# A Molecular Phenomics Approach to Complex Multisystemic Diseases

**Sofina Begum BSc (Hons), MRes**

Department of Metabolism, Digestion and Reproduction,

Faculty of Medicine

Imperial College London

May 2021

**A thesis submitted to Imperial College London in partial requirements for the degree of Doctor of Philosophy**

**Supervisors**

**Professor Elaine Holmes**

Department of Metabolism, Digestion and Reproduction, Imperial College London

**Professor Jeremy Nicholson**

Department of Metabolism, Digestion and Reproduction, Imperial College London

**Dr. Isabel Garcia-Perez**

Department of Metabolism, Digestion and Reproduction, Imperial College London

# Declaration of Originality

I, Sofina Begum, confirm that the work presented in this thesis is my own work. Where information has been derived from other sources, I confirm that this has been indicated. Portions of this thesis have been published in peer reviewed journals; this has been indicated where this is the case.

# Copyright Declaration

'The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.'

# Production of Data

All spectroscopic data presented in this thesis were acquired from the Australian National Phenome Centre, Murdoch University, Health Futures Institute, 5 Robin Warren Drive, Murdoch, Western Australia 6150, except for the PBC chapter where the spectroscopic were acquired by Dr. Alexandros Pechliavanis in collaboration with the National Phenome Centre, Imperial College London and the 16S RNA data were conducted by Dr. Julie McDonald, at the Imperial College London. All informatics, data analyses and data integration were performed by myself.

# Acknowledgements

Firstly, I'd like to sincerely thank my supervisor Prof. Elaine Holmes for supporting me through my PhD. For your patience (apparently there's a lot of it), mentorship, tutelage, locking me away for a year and drinking tea with me. For pushing me forward even when things looked bleak. This would not have been possible without you. To Prof. Jeremy Nicholson, thank you for your mentorship, analytical expertise and for teaching me how to be assertive; I've grown a lot in the last 3 years.

Thank you to Prof. Laki Buluwela, Ms. Nousheen Tariq and the Imperial College London MRC DTP, for supporting my PhD. I've been able to achieve so much through the course of my PhD, both personally and academically, because of your support. To Dr. Isabel Garcia-Perez and Dr. Joram Posma, thank you for mentoring me with humour. To the team at the Australian National Phenome Centre, thank you for your infectious enthusiasm. Thank you for sharing this journey with me. I would like to thank Dr. Mark Fear, Dr Luke Whiley, Dr Ruey Leng Loo, Dr. Reika Masuda, Dr. Sam Lodge and Prof Julien Wist for their analytical expertise and help in reviewing my thesis.

To my friends, saddle club, siblings and family, thank you for continually supporting me even when I've been absent for months and months. I miss you too. To my mother and adoptive mothers around the world, thank you for always enduring and supporting me. Here's to another chapter?

# ABSTRACT

The age of -omics has set a precedent for precision medicine, yet one of the most prevalent challenges is multi-omics data integration, with no platform yet that can efficiently integrate clinical and multi-omic data. This thesis uses several integrative data strategies to deconvolute complex multisystemic disease states, utilising clinical, cytokine profiling, metagenomic and metabonomic data to gain greater depth and understanding. High-throughput metabolic profiling methods have been employed for in-depth systemic analysis of SARS-CoV-2 infection, Primary Biliary Cholangitis and paediatric burn injury. Correlation analysis and derived clustering methods were utilised for data integration, following multivariate analyses and reveal disease state specific signatures. Similarly, different biofluids (plasma, urine, faeces) contributed specific complementary information. Several systemic biomarkers were identified for each disease, each with hallmarks of organ dysfunction. This work highlights the value of data integration in heavily confounded data and disease.

# Abbreviations

**SARS;** Severe Acute Respiratory Syndrome
**SARS-CoV;** Severe Acute Respiratory Syndrome novel coronavirus
**MERS;** Middle East Respiratory Syndrome
**MERS-CoV;** Middle East Respiratory Syndrome novel coronavirus
**COVID-19;** coronavirus disease 2019
**SARS-CoV-2;** SARS coronavirus 2
**WHO;** World Health Organisation
**RBD;** receptor binding protein
**ACE2;** Angiotensin Converting Enzyme 2
**ATP;** adenosine triphosphate
**PPE;** personal protective equipment
**ARDS;** acute respiratory distress syndrome
**rRT-PCR;** reverse transcriptase-polymerase chain reaction
**DAMPS;** damage associated molecular patterns
**IL-;** interleukin
**MIP;** macrophage inflammatory protein
**IFN;** interferon
**LDL;** low-density lipoprotein
**VLDL;** very low-density lipoprotein
**HDL;** high-density lipoprotein
**ALT;** Alanine Aminotransferase
**ALP;** Alkaline Phosphatase
**GGT;** Gamma Glutamyl Transferase
**PCA;** principal component analysis
**PLS-DA;** partial least square discriminant analysis
**OPLS-DA;** orthogonal partial least square discriminant analysis
**VIP;** variable importance in projection
**SpC;** supramolecular phospholipid composite
**BMI;** body mass index
**ppm;** parts per million

# Contents Page

**Title Page**.....1
**Declarations**.....2
**Acknowledgements**.....4
**Abstract**.....5
**Abbreviations**.....6
**Contents Page**.....7
**List of Figures and Tables**.....9

**Chapter One: Introduction**.....**13**

**Chapter Two: Methods**.....**34**

**Chapter Three: Integrative Modelling of Clinical And Metabolic Phenomic Data - A Systems Approach To Characterising SARS-CoV-2 Infection**.....**40**

**3.1Introduction**.....41
**3.2 Materials and Methods**.....52
**3.3 Results**.....61
**3.4 Discussion**.....104
**3.5 Conclusions**.....117

**Chapter Four: UK- Primary Biliary Cholangitis (UK-PBC) Nested Cohort Study**.....**118**

**1 Introduction**.....119
**2 Materials and Methods**.....132
**3 Results**.....141
**4 Discussion**.....156
**5 Conclusions**.....164

**Chapter Five: Identification of a Molecular Signature Reflecting Prior Burn-Injury in Children**.....**165**

**1 Introduction**.....166
**2 Materials and Methods**.....184
**3 Results**.....193
**4 Discussion**.....236
**5 Conclusions**.....245

**Chapter Six: Discussion**.....**246**

**Chapter Seven: References**.....**257**

**Appendix I: Published paper I……279**

**Appendix I: Published paper II……281**

**Appendix I: Published paper III……283**

**Appendix I: Published paper IV……285**

**Appendix I: Published paper V……287**

**Appendix I: Published paper VI……289**

**Appendix I: Published paper VII……291**

**Appendix II: Paper I under review……293**

**Appendix III: Paper I under review……296**

# List List of Figures and Tables

**Chapter One: Introduction**

**Table 1.1:** Summary of NMR advantages and limitations……22
**Figure 1.1:** Schematic of how variance is calculated**.** Principal components analysis (PCA) specifies a new coordinate system in which every point is assigned an (x,y) value……25

**Chapter Two: Methods**

**Table 2.1:** Side-by-side comparison of all datasets available from each study……34
**Figure 2.1:** Outline of analytical techniques used for the Western Australia COVID-19 study ……35
**Figure 2.2:** Outline of analytical techniques used for the UK-PBC Nested Cohort study……37
**Table 2.2** UK-PBC study cohort outline and criteria……38
**Figure 2.3:** Outline of analytical techniques used for the Paediatric Burns study……39

**Chapter Three: Integrative Modelling of Clinical And Metabolic Phenomic Data - A Systems Approach To Characterising SARS-CoV-2 Infection**

**3.1 Introduction**

**Figure 3.1**: Viral binding and replication process. ……42
**Figure 3.2:** The antibody response to SARS-COV-2 and disease severity……43
**Table 3.1:** Summary of exemplar COVID-19 metabolic profiling studies……51

**3.2 Methods**

**Table 3.2**: WA SARS-CoV-2 Study Cohort demographics……53
**Figure 3.3:** Symptoms presented during peak infection for SARS-CoV-2 positive patients……53

**3.3 Results**

**Figure 3.4:** Blood biochemistry results from peak of SARS-CoV-2 infection, log scaled with upper limit of normal (ULN)…… ……62
**Figure 3.5: Standard 1D NMR spectra aliphatic ($\delta$ 0.5-4.5)**: Exemplar spectra from 2 SARS-CoV-2 positive patients compared to a healthy control……63
**Figure 3.6: Standard 1D NMR spectra aromatic region ($\delta$ 5.5-9.5):** Exemplar spectra from 2 SARS-CoV-2 positive patients compared to healthy control……64
**Figure 3.7: CPMG NMR spectra aliphatic region ($\delta$ 0.5-4.5):** Exemplar spectra from 2 SARS-CoV-2 patients compared to a healthy control……65
**Figure 3.8: CPMG NMR spectra aromatic region ($\delta$ 5.5-9.5):** Exemplar spectra from 2 SARS-CoV-2 patients compared a healthy control……66
**Figure 3.9: DIRE NMR spectra aliphatic region ($\delta$ 0.5-4.5 and aromatic regions ($\delta$ 5.5-9.5):** Exemplar spectra from SARS-CoV-2 positive patients compared to healthy controls……67
**Figure 3.10:** PCA and OPLS-DA of standard 1D NMR spectra between SARS-CoV-2 patients and healthy controls. ……68
**Figure 3.11:** PCA and OPLS-DA modelling on 1D standard NMR spectra based on sex differences between SARS-CoV-2 patients and healthy controls.
……69
**Figure 3.12:** PCA and OPLS-DA modelling on 1D standard NMR based on BMI differences in SARS-CoV-2 patients……70
**Figure 3.13:** PCA and OPLS-DA of CPMG NMR spectra……71
**Figure 3.14:** PCA and OPLS-DA of DIRE NMR spectra between SARS-CoV-2 patients and healthy controls……72
**Figure 3.15:** PCA and OPLS-DA modelling of DIRE NMR spectra based on sex differences between SARS-CoV-2 patients and healthy controls……73
**Figure 3.16:** PCA and OPLS-DA modelling on DIRE NMR spectra based on BMI differences in SARS-CoV-2 patients……74
**Figure 3.17:** PCA and OPLS-DA of quantified lipoprotein subfractions. ……78
**Figure 3.18:** OPLS-DA Variable importance of Projection (VIP)…… ……79

**Figure 3.19:** Quantified lipoprotein subclass correlation analysis against quantified glycoprotein A, glycoprotein B and supramolecular compound concentrations in A) healthy controls and B) SARS-CoV-2 positive patients……82
**Figure 3.20:** Cytokine quantification from plasma. ……83
**Figure 3.21:** Hierarchically clustered immuno-metabolic spearman's correlation mapping between cytokine and chemokine measurements and quantified glycoprotein A, glycoprotein B and supramolecular compound concentrations in A) healthy controls and B) SARS-CoV-2 positive patients.
……84
**Figure 3.22:** Correlation mapping of quantified lipoproteins against hierarchically clustered cytokines in SARS-CoV-2 patients. …… 85
**Figure 3.23:** Significant plasma metabolites and metabolite ratios between SARS-CoV-2 (+) patients and healthy controls from UHPLC-MS biogenic amine and tryptophan assays…… 88
**Figure 3.24:** PCA and OPLS-DA of quantified biogenic amine and tryptophan pathway metabolites……89
**Figure 3.25:** OPLS-DA Variable importance of Projection (VIP) plot of all quantified biogenic amine and tryptophan metabolites (figure 3.24C). ……90
**Figure 3.26:** OPLS-DA models and respective loadings plot all quantified biogenic amine and tryptophan metabolites discriminating healthy controls and SARS based on sex and BMI differences, but coloured according to infection status…… ……93
**Figure 3.27:** OPLS-DA model based on sex differences- Variable importance of Projection (VIP) of all quantified biogenic amine and tryptophan metabolites (figure 21A). ……94
**Figure 3.28:** OPLS-DA model based on body mass index (BMI) - Variable importance of Projection (VIP) of all quantified biogenic amine and tryptophan metabolites (figure 21C) ……95
**Table 3.3:** Comparisons between variable importance loadings from OPLS-DA models of biogenic amines and tryptophan pathway metabolites……96
**Figure 3.29:** Hierarchically clustered Spearman's correlation of biogenic amines and tryptophan pathway metabolites against measured cytokines……98
**Figure 3.30:** Hierarchical edge clustering of significant metabolites (pvalue >0.05) and array of cytokines…99
**Figure 3.31:** Metabolite set enrichment analysis (MSEA) and network analysis……102

**Chapter Four: UK- Primary Biliary Cholangitis (UK-PBC) Nested Cohort Study**

**4.2 Methods**
**Table 4.1:** UK-PBC study cohort outline and criteria……132
**Figure 4.1:** Ethanol resonances detected at 1.1-1.3 ppm and 2.9-3.1 ppm……135
**Figure 4.2:** Ethanol-related resonances detected at 1.1-1.3 ppm and 2.9-3.1 ppm. A) Skyline plot of ethanol regions). B) Magnification of ethyl glucuronide and ethanol C) Magnification of JRES in Fig. 2.1, ethyl glucuronide and ethanol……137
**Table 4.2:** Number of samples containing ethanol and ethanol glucuronide signals……137

**4.3 Results**
**Figure 4.3:** Urine NMR multivariate analysis following removal of detected Ethanol resonances and regression analysis using Bilirubin as a regressor……141
**Figure 4.4:** Exemplar spectra with magnified region containing the C18 axial signals from bile acids……143
**Figure 4.5:** Bile Acid region- supervised and unsupervised analysis. A) PCA of all cohorts based on Bile Acid defined region……144
**Figure 4.6:** Example of highly variable chemical shifts observed in urine NMR spectra. Shift amongst 3 urine samples shown for assigned creatinine……145
**Figure 4.7:** PCA scores plot of the spectral region 7.5-8.0 ppm generated using control (Cohort E) and drug naïve PBC (cohort B) urine samples showing clustering of a subset of the control samples that contain relatively higher levels of hippurate……
……146
**Figure 4.8:** PCA scores plot of the spectral region 6.5-8.0 ppm generated from the Cohort A containing urine samples from patients with PBC treated with UDCA. Patients who responded positively to therapy are denoted in blue and non-responders are plotted in red……147

**Figure 4.9:** Exemplar $^{1}$H J-Resolved NMR urine spectrum showing assignment for hippurate, PAG and 4-cresylsulphate ……..147
**Figure 4.10:** Exemplar spectra showing variable levels of phenylacetylglutamine from the aromatic (high frequency) region $\delta$5.5-9.0……..148
**Table 4.3:** Number and percentage of samples quantified for the three selected gut microbial metabolites……..150
**Figure 4.11:** Signals relating to 4-cresylsulphate coded according to the degree of certainty of their presence in the spectrum based on the correlation coefficient with an authentic reference pattern……..150
**Figure 4.12:** Relative integrals of Hippurate, PAG and 4-cresyl sulphate stratified by cohort. No significant differences were found between PBC disease groups drug naïve (Cohort B) or UDCA treated (Cohort A) and the control group (Cohort E) ……..151
**Figure 4.13:** Relative spectral integrals of Hippurate, PAG and 4-cresyl sulphate in Cohort A, stratified by responder / non-responder. No significant differences were found between any group comparison……..152
**Figure 4.14:** Correlation matrix between Hippurate, PAG, 4-cresylsulphate and derived ratios and the microbial parameters for each of Cohort A (PBC treated), Cohort B (PBC naïve) and E (controls)……..154
**Figure 4.15:** Correlation matrix based on UDCA response (Cohort A) showing associations between Hippurate, PAG, 4-cresylsulphate and derived ratios and the microbial parameters, stratified according to response to UDCA treatment……155

**Chapter Five: Identification of a Molecular Signature Reflecting Prior Burn-Injury in Children**

**5.1 Introduction**
**Figure 5.1:** Structural framework with skin and underlying immune and vascular properties……..169
**Table 5.1:** Summary of burn injury and systemic damage studies……..178
**Table 5.2:** Summary of metabonomic elucidated biomarkers associated with burn injury and trauma……..181

**5.3 Results**
**Figure 5.2:** Significant quantified metabolites……194
**Figure 5.3:** Non-signficant quantified metabolites……195
**Figure 5.4:** Multivariate analysis of quantified biogenic amine and tryptophan pathway metabolites……196
**Figure 5.5:** Ranked variable importance of projection (VIP) plot based OPLS-DA model for burn injury and healthy control; listing the most significant variables contributing to the model ……198
**Figure 5.6:** Venn diagram comparing significant metabolites from univariate analysis and supervised multivariate analysis between burn injury and non-burn controls……..198
**Table 5.3:** List of all calculated metabolite ratios……201
**Figure 5.7:** Correlation heatmaps integrating immune cell phenotyping (mass cytometry) and circulating cytokines. A) Measurements from burn injury patients. ……202
**Figure 5.8:** Correlation heatmaps integrating biogenic amines and tryptophan pathway metabolite concentrations and circulating cytokines…… ……203
**Figure 5.9:** Correlation heatmaps integrating immune cell phenotyping (mass cytometry) and metabolites...204
**Figure 5.10: Metabolite set enrichment analysis (MSEA)**- associated pathways and networks associated with the abundance of metabolites in burn injury patients compared to healthy controls……205
**Figure 5.11: Network analysis of MSEA results**- This figure illustrates the connections between each pathway network and associations related to MSEA of metabolite abundance and proportional significance between burn injury and healthy controls……205
**Figure 5.12:** PCA and OPLS-DA of standard 1D NMR spectra between burn injury patients and healthy controls…… ……207
**Figure 5.13: PCA and OPLS-DA of CPMG NMR spectra between burn injury patients and healthy controls**……208
**Figure 5.14:** Loadings from CPMG spectra OPLS-DA model, discriminating burn injury and non-burn controls, highlighting the metabolites with the most influence in the model projection and corresponding STOCSYs…… ……209-211

**Figure 5.15:** PCA and OPLS-DA of DIRE NMR spectra between burn injury patients and healthy controls……213
**Figure 5.16:** PCA and OPLS-DA of DIRE NMR spectra, discriminating against vaccine response and non-response……214
**Table 5.4:** Calculated concentrations of DIRE integrals comparing burn injury and non-burn controls……215
**Figure 5.17:** Violin plots showing distribution of calculated DIRE integrals……215
**Figure 5.18:** PCA and OPLS-DA of quantified lipoprotein subfractions……216
**Figure 5.19: OPLS-DA Variable importance of Projection (VIP):** Arbitrary threshold of 1 applied, showing the most discriminant metabolites in the OPLS-DA model discriminating burn injury and healthy controls……217
**Figure 5.20:** PCA and OPLS-DA of quantified lipoprotein subfractions, discriminating against vaccine response and non-response……218
**Figure 5.21: OPLS-DA Variable importance of Projection (VIP):** Arbitrary threshold of 1 applied, showing the most discriminant metabolites in the OPLS-DA model discriminating vaccine response……219
**Figure 5.22A: Metabolite-metabolite spearman's correlations.** Quantified biogenic amine and tryptophan metabolites and DIRE integrals correlated against one another, showing the signature present in non-burn controls……227
**Figure 5.22B: Metabolite-metabolite spearman's correlations.** Quantified biogenic amine and tryptophan metabolites and DIRE integrals correlated against one another, showing the signature present in burn injury patients……228
**Figure 5.23:** Spearman's correlation between measured cytokines and DIRE quantified integrals……229
**Figure 5.24: Lipoprotein-DIRE integrals spearman's correlations**……230
**Figure 5.25A: Lipoprotein - cytokine spearman's correlations.** Quantified lipoproteins and cytokines correlated against one another, showing the signature present in non-burn controls……232
**Figure 5.25B: Lipoprotein - cytokine spearman's correlations.** Quantified lipoproteins and cytokines correlated against one another, showing the signature present in burn injury patients……233
**Figure 5.26: Hierarchical edge correlation between significant lipoproteins, small molecules and cytokines**……235

**Chapter Six: Discussion**

**Figure 6.1:** Exemplar systemic-metabolite associations from SARS-CoV-2 infection……248
**Figure 6.2**: Exemplar metabolite-microbiome biomarkers from PBC UDCA responders……249
**Figure 6.3:** Exemplar systemic-metabolite associations from acute paediatric burn injury……250

# CHAPTER ONE: INTRODUCTION

## 1.1 Aims and Objectives

The aim of this thesis is to investigate complex immune-mediated metabolic signatures in multisystemic diseases, using a variety of analytical metabolic profiling techniques and to develop a framework for integration of immune metabolic signatures with clinical datasets. Key objectives are:

- To identify disease specific signatures using nuclear magnetic spectroscopy (NMR) and ultra-high performance liquid chromatography mass spectrometry (UHPLC-MS).
- To identify specific immuno-metabolic signatures through the integration of cytokine data with spectroscopic data.
- To develop generic strategies for stratifying and integrating multimodal data to gain a systems overview of complex diseases.

## 1.2 Metabolism and Metabonomics

Metabolism can be widely defined as the sum of biochemical processes in living organisms that produce or consume energy [1] and pervades every aspect of biology. Metabolites are the end products of cellular regulatory processes and are considered to be the terminal response of biological systems to changes in genetic or environmental factors. Over 11, 000 reactions and 18, 000 metabolites are now annotated in KEGG (Kyoto Encyclopedia of Genes and Genomes) [2] as of May 2021, with annotation ongoing. Core metabolism can be reduced to pathways involving an abundance of nutrients such as carbohydrates, fatty acids

and amino acids, essential components for global homeostasis and macromolecular synthesis in humans [3].

The 3 primary classes of pathways involved in core metabolism are as follows:

- Anabolism: synthesis of simple molecules or polymerisation to form more complex macromolecules
- Catabolism: degradation of molecules to release energy
- Waste disposal: eliminate toxic waste produced by other classes [3].

The identification of regulatory mechanisms that link cell signalling to the orchestration of metabolic pathways or enable cells to sense energy changes, could be key in understanding the underlying mechanism to a multitude of diseases. Identifying disease-related metabolites, patterns and biomarkers is critical not only for improving clinical diagnoses but also for gaining a better understanding of metabolic pathological processes, which in turn can drive the development of personalised medicine, and the development of therapeutics with greater efficacy for specific populations [4]. Metabolites rarely function in isolation and operate in pathways that bridge the relationship between genotypes and phenotypes, influenced by the combined power of the genomes and phenome. The metabolic consequence of a single disease is not restricted to the one or two metabolites but generally the effects of the disease spread functionally throughout an organism, and thus comprehensive, global profiling is an invaluable tool.

The suffix of “omics”, when applied to a molecular term implies a comprehensive screening of a set of molecules. Over the last 3 decades, development of other omics beyond

genomics have taken place, with technologies capable of interrogating transcripts, proteins and metabolites. Metabonomics is defined as “the quantitative measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification” [1]. A particular strength of metabonomics is the ability to elucidate population differences due to genetics, disease and environmental changes, including nutrition [1]. Metabonomics typically utilises spectroscopic technologies such as nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS); MS is typically coupled to a molecular separation technique such as gas chromatography (GC-MS) [6], liquid chromatography (LC-MS) [7] and ultra-high performance liquid chromatography (UHPLC-MS) to generate molecular profiles for a variety of biofluids and tissues, including blood, urine, faeces, and whole tissue extracts [8]. It has become an increasingly popular tool in drug development and molecular medicine due to the capacity to profile phenotypes and address environmental changes on pathophysiology [5]. The complexity of the data as with other ‘omics’ necessitates computational modelling to accommodate the complexity of the informatic dense spectra. Thus, advancements in technology resolution and bioinformatics have further facilitated this growth. To probe changes in gut microbiome composition, metagenomics is used, with 16 ribosomal RNA (16S rRNA) gene sequencing being frequently employed in combination with spectrometry methods for a comprehensive screen [9].

This thesis uses a combination of NMR, UHPLC-MS, 16s rRNA sequencing and cytokine profiling techniques to interrogate the metabolic profiles of three diseases, each with unique hypermetabolic signatures: namely SARS-CoV-2 infection, Primary Biliary Cholangitis and acute burn injury in children.

### 1.3 Nuclear Magnetic Resonance (NMR) Spectroscopy

Nuclear Magnetic Resonance (NMR) remains a key analytical technology in metabolic profiling for several key attributes. NMR is automatable, non-destructive, nonbiased, can be used in a quantitative or qualitative mode, and can be used in a capacity that requires little or no chromatographic separation, or chemical derivatization. Additionally, NMR is capable of detecting compounds less amenable to LC-MS, such as sugars, organic acids, alcohols, polyols and highly polar compounds [10]. These features make NMR spectroscopy an ideal technology for high-throughput, large scale epidemiological studies.

NMR is a phenomenon in which nuclei exhibit a property of spin when subjected to a strong continuous magnetic field and behave like bar magnets on exposure to radiofrequency pulses, the nuclei can shift energy levels and produce a measurable signal that is characteristic of the magnetic field at the nucleus [11]. Solution-state NMR spectroscopy is a powerful and efficient tool for metabolic profiling, with the ability to generate information-dense spectra that have been proven to be effective in metabolism studies and medical research [1]. $^{1}$H NMR spectroscopy is generally utilised for biological samples, in which almost all molecules contain protons and thus will be the technique focused on in this thesis, to provide access to metabolite information. Properties that can aid structural elucidation of molecules include chemical shift, integration (quantification), peak multiplicity and relaxation [11]. Aside from inorganic ions, all small molecules that contain hydrogen atoms, can be measured due to the high sensitivity of abundant $^{1}$H nuclei.

### 1.3.1 $^{1}$H NMR Chemical Shift

Briefly, every chemically distinct hydrogen nucleus in each metabolite in a biological sample, will exhibit an NMR signal with a characteristic resonance frequency; measured as a chemical shift relative to a standard compound, measured in parts per million (ppm) [12]. This characteristic signal occurs as a result of each proton which, although exposed to the same external magnetic field, differs slightly in that the local chemical and electronic environment wherein surrounding electrons will shield the proton to a small extent.

When analysing urine in metabonomic studies, it is common practise to add a reference material such as 3-(trimethylsilyl)-2,2',3,3'-tetradeuteropropionic acid (usually abbreviated to TSP) or deuterated forms of 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) to define as 0 ppm such that spectra can be aligned and interpreted. The chemical shift of an NMR signal from a hydrogen nucleus in a. metabolite is independent of the applied field strength and is highly reproducible for that those specific nuclei. Biofluids such as a serum or plasma from blood, are typically referenced to an anomer of glucose at $\delta$ 5.23, as some components of serum or plasma may bind to TSP or DSS [13]. However, this signal is highly sensitive to temperature, and thus care must be taken to ensure that homogenous conditions. Solvent induced chemical shift variation can be useful analytically, to distinguish signals that may be otherwise overlapped [14].

