/10.15167/2421-4248/jpmh2017.58.1.728>
- Barry, C. E. (2001). Mycobacterium smegmatis: an absurd model for tuberculosis? *Trends in Microbiology*, *9*(10), 473–474. [https://doi.org/10.1016/s0966-842x\(01\)02169-2](https://doi.org/10.1016/s0966-842x(01)02169-2)
- Borah, K., Beyß, M., Theorell, A., Wu, H., Basu, P., Mendum, T. A., ... McFadden, J. (2019). Intracellular Mycobacterium tuberculosis Exploits Multiple Host Nitrogen Sources during Growth in Human Macrophages. *Cell Reports*, *29*(11), 3580-3591.e4. <https://doi.org/10.1016/j.celrep.2019.11.037>
- Burg, M. B., & Ferraris, J. D. (2008, March 21). Intracellular organic osmolytes: Function and regulation. *Journal of Biological Chemistry*.
<https://doi.org/10.1074/jbc.R700042200>
- Chong, J., Soufan, O., Li, C., Caraus, I., Li, S., Bourque, G., ... Xia, J. (2018). MetaboAnalyst 4.0: Towards more transparent and integrative metabolomics analysis. *Nucleic Acids Research*, *46*(W1), W486–W494. <https://doi.org/10.1093/nar/gky310>
- Cordero, P. R. F., Bayly, K., Man Leung, P., Huang, C., Islam, Z. F., Schittenhelm, R. B., ... Greening, C. (2019). Atmospheric carbon monoxide oxidation is a widespread mechanism supporting microbial survival. *ISME Journal*, *13*(11), 2868–2881.

<https://doi.org/10.1038/s41396-019-0479-8>

- Drapal, M., Perez-Fons, L., Wheeler, P. R., & Fraser, P. D. (2014). The application of metabolite profiling to *Mycobacterium* spp.: Determination of metabolite changes associated with growth. *Journal of Microbiological Methods*, *106*, 23–32.
<https://doi.org/10.1016/j.mimet.2014.07.037>
- Drapal, M., Wheeler, P. R., & Fraser, P. D. (2016). Metabolite analysis of *Mycobacterium* species under aerobic and hypoxic conditions reveals common metabolic traits. *Microbiology (United Kingdom)*, *162*(8), 1456–1467.
<https://doi.org/10.1099/mic.0.000325>
- Ehrt, S., & Schnappinger, D. (2009). Mycobacterial survival strategies in the phagosome: Defence against host stresses. *Cellular Microbiology*. <https://doi.org/10.1111/j.1462-5822.2009.01335.x>
- Falkinham, J. O. (2009). The biology of environmental mycobacteria. *Environmental Microbiology Reports*. <https://doi.org/10.1111/j.1758-2229.2009.00054.x>
- Fan, T. W. M., & Lane, A. N. (2016, February 1). Applications of NMR spectroscopy to systems biochemistry. *Progress in Nuclear Magnetic Resonance Spectroscopy*. Elsevier B.V. <https://doi.org/10.1016/j.pnmrs.2016.01.005>
- Furuya, T., Nakao, T., & Kino, K. (2015). Catalytic function of the mycobacterial binuclear iron monooxygenase in acetone metabolism. *FEMS Microbiology Letters*, *362*(19).
<https://doi.org/10.1093/femsle/fnv136>
- Ganji, R., Dhali, S., Rizvi, A., Rapole, S., & Banerjee, S. (2016). Understanding HIV-*Mycobacteria* synergism through comparative proteomics of intra-phagosomal mycobacteria during mono- and HIV co-infection. *Scientific Reports*, *6*, 22060.
<https://doi.org/10.1038/srep22060>
- Gouzy, A., Larrouy-Maumus, G., Bottai, D., Levillain, F., Dumas, A., Wallach, J. B., ... Neyrolles, O. (2014). *Mycobacterium tuberculosis* Exploits Asparagine to Assimilate Nitrogen and Resist Acid Stress during Infection. *PLoS Pathogens*, *10*(2).
<https://doi.org/10.1371/journal.ppat.1003928>
- Gouzy, A., Larrouy-Maumus, G., Wu, T. Di, Peixoto, A., Levillain, F., Lugo-Villarino, G., ... Neyrolles, O. (2013). *Mycobacterium tuberculosis* nitrogen assimilation and host colonization require aspartate. *Nature Chemical Biology*, *9*(11), 674–676.

<https://doi.org/10.1038/nchembio.1355>

- Harper, C. J., Hayward, D., Kidd, M., Wiid, I., & van Helden, P. (2010). Glutamate dehydrogenase and glutamine synthetase are regulated in response to nitrogen availability in *Mycobacterium smegmatis*. *BMC Microbiology*, *10*, 138. <https://doi.org/10.1186/1471-2180-10-138>
- Holmström, K. -O, Welin, B., Mandal, A., Kristiansdottir, I., Teeri, T. H., Lamark, T., ... Palva, E. T. (1994). Production of the *Escherichia coli* betaine-aldehyde dehydrogenase, an enzyme required for the synthesis of the osmoprotectant glycine betaine, in transgenic plants. *The Plant Journal*, *6*(5), 749–758. <https://doi.org/10.1046/j.1365-313X.1994.6050749.x>
- Kanehisa, M., & Goto, S. (2000, January 1). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*. Oxford University Press. <https://doi.org/10.1093/nar/28.1.27>
- Kapopoulou, A., Lew, J. M., & Cole, S. T. (2011). The MycoBrowser portal: A comprehensive and manually annotated resource for mycobacterial genomes. *Tuberculosis*, *91*(1), 8–13. <https://doi.org/10.1016/j.tube.2010.09.006>
- Kim, S. G., Bae, H. S., & Lee, S. T. (2001). A novel denitrifying bacterial isolate that degrades trimethylamine both aerobically and anaerobically via two different pathways. *Archives of Microbiology*, *176*(4), 271–277. <https://doi.org/10.1007/s002030100319>
- Koeken, V. A. C. M., Lachmandas, E., Riza, A., Matzaraki, V., Li, Y., Kumar, V., ... Van Crevel, R. (2019). Role of Glutamine Metabolism in Host Defense against *Mycobacterium tuberculosis* Infection. *Journal of Infectious Diseases*, *219*(10), 1662–1670. <https://doi.org/10.1093/infdis/jiy709>
- Koeth, R. A., Levison, B. S., Culley, M. K., Buffa, J. A., Wang, Z., Gregory, J. C., ... Hazen, S. L. (2014). γ -butyrobetaine is a proatherogenic intermediate in gut microbial metabolism of L-carnitine to TMAO. *Cell Metabolism*, *20*(5), 799–812. <https://doi.org/10.1016/j.cmet.2014.10.006>
- Koliwer-Brandl, H., Syson, K., van de Weerd, R., Chandra, G., Appelmelk, B., Alber, M., ... Kalscheuer, R. (2016). Metabolic Network for the Biosynthesis of Intra- and Extracellular α -Glucans Required for Virulence of *Mycobacterium tuberculosis*. *PLoS Pathogens*, *12*(8). <https://doi.org/10.1371/journal.ppat.1005768>

- Leonardi, R., & Jackowski, S. (2007). Biosynthesis of Pantothenic Acid and Coenzyme A. *EcoSal Plus*, 2(2). <https://doi.org/10.1128/ecosalplus.3.6.3.4>
- Lew, J. M., Kapopoulou, A., Jones, L. M., & Cole, S. T. (2011). TubercuList - 10 years after. *Tuberculosis*, 91(1), 1–7. <https://doi.org/10.1016/j.tube.2010.09.008>
- Li, Xiaojing, Wu, J., Han, J., Hu, Y., & Mi, K. (2015). Distinct responses of mycobacterium smegmatis to exposure to low and high levels of hydrogen peroxide. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0134595>
- Li, Xinfeng, Mei, H., Chen, F., Tang, Q., Yu, Z., Cao, X., ... He, J. (2017). Transcriptome landscape of Mycobacterium smegmatis. *Frontiers in Microbiology*, 8(DEC), 2505. <https://doi.org/10.3389/fmicb.2017.02505>
- Loebel, R. O., Shorr, E., & Richardson, H. B. (1933). The Influence of Adverse Conditions upon the Respiratory Metabolism and Growth of Human Tubercle Bacilli 1. *Journal of Bacteriology*, 26(2), 167–200. <https://doi.org/10.1128/jb.26.2.167-200.1933>
- Man, D. K.-W., Kanno, T., Manzo, G., Robertson, B. D., Lam, J. K. W., & Mason, A. J. (2018). Rifampin- or Capreomycin-Induced Remodeling of the Mycobacterium smegmatis Mycolic Acid Layer Is Mitigated in Synergistic Combinations with Cationic Antimicrobial Peptides. *MSphere*, 3(4), e00218-18. <https://doi.org/10.1128/msphere.00218-18>
- McIntire, W. S. (1990). Trimethylamine dehydrogenase from bacterium W3A1. *Methods in Enzymology*, 188(C), 250–260. [https://doi.org/10.1016/0076-6879\(90\)88042-9](https://doi.org/10.1016/0076-6879(90)88042-9)
- Nambi, S., Long, J. E., Mishra, B. B., Baker, R., Murphy, K. C., Olive, A. J., ... Sassetti, C. M. (2015). The Oxidative Stress Network of Mycobacterium tuberculosis Reveals Coordination between Radical Detoxification Systems. *Cell Host and Microbe*, 17(6), 829–837. <https://doi.org/10.1016/j.chom.2015.05.008>
- O'Toole, R., Smeulders, M. J., Blokpoel, M. C., Kay, E. J., Lougheed, K., & Williams, H. D. (2003). A two-component regulator of universal stress protein expression and adaptation to oxygen starvation in Mycobacterium smegmatis. *Journal of Bacteriology*, 185(5), 1543–1554. <https://doi.org/10.1128/JB.185.5.1543-1554.2003>
- Pavelka, M. S., & Jacobs, W. R. (1996). Biosynthesis of diaminopimelate, the precursor of lysine and a component of peptidoglycan, is an essential function of Mycobacterium smegmatis. *Journal of Bacteriology*, 178(22), 6496–6507.

<https://doi.org/10.1128/jb.178.22.6496-6507.1996>

Pelosi, A., Smith, D., Brammananth, R., Topolska, A., Billman-Jacobe, H., Nagley, P., ... Coppel, R. L. (2012). Identification of a novel gene product that promotes survival of mycobacterium smegmatis in macrophages. *PLoS ONE*.

<https://doi.org/10.1371/journal.pone.0031788>

Piddington, D. L., Kashkouli, A., & Buchmeier, N. A. (2000). Growth of Mycobacterium tuberculosis in a defined medium is very restricted by acid pH and Mg²⁺ levels.

Infection and Immunity, 68(8), 4518–4522. <https://doi.org/10.1128/IAI.68.8.4518-4522.2000>

Pierre-Audigier, C., Jouanguy, E., Lamhamedi, S., Altare, F., Raugier, J., Vincent, V., ... Casanova, J. L. (1997). Fatal disseminated Mycobacterium smegmatis infection in a child with inherited interferon γ receptor deficiency. *Clinical Infectious Diseases*, 24(5), 982–984. <https://doi.org/10.1093/clinids/24.5.982>

Price, C. T. D., Bukka, A., Cynamon, M., & Graham, J. E. (2008). Glycine betaine uptake by the proXVWZ ABC transporter contributes to the ability of Mycobacterium tuberculosis to initiate growth in human macrophages. *Journal of Bacteriology*, 190(11), 3955–3961. <https://doi.org/10.1128/JB.01476-07>

Rao, M., Streur, T. L., Aldwell, F. E., & Cook, G. M. (2001). Intracellular pH regulation by Mycobacterium smegmatis and Mycobacterium bovis BCG. *Microbiology*, 147(4), 1017–1024. <https://doi.org/10.1099/00221287-147-4-1017>

Rao, P. K., & Li, Q. (2009, September). Protein turnover in mycobacterial proteomics. *Molecules*. <https://doi.org/10.3390/molecules14093237>

Roxas, B. A., & Li, Q. (2009). Acid stress response of a mycobacterial proteome: insight from a gene ontology analysis. *International Journal of Clinical and Experimental Medicine*, 2(4), 309–328. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/20057975>

Tambellini, N., Zaremborg, V., Turner, R., & Weljie, A. (2013). Evaluation of Extraction Protocols for Simultaneous Polar and Non-Polar Yeast Metabolite Analysis Using Multivariate Projection Methods. *Metabolites*, 3(3), 592–605. <https://doi.org/10.3390/metabo3030592>

Tyagi, P., Dharmaraja, A. T., Bhaskar, A., Chakrapani, H., & Singh, A. (2015).

- Mycobacterium tuberculosis has diminished capacity to counteract redox stress induced by elevated levels of endogenous superoxide. *Free Radical Biology and Medicine*, 84, 344–354. <https://doi.org/10.1016/j.freeradbiomed.2015.03.008>
- Ulrich, E. L., Akutsu, H., Doreleijers, J. F., Harano, Y., Ioannidis, Y. E., Lin, J., ... Markley, J. L. (2008). BioMagResBank. *Nucleic Acids Research*, 36(SUPPL. 1), D402-408. <https://doi.org/10.1093/nar/gkm957>
- Vandal, O. H., Nathan, C. F., & Ehrt, S. (2009, August 1). Acid resistance in Mycobacterium tuberculosis. *Journal of Bacteriology*. American Society for Microbiology Journals. <https://doi.org/10.1128/JB.00305-09>
- Voskuil, M. I., Bartek, I. L., Visconti, K., & Schoolnik, G. K. (2011). The response of Mycobacterium tuberculosis to reactive oxygen and nitrogen species. *Frontiers in Microbiology*, 2(MAY). <https://doi.org/10.3389/fmicb.2011.00105>
- Wang, R., Prince, J. T., & Marcotte, E. M. (2005). Mass spectrometry of the M. smegmatis proteome: Protein expression levels correlate with function, operons, and codon bias. *Genome Research*, 15(8), 1118–1126. <https://doi.org/10.1101/gr.3994105>
- Wanichthanarak, K., Fahrmann, J. F., & Grapov, D. (2015). Genomic, Proteomic, and Metabolomic Data Integration Strategies. *Biomarker Insights*, 10s4(Suppl 4), BMI.S29511. <https://doi.org/10.4137/BMI.S29511>
- Weljie, A. M., Newton, J., Mercier, P., Carlson, E., & Slupsky, C. M. (2006). Targeted profiling: Quantitative analysis of ¹H NMR metabolomics data. *Analytical Chemistry*, 78(13), 4430–4442. <https://doi.org/10.1021/ac060209g>
- Whatmore, A. M., Chudek, J. A., & Reed, R. H. (1990). The effects of osmotic upshock on the intracellular solute pools of Bacillus subtilis. *Journal of General Microbiology*, 136(12), 2527–2535. <https://doi.org/10.1099/00221287-136-12-2527>
- Wishart, D. S., Tzur, D., Knox, C., Eisner, R., Guo, A. C., Young, N., ... Querengesser, L. (2007). HMDB: The human metabolome database. *Nucleic Acids Research*, 35(SUPPL. 1), D521-526. <https://doi.org/10.1093/nar/gkl923>
- Worley, B., & Powers, R. (2013). Multivariate Analysis in Metabolomics. *Current Metabolomics*, 1(1), 92–107. <https://doi.org/10.2174/2213235x11301010092>
- Xia, J., Bjorndahl, T. C., Tang, P., & Wishart, D. S. (2008). MetaboMiner - Semi-automated identification of metabolites from 2D NMR spectra of complex biofluids. *BMC*

Bioinformatics, 9, 507. <https://doi.org/10.1186/1471-2105-9-507>

Xia, J., Wishart, D. S., & Valencia, A. (2011). MetPA: A web-based metabolomics tool for pathway analysis and visualization. In *Bioinformatics* (Vol. 27, pp. 2342–2344). Oxford University Press. <https://doi.org/10.1093/bioinformatics/btq418>

Yamada, H., Yamaguchi, M., Igarashi, Y., Chikamatsu, K., Aono, A., Murase, Y., ... Mitarai, S. (2018). Mycolicibacterium smegmatis, Basonym Mycobacterium smegmatis, Expresses Morphological Phenotypes Much More Similar to Escherichia coli Than Mycobacterium tuberculosis in Quantitative Structome Analysis and CryoTEM Examination. *Frontiers in Microbiology*, 9(SEP). <https://doi.org/10.3389/fmicb.2018.01992>

Yancey, P. H., & Somero, G. N. (1979). Counteraction of urea destabilization of protein structure by methylamine osmoregulatory compounds of elasmobranch fishes. *Biochemical Journal*, 183(2), 317–323. <https://doi.org/10.1042/bj1830317>

Yancey, Paul H., Blake, W. R., & Conley, J. (2002). Unusual organic osmolytes in deep-sea animals: Adaptations to hydrostatic pressure and other perturbants. In *Comparative Biochemistry and Physiology - A Molecular and Integrative Physiology* (Vol. 133, pp. 667–676). Elsevier Inc. [https://doi.org/10.1016/S1095-6433\(02\)00182-4](https://doi.org/10.1016/S1095-6433(02)00182-4)

You, D., Xu, Y., Yin, B. C., & Ye, B. C. (2019). Nitrogen regulator GlnR controls redox sensing and lipids anabolism by directly activating the whiB3 in Mycobacterium smegmatis. *Frontiers in Microbiology*. <https://doi.org/10.3389/fmicb.2019.00074>

Yousf, S., Sardesai, D. M., Mathew, A. B., Khandelwal, R., Acharya, J. D., Sharma, S., & Chugh, J. (2019). Metabolic signatures suggest o-phosphocholine to UDP-N-acetylglucosamine ratio as a potential biomarker for high-glucose and/or palmitate exposure in pancreatic β -cells. *Metabolomics*, 15(4). <https://doi.org/10.1007/s11306-019-1516-3>

Zahrt, T. C., & Deretic, V. (2002). Reactive nitrogen and oxygen intermediates and bacterial defenses: Unusual adaptations in Mycobacterium tuberculosis. *Antioxidants and Redox Signaling*. <https://doi.org/10.1089/152308602753625924>

Zhao, P., Li, J., Li, Y., Tian, Y., Yang, L., & Li, S. (2017). Integrating Transcriptomics, Proteomics, and Metabolomics Profiling with System Pharmacology for the Delineation of Long-Term Therapeutic Mechanisms of Bufei Jianpi Formula in Treating COPD. *BioMed Research International*, 2017. <https://doi.org/10.1155/2017/7091087>

Chapter 6

***Cold storage reveals distinct metabolic perturbations in processing
and non-processing cultivars of potato (*Solanum tuberosum* L.)***

6.1 Introduction

Potato (*Solanum tuberosum*), the largest popular non-grain vegetable food crop worldwide (Hardigan et al., 2017), belongs to the Solanaceae family and ranks third most wanted food crop after wheat and rice. Potatoes can be cultivated in diverse environments and are currently grown in more than 100 different countries (Mankotia & Sharma, 2020). As per the recent estimates of the UN-FAO (Food and Agriculture Organizations of the United Nations), the global potato production was over 368 million metric tonnes in the year 2018 alone (www.fao.org/). India ranks as the second-largest potato producing nation of the world, with the reported production of around 52 million metric tons in 2019 (www.indiastat.com/). The majority of potato cultivation in India happens in the states of Uttar Pradesh, West Bengal, Punjab, Karnataka, Bihar, Assam, and Madhya Pradesh. Potatoes are not cultivated only as a vegetable for cooking purposes but are processed into a variety of forms such as potato chips, French fries and are also used for the production of starch.

Amongst all major food crops, potato is nutritionally superior in producing the highest amount of food, dry matter, protein, and other nutrients per unit land and time (growing time) (Rajiv & Kavar, 2016). After harvest, potato tubers need to be stored under cold conditions to suppress sprouting, prevent the growth of tuber-borne pathogens, avoid losses due to shrinkage, retention of dry matter, extend post-harvest shelf life, and to maintain a year-round supply for consumption purpose (Bianchi, Scalzo, Testoni, & Maestrelli, 2014; Hou et al., 2017; Singh & Saldaña, 2011; D. Zhang, Mu, Sun, Chen, & Zhang, 2017). However, cold-stored potato tubers exhibit a phenomenon called as cold-induced sweetening (CIS) in which starch is rapidly degraded and reducing sugars (RS) such as glucose and fructose are formed via hydrolysis of sucrose (Dale & Bradshaw, 2003; Galani Yamdeu, Gupta, Patel, Shah, & Talati, 2016).

CIS is a complex phenomenon, wherein the accumulation of RS is regulated by several metabolic pathways, such as starch synthesis and degradation, hexogenesis, glycolysis, and anaerobic respiration (Liu et al., 2017; Malone, Mittova, Ratcliffe, & Kruger, 2006a; J. R. Sowokinos, 2001a). During cold storage, the starch in amyloplast undergoes either hydrolytic or phosphorolytic degradation to produce hexose phosphates (hexose-P) or free sugars, which are then exported into the cytoplasm (Smith, Zeeman, & Smith, 2005; Weber, 2004). In the cytosol, these metabolites follow either hexogenesis pathway leading to the formation of sucrose or undergo glycolysis, which further enter into mitochondrial respiration (Xia Chen et al., 2012; Greiner, Rausch, Sonnewald, & Herbers, 1999; J. R. Sowokinos, 2001b). However, it has been suggested that at low temperatures, the entry of hexose-P into glycolysis is restricted, diverting the products of starch breakdown into sucrose synthesis (Malone, Mittova,

Ratcliffe, & Kruger, 2006b). The sucrose, thus formed, is then transported into vacuole where enzyme invertase hydrolyses sucrose into glucose and fructose, the two reducing sugars primarily responsible for cold-induced sweetening in potato.

The increase in RS content in potato tubers during storage poses severe problems for processors. Exposure of RS to high temperatures during frying results in a non-enzymatic chemical reaction – the Maillard reaction – between reducing sugars and amino acids, which leads to the formation of different flavours and dark-pigmented products (Amjad, Javed, Hameed, Hussain, & Ismail, 2020; Dale & Bradshaw, 2003; J. R. Sowokinos, 2001b; Tamanna & Mahmood, 2015). This process has attracted more attention because of the formation of acrylamide, a potential neurotoxin and carcinogen, through the Maillard reaction between the amino group of asparagine and carbonyl group of sugars via an N-glycoside intermediate (Mottram, Wedzicha, & Dodson, 2002). Thus, the acrylamide levels in fried potato products are directly proportional to the RS content in the potato tubers and are increased by prior storage of the potatoes at low temperatures (Jin et al., 2016).