Chemical shift variation is typically observed to a certain degree and is normal for values to be within $\delta$0.03 between samples [8]. Chemical shifts in in NMR spectroscopy are highly sensitive to structural and environmental change. Sensitivity to pH change can be used for

metabolite identification, particularly for those with ionisable functional groups, for example citrate [15].

Different pulse sequences serve the purpose of different resolutions to best exploit accessing different molecular information and experiments are generally run sequentially. In addition to standard 1D (one-dimensional) NMR, 2D (two-dimensional) NMR spectroscopy can be used to assess spin patterns to aid metabolite identification and structural assignments. 2D experiments commonly acquired alongside standard 1D NMR acquisition include J-resolved spectroscopy (JRES), correlation spectroscopy (COSY), and total correlation spectroscopy (TOCSY). To summarise:

- 1D Standard NMR spectra which use the first increment of the Nuclear Overhauser Effect (NOE) to achieve saturation of the water resonance, contain signals of both high molecular weight molecules such as lipids and metabolites and low molecular weight [16].
- Carr-Purcell-Meiboom-Gill (CPMG) spectra attenuate signals from small molecules by via $T_2$ filtering (transverse relaxation). $T_2$ filters exploit short relaxation times to reduce contribution from macromolecules with the broad signals such as those from larger molecules like lipids [17].
- 2D *J*-resolved (JRES)spectra separate signals into two different spectral dimensions wherein the second dimension contains information on the J coupling or peak multiplicity and structure. Proton-decoupled 1D spectrum can observed from the skyline projection from the chemical shift dimension, enabling metabolite identifications and quantifications. From the *J*-coupling dimension, *J*-coupling information provides information that aids metabolite identification as *J*-couplings

are non-reactive to physiological factors such as temperature or pH compared to chemical shifts. JRES spectra provide insight into spin-spin coupling patterns, providing information on which protons in a molecule may be closely linked together in terms of chemical structure and bonds [18].

- $^{1}$H-$^{1}$H Correlation Spectroscopy (COSY) shows the correlation between coupled hydrogens in the $^{1}$H spectrum. The 1D $^{1}$H spectrum can be plotted along both 2D axes. COSY is used to confirm metabolites and to identify protons that are adjacent to each other in a molecule [19].
- Total Correlation Spectroscopy (TOCSY) generates correlations between all protons within a spin system, yielding information on bond correlations through spin-spin coupling. TOCSY provides complementary information to the COSY since it can detect protons that are several bonds distant (typically 5 or 6) from each other [20].

Development and establishment of NMR techniques such as the Bruker In-Vitro Diagnostics research (IVDr) methods have enabled accurate quantification of lipoproteins and small molecules in blood [42]. The IVDr system integrates advanced hardware, software, automation and standard operating procedures (SOPs) to acquire this screen. These methods have been reported to be highly reproducible, with appropriate accuracy for clinical and regulatory settings [42].

### 1.3.1 Spin-spin coupling in $^{1}$H NMR spectra

Spin-spin coupling (or J-coupling) is a phenomenon that is a result of magnetic interaction between nuclei, resulting in NMR resonances to be split into components with characteristic

relative intensities and scales [21]. Spin-spin coupling aids spectral interpretation and metabolite identification. Frequency separation results in the separation of peaks into a second dimension that would otherwise overlap, and also facilitates identification through the ratio of frequency separation to the coupling constant [11].

### 1.3.2 Advantages and Limitations of NMR

The advantages and limitations of NMR are usually attributed to one fundamental factor: its non-selective and broad analytical spectrum. Table 1.1 briefly summaries the advantages and limitations of the technology. Whereas high reproducibility is the main advantage of NMR spectroscopy, the sensitivity is intrinsically low compared to other techniques such as mass spectrometry (MS). Sensitivity however can be improved with a larger number of scans, resulting in longer data acquisition times, and higher magnetic field strength. NMR can be enhanced by hyper polarisation, a method whereby signals can be enhanced several thousand times over the normal concentration [22]. Due to the non-selective nature of NMR, there is often a large overlap between multiple metabolites, which poses a significant metabolite identification challenge. However, this also enables relatively fast simultaneous measurement of all measurable metabolites using NMR spectroscopy, where all metabolites at a detectable concentration level can be attained in a single observation [23].

NMR is a non-destructive technique, and thus several analyses can be carried out on the same sample. This is ideal for epidemiological studies, for which large sample volumes are difficult to acquire due to cost, time and logistic limitations such as availability of freezer storage. It is also ideal for paediatric samples, due to difficulty to attain large volumes of

plasma and urine given their weight and size [24]. Studies investigating the impact of freeze thaw cycles on NMR analysis suggest that there is minimal change in metabolite concentrations and lipoprotein measurements in urine, and minor changes to serum metabolites and lipoprotein composition over 5 freeze-thaw cycles [25]. This is useful in the analysis of longitudinal studies which may have been collected over a series of decades and thus samples may have been subjected to different handling and temperature changes.

Standard 1D NMR data acquisition is relatively quick, with typically an experimental time of 12 minutes for spectral acquisition from a single sample analysed with 32 scans on a 600MHz spectrometer, generating 3 spectra, standard 1D spectra, a spin echo spectrum attenuating low molecular weight components and a diffusion edited spectrum attenuating macromolecular components.

Enhancement of technology such as the development of cryoprobes has vastly improved signal to noise detection, the resolution improved up to four-fold [20]. Cryogenic probes function by cooling of the detector coil and amplifier to 20K, reducing thermal noise, and enabling a clearer resolution of nuclei with naturally low abundance. At thermal equilibrium, two states required for energy transition and subsequently signal detection are extremely close together and thus requires very little energy for the nuclei to be promoted from one state to another just by normal thermal conditions [26].

**Table 1.1: Summary of NMR advantages and limitations**

| Advantages | Limitations |
|---|---|
| **Quick and easy sample preparation** | **Non-selective and so requires convolution to overcome peak overlap** |
| **Robust against freeze thaw cycles for metabolite composition and lipoprotein quantification** | **Metabolites in serum have minor alterations in freeze thaw cycles** |
| **Comprehensive and untargeted in one measurement** | **Untargeted which results in some metabolites detected ambiguously** |
| **Bruker IVDr Lipoprotein Subclass Analysis (B.I. LISA) methods enable comprehensive quantification of 112 lipoproteins, in addition to 20 small molecules** | **Variance in urinary excretion volumes and subsequent variations in metabolite concentrations remains a significant unresolved issue** |
| **Inherently quantitative: signal intensity is directly proportional to the number of nuclei** | **Complete assignments may be difficult due to the complexity of signals and overlap.** |
| **Minimal sample preparation** | **Sensitive to changes in pH, osmolality or complexation with metal ions resulting in variability in peak chemical shift** |

## 1.4 Targeted Liquid Chromatography Mass Spectrometry Assays

Both NMR and mass spectrometry (MS) are complementary to one another. In order to provide the broadest and most comprehensive coverage for metabolic profiling, employing both techniques is the optimal strategy but may be constrained by financial or sample volume limitations. The application of MS is broadly expanding due to the speed, sensitivity and wide dynamic range of this technology which often utilises a separation method for enhanced dispersion and detection. Liquid chromatography MS (LC-MS) has become the dominating mass spectrometry technique in metabonomics, with a broad coverage of detection for metabolites. For many assays no sample derivatisation is required as samples are not required to be volatilised due to the mobile phase changing from gas to liquid.

Ultra-high-performance liquid chromatography MS (UHPLC-MS) employs the use of sub-2 µm or 2.6–2.8 µm (fused-core) stationary phase particles at the chromatography column, with the ability to withstand high solvent flow rates and high pressures ranging between 6000–19 000 psi, reducing broadness of peaks typically encountered using standard particle sizes, thereby yielding shorter analytical run times and increasing peak capacity. The type of mass analyser coupled to the separation technique can further increase sensitivity.

Quadrupole time-of-flight (QToF) mass spectrometers employ a hybrid technique that combines time-of-flight mass analysers with a quadrupole technology. The third quadrupole is replaced with a time-of-flight tube where mass separation occurs with lighter mass ions exhibiting a shorter time-of-flight and heavier mass ions exhibiting a longer time-of-flight. This enables the comprehensive identification and quantification of compounds with exact mass. Triple quadrupole (TQ) mass analysers facilitate the accurate measurement of concentrations of target compounds that are present at very low levels in complex samples. The improved ionization, and reduce peak overlap, leads to easier structural determination and confirmation of candidate biomarkers [27].

Targeted metabonomics has broad applications for the screening of specific pathways and their biological interpretation and provides a concise analytical window for disease screening. This provides a useful tool for rapid screening of metabolites with either relative or absolute metabolite quantification. In comparison to NMR assignments, targeted UHPLC-MS assays utilise standards for calibration and the identification of compounds in biofluids [28]. Whilst some targeted metabolic profiling kits can quantify up to 800 metabolites [29],

it is difficult to obtain all required chemical standards for metabolites of interest and thus the coverage of targeted metabolic profiling is often limited [30].

### 1.5 Computational Methods in Metabolic Profiling

The primary objective of multivariate analysis in metabonomics is to employ pattern recognition algorithms, for reduction and deconvolution of the multidimensional metabolic data. Unsupervised and supervised statistical methods are employed for data reduction and biological interpretation of complex multidimensional data. This in turn enables further discrimination between groups, highlighting significant variables and features that may attribute to a group or an individual.

#### 1.5.1. Principal Components Analysis

Principal components analysis (PCA) forms the initial step in multivariable analysis, assessing the variable space and, in doing so, reducing the dimensionality, in order to facilitate the detection of patterns of parameters or variables (in this instance spectral signals or data points) [31]. PCA predicts an approximation of a given data matrix X, as a product of two smaller matrices T and P, which capture the most prominent features of X. For a spectral dataframe, this is perceived as matrix X, with N samples and K variables (spectral points), translating as N points in a K dimensional space. The dimensional space (K) can be reduced to a subspace defined by a small number of principal components (PC). For each PC, two sets of vectors are calculated, score (T) and loadings (P). Scores capture inter-sample variation and thus are useful for assessing outliers and trends in the data. Loadings capture

inter-variable variation, and thus are useful for assessing the influence of variables on the distribution of the scores and the relationship between variables.

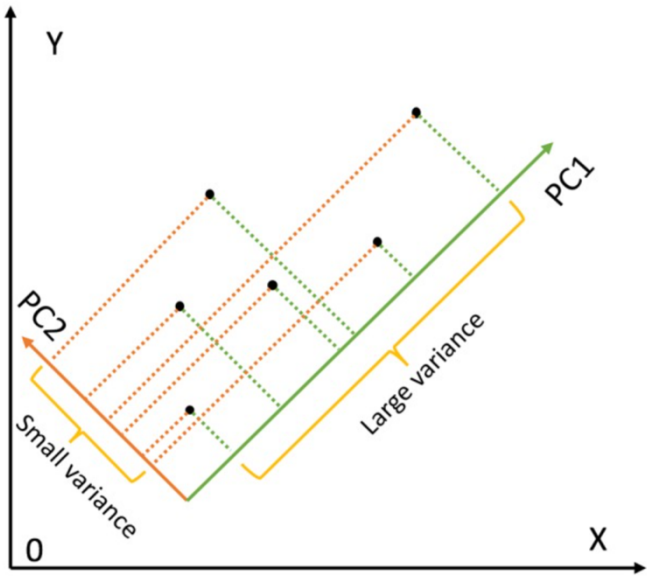



**Figure 1.1: Schematic of how variance is calculated.** Principal components analysis (PCA) specifies a new coordinate system in which every point is assigned an (x,y) value. Figure credited to: Zhang Z, Castelló A. Principal components analysis in clinical studies. Annals of translational medicine. 2017 Sep;5(17).

The number of PCs fitted in the model to attributed to the percentage of variance or goodness of fit ($R^2X$) of the model and the predicted variation or goodness of prediction ($Q^2X$) using a leave one out seven-fold cross-validation procedure. This method of validation is useful in assessing the robustness of the model and to useful for assess whether or not the model has been overfitted, which can be a risk when the dataset contains few samples and many variables.

### 1.5.2. Partial Least Squares and Partial Least Squares-Discriminant Analysis

Partial Least Squares (PLS) is a regression method which combines and generalises PCA and multiple linear regression, and aims to describe relationships between a descriptor matrix

(X) and corresponding response matrix (Y) [32]. PLS then deconvolutes the predictor matrix by sequentially extracting orthogonal components and subsequently summarises the model through computing scores and loadings for both spaces. PLS maximises the covariance between the X scores and Y scores for each component, by calculating the contribution of each X-variable to each Y component. Partial least squares-discriminant analysis (PLS-DA) extends from PLS, and primarily focuses on discriminating classes based on the X variables. PLS-DA model is fitted between a descriptor matrix (X) and the predicted response (Y), defining a class or dummy variable, with the discrimination plane seeking separation of variables based on the class.

### 1.5.3 Orthogonal Partial Least Squares-Discriminant Analysis (OPLS-DA)

Orthogonal Partial Least Squares (OPLS) combines orthogonal signal correction with PLS which removes systematic variation that is correlated to the biological class of interest from the data. OPLS maximises the covariance between X variables from Y response matrix, removing the X variation that has zero correlation to Y, or to simply state removes X variation that is orthogonal to Y. This in turn reduces the number of variables in the OPLS model, subsequently reducing the complexity of the model and enhancing biological interpretation. Orthogonal Partial Least Squares-Discriminant Analysis (OPLS-DA) includes a dummy-Y variable (a variable or class discriminator), and predicts two modelled variations, the Y predictive and the Y-orthogonal component. The regression coefficients can be 'back-scaled' to retrieve the correlation and covariance in the loadings, which can display directly the discrimination between the groups and subsequently identify "biomarkers". The regression coefficients are relative to the covariance between X and Y if the data are mean-

centred. However, if the data are scaled to unit-variance, regression coefficients are relative to the correlation. The variable importance in projection (VIP) index is a normalised average of the variables with the highest influence on the model, measured by the contribution of each predictor to the construction of the components, and thus is useful when interrogating the results from OPLS-DA [33].

### 1.5.4 Statistical Total Correlation Spectroscopy (STOCSY)

Statistical Total Correlation Spectroscopy (STOCSY) computes a correlation matrix (C) between intensities of all datapoints in a spectral dataset and can be obtained according to the following formula:

$$C = 1/(n-1)X_1{}^t X_2$$

whereby the autoscaled experimental matrices of $n \times v_1$ and $n \times v_2$, respectively; n is the number of spectra (one for each sample) and $v_1$ and $v_2$ are the number of variables in the spectra for each matrix. C is therefore a matrix of $v_1 \times v_2$, wherein each value is a correlation coefficient between two variables of the matrices $X_1$ and $X_2$[20]. As the signal intensities from a molecule will always be measured with the same ratio, given the same experimental conditions, then it is thought that theoretically the relative intensities will have absolute correlation (correlation coefficient r=1). Given the complexity of biofluids, *r* will always be inferior to 1, due to spectral noise and observed peak overlap. The signal intensity is used to represent covariance and is presented with spectral graphical representation for ease of interpretation and assignment of metabolites, but additionally aids structural elucidation

[20]. OPLS-DA is often used in conjunction with STOCSY, for metabolite identification, facilitating direct identification of discriminant metabolites.

### 1.6 Multimodal data in complex systems diseases

The development of precision medicine strategies via systems medicine techniques is a multi-disciplinary endeavour involves quantification, integration, analyses and interpretation of a combination of clinical and non-clinical data. This endeavour is essential to gain a deeper understanding of mechanisms and physiological changes that drive disease development. Substantial developments in computational deconvolution and integration of large, high throughput datasets utilise pattern recognition strategies to integrate different windows of resolution in the data in the search for patterns related to biological processes or for disease stratification [34].

Studies have shown the prerequisite ability for integrative data strategies and the application to high-throughput data to show invaluable insight into disease trajectories. Using unsupervised techniques, one study by Argelaguet et al., presented a framework for agnostic integration of heterogenous data, and presented the benefits of such work by identifying several major drivers of disease variation [35].

The window of analysis provided by different datasets such as cytokine and microbiome data, can be essential in the understanding of different disease groups. For disease group stratification, employing the use of methods such as correlation analysis can help identify biomarker signatures [36].

### 1.6.1 Correlation analysis

Correlation analysis is a group of techniques to investigate potential two-way (normally linear) associations between two continuous variables, resulting in positive, neutral or negative correlation [37]. Correlation is calculated by measuring the correlation coefficient, which represents the strength of the potential linear association between the variables of interest. A coefficient of 0 indicates no correlation, whereas a coefficient of -1 or +1 indicates a near perfect negative or positive linear correlation respectively. Both scenarios can inform on the order within a molecular system If the coefficient is a positive number, this implies that both variables vary in phase (synchronically) or increase simultaneously, whereas a negative coefficient, implies that the variables are inversely related (vary in anti-phase), with one value increasing and the other decreasing [38].

Two types of correlation co-efficient are frequently employed: Pearson's product moment correlation coefficient and Spearman's rank correlation coefficient, with other techniques employed for multiple variable correlation.

Pearson's product moment correlation coefficient is represented as ϱ for a population parameter and as r for a sample of that population. Pearson's correlation is employed when Gaussian (or normal) distribution is present for both variables of interest. This coefficient is affected by the presence of extreme values in an otherwise normal distribution, which may result in the exaggeration or inhibition of the true correlation strength, and thus is not an adequate method when either or both variables are not normally distributed. For

correlation between variables x and y, the formula for calculating the sample Pearson's correlation coefficient is denoted as:

$$r = \frac{\sum_{i=1}^{n}(x_i - x)(y_i - y)}{\sqrt{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]\left[\sum_{i=1}^{n}(y_i - \bar{y})^2\right]}}$$

where $x_i$ and $y_i$ are the values of x and y for the "ith" individual [28].

Spearman's rank correlation coefficient is denoted as $\varrho_s$ as a population parameter and as $r_s$ for sample statistics. Spearman's correlation is employed in the absence of a Gaussian distribution, and where a population or variable skew exists [37]. For correlation between variables x and y, the formula for calculating the sample Spearman's correlation coefficient is denoted by:

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2 - 1)}$$

whereby $d_i$ stands for the difference in ranks of the "ith" element and n is equal to the number of data points of the two variables.

### 1.6.2 Hierarchical Clustering

Correlation analysis is often paired with hierarchical clustering analysis (HCA), a method used to investigate related structures within a given dataset [38]. This allows grouping of features and is a useful tool when biologically interpreting large datasets. Two clusters that

are close together are considered to be agglomerated by a single metric, thus signifying a close relationship. Methods such as hierarchical edge clustering (HEC) extrapolate from the original hierarchical clustering method, and present pair-wise associations or dissociations between two variables. Edge bundles are effective in presenting linkages in a network and projects onto the map of controls [39]. This method is beneficial in the integration of multi-modal datasets given the deviation between variables.

### 1.6.3 Cytokine multiplex assays

Cytokines are small, secreted proteins, encompassing interleukins, chemokines, interferons, and tumour necrosis factors. Cytokines behave as signalling molecules to aid cellular communication, and to stimulate the relocation of cells to sites of trauma, infection and inflammation. They are released in a multitude of pathways, including paracrine, autocrine and endocrine, and implicated in a number of disease processes and immune system disorders [320]. The evaluation of cytokines in disease populations, can be crucial in the understanding of downstream causal effects, mediated by immune changes, and can be integrated with other profiling data to establish what changes occur in a pro-inflammatory and anti-inflammatory disease state.

Technological advancements moving beyond traditional techniques for cytokine profiling as a measure of immune response, now enable comprehensive screening of multiple cytokines at one time. Flow cytometric multiplex arrays, also known as bead-based multiplex assays, employ the use of bead sets that are unique under flow cytometry. Each bead set is coated with a specific antibody, and corresponding fluorescence detection antibodies bind to the

specific capture antibody complex on the bead set, thus enabling sequential and comprehensive profiling in one experimental process [40]. However, despite the effectiveness of the assay, some discrepancies have been noted in the analysis of serum and plasma due to sample preparation dissimilarities and the level of accuracy required for preparation for reproducibility. Despite this, the assay is emerging as a gold standard for inflammatory cytokine profiling.

### 1.6.4 16S Ribosomal Ribonucleic Acid (16s rRNA) Sequencing

The influence of the gut microbiome on health and disease is emerging as a key factor in a host of diseases, ranging from obesity, inflammatory bowel disease and hepatocellular carcinoma [41, 42]. It is well acknowledged that the bacterial composition of the human microbiome is perturbed throughout an individual's lifespan, in response to different environmental factors and conditions. Whilst some bacteria behave synergistically within humans, aiding with functions such as the digestion of complex molecules, some bacteria when overwhelmingly present, can be disease causing [321]. Deconvoluting the microbiome composition is not only of clinical importance but will further knowledge on how bacteria interact with the human host.

16s ribosome ribonucleic acid (16S rRNA) sequencing refers to the 16S RNA component of the 30S subunit of a prokaryote ribosome. Through exploiting this technology, this enables evaluation of the 16S gene at a taxonomic resolution at species and strain level. PCR-amplified 16S sequences are classically clustered based on similarity to generate a taxonomical unit. In this thesis, a newer method of classification is adopted, called Amplicon

Sequence Variants (ASVs) [43]. ASV workflows partition the reads on the basis of error capturing models, to counter and correct sequencing errors, in addition to considering abundance and sequence similarity [44]. Through this approach ASV annotation is able to detect minute biological sequence variations and counteracts the effect of any technical errors. However, it should be considered that some genuine biological variations can be classified as technical errors. Due to the sensitivity of the profiling method, it is worthwhile to interrogate the data with additional microbiome information containing data, such as metabolites detected from urine and blood.

# CHAPTER TWO: METHODS

For the purpose of this thesis, three different metabonomics studies were conducted, each with a series of different data integrations. Each study provides a unique metabolic profiling screening, enhanced with integration to composites of other biological processes (i.e cytokine profiling or microbiome profiling). Briefly discussed are the studies used in this thesis. Methods will be described in more detail for each study in the corresponding chapter. Table 2.1 shows a comparison of all datasets used from each study.

**Table 2.1: Side-by-side comparison of all datasets available from each study**

| Western Australia COVID-19 Study | UK- Primary Biliary Cholangitis Nested Cohort Study | Paediatric Burns Study |
|---|---|---|
| **Untargeted $^{1}$H NMR** (1D, 2D, DIRE) **Quantified 112 lipoproteins Quantified Small Molecules** | **Untargeted $^{1}$H NMR** (1D, 2D) | **Untargeted $^{1}$H NMR** (1D, 2D, DIRE) **Quantified 112 lipoproteins Quantified Small Molecules** |
| **Targeted UHPLC-MS Amino acid assay** | | **Targeted UHPLC-MS Amino acid assay** |
| **Targeted UHPLC-MS Tryptophan pathway assay** | | **Targeted UHPLC-MS Tryptophan pathway assay** |
| | **Targeted UHPLC-MS Bile acid assay** | |
| | **16s rRNA sequencing** | |
| **Clinical Biochemistry and Pathology** | **Clinical Biochemistry and Pathology** | |
| **Cytokine profiling assay** | | **Cytokine profiling assay** |
| | | **Mass Cytometry** |
| | | **Vaccine titre testing** |

### 2.1.1 Western Australian COVID-19 Study

The Western Australia (WA) COVID-19 metabonomics study aims to utilise metabonomics for the comprehensive profiling of pathological signatures of COVID-19. Further information on the pathogenesis and aetiology is provided in chapter three, section one. A brief schematic of the analyses conducted is shown in figure 2.1.

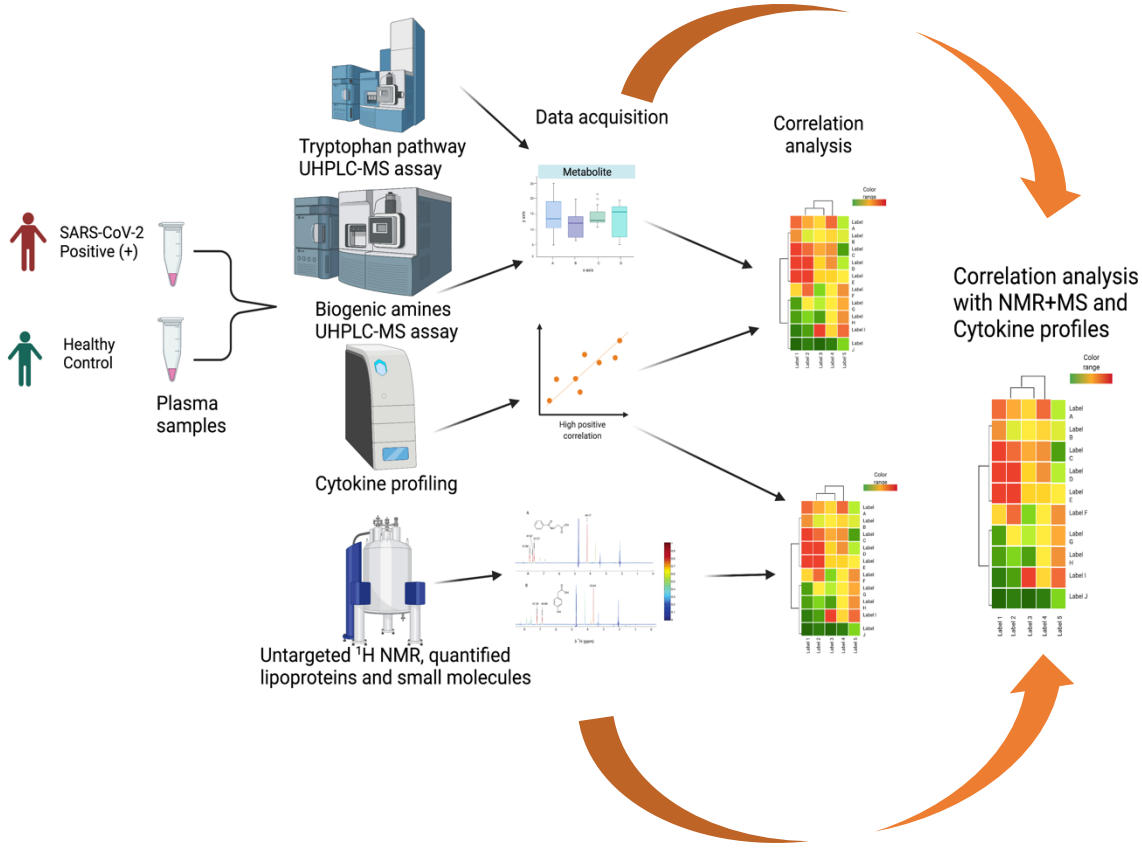



**Figure 2.1: Outline of analytical techniques used for the Western Australia COVID-19 study**

Analysis was carried out in collaboration with the Australian National Phenome Centre. Briefly, plasma samples were collected from 17 adults who presented COVID-19 disease like symptoms and subsequently tested positive for SARS-CoV-2 infection from upper and/lower respiratory tract swabs by RT-PCR and 25 healthy controls recruited from the population who had not exhibited any COVID-19 disease symptoms and were serologically tested negative

with respect to IgG antibodies as a part of this research study. All samples used in the study were collected from single time-point collections, with the exception for one patient who tested positive for SARS-CoV-2 and provided three samples during their hospital stay. Serological testing for SARS-CoV-2 antibodies were performed by the PathWest testing laboratories, Western Australia using a commercial point-of-care SARS-CoV-2 IgA/IgG specific test. The study was coordinated at Fiona Stanley Hospital within the Western Australian South Metropolitan Health Service catchment as part of the International Severe Acute Respiratory and Emerging Infection Consortium (ISARIC)/ World Health Organisation (WHO) pandemic trail framework (SMHS Research Governance Office PRN:3976 and Murdoch University Ethics no. 2020/052). Healthy controls were enrolled as volunteers, provided study details and written consent was obtained prior to data collection in accordance with the ethical governance (Murdoch University Ethics no. 2020/053).

### 2.1.2: UK-Primary Biliary Cholangitis Nested Cohort Study

The UK-Primary Biliary Cholangitis (UK-PBC) Nested Cohort study aims to use metabolic profiling via untargeted NMR and targeted UHPLC-MS bile acid profiling. Analysis was carried out in collaboration with Dr. Alexandros Pechliavanis, Imperial College London. Further information on the pathogenesis and aetiology is provided in chapter four, section one. Figure 2.2 is a brief schematic of the analyses covered in this thesis for the UK-PBC nested cohort study.

The UK-PBC Nested Cohort was recruited from 18 different research centres located across the United Kingdom, with a final sampling number of >600 patients. Recruited PBC patients were assigned to cohorts on the basis of specific criteria (Table 2.2), with the primary definer being the use of UDCA for treatment (Cohort A), and PBC patients who were treatment-naïve, i.e., those who were not on UDCA (Cohort B). The primary objective of this study was to investigate changes in metabolic profiles based on UDCA treatment and response, and to investigate changes between PBC and a normal state of health. For this thesis only urine samples were analysed. Age and sex-matched healthy controls recruited via the NIHR Bioresource, with samples collected from participants at a research centre using standardised methods. Sample collection and processing was concentrated in each of these centres, using standardised protocols and methods to ensure consistent sample collection and handling.

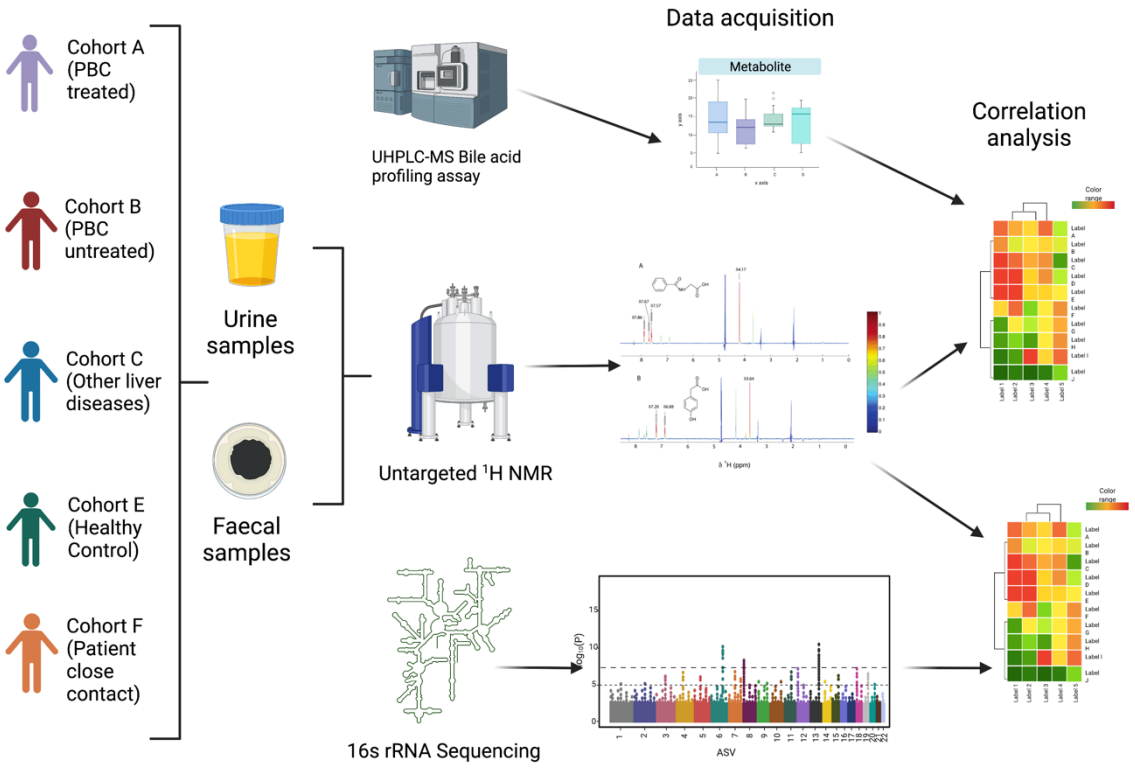



**Figure 2.2: Outline of analytical techniques used for the UK-PBC Nested Cohort study.**

Informed consent was obtained from all participating patients and patients deemed unsuitable for the criteria of the study were excluded and informed (REC reference 14/NW/1146). Clinical research nurses collected comprehensive medical records and validated questionnaires characterising symptoms were collected, in addition to socio-geographical and alcohol consumption data.

**Table 2.2:** UK-PBC study cohort outline and criteria

| Cohort | Criteria |
|---|---|
| **A** | **Patients with diagnosed PBC, incipient PBC or PBC/Autoimmune Hepatitis who ARE receiving treatment with UDCA** |
| **B** | **Patients with diagnosed PBC, incipient PBC or PBC/Autoimmune Hepatitis who ARE NOT receiving treatment with UDCA (treatment naïve)** |
| **C** | **Patients with liver diseases other than PBC** |
| **E** | **Healthy age and sex matched controls from NIHR Bioresource** |
| **F** | **Healthy best friend &/or same sex sibling controls** |

### 2.1.3 Paediatric Burns Study

The Paediatric Burns Study aims to utilise metabolic profiling to investigate immune driven changes in children with prior burn injury, with prior investigations determining a diminished immune response to childhood vaccination. Further information on the pathogenesis and aetiology is provided in chapter five, section one. Figure 2.3 is a brief schematic of all analysis, showing the pipeline and methods used.

Briefly, Paediatric burn patients were recruited for the WA Paediatric Burns Biobank pilot study from 2 separate hospital sites; 3 years post-burn injury, aged 0-4 at the time of original injury presentation. Age and sex matched healthy controls were recruited from a pool of participants. All samples were collected and obtained following informed consent of a parent or guardian and collection was conducted with ethical approval from the Child and Adolescent Health Service WA (2015219EP;1111EP;768EP). All patients had no prior pre-existing illness and were not receiving treatment or medical interventions of any kind.

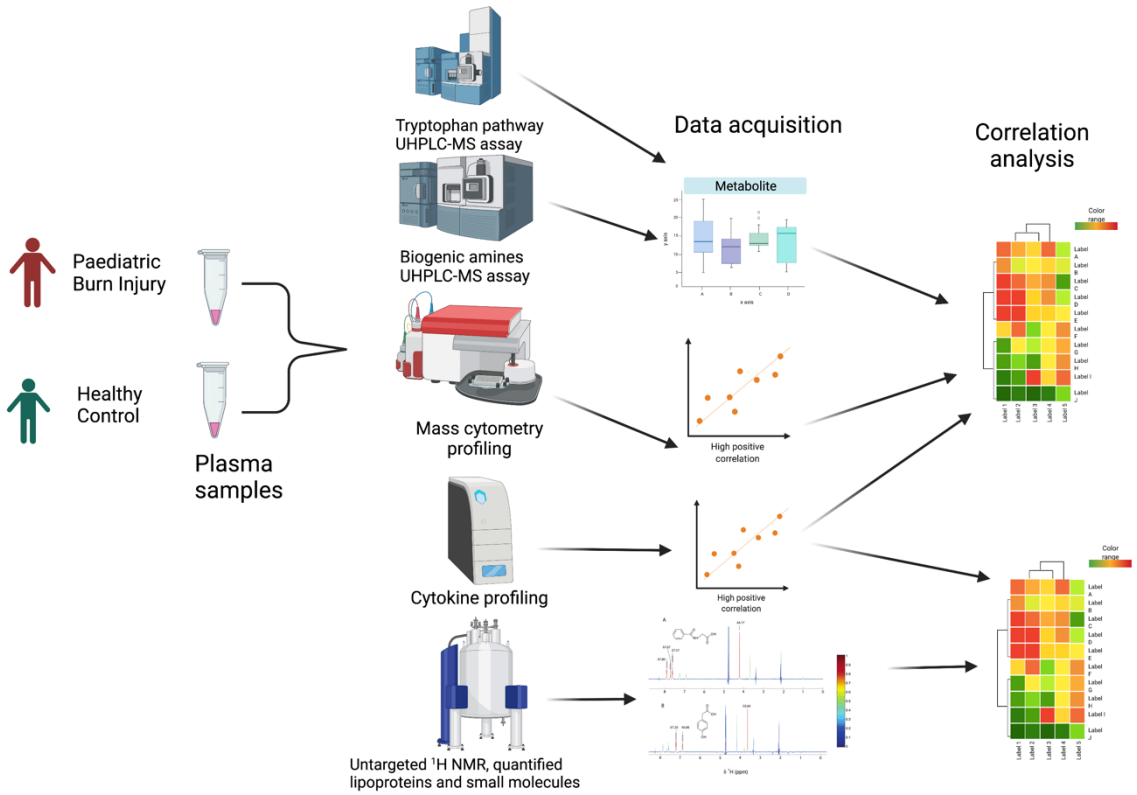



**Figure 2.3: Outline of analytical techniques used for the Paediatric Burns study.**

Methodological details unique to each study will be discussed in further detail in the corresponding chapter.

# CHAPTER THREE: INTEGRATIVE MODELLING OF CLINICAL AND METABOLIC PHENOMIC DATA - A SYSTEMS APPROACH TO CHARACTERISING SARS-CoV-2 INFECTION

The aim of this work is to characterise and identify metabolic signatures in SARS-CoV-2 positive patients compared to healthy controls, in order to understand the underlying molecular mechanisms driving the disease. This work also aims to identify potential pathological signatures of SARS-CoV-2 associated with inflammation.

# 3.1 Introduction:

### 3.1.1 Disease pathogenesis

In the last two decades, novel coronavirus infections have emerged as a severe global threat to public health. SARS (Severe Acute Respiratory Syndrome) novel coronavirus (SARS-CoV) was first detected and identified in March 2003, spreading initially from the Guangdong Province, China, reaching to 29 countries with 8422 cases and 916 deaths [54]. MERS (Middle East Respiratory Syndrome) novel coronavirus (MERS-CoV) first emerged in July 2012, initially detected in Saudi Arabia, reaching 27 countries with 2494 cases and 858 deaths since [55]. The current coronavirus disease 2019 (COVID-19) pandemic resulting from the novel SARS coronavirus 2 (SARS-CoV-2) was first detected in December 2019 in Wuhan, China, with the COVID-19 outbreak declared a pandemic by the World Health Organisation (WHO) in March 2020. As of May 2021, more than 200 countries have reported cases, with nearly 157 million cases and over 3 million deaths reported globally [56]. The quickly evolving pandemic now poses a severe challenge to healthcare systems worldwide, with medical infrastructures straining under the increased health crisis [57].

Coronaviruses are highly complex viruses that can cause respiratory, digestive and nervous system diseases in humans and many other animal species. The virus consists of an enveloped, positive sense single-stranded RNA genome, capable of infecting humans and other animal species [57]. Coronaviruses possess a distinct morphology as spherical virions with a central capsule; the name being derived from the unique crown-like surface projections [58]. Coronaviruses have been defined to four separate genera; alpha, beta, gamma and delta, all of which have zoonotic origins, with the viral genome of SARS-CoV-2

mapping with 96% similarity to bat-related coronaviruses [59]. The enveloped viral genome of SARS-CoV-2 contains four structural proteins; of which the receptor binding protein (RBD) in the spike protein binds to the Angiotensin Converting Enzyme 2 (ACE2) receptor on the surface of the host cell, which is initially expressed in the airway epithelial cells and vascular endothelial cells [60]. This subsequently mediates a fusion between the virion and the cell membranes in the host cell as shown in figure 3.1[60]. The active replication and release of the virus causes the host cell to undergo pyroptosis (highly inflammatory form of lytic apoptosis) and release damage-associated molecular patterns such as adenosine triphosphate (ATP), nucleic acids and ASC oligomers [61]. SARS- CoV-2 virus belongs to the beta-coronavirus genus; the subtype considered the most threatening to human health, with the ability to cause significant morbidity and mortality [62].

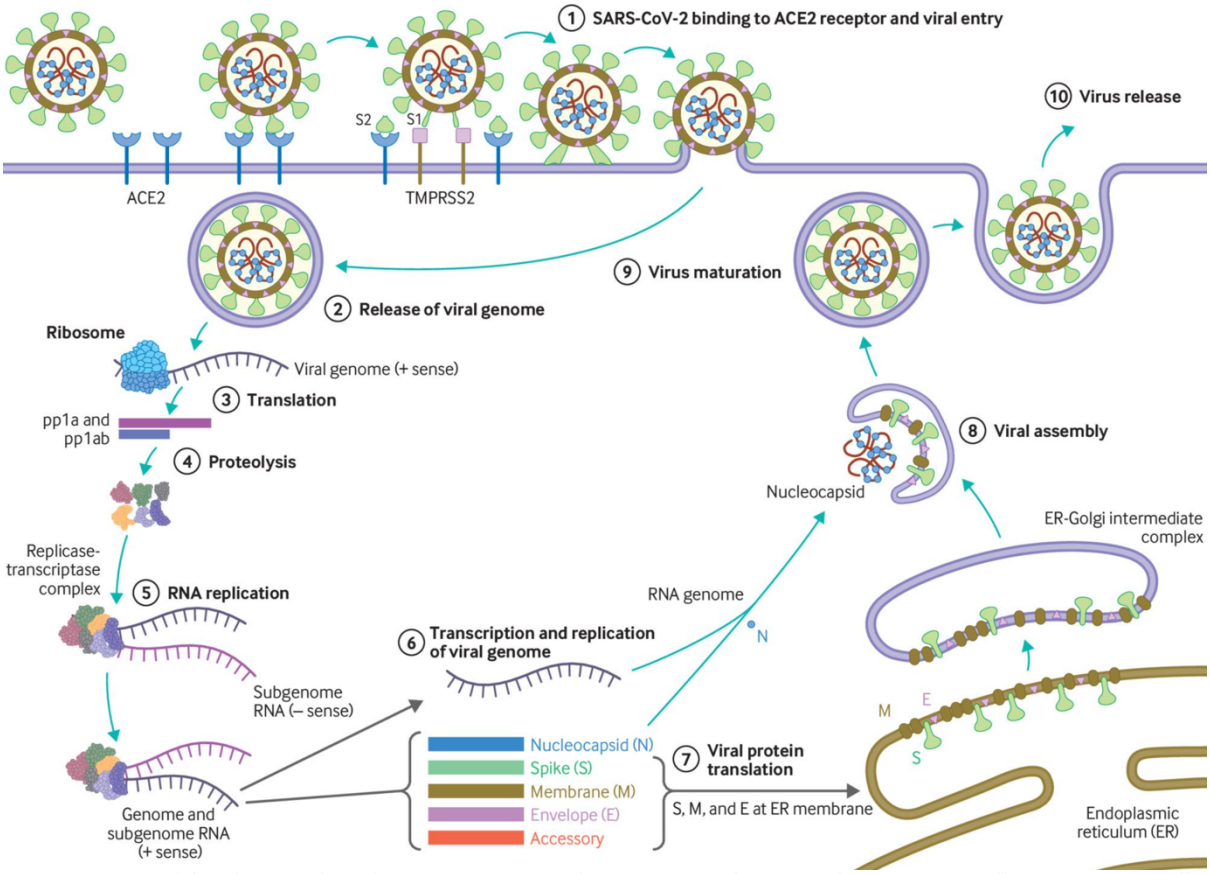



**Figure 3.1**: **Viral binding and replication process. The virus attaches to a host target cell receptor initially expressed in airway epithelial and vascular endothelial cells.** Reproduced from *Cevik M, Kuppalli K, Kindrachuk J, Peiris M. Virology, transmission, and pathogenesis of SARS-CoV-2. BMJ. 2020 Oct 23;371.*

### 3.1.2 Disease transmission

SARS-CoV-2 has a wide range of transmission routes, such as short-distance droplet transmission and contact transmission, in addition to being transmitted through aerosols in enclosed spaces and urine, and mother to child transmission may also exist through the placental barrier or through breastmilk [63, 64]. In addition, experimental studies have indicated the ability of SARS-CoV-2 to survive in stool, with transgenic mice expressing the human ACE2 receptor, found to be infected and exhibit pathological changes, suggesting that the faecal-oral transmission may be a viable transmission mode [65]. Evidence of person-to-person transmission of SARS-CoV-2 has resulted in extensive protective measures such as utilising personal protective equipment (PPE) and national lockdowns being implemented globally in order to control the COVID-19 outbreak [66].

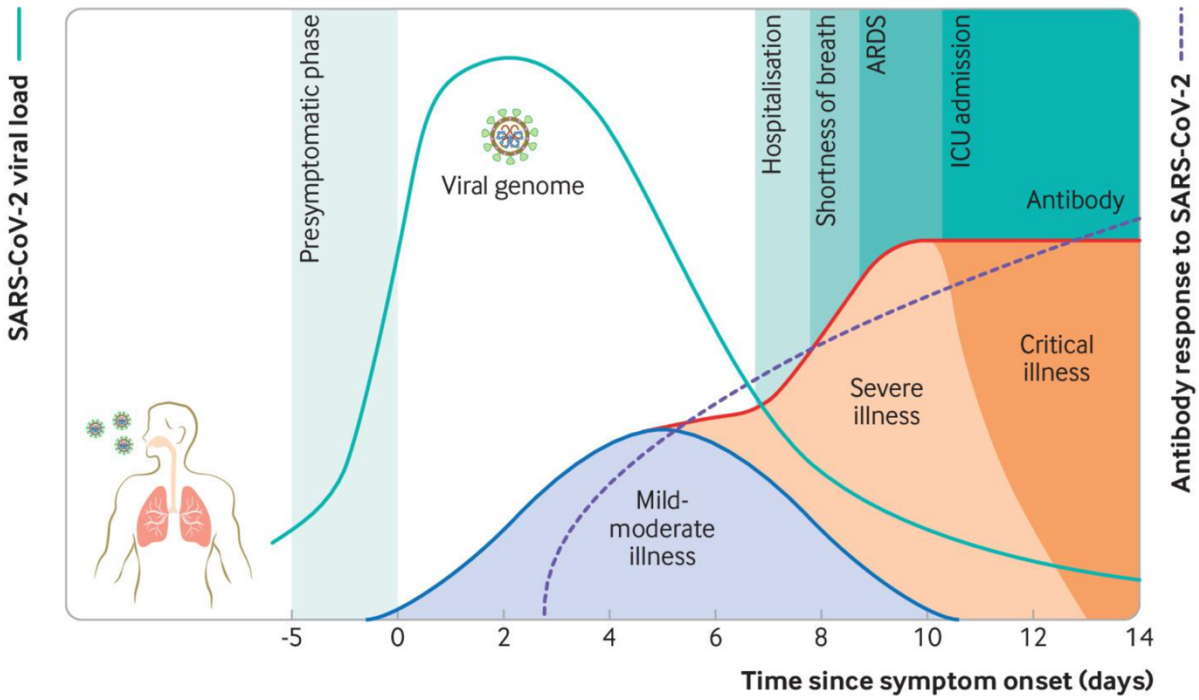



**Figure 3.2: The antibody response to SARS-COV-2 and disease severity.** Reproduced from *Cevik M, Kuppalli K, Kindrachuk J, Peiris M. Virology, transmission, and pathogenesis of SARS-CoV-2. BMJ. 2020 Oct 23;371.*

Whilst SARS-CoV-2 is more transmissible than SARS-CoV or MERS-CoV with an average R0 rate (rate of infection reproduction) of 1.8 – 3.6, the disease is less fatal, with a lower mortality rate for SARS-CoV-2 than for SARS-CoV[67]. Studies investigating the mortality rate of SARS, found that between 23-32% of all patients would develop severe disease and death was a likely outcome for over 25% [68]. A more mild-moderate disease response is seen with a reduction in antibody titres, however with an enduring elevated antibody response, disease severity increases (figure 2) [69]. Epidemiological studies have revealed that of more than 70,000 cases of COVID-19, 80.9% of patients presented mild or moderate illness [70]. SARS-CoV-2 disease is usually mild, unless combined with other comorbidities or other pre-existing diseases, with approximately 11% of cases requiring acute medical attention, and the rest requiring more intensive treatment [71].

The pathophysiology of SARS-CoV-2 infection strongly resembles that of other coronavirus infections; inducing an aggressive inflammatory response, a “cytokine storm”, that can result in extensive lung damage and scarring [72]. Disease severity in patients is not only a consequence of the viral infection, load and variant but is also dependent on host response. Therefore, the health status and age of a patient play a key role in disease progression [73]. COVID-19 has a higher fatality rate in men (2.8%) in comparison with women (1.7%); which may be attributed to the fact that ACE2 is located on the X chromosome, suggesting that there may be some alleles that present resistance to COVID-19 [74]. Conversely, this could also be attributed to the sex hormones having different immunoregulatory functions, which in turn could influence immunoprotection or disease severity [75].

### 3.1.3 Symptoms and Diagnosis

The non-specific and diverse nature of COVID-19 symptoms have made public awareness and testing very difficult, resulting in rapid and widespread infections dispersed across more than 200 countries [71]. COVID-19 symptoms are quite diverse in presentation ranging from "flu-like" symptoms such as mild sore throat and fatigue, to fever and anosmia in the majority of cases, with some people presenting as asymptomatic. The presentation of symptoms of COVID-19 infection appear after an approximate incubation period of 4.9-7 days [76]. In cases resulting in mortality, the period of onset of COVID-19 symptoms to death typically ranged from 6-41 days [77]. This period was found to be largely dependent on the age and pre-existing conditions of the patient, if any and the overall general health of the individual [78].

There are typically three key clinical stages in the disease trajectory for SARS-CoV-2; stage 1, an asymptomatic incubation period with or without detectable virus; stage II, a non-severe symptomatic period in the presence of the virus; and stage III, a severe respiratory symptomatic stage with a high viral load and severe immune dysregulation [79]. It is estimated that within 5-6 days after symptom onset, the SARS-CoV-2 viral load peaks; much earlier compared to the 10 days after symptom onset viral load peak seen with SARS-CoV [80]. In cases of severe COVID-19, progression to acute respiratory distress syndrome (ARDS) is reached after 8-9 days following symptom onset [81]. ARDS is characterised by breathing difficulty and low blood oxygen concentrations, rendering patients vulnerable to secondary infections [28]. ARDS is the primary cause for death in COVID-19 cases; in many cases respiratory failure is directly responsible for death [82]. The immune system releases a vast and prolonged level of cytokines, in response to the viral infection and/or further secondary

infections, the excessive release of which can evolve to a cytokine storm (hypercytokinemia) [83]. This uncontrolled inflammatory response can elicit multi-organ damage leading to organ failure; in particular the cardiac, hepatic and renal systems; with high mortality in patients who progressed to renal failure [84].

Recent studies have suggested that COVID-19 disease can also predispose and trigger a type of new-onset diabetes and liver dysfunction [85,86] in some patients; with as high as 77% of SARS-CoV-2 positive patients reported to have abnormal liver function tests in the acute phase, notwithstanding the level of respiratory disease severity [87]. Further pathologies reported to be associated with COVID-19 disease include pulmonary embolism, neurological inflammation and encephalitis, renal damage, gastrointestinal disorders, heart disease and stroke events complicating clinical presentations and diagnosis [88-94]. The findings suggest that COVID-19 is a complex systemic as well as an acute respiratory disease with unknown long-term repercussions.

Currently, rapid diagnostic tests for SARS-CoV-2 exposure are carried out *via* qPCR and immunoassay methods, to detect IgA, IgG and IgM following seroconversion, which remains the international gold standard. However, given the half-life of cytokines, the sampling window for virus detection is relatively small, in addition to the fact that not all SARS-CoV-2 positive patients seroconvert leading to growing concerns regarding testing accuracy and the potential for a high false negative reporting rate [95]. SARS-CoV-2 rapid tests currently use real-time reverse transcriptase-polymerase chain reaction (rRT-PCR) for the qualitative detection of SARS-CoV-2 from nasopharyngeal swabs to confirm clinical diagnosis [96]. They detect genetic components of the virus, the RNA, but this is only possible if the virus is present

at a detectable level whilst an individual is actively infected, and if the virus is not a substantially different variant to the known test variants. Genetic diversity and rapid evolution of the SARS-CoV-2 virus have been widely observed [97]. Whilst there has been some progress in the development of other diagnostic tests, given the multi-dimensional and patient variability, this has proven challenging.

### 3.1.4 Immune dysregulation

Interestingly, it has been found that not all people exposed to SARS-CoV-2 will be infected and not all develop severe respiratory illness [98]. The status of a person's immune system plays a key role in the response and outcome of the disease trajectory [99]. It has been clinically observed that SARS-CoV-2 infections are typically immunologically two phased. During the incubation and non-severe stages, a specific adaptive immune response is required for virus clearance, in order to halt disease progression to severe stages [100]. For the development of an endogenous, protective immune response at incubation and non-severe stages, the individual must be of good health and with a genetic background that elicits specific antiviral immunity [101]. In the circumstance of a compromised immune response, the virus will rapidly propagate, and extensive tissue damage will occur in affected tissues; with studies finding that the virus a particular has affinity for organs that have high ACE2 expression, such as the intestine and kidney [102]. Cellular damage in turn, induces an innate inflammation in the lungs that is mediated by pro-inflammatory macrophages and granulocytes [103].