The potato industry faces a massive loss due to CIS as chips and French fries get discolored, a parameter which is primarily determined by RS content in potato tubers. As a result, CIS is known to be one of the critical parameters in potato cultivation; the selection and breeding of CIS-resistant potato tubers have become a priority in potato breeding programs (Colman, Massa, Carboni, & Feingold, 2017; Hamernik, Hanneman, & Jansky, 2009; Xiong, Tai, & Seabrook, 2002). Potato cultivars resistant to CIS can deliver a wide range of advantages, such as the reduced requirement for sprout inhibitors, decreased dry matter (DM) losses, reduced pathogen levels during storage, and reduced chilling injury during handling, harvest, transport, and storage (Ali et al., 2016). However, potato breeding programs are often complicated due to tetrasomic inheritance and high-level heterozygosity of cultivated potato (Muthoni, Kabira, Shimelis, & Melis, 2015). In this regard, the metabolic stability of potato tubers over the long-term cold storage has been regarded as one of the prime traits to be examined for breeding programs worldwide (Brummell et al., 2011), wherein the selection of potato genotypes at early generations, using biochemical information through marker-trait associations are advantageous (Slater et al., 2014). Although CIS is well studied, only the genes operating in carbohydrate metabolism pathway have been cloned and functionally characterized (Baldwin et al., 2011; Datir et al., 2012a; Draffehn, Meller, Li, & Gebhardt, 2010; Li et al., 2008; C. M. Menéndez et al., 2002). In addition to this, little information is available about the metabolic events associated with the CIS process under cold storage conditions.

Metabolomics has been identified as a powerful technique to characterize crop plants as it detects a broad range of metabolites from a single extract and can expedite the selection of elite crop traits and improve breeding materials (Kumar, Bohra, Pandey, Pandey, & Kumar, 2017a). In potato breeding, metabolomic studies have been crucial primarily because tuber quality traits such as content and quality of starch, chipping quality, flesh colour, flavour, and glycoalkaloid content, etc., are related to a diverse set of metabolites and changes in metabolic networks (Carreno-Quintero et al., 2012; Chaparro, Holm, Broeckling, Prenni, & Heuberger, 2018; Dobson et al., 2008). Indeed, metabolomics approaches have previously been used to detect various metabolites present in potato tubers to assess the changes in composition that occur in genetically modified potato tubers, to understand the potato tuber life cycle, and to detect anthocyanin and polyphenol profiles of colored potato varieties (Defernez et al., 2004; Oertel et al., 2017; Shepherd et al., 2010a; Uri, Juhász, Polgár, & Bánfalvi, 2014). For instance, metabolic profiles in different life cycle stages of potato tubers have been characterized to link temporal changes in metabolites to their acrylamide-forming potential (Shepherd et al., 2010b). Likewise, metabolite characterization of six Hungarian commercial potato cultivars representing three major cooking types mainly differing in their starch contents and dormancy periods at harvest and storage (20–22°C) in the dark has revealed changes in the metabolite levels (Uri et al., 2014). Therefore, variations in metabolite profiling among different potato cultivars offer the potential to develop potato cultivars with improved processing characteristics.

Potato processing is fast emerging as an important industry in India, and therefore, the demand for processed potato products such as chips, wedges, French fries, cutlets, flakes, etc. is continuously growing (Marwaha, Pandey, Kumar, Singh, & Kumar, 2010). To meet the requirement for processing potatoes, specific morphological and biochemical attributes are necessary for potato varieties. Morphological traits mainly include the size and shape of tubers, internal and external defects, whereas biochemical markers include dry matter, specific gravity, reducing sugars, free amino acids, phenol content, etc. Potato cultivars with good resistance to CIS, high specific gravity, and dry matter (DM) content along with low RS content are ideal for processing purposes (R Rana, 2007; S Kaur, 2014). In this regard, Atlantic and Frito Lay-1533, which are commercially grown processing cultivars, have been ranked as the best varieties for processing purpose, primarily due to high specific gravity and DM content and low RS content (R Rana, 2007; Raigond, Mehta, & Singh, 2018).

On the other hand, Indian popular potato cultivars Kufri Pukhraj and Kufri Jyoti, along with one locally grown table purpose cultivar (PU1), have been found inferior for processing

purpose due to high RS and DM content (Kaur, 2017; P Aggarwal, 2017; Raigond et al., 2018; S Kaur, 2014). Kufri Pukhraj and Kufri Jyoti are the most popular potato cultivars among Indian farmers due to their medium and average storability. Therefore, to advance our knowledge about the biochemical variation and the metabolic events associated with the CIS process of potato tubers under cold storage, a comprehensive analysis needs to be carried out in potato cultivars differing in their storability as well as processing attributes.

In this study, an untargeted ^1H nuclear magnetic resonance (NMR)-based approach was conducted to assess the alterations in the metabolic profiles of five different potato cultivars namely, Atlantic, Frito Lay-1533, Kufri Pukhraj, Kufri Jyoti, and PU1, under cold storage at 4 °C. These potato cultivars are differing in their CIS abilities and processing characteristics. The key purpose of this work was to investigate the variations in metabolic profiles of different potato cultivars at fresh harvest and after cold storage to further advance the knowledge of metabolic events associated with the CIS phenomenon. From this study, key metabolites were identified that could potentially be used in breeding programs for the development of CIS-resistant cultivars with improved processing characteristics and thereby would enhance the quality of potato tubers.

6.2 Materials and Methods

6.2.1 Plant Material

Two readily available exotic potato cultivars – Frito Lay-1533 and Atlantic (Pepsi Foods Pvt. Ltd. Channo, Sangrur) – suitable for processing purpose; and two Indian non-processing cultivars – Kufri Jyoti and Kufri Pukhraj (Central Potato Research Institute, Shimla) (Kaur, 2017; P Aggarwal, 2017; RS Marwaha, 2005; S Kaur, 2014; Sharma, 2012); along with one locally grown potato cultivar – (PU1), which is reported to possess poor storability and contain high RS content after cold storage (Datir, Mirikar, & RaviKumar, 2019) were used in the current study. These cultivars were obtained from BT Company and Jai Kisan Farm Products and Cold Chains Pvt. Ltd, India, Pune. All these cultivars vary in their CIS abilities and processing characteristics (Raigond et al., 2018; RS Marwaha, 2005; Sharma, 2012). The physiological information of these potato varieties has been tabulated in Table 6.1.

6.2.2 Potato Plantation and Harvesting

Tubers from all the five different potato cultivars (listed above) were grown in triplicates under open-door conditions (natural conditions and sunlight) in PB5 Polythene bags

containing potting mix (50% shredded pine bark, 20% crusher dust, 10% cow dung, 20% soil supplemented with sand and slow-release fertilizer) on 20th November 2017, at the Department of Biotechnology, SPPU, Pune, India. The plants were watered regularly, and on 5th March 2018, the tubers were harvested after full senescence. In each replicate, two medium to large-sized potato tubers were selected randomly for each cultivar and were placed in a paper bag for analysis. The tubers were cleaned under running tap water, cored and de-skinned. The three freshly harvested (FH) tubers (one from each replication) from each cultivar were immediately processed for the analysis (referred to as treatment (a)). The remaining three freshly harvested tubers were kept in the dark location for one month at 4 °C in paper bags (referred to as treatment (b)). After one month of cold storage (CS), these tubers from each cultivar were processed for the metabolite extraction for NMR analysis. Prior to metabolite extraction, all potato tuber samples were subjected to freeze-drying (Operon, FDB-5503, Korea) for one week.

Table 6.1: information on potato cultivars.

Variety /Cultivar	Source	Type	Shape	Colour		Storage behaviour
				Skin	Flesh	
Frito Lay-1533	Pepsi Foods Pvt. Ltd. Channo, Sangrur	Processing	Oval	Light russet	White	Good
Atlantic	Pepsi Foods Pvt. Ltd. Channo, Sangrur	Processing	Round	Brownish-yellow	White	Good
Kufri Pukhraj	Central Potato Research Institute, Shimla	Non-processing/Table purpose	Oval	Brown	Cream	Average
Kufri Jyoti	Central Potato Research Institute, Shimla	Non-processing/Table purpose	Round	Brownish yellow	Cream	Medium
PU1	No information	Non-processing/Table purpose	Oval	Brownish yellow	Cream	Medium

The information is adapted from Kaur and Aggarwal, 2014; Kaur and Khurana, 2017; Marwaha, et al., 2005; Raigond et al., 2018. The storage behaviour of local potato cultivar PU1 was medium/poor (noted during this study).

6.2.3 Metabolite extraction

After freeze-drying, peeled potato tubers were grounded into a very fine powder. Approximately 200 mg of freeze-dried powdered sample was homogenized in 200 μ l Phosphate Buffer Saline (PBS) in 2 ml of Eppendorf tubes. After vortexing for 5 min, 400 μ l ice-cold methanol (Sigma, HPLC grade) was added to each tube and then vortexed again for 5 min. After a 12 h incubation period at -20°C , followed by centrifugation at $16,000 \times g$ (Eppendorf centrifuge 5415 C, Hamburg, Germany) for 20 min at 4°C , the supernatants were collected and transferred to new 1.5ml Eppendorf tubes and were subjected to lyophilization (Operon, FDB-5503, Korea). A total of 30 distinct samples (three replicates for each freshly harvested and cold storage tubers for all five different cultivars) were used for NMR data measurement. The lyophilized tuber samples were reconstituted into 580 μ l 100% NMR buffer (20 mM sodium phosphate, pH 7.4 in D_2O containing 0.4 mM DSS (2,2-dimethyl-2-silapentane-5-sulfonic acid). After vortexing and centrifugation (4000 g for 2 min) at room temperature, the supernatants were transferred to 5mm NMR tubes for NMR analysis.

6.3 NMR spectroscopy and spectral processing

All the NMR spectra were acquired at 298 K using the NOESY-presaturation pulse sequence `noesygppr1d` as described in detail in Chapter 2. A total of 64 scans were acquired in 32K data points using a spectral width of 7200 Hz and an acquisition period of 6.95 s for each spectrum. ^1H - ^1H total correlation spectroscopy (TOCSY) experiment was performed to further assist in the resonance assignment of metabolites. The ^1H NMR spectra of all the potato samples were processed using Bruker's NMR data processing software Topspin (v3.5) (www.bruker.com/bruker/topspin). Standard pre-processing steps such as phasing, baseline correction, line broadening was carried out for all the individual ^1H NMR spectra. All the ^1H chemical shifts were directly referenced with respect to the chemical shift of methyl singlet of DSS internal reference.

6.4 Quantification assessment and statistical analysis

After spectral processing, identification and quantification of metabolites were carried out with the Chenomx NMR Suite 8.1 (Edmonton, AB, Canada) software as described in detail in Chapter 2. All the identified metabolites were further confirmed with biological magnetic resonance data bank (BMRB) (Ulrich et al., 2008) and human metabolome database (HMDB) (Wishart et al., 2007). In addition, two-dimensional ^1H - ^1H TOCSY was used for further metabolite confirmation via a semi-automated software – MetaboMiner (Xia, Bjorndahl, Tang, & Wishart, 2008). All the identified metabolites were then quantified using the profiler module of Chenomx software, which enables metabolites quantification relative to an internal standard of known concentration (400 μM).

To assess the effect of cold-treatment on the metabolic profiles of different potato cultivars differing in their CIS trait multivariate statistical analysis was carried out to discern the differences between metabolic profiles of fresh and cold storage potato tubers. Supervised Partial Least Squares Discriminant Analysis (PLS-DA) was performed using a free web server, MetaboAnalyst, to identify the metabolites that significantly alter between FH and CS tubers in all five cultivars studied and are responsible for the intergroup discrimination. Prior to chemometric analysis, the metabolomics data were subjected to the Pareto-scaling approach. To identify the key metabolites responsible for the differential clustering of score plots in the PLS-DA model, the variable importance of projection (VIP) score plot was generated from PLS-DA analysis. Further, one-way ANOVA analysis followed by Fishers's least significant difference (LSD) post-hoc tests were used to assess the significance of changes in the levels of

metabolite concentrations for both the treatments in all potato cultivars simultaneously. In ANOVA, $q\text{-value} \leq 0.05$ has been used to determine the significance of differences in metabolite levels. Pair-wise analysis of all five cultivars in FH and CS treatments were carried out using Volcano plot utility of MetaboAnalyst that screens significant metabolites based on fold-change (FC) and FDR-adjusted $p\text{-value}$ ($q\text{-value}$). A union set of significant metabolites identified from volcano plot analysis and VIP score were used to short-list critical metabolites associated with CS treatment in each cultivar to reduce the chances of losing any important metabolites of interest. Box and Whisker plots were generated for all discriminatory metabolites identified from VIP score plot and volcano plot analysis was used to visualise the comparative alteration of a particular metabolite for both freshly harvested and cold storage tuber samples across replicates and in different potato cultivars. Metabolites, e.g., ascorbate, with low signal-to-noise ($s/n \leq 15$) in NMR analysis, although identified with confidence, were excluded from box and whisker plot analysis. In addition, metabolite-metabolite correlations were obtained using Pearson's correlation coefficient analysis to identify all the significant correlations ($p\text{-value} \leq 0.05$) in freshly harvested and cold storage tuber samples for each cultivar.

6.5 Metabolic pathway analysis

Metabolic pathway analysis of all the discriminating metabolites was conducted to understand the biological significance of the metabolic changes and to identify target pathways that predominantly get affected under cold stress in all five cultivars. Metabolic pathway analysis of significantly differential metabolites identified through volcano plot analysis ($q\text{-value} \leq 0.05$, $FC \geq 1.5$) and VIP score plot ($VIP \geq 1$) was performed using KEGG, MetaboAnalyst, and reference pathways from the previous literature (Malone et al., 2006b; J. R. Sowokinos, 2001b). Metabolic Pathway Analysis (MetPa) was carried out using MetaboAnalyst (Chong et al., 2018; Xia, Wishart, & Valencia, 2011).

6.6 Results

6.6.1 Global profiling of metabolites in different potato cultivars – Processing versus non-processing cultivars

Untargeted NMR-based metabolic profiling was performed on 30 potato samples spanning five different potato cultivars: Atlantic, FL-1533, Kufri Pukhraj, Kufri Jyoti, and PU1, which differ in their CIS abilities and processing characteristics. A total of 41 metabolites

were detected from the aqueous phase of the metabolic extracts of all the samples using standard one dimensional ^1H NMR. All the metabolites were identified using the profiler module of Chenomx NMR Suite 8.1 (Edmonton, AB, Canada) software. The organic phase and water-insoluble compounds were excluded from our analysis as they gave broad signals in ^1H NMR. Six NMR peaks remained unassigned and have been annotated as unknowns - U1, U2, U3, U4, and U5 in Figure 6.1. Resonance assignment of these metabolites was further confirmed using 2D ^1H - ^1H TOCSY NMR spectroscopy (Figure 6.2). The NMR spectra displayed signals mainly from amino acids, sugars, sugar alcohol, and organic acids. Further, secondary metabolites such as phenolic compound 5-caffeoylquinic acid (chlorogenic acid) and alkaloid trigonelline were also identified from the ^1H NMR spectra. All the 41 metabolites and their respective ^1H chemical shifts (in reference to DSS) have been tabulated in Appendix Table 6.4.

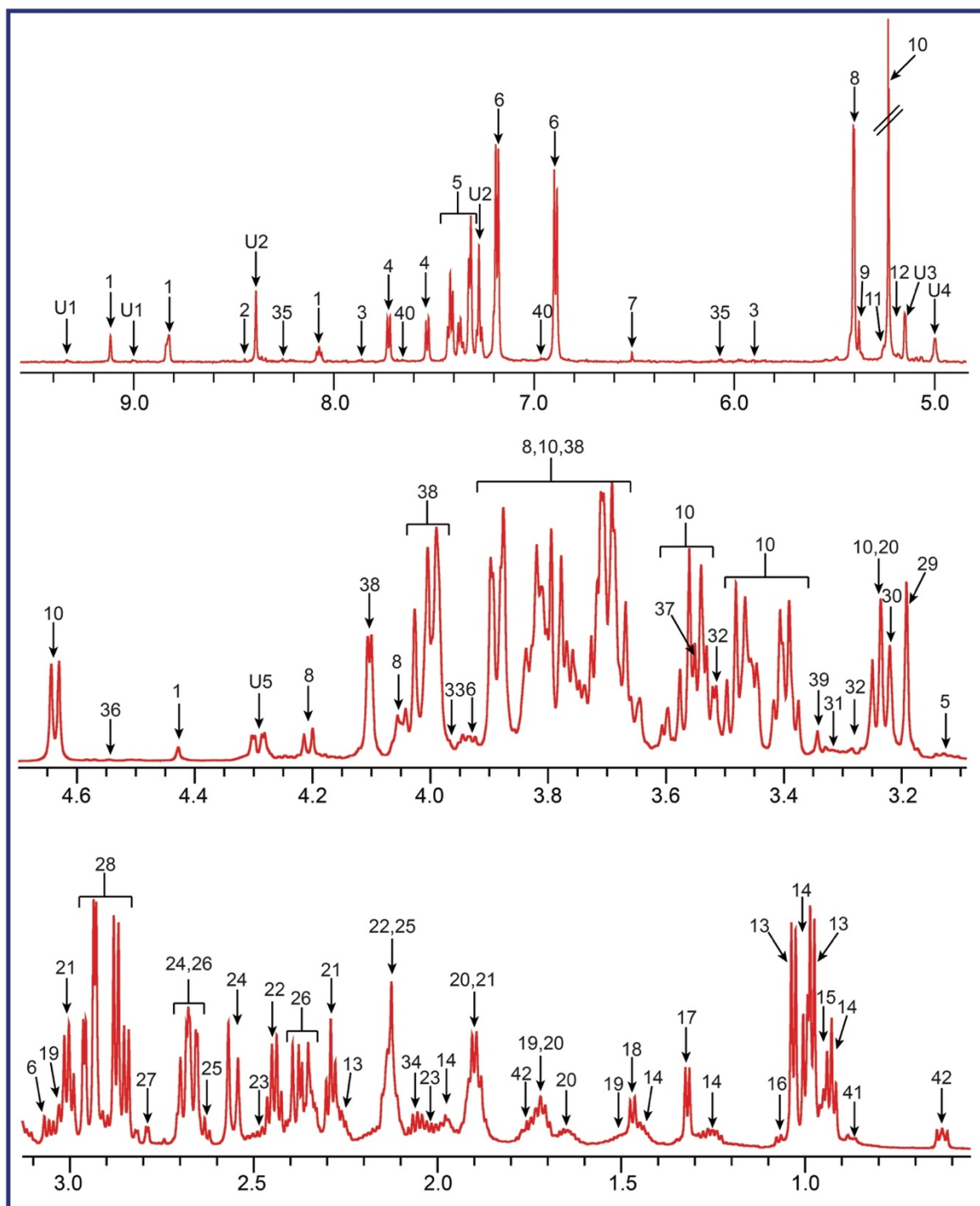


Figure 6.1: $^1\text{H-NMR}$ spectrum of the methanolic extract of Kufri Pukhraj potato cultivar (cold storage); 1, Trigonelline; 2, Formate; 3, Uridine; 4, Tryptophan; 5, Phenylalanine; 6, Tyrosine; 7, Fumarate; 8, Sucrose; 9, Allantoin; 10, Glucose; 11, Galactose; 12, Mannose; 13, Valine; 14, Isoleucine; 15, Leucine; 16, 3-hydroxyisobutyrate; 17, Threonine; 18, Alanine; 19, Lysine; 20, Arginine; 21, 4-Aminobutyrate; 22, Glutamine; 23, Pyroglutamate; 24, Citrate; 25, Methionine; 26, Malate; 27, Aspartate; 28, Asparagine; 29, Choline; 30, sn-glycero-3-phosphocholine; 31, Proline; 32, Myo-inositol; 33, Serine; 34, Glutamate; 35, Adenosine; 36, Ascorbate; 37, Glycine; 38, Fructose; 39, Methanol; 40, Chlorogenate; 41, LDL; 42, DSS; 43, U1; 44, U2; 45, U3; 46, U4; 47, U5.

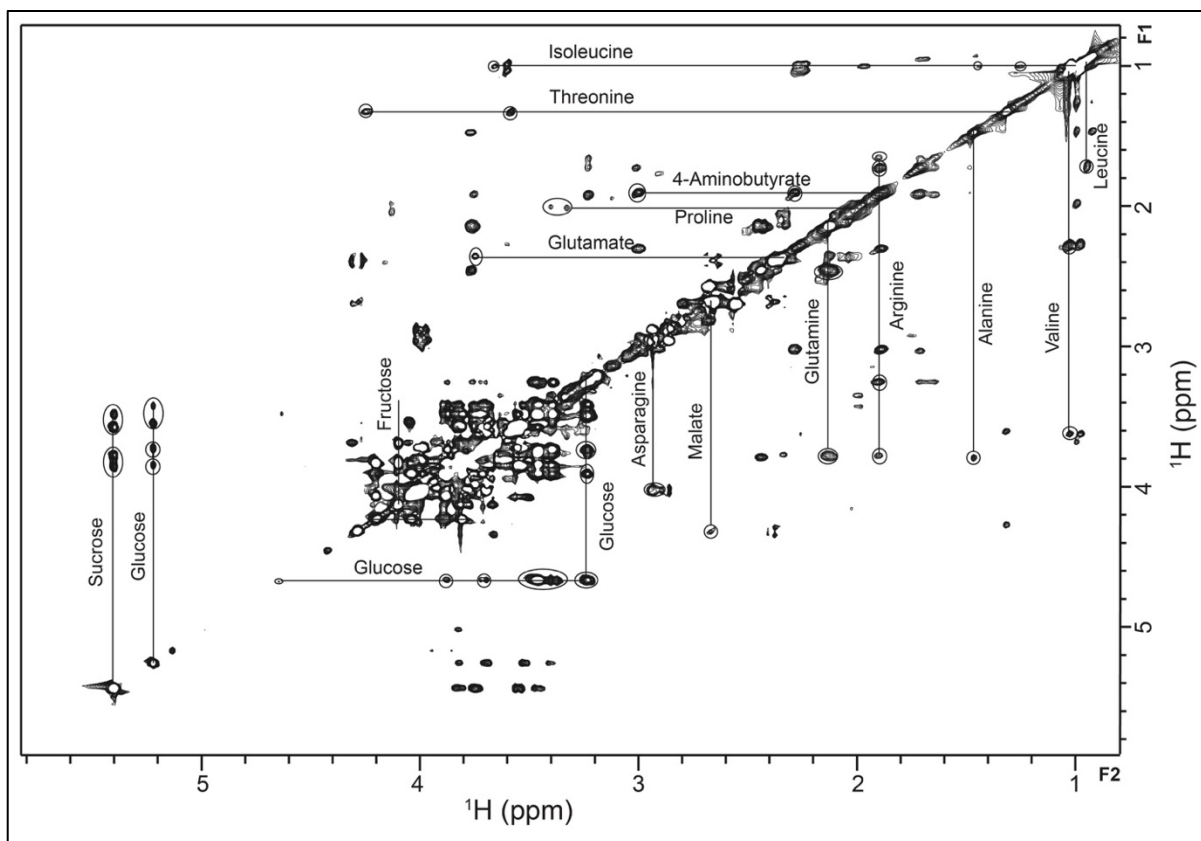


Figure 6.2: ^1H - ^1H TOCSY correlation spectrum of the methanolic extract of Kufri Pukhraj potato cultivar (cold storage). Cross-peaks in the TOCSY spectrum were used to re-confirm the resonance assignments enlisted in Appendix Table 6.4. Representative peaks have been labelled in the spectrum.