Following the binding of the virion to the host cell induction of a highly lytic apoptosis state within the host cell, occurs with release of damage associated molecular patterns (DAMPS), such as ATP, nucleic acids and ASC oligomers (inflammasomes); these signals are recognised by neighbouring epithelial cells, endothelial cells and alveolar macrophages that trigger the start of the cytokine cascade. The resulting production of pro-inflammatory cytokines and chemokines (IL-6, IP-10, macrophage inflammatory protein 1α (MIP1α), MIP1β, and MCP1) attracts monocytes, macrophages and T cells, to the site of infection, with T cells additionally producing IFNγ, the amalgamation of which further promotes a state of inflammation. In a defective immune response, this may lead to further accumulation of immune cells in the lungs; thereby promoting the production of pro-inflammatory cytokines in the sub-locale, resulting in lung damage. The effects of the cytokine storm subsequently target other organs, resulting in multiple organ damage. Organ damage is further exacerbated by non-neutralising antibodies produced by B cells; through antibody-dependent enhancement [104].

Genetic differences are a key contributor to the variable immune response seen between individuals, with extensive immune profiling studies of COVID-19 patients revealing a highly heterogenous immune response. Current comparison of immune profiling studies is complicated by the different cohort compositions and sizes, and importantly, dissimilar clinical phenotypes and parameters used for investigation. Studies have suggested that in the most severe cases of COVID-19, with patients eventually developing severe respiratory failure, patients present one of two immune phenotypes [104]. In approximately 75% of patients, immune dysregulation is characterised by a substantial decrease of HLA-class II molecules on CD14+ monocytes in the absence of elevated ferritin. This mechanism is triggered by monocyte hyperactivation, excessive IL-6 production and profound

lymphopenia, without an increase in IL-1β levels [105]. In the second observed phenotype, the immune dysfunction is associated with macrophage activation syndrome (MAS), associated with elevated ferritin, comparatively less reduction of HLA class II molecules on monocytes and triggered by IL-1β [105].

The ensuing cytokine storm and subsequent disease severity are the direct consequence of the immune dysregulation.

### 3.1.5 Global pandemic and international health crisis

There is currently no proven treatment or cure for SARS-CoV-2 infection and disease; although several studies have found that some therapeutics, such as the steroid dexamethasone and anti-viral Remdesivir [106-108], have been found to reduce mortality and recovery times in some severe cases. Several vaccines have been developed (eg, Pfizer, Moderna, Astrazeneca, Novavax), using different strategies such as mRNA, whole virus and protein subunit [109]. SARS-CoV-2 variants have been detected globally, some with higher infection rates such as the B.1.1.7 variant, which additionally has been found to have a higher rate of morbidity and mortality [110]. Further variants, such 501Y.V2, have been found to have substantial changes in the RNA genome, leaving targeted vaccines for variants with a diminished protective response. Different variants have been found to be associated with different viral loads, suggesting differential rates of transmission [111]. Despite the successful development of several vaccines (Astrazenica, Pfizer, Moderna, Novavax) and nationwide rollout across several countries, some variant resistance has already been found to these vaccines, suggesting impaired efficacy.

Whilst several countries have controlled the first wave of the infection well, several countries are facing their third or fourth wave of infection. Incomplete and inadequate testing and ineffective quarantine measures for travel, have resulted in several outbreaks in countries that had previously controlled the virus well [112]. Due to the presentation and differential viral load of the virus, diagnosis and tracking is very difficult. In order to prevent complete exhaustion of medical facilities, diagnosis and disease severity prediction of COVID-19 are crucial. Improved and quicker management would improve both short- and long-term health outcomes and reduce the international financial burden on the healthcare system. SARS-CoV-2/COVID-19 metabolic literature is still relatively sparse although recent reports indicate that new onset-diabetes and liver injury [113, 114] are common but previously overlooked side-effects of SARS-CoV infection and these abnormalities would be expected to yield highly characteristic metabolic signatures. This may indicate a prospective burden on health systems, with a cohort of patients suffering from the long-term effects of SARS-CoV-2 disease. A recent report, published from the data in this thesis, indicates that multiple metabolic and proteomic disruptive signatures are caused by the disease [114].

### 3.1.6 SARS-CoV-2 Metabolic profiling studies to date

To date, there have been very few comprehensive metabolic profiling studies on COVID-19, with the majority of research focusing on genomic technologies and seeking to establish a link between the human genome and susceptibility to the virus. Due to the global pandemic and uncertainty regarding laboratory standards on treatment of blood samples from COVID-

19 patients for certain analyses, the research output is convoluted due to experimental inconsistencies.

**Table 3.1:** Summary of exemplar COVID-19 metabolic profiling studies

| Technique & Biofluid, Species | Number of participants | Biomarkers | Reference |
|---|---|---|---|
| NMR and targeted UPLC-MS, Plasma, human | 17 patients, 25 control | A-1-Acid, Glycoprotein, Kynurenine/Tryptophan Ratio, Glutamine/Glutamate Ratio, Fischer's Ratio | Kimhofer et al., 2020 [114] |
| NMR, Plasma, Human | 30 patients, 30 control | 3-hydroxybutyrate, Mannose, Pyruvate, Glycoproteins, Isoleucine, Leucine, Phenylalanine, Creatinine, Glucose, Valine, Lactate, Tyrosine, Formate, Acetate, Histidine, Glutamine, Alanine, Citrate, Creatinine, VLDLs, Free Cholesterols, APOA1, APO B100 | Meoni et al., 2021 [115] |
| NMR, Plasma, Human | 263 patients, 280 control | Acetoacetic Acid, 3-Hydroxybutyric Acid, Acetone, 2-Hydroxybutyric Acid, Methionine, Isoleucine, Histidine, Lysine, Glutamine, Tyrosine, Phenylalanine, Phospholipids, Triglycerides | Bruzzone et al., 2020 [116] |
| UPLC-MS, Serum, Human | 46 patients, 53 control | Mannose, Serotonin, Sphingosine, 1-(1-Enyl-Palmitoyl)-2-Arachidonyl-GPC, Kynurenine, Choline, Arachidonate, Glutamate, Arginine, N-(I-Arginino)Succinate, Citrulline, Ornithine, Glutamine, 2-Oxoglutarate, N-Acetyl-L-Glutamate, Urea, Fumarate, Argininate, Asymmetric Dimethylarginine, Symmetric Dimethylarginine, Homoarginine, N-Acetylarginine | Shen et al., 2020 [117] |
| LC-MS, Serum, Human | 33 patients, 16 control | Creatine, Creatinine, Polyamine, Circulating Amino Acids | Thomas et al., 2020 [118] |
| LC-MS/MS, Serum, Human | 120 patients (mild and severe) | Cytosine, Kynurenine, Nicotinuric Acid, Hydroxybutyrylcarnitine, Hexenoylcarnitine, Hydroxyoctanoyl | Roberts et al., 2020 [119] |
| UHPLC-MS-MS, Serum, Human | 37 patients, 17 control | Arginine, Kynurenine, Tyrptophan, Carnitine, Glycocholic aicds, Taurocholic acid, Thiamine, 2-Aminoacetanoic Acid | Xiao et al., 2021 [120] |

## 3.2 Materials and Methods

### 3.2.1 Patient enrolment and sample collection

Plasma samples were collected from 17 adults who presented COVID-19 disease symptoms and subsequently tested positive for SARS-CoV-2 infection from upper and/lower respiratory tract swabs by RT-PCR and 25 healthy controls recruited from the population who had not exhibited any COVID-19 disease symptoms and were serologically tested negative with respect to IgA/IgG antibodies as a part of this research study. SARS-CoV-2 patients had a wide range of comorbidities (Table 3.2) and symptoms (Figure 3.3) were recorded at the time of sampling in this small cohort. All samples used in the study were collected from single time-point collections, with the exception for one patient who tested positive for SARS-CoV-2 and provided three samples during their hospital stay.

Serological testing for SARS-CoV-2 antibodies were performed by the PathWest testing laboratories, Western Australia on 10µL of plasma samples using a commercial point-of-care SARS-CoV-2 IgA/IgG specific test. The study and associated metadata was coordinated at Fiona Stanley Hospital within the Western Australian South Metropolitan Health Service catchment as part of the International Severe Acute Respiratory and Emerging Infection Consortium (ISARIC)/ World Health Organisation (WHO) pandemic trail framework (SMHS Research Governance Office PRN:3976 and Murdoch University Ethics no. 2020/052). Healthy controls were enrolled as volunteers, provided study details and written consent was obtained prior to data collection in accordance with the ethical governance (Murdoch University Ethics no. 2020/053).

**Table 3.2**: WA SARS-CoV-2 Study Cohort demographics

| | SARS-CoV-2 (n=17) | Healthy Controls (n=25) |
|---|---|---|
| Sex - M | 11 (68.75%) | 17 [70.5%] |
| Age (SD) | 69.4 [±10.5] | 48.8 [± 15.3] |
| BMI (SD) | 32.9 [±9.0] | 25.9 [± 2.8] |
| Diabetic | 3 (12.5%) | 0 |
| Hypertension | 9 [47%] | 4 (16%) |
| Asthma | 3 [17.6%] | 0 |
| COPD | 1 [5.9%] | 0 |
| Arthritis | 4 [23.6%] | 0 |
| Glaucoma | 2 [11.8%] | 0 |
| Dyslipidaemia | 1 [5.9%] | 1 (4%) |
| Chronic Renal Disease | 1 [5.9%] | 0 |
| Chronic Heart Disease | 4 [23.5%] | 0 |

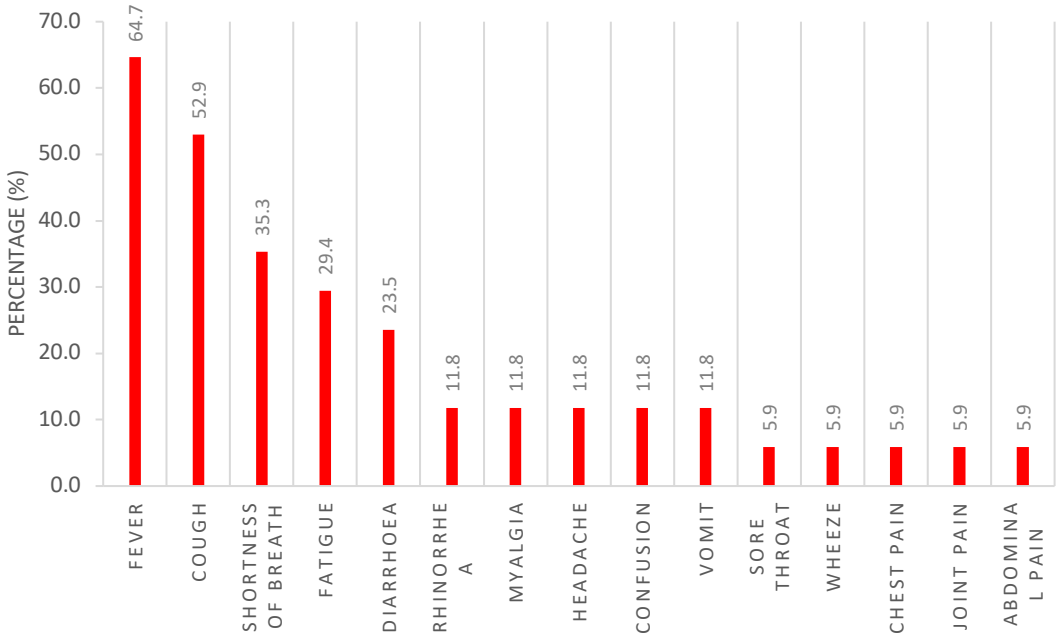



**Figure 3.3:** Symptoms presented during peak infection for SARS-CoV-2 positive patients.

### 3.2.2 Sample analysis: $^{1}$H NMR Spectroscopy of blood plasma and lipoprotein quantification

$^{1}$H NMR spectroscopy of plasma was completed according to a previously published method [68], briefly described as follows: Blood samples were centrifuged at 13,000g for 10 min at 4°C. The plasma supernatant was mixed with 75 mM pH 7.4 sodium phosphate, buffer in 1:1 ratio, and 600 µL were transferred into a Bruker SampleJet™ 5mm NMR tube. For quality control, a pooled study sample of 50 µL supernatant from all individual study samples was made and transferred into a 5 mm SampleJet™ NMR tube using the prior method.

NMR analysis was performed on a 600 MHz Bruker Avance III HD spectrometer, with a BBI probe and fitted with an automated Bruker SampleJet™ robot cooling system set to 5°C. NMR analytical calibration was run prior to the experimental run using the protocol described by Dona et al., [45]. Experiments were completed according to the Bruker *in vitro* Diagnostics research (IVDr) methods [46]. For each sample, three experiments were run in tandem with a total analysis time of 12.5 minutes; $^{1}$H 1D experiment with solvent pre-saturation [47] (32 scans, 98304 data points, spectral width of 18028.85 Hz), a 1D Carr-Purcell-Meiboom-Gill (CPMG) spin-echo experiment (32 scans, 73728 data points, spectral width of 12019.23Hz) and lastly a diffusion edited relaxation (DIRE) experiment (2 scans with 40 $t_1$ increments). All acquired data were processed in automation using Bruker Topspin™ 3.6.2, and ICON™ NMR for phasing, baseline correction and TSP calibration (δ=0).

112 lipoprotein fractions were quantified for each sample, using the Bruker IVDr Lipoprotein Subclass Analysis (B.I.-LISA ™) method [46]. This is completed by mathematically interrogating and quantifying the -$CH_2$ (1.25 ppm) and -$CH_3$ (0.8 ppm) peaks of the 1D

spectrum after normalization to the Bruker QuantRef ™ manager within Topspin™ using a PLS-2 regression model. The various lipoprotein subclasses included different fractions of intermediate-density lipoprotein (IDL, density 1.006–1.019 kg/L), very low-density lipoprotein (VLDL, 0.950–1.006 kg/L), low-density lipoprotein (LDL, density 1.09–1.63 kg/L), and high-density lipoprotein (HDL, density 1.063–1.210 kg/L). The LDL sub-fraction was systematized into six density classes (LDL-1 1.019–1.031 kg/L, LDL-2 1.031–1.034 kg/L, LDL-3 1.034–1.037 kg/L, LDL-4 1.037–1.040 kg/L, LDL-5 1.040–1.044 kg/L, LDL-6 1.044–1.063 kg/L), and the HDL sub-fractions into four density classes (HDL-1 1.063–1.100 kg/L, HDL-2 1.100–1.125 kg/L, HDL-3 1.125–1.175 kg/L, and HDL-4 1.175–1.210 kg/L). The 1-acid glycoprotein N-acetyl-glucosamino (N-acetyl) signal integrals were calculated as Glyc A (2.06ppm) and Glyc B (2.10 ppm), as determined from the DIRE spectra.

### 3.2.3 Amino Acid/Biogenic Amine Quantification

LC-MS analysis was performed using a Waters Acquity UPLC® (Waters Corp., Milford, MA, USA) coupled to a Waters Xevo TQ-S MS (Waters Corp., Wilmslow, UK). Peak integration and calculation of resultant quantification data were performed using the TargetLynx package in MassLynx v4.2 (Waters Corp., Milford, MA, USA). Calibration curves were linearly fitted with a weighting factor of $1/x^2$.

Fully quantitative analysis for 35 amino acids were performed based on a UPLC-triple quadrupole MS method following derivatisation [51]. Unlabelled amino acid standards and ammonium formate were purchased from Sigma-Aldrich (MO, USA). Stable isotope labelled internal standard non-canonical and canonical amino acid mixes were purchased from

Cambridge Isotope Laboratories (MA, USA). Water, acetonitrile, methanol and isopropanol (Optima grade) were purchased from Thermo Fisher Scientific. Calibrators and quality controls were prepared from a stock solution of physiological amino acids (acids, basics and neutrals) at 500 uM. Asparagine and glutamine were prepared freshly at 0.5 mM due to instability. For calibration, a working stock solution of all amino acids was prepared at 400 uM in water and separately diluted to 200, 100, 40, 20, 10, 4, 2 and 1 uM and for quality controls, at 300, 75, 15 and 3 uM.

The stable isotope labelled (SIL) internal standard solution (12.5 uM in water) was prepared from stocks and was added to each sample. Methanol was then added to effect protein precipitation. Following centrifugation, the extract supernatant was taken for a derivatization step with AccQTag$^{TM}$ reagent (Waters Corp., Milford, MA, US). Finally, samples were then diluted 1:50 with LC-MS grade water for analysis LC-MS analysis.

Amino acid analysis was performed using a Waters Acquity I-class UHPLC system (Waters Corp., Milford, MA, USA) coupled to a Waters TQ-XS triple quadrupole mass analyser operating with positive electrospray ionisation (ESI) and selected reaction monitoring (SRM) (Waters Corp., Wilmslow, UK). Chromatographic separation was completed using an Acquity UPLC HSS T3 1.8 um 2.1 x 150 mm column (Waters, Milford, MA, USA). Eluent A contained 2 mM ammonium formate in water and eluent B comprised of 2 mM ammonium formate acetonitrile/water 95/5 (v/v). Flow rate was set at 0.6 mL/min and column temperature were maintained at 45°C. The autosampler compartment was then cooled to 4°C and using the full-loop injection mode, a 2 uL injection volume was performed.

### 3.2.4 Quantification of Tryptophan pathway metabolites

LC-MS analysis was performed using a Waters Acquity UPLC® (Waters Corp., Milford, MA, USA) coupled to a Waters Xevo TQ-S MS (Waters Corp., Wilmslow, UK). Peak integration and calculation of resultant quantification data were performed using the TargetLynx package in MassLynx v4.2 (Waters Corp., Milford, MA, USA). Calibration curves were linearly fitted with a weighting factor of $1/x^2$.

The quantification of tryptophan and 14 products of its metabolic pathway was performed as previously reported [52]. Briefly, tryptophan metabolites were quantified from 50 mL of plasma. Sample extraction was performed on a Biomek i5 sample automation system. SIL internal standards (20 mL) were added to all samples prior to protein precipitation via the addition of 250 mL of methanol containing 2 mM ammonium formate. Following mixing, samples were then transferred to a Phenomenex PHREE™ phospholipid removal solid phase extraction plate (Phenomenex, NSW, Aus). PHREE plates were then washed with an additional 150 mL of methanol containing 2 mM ammonium formate. Eluent collection plates were dried using a SpeedVac vacuum concentrator (Thermo Fisher, Massachusetts, USA). Dried extracts were re-suspended in 100 mL of water with 0.1% formic acid prior to LC-MS analysis.

### 3.2.5: Cytokine Luminex Multiplex Assay

To quantify and assess the immune response, xMAP cytokine assays were carried out when sample volume was sufficient (n = 17 SARS-CoV-2 positive samples and n = 25 healthy) in 96-well polystyrene microplates using a MagPix detection system (Luminex, USA). A total of 34

cytokines and chemokines were quantified using a multiplexed Human ProcartaPlex Panel 1A (Life Technologies, USA) according to manufacturer's instructions.

Briefly, capture bead mixes were incubated overnight with 25 µL of serum sample at 4 °C and then washed out with wash buffer. Samples and standards were then incubated with detection antibodies for 30 min at room temperature, followed by three washes with the wash buffer; all washes were performed using a Bio-Plex Pro Wash station (Bio-Rad Laboratories). Streptavidin-PE was added, and the plate incubated for 30 min at room temperature (20 °C). Plates were then washed again before the reading buffer was added. All buffers and reagents were supplied with the Luminex detection kit (Life Technologies, USA). The plate was incubated for further 5 min at room temperature before analysis. All incubation steps were carried out using an orbital shaker at 500× rpm. Data were analyzed using Luminex xPonent software. A total of 14 cytokines were deemed to have sufficient detection and coverage and were selected to be used for downstream analysis.

### 3.2.6: Data Processing and Statistical Analysis of NMR data

Following spectral preprocessing and normalisation in *R*, using the "Metabomate v0.5" package, NMR data was exported into SIMCA 14.0.1 (Umetrics, Umea, Sweden) for multivariate analysis. Initial Principal Component Analysis (PCA) was carried out to identify extreme outliers and inspect sample variability. Supervised orthogonal partial least square discriminant analysis (OPLS-DA) [121] was performed to investigate and discriminate metabolites between groups. Statistical significance and robustness of OPLS-DA models were then validated by permutation testing of the Y matrix for 1000 permutations.

Permutation testing comprises of the repeated randomisation of class labels in the dataset, with classified models then recalculated using this new permutated dataset. Permutation testing evaluates the null hypothesis; if the null hypothesis is accepted, then no difference will be found in the quality-of-fit value between the original model dataset and the permutated set, and the given classification model is significant. The obtained p-values were adjusted using Benjamini-Hochberg (BH) correction [48] with a false discovery rate (FDR) threshold of 0.05.

Spearman's correlation was carried to interrogate the relationship between each dataset, hierarchically clustered and mapped using ggplot2 and corrplot visualisations. Ratios of significance for biological interpretation were tested for significance using Mann Whitney U testing, corrected for multiple testing (Benjimini-Hochberg, pvalue<0.01).

### 3.2.7: Data Processing and Statistical Analysis of quantified biogenic amine and tryptophan pathway metabolites

Peak integration and calculation of resultant quantification data were performed using the TargetLynx package in MassLynx v4.2 (Waters Corp., Milford, MA, USA). Once data were scaled to account for 1:1 dilution of sample with water and converted to appropriate scales across both assays, data were then modelling using principal components analysis (PCA) (Simca V.15.0., Umetrics, Sweden) to investigate variation. Scores plots were used to examine the relationship between healthy controls and SARS-CoV-2 positive patients. To further interrogate the relationship and differences between healthy controls and SARS-CoV-2 patients, orthogonal partial least-squares discriminant analysis (OPLS-DA). To test the

the robustness of the model Subsequent statistical analysis was conducted using R; with "ggplot2" used to visualise log-transformed metabolite concentrations. Statistical significance of individual metabolites was evaluated using Wilcoxon Rank Sum testing, corrected for multiple testing (Benjamini-Hochberg, pvalue<0.01).

Spearman's correlation was carried to interrogate the relationship between each dataset, hierarchically clustered and mapped using ggplot2 visualisations. Ratios of significance for biological interpretation were tested for significance using Mann Whitney U testing, corrected for multiple testing (Benjimini-Hochberg, pvalue<0.01). Metaboanalyst (V5.0) was used to perform metabolite set enrichment analysis (MSEA), data were mean centred, against a reference metabolome extracted from SMPDB (Small Molecule Pathway Database, www.smpdb.ca) comprising of 99 metabolite sets based on normal human metabolic pathways, and a minimum threshold of 3 hit was applied. Differential enrichment was ranked based on p-values (Fisher's exact test). Hierarchical edge clustering was applied to calculated spearmans correlation of significant MS quantified metabolites and cytokines, using the R package, "edgebundleR", to investigate the link between significant metabolites and cytokines.

# 3.3 Results

### 3.3.1: Clinical pathology results

Analysis of preliminary results from clinical metadata collected during SARS-CoV-2 infection, revealed a pathological signature far deviated from the upper limit of normal ULN) for a significant proportion of measures (Figure 3.4). Deviation of biochemistry results from the ULN is shown for the majority of parameters, suggesting a vast systemic response. During the peak of SARS-CoV-2 positivity is D-Dimer, Procalcitonin, White Cell Count, Lymphocytes, Monocytes, Eosinophils and Basophils were found to be depleted. Measures found with less marked elevation or within normal limits include Bilirubin, Potassium, Blood Urea Nitrogen, Mean Cell Haemoglobin, and Neutrophils. All other parameters presented as elevated including C reactive protein, Total Protein, Albumin, Globulin, Alanine Aminotransferase (ALT), Alkaline Phosphatase (ALP), and Gamma Glutamyl Transferase (GGT).

### 3.3.2 NMR data analysis and SARS-CoV-2 signatures

PCA scores plots from acquired spectra (1D (figure 3.10A), CPMG (figure 3.13A), and DIRE (figure 3.14A) show similar results with distinct separation of healthy and SARS-CoV-2 positive patients, and some overlap in the component space. Exemplar spectra from all 3 spin acquisitions of SARS-CoV-2 positive patients and healthy control analysis are shown and with metabolite annotation in figures 3.5– 3.9, showing markedly different spectra comparing SARS-CoV-2 patients to healthy controls.

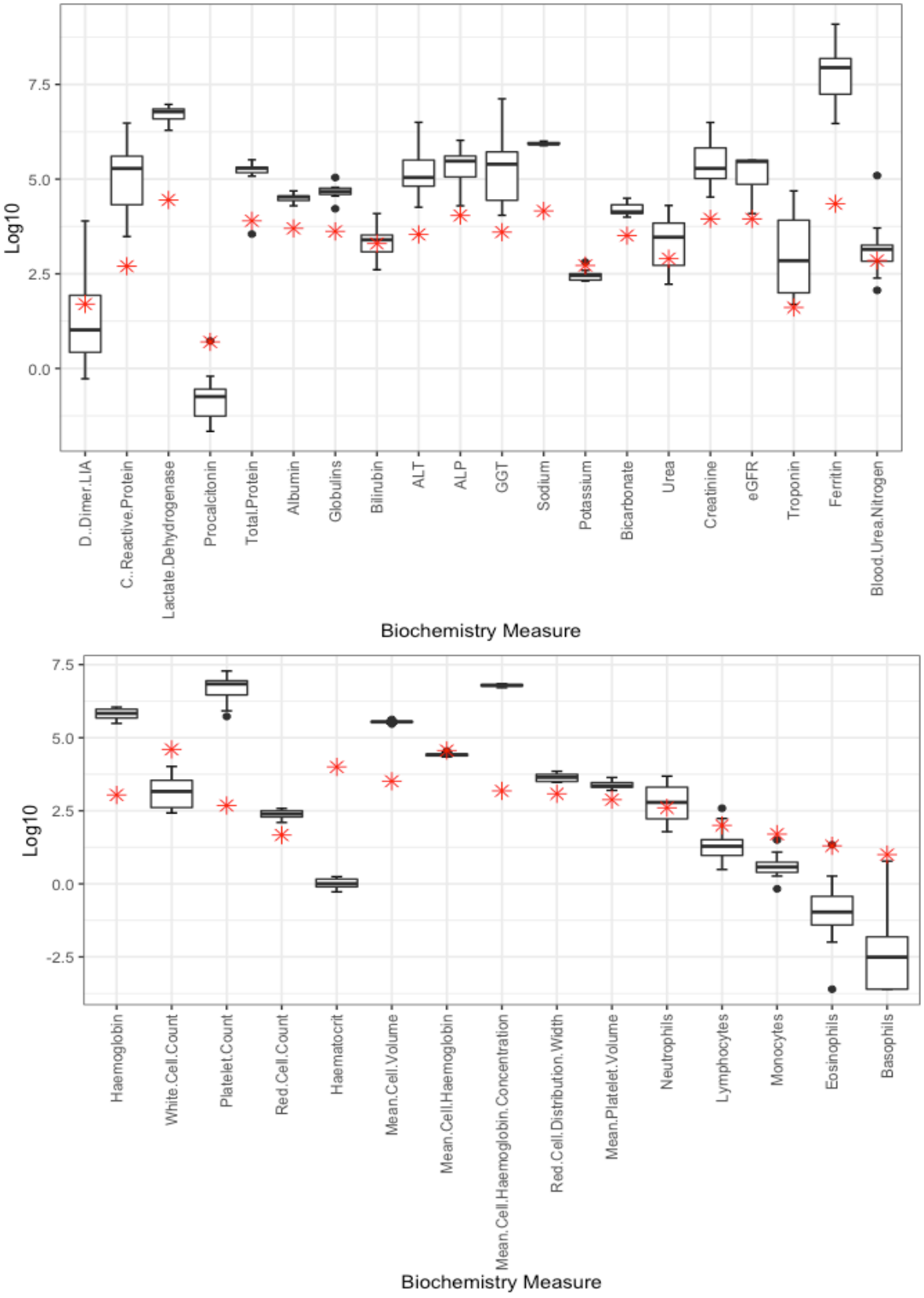



**Figure 3.4:** Blood biochemistry results from peak of SARS-CoV-2 infection, log scaled with upper limit of normal (ULN) denoted by red asterisk (*)

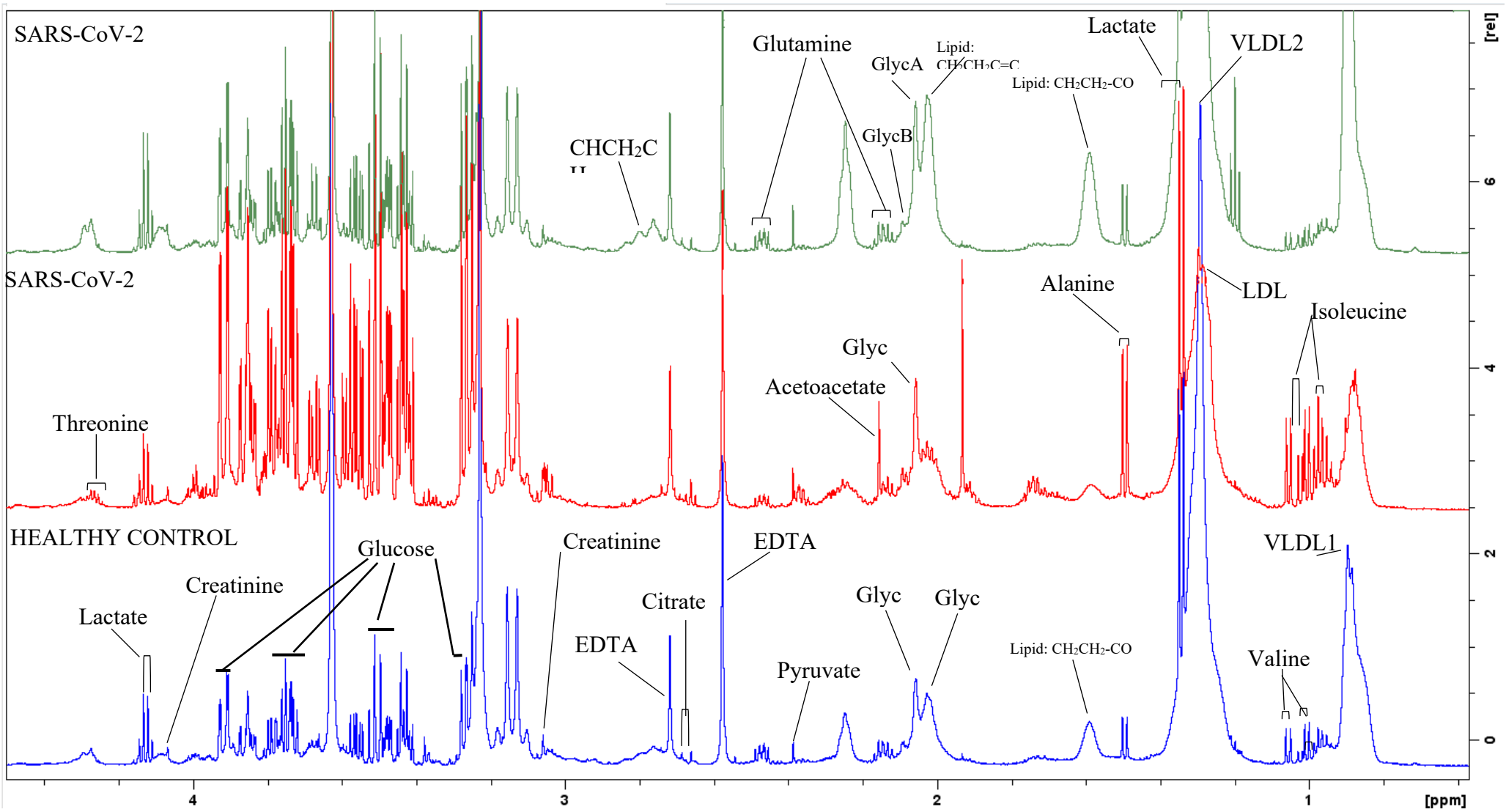



**Figure 3.5: Standard 1D NMR spectra aliphatic ($\delta$ 0.5-4.5)**: Exemplar spectra from 2 SARS-CoV-2 positive patients compared to a healthy control.

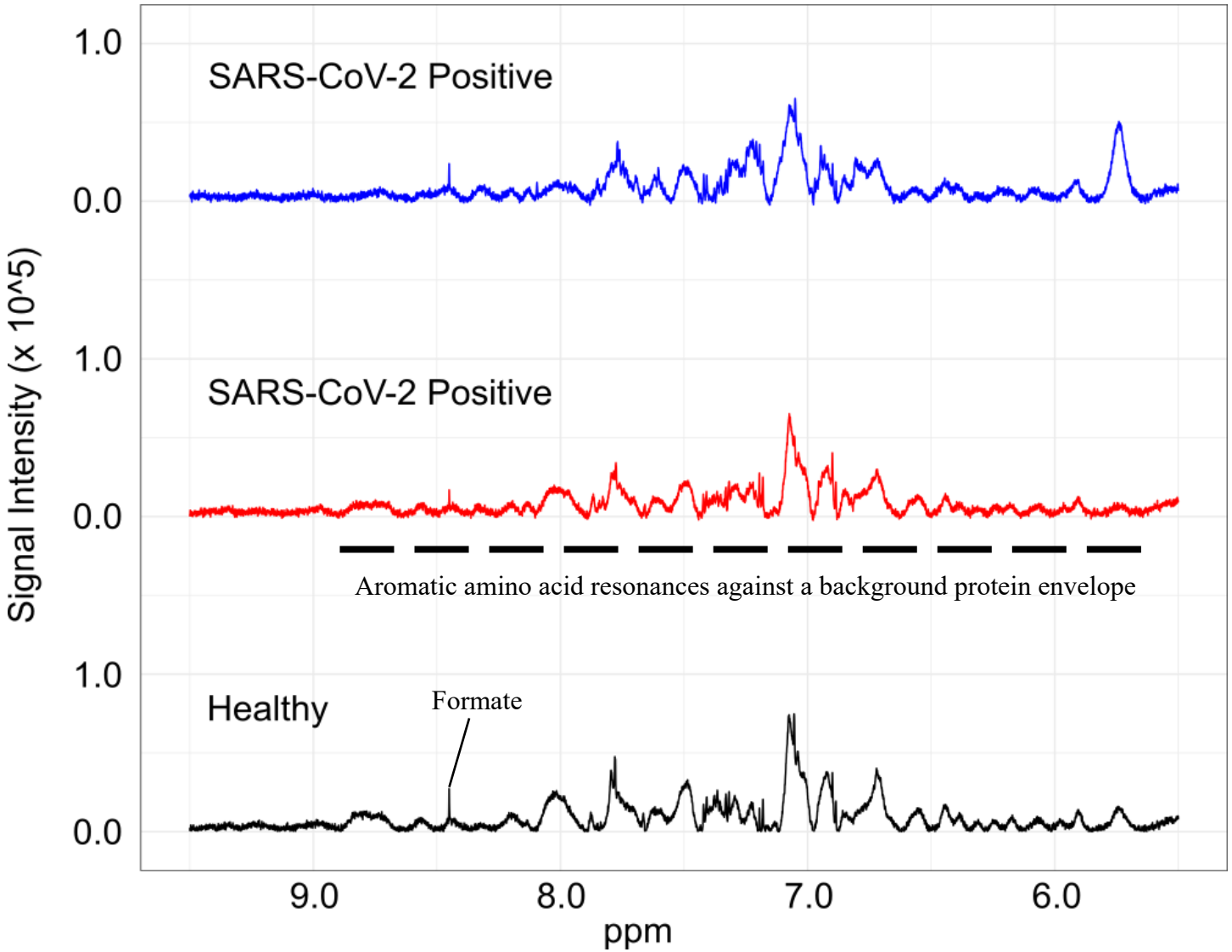



**Figure 3.6: Standard 1D NMR spectra aromatic region ($\delta$ 5.5-9.5):** Exemplar spectra from 2 SARS-CoV-2 positive patients compared to healthy control.

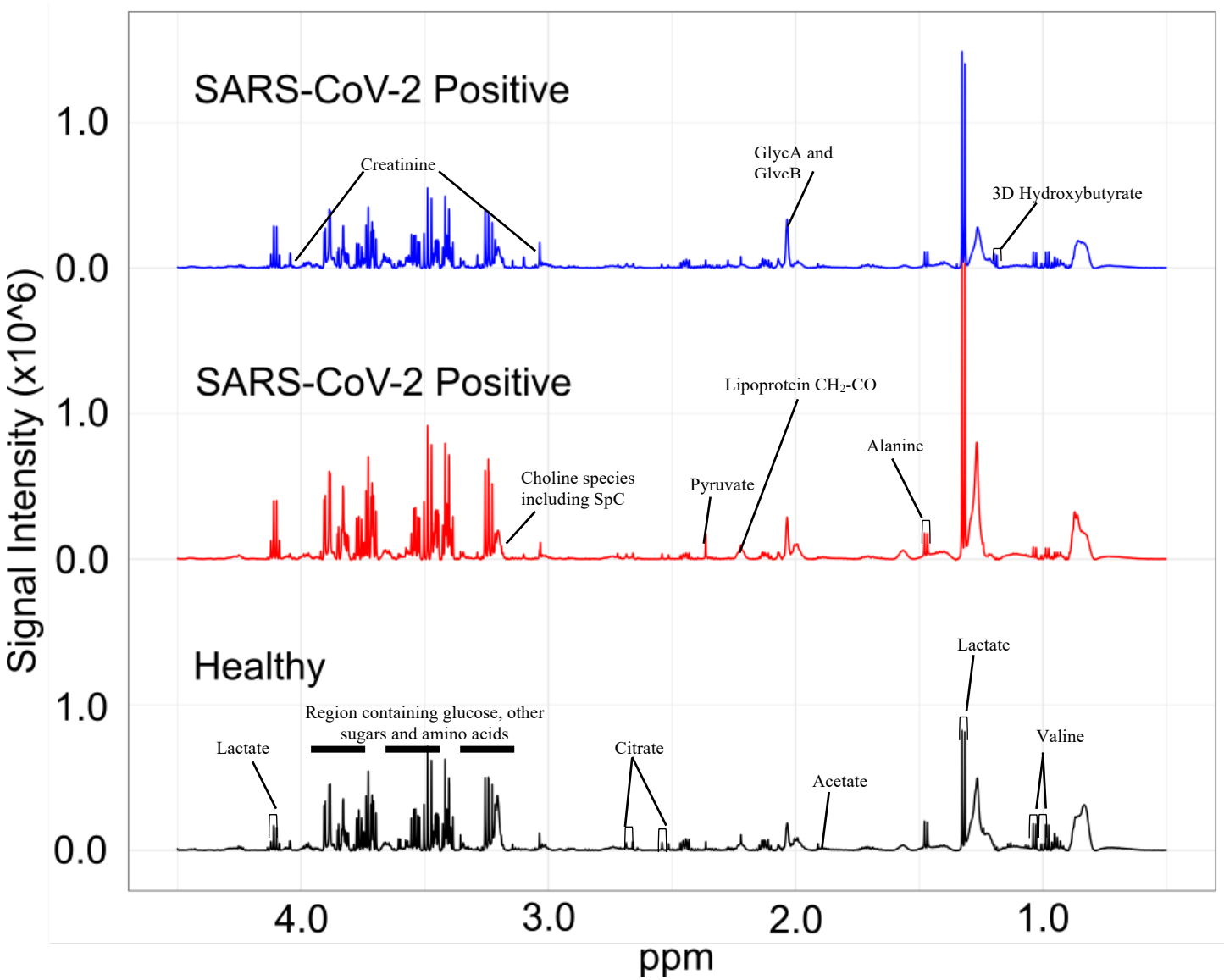



**Figure 3.7: CPMG NMR spectra aliphatic region ($\delta$ 0.5-4.5):** Exemplar spectra from 2 SARS-CoV-2 patients compared to a healthy control.

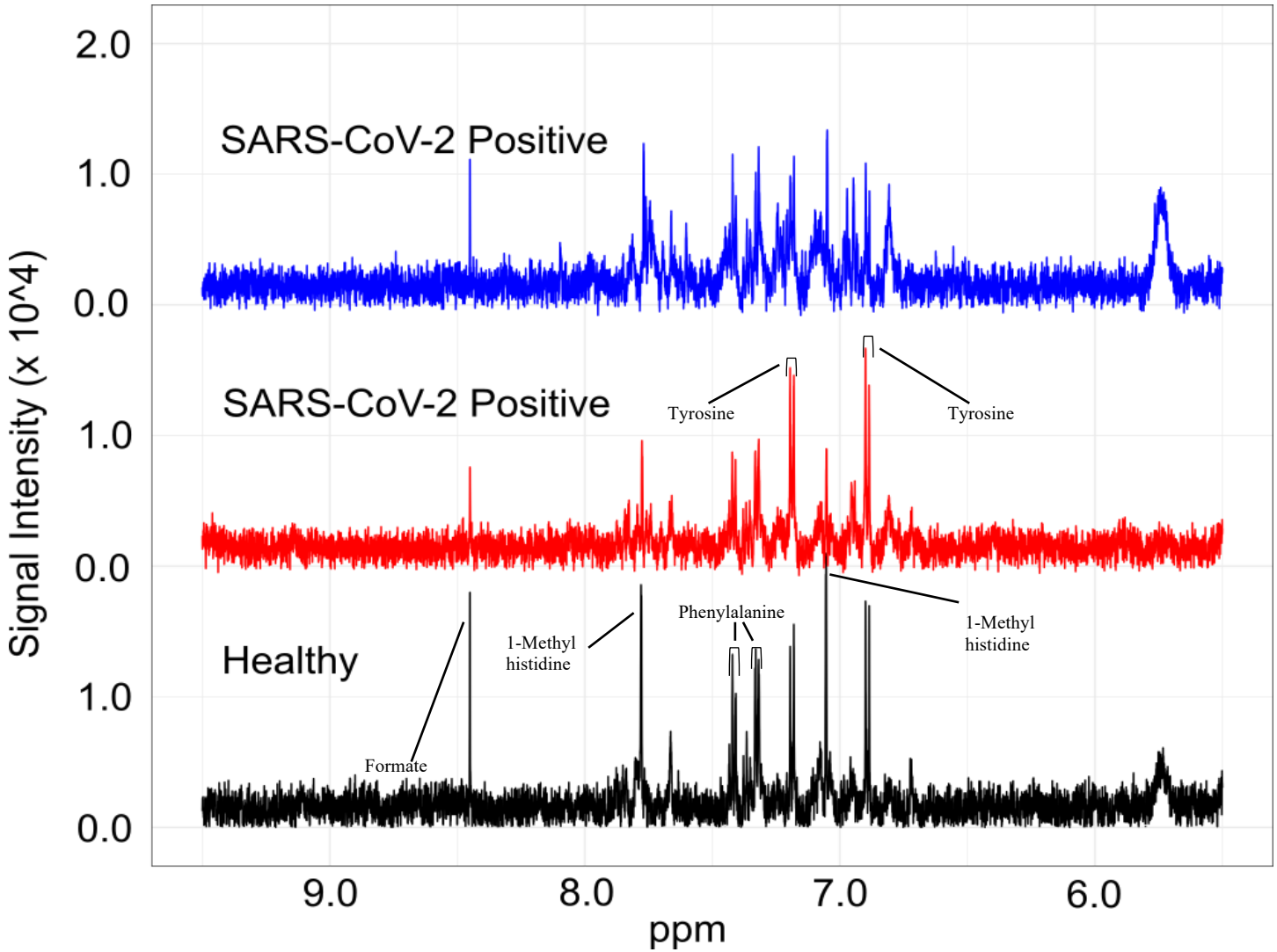



**Figure 3.8: CPMG NMR spectra aromatic region ($\delta$ 5.5-9.5):** Exemplar spectra from 2 SARS-CoV-2 patients compared a healthy control. The CPMG pulse sequence suppresses the larger molecular components such as lipoproteins and increases the visibility of low molecular weight molecular and the more mobile lipoprotein fractions.

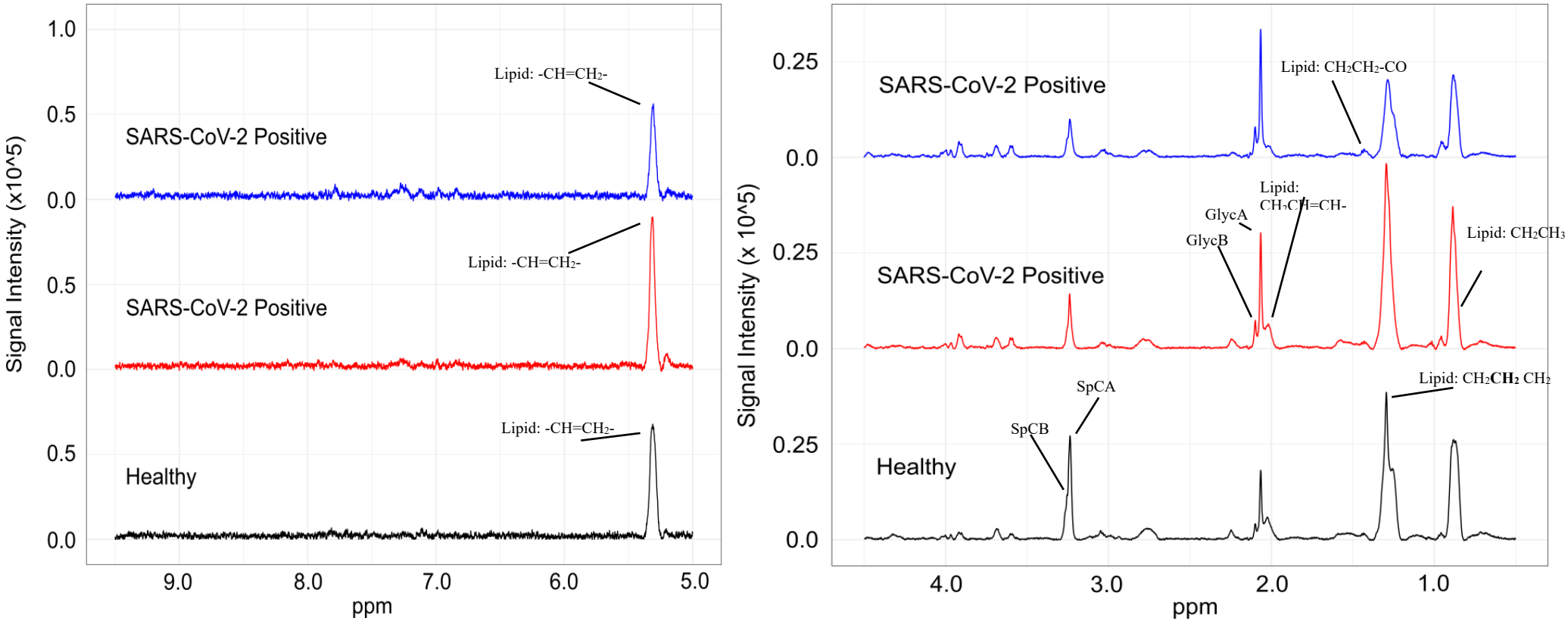



**Figure 3.9: DIRE NMR spectra aliphatic region ($\delta$ 0.5-4.5 and aromatic regions ($\delta$ 5.5-9.5):** Exemplar spectra from SARS-CoV-2 positive patients compared to healthy controls.

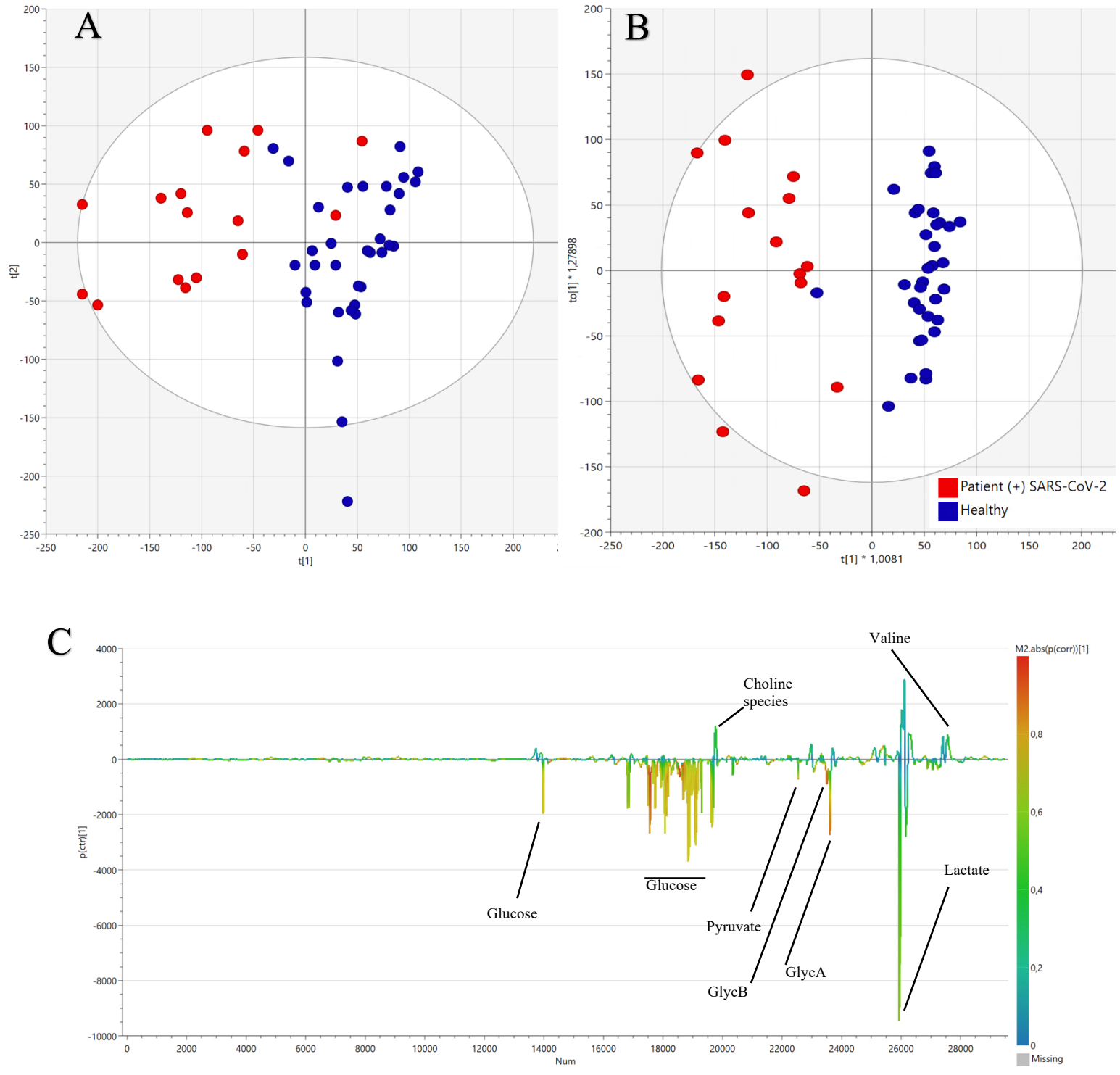



**Figure 3.10: PCA and OPLS-DA of standard 1D NMR spectra between SARS-CoV-2 patients and healthy controls.** A) PCA scores plot of SARS-CoV-2 positive (+) patients and healthy control profiles $R^2X$ 0.677, $Q^2Y$ 0.428). B) OPLS-DA scores plot of SARS-CoV-2 positive (+) patients and healthy controls ($R^2X$ 0.348, $R^2Y$ 0.856, $Q^2Y$ 0.742). C) Loadings from OPLS-DA modelling, highlighting the metabolites with the most influence in the model. Colour is equivalent to the absolute value of the correlation loading.