6.6.2 Cold treatment influences the metabolome of potato cultivars

Multivariate data analysis was conducted to explore the effect of cold-treatment on the metabolome of different cultivars. PLS-DA analysis, which was applied to the data matrix consisting of 38 columns and 30 rows, displays a clear separation on the scores plot, accounting for a 48.1% of total variance along PC1 and 28.5% of total data variance along PC2. The scores corresponding to treatment-a (fresh harvest) and treatment-b (cold storage) for each cultivar were grouped separately in PLS-DA analysis, indicating discrimination of samples upon cold treatment. It was also observed from the PLS-DA score plot analysis that the samples corresponding to FH tubers of Atlantic and FL-1533 processing cultivars were clustered together while samples from FH tubers of non-processing cultivars (Kufri Jyoti and Kufri Pukhraj) were more similar to each other (Figure 6.3).

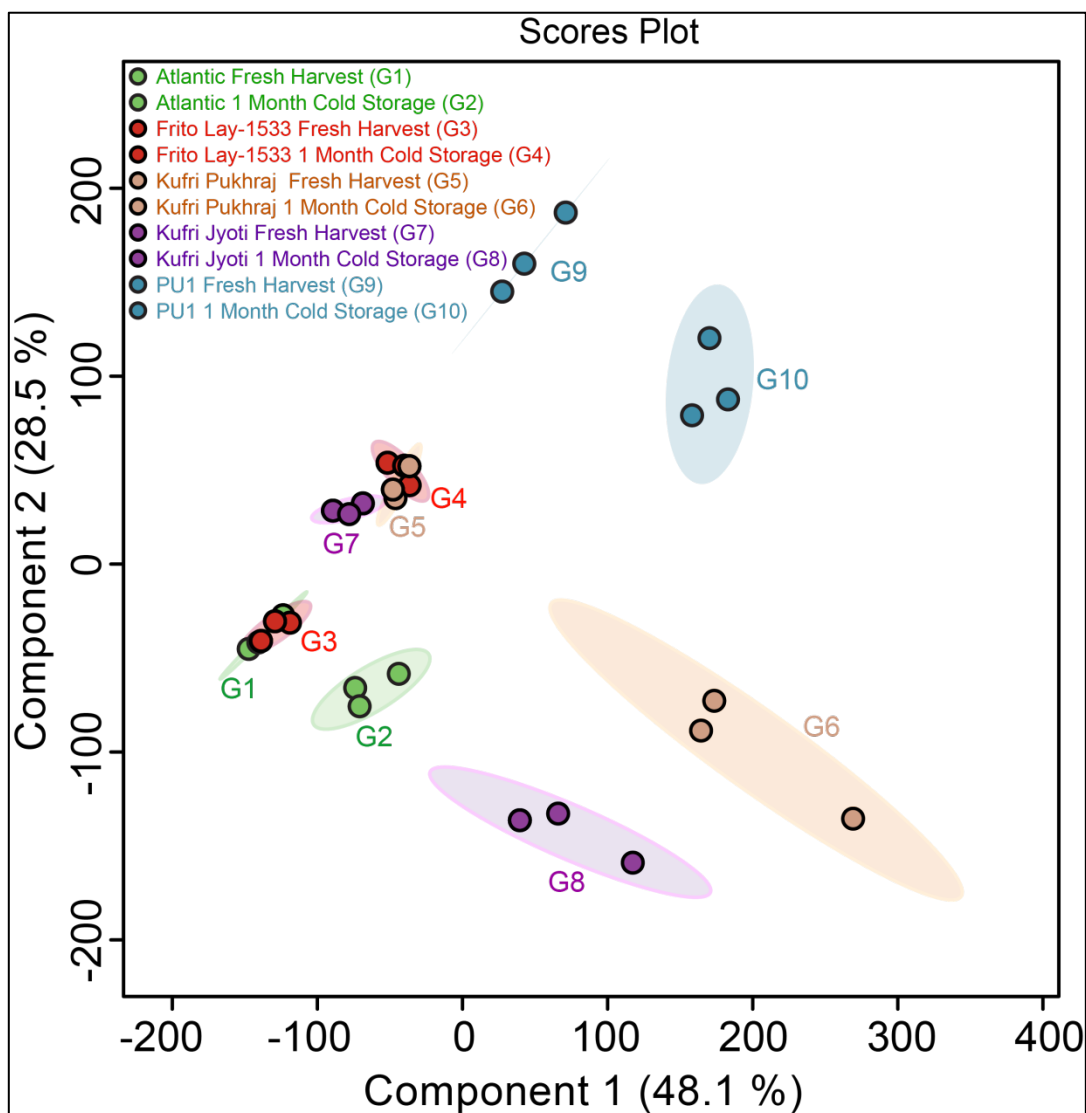


Figure 6.3: Scores plot of PLS-DA generated from MetaboAnalyst software for the different potato cultivars – Atlantic (G1, G2), Frito Lay-1533 (G3, G4), Kufri Pukhraj (G5, G6), Kufri Jyoti (G7, G8), and PU1 (G9, G10) at fresh harvest and one-month cold storage at 4°C. Three replicates were used for each treatment in all five different potato cultivar. Each ellipse on the score plot represents 95% confidence limits of a normal distribution for each group. Colour legends have been mentioned in the figure.

Further pair-wise comparison was performed to explore the effect of cold storage on the metabolic profiles of each cultivar separately. PLS-DA analysis between fresh harvest and cold-stressed tubers in processing cultivars showed a differential clustering of groups on their respective score plots with 84.2% variation in PC1 and 10% variation in PC2 for Atlantic cultivar (Figure 6.4A), 84.4% variation in PC1, and 9.5% variation in PC2 for Frito Lay 1533 cultivar. (Figure 6.4B). Similarly, pairwise PLS-DA analysis of non-processing cultivars between FH and CS showed 92.8% variation in PC1 and 3.8% variation in PC2 for Kufri Jyoti (Figure 6.4C), 90.4% variation in PC1 and 6.1% variation in PC2 for Kufri Pukhraj (Figure

6.4D), and 92.9% variation in PC1 and 3.4% variation in PC2 for local cultivar PU1 (Figure 6.4E). Such large differences observed due to variation in the metabolite content in the different cultivars at the two time-points could be attributed to the genetic make-up of each cultivar used in the present study. From the PLS-DA model analysis, maximum segregation was observed between FH and CS classes in non-processing cultivars compared to processing cultivars. This group separation was based on the first two components, which display a distinct demarcation between the FH and CS groups. VIP scores obtained from the PLS-DA model were used to identify the discriminatory metabolites contributed most to the group separation in the PLS-DA models for each cultivar under FH and CS conditions. The VIP score plot, which is based on PLS loadings, reflects the influence of each variable on the overall model. The metabolites having a VIP score ≥ 1 were considered important in classification. The discriminatory metabolites were organised in the descending order of the VIP score in component 1. Based on the VIP score (VIP ≥ 1), the key metabolites discriminating FH from CS tubers in Atlantic cultivar were fructose, glucose, sucrose, asparagine, and methanol (Figure 6.5A). While sucrose, citrate, glucose, fructose, and glutamine were the most important metabolites that altered significantly in Frito Lay-1533 (Figure 6.5B) upon cold storage. The discriminatory metabolites (VIP ≥ 1) altered significantly under cold storage in Kufri Jyoti (Figure 6.5C) include fructose, glucose, sucrose, malate, and asparagine. In Kufri Pukhraj (Figure 6.5D), the discriminatory metabolites identified were fructose, glucose, malate, asparagine, methanol, proline, and glutamine. Similarly, in PU1 cultivar (Figure 6.5E), the discriminatory metabolites include methanol, fructose, sucrose, glucose, 4-aminobutyrate, and proline. We observed most of the discriminating metabolites majorly belong to starch and amino acid metabolism pathways. These metabolites are largely contributing to the differential clustering of score plots in the PLS-DA analysis.

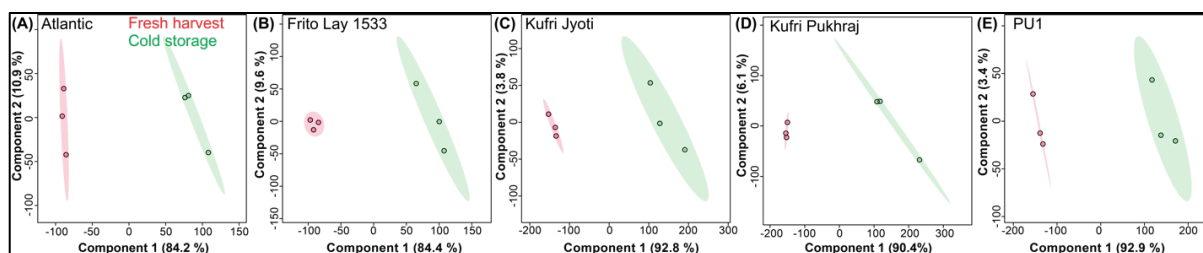


Figure 6.4: PLS-DA scores plots for pair-wise comparison analysis of metabolites between potato tubers samples at fresh harvest (Red) and cold storage at 4°C for 1 month (Green) in A) Atlantic, B) Frito Lay-1533, C) Kufri Jyoti, D) Kufri Pukhraj, and E) PU1. Each ellipse on the score plot represents 95% confidence limits of a normal distribution for each group.

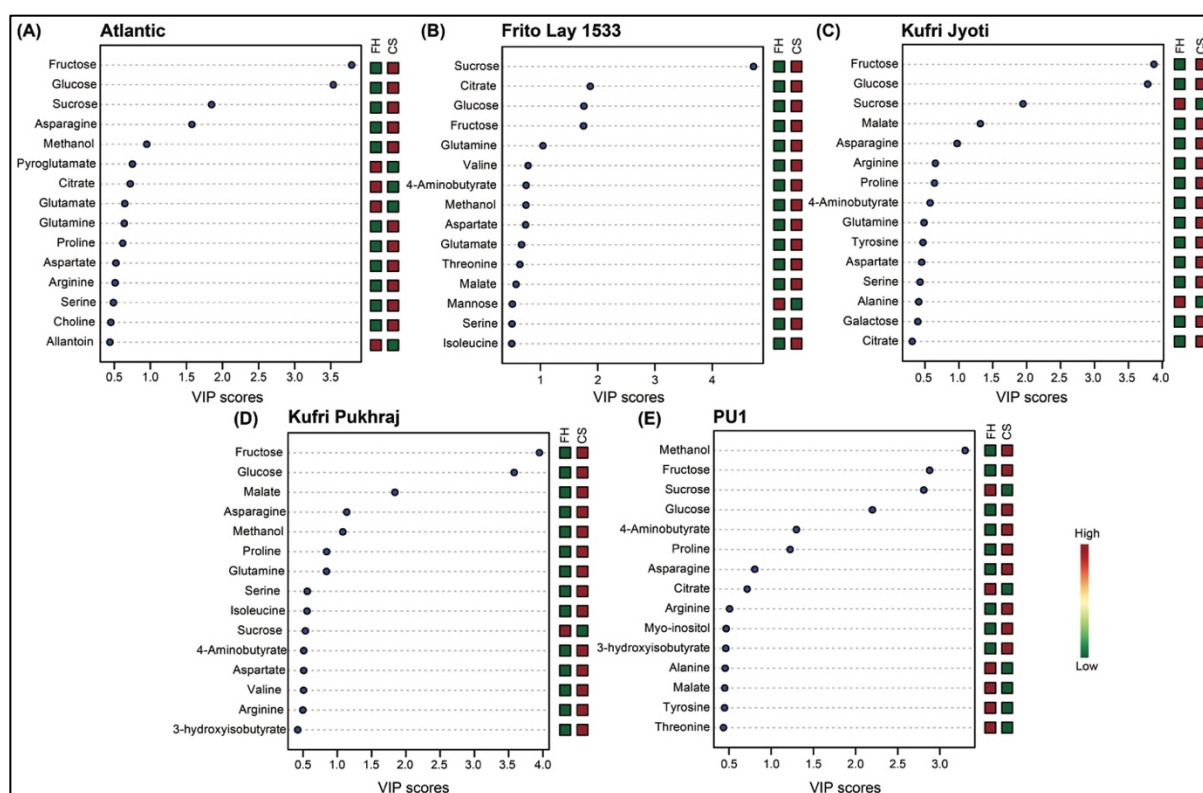


Figure 6.5: VIP scores obtained after pair-wise PLS-DA analysis for A) Atlantic, B) Frito Lay-1533, C) Kufri Jyoti, D) Kufri Pukhraj, and E) PU1. A VIP score of ≥ 1.0 is considered significant. VIP plots display the most significantly altered metabolites detected by PLS-DA model analyses arranged in descending order of VIP score in component 1. The coloured boxes on the right indicate the relative concentrations of the corresponding metabolite in each condition used. The relative concentration of metabolites has been represented by a coloured scale from green to red, indicating the low and high.

6.6.3 Metabolic perturbations in processing, non-processing, and local cultivars under CS treatments.

The effects of CS treatment were further evaluated for each cultivar with a volcano plot analysis to identify the metabolites whose levels vary significantly under CS condition.

Volcano plot analysis uses fold change and p-values to derive the significantly dysregulated metabolites.

Cold storage of Atlantic cultivar resulted in significant up-regulation of fructose, glucose, galactose, and methanol, while fumarate and glutamate (Figure 6.6A) were found to be significantly downregulated in cold storage treatment. Increased levels of fructose, glucose, sucrose, galactose, fumarate, trigonelline, citrate, aspartate, and glutamate were observed in Frito lay upon CS treatment, whereas the levels of mannose and 3-hydroxyisobutyrate (Figure 6.6B) were reduced substantially. Similarly, in non-processing Kufri Jyoti cultivar, the levels of glucose, fructose, ascorbate, mannose, galactose, aspartate, malate, fumarate, leucine, proline, and serine were increased, while sucrose and alanine (Figure 6.6C) decreased in CS treatment. CS treatment in Kufri Pukhraj cultivar was associated with significantly elevated fructose, glucose, 3-hydroxyisobutyrate, ascorbate, mannose, malate, leucine, aspartate, serine, proline, isoleucine, adenosine, and arginine; while as chlorogenate and formate (Figure 6.6D) levels undergo down-regulation. In PU1, cultivar, formate, tryptophan, and sucrose were significantly decreased while 3-hydroxyisobutyrate, methanol, fructose, glucose, proline, 4-aminobutyrate, trigonelline, myo-inositol, arginine, aspartate, uridine, and sn-glycero-3-phosphocholine (Figure 6.6E) were increased upon cold storage treatment.

From the volcano plots and VIP score plots, it was apparent that the number of significantly dysregulated metabolites varies in each cultivar under CS condition, with mostly metabolites showing upregulation in their concentration levels under CS conditions. Further, it was also observed that non-processing cultivars exhibit more metabolite alteration than processing cultivars. All these results indicate a large change in metabolite content occurs during cold storage. A union set of significant metabolites identified from volcano plot analysis (Figure 6.6) and VIP score (Figure 6.5) were used to culminate critical metabolites (Figure 6.7) associated with CS treatment in each cultivar to reduce the chances of losing any important metabolites of interest (Feng et al., 2016). The levels of specific metabolites were observed to be varying distinctly depending on potato cultivars (Figure 6.7), suggesting that the particular metabolite could play a crucial role in predicting the CIS potential of potato cultivars. The molecular events controlling such metabolic perturbations in potato tubers after cold storage need to be investigated in further detail in the future.

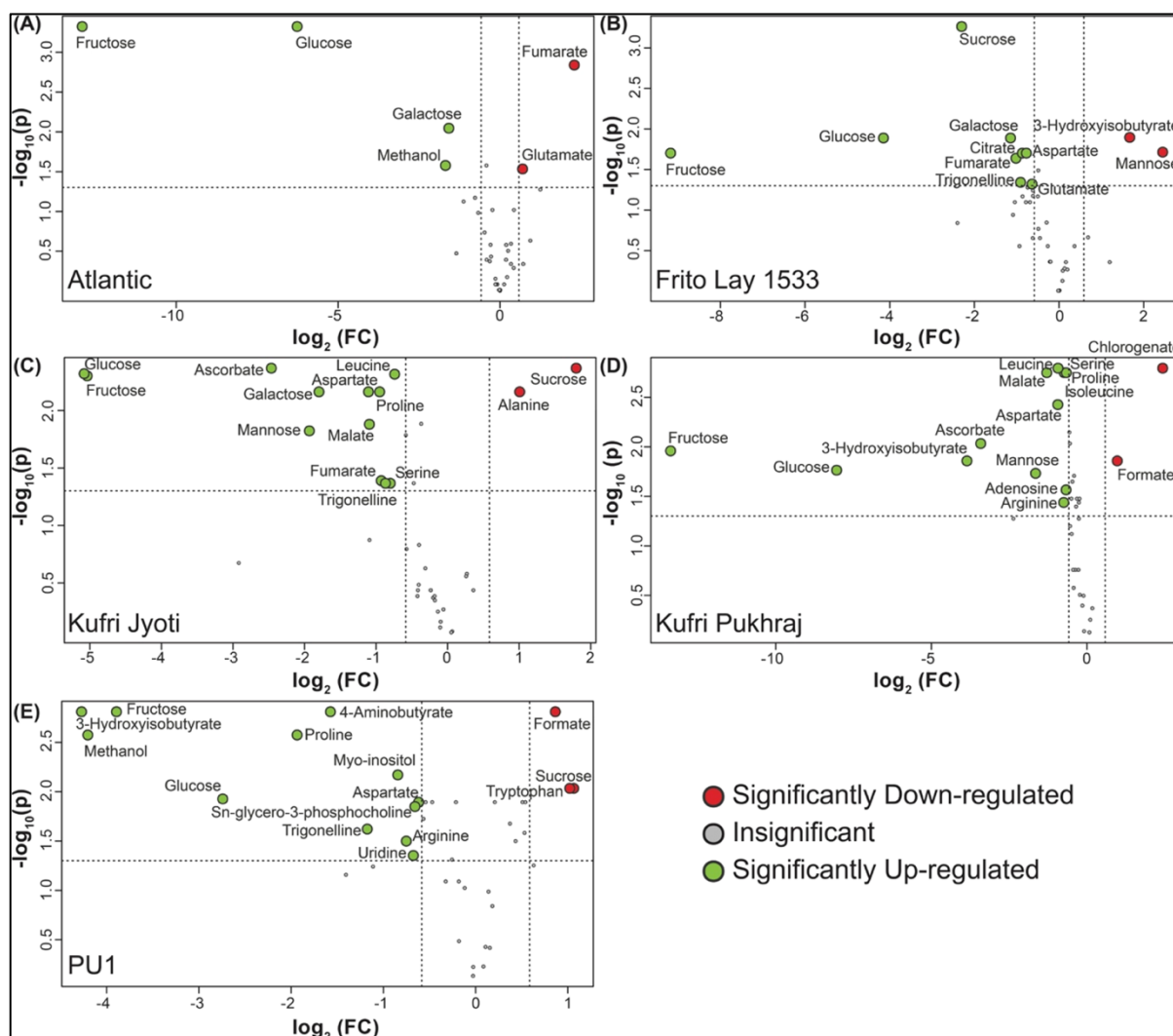


Figure 6.6: Volcano plots pairwise metabolite comparison between freshly harvested (FH) and cold storage (CS) tubers. The different cultivars used for the study have been depicted as A) Atlantic, B) Frito Lay-1533, C) Kufri Jyoti, D) Kufri Pukhraj, and E) PU1. In the volcano plot, vertical dotted lines indicate the threshold of ± 1.5 -fold changes in concentration of metabolites on a \log_2 scale, and the horizontal dotted line corresponds to the threshold of FDR-corrected p-value significance ($p\text{-value} \leq 0.05$) on a \log_{10} scale. The significantly down-regulated metabolites upon cold storage have been marked in red and the ones up-regulated have been marked in green.

Cold storage has been known to induce modifications in plants metabolome (Bustamante et al., 2016; Hatoum, Annaratone, Hertog, Geeraerd, & Nicolai, 2014). The alterations in the levels of metabolites upon cold storage treatment can be correlated with the differential enzyme activities or cold signalling and altered gene expression patterns. For instance, cold storage induced alterations in metabolite levels observed in six peach fruits were reflected as transcriptional and/or posttranscriptional responses (Bustamante et al., 2016). Integration of metabolomics with quantitative genetics found an effective approach to identify the candidate genes underlying metabolite variations, which offers trait-specific markers to

improve commercially important traits (Kumar, Bohra, Pandey, Pandey, & Kumar, 2017b). Previous studies have reported the association of phenotypic traits with primary metabolites in potatoes through metabolic quantitative trait loci (QTL) analyses and have revealed that the CIS trait is expected to be associated with primary metabolites (Carreno-Quintero et al., 2012). Untargeted GC-MS metabolic profiling for 26 starch and CIS-related traits on 97 potato genotypes detected 139 polar metabolites, of which QTL were identified for 72 % of the detected compounds. It was also observed that the QTLs for starch phosphorylation co-localize with mQTLs of few amino acids (Carreno-Quintero et al., 2012). Therefore, we propose that the metabolite variability obtained in the present study, when combined with the genetic information, would facilitate in the discovery of metabolite biomarkers for CIS and help in determining potato processing related traits in the cultivars used in the study.