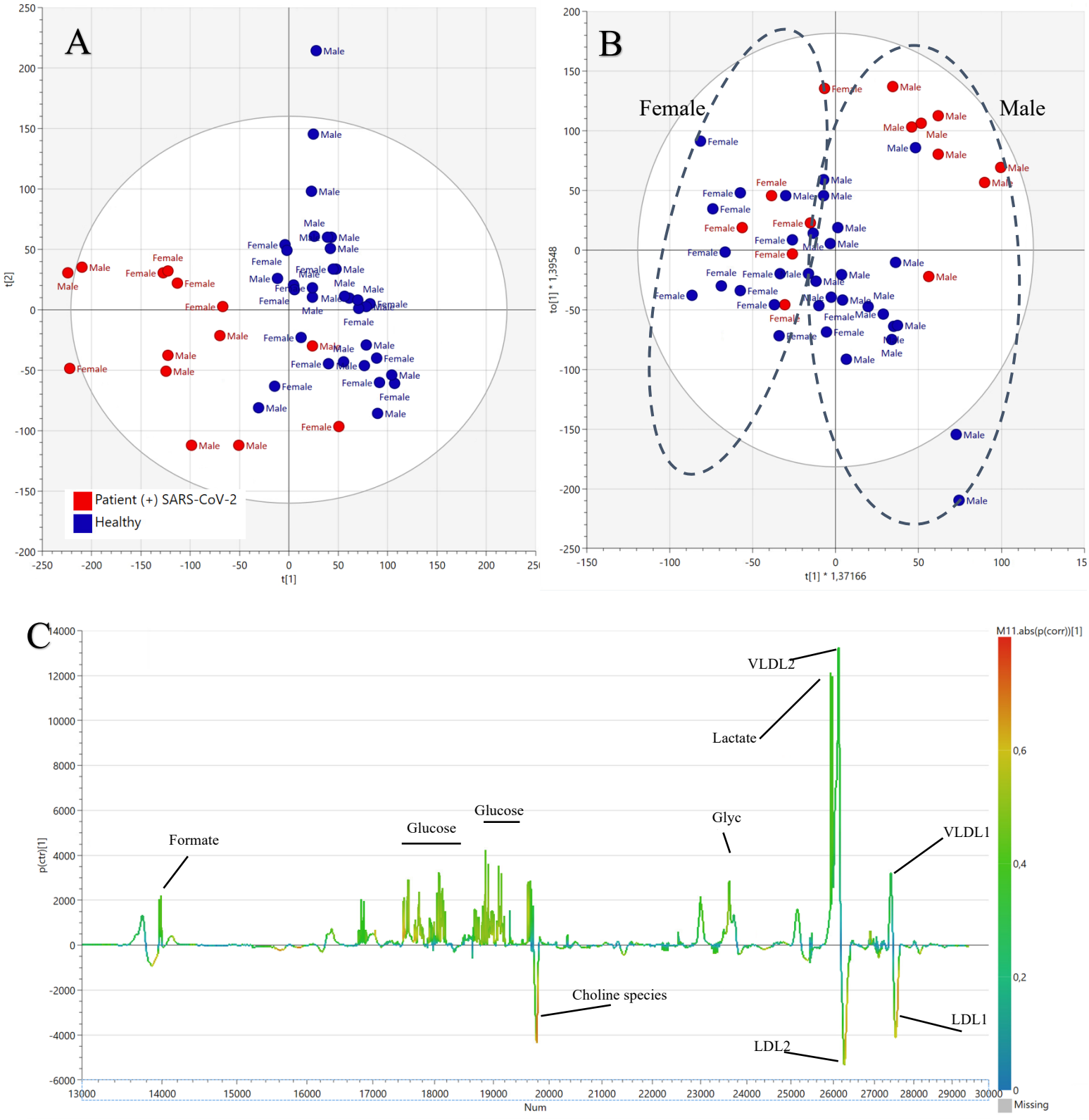



**Figure 3.11: PCA and OPLS-DA modelling on 1D standard NMR spectra based on sex differences between SARS-CoV-2 patients and healthy controls.** A) PCA scores plot of SARS-CoV-2 positive (+) patients and healthy control profiles based on sex classification ($R^2X$ 0.655, $Q^2Y$ 0.398). B) OPLS-DA scores plot of SARS-CoV-2 positive (+) patients and healthy controls based on sex differences ($R^2X$ 0.245, $R^2Y$ 0.562, $Q^2$ 0.0184) with missing data variables removed. C) Loadings from OPLS-DA modelling, highlighting the metabolites with the most influence in the model. Colour is equivalent to the absolute value of the

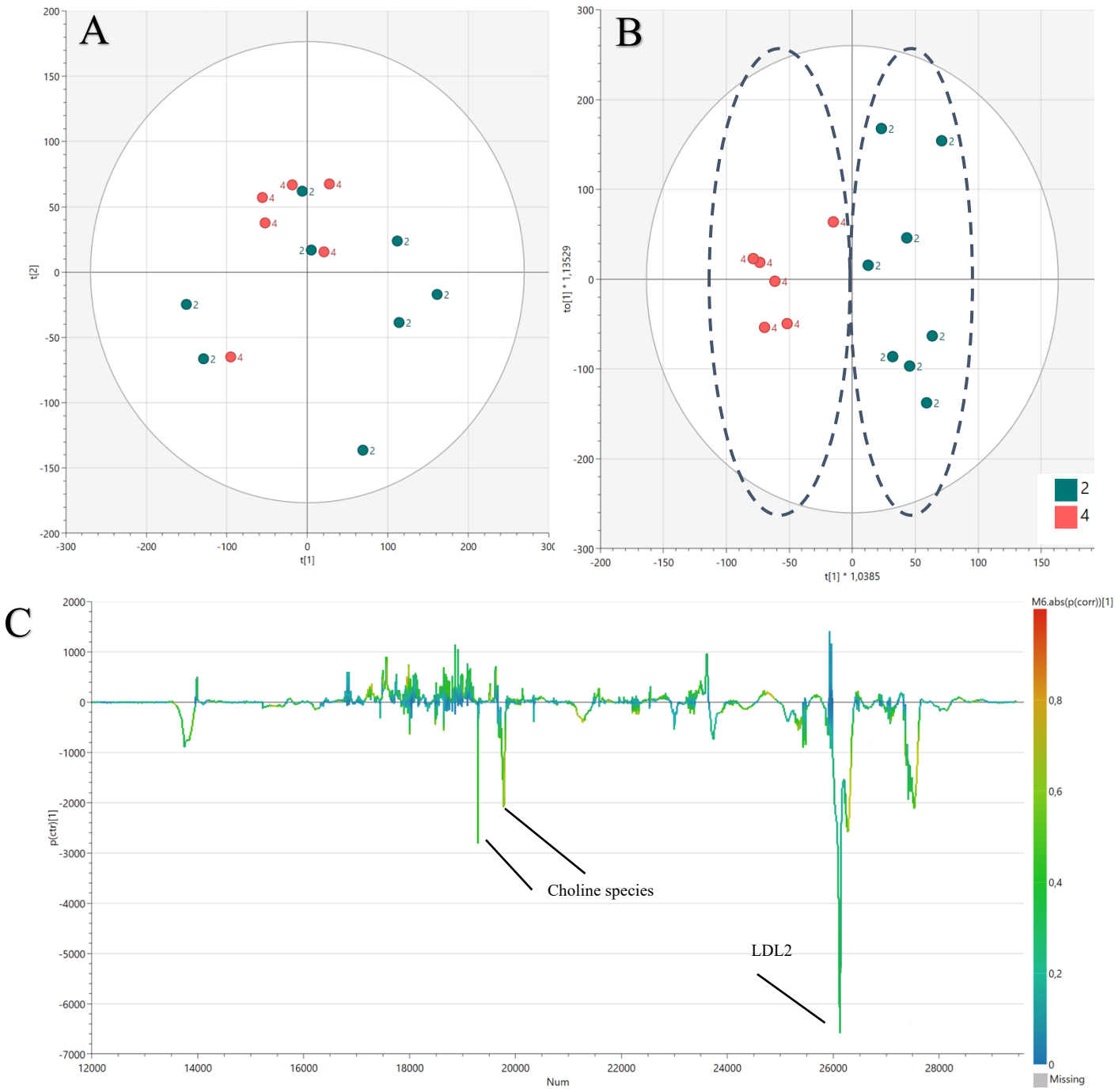



**Figure 3.12: PCA and OPLS-DA modelling on 1D standard NMR based on BMI differences in SARS-CoV-2 patients.** A) PCA scores plot of SARS-CoV-2 positive (+) patients BMI classification (2 = <30, 4= >30, restricted thresholds due to sample size) ($R^2X$ 0.418, $Q^2Y$ 0.142). B) OPLS-DA scores plot of SARS-CoV-2 positive (+) patients based on BMI classification (2 = <30, 4= >30, restricted thresholds due to sample size) ($R^2X$ 0.381, $R^2Y$ 0.864, $Q^2$ 0.0763). Samples with missing data variables were removed from analysis. C) Loadings from OPLS-DA modelling, highlighting the metabolites with the most influence in the model. Colour is equivalent to the absolute value of the correlation loading.

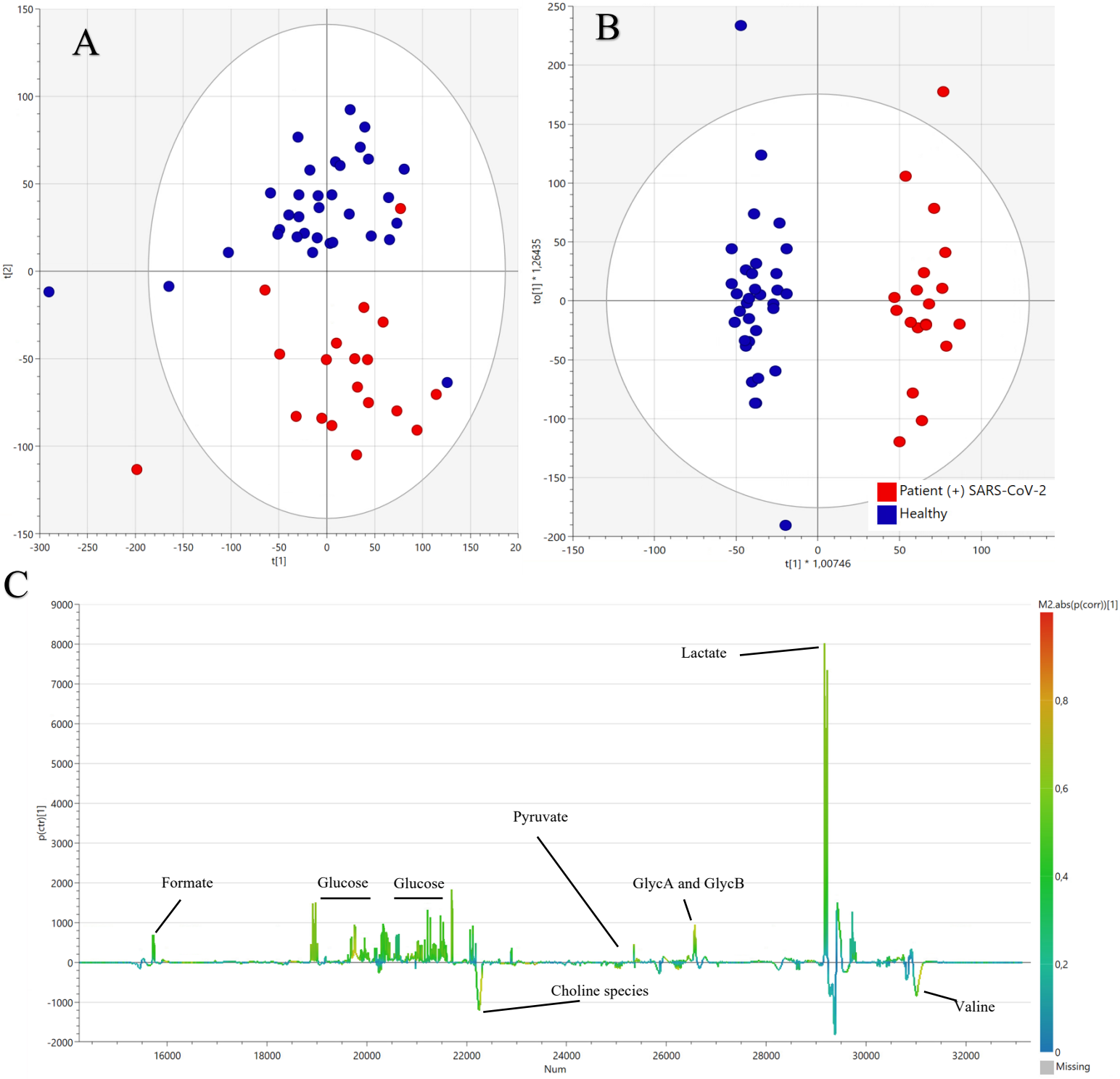



**Figure 3.13: PCA and OPLS-DA of CPMG NMR spectra.** A) PCA scores plot of SARS-CoV-2 positive (+) patients and healthy control profiles ($R^2X$ 0.429, $Q^2Y$ 0.206). B) OPLS-DA scores plot of SARS-CoV-2 positive (+) patients and healthy control profiles ($R^2X$ 0.284, $R^2Y$ 0.959, $Q^2$ 0.724). C) Loadings from OPLS-DA modelling, highlighting the metabolites with the most influence in the model. Colour is equivalent to the absolute value of the correlation loading.

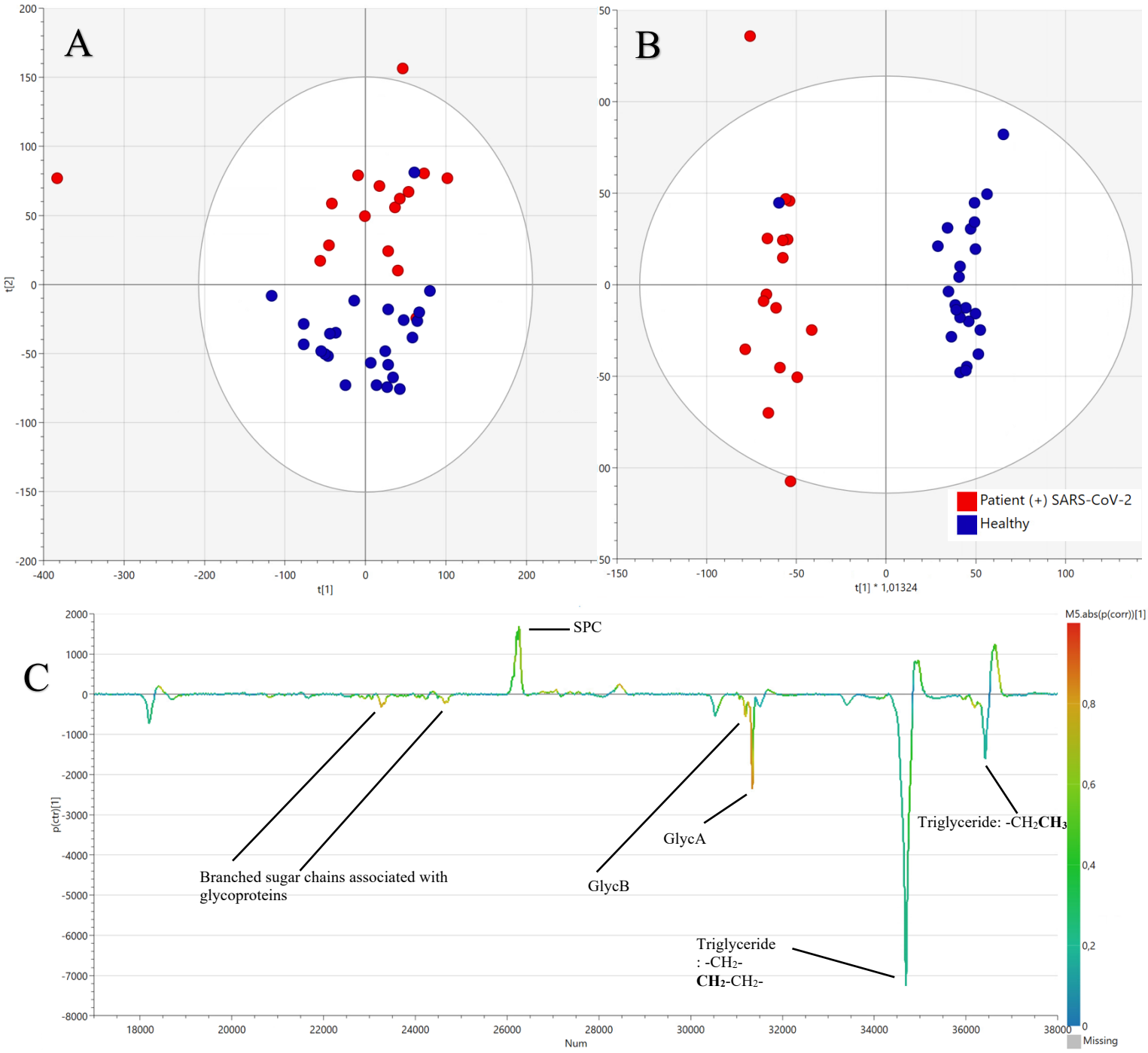



**Figure 3.14: PCA and OPLS-DA of DIRE NMR spectra between SARS-CoV-2 patients and healthy controls.** A) PCA scores plot of SARS-CoV-2 positive (+) patients and healthy control profiles $R^2X$ 0.339, $Q^2Y$ 0.077). B) OPLS-DA scores plot of SARS-CoV-2 positive (+) patients and healthy controls ($R^2X$ 0.122, $R^2Y$ 0.975, $Q^2$ 0.682). C) Loadings from OPLS-DA modelling, highlighting the metabolites with the most influence in the model. Colour is equivalent to the absolute value of the correlation loading.

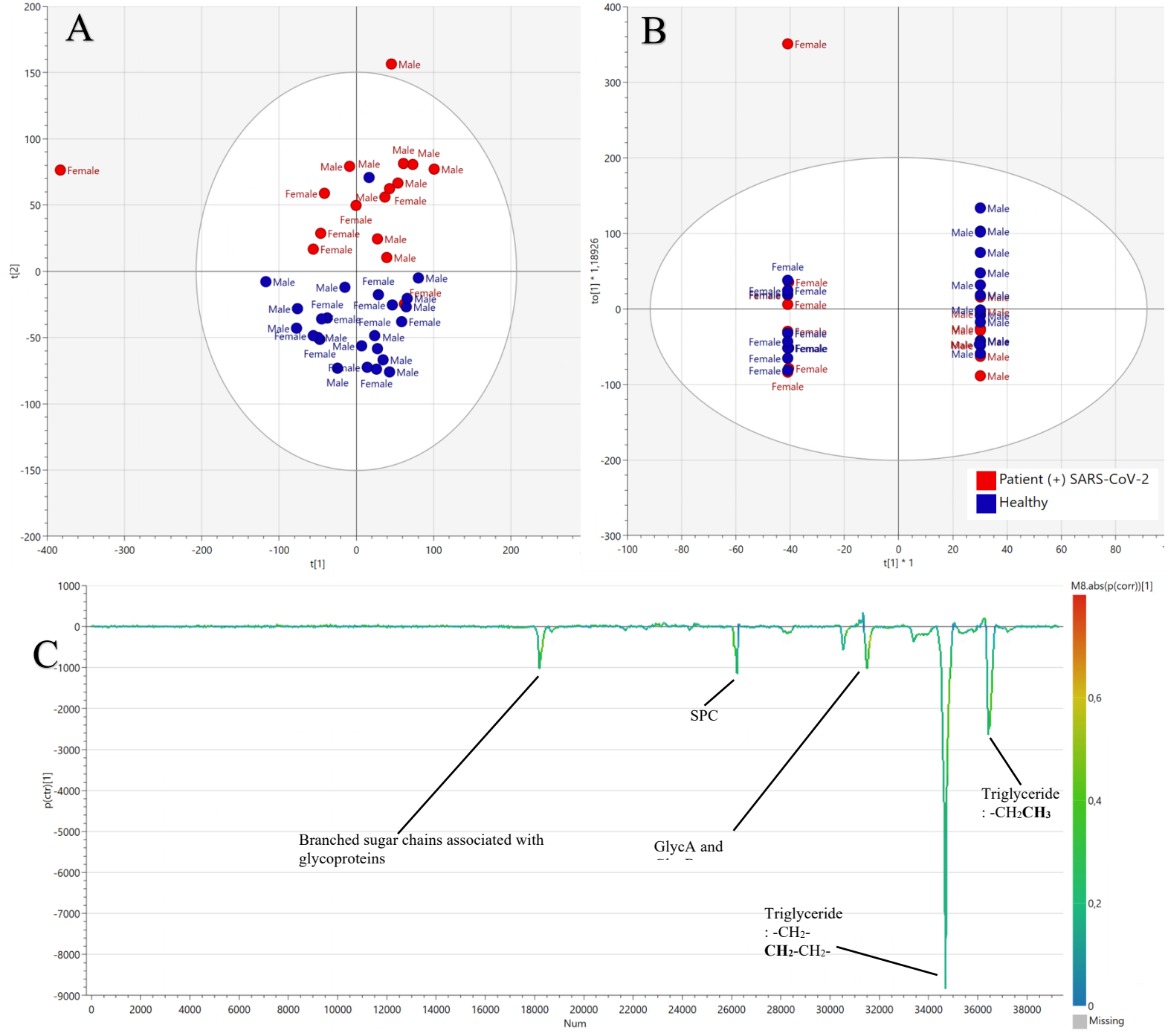



**Figure 3.15: PCA and OPLS-DA modelling of DIRE NMR spectra based on sex differences between SARS-CoV-2 patients and healthy controls.** A) PCA scores plot of SARS-CoV-2 positive (+) patients and healthy control profiles based on sex classification ($R^2X$ 0.339, $Q^2Y$ 0.077). B) OPLS-DA scores plot of SARS-CoV-2 positive (+) patients and healthy controls based on sex differences ($R^2X$ 0.235, $R^2Y$ 0.562, $Q^2$ 0.0014) with missing data variables removed. C) Loadings from OPLS-DA modelling, highlighting the metabolites with

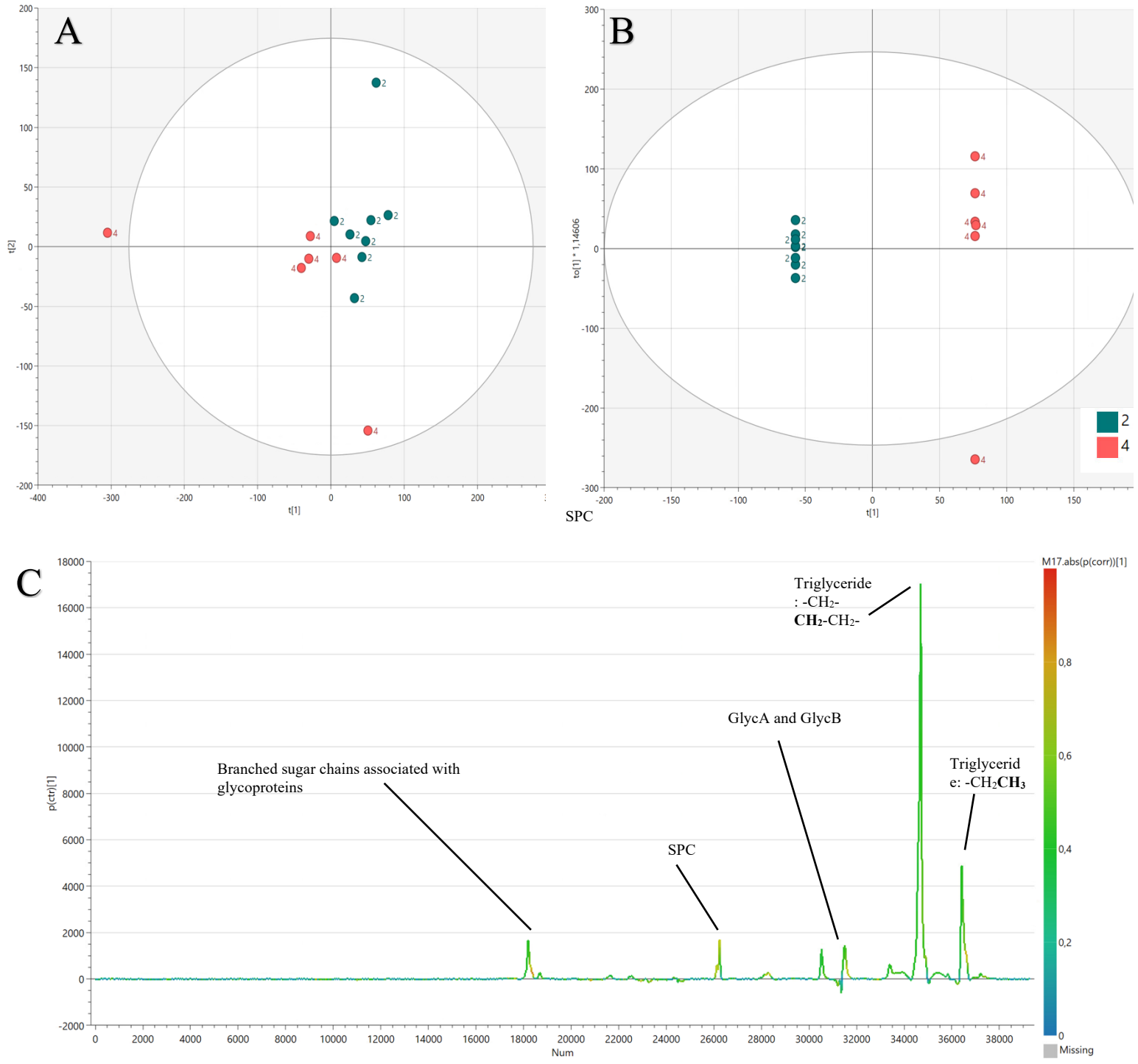



**Figure 3.16: PCA and OPLS-DA modelling on DIRE NMR spectra based on BMI differences in SARS-CoV-2 patients** A) PCA scores plot of SARS-CoV-2 positive (+) patients BMI classification (2 = <30, 4= >30, restricted thresholds due to sample size) ($R^2X$ 0.322, $Q^2Y$ -0.007). B) OPLS-DA scores plot of SARS-CoV-2 positive (+) patients based on BMI classification (2 = <30, 4= >30, restricted thresholds due to sample size) ($R^2X$ 0.522, $R^2Y$ 1, $Q^2$ 0.0763). Samples with missing data variables were removed from analysis. C) Loadings from OPLS-DA modelling, highlighting the metabolites with the most influence in the model. Colour is equivalent to the absolute value of the correlation loading.

Clear separation between SARS-CoV-2 patients and healthy controls is shown in the principal components score plot for the model based on standard 1D NMR spectra ($R^2X$ 0.677, $Q^2Y$ 0.428), with 2 healthy controls clustering with the SARS-CoV-2 patients. The standard 1D NMR spectra (figure 3.5 and 3.6) present the strongest PCA modelling in terms of clustering (figure 3.10A, $R^2X$ 0.677, $Q^2Y$ 0.428), which indicates that both high molecular weight and small molecules are important components of the infection signature as CPMG involved in signal suppression of the larger molecular weight compounds (figure 3.1 and 3.2) (Figure 3.13A, $R^2X$ 0.429, $Q^2Y$ 0.206) and DIRE attenuates the larger molecules such as glycoproteins (figure 3.9) (figure 3.14A, $R^2X$ 0.339, $Q^2Y$ 0.077).

Clustering is observed with standard 1D NMR spectra in OPLS-DA modelling (figure 3.10B) ($R^2X$ 0.348, $R^2Y$ 0.856, $Q^2Y$ 0.742). Group discrimination is seen with CPMG (figure 3.13B) ($R^2X$ 0.284, $R^2Y$ 0.959, $Q^2$ 0.724) and DIRE (figure 3.14B) ($R^2X$ 0.122, $R^2Y$ 0.975, $Q^2Y$ 0.682) spectra. Standard 1D spectra OPLS-DA models show two distinct clusters, with healthy controls to the left and SARS-CoV-2 patients to the right, but with one healthy control grouping with the SARS-CoV-2 positive group. This healthy control was later found to be positive for IgG antibodies using a SARS-CoV-2 antibody detection kit. Permutation testing of the model (1000 permutations) illustrated a robust model. The most highly discriminating metabolites from the model coefficients are highlighted in the loadings plot (figure 3.10C), and includes elevated lactate, GlycA and glucose, and reduced valine, SpC and other mobile lipids: elements also highlighted in loadings from CPMG spectra modelling (figure 3.13C). DIRE spectra also gave good discrimination in OPLS-DA modelling, despite the lack of spectral complexity, and captured the contribution (figure 3.14C) of GlycA and GlycB found to be significantly increased in SARS-CoV-2 patients, and supramolecular phospholipid

composite (SpC) signals [75], found to be significantly lower in SARS-CoV-2 patients (pvalue <0.05).