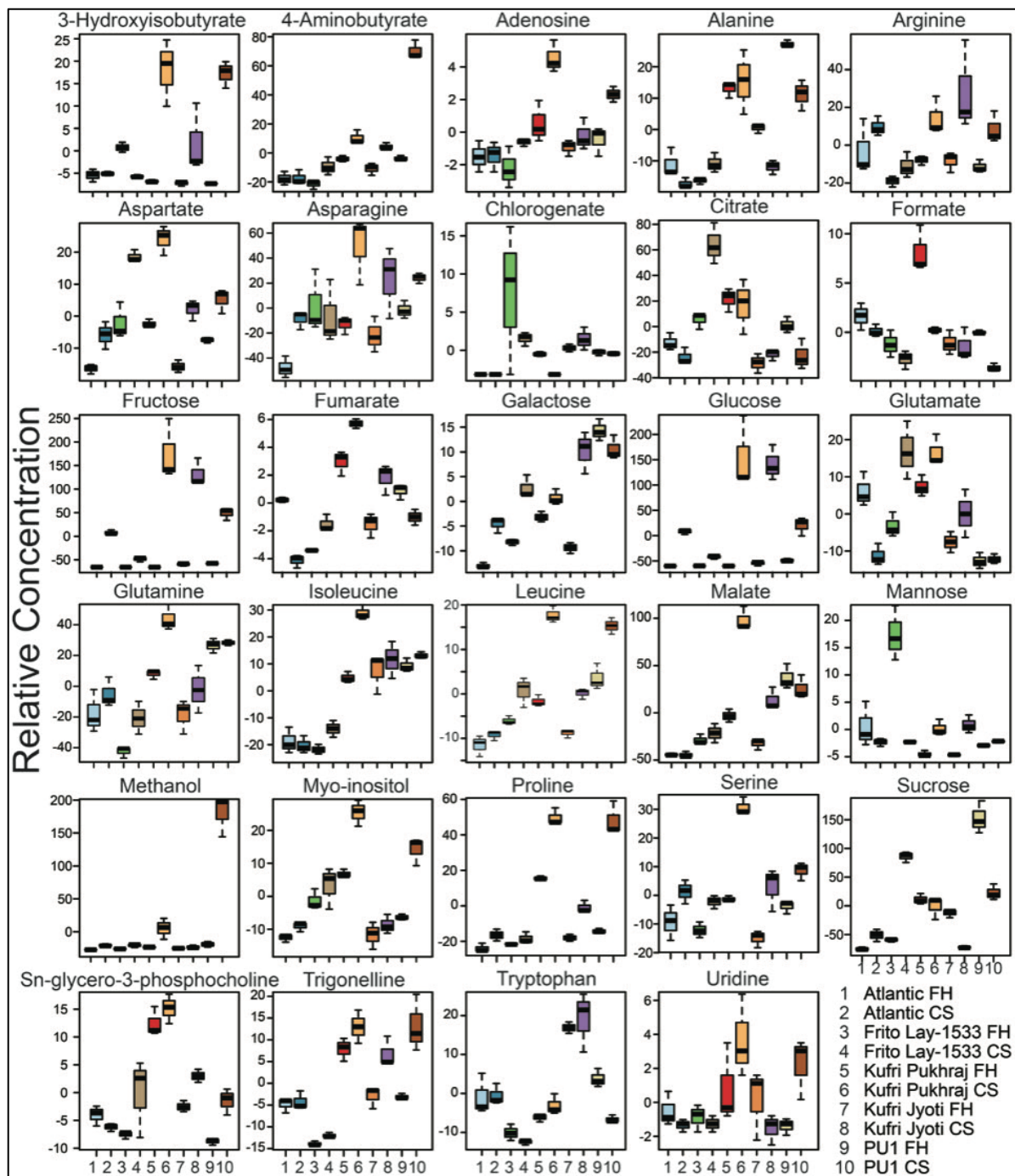


Figure 6.7: Box-Whisker plots displaying the comparative alteration in the concentration of discriminating metabolites (identified from VIP score plot and Volcano plot analysis) for both freshly harvested and cold storage tuber samples across replicates and in different potato cultivars used in the study. FH – fresh harvest and CS – cold storage at 4 °C.

6.6.4 Metabolic correlation network analysis

Metabolite-metabolite correlation analysis (MMCA) among the identified metabolites was conducted in all potato cultivars using Pearson’s correlation coefficient analysis at both the time points – fresh harvest and cold storage. The aim of this correlation analysis was to

identify metabolites that are correlated to each other in a given treatment. Specifically, metabolite-metabolite correlations within FH and CS samples were compared for each cultivar. The correlation plots for each treatment within a particular cultivar were combined in a single plot (separated by the diagonal) to visualize the differences in metabolite-metabolite correlations within FH and CS conditions, with the upper-right half of the plot marked in white corresponding to correlations in FH samples while lower-left half of the plot marked in light blue corresponding to metabolite-metabolite correlations under CS condition for each cultivar. A total of 1444 correlations were analysed in each treatment within each cultivar. The number of significant correlation ($p\text{-value} \leq 0.05$) for each treatment within each cultivar has been tabulated in Table 6.2. Overall 49, 51, 36, 48, and 45 significant metabolite-metabolite correlation were identified in fresh harvest tuber samples for Atlantic, FL-1533, Kufri Pukhraj, Kufri Jyoti, and PU1, respectively (upper-right half of the plot marked with a white triangle in Figure 6.8). Upon CS treatment, the number of significant correlations for Atlantic, Frito Lay 1533, Kufri Pukhraj, Kufri Jyoti, and PU1 changed to 55, 53, 52, 53, and 31 (lower-left half of the plot marked with a blue triangle in Figure 6.8). The number of positive versus negative correlations also changed upon CS treatment in all cultivars (Table 6.2). Amino acids and sugars display the most number of significant metabolite-metabolite correlations, most probably indicating their centrality in primary metabolism. Combining significant metabolic correlations and metabolic alterations in FH and CS tuber profiling might help to predict the CIS status of the particular potato genotype. Previous studies have shown that the metabolic variability and metabolite-metabolite correlations associate with market classes such as russet, red, yellow, chip, and specialty in 60 unique potato genotypes. These investigations concluded that metabolite diversity and correlations data can support the potential to breed new cultivars for improved health traits (Y. Lin et al., 2013). Thus, the metabolite alterations (Figure 6.7) and metabolite-metabolite correlations (Figure 6.8) obtained, especially after cold storage combined with the correlative behaviour of genes at the transcript level, will further help in validating the role of a specific group of metabolites in the CIS phenomenon.

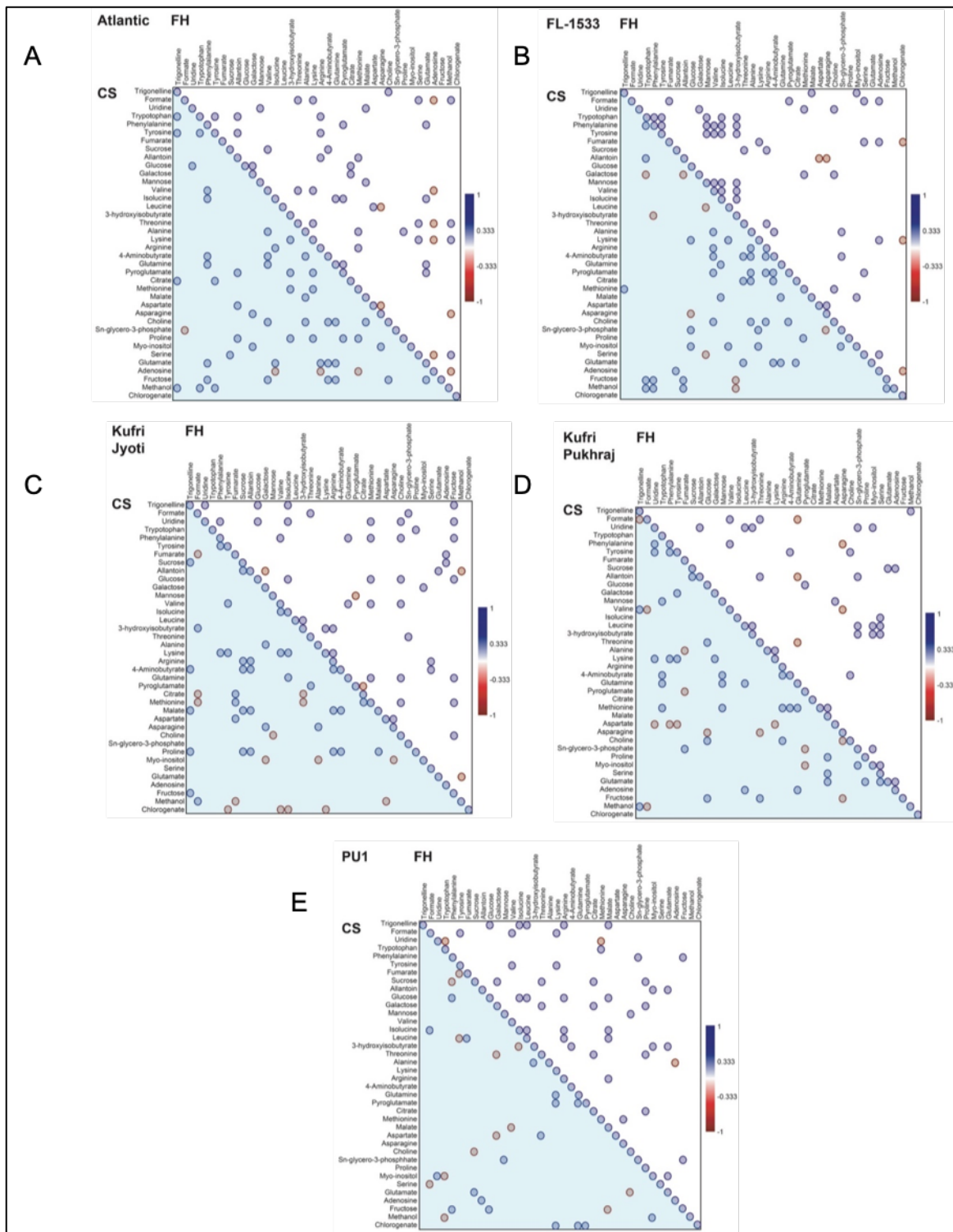


Figure 6.8: Correlation plots between fresh harvest (FH; upper-right half of the plot marked in white) and cold storage (CS; lower-left half of the plot marked in light blue) metabolites for A) Atlantic, B) Frito Lay-1533, C) Kufri Jyoti, D) Kufri Pukhraj, and E) PU1. Both the upper half and lower half diagonal correlation plots are independent and must be read independently. Blue and red circles represent positively and negatively significant ($p\text{-value} \leq 0.05$) metabolite-metabolite correlations, respectively.

Table 6.2: The number of correlation observed for the metabolites in the different cultivars.

Cultivar – time point	Positive	Negative	Total
Atlantic FH	40	9	49
Atlantic CS	51	4	55
Frito Lay 1533 FH	46	5	51
Frito Lay 1533 CS	44	9	53
Kufri Pukhraj FH	31	5	36
Kufri Pukhraj CS	37	15	52
Kufri Jyoti FH	43	5	48
Kufri Jyoti CS	38	15	53
PU1 FH	42	3	45
PU1 CS	18	13	31

6.7 Metabolic pathway analysis and potential metabolite insights of CIS

CIS, caused by the accumulation of reducing sugars, is a complex genetic trait that involves several metabolic pathways regulated by multiple candidate genes. This suggests that a group of metabolites could be involved in the regulation of the CIS phenomenon. To gain insight into the metabolic mechanism of CIS in potatoes, metabolic pathway analysis of significantly differential metabolites identified through volcano plot analysis ($q\text{-value} \leq 0.05$, $FC \geq 1.5$) and VIP scores plot ($VIP \geq 1$) were performed using KEGG, MetaboAnalyst, and reference pathways (Malone et al., 2006b; J. R. Sowokinos, 2001b) from the previous literature. Several physiological and biochemical mechanisms occur in potatoes during CIS, and the amount of sugar in potato tubers is influenced by several candidate genes operating in starch degradation, hexogenesis, and anaerobic respiration (J. R. Sowokinos, 2001b; H. Zhang et al., 2017). Metabolic pathway analysis (MetPa) of significantly altered metabolites revealed the metabolic pathways perturbed under CS in potato cultivars were alanine, aspartate, and glutamate metabolism; glyoxylate and dicarboxylate metabolism; amino acid metabolism; arginine and proline metabolism; glycine, serine, and threonine metabolism; the TCA cycle, fructose, and mannose metabolism, galactose metabolism, Tryptophan metabolism, glycolysis; Starch and sucrose metabolism; and Inositol phosphate metabolism (Table 6.3). All the identified pathways were then constructed using the KEGG metabolic network and have been pictorially depicted in Figure 6.9.

Table 6.3: Metabolic pathway analysis revealing the significantly impacted metabolic pathways in potato tubers under cold storage

Metabolic Pathways	Hits	Raw p	-log10(p)	FDR	Impact
Alanine, aspartate, and glutamate metabolism	7	1.33E-07	6.8748	6.34E-06	0.777
Starch and sucrose metabolism	2	0.078372	1.1058	0.43796	0.391
Arginine and proline metabolism	4	0.0024371	2.6131	0.033075	0.246
Fructose and mannose metabolism	1	0.35292	0.45233	1	0.037
Glycine, serine, and threonine metabolism	3	0.031243	1.5052	0.26983	0.181
Citrate cycle (TCA cycle)	3	0.0078731	2.1039	0.093493	0.178
Tryptophan metabolism	1	0.39415	0.40434	1	0.172
Arginine biosynthesis	5	2.23E-05	4.6512	0.00053029	0.170
Pyruvate metabolism	2	0.078372	1.1058	0.43796	0.155
Glyoxylate and dicarboxylate metabolism	6	1.91E-05	4.7187	0.00053029	0.138
Butanoate metabolism	2	0.049273	1.3074	0.36007	0.136
Aminoacyl-tRNA biosynthesis	11	4.68E-10	9.3301	4.44E-08	0.111
Inositol phosphate metabolism	1	0.45729	0.33981	1	0.103
Tyrosine metabolism	1	0.32394	0.48954	1	0.070
Glycerophospholipid metabolism	1	0.55528	0.25549	1	0.059
Carbon fixation in photosynthetic organisms	3	0.0090535	2.0432	0.095564	0.058
Glutathione metabolism	2	0.11157	0.95247	0.58882	0.050
Galactose metabolism	5	0.00018372	3.7358	0.0034907	0.049
Phenylpropanoid biosynthesis	1	0.5351	0.27156	1	0.047
Phosphatidylinositol signalling system	1	0.43284	0.36368	1	0.033
Valine, leucine and isoleucine degradation	3	0.042013	1.3766	0.33261	0.028
Pyrimidine metabolism	2	0.1939	0.71243	0.87715	0.028
Flavonoid biosynthesis	1	0.6441	0.19105	1	0.021
Purine metabolism	2	0.39372	0.40481	1	0.001
Nitrogen metabolism	3	0.0017061	2.768	0.027013	0.000
Cyanoamino acid metabolism	3	0.016466	1.7834	0.15643	0.000
Valine, leucine, and isoleucine biosynthesis	2	0.078372	1.1058	0.43796	0.000
Monobactam biosynthesis	1	0.15916	0.79816	0.7958	0.000
Lysine biosynthesis	1	0.17724	0.75143	0.84191	0.000
Nicotinate and nicotinamide metabolism	1	0.24589	0.60926	0.9733	0.000
Cysteine and methionine metabolism	2	0.25786	0.58861	0.97987	0.000
Sulfur metabolism	1	0.2781	0.5558	1	0.000
Amino sugar and nucleotide sugar metabolism	2	0.29021	0.53729	1	0.000
Sphingolipid metabolism	1	0.30898	0.51007	1	0.000
beta-Alanine metabolism	1	0.32394	0.48954	1	0.000
Phenylalanine, tyrosine, and tryptophan biosynthesis	1	0.3807	0.41942	1	0.000
Ascorbate and aldarate metabolism	2	0.054707	1.262	0.37122	0.216

6.8 Discussion

6.8.1 Sugar metabolism under cold storage condition

Previous studies report that the low temperature mediated accumulation reducing of sugars in potato tubers is derived mostly from the degradation of starch, thereby leading to an imbalance between starch degradation and sucrose metabolism in tubers. The enzymes associated with these processes such as starch synthases, pyruvate kinase, invertase, AGPase, sucrose synthase, alpha-amylase, beta-amylase, sucrose phosphate synthase, etc., have been studied extensively at both biochemical and transcript levels (Brummell et al., 2011; X. Chen, Salamini, & Gebhardt, 2001; Datir et al., 2019; Jansky & Fajardo, 2014; C. Menéndez, 2002; J. R. Sowokinos, 2001b; Joseph R. Sowokinos, Hayes, & Thill, 2018; H. Zhang et al., 2017). The increased levels of sucrose, glucose, and fructose in Atlantic and Frito Lay 1533 tubers under cold storage condition indicates that sucrose has been broken down; however, the replenishing of sucrose concentrations through the breakdown of starch was occurring more rapidly in these cultivars. The sucrose concentration was significantly reduced in Kufri Jyoti, and local cultivar PU1 and slightly decreased in Kufri Pukhraj (Figure 6.7).

While cold-treatment largely enhances the levels of reducing sugars in all cultivars, the increase was significantly prominent in the non-processing cultivars, which was quite apparent compared to processing cultivars used in this study. Previous studies have also shown that the accumulation of reducing sugars in tubers of cold-sweetening susceptible and cold-sweetening resistant potato cultivars was found to be under the control of vacuolar invertase and invertase inhibitor (Brummell et al., 2011; Datir et al., 2012b, 2019). The significantly increased levels of reducing sugars (glucose and fructose) and decreased levels of sucrose in the non-processing cultivars – Kufri Pukhraj, Kufri Jyoti, and PU1 under CS condition (Figure 6.7)- suggest enhanced vacuolar invertase activity in these cultivars (Y. Lin et al., 2013). Indeed, decreased levels of RS in cold-sweetening resistant cultivars upon cold storage have been previously reported, which was accompanied by enhanced expression of the vacuolar invertase inhibitor (Datir et al., 2019). The difference in levels of reducing sugars in different cultivars after CS conditions could also be attributed to allelic variations in the vacuolar invertase inhibitor gene in these cultivars (Datir et al., 2019). However, these observations need to be further validated at the transcript level using a qRT-PCR expression of vacuolar invertase inhibitor gene before and after cold storage in these cultivars.

6.8.2 Amino acid metabolism under cold storage condition

Amino acid metabolism plays a crucial role during abiotic stress conditions (heat, cold, and drought) in plants and has been found to minimize the adverse effects of cold temperature (Hasanuzzaman et al., 2019; Krasensky & Jonak, 2012; Zhao et al., 2009). We also observed the dominance of amino acids in the metabolite-metabolite correlation analysis, where maximum significant correlations were shown by amino acids, thereby hinting towards an important role of amino acids in potato tubers under cold storage conditions. The amino acids that were found to be significantly dysregulated under CS conditions were isoleucine, glutamate, glutamine, leucine, alanine, arginine, proline, tryptophan, aspartate, asparagine, and serine. Accumulation of organic osmolytes such as proline and 4-aminobutyrate (GABA) was observed in all cultivars under CS condition, but tubers of non-processing cultivars (Kufri Jyoti, Kufri Pukhraj, and PU1) exhibited relatively higher levels of both these metabolites (Figure 6.7) compared to processing cultivars (Atlantic and Frito Lay 1533). Previously it has been reported that cold-stress induced accumulation of proline and GABA is associated with increased resistance to chilling stress in some fruits species (Mazzucotelli, Tartari, Cattivelli, & Forlani, 2006; Shang, Cao, Yang, Cai, & Zheng, 2011; Z. Zhang et al., 2017). Besides acting as an osmoprotectant, proline also acts as a metal chelator, antioxidative defense molecule, and signaling molecule during stress (Hayat et al., 2012; Liang, Zhang, Natarajan, & Becker, 2013). The GABA shunt pathway was up-regulated under CS condition, as seen by the elevated levels of GABA and glutamate (a predominant source of GABA in plants). GABA treatment has been demonstrated as a promising approach not only for reducing the enzymatic browning but also for maintaining the quality of fresh-cut potatoes (Gao, Zeng, Ren, Li, & Xu, 2018).

Arginine metabolism is known to be involved in abiotic stress conditions in potato tubers. Arginine levels showed a significant increase in all potato cultivars under CS condition (Figure 6.7). Putrescine, an arginine derived molecule, is known to play an important role in inducing cold-acclimated freezing tolerance in potatoes (Kou et al., 2018). The branched-chain amino acids – leucine and isoleucine – also showed an accumulated in all cultivars after cold storage. Previous studies report the accumulation of branched-chain amino acids under abiotic stress conditions, including cold stress (Obata & Fernie, 2012). However, to date, there has been no conclusive evidence on the role of several of these metabolites in the CIS process of potato tubers. This is particularly important because a comprehensive analysis of the biochemical pathways affected during CIS remains had not been performed earlier.

From the metabolomics data, significantly elevated levels of glutamine and asparagine (one of the precursors for Maillard reaction) were observed in non-processing cultivars – Kufri

Pukhraj, Kufri Jyoti, and PU1 after cold storage. Among all cultivars, Kufri Pukhraj displayed higher levels of glutamine and asparagine after CS. The increased levels of glutamine can be attributed to the enhanced expression of glutamine synthetase (Roessner-Tunali et al., 2003). Enzymes, branched-chain amino acid aminotransferase and glutamine synthetase, involved in glutamine biosynthesis have been found to be potentially involved in governing potato tuber quality traits. The alterations in the contents of amino acid metabolite have also been shown to be distinct in different potato cultivars. In addition to RS content, the total and individual amino acid content, the asparagine content, organic acid levels, and other metabolites have been found to be essential processing parameters in potatoes (Muttucumaru et al., 2017). Levels of acrylamide have previously been associated with RS and free-asparagine content (Raffan & Halford, 2019). Breeders aim to select potato cultivars with low concentrations of RS and are asparagine-free for processing purposes to mitigate the health concerns related to acrylamide. In this regard, Frito Lay 1533 was found to be the best cultivar for processing purposes with the lowest amount of reducing sugars and asparagine under CS condition. Similarly, Kufri Pukhraj was found with the highest content of RS and asparagine upon cold storage as compared to other cultivars.

6.8.3 TCA cycle metabolites under cold storage condition

Organic acids, particularly TCA cycle intermediates, are known to play numerous and diverse functions within and beyond cellular metabolism in plants (W. F. Zhang et al., 2019; Y. Zhang & Fernie, 2018). From our metabolomics data, the TCA cycle intermediates, namely, citrate, malate, and fumarate were observed to be significantly dysregulated upon CS. The levels of citrate and fumarate were significantly up-regulated in Frito Lay 1533, whereas malate levels were significantly increased in Kufri Jyoti and Kufri Pukhraj upon CS (Figure 6.7). Previous reports have shown the citrate levels are enhanced during low-temperature exposure in banana, tomato, and Ponkan fruits, thereby suggesting that accumulation of citrate may be involved in cold tolerance in these fruits (Bugaud, Alter, Daribo, & Brillouet, 2009; Q. Lin et al., 2016; W. F. Zhang et al., 2019). The increased expression of ATP-citrate synthase and isocitrate dehydrogenase promoted by cold stress and hence the increased content of citrate after cold storage at 4°C for 28 days has been reported in tomato fruits (W. F. Zhang et al., 2019). Cold storage resulted in an increase in multiple isoforms of malate dehydrogenase, isocitrate dehydrogenase, and 2-oxoglutarate dehydrogenase genes in cold-stored (4°C) litchi fruit, which accelerated the fruit senescence and hence, might be associated with cold susceptibility or respiratory burst (Yun et al., 2016). Besides having an important role in cold

tolerance, citrate and malate have been commonly used in the food industry as chemical inhibitors to prevent or reduce enzymatic browning. The discoloration (brown color) of fresh-cut potato slices after mechanical operations has been attributed to the action of a group of enzymes called polyphenol oxidases (PPO) (Severini, Baiano, De Pilli, Romaniello, & Derossi, 2003). Citric acid is a popular anti-browning agent due to its inhibitory effect on PPO via a reduction in pH. In addition to this, citrate chelates copper at the active site of PPO and thereby inhibits its activity (Suttirak & Manurakchinakorn, 2010). Since the levels of citrate were significantly elevated in Frito Lay 1533 compared to other cultivars under CS, it implies that increased citrate levels in FL-1533 might be possibly associated with CIS resistance with good processing attributes. Therefore, potato cultivars varying in their CIS ability could be further examined for the quantitative variations in citrate and malate levels.

Our analysis revealed several metabolites such as fumarate, adenosine, sn-glycero-3-phosphocholine, 3-hydroxyisobutyrate, trigonelline, and chlorogenate were significantly dysregulated under cold storage conditions, indicating the important effect of these metabolites in cold stress and CIS process. Significantly elevated levels of methanol were observed in PU1, Kufri Pukhraj, and Atlantic cultivar after CS. Processed potato tuber textural properties are affected by several factors, including starch content and distribution within the tuber, cell size, cell-wall structure and composition, the breakdown of the cell wall middle lamella during cooking, and the correlation between pectin methyltransferase activity and level of pectin methyl esterification (Ross et al., 2011). After saponification, the proportion of released methanol content is the measure of the degree of pectin methylation for the cell wall and was found indirectly associated with the textural properties of potato tuber (Ross et al., 2011). Therefore, the methanol content can be used as a potential marker for textural differences in potato tubers. However, we cannot rule out the possibility that because methanolic extracts have been used for the metabolomics studies, the methanol content found in these tubers might represent the exogenous levels in NMR. Levels of Myo-inositol, which is known to play important roles in abiotic stresses, including cold stress in plants, was significantly up-regulated in local cultivar PU1 upon CS (Valluru & Van den Ende, 2011). However, more studies are required to understand further the function of these metabolites in the CIS process.

CIS is a complex trait involving a specific group of metabolites in a network of intricate metabolic pathways. To gain further insights into the CIS phenomenon, the importance and the expression pattern of these metabolites in these cultivars after CS should be investigated further by identifying candidate genes that are putatively associated with the biosynthesis of these metabolites. Further, it is important to determine the chemical composition of tubers to identify

and develop potato genotypes with good processing attributes (Pal, Bhattacharya, Konar, Mazumdar, & Das, 2008). Earlier reports have mentioned that different primary metabolites have been used as biomarkers in quality assessment analysis for predicting agronomically essential phenotypes such as black spot bruising and chip quality (Inostroza-Blancheteau et al., 2018; Steinfath et al., 2010).

One of the current and significant challenges in the potato industry is the maintenance of tuber quality during storage. Thus, the biochemical information and metabolite accumulation in response to cold stress in potato tubers could be used to select candidate metabolites for predicting the magnitude of cold-induced sweetening of different potato genotypes. Moreover, such predictive candidate metabolites can be used in selection for potato breeding and for tailoring storage conditions for the harvested tubers.

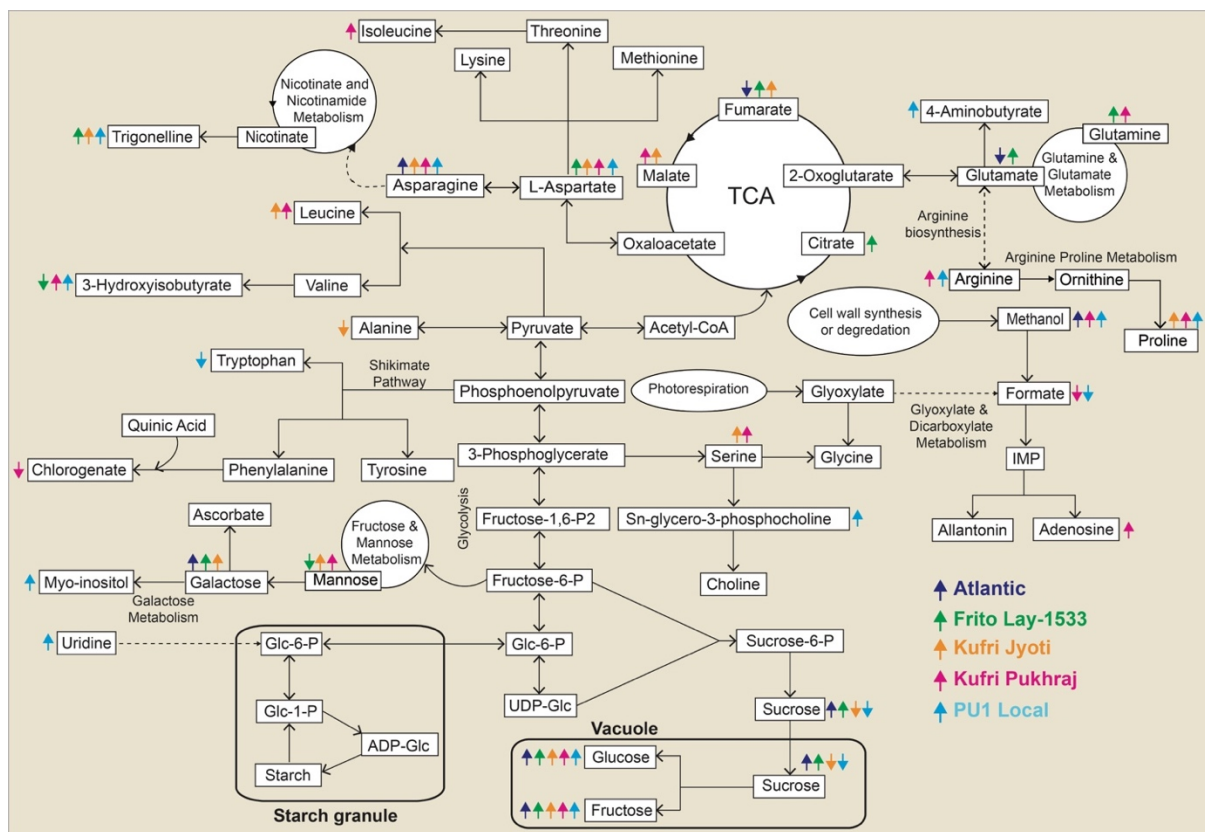


Figure 6.9: Pictorial depiction of metabolic pathways affected during cold-induced sweetening in the different cultivars of potato. The significant metabolites altered under cold storage have been marked with arrows wherein an ↑ indicates upregulated metabolites, and the ↓ arrow indicates down-regulated metabolites. TCA – tricarboxylic acid.

6.9 Conclusions

In this study, we have identified the alterations in metabolic events in different potato cultivars after cold storage conditions. Cold-treatment has a major influence on the overall content of several metabolites such as glucose, fructose, sucrose, asparagine, glutamine, citrate, malate, proline, GABA, etc. These candidate metabolites can serve as ideal targets to unravel the biological basis of CIS trait in potato tubers and could thus help in improving the agronomically important phenotypes. For instance, the low levels of acrylamide precursors (glucose, fructose, and asparagine) accompanied by a high level of citrate (anti-browning agent) observed in processing cultivar (FL-1533) and high levels of acrylamide precursors observed in non-processing cultivars (Kufri Pukhraj and Kufri Jyoti) can be used for the selection and breeding of potato tubers with good health benefits and well processing characteristics. Similarly, the alterations in the contents of GABA and glutamine can be utilized in determining the enzymatic browning of fresh-cut potatoes and tuber quality, respectively. The metabolite-metabolite correlations observed between sugars, amino acids, and organic acids suggest the close biochemical relationship and the interdependence of their associated metabolic pathways such as glycolysis, amino acid metabolism, and the tricarboxylic acid (TCA) cycle in the CIS process of potato tubers. The phenomenon of CIS is associated with the presence or absence of a specific group of metabolites, and the relative content of these metabolites may have a substantial effect on the CIS status of potato tubers. Taken together, the results from our study have helped in the identification of metabolic events that can be potentially used in the selection and development of CIS resistant potato genotypes with good processing attributes and health benefits.

References

- Ali, A., Ahmad Nasir, I., Muzaffar, A., Shahzad Iqbal, M., Qayyum Rao, A., & Husnain, T. (2016). Screening of Potato Germplasm Resistant Against Low Temperature Sweetening. *Journal of Food Quality*. <https://doi.org/10.1111/jfq.12196>
- Amjad, A., Javed, M. S., Hameed, A., Hussain, M., & Ismail, A. (2020). Changes in sugar contents and invertase activity during low temperature storage of various chipping potato cultivars. *Food Science and Technology*. <https://doi.org/10.1590/fst.00219>
- Baldwin, S. J., Dodds, K. G., Auvray, B., Genet, R. A., Macknight, R. C., & Jacobs, J. M. E. (2011). Association mapping of cold-induced sweetening in potato using historical phenotypic data. *Annals of Applied Biology*. <https://doi.org/10.1111/j.1744-7348.2011.00459.x>
- Bianchi, G., Scalzo, R. Lo, Testoni, A., & Maestrelli, A. (2014). Nondestructive Analysis to Monitor Potato Quality during Cold Storage. *Journal of Food Quality*, 37(1), 9–17. <https://doi.org/10.1111/jfq.12068>
- Brummell, D. A., Chen, R. K. Y., Harris, J. C., Zhang, H., Hamiaux, C., Kralicek, A. V., & McKenzie, M. J. (2011). Induction of vacuolar invertase inhibitor mRNA in potato tubers contributes to cold-induced sweetening resistance and includes spliced hybrid mRNA variants. *Journal of Experimental Botany*, 62(10), 3519–3534. <https://doi.org/10.1093/jxb/err043>
- Bugaud, C., Alter, P., Daribo, M. O., & Brillouet, J. M. (2009). Comparison of the physico-chemical characteristics of a new triploid banana hybrid, FLHORBAN 920, and the Cavendish variety. *Journal of the Science of Food and Agriculture*, 89(3), 407–413. <https://doi.org/10.1002/jsfa.3459>
- Bustamante, C. A., Monti, L. L., Gabilondo, J., Scossa, F., Valentini, G., Budde, C. O., ... Drincovich, M. F. (2016). Differential metabolic rearrangements after cold storage are correlated with chilling injury resistance of peach fruits. *Frontiers in Plant Science*, 7(September2016), 1478. <https://doi.org/10.3389/fpls.2016.01478>
- Carreno-Quintero, N., Acharjee, A., Maliepaard, C., Bachem, C. W. B., Mumm, R., Bouwmeester, H., ... Keurentjes, J. J. B. (2012). Untargeted metabolic quantitative trait loci analyses reveal a relationship between primary metabolism and potato tuber quality. *Plant Physiology*, 158(3), 1306–1318. <https://doi.org/10.1104/pp.111.188441>
- Chaparro, J. M., Holm, D. G., Broeckling, C. D., Prenni, J. E., & Heuberger, A. L. (2018). Metabolomics and Ionomics of Potato Tuber Reveals an Influence of Cultivar and Market Class on Human Nutrients and Bioactive Compounds. *Frontiers in Nutrition*, 5,

36. <https://doi.org/10.3389/fnut.2018.00036>
- Chen, X., Salamini, F., & Gebhardt, C. (2001). A potato molecular-function map for carbohydrate metabolism and transport. *Theoretical and Applied Genetics*, *102*(2–3), 284–295. <https://doi.org/10.1007/s001220051645>
- Chen, Xia, Song, B., Liu, J., Yang, J., He, T., Lin, Y., ... Xie, C. (2012). Modulation of gene expression in cold-induced sweetening resistant potato species *Solanum berthaultii* exposed to low temperature. *Molecular Genetics and Genomics*. <https://doi.org/10.1007/s00438-012-0688-6>
- Chong, J., Soufan, O., Li, C., Caraus, I., Li, S., Bourque, G., ... Xia, J. (2018). MetaboAnalyst 4.0: Towards more transparent and integrative metabolomics analysis. *Nucleic Acids Research*, *46*(W1), W486–W494. <https://doi.org/10.1093/nar/gky310>
- Colman, S. L., Massa, G. A., Carboni, M. F., & Feingold, S. E. (2017). Cold sweetening diversity in Andean potato germplasm from Argentina. *Journal of the Science of Food and Agriculture*, *97*(14), 4744–4749. <https://doi.org/10.1002/jsfa.8343>
- Dale, M. F. B., & Bradshaw, J. E. (2003, July 1). Progress in improving processing attributes in potato. *Trends in Plant Science*. Elsevier Ltd. [https://doi.org/10.1016/S1360-1385\(03\)00130-4](https://doi.org/10.1016/S1360-1385(03)00130-4)
- Datir, S. S., Latimer, J. M., Thomson, S. J., Ridgway, H. J., Conner, A. J., & Jacobs, J. M. E. (2012a). Allele diversity for the apoplastic invertase inhibitor gene from potato. *Molecular Genetics and Genomics*. <https://doi.org/10.1007/s00438-012-0690-z>
- Datir, S. S., Latimer, J. M., Thomson, S. J., Ridgway, H. J., Conner, A. J., & Jacobs, J. M. E. (2012b). Allele diversity for the apoplastic invertase inhibitor gene from potato. *Molecular Genetics and Genomics*, *287*(6), 451–460. <https://doi.org/10.1007/s00438-012-0690-z>
- Datir, S. S., Mirikar, D., & RaviKumar, A. (2019). Sequence diversity and in silico structure prediction of the vacuolar invertase inhibitor gene from potato (*Solanum tuberosum* L.) cultivars differing in sugar content. *Food Chemistry*, *295*, 403–411. <https://doi.org/10.1016/j.foodchem.2019.05.096>
- Defernez, M., Gunning, Y. M., Parr, A. J., Shepherd, L. V. T., Davies, H. V., & Colquhoun, I. J. (2004). NMR and HPLC-UV profiling of potatoes with genetic modifications to metabolic pathways. *Journal of Agricultural and Food Chemistry*. <https://doi.org/10.1021/jf049522e>
- Dobson, G., Shepherd, T., Verrall, S. R., Conner, S., Mcnicol, J. W., Ramsay, G., ... Stewart, D. (2008). Phytochemical diversity in tubers of potato cultivars and landraces using a

- gc-ms metabolomics approach. *Journal of Agricultural and Food Chemistry*, 56(21), 10280–10291. <https://doi.org/10.1021/jf801370b>
- Draffehn, A. M., Meller, S., Li, L., & Gebhardt, C. (2010). Natural diversity of potato (*Solanum tuberosum*) invertases. *BMC Plant Biology*. <https://doi.org/10.1186/1471-2229-10-271>
- Feng, Q., Liu, Z., Zhong, S., Li, R., Xia, H., Jie, Z., ... Li, G. (2016). Integrated metabolomics and metagenomics analysis of plasma and urine identified microbial metabolites associated with coronary heart disease. *Scientific Reports*, 6, 22525. <https://doi.org/10.1038/srep22525>
- Galani Yamdeu, J. H., Gupta, P. H., Patel, N. J., Shah, A. K., & Talati, J. G. (2016). Effect of Storage Temperature on Carbohydrate Metabolism and Development of Cold-Induced Sweetening in Indian Potato (*Solanum Tuberosum* L.) Varieties. *Journal of Food Biochemistry*, 40(1), 71–83. <https://doi.org/10.1111/jfbc.12190>
- Gao, H., Zeng, Q., Ren, Z., Li, P., & Xu, X. (2018). Effect of exogenous γ -aminobutyric acid treatment on the enzymatic browning of fresh-cut potato during storage. *Journal of Food Science and Technology*. <https://doi.org/10.1007/s13197-018-3442-1>
- Greiner, S., Rausch, T., Sonnewald, U., & Herbers, K. (1999). Ectopic expression of a tobacco invertase inhibitor homolog prevents cold-induced sweetening of potato tubers. *Nature Biotechnology*. <https://doi.org/10.1038/10924>
- Hamernik, A. J., Hanneman, R. E., & Jansky, S. H. (2009). Introgression of wild species germplasm with extreme resistance to cold sweetening into the cultivated potato. *Crop Science*, 49(2), 529–542. <https://doi.org/10.2135/cropsci2008.04.0209>
- Hardigan, M. A., Laimbeer, F. P. E., Newton, L., Crisovan, E., Hamilton, J. P., Vaillancourt, B., ... Buell, C. R. (2017). Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.1714380114>
- Hasanuzzaman, M., Fujita, M., Oku, H., Islam, M. T., Ali, Q., Athar, H.-R., ... Hussain, S. M. (2019). Role of Amino Acids in Improving Abiotic Stress Tolerance to Plants. In *Plant Tolerance to Environmental Stress* (pp. 175–204). CRC Press. <https://doi.org/10.1201/9780203705315-12>
- Hatoum, D., Annaratone, C., Hertog, M. L. A. T. M., Geeraerd, A. H., & Nicolai, B. M. (2014). Targeted metabolomics study of “Braeburn” apples during long-term storage. *Postharvest Biology and Technology*, 96, 33–41.

<https://doi.org/10.1016/j.postharvbio.2014.05.004>

- Hayat, S., Hayat, Q., Alyemeni, M. N., Wani, A. S., Pichtel, J., & Ahmad, A. (2012, November). Role of proline under changing environments: A review. *Plant Signaling and Behavior*. Taylor & Francis. <https://doi.org/10.4161/psb.21949>
- Hou, J., Zhang, H., Liu, J., Reid, S., Liu, T., Xu, S., ... Xie, C. (2017). Amylases StAmy23, StBAM1 and StBAM9 regulate cold-induced sweetening of potato tubers in distinct ways. *Journal of Experimental Botany*, 68(9), 2317–2331. <https://doi.org/10.1093/jxb/erx076>
- Inostroza-Blancheteau, C., de Oliveira Silva, F. M., Durán, F., Solano, J., Obata, T., Machado, M., ... Nunes-Nesi, A. (2018). Metabolic diversity in tuber tissues of native Chiloé potatoes and commercial cultivars of *Solanum tuberosum* ssp. *tuberosum* L. *Metabolomics*, 14(10). <https://doi.org/10.1007/s11306-018-1428-7>
- Jansky, S. H., & Fajardo, D. A. (2014). Tuber starch amylose content is associated with cold-induced sweetening in potato. *Food Science and Nutrition*, 2(6), 628–633. <https://doi.org/10.1002/fsn3.137>
- Jin, Y.-I., Park, K.-H., Chang, D.-C., Cho, J.-H., Cho, K.-S., Im, J.-S., ... Chung, I.-M. (2016). Factors Influencing the Acrylamide Content of Fried Potato Products. *Korean Journal of Environmental Agriculture*. <https://doi.org/10.5338/kjea.2016.35.4.31>
- Kaur, R. (2017). Growth, Yield and Quality of Different Processing Cultivars of Potato (*Solanum tuberosum* L.). *International Journal of Pure & Applied Bioscience*, 5(6), 594–599. <https://doi.org/10.18782/2320-7051.2863>
- Kou, S., Chen, L., Tu, W., Scossa, F., Wang, Y., Liu, J., ... Xie, C. (2018). The arginine decarboxylase gene ADC1, associated to the putrescine pathway, plays an important role in potato cold-acclimated freezing tolerance as revealed by transcriptome and metabolome analyses. *Plant Journal*. <https://doi.org/10.1111/tpj.14126>
- Krasensky, J., & Jonak, C. (2012, February). Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks. *Journal of Experimental Botany*. Europe PMC Funders. <https://doi.org/10.1093/jxb/err460>
- Kumar, R., Bohra, A., Pandey, A. K., Pandey, M. K., & Kumar, A. (2017a, August 7). Metabolomics for plant improvement: Status and prospects. *Frontiers in Plant Science*. Frontiers Media S.A. <https://doi.org/10.3389/fpls.2017.01302>
- Kumar, R., Bohra, A., Pandey, A. K., Pandey, M. K., & Kumar, A. (2017b, August 7). Metabolomics for plant improvement: Status and prospects. *Frontiers in Plant Science*. Frontiers Media S.A. <https://doi.org/10.3389/fpls.2017.01302>

- Li, L., Paulo, M. J., Strahwald, J., Lübeck, J., Hofferbert, H. R., Tacke, E., ... Gebhardt, C. (2008). Natural DNA variation at candidate loci is associated with potato chip color, tuber starch content, yield and starch yield. *Theoretical and Applied Genetics*.
<https://doi.org/10.1007/s00122-008-0746-y>
- Liang, X., Zhang, L., Natarajan, S. K., & Becker, D. F. (2013, September 20). Proline mechanisms of stress survival. *Antioxidants and Redox Signaling*. Mary Ann Liebert, Inc. <https://doi.org/10.1089/ars.2012.5074>
- Lin, Q., Qian, J., Zhao, C., Wang, D., Liu, C., Wang, Z., ... Chen, K. (2016). Low temperature induced changes in citrate metabolism in Ponkan (*Citrus reticulata* Blanco cv. Ponkan) fruit during maturation. *PLoS ONE*, *11*(6), e0156703.
<https://doi.org/10.1371/journal.pone.0156703>
- Lin, Y., Liu, J., Liu, X., Ou, Y., Li, M., Zhang, H., ... Xie, C. (2013). Interaction proteins of invertase and invertase inhibitor in cold-stored potato tubers suggested a protein complex underlying post-translational regulation of invertase. *Plant Physiology and Biochemistry*, *73*, 237–244. <https://doi.org/10.1016/j.plaphy.2013.09.012>
- Liu, T., Fang, H., Liu, J., Reid, S., Hou, J., Zhou, T., ... Xie, C. (2017). Cytosolic glyceraldehyde-3-phosphate dehydrogenases play crucial roles in controlling cold-induced sweetening and apical dominance of potato (*Solanum tuberosum* L.) tubers. *Plant Cell and Environment*. <https://doi.org/10.1111/pce.13073>
- Malone, J. G., Mittova, V., Ratcliffe, R. G., & Kruger, N. J. (2006a). The response of carbohydrate metabolism in potato tubers to low temperature. *Plant and Cell Physiology*. <https://doi.org/10.1093/pcp/pcj101>
- Malone, J. G., Mittova, V., Ratcliffe, R. G., & Kruger, N. J. (2006b). The response of carbohydrate metabolism in potato tubers to low temperature. *Plant and Cell Physiology*, *47*(9), 1309–1322. <https://doi.org/10.1093/pcp/pcj101>
- Mankotia, S., & Sharma, S. (2020). Potato (*Solanum tuberosum* L.) Yield, NPK Concentration and Uptake as Influenced by Variable Levels of Drip Irrigation and Fertigation. *Int.J.Curr.Microbiol.App.Sci*, *9*(5), 1277–1289.
<https://doi.org/10.20546/ijcmas.2020.905.142>
- Marwaha, R. S., Pandey, S. K., Kumar, D., Singh, S. V., & Kumar, P. (2010). Potato processing scenario in India: Industrial constraints, future projections, challenges ahead and remedies - A review. *Journal of Food Science and Technology*.
<https://doi.org/10.1007/s13197-010-0026-0>
- Mazzucotelli, E., Tartari, A., Cattivelli, L., & Forlani, G. (2006). Metabolism of γ -

- aminobutyric acid during cold acclimation and freezing and its relationship to frost tolerance in barley and wheat. *Journal of Experimental Botany*.
<https://doi.org/10.1093/jxb/erl141>
- Menéndez, C. (2002). Cold Sweetening in Diploid Potato: Mapping Quantitative Trait Loci and Candidate Genes. *Genetics*, *162*, 1423–1434.
- Menéndez, C. M., Ritter, E., Schäfer-Pregl, R., Walkemeier, B., Kalde, A., Salamini, F., & Gebhardt, C. (2002). Cold sweetening in diploid potato: Mapping quantitative trait loci and candidate genes. *Genetics*.
- Mottram, D. S., Wedzicha, B. L., & Dodson, A. T. (2002). Food chemistry: Acrylamide is formed in the Maillard reaction. *Nature*, *419*(6906), 448–449.
<https://doi.org/10.1038/419448a>
- Muthoni, J., Kabira, J., Shimelis, H., & Melis, R. (2015). Tetrasomic inheritance in cultivated potato and implications in conventional breeding. *Australian Journal of Crop Science*.
- Muttucumar, N., Powers, S. J., Elmore, J. S., Dodson, A., Briddon, A., Mottram, D. S., & Halford, N. G. (2017). Acrylamide-forming potential of potatoes grown at different locations, and the ratio of free asparagine to reducing sugars at which free asparagine becomes a limiting factor for acrylamide formation. *Food Chemistry*, *220*, 76–86.
<https://doi.org/10.1016/j.foodchem.2016.09.199>
- Obata, T., & Fernie, A. R. (2012). The use of metabolomics to dissect plant responses to abiotic stresses. *Cellular and Molecular Life Sciences*. <https://doi.org/10.1007/s00018-012-1091-5>
- Oertel, A., Matros, A., Hartmann, A., Arapitsas, P., Dehmer, K. J., Martens, S., & Mock, H. P. (2017). Metabolite profiling of red and blue potatoes revealed cultivar and tissue specific patterns for anthocyanins and other polyphenols. *Planta*.
<https://doi.org/10.1007/s00425-017-2718-4>
- P Aggarwal, S. K. V. V. (2017). Processing quality traits of different potato (*Solanum tuberosum* L.) genotypes in India. *Pharma Innov.*, *6*, 27–30.
- Pal, S., Bhattacharya, A., Konar, A., Mazumdar, D., & Das, A. K. (2008). Chemical composition of potato at harvest and after cold storage. *International Journal of Vegetable Science*, *14*(2), 162–176. <https://doi.org/10.1080/19315260801934779>
- R Rana, S. P. (2007). Processing Quality Potatoes in India: An Analysis of Industry's Demand. *Processed Food Ind.*, *10*, 26–35.
- Raffan, S., & Halford, N. G. (2019). Acrylamide in food: Progress in and prospects for genetic and agronomic solutions. *Annals of Applied Biology*.