### 3.3.3 Sex differences and Body Mass Index in relation to SARS-CoV-2 pathology

To investigate the impact of sex differences on the models, PCA and OPLS-DA modelling was conducted on standard 1D and DIRE spectra, using sex as a discriminator rather than infection. Clustering of sexes was seen within groups in PCA (figure 3.11A), with clear grouping of males and females. OPLS-DA modelling (figure 3.11B) ($R^2X$ 0.245, $R^2Y$ 0.562, $Q^2Y$ 0.0184) shows sex-differentiated clusters with some clusters of SARS-CoV-2 patients in the upper quartiles. However, although sex differences were apparent in the scores plot, the model was far weaker than those generated for infection status.

The most discriminatory metabolites between the two sexes were driven by higher levels of lipoprotein species (VLDL1, VLDL2), lactate, GlycA, and glucose and decreased LDL1, LDL2 and choline species in women compared to men (figure 3.11C). Similar PCA patterns were also observed in DIRE spectra (figure 3.15A), but with weaker models when using OPLS-DA ($R^2X$ 0.235, $R^2Y$ 0.562, $Q^2$ 0.0014). Again however, it is clear that SARS-CoV-2 infection exerts a far stronger influence in the inherent clustering in the PCA scores plot, indicating sex is not a substantial confound in DIRE analysis, with the metabolic consequences of infection dominating the data.

The influence of Body Mass Index (BMI) on models was also investigated due to the known association between obesity and SARS-CoV-2 morbidity, but no pattern was detected in

healthy controls due to low numbers. In SARS-COV-2, a BMI of >30 was differentiated from normal weight. PCA modelling (figure 3.12A, $R^2X$ 0.418, $Q^2Y$ 0.142) identified a weak metabolic signature for BMI when modelling standard 1D spectra, and similarly weak models were found for DIRE spectra (figure 3.16A, $R^2X$ 0.322, $Q^2Y$ -0.007) OPLS-DA models due to low numbers (figure 3.12B and 3.16B). The loadings from these models showed that BMI was not a confounding variable in the modelling of SARS-CoV-2 infection, due to the similarity of loadings to the OPLS-DA that discriminates SARS-CoV-2 and healthy.

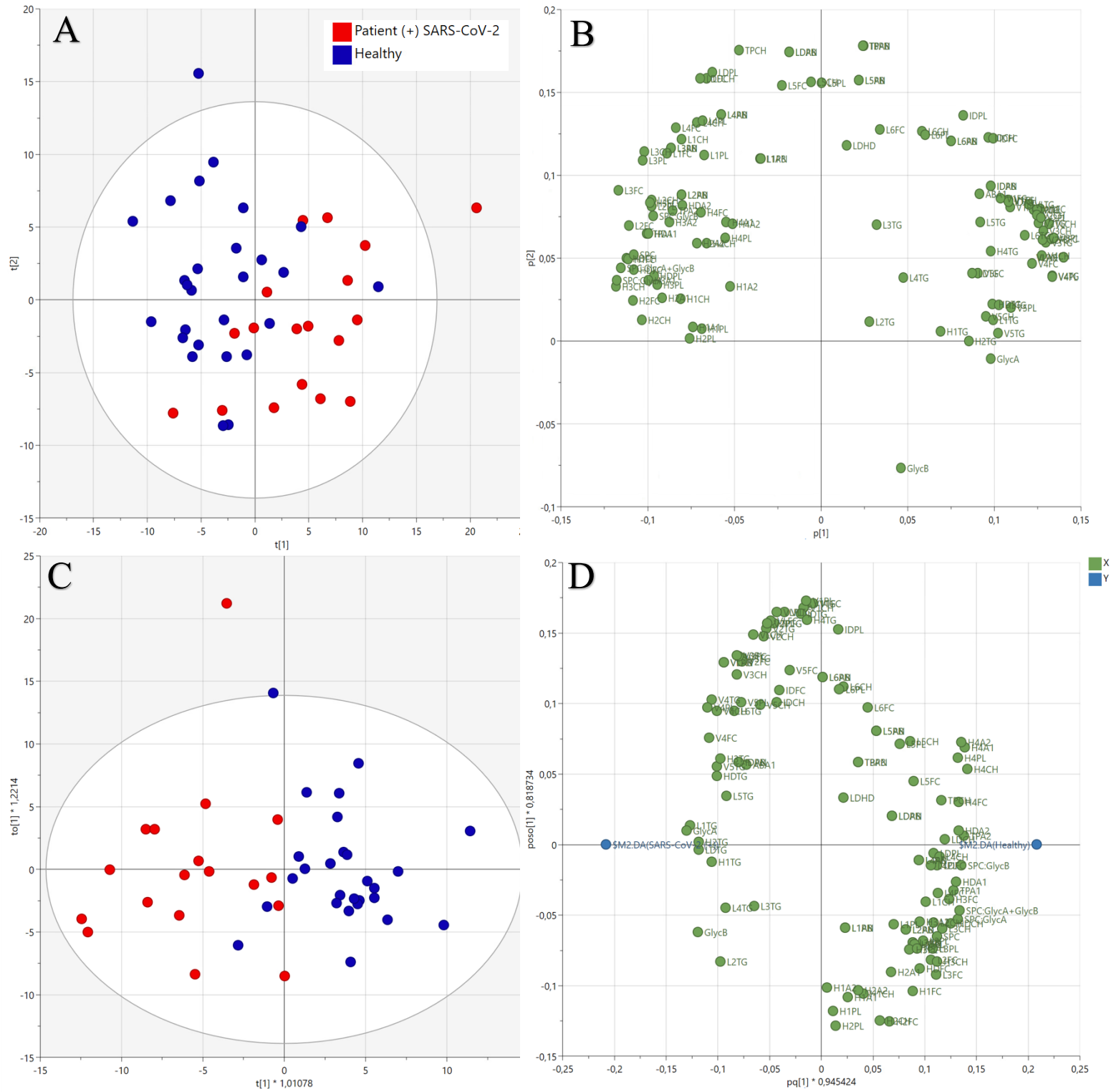



**Figure 3.17: PCA and OPLS-DA of quantified lipoprotein subfractions.** A) PCA scores plot of SARS-CoV-2 positive (+) patients and healthy control profiles ($R^2Y$ 0.987, $Q^2Y$ 0.95). B) PCA loadings plot of SARS-CoV-2 positive (+) patients and healthy control profiles. C) Scores plot of OPLS-DA model, presenting the most significant metabolites influencing the model clustering ($R^2Y$ 0.422, $Q^2Y$ 0.885). D) Loadings plot of OPLS-DA model, presenting the most significant metabolites influencing the model clustering and discrimination between SARS-CoV-2 and healthy controls.

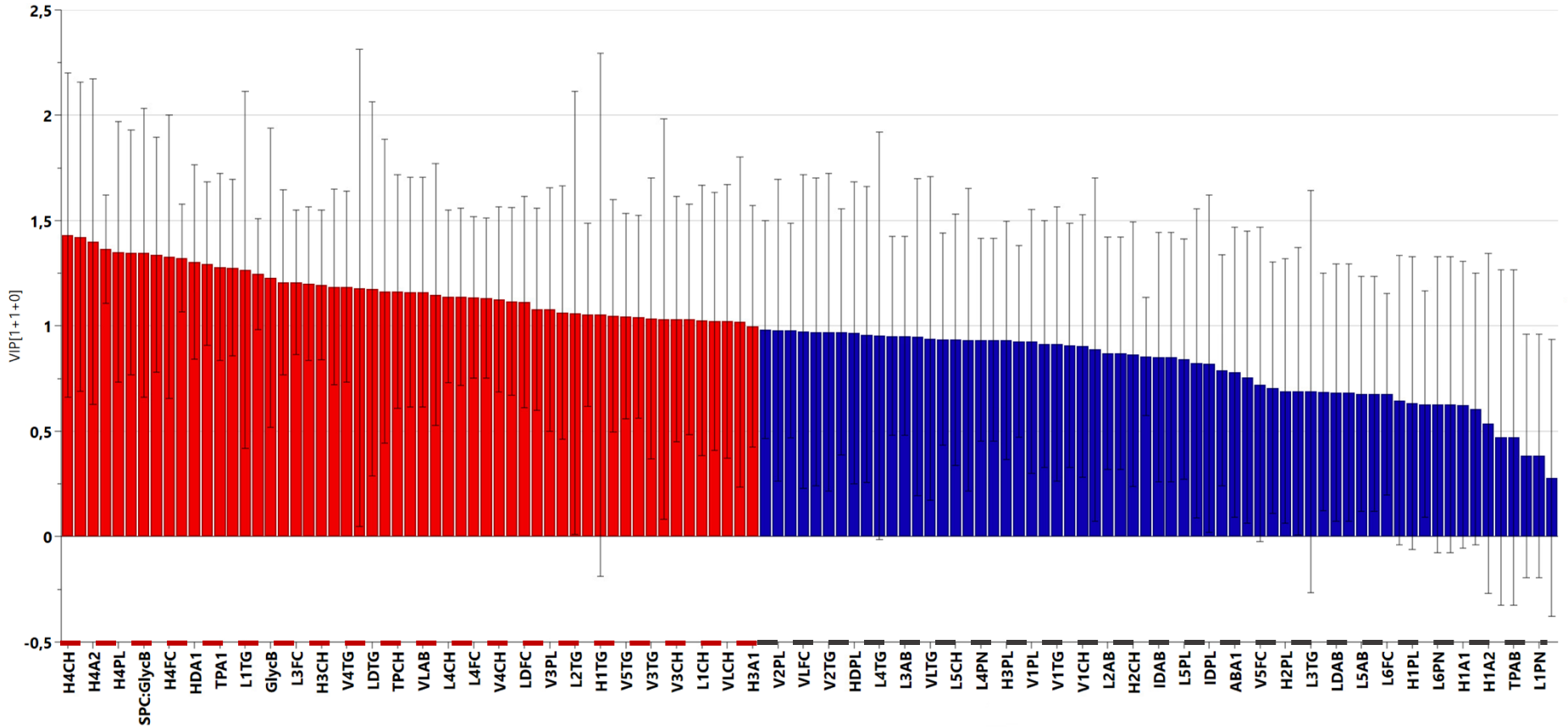



**Figure 3.18: OPLS-DA Variable importance of Projection (VIP):** Arbitrary threshold of 1 applied, showing the most discriminant metabolites in the OPLS-DA model. Underlined in red are the lipoprotein parameters the most model substantial and underlined in blue are the lipoprotein parameters with decreasing model influence.

### 3.3.4 Lipoprotein quantification and analysis

The Bruker IVDr quantification methods were applied to the NMR data as according to the standard settings. PCA scores plot (figure 3.17A, $R^2X$ 0.987, $Q^2Y$ 0.95) shows a clear clustering of healthy controls and SARS-CoV-2 patients with a certain degree of overlap, more so than standard 1D spectral acquisition, which would be expected due to the reduced resolution and reduced number of acquired datapoints in these spectral data.

OPLS-DA modelling of lipoprotein subfractions (figure 3.17B, $R^2X$ 0.422, $Q^2Y$ 0.885) showed clear separation of healthy controls and SARS-CoV-2 patients with the exception of single healthy control which sits within the SARS-CoV-2 cluster. Upon further investigation of this individual, they were found to have pre-existing type 2 diabetes and a BMI of 37.4. The OPLS-DA loadings plot (figure 3.17D) indicates a model strongly influenced by higher concentrations of several LDL subfractions in the healthy control group: LDL cholesterols (LDCH), LDL phospholipids (LDPL), LDL-free cholesterols. (LDFC), LDL apolipoprotein B, HDL cholesterol (HDCH) and phospholipids (HDPL). In comparison, in SARS-CoV-2 patients, a higher concentration of VLDL and LDL cholesterols and triglycerides were observed. The Variable Importance of Projection (VIP) plot highlights several fractions of HDLs and the SpC:GlycB ratio as the most significant components of the OPLS-DA model (figure 3.18).

### 3.3.5 Lipoprotein and Glycoprotein integration

Modelling of DIRE spectra revealed a significant signature for SARS-CoV-2 infection. The key metabolites contributing to this model were *N*-acetylglycoprotein A and B and a further peak

at $\delta$ 3.2-3.3, identified as linoleoyl-phosphatidylcholine, collectively termed the SPC signal [127]. Lysophosphatidylcholines are known to bind to a-1-acid glycoproteins, a significant contributor to the GlycA signal and thus is expected to have the same diffusion properties. Statistical correlations were carried out to investigate the significance of GlycA, GlycB and SpC on lipoprotein quantification, values were correlated, and signatures were assessed in healthy controls and compared to SARS-CoV-2 patients.

Based on correlations, another component of the SPC signal is likely to be from contribution from LDL, HDLs and phospholipids in the HDL lipoprotein fraction (figure 3.19A and 3.19B). The triglyceride signals highlighted in both DIRE OPLS-DA loadings (figure 3.14C and 3.16C), has been suggested to be an indicator of BMI, rather than infection status. Overall a strong correlation signature distinguishes SARS-CoV-2 infection from healthy controls, and is heavily influenced by the total SP, total plasma, and total HDL apolipoprotein A1 and A2 levels. This is in part due to the fact that a substantial proportion of plasma apolipoproteins are carried on HDLs, which is reduced SARS-CoV-2 infection. It is noteworthy that there is a strong correlation between total SpC and HDL3 and HDL4 and weaker correlation observed between SpC and HDL1 and HDL2 fractions (phospholipids and cholesterols), as they are decreased in SARS-CoV-2 infection. This suggests a significant proportion of the total SpC is present in the HDL3 and HDL4 fractions. Weaker positive correlation is also observed between SpC and LDL3, LDL4, LDL5 and LDL6 comparatively in SARS-CoV-2 patients. Interestingly, stronger positive correlations are seen between smaller LDL subfractions such as particle number and phospholipids in healthy controls, which could be in relation to depleted SpC in SARS-CoV-2 infection.

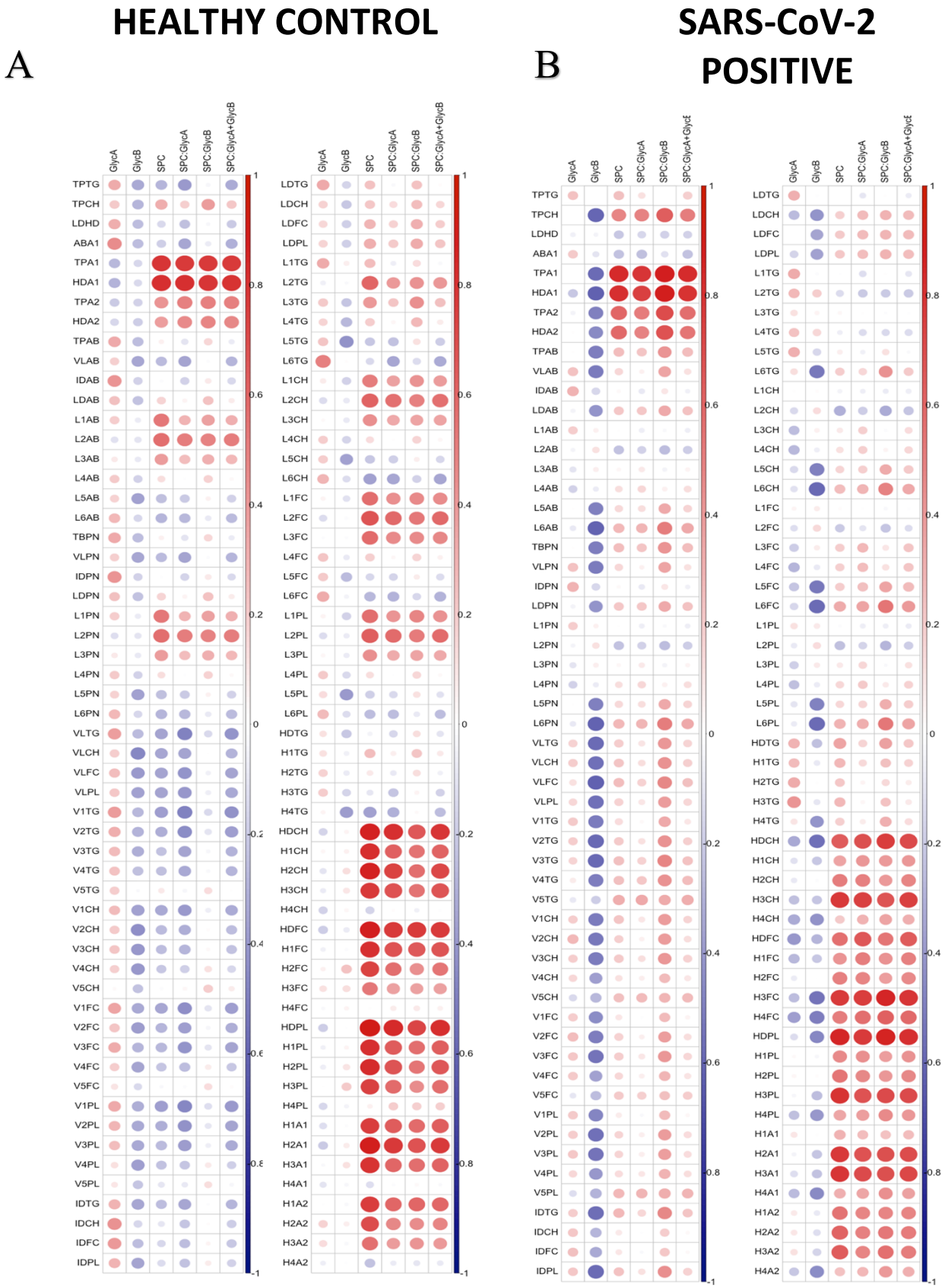



**Figure 3.19:** Quantified lipoprotein subclass correlation analysis against quantified glycoprotein A, glycoprotein B and supramolecular compound concentrations in A) healthy controls and B) SARS-CoV-2 positive patients

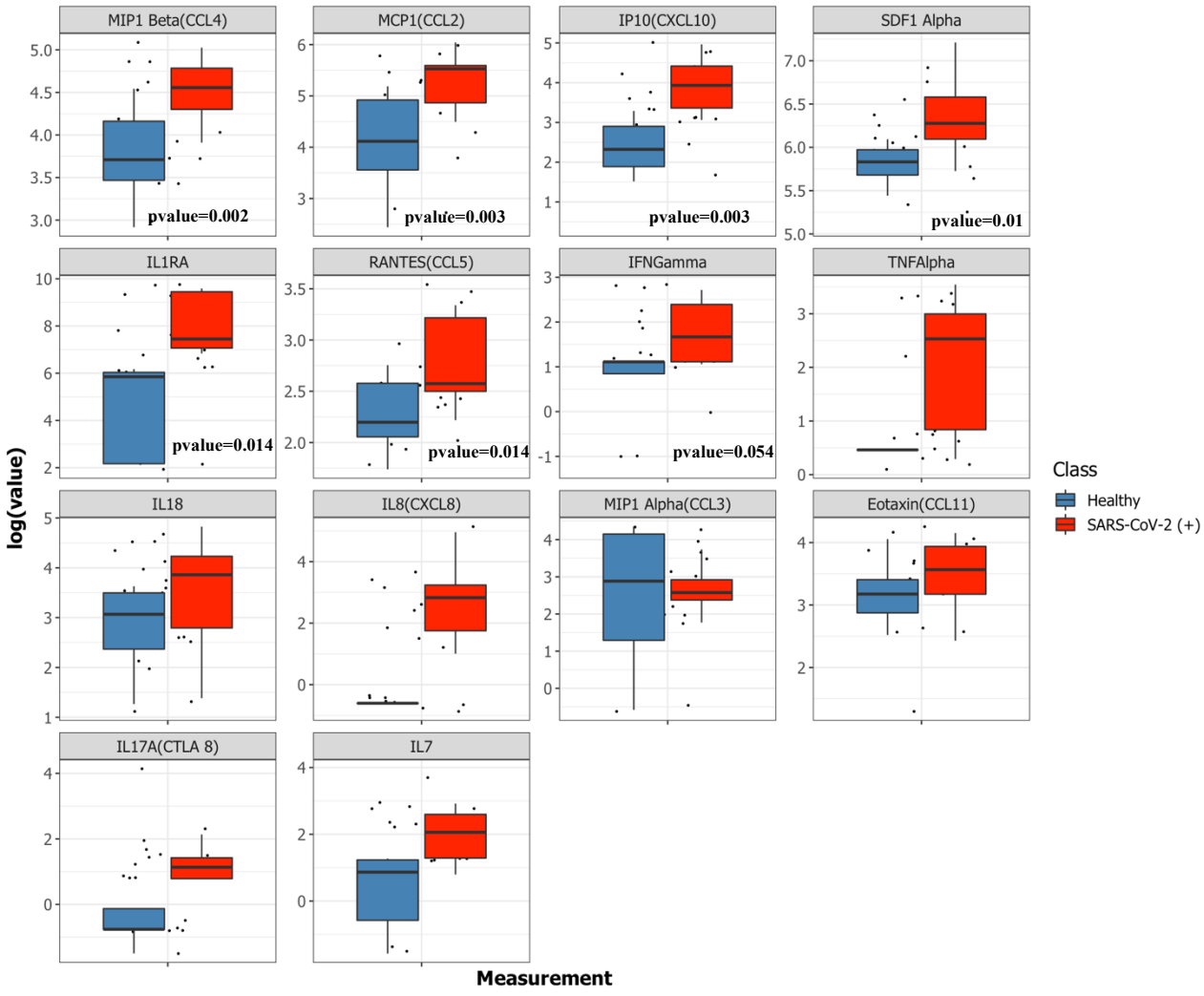



**Figure 3.20: Cytokine quantification from plasma.** 14 cytokine measurements from healthy controls and SARS-CoV-2 patients, deemed with sufficient coverage and detection from the cytokine assay. 7 of 14 were statistically significant (pvalue >0.05).

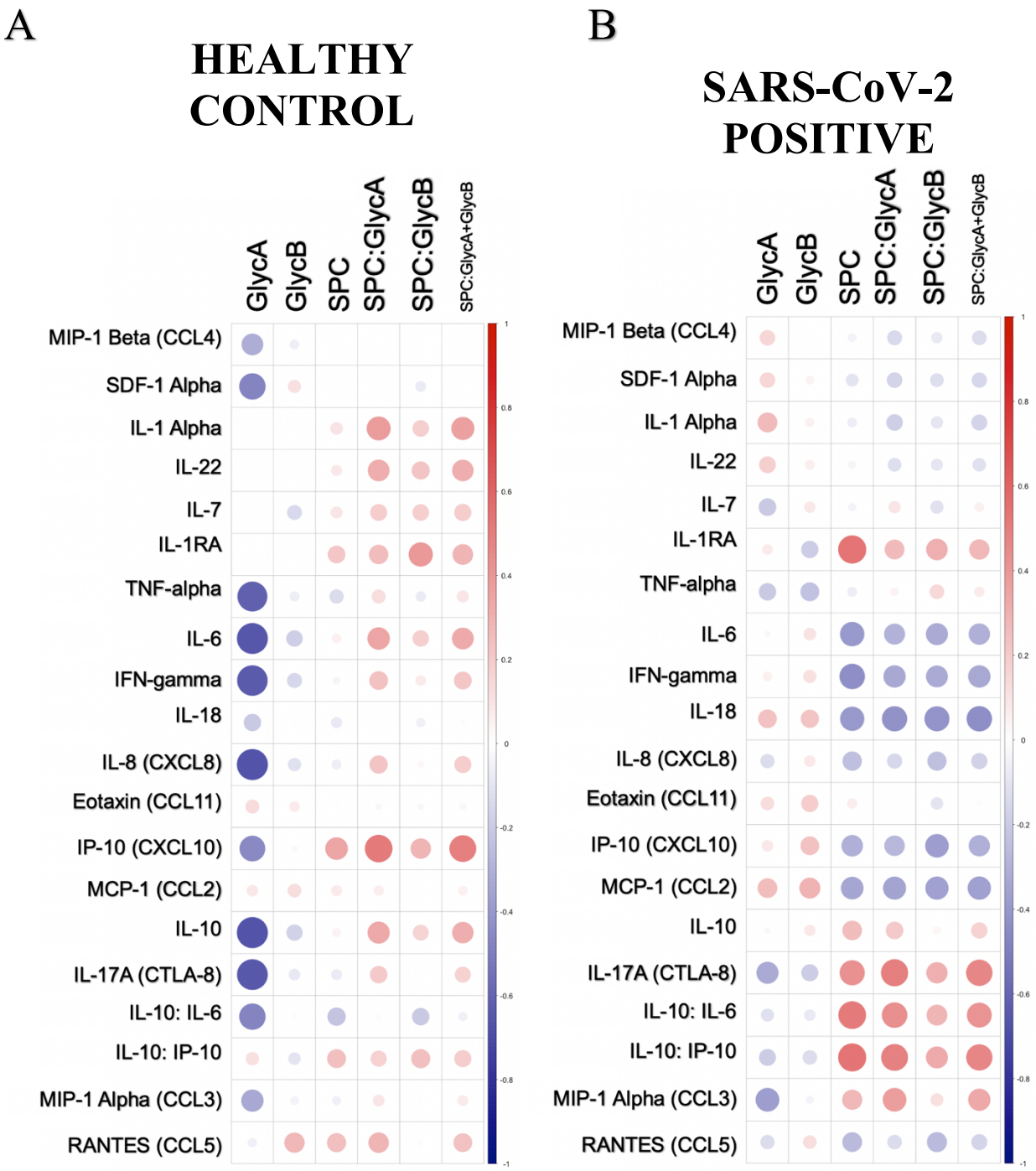



**Figure 3.21:** Hierarchically clustered immuno-metabolic spearman's correlation mapping between cytokine and chemokine measurements and quantified glycoprotein A, glycoprotein B and supramolecular compound concentrations in A) healthy controls and B) SARS-CoV-2 positive patients. Only correlation values with a p-value significance of >0.05 are shown, and blank squares indicate a non-significant result. Hierarchically clustered according to results from SARS-CoV-2 results and order applied to healthy controls for direct motif comparison.