<https://doi.org/10.1111/aab.12536>

Raigond, P., Mehta, A., & Singh, B. (2018). Sweetening During Low-Temperature and Long-Term Storage of Indian Potatoes. *Potato Research*, *61*(3), 207–217.

<https://doi.org/10.1007/s11540-018-9369-0>

Rajiv, & Kwar, P. G. (2016). Enriched potato for mitigating hidden hunger. In

Biofortification of Food Crops. https://doi.org/10.1007/978-81-322-2716-8_32

Roessner-Tunali, U., Urbanczyk-Wochniak, E., Czechowski, T., Kolbe, A., Willmitzer, L., & Fernie, A. R. (2003). De Novo Amino Acid Biosynthesis in Potato Tubers Is Regulated by Sucrose Levels. *Plant Physiology*, *133*(2), 683–692.

<https://doi.org/10.1104/pp.103.024802>

Ross, H. A., Wright, K. M., McDougall, G. J., Roberts, A. G., Chapman, S. N., Morris, W.

L., ... Taylor, M. A. (2011). Potato tuber pectin structure is influenced by pectin methyl esterase activity and impacts on cooked potato texture. *Journal of Experimental Botany*, *62*(1), 371–381. <https://doi.org/10.1093/jxb/erq280>

RS Marwaha, S. P. S. S. S. K. (2005). Processing and nutritional qualities of Indian and exotic potato cultivars as influenced by harvest date, tuber curing, pre-storage holding period, storage and reconditioning under short days. *Adv. Horticultural Sci.*, *19*, 130–140.

S Kaur, P. A. (2014). Studies on Indian Potato Genotypes for their Processing and Nutritional Quality Attributes. *Int. J. Curr. Microbiology Appl. Sci.*, *3*, 172–177.

Severini, C., Baiano, A., De Pilli, T., Romaniello, R., & Derossi, A. (2003). Prevention of enzymatic browning in sliced potatoes by blanching in boiling saline solutions. *LWT - Food Science and Technology*. [https://doi.org/10.1016/S0023-6438\(03\)00085-9](https://doi.org/10.1016/S0023-6438(03)00085-9)

Shang, H., Cao, S., Yang, Z., Cai, Y., & Zheng, Y. (2011). Effect of exogenous γ -aminobutyric acid treatment on proline accumulation and chilling injury in peach fruit after long-term cold storage. *Journal of Agricultural and Food Chemistry*.

<https://doi.org/10.1021/jf104424z>

Sharma, A. (2012). STORABILITY AND SPROUTING BEHAVIOUR OF MICRO-TUBERS OF SOME INDIAN POTATO CULTIVARS. *Potato J.*, *39*, 31–38.

Shepherd, L. V. T., Alexander, C. A., Sungurtas, J. A., McNicol, J. W., Stewart, D., & Davies, H. V. (2010a). Metabolomic analysis of the potato tuber life cycle.

Metabolomics. <https://doi.org/10.1007/s11306-009-0183-1>

Shepherd, L. V. T., Alexander, C. A., Sungurtas, J. A., McNicol, J. W., Stewart, D., & Davies, H. V. (2010b). Metabolomic analysis of the potato tuber life cycle.

- Metabolomics*, 6(2), 274–291. <https://doi.org/10.1007/s11306-009-0183-1>
- Singh, P. P., & Saldaña, M. D. A. (2011). Subcritical water extraction of phenolic compounds from potato peel. *Food Research International*, 44(8), 2452–2458.
<https://doi.org/10.1016/j.foodres.2011.02.006>
- Slater, A. T., Cogan, N. O. I., Hayes, B. J., Schultz, L., Dale, M. F. B., Bryan, G. J., & Forster, J. W. (2014, October 25). Improving breeding efficiency in potato using molecular and quantitative genetics. *Theoretical and Applied Genetics*. Springer Verlag.
<https://doi.org/10.1007/s00122-014-2386-8>
- Smith, A. M., Zeeman, S. C., & Smith, S. M. (2005). STARCH DEGRADATION. *Annual Review of Plant Biology*, 56(1), 73–98.
<https://doi.org/10.1146/annurev.arplant.56.032604.144257>
- Sowokinos, J. R. (2001a). Biochemical and molecular control of cold-induced sweetening in potatoes. *American Journal of Potato Research*. <https://doi.org/10.1007/BF02883548>
- Sowokinos, J. R. (2001b). Biochemical and molecular control of cold-induced sweetening in potatoes. *American Journal of Potato Research*. Potato Association of America.
<https://doi.org/10.1007/BF02883548>
- Sowokinos, Joseph R., Hayes, R. J., & Thill, C. A. (2018). Coordinated Regulation of Cold Induced Sweetening in Tetraploid Potato Families by Isozymes of UDP-Glucose Pyrophosphorylase and Vacuolar Acid Invertase. *American Journal of Potato Research*.
<https://doi.org/10.1007/s12230-018-9653-1>
- Steinfath, M., Strehmel, N., Peters, R., Schauer, N., Groth, D., Hummel, J., ... van Dongen, J. T. (2010). Discovering plant metabolic biomarkers for phenotype prediction using an untargeted approach. *Plant Biotechnology Journal*, 8(8), 900–911.
<https://doi.org/10.1111/j.1467-7652.2010.00516.x>
- Suttirak, W., & Manurakchinakorn, S. (2010). Potential Application of Ascorbic Acid, Citric Acid and Oxalic Acid for Browning Inhibition in Fresh-Cut Fruits and Vegetables. *Walailak Journal of Science and Technology*.
- Tamanna, N., & Mahmood, N. (2015). Food processing and maillard reaction products: Effect on human health and nutrition. *International Journal of Food Science*.
<https://doi.org/10.1155/2015/526762>
- Ulrich, E. L., Akutsu, H., Doreleijers, J. F., Harano, Y., Ioannidis, Y. E., Lin, J., ... Markley, J. L. (2008). BioMagResBank. *Nucleic Acids Research*, 36(SUPPL. 1), D402-408.
<https://doi.org/10.1093/nar/gkm957>
- Uri, C., Juhász, Z., Polgár, Z., & Bánfalvi, Z. (2014). A GC-MS-based metabolomics study

- on the tubers of commercial potato cultivars upon storage. *Food Chemistry*, *159*, 287–292. <https://doi.org/10.1016/j.foodchem.2014.03.010>
- Valluru, R., & Van den Ende, W. (2011, October). Myo-inositol and beyond - Emerging networks under stress. *Plant Science*. <https://doi.org/10.1016/j.plantsci.2011.07.009>
- Weber, A. P. M. (2004). Solute transporters as connecting elements between cytosol and plastid stroma. *Current Opinion in Plant Biology*. <https://doi.org/10.1016/j.pbi.2004.03.008>
- Wishart, D. S., Tzur, D., Knox, C., Eisner, R., Guo, A. C., Young, N., ... Querengesser, L. (2007). HMDB: The human metabolome database. *Nucleic Acids Research*, *35*(SUPPL. 1), D521-526. <https://doi.org/10.1093/nar/gkl923>
- Xia, J., Bjorndahl, T. C., Tang, P., & Wishart, D. S. (2008). MetaboMiner - Semi-automated identification of metabolites from 2D NMR spectra of complex biofluids. *BMC Bioinformatics*, *9*, 507–507. <https://doi.org/10.1186/1471-2105-9-507>
- Xia, J., Wishart, D. S., & Valencia, A. (2011). MetPA: A web-based metabolomics tool for pathway analysis and visualization. In *Bioinformatics* (Vol. 27, pp. 2342–2344). Oxford University Press. <https://doi.org/10.1093/bioinformatics/btq418>
- Xiong, X., Tai, G. C. C., & Seabrook, J. E. A. (2002). Effectiveness of selection for quality traits during the early stage in the potato breeding population. *Plant Breeding*, *121*(5), 441–444. <https://doi.org/10.1046/j.1439-0523.2002.00699.x>
- Yun, Z., Qu, H., Wang, H., Zhu, F., Zhang, Z., Duan, X., ... Jiang, Y. (2016). Comparative transcriptome and metabolome provides new insights into the regulatory mechanisms of accelerated senescence in litchi fruit after cold storage. *Scientific Reports*, *6*. <https://doi.org/10.1038/srep19356>
- Zhang, D., Mu, T., Sun, H., Chen, J., & Zhang, M. (2017). Comparative study of potato protein concentrates extracted using ammonium sulfate and isoelectric precipitation. *International Journal of Food Properties*, *20*(9), 2113–2127. <https://doi.org/10.1080/10942912.2016.1230873>
- Zhang, H., Hou, J., Liu, J., Zhang, J., Song, B., & Xie, C. (2017, January 1). The roles of starch metabolic pathways in the cold-induced sweetening process in potatoes. *Starch/Staerke*. Wiley-VCH Verlag. <https://doi.org/10.1002/star.201600194>
- Zhang, W. F., Gong, Z. H., Wu, M. B., Chan, H., Yuan, Y. J., Tang, N., ... Deng, W. (2019). Integrative comparative analyses of metabolite and transcript profiles uncovers complex regulatory network in tomato (*Solanum lycopersicum* L.) fruit undergoing chilling injury. *Scientific Reports*, *9*(1). <https://doi.org/10.1038/s41598-019-41065-9>

- Zhang, Y., & Fernie, A. R. (2018). On the role of the tricarboxylic acid cycle in plant productivity. *Journal of Integrative Plant Biology*. <https://doi.org/10.1111/jipb.12690>
- Zhang, Z., Zhu, Q., Hu, M., Gao, Z., An, F., Li, M., & Jiang, Y. (2017). Low-temperature conditioning induces chilling tolerance in stored mango fruit. *Food Chemistry*, 219, 76–84. <https://doi.org/10.1016/j.foodchem.2016.09.123>
- Zhao, D. Y., Shen, L., Fan, B., Liu, K. L., Yu, M. M., Zheng, Y., ... Sheng, J. P. (2009). Physiological and genetic properties of tomato fruits from 2 cultivars differing in chilling tolerance at cold storage. *Journal of Food Science*, 74(5). <https://doi.org/10.1111/j.1750-3841.2009.01156.x>

Appendix 1

Supporting Tables

Table 3.3: List of chemical shift assignment of abundant water-soluble metabolites as extracted from *INS-1E* cells

S. No.	Metabolite (PubChem CID Code)	Chemical shift (multiplicity)	Assignment
1	Acetate (175)	1.90 (s)	CH ₃
2	Alanine (5950)	3.805 (q)	α -CH
		1.46 (d)	β -CH ₃
3	ADP (6022)	8.534 (s)	C ₇ H
		8.261 (s)	C ₁₂ H
		6.122 (d)	C ₂ H
		4.52 (m)	C ₃ H
		4.35 (m)	C ₅ H
		4.00 (m)	C ₁₇ H
4	AMP (6083)	8.596 (s)	C ₇ H
		8.25 (s)	C ₁₂ H
		6.12 (d)	C ₂ H
		4.49 (dd)	C ₃ H
		4.355 (dd)	C ₅ H
		4.02 (dd)	C ₁₇ H
5	Aspartate (5960)	2.66 (dd)	β -CH ₂
		2.79 (dd)	β -CH ₂
		3.916 (dd)	α -CH
6	Asparagine (6267)	2.84 (m)	β -CH ₂
		2.94 (m)	β -CH ₂
		3.99 (dd)	α -CH
7	ATP (5957)	8.52 (s)	C ₇ H
		8.25 (s)	C ₁₂ H
		6.128 (d)	C ₂ H
		4.56 (t)	C ₃ H
		4.41 (m)	C ₄ H
		4.23 (m)	C ₁₇ H
8	Choline (305)	3.2 (s)	N(CH ₃) ₃
		3.52 (m)	NCH ₂
		4.06 (m)	OCH ₂
9	Citrate (311)	2.67 (d)	CH ₂

		2.76 (d)	CH ₂
10	Citrulline (9750)	1.556 (m)	γ-CH ₂
		1.86 (m)	β-CH ₂
		3.126 (m)	α-CH ₂
		3.74 (m)	α-CH
11	Creatine (586)	3.02 (s)	CH ₃
		3.90 (s)	CH ₂
12	DSS (74873)	0	Si(CH ₃) ₃
		0.62 (t)	γ-CH ₂
		1.75 (m)	β-CH ₂
		2.91 (t)	α-CH ₂
13	Ethanol (702)	1.18 (t)	CH ₃
		3.64 (q)	CH ₂
14	Formate (283)	8.44 (s)	CH
15	Fumarate (5460307)	6.50 (s)	CH
16	Glucose (5793)	3.23 (dd)	C ₃ H
		3.398 (m)	C ₅ H
		3.458 (m)	C ₆ H
		3.524 (m)	C ₃ H
		3.728 (m)	C ₄ H/C ₁₁ H
		3.824 (m)	C ₁₁ H/C ₆ H
		3.889 (dd)	C ₁₁ H
		4.634 (d)	C ₂ H
		5.226 (d)	C ₂ H
17	Glutamate (33032)	2.03 (m)	β-CH ₂
		2.10 (m)	β-CH ₂
		2.34 (m)	γ-CH ₂
		3.75 (dd)	α-CH
18	Glutamine (5961)	2.14 (m)	β-CH ₂
		2.459 (m)	γ-CH ₂
		3.76 (t)	α-CH
19	Glutathione (124886)	4.56 (q)	C ₃ H
		3.78 (m)	C ₁₄ H/C ₁₁ H
		2.97 (dd)	C ₂ H
		2.54 (m)	C ₉ H

		2.15 (m)	C ₁₀ H
20	Glycerol (753)	3.55 (dd)	CH ₂
		3.64 (dd)	CH ₂
		3.77 (dd)	CH ₂
21	Glycine (750)	3.56 (s)	CH ₂
22	GTP (6830)	4.18 (m)	C ₁₇ H
		4.24 (m)	C ₁₇ H
		4.36 (m)	C ₅ H
		4.58 (dd)	C ₄ H
		5.92 (d)	C ₂ H
		8.13 (s)	C ₇ H
23	Histidine (6274)	7.874 (s)	C ₂ H
		7.10 (s)	C ₅ H
		3.98 (dd)	α -CH
		3.21 (dd)	β -CH ₂
		3.11 (dd)	β -CH ₂
24	Hypoxanthine (790)	8.20 (s)	C ₇ H
		8.195 (s)	C ₂ H
25	IMP (8582)	8.55 (s)	C ₇ H
		8.21 (s)	C ₁₂ H
		6.13 (d)	C ₂ H
		4.5 (m)	C ₄ H
		4.35 (m)	C ₅ H
		4.02 (m)	C ₁₇ H
26	Inosine (6021)	8.34 (s)	C ₁₂ H
		8.23 (s)	C ₇ H
		6.09 (d)	C ₂ H
		4.77 (dd)	C ₃ H
		4.427 (dd)	C ₄ H
		4.257 (dd)	C ₅ H
		3.90 (dd)	C ₁₇ H
		3.84 (dd)	C ₁₇ H
27	Isocitrate (1198)	2.5 (dq)	γ -CH ₂
		2.99 (m)	β -CH
		4.05 (d)	α -CH

28	Isoleucine (6306)	0.926 (t)	δ -CH ₃
		0.992 (d)	β -CH ₃
		1.248 (m)	γ -CH ₂
		1.457 (m)	γ -CH ₂
		1.968 (m)	β -CH
		3.661 (d)	α -CH
29	Isopropanol (3776)	1.16 (m)	CH ₃
		4.01 (m)	CH
30	Lactate (612)	4.096 (q)	CH
		1.313 (d)	γ -CH ₃
31	LDL	0.84 (t)	CH ₃ (CH ₂) _n
		1.25 (m)	(CH ₂) _n
32	Leucine (6106)	0.942 (d)	δ -CH ₃
		0.954 (d)	δ -CH ₃
		1.68 (m)	β -CH ₂
		1.70 (m)	γ -CH
		3.71 (t)	α -CH
33	Malate (525)	2.37 (dd)	β -CH ₂
		2.66 (dd)	β -CH ₂
		4.29 (dd)	α -CH
34	Myo-inositol (892)	3.27 (t)	C ₅ H
		3.54 (dd)	C ₁ H/C ₃ H
		3.62 (dd)	C ₄ H/C ₆ H
		4.05 (t)	C ₂ H
35	NAD (5892)	4.22 (m)	C ₂₅ H
		4.35 (m)	C ₁₇ H
		4.37 (m)	C ₃ H
		4.43 (m)	C ₃₀ H
		4.49 (m)	C ₂₉ H
		4.53 (m)	C ₂₆ H
		6.02 (d)	C ₂ H
		6.08 (d)	C ₂₈ H
		6.12 (d)	C ₂ H
		8.14 (s)	C ₁₂ H
		8.20 (m)	C ₃₈ H

		8.40 (s)	C ₇ H
		8.83 (d)	C ₃₉ H
		9.15 (d)	C ₃₇ H
		9.33 (s)	C ₃₅ H
36	Niacinamide (936)	7.58 (dd)	C ₅ H
		8.24 (dd)	C ₄ H
		8.7 (dd)	C ₆ H
		8.92 (s)	C ₂ H
37	o-Phosphocholine (1014)	3.685 (s)	N(CH ₃) ₃
		3.58 (m)	NCH ₂
		4.15 (m)	OCH ₂
38	Proline (145742)	4.116 (t)	α-CH
		3.405 (m)	δ-CH ₂
		3.33 (m)	δ-CH ₂
		2.339 (m)	β-CH ₂
		2.07 (m)	β-CH ₂
		2.01 (m)	γ-CH ₂
39	Phenylalanine (6140)	7.42 (m)	C ₃ H/C ₅ H
		7.36 (m)	C ₂ H/C ₆ H
		7.32 (m)	C ₄ H
		3.98 (dd)	α-CH
		3.27 (m)	β-CH ₂
		3.13 (m)	β-CH ₂
40	Pyruvate (107735)	2.36 (s)	CH ₂
41	Succinate (160419)	2.385 (s)	CH
42	Taurine (1123)	3.25 (t)	S-CH ₂
		3.42 (t)	N-CH ₂
43	Tyrosine (6057)	7.17 (d)	C ₂ H/C ₆ H
		6.87 (d)	C ₃ H/C ₅ H
		3.93 (dd)	α-CH
		3.18 (dd)	β-CH ₂
		3.04 (dd)	β-CH ₂
44	UDP-Glucose (8629)	7.931 (d)	C ₃₂ H
		5.988 (d)	C ₃₁ H

		5.988 (d)	C ₂₁ H
		5.595 (dd)	C ₂ H
		4.36 (m)	C ₂₂ H
		4.23 (m)	C ₁₈ H
		3.91 (m)	C ₄ H
		3.54 (dd)	C ₃ H
45	UDP-N-Acetylglucosamine (445675)	7.931 (d)	C ₃₅ H
		5.988 (d)	C ₂₆ H
		5.988 (d)	C ₃₄ H
		5.505 (dd)	C ₂ H
		4.35 (m)	C ₂₇ H
		4.23 (m)	C ₂₃ H
		3.91 (m)	C ₆ H
		3.54 (dd)	C ₅ H
		2.06 (s)	C ₂₁ H
46	UMP (6030)	8.07 (d)	C ₁₀ H
		5.98 (d)	C ₁₁ H
		5.98 (d)	C ₂ H
		4.41 (t)	C ₃ H
		4.34 (t)	C ₄ H
		4.26 (m)	C ₅ H
		3.97 (m)	C ₁₄ H
47	Uracil (1174)	5.79 (d)	C ₆ H
		7.54 (d)	C ₅ H
48	Uridine (6029)	7.85 (d)	C ₁₁ H
		5.9 (d)	C ₂ H
		5.89 (d)	C ₁₀ H
		4.34 (dd)	C ₃ H
		4.22 (dd)	C ₄ H
		4.12 (m)	C ₅ H
		3.9 (dd)	C ₁₄ H
		3.8 (dd)	C ₁₄ H
49	Valine (6287)	0.996 (d)	CH ₃
		1.047 (d)	CH ₃

		2.281 (m)	β -CH
		3.617 (d)	α -CH

Table 3.4: Summary of metabolic studies performed in pancreatic β -cells following high-glucose and FFA exposure

Cellular system	Stimulation conditions using glucose	Extraction protocol	Platform used	Major metabolites identified
INS 832/13 (Fernandez et al. 2008)	2.8 mM or 16.7 mM glucose for 48 h	Methanol-chloroform extraction	GC/MS (2008)	Alanine; TCA cycle intermediates: Citrate, malate, fumarate increase; Pentose phosphate pathway: Ribose-5-phosphate; Amino acids: Alanine, hydroxyproline increase; Glutamine, serine and ornithine decrease
<i>INS-1E</i> (Nyblom et al. 2008)	5.5, 11, 20 and 27 mM glucose for 5 days	Lipid extraction using 0.88% KCl and Methanol-chloroform (2:1) extraction	HR-MAS NMR and GC/MS (2008)	<i>De novo</i> synthesis of fatty acid accumulation (oleate, palmitate, stearate, octadecenoate, and palmioleate); Decrease in o-phosphocholine and glycerophosphocholine
INS-1 832/13 (Spégel et al. 2011)	2.8 or 16.7 mM glucose for 1 h	Methanol-chloroform extraction	GC/MS (2011)	Increase in late glycolytic and TCA cycle intermediates like fumarate, malate; lactate
<i>INS-1E</i> and Human Islets (Goehring et al. 2011)	3 or 16 mM glucose for 5 m, 45 m, 2 h, 4 h, 8 h, 24 h and 48 h	Methanol: Chloroform:Water (2.5:1:1)	GC (TOF) - MS (2011)	Increase in glycolysis, TCA cycle, Polyol synthesis, metabolites from pentose phosphate pathway, DNA/RNA synthesis pathway and glycerol synthesis pathway meabolites.
INS-1 832/13 (Huang and Joseph 2012)	2 h in KRB supplemented with 2 mM glucose for 2 h followed by 2 mM or	Methanol extraction	GC/MS (2012)	Glycolytic, TCA pathway, pentose phosphate, sorbitol-aldose reductase pathways intermediates increase; NADPH/NADP ⁺ ratio increase; leucine, lysine, tyrosine, alanine, and

	16.7 mM glucose for 2 h			serine were increased; aspartate decreased; palmitic acid, oleic acid increase
INS-1 832/13 and rat islets (Spégel et al. 2013)	2.8 mM glucose for 2 h followed by stimulation for 0, 3, 6, 10 and 15 min in 16.7 mM glucose	Methanol: water extraction	GC/MS (2013)	Increase in glycolytic and TCA cycle intermediates like lactate, fumarate, malate, α -KG, succinate; aspartate, serine decrease; glutamate alanine increase; ribose-5P increase; Increase NADPH/NADP ⁺ ratio
INS-1 832/13 (Lorenz et al. 2013)	3 mM glucose for 20 h followed by 0, 2, 5, 10 and 20 mM glucose for 30 min; U-13C Glucose was used to probe flux in specific pathways.	Methanol-chloroform-water based extraction	LC-TOF-MS (2013)	16:0 CoA, Acetyl CoA, Malonyl CoA increase; Increase in glycolytic and TCA intermediates like citrate, isocitrate, succinate, α -KG, malate; decrease in glutamate, aspartate
BRIN-BD11 and <i>INS-1E</i> (Wallace et al. 2013)	11.1 for 20 min followed by 25 mM glucose for 20 h	Methanol-chloroform extraction for GC/MS and perchloric acid extraction for NMR	GC/MS, ¹ H NMR and ¹³ C NMR (2013)	Reduction in pyruvate oxidation; reduced arachidonic acid, EPA, cis-8,11,14-eicosatrienoic and polyunsaturated fatty acids (PUFA); decrease in phospho-choline, creatine, aspartate, and alanine and an increase in glycine and glucose; decrease in TCA cycle metabolism; glutathione
INS-1 832/13 (Göhring et al. 2014)	2.8 mM and 16 mM glucose for 48 h	Methanol extraction	GC/MS (2014)	Increase in Glycolytic and TCA cycle intermediates like lactate, Alanine, Malate, citrate, isocitrate;
INS-1 832/13 (Spégel et al. 2015)	2.8 mM glucose for 2 h followed by 0 and 60 min after stimulation with 16.7 mM glucose	Methanol-chloroform extraction	GC/MS (2015)	Increase in glycolysis and TCA cycle intermediates (citrate, isocitrate, α -KG, fumarate, and malate); Glutamate Increased; aspartate decrease; branched chain amino acids
INS-1 832/13 (Mugabo et al. 2017)	2 mM glucose and 2 mM glutamine for 2 h followed by different concentrations of	80% methanol, 13.7 mM ammonium	LC-ESI-MS/MS	Increase in glycolysis-related and Krebs cycle intermediates (citrate/isocitrate, fumarate, malate); Increase in NADH, ATP; Decrease

	glucose (5-20 mM) for 1 h	acetate pH 9.0; Polar metabolites - chloroform:hexane (3:1, v/v)		in aspartate; Increase production and release of FFA; Increase glycerol, Triglycerides, glycogen
<i>INS-1E</i> (Yousf et al. 2019)	5 mM glucose for 24 h followed by stimulation for 24 h in 16 mM glucose and 0.5 mM palmitate	Methanol-extraction	NMR (2019)	Increase in glycolysis-related and Krebs cycle intermediates (lactate, citrate/isocitrate, fumarate); Increased flux of Hexosamine pathway; Decrease in aspartate, glutathione, o-phosphocholine, taurine.
Cellular system	Stimulation conditions using lipids	Extraction protocol	Platform used	Major metabolites identified
MIN6 (Boslem et al. 2011)	48 h with 0.4 mM palmitate	Lipid extraction using chloroform/methanol	MS (2011)	Metabolites of sphingolipids like Phosphatidylethanolamine, glucosylceramine; ceramide increase
INS-1 (El-Azzouny et al. 2014)	Glucose-free media with BSA or palmitate for 30 min followed by stimulation with 16.6 mM glucose for 5-60 min.	Methanol:chloroform:water	LC/MS (2014)	A decrease in glycolytic and TCA cycle intermediates (Malonyl CoA); reduction in NADH/NAD ⁺ ; Increase in <i>de novo</i> synthesized glycerolipids (Diacylglycerol); increase in sphingosine-1-phosphate and N-acyl amide levels

Table 4.2: Serum metabolites (as identified from Figure 4.1) were assigned to their respective proton chemical shifts (ppm)

Metabolite (Pubchem CID)	¹ H Chemical shift (ppm) and Multiplicity
3-hydroxyisobutyrate (87)	1.07(d), 2.47(m)
Acetate (175)	1.90(s)
Acetone (180)	2.221(s)
Alanine (5950)	1.46(d), 3.805(q)
Arginine (6322)	1.64(m), 1.72(m), 1.89(m), 3.23(t)
Aspartate (5960)	2.66(dd), 2.79(dd), 3.916(dd)
Betaine (247)	3.263(s), 3.885(s)
Choline (305)	3.2(s), 3.52(m), 4.06(m)
Citrate (311)	2.67(d), 2.76(d)
Creatinine (588)	3.03 (s), 4.05 (s)
DSS (74873)	0(s), 0.62(t), 1.75(m), 2.91(t)
Formate (283)	8.44(s)
Fructose (2723872)	4.11(d), 4.01(m), 3.886(dd), 3.78(m), 3.70(m), 3.54(m)
Glucose (5793)	5.22(d), 4.634(d), 3.398(m), 3.889(dd), 3.824(m), 3.728(m), 3.524(m), 3.458 (m), 3.23(dd)
Glutamate (33032)	2.03(m), 2.10(m), 2.34(m), 3.75(dd)
Glutamine (5961)	2.14(m), 2.459(m), 3.76(t)
Glycerol	3.768 (tt), 3.64 (dd), 3.546 (dd)
Glycine (750)	3.56(s)
Isoleucine (6306)	0.926(t), 0.992(d), 1.248(m), 1.457(m), 1.968(m), 3.661 (d)
Isopropanol (3776)	1.16 (d), 4.01 (m)
Lactate (612)	4.096(q), 1.313(d)
LDL	0.87(m), 1.25 (m)
Leucine (6106)	0.942(d), 0.954(d), 1.68(m), 1.70(m), 3.71(t)
Lysine (5962)	3.75(t), 3.01(t) 1.91(m), 1.72(m), 1.49(m), 1.42(m)
Methanol (887)	3.34(s)
Methionine (6137)	3.85(t), 2.17(m), 2.63(t), 2.12(s)
o-Phosphocholine	3.21 (s), 3.58 (m), 4.15 (m)
Phenylalanine (6140)	7.32(m), 7.42(m), 7.36(m), 3.98(dd), 3.27(m), 3.13(m)
Proline (145742)	4.116(t), 3.405(m), 3.33(m), 2.339(m), 2.07(m), 2.01(m)
Pyroglutamate (7405)	2.02(m), 2.40(m), 2.49(m), 4.16(m)
Serine (5951)	3.98(m), 3.94(m), 3.83(m)
SN-glycero-3-phosphocholine (657272)	3.21(s), 4.31(m)
Succinate (160419)	2.385(s)
Tryptophan (6305)	7.72(d), 7.53(d), 7.31(s), 7.27(t), 7.19(t), 4.05(m), 3.29(m)
Tyrosine (6057)	7.18(d), 6.89(d), 3.18(dd), 3.04(dd)
Valine (6287)	3.60(d), 2.26(m), 1.03(d), 0.97(d)

Table 4.3: p-values and FDR-values of Control vs. T2DM group obtained from ANOVA.

Metabolite	p-value	-log₁₀(p)	FDR
Glucose	2.16E-33	32.666	6.26E-32
Serine	1.99E-16	15.701	2.89E-15
Pyroglutamate	4.83E-16	15.316	4.67E-15
Proline	1.21E-12	11.918	8.77E-12
Glutamate	3.80E-12	11.421	2.20E-11
Isoleucine	6.35E-12	11.197	3.07E-11
Methionine	1.25E-10	9.903	5.19E-10
o-Phosphocholine	5.74E-10	9.241	2.08E-09
Citrate	8.27E-08	7.083	2.66E-07
Choline	2.06E-07	6.686	5.98E-07
Betaine	2.48E-07	6.605	6.55E-07
Alanine	3.04E-07	6.517	7.35E-07
Leucine	1.48E-06	5.831	3.29E-06
Glycerol	8.05E-06	5.094	1.67E-05
3-hydroxyisobutyrate	8.96E-06	5.048	1.73E-05
Valine	5.21E-05	4.283	9.44E-05
SN-glycero-3-phosphocholine	0.00028953	3.538	0.0004939
Glycine	0.00062976	3.201	0.0010146
Formate	0.0015834	2.800	0.0024168
Lysine	0.001731	2.762	0.0025099
Aspartate	0.0042584	2.371	0.0058807
Arginine	0.018809	1.726	0.024793
Lactate	0.039626	1.402	0.04831
Tryptophan	0.039981	1.398	0.04831

Table 4.4a: d(i) values, p-values and FDR-values of Control vs. T2DM group obtained from SAM analysis

Metabolite	d.value	p-value	q.value
Glucose	12.4	0	0
Serine	-8.8	0	0
Pyroglutamate	-8.1	0	0
Glutamate	7.4	0	0
Proline	7.2	0	0
Methionine	6.9	0	0
Isoleucine	6.9	0	0
Citrate	5.7	0	0
Betaine	-5.6	0	0
o-Phosphocholine	-4.9	0	0
Alanine	4.9	0	0
3-hydroxyisobutyrate	4.6	0	0
Formate	-4	0	0
Choline	-4	0	0
Leucine	3.9	0	0
Glycine	-3.7	0	0
Valine	3.7	0	0
Lysine	3.5	0	0
SN-glycero-3-phosphocholine	-3.5	0.00034483	1.65E-05
Glycerol	3.3	0.0010345	4.70E-05
Aspartate	-3	0.0037931	0.0001643
Arginine	-2.4	0.016897	0.0006986
Tryptophan	2.3	0.024483	0.00096825

Table 4.4b: d(i) values, p-values, and FDR-values of Control vs. Prediabetic group obtained from SAM analysis.

Metabolite	d.value	p-value	q.value
Glucose	6.9	0	0
Pyroglutamate	-4.7	0	0
o-Phosphocholine	-4	0	0
Serine	-3.9	0	0
Choline	-3.6	0	0
SN-glycero-3-phosphocholine	-2.4	0.0034483	0.00037104
Lactate	-2.1	0.0086207	0.0006184
Acetate	-1.8	0.02069	0.0013357
Betaine	-1.8	0.022759	0.0013357

Table 5.2: *Mycobacterium smegmatis* metabolites (as identified from Figure 5.1) were assigned to their respective proton chemical shifts (ppm)

S.No	Metabolites	Chemical shift (ppm)
1	1-Methylnicotinamide	9.27(s), 8.96(d), 8.89(d), 8.17(t)
2	2-Aminobutyrate	3.706(t), 1.887(m), 0.968(t)
3	2-Hydroxy-3-methylvalerate (HMVA)	3.88(d), 1.705(m), 1.351(m), 1.162(m), 0.932(d), 0.872(t)
4	3-Hydroxyisovalerate	1.233(s), 2.35(s)
5	3-Methyl-2-Oxovalerate	0.88(t), 1.086(d), 1.444(m), 1.687(m), 2.922(m)
6	Acetate	1.90(s)
7	Acetone	2.221(s)
8	ADP	8.534(s), 8.261(s), 6.122(d), 4.52(m), 4.35(m), 4.00(m)
9	Alanine	1.46(d), 3.805(q)
10	AMP	8.596(s), 8.25(s), 6.12(d), 4.49(t), 4.355(m), 4.02(m)
11	Asparagine	2.777(dd), 2.89(dd), 3.98(dd)
12	Aspartate	2.66(dd), 2.79(dd), 3.916(dd)
13	ATP	8.52(s), 8.25(s), 6.128(d), 4.56(t), 4.41(m), 4.23(m), 4.30(m)
14	Beta-alanine	3.16(t), 2.54(t)
15	Betaine	3.263(s), 3.885(s)
16	Caprate	0.839(t), 1.281(m), 1.528(m), 2.519(m)
17	Cholate	0.711(s), 0.905(s), 0.960(d), 1.00(m), 1.16(m), 1.130(m), 1.140(m), 1.491(m), 1.568(m), 1.623(m), 1.75(m), 1.89(m), 2.01(m), 2.05(m), 2.23(m), 3.501(m), 3.39(m), 4.06(m)
18	CDP	8.083 (s), 7.987 (s), 6.102 (d), 5.99 (d)
19	Citrate	2.53(d), 2.666(d)
20	Citrulline	1.526(m), 1.59(m), 1.84(m), 1.889(m), 3.147(m), 3.126(m), 3.74(m)
21	Dimethylamine	2.718(s)
22	DSS	0.0(s), 0.62(t), 1.75(m), 2.91(t)
23	dTTP	7.68(s), 6.333(m), 4.618(m), 4.22(m), 4.17(m), 2.38(m), 1.918(s)
24	Ethanol	1.185(t), 3.664(q)
25	Fructose 1,6-biphosphate	4.19 (m), 3.80 (m), 3.70 (m)
26	Formate	8.44(s)
27	Fumarate	6.50(s)
28	Glucose-1-Phosphate	5.45(dd), 3.908(m), 3.86(m), 3.76(m), 3.487(m), 3.398(t)
29	Glutamate	2.03(m), 2.10(m), 2.34(m), 3.75(dd)

30	Glutamine	2.141(m), 2.459(m), 3.76(t)
31	Glycerol	3.768 (tt), 3.64 (dd), 3.546 (dd)
32	GTP	8.110(s), 6.12(s),5.96(s), 4.54(m), 4.35(m), 4.24(m)
33	Homoserine	2.01(m), 2.16(m), 3.77(m), 3.85(dd)
34	IMP	8.553(s), 8.231(s), 6.136(d), 4.51(m), 4.36(m), 4.01(m)
35	isoleucine	0.926(t), 0.992(d), 1.248(m), 1.457(m), 1.968(m), 3.661(d)
36	Lactate	4.096(q), 1.313(d)
37	Leucine	3.721(m), 1.701(m), 0.94(m)
38	Lysine	3.74(t), 3.02(t), 1.89(m), 1.71(m), 1.46(m)
39	Malonate	3.09(s)
40	Maltose	5.41(d), 5.39(d), 5.211(d), 3.96(m), 3.93(m), 3.84(m), 3.76(m), 3.70(m), 3.66(m),3.62(m),
41	Methylamine	2.573(s)
42	N-Acetyl-glucosamine	8.177(d), 8.077(m), 5.19(d), 3.90(m), 2.05(s), 2.03(s)
43	NAD	9.314(s), 9.121(d), 8.824(d), 8.406(s), 8.184(m), 8.154(s), 6.07(d), 6.021(d), 4.522(m),
44	NADP+	9.281(s), 9.08(d), 8.80(d), 8.41(s), 8.18(m), 8.13(s), 6.112(d), 6.022(d), 4.99(q), 4.60(t),
45	NADPH	8.459 (s), 8.248 (s), 6.96 (s), 6.61 (br), 6.18 (d), 5.965 (d), 4.597 (t)
46	Phenyl alanine	7.42(m), 7.36(m), 7.32(m), 3.98(dd), 3.27(m), 3.13(m)
47	Succinate	2.385(s)
48	Threonine	4.244(m), 3.582(d), 1.313(d)
49	Trehalose	5.18(d), 3.84(m), 3.80(m), 3.76(m), 3.64(dd),3.43(t)
50	Tryptophan	7.72(d), 7.53(d), 7.32(s), 7.24(m), 7.19(m), 4.04(dd), 3.47(dd), 3.29(dd)
51	Tyrosine	7.17(d), 6.87(d), 3.93(dd), 3.18(dd),3.04(dd)
52	UDP Galactose	7.93(d), 5.97(m), 5.63(dd), 4.36(m), 4.25(m), 4.15(m),4.02(d), 3.90(dd), 3.805(dt), 3.75(m)
53	UDP-Glucose	7.925(d), 5.977(m), 5.593(dd), 4.36(m), 4.27(m), 4.24(m), 4.189(m), 3.877(m), 3.77(m),
54	UDP-N-Acetylglucosamine	8.177(d), 8.077(m), 5.19(d), 4.70(d), 3.904(m), 3.841(m), 3.784(m), 3.74(m),
55	UMP	8.07(d), 5.98(m), 4.41(t), 4.34(t), 4.26(m), 3.97(m)
56	Valine	0.996(d), 1.047(d), 2.281(m),3.617(d)

Table 5.3a: Fold change (FC), p-values and FDR-values of Control vs Acidic stress.

S.No	Metabolite	log ₂ (FC)	FDR	-log ₁₀ (p)
1	Alanine	1.97	1.52E-09	8.82
2	L-Tyrosine	1.83	1.52E-09	8.82
3	L-Phenylalanine	0.62	1.52E-09	8.82
4	D-Maltose	-1.62	1.31E-08	7.88
5	L-Valine	1.76	4.36E-08	7.36
6	L-Homoserine	1.03	9.03E-08	7.04
7	D-Glutamic acid	0.87	9.03E-08	7.04
8	/--erythro-Isoleucine	1.43	2.43E-07	6.61
9	1-Methylnicotinamide	1.60	4.54E-07	6.34
10	Beta-Alanine	1.67	1.04E-05	4.98
11	L-Glutamine	0.77	1.04E-05	4.98
12	L-Threonine	1.70	1.86E-05	4.73
13	Thymidine 5'-triphosphate	0.76	3.35E-05	4.47
14	Succinic acid	0.73	3.38E-05	4.47
15	Uridine diphosphate galactose	0.60	7.73E-05	4.11
16	Taurine	-0.82	0.00010462	3.98
17	L-Aspartic acid	0.82	0.0001951	3.71
18	Capric acid	-1.12	0.00019709	3.71
19	Uridine 5'-monophosphate	0.71	0.00027747	3.56
20	L-Alpha-aminobutyric acid	0.81	0.0013696	2.86
21	L-Leucine	0.59	0.0034988	2.46
22	Cholic acid	0.75	0.0072496	2.14

Table 5.3b: Fold change (FC), p-values, and FDR-values of Control vs. Oxidative stress.

S.No	Metabolite	log ₂ (FC)	FDR	-log ₁₀ (p)
1	NADP	8.84	1.05E-10	9.98
2	Betaine	-8.88	7.91E-08	7.10
3	D-Maltose	-8.54	4.19E-07	6.38
4	Adenosine monophosphate	8.77	6.49E-07	6.19
5	1-Methylnicotinamide	8.56	6.49E-07	6.19
6	D-Glutamic acid	9.19	7.41E-07	6.13
7	L-Aspartic acid	-5.87	8.73E-07	6.06
8	L-Lysine	-9.05	1.09E-06	5.96
9	3-Methyl-2-oxovaleric acid	-8.00	1.27E-06	5.90
10	UDP-N-acetylglucosamine	-8.63	1.71E-06	5.77
11	NAD	6.11	7.46E-06	5.13
12	Adenosine triphosphate	3.86	1.87E-05	4.73
13	Capric acid	-3.53	2.99E-05	4.52
14	Methylamine	-8.72	6.51E-05	4.19
15	2-Hydroxy-3-methylpentanoic acid	-7.65	0.00014706	3.83
16	Nicotinamide N-oxide	3.65	0.00024024	3.62
17	Isoleucine	-4.00	0.0035162	2.45
18	Dimethylamine	-2.01	0.0064698	2.19
19	L-Glutamine	-2.23	0.0086985	2.06
20	Malonic acid	2.29	0.030726	1.51
21	Inosinic acid	1.27	0.038132	1.42

Table 5.3c: Fold change (FC), p-values, and FDR-values of Control vs. starvation stress.

S.No	Metabolite	log ₂ (FC)	FDR	-log ₁₀ (p)
1	3-Hydroxyisovaleric acid	-10.05	2.14E-14	13.67
2	L-Phenylalanine	9.30	1.02E-13	12.99
3	D-Maltose	9.87	3.95E-13	12.40
4	L-Asparagine	13.23	4.96E-13	12.31
5	Uridine diphosphategalactose	9.59	8.91E-13	12.05
6	Ethanol	-11.57	2.58E-12	11.59
7	Adenosine monophosphate	11.28	4.58E-12	11.34
8	Thymidine 5'-triphosphate	9.96	4.85E-12	11.32
9	Acetone	-10.38	4.15E-11	10.38
10	1-Methylnicotinamide	10.08	5.63E-10	9.25
11	Taurine	-11.47	8.78E-10	9.06
12	Uridine diphosphate-N-acetylglucosamine	9.62	1.81E-09	8.74
13	Uridine 5'-monophosphate	9.97	5.81E-09	8.24
14	L-Aspartic acid	12.42	6.92E-09	8.16
15	Beta-Alanine	11.79	1.17E-08	7.93
16	L-Glutamine	12.29	1.67E-08	7.78
17	Alanine	11.11	2.31E-08	7.64
18	L-Tyrosine	-10.78	5.26E-08	7.28
19	Malonic acid	-12.12	2.01E-07	6.70
20	Capric acid	-10.70	2.01E-07	6.70
21	NADP	9.28	2.43E-07	6.61
22	L-Lysine	-10.72	2.71E-07	6.57
23	Citric acid	12.11	3.13E-07	6.50
24	Succinic acid	11.20	5.05E-07	6.30
25	NAD	10.87	5.28E-07	6.28
26	2-Hydroxy-3-methylpentanoic acid	-11.45	5.35E-07	6.27
27	Uridine diphosphate glucose	10.05	1.30E-06	5.89
28	Glucose 1-phosphate	10.11	6.70E-06	5.17
29	ADP	7.67	7.19E-06	5.14
30	Betaine	-5.11	1.23E-05	4.91
31	Inosinic acid	8.74	1.25E-05	4.90
32	Adenosine triphosphate	9.86	2.34E-05	4.63
33	/--erythro-Isoleucine	3.67	2.34E-05	4.63
34	L-Leucine	-11.25	2.66E-05	4.58
35	Sucrose	9.95	0.00010247	3.99
36	Nicotinamide N-oxide	2.18	0.00040799	3.39
37	3-Methyl-2-oxovaleric acid	-8.53	0.00047503	3.32
38	Formic acid	-9.65	0.00062956	3.20
39	D-Glutamic acid	-4.25	0.00063653	3.20

40	Trehalose	5.66	0.0024417	2.61
41	Methylamine	-10.65	0.0025862	2.59
42	L-Homoserine	2.50	0.0054186	2.27
43	L-Threonine	10.75	0.0056602	2.25
44	Acetic acid	-1.76	0.0062837	2.20
45	L-Valine	1.34	0.028662	1.54
46	Citrulline	-1.06	0.047419	1.32
47	L-Lactic acid	-2.16	0.049403	1.31

Table 5.4: Primers used in the study

Gene name	Sequence (5'-3')	Annealing Temperature (°C)
Trehalose biosynthesis and utilization pathway in <i>Mycobacterium smegmatis</i>		
MSMEG_6515treSRTFp	TCGGTATCGACGGTTTCC	57
MSMEG_6515treSRTRp	GGAAGTGGAACGCCATGT	57
MSMEG_4916glgERTFp	AGATCCTGCAGATGTTCGAA	57
MSMEG_4916glgERTRp	AACTGCTGCGCGCCTCCA	59
MSMEG_4918glgBRTFp	AACCTCATCGACTACCGC	59
MSMEG_4918glgBRTRp	AGTGGTTGAACTCGCCGAT	57
MSMEG_6514Mak-Pep2RTFp	ACGCCATCTTCAAGCTCTT	57
MSMEG_6514Mak-Pep2RTRp	TGCTGGTGGTCGCCATGT	57
MSMEG_4535TreHRTFP	CAGTGGATCAACGTCGGTG	57
MSMEG_4535TreHRTRP	GACAGGTGGTGCAGCTCTTC	57
Endogenous controls <i>Mycobacterium smegmatis</i>		
MSMEG_2758 sigAFP	GAAGACACCGACCTGGAAC	55
MSMEG_2758 sigA RP	GACTCTTCCTCGTCCCACAC	55
Trimethylamine Biosynthesis synthesis pathways in <i>Mycobacterium smegmatis</i>		
2,4-DCA MSMEG5124FP	AGATCGTCTGTTTGCAAC	58
2,4-DCAMSMEG5124RP	GATCACTCAGCAGCGGAC	58
YeaXMSMEG_4371RTFP	AGAGATCCACGAGCAGTT	57
YeaXMSMEG_4371RTRP	CTGCGGGAGCACCTCGAC	57
TR2Fe-2SMSMEG_0657FP	TGCTCGTCCATCCATCCCG	57
TR2Fe-2SMSMEG_0657RP	GACGGCCAGCGGTGTCAT	57

Table 6.4: List of chemical shift assignment of abundant water-soluble metabolites extracted from potato tuber samples

S. No.	Metabolite	CID Code	Chemical shift	Assignment
1	3-hydroxyisobutyrate	87	1.07 (d)	β -CH ₃
			2.47 (m)	α -CH
2	4-Aminobutyrate	119	1.89 (m)	β -CH ₂
			2.29 (t)	α -CH ₂
			3.00 (t)	γ -CH ₂
3	Adenosine	60961	6.07 (d)	C ₂ H
4	Alanine	5950	3.80 (q)	α -CH
			1.46 (d)	β -CH ₃
5	Allantoin	204	5.37 (s)	CH
6	Arginine	6322	1.68 (m)	γ -CH ₂
			1.89 (m)	β -CH ₂
			3.23 (t)	δ -CH ₂
			3.76 (t)	α -CH
7	Ascorbate	54670067	4.51 (d)	C ₄ H
8	Asparagine	6267	2.84 (m)	β -CH ₂
			2.94 (m)	β -CH ₂
			3.99 (dd)	α -CH
9	Aspartate	5960	2.66 (dd)	β -CH ₂
			2.79 (dd)	β -CH ₂
			3.91 (dd)	α -CH
10	Chlorogenate	1794427	6.40 (d)	C ₁₅ H
			6.95 (d)	C ₂₁ H
			7.13 (dd)	C ₂₂ H
			7.19 (d)	C ₁₈ H
			7.66 (d)	C ₁₆ H
11	Choline	305	3.20 (s)	N(CH ₃) ₃
			3.52 (m)	NCH ₂
			4.06 (m)	OCH ₂
12	Citrate	311	2.67 (d)	CH ₂
			2.76 (d)	CH ₂
13	DSS	74873	0.00 (s)	Si(CH ₃) ₃
			0.62 (t)	γ -CH ₂
			1.75 (m)	β -CH ₂

			2.91 (t)	α -CH ₂
14	Formate	283	8.44 (s)	CH
15	Fructose	2723872	3.54 (m)	C ₁ H
			3.70 (m)	C ₁ H
			3.88 (dd)	C ₃ H
			4.01 (m)	C ₄ H
			4.11 (d)	C ₅ H
16	Fumarate	5460307	6.50 (s)	CH
17	Galactose	6036	4.57 (d)	C ₂ H
			5.25 (d)	C ₂ H
18	Glucose	5793	3.23 (dd)	C ₃ H
			3.39 (m)	C ₅ H
			3.45 (m)	C ₆ H
			3.52 (m)	C ₃ H
			3.72 (m)	C ₄ H/C ₁₁ H
			3.82 (m)	C ₁₁ H/C ₆ H
			3.88 (dd)	C ₁₁ H
			4.63 (d)	C ₂ H
			5.22 (d)	C ₂ H
19	Glutamate	33032	2.03 (m)	β -CH ₂
			2.10 (m)	β -CH ₂
			2.34 (m)	γ -CH ₂
			3.75 (dd)	α -CH
20	Glutamine	5961	2.14 (m)	β -CH ₂
			2.459 (m)	γ -CH ₂
			3.76 (t)	α -CH
21	Glycine	750	3.56 (s)	CH ₂
22	Isoleucine	6306	0.92 (t)	δ -CH ₃
			0.99 (d)	β -CH ₃
			1.24 (m)	γ -CH ₂
			1.45 (m)	γ -CH ₂
			1.97 (m)	β -CH
			3.66 (d)	α -CH
23	LDL		0.84 (t)	CH ₃ (CH ₂) _n
			1.25 (m)	(CH ₂) _n
24	Leucine	6106	0.94 (d)	δ -CH ₃

			0.95 (d)	δ -CH ₃
			1.68 (m)	β -CH ₂
			1.70 (m)	γ -CH
			3.71 (t)	α -CH
25	Lysine	5962	1.49 (m)	γ -CH ₂
			1.72 (m)	δ -CH ₂
			1.89 (m)	β -CH ₂
			3.01 (t)	ϵ -CH ₂
			3.74 (t)	α -CH
26	Malate	525	2.37 (dd)	β -CH ₂
			2.66 (dd)	β -CH ₂
			4.29 (dd)	α -CH
27	Mannose	18950	4.89 (d)	C ₂ H
			5.17 (d)	C ₂ H
28	Methanol	887	3.34 (s)	CH ₃
29	Methionine	6137	2.12 (s)	S-CH ₃
			2.17 (m)	β -CH ₂
			2.63 (t)	S-CH ₂
			3.85 (t)	α -CH
30	Myo-inositol	892	3.27 (t)	C ₅ H
			3.54 (dd)	C ₁ H/C ₃ H
			3.62 (dd)	C ₄ H/C ₆ H
			4.05 (t)	C ₂ H
31	Phenylalanine	6140	7.42 (m)	C ₃ H/C ₅ H
			7.36 (m)	C ₂ H/C ₆ H
			7.32 (m)	C ₄ H
			3.98 (dd)	α -CH
			3.27 (m)	β -CH ₂
32	Proline	145742	4.11 (t)	α -CH
			3.40 (m)	δ -CH ₂
			3.33 (m)	δ -CH ₂
			2.34 (m)	β -CH ₂
			2.07 (m)	β -CH ₂
			2.01 (m)	γ -CH ₂
33	Pyroglutamate	7405	2.02 (m)	C ₄ H
			2.40 (m)	C ₃ H

			2.49 (m)	C ₄ H
			4.16 (m)	C ₅ H
34	Serine	5951	3.83 (m)	α-CH
			3.94 (m)	β-CH ₂
			3.98 (m)	β-CH ₂
35	sn-glycero-3-phosphocholine	657272	3.21 (s)	C ₁₃ H/C ₁₄ H
			4.31 (m)	C ₃ H
36	Sucrose	5988	3.46 (t)	C ₁₀ H
			3.55 (m)	C ₁₂ H
			3.75 (m)	C ₁₁ H
			3.81 (m)	C ₁₇ H/C ₁₉ H
			3.88 (m)	C ₅ H
			4.04 (t)	C ₄ H
			4.20 (d)	C ₃ H
			5.40 (d)	C ₇ H
37	Threonine	6288	1.32 (d)	γ-CH ₃
			3.58 (d)	α-CH
			4.24 (m)	β-CH ₂
38	Trigonelline	5570	4.42 (s)	C ₉ H
			8.07 (t)	C ₄ H
			8.82 (m)	C ₅ H/C ₃ H
			9.11 (s)	C ₁ H
39	Tryptophan	6305	3.29 (m)	CH ₂
			4.05 (m)	CH
			7.19 (t)	C ₅ H/C ₆ H
			7.27 (t)	C ₅ H/C ₆ H
			7.31 (s)	C ₂ H
			7.53 (d)	C ₇ H
			7.72 (d)	C ₄ H
40	Tyrosine	6057	7.17 (d)	C ₂ H/C ₆ H
			6.87 (d)	C ₃ H/C ₅ H
			3.93 (dd)	α-CH
			3.18 (dd)	β-CH ₂
			3.04 (dd)	β-CH ₂
41	Uridine	6029	7.85 (d)	C ₁₁ H

			5.9 (d)	C ₂ H
			5.89 (d)	C ₁₀ H
			4.34 (dd)	C ₃ H
			4.22 (dd)	C ₄ H
			4.12 (m)	C ₅ H
			3.9 (dd)	C ₁₄ H
			3.8 (dd)	C ₁₄ H
42	Valine	6287	0.99 (d)	CH ₃
			1.04 (d)	CH ₃
			2.28 (m)	β-CH
			3.61 (d)	α-CH
43	U1		8.22 (m)	
			9.00 (d)	
			9.10 (d)	
			9.33 (s)	
44	U2		7.27 (s)	
			8.39 (s)	
45	U3		3.85	
			3.95	
			4.02	
			5.14 (d)	
46	U4		3.82	
			3.89	
			3.97	
			4.03	
			4.99	
47	U5		3.66	
			4.29 (dd)	

Appendix 2

Questionnaire, Consent form and Approvals

CONSENT FORM

I have been fully informed about the studies regarding “**Micronutrient Deficiency and Type 2 Diabetes in the Indian population**” conducted by Armed Forces Medical College, Pune and University of Pune. I have agreed to participate in the study voluntarily, of my own free will, without any pressure or monetary inducement. It has been explained to me that if I decide to opt out of the study at a later date for any reason, I may do so and will not affect the routine care given to me in anyway.

I understand that apart from routine checkups such as height, weight, waist-hip circumference, blood pressure, etc., I will have to undergo urine examination and several blood test including miRNA and DNA studies. I may in the future be asked to attend for additional investigations but will be free to refuse these if I do not want to take part.

It is understood that

- these studies would only be for biochemical and epigenetic analysis of diabetes and diabetes related conditions;
- that studies will all be done anonymously; and
- the results of the studies will not be available to me.

The cost of tests to be carried out at Department of Biotechnology and Armed Forces Medical College will be borne by Department of Biotechnology, University of Pune.

I have read the consent form carefully. All my queries and doubts have been satisfied. I am willing to participate in the study voluntarily and will cooperate fully for the various tests that will be conducted during the study.

Name _____

Name of the witness _____

Signature & date

Signature & date

I. GENERAL INFORMATION

1. Name of the subject: _____

2. Sample Code: _____ Patient _____ Control

3. Address: _____

4. Mobile: _____

5. E-mail Address: _____

6. Date of Birth: _____ (Known / derived) ____/____/_____

7. Occupation details: _____

8. Gender: _____ Male _____ Female

II. PERSONAL INFORMATION

1. Weight _____ Kgs

2. Height _____ Ft _____ Inches

3. Waist Measurement _____ Inches

4. BMI _____

5. Diet

a. Lacto-vegetarian _____ Yes _____ No

b. Ovo-vegetarian _____ Yes _____ No

c. Non-vegetarian _____ Yes _____ No

d. Smoking _____ Yes _____ No

e. Tobacco _____ Yes _____ No

f. Alcohol _____ Yes _____ No

g. Drugs _____ Yes _____ No

6. Family History of Diabetes: _____ Yes _____ No _____ Don't Know

	Alive	Diabetes	Age at Diagnosis (y)
1. Father			
2. Mother			
3. Brother (s)			
4. Sister (s)			
5. Children			
a. Son			
b. Daughter			
6. Paternal Grandfather			
7. Paternal Grandmother			
8. Paternal Uncle			
9. Paternal Aunt			
10. Maternal Grandfather			
11. Maternal Grandmother			
12. Maternal Uncle			
13. Maternal Aunt			
14. Spouse			

8. Exposure to sunlight _____ Hours

9. Any other illness in the past _____ Yes _____ No

Doctor's Name:

Name of the person taking information:

Date:

Place:

Signature:

III. BIOCHEMICAL PARAMETERS

Laboratory Investigations

1. Fasting Glucose:
2. 2 hr post-plasma prandial glucose:
3. Hemoglobin:
4. Fasting serum Cholesterol:
5. Fasting Triglycerides:
6. Fasting HDL Cholesterol:
7. HbA1c
8. Creatinine
9. Fasting Urine Glucose:
10. Fasting Urine Protein:
11. HOMA-IR
12. Plasma Insulin
13. Vitamin B12
14. Vitamin D3
15. Folate

Name of the Laboratory:

Name of pathologist performing the assay:

Recorder's name:

Recorder's Signature:

IV. MOLECULAR STUDIES

1. DNA

a. Concentration _____ ng/ul

b. 260/280 _____

2. PLASMA _____ ul

3. SERUM _____ ul

4. RNA

a. Concentration _____ ng/ul

b. 260/280 _____

5. miRNA

a. Concentration _____ ng/ul

b. 260/280 _____

Prof. R.K.Mutatkar
Chairman
rkm@unipune.ernet.in
9422009630



Savitribai Phule University of Pune
Ethics Committee

Interdisciplinary School
of Health Sciences
SPPU,
Pune 411007.

SPPU/SHS/2014/1195

To
Dr Shilpi Sharma
Department of Biotechnology
Savitribai Phule Pune University
Pune.

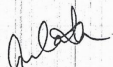
17th December 2014
13

Subject : Your proposal titled "Micronutrient deficiency and diabetes in the Indian population--- understanding the role of non-coding RNAs".

Dear Madam,

Following is the decision of the Ethics committee:

1. EC approved the studies in principle for basic research, with the condition that any potential patent/sample transfer/ commercial applications would need to be resubmitted for IEC approval.


Secretary,
Ethics Committee
Savitribai Phule Pune University

ARMED FORCES MEDICAL COLLEGE
CERTIFICATE FROM ETHICS COMMITTEE REVIEWERS

Title: Micronutrient deficiency and diabetes in the Indian population.

Name of the Principal worker: Col HS Batra

- 3. Research design:
 - (g) Scientifically sound enough to address the objectives stated: ✓ Y/N
 - (h) Relevant to contribute to further knowledge: ✓ Y/N
 - (i) Research of Armed Forces significance: Y/N *NA*
- 12. Subject/ participant selection
 - (e) Inclusion/ exclusion criteria addressed appropriately ✓ Y/N
 - (f) Spl groups/ vulnerable persons adequately protected: ✓ Y/N
- 13. Risks and benefits
 - (e) Physical/social/ psychological risk/ discomfort minimized/ addressed: ✓ Y/N
 - (f) Overall risk benefit ratio: Acceptable/ Unacceptable ✓
- 14. Privacy and confidentiality
 - (e) Privacy of research participant maintained: ✓ Y/N
 - (f) Confidentiality of research participant maintained: ✓ Y/N
- 15. Informed consent form (ICF)
 - (c) ICF component/ process addressed properly: ✓ Y/N
- 16. Conflict of interest: Y/N ✓

DECISION: APPROVED/ NOT APPROVED

DECISION: UNANIMOUS/ DISSENT

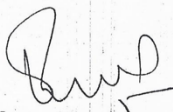
IF DISSENT: REASONS FOR DISSENT

COMMENTS IF ANY:

SIGNATURES OF ETHICS COMMITTEE MEMBERS

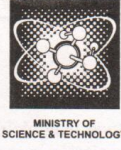
INSTITUTIONAL ETHICAL COMMITTEE
CLEARANCE ACCORDED / NOT ACCORDED ✓

Unit: AFMC, Pune
Date: 2014


(Colvani Suryam)

Secretary,
Medical Research Cell &
Institutional Ethical Committee
Armed Forces Medical College
Pune - 411 001

08 Sep 2014



No. BT/BS/17/582/2014-PID

Dated: 22.06.2018

To

Dr. Siddhesh Kamat,
Assistant Professor,
Member Secretary-IBSC,
Indian Institute of Science and Education Research (IISER),
Dr. Homi Bhabha Road, Pashan, Pune-411 008.

Subject: Applications submitted by Indian Institute of Science and Education Research (IISER), Pune for the information and records of Review Committee on Genetic Manipulation (RCGM) on the projects involving recombinant DNA technology.

Gentleman,

It is informed that the following applications to carry out research work were considered & noted by the Review Committee on Genetic Manipulation (RCGM) in its **165th meeting held on 22.05.2018:**

- i) Randomized phase 2 trial of Rosuvastatin with neo-adjuvant chemo-radiation on patients with rectal cancer.
- ii) Regulation and role of Aurora Kinase A in Ra1 mediated oncogenic Ras dependent anchorage independence.
- iii) Role of Azadiradione in the regulation of diet induced metabolic syndrome in rat model.
- iv) Metabolomic analysis of skeletal muscle cells (L6) and adipocyte cells (3T3 L1) in response to glucolipotoxic conditions with and without small molecules and natural compounds used as potential anti-diabetic agents
- v) Identification of metabolic biomarkers from Micronutrient deficient Type 2 Diabetic population from India.
- vi) Development of a stable and inducible CRISPR-Cas system for high throughput site specific genome editing in *Plasmodium falciparum*.
- vii) *In vivo* imaging of enzyme activity with exquisite specificity using Activity based Reporter Gene Technology (AbRGT).
- viii) *In vivo* screening and analysis of Bar domain family in membrane cytoskeletal interactions in epithelial morphogenesis and synaptic membrane remodelling in *Drosophila*.
- ix) Fibulin-1, a novel regulator of EGFR signalling in lung cancer.
- x) systematic editing of the *Drosophila* genome by Crispr/ Cas9 to uncover biological roles for SUMOylation in the immune response.
- xi) The effect of various types of cofactors influencing YAP/TAZ related cancers in humans.
- xii) Protein folding and aggregation.
- xiii) Understanding the mechanism of protein folding in neuro-degenerative disease.
- xiv) The process of protein aggregation in cells with alpha-synuclein as a model system.
- xv) The use of mouse prion protein as a model system to address the fundamental questions of protein folding, unfolding and aggregation.
- xvi) Hsp40, Hsp70 and Hsp60 chaperone modulated aggregation kinetics of tau protein.
- xvii) Mapping signalling pathways by bioactive lysophosphatidylserine lipids.
- xviii) The role of Dmon 1 in synaptic signalling and glutamate receptor regulation.

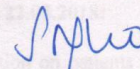
- xix) Molecular and chemical ecology of the biological control of diamondback moth: revealing the significance of the infochemical communication between bottom-up and top-down factors.
- xx) Molecular and chemical ecology of multitrophic interactions involving plants, herbivores and their natural enemies.
- xxi) Molecular and chemical ecology of the biological pest control: plant-mediated RNAi based investigation of counter adaptations of diamondback moth and its parasitoid against Brassicaceae's glucosinolate-myrosinase defence system.
- xxii) Is chlorogenic acid, brinjal's major phenolic, a key to the biocontrol of brinjal shoot and fruit borer.
- xxiii) Understanding chemical and molecular basis of host specialization in a native host plant-insect herbivore system.

2. You are required to comply with the Regulations and Guidelines for Recombinant DNA Research and Biocontainment, 2017.

3. Please provide the information on the above project for updation on the website.

Please acknowledge the receipt of the letter.

Yours faithfully,



(S. R. Rao)
Member Secretary, RCGM &
Scientist-'H', DBT

- i) Biocontrol phase 2 trial of *Rosuvastatin* with neo-agrotrans chemo-... with breast cancer.
- ii) Significance and role of Aurora Kinase A in Ras mediated... oncogene independence.
- iii) Role of Aspirin/Aspirin in the regulation of diet induced metabolic syndrome in rat model.
- iv) Metabolomic analysis of skeletal muscle cells (C2C12) and adipocyte cells (3T3-L1) in response to glucocorticoid conditions with and without small molecules and natural compounds as potential anti-diabetic agents.
- v) Identification of metabolic biomarkers from Microarray data deficient Type 2 Diabetes population from India.
- vi) Development of a stable and inducible CRISPR-Cas system for high throughput genome editing in *Pichia pastoris*.
- vii) In vivo imaging of enzyme activity with exquisite specificity using Activity based Reporter Gene Technology (AbRGT).
- viii) In vivo screening and analysis of Bar domain family in membrane cytoskeletal interactions in epithelial neurofibromin 2 and cyclin dependent kinase 5 in Drosophila.
- ix) ...
- x) ...
- xi) ...
- xii) ...
- xiii) ...
- xiv) ...
- xv) ...
- xvi) ...
- xvii) ...
- xviii) ...
- xix) ...
- xx) ...
- xxi) ...
- xxii) ...
- xxiii) ...