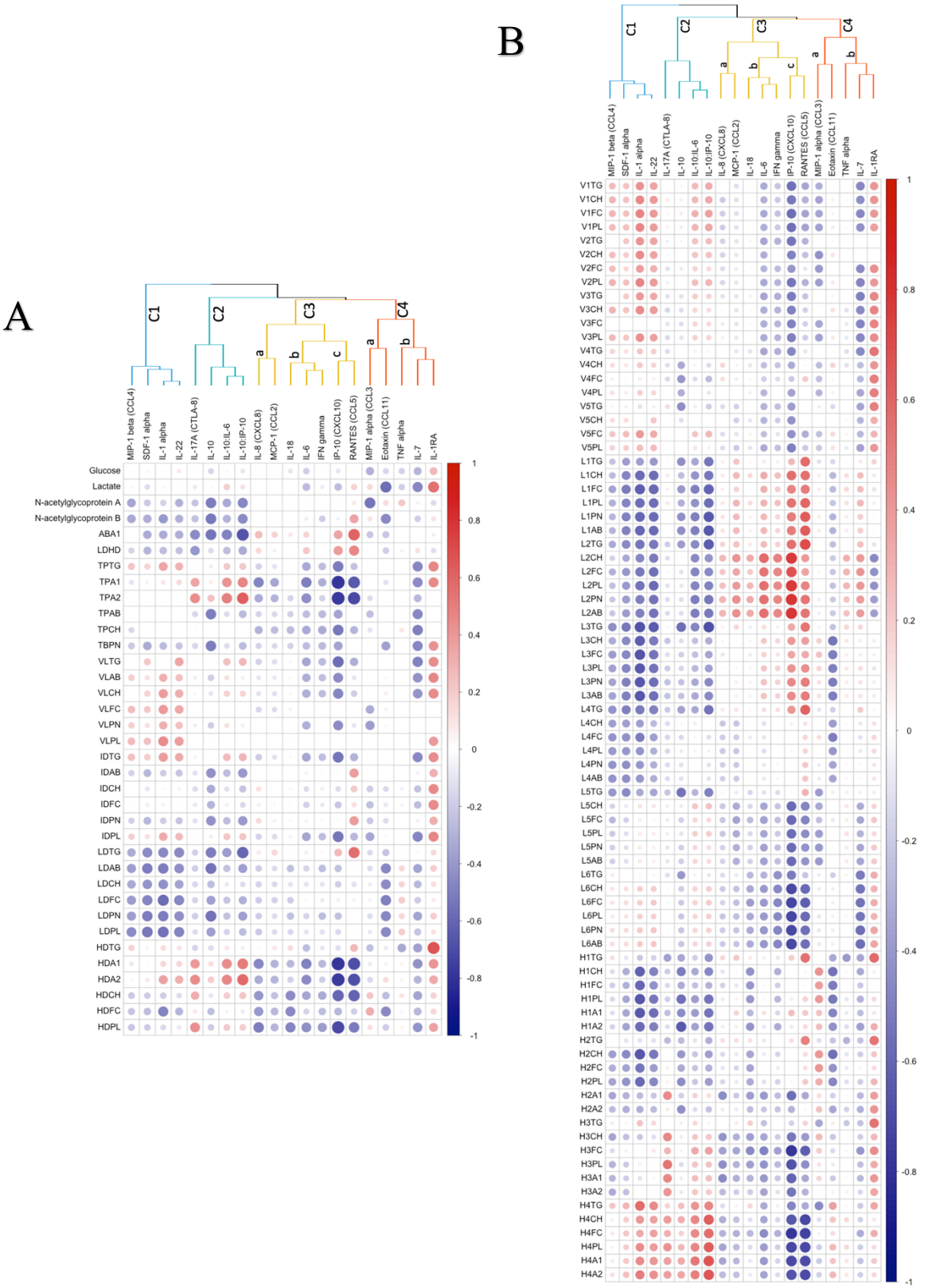



**Figure 3.22:** Correlation mapping of quantified lipoproteins against hierarchically clustered cytokines in SARS-CoV-2 patients. Ordering of lipoproteins is according to size of lipoprotein and then the size of the subfraction. Clustering is shown as 4 main branches. A) Subfractions Glucose – HDPL. B) Subfractions V1TG- H4A2. Only correlation values with a p-value significance of >0.05 are shown, and blank squares indicate a non-significant result.

### 3.3.6: Circulating cytokines and NMR metabolites

From the panel of 42 cytokines, 14 were considered to have sufficient limit of detections. Of the 14 cytokines, 7 were found to be significant (pvalue<0.05)(figure 3.20). MIP1-Beta, MCP-1, IP10, SDF1 alpha, IL1RA, RANTES and IFNγ were found to be significantly elevated in SAR-CoV-2 infection. This response was as expected due to known inflammatory response associated with SARS-CoV-2 pathology, and the ensuing "cytokine storm". However, it is of note that the majority of the patients in the cohort were mildly affected. When comparing correlation of cytokines to the detected SpC and GlycA and GlycB components (figure 3.21), weaker correlations are generally seen in healthy controls, which would be expected given the inactivated immune state of healthy controls, confirmed by the negative correlation of GlycA (an inflammatory marker) across the majority of cytokine measures. SpC shows strong correlation to IL-1RA, IL17A, IL-10:IL-6 and IL-10:IP-10 in SARS-CoV-2 infections. Generally, proinflammatory cytokines such as IL6, IFNγ were seen to anti-correlate with SpC in SARS-CoV-2 patients.

To assess the association between cytokines and lipoproteins correlation analysis was used. Figure 3.,22 presents the impact higher levels of HDL. and total apolipoprotein subfractions in healthy controls as well as the impact of elevated triglycerides, particularly LDL triglycerides and the ABA1 ratio in SARS-CoV-2 infection. Four primary clusters of cytokine-lipoprotein correlations were observed. C1 consists of MIP-1B, SDF-1a, IL-1a and IL-22, associated with reduced levels of LDL particles and increased VLDL particles. MIP-1B was the most significant discriminatory cytokine between healthy and SARS-CoV-2 infection, reflecting innate cell immune cell inflammation and is consistent with monocyte and macrophage activation.

Cluster C2 comprised of IL-17 and IL-10, and weaker ratios, generally showed weaker correlation to lipoproteins. IL-8 and MCP-1 in cluster 3a, was found to be weakly anticorrelated with the majority of lipoproteins, but not LDL1 and LDL2 subfractions. In cluster 3c, IP-10 and RANTES present the strongest correlation of all the cytokine, with the lipoprotein panel showing correlation with L1 and L2 LDL subfractions and anti-correlated with HDL, VLDL, and L6 LDL parameters. GlycA, a marker of systemic inflammation, was seen to be directly correlated to RANTES.

In cluster C3, RANTES and IP-10 are associated with inflammatory and cytotoxic T cell responses, crucial components of the innate immune system, and their correlation with GlycA suggests that the increase in these circulating cytokines is a reflection of broader systemic immune response.

Cluster 4b shows IL-7 and IL-1RA, which are significantly elevated in SARS-CoV-2 patients and TNF-α. TNF-α and IL-1RA would suggest good prognosis with reduced inflammation and increased removal of the infected cells. However, the counterintuitive function of these. cytokines are insufficient to counter the level of inflammatory cytokines released in response to the virus. IL-1RA is strongly positively correlated with HDL and VLDL lipoprotein particles. HDL is generally seen as an anti-inflammatory lipoprotein with protective effects against oxidized lipids [122] and has been shown to negatively regulate T cell activation, thus the correlation of HDL with IL-1RA supports the hypothesis that IL-1RA reduces the inflammatory cascade.

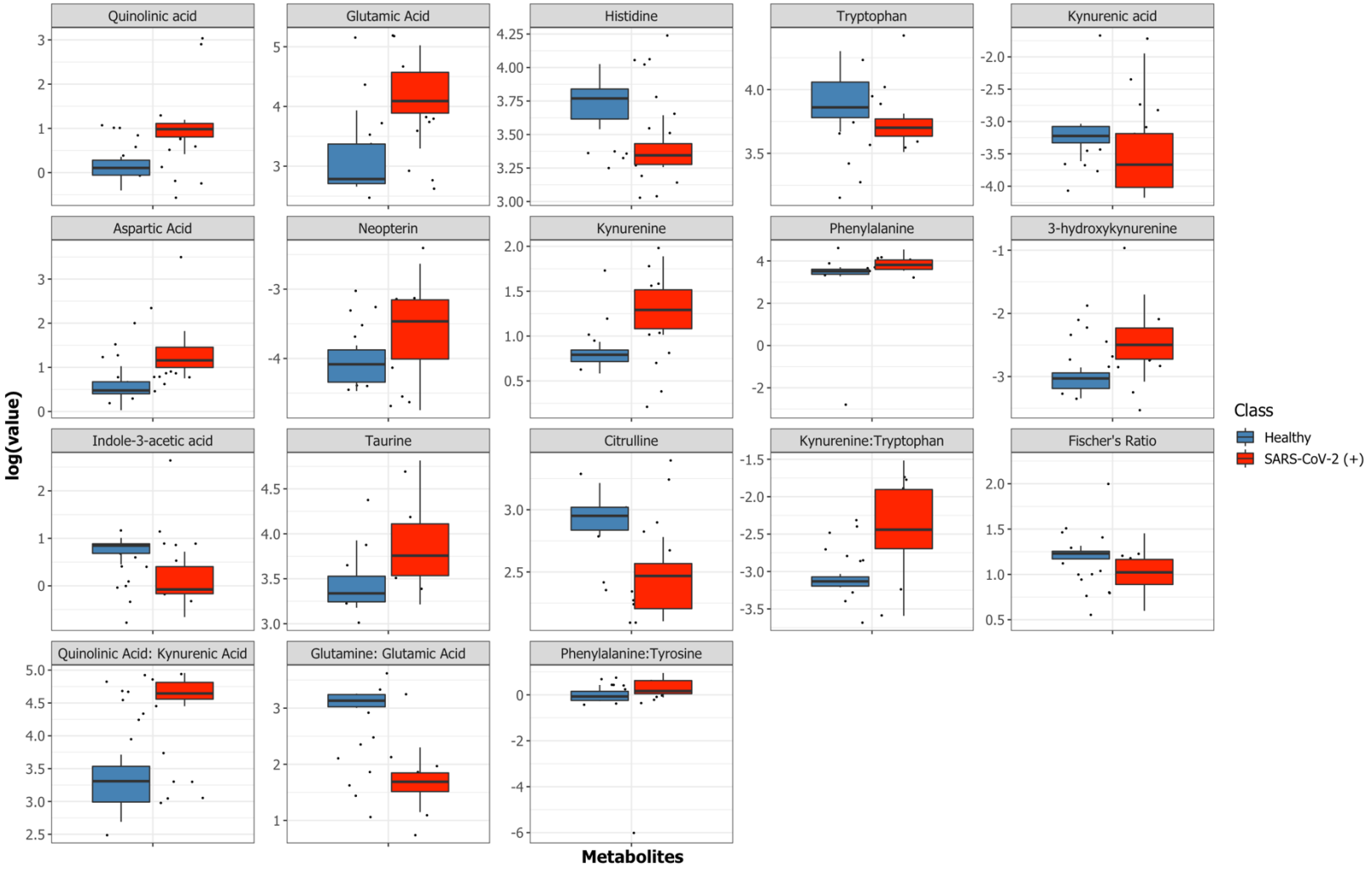



**Figure 3.23: Significant plasma metabolites and metabolite ratios between SARS-CoV-2 (+) patients and healthy controls from UHPLC-MS biogenic amine and tryptophan assays.** 13 out of 41 metabolites were found to be statistically significant between SARS-CoV-2 patients and healthy controls, and a further 6 metabolite ratios were found to be significant (Mann-Whitney, $p<0.05$).

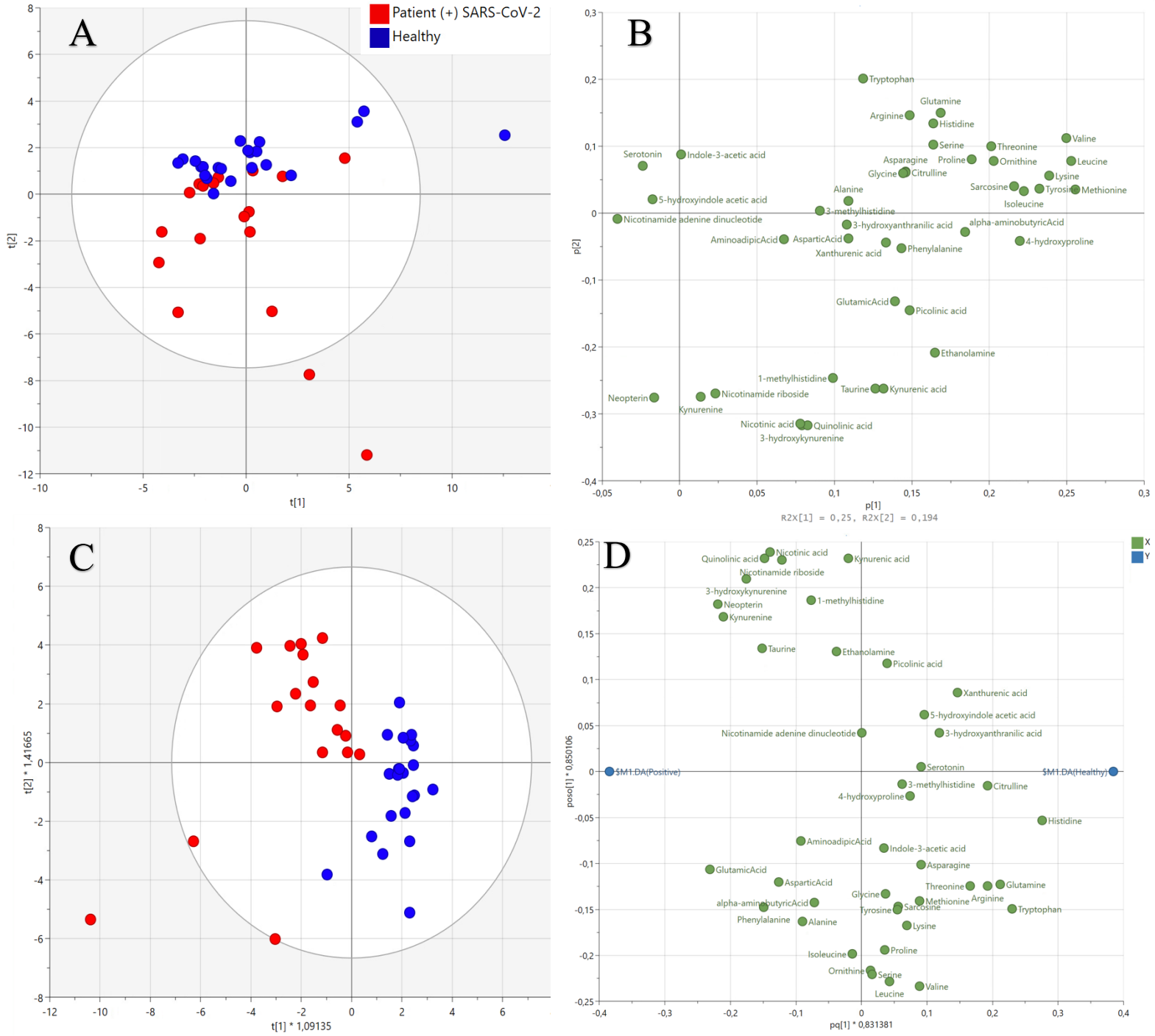



**Figure 3.24: PCA and OPLS-DA of quantified biogenic amine and tryptophan pathway metabolites.** A) PCA scores plot of SARS-CoV-2 positive (+) patients and healthy control profiles. B) Loadings plot of PCA model ($R^2X$ 0.541, $Q^2Y$ 0.239), presenting the most significant metabolites influencing the model clustering. C) OPLS-DA scores plot of SARS-CoV-2 positive (+) patients and healthy control profiles ($R^2X$ 0.471, $R^2Y$ 0.889, $Q^2Y$ 0.753) D) Loadings plot of OPLS-DA model based on class (SARS-CoV-2 positive or healthy control), presenting the most significant metabolites influencing the model clustering.

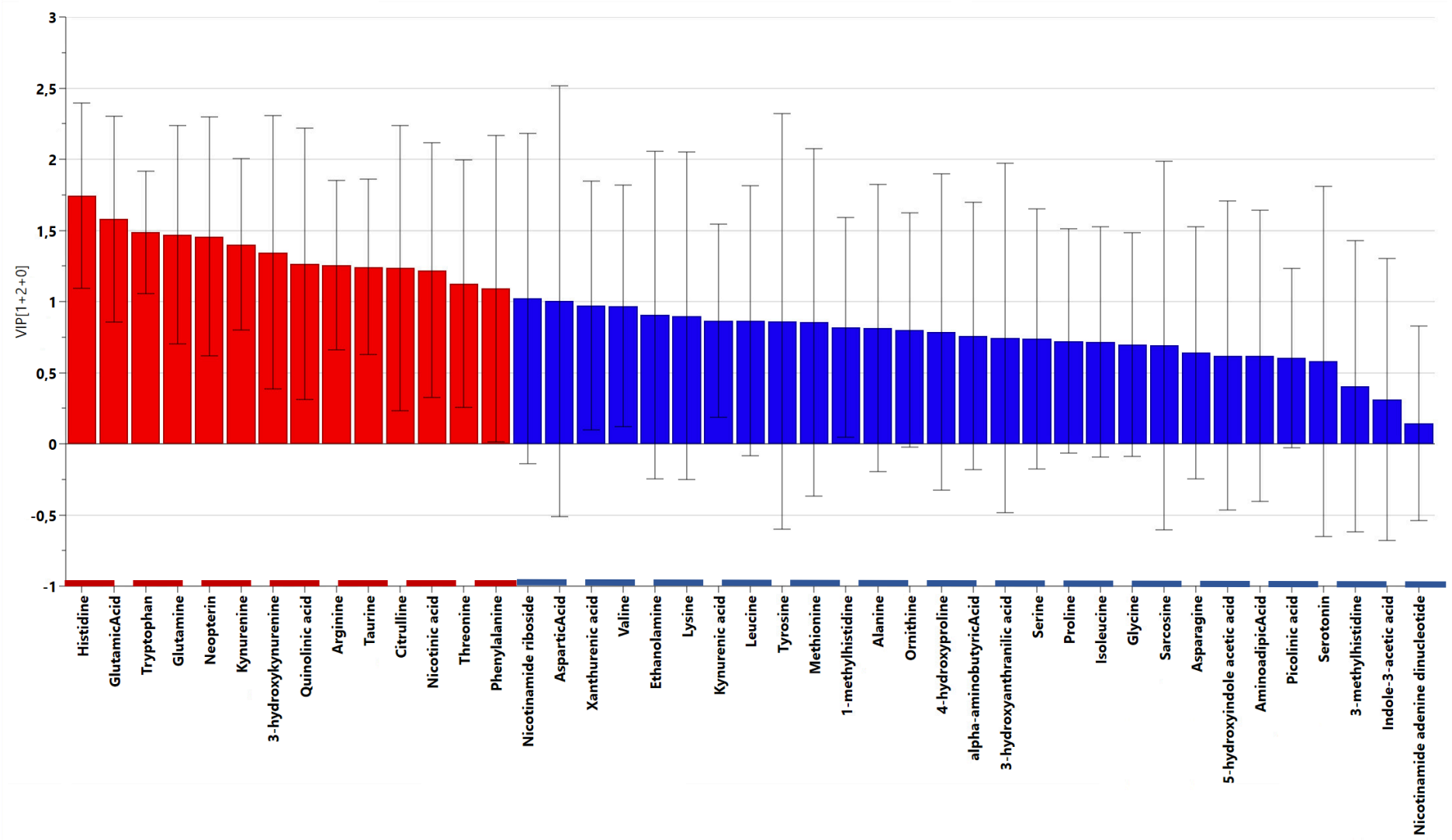



**Figure 3.25: OPLS-DA Variable importance of Projection (VIP) plot of all quantified biogenic amine and tryptophan metabolites (figure 3.24C).** Arbitrary threshold of 1 applied, showing the most discriminant metabolites in the OPLS-DA model.

### 3.3.7 Biogenic amines and tryptophan pathway metabolites in SARS-CoV-2 infection

Targeted UHPLC-MS assays were used to investigate the metabolic response of two distinct metabolic pathways (amino acids and tryptophan catabolism) in SARS-CoV-2 infection. 14 of 42 quantified metabolites from the profiling were found to be significant between SARS-CoV-2 patients and healthy controls (figure 3.23) (pvalue <0.05). Quinolinic acid, glutamic acid, nicotinic acid, aspartic acid, neopterin, kynurenine, phenylalanine, 3-hydroxykynurenine and taurine were all found to be significantly elevated in SARS-CoV-2 infection. Histidine, tryptophan, indole-3-acetic acid, and citrulline were found to be significantly depleted in SARS-CoV-2 infection. Fischer's ratio, a marker of liver function, was calculated as a ratio of branched chain amino acids (valine, leucine and isoleucine) against aromatic amino acids (tryptophan and phenylalanine) and the glutamine/glutamic acid ratio was found to be significantly (pvalue<0.05) depleted in SARS-CoV-2 infection (figure 3.23). The quinolinic/kynurenic acid ration, 3-hydroxykynrenine/kynurenic acid ratio, kynurenine/tryptophan ratio and phenylalanine ratio were all found to be significantly increased in SARS-CoV-2 patients (pvalue <0.05).

PCA analysis (figure 3.24A) ($R^2X$ 0.541, $Q^2Y$ 0.239) shows grouping of SARS-CoV-2 patients, with some overlay with healthy controls. OPLS-DA (figure 3.24C) shows similar patterns but highlights the single healthy control with elevated IgG antibodies, also seen in figure 3.10A, who clusters with the SARS-CoV-2 patients. 3 patients are strong outliers from the SARS-CoV-2 cohort and were found to be severe cases with higher concentrations of neopterin and quinolinic acid. Loadings from PCA and OPLS-DA modelling ($R^2X$ 0.471, $R^2Y$ 0.889, $Q^2Y$

0.753) present the most significant metabolites in the corresponding model. Figure 3.24B and 3.24D both emphasize increased levels of neopterin, quinolinic acid, glutamic acid, kynurenine and taurine in SARS-CoV-2 patients, whereas lower levels of tryptophan and histidine are responsible for the differential clustering of groups, also further highlighted in the variable importance of project plot (figure 3.25).

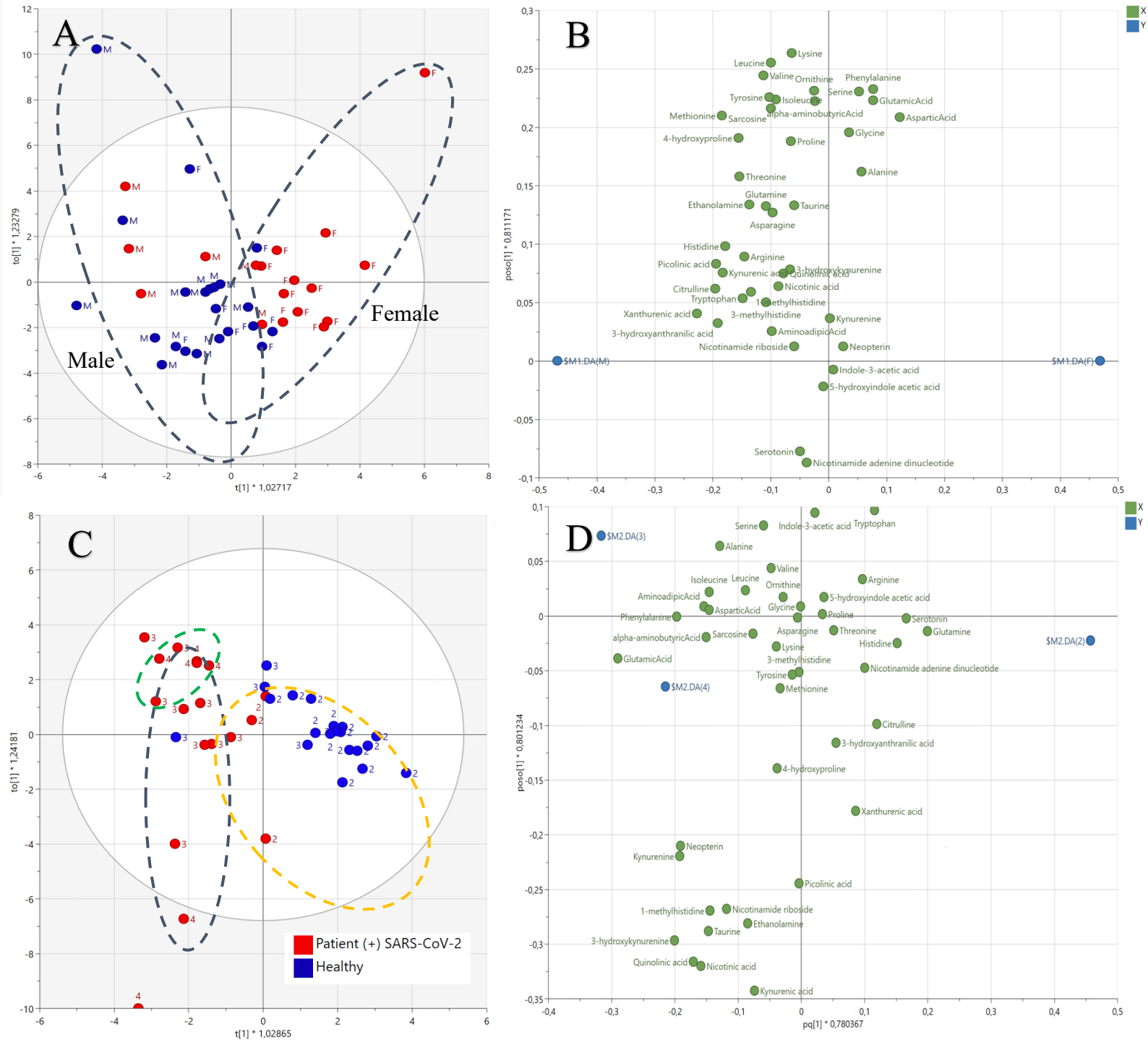



**Figure 3.26: OPLS-DA models and respective loadings plot all quantified biogenic amine and tryptophan metabolites discriminating healthy controls and SARS based on sex and BMI differences, but coloured according to infection status.** A) OPLS-DA scores plot of SARS-CoV-2 positive (+) patients and healthy control profiles based on sex classification ($R^2X$ 0.332, $R^2Y$ 0.473 $Q^2Y$ 0.24). B) Loadings plot of OPLS-DA model based on sex differences, presenting the most significant metabolites influencing the model clustering. C) OPLS-DA scores plot of SARS-CoV-2 positive (+) patients and healthy control profiles based on BMI classification (2= 18.5-24.9, 3= 25.2-29.9, 4= >30) ($R^2X$ 0.443, $R^2Y$ 0.424 $Q^2Y$ 0.126). BMI was not available for all patients and those datapoints were removed from modelling.D)Loadings plot of OPLS-DA model based on BMI differences, presenting the most significant metabolites influencing the model clustering.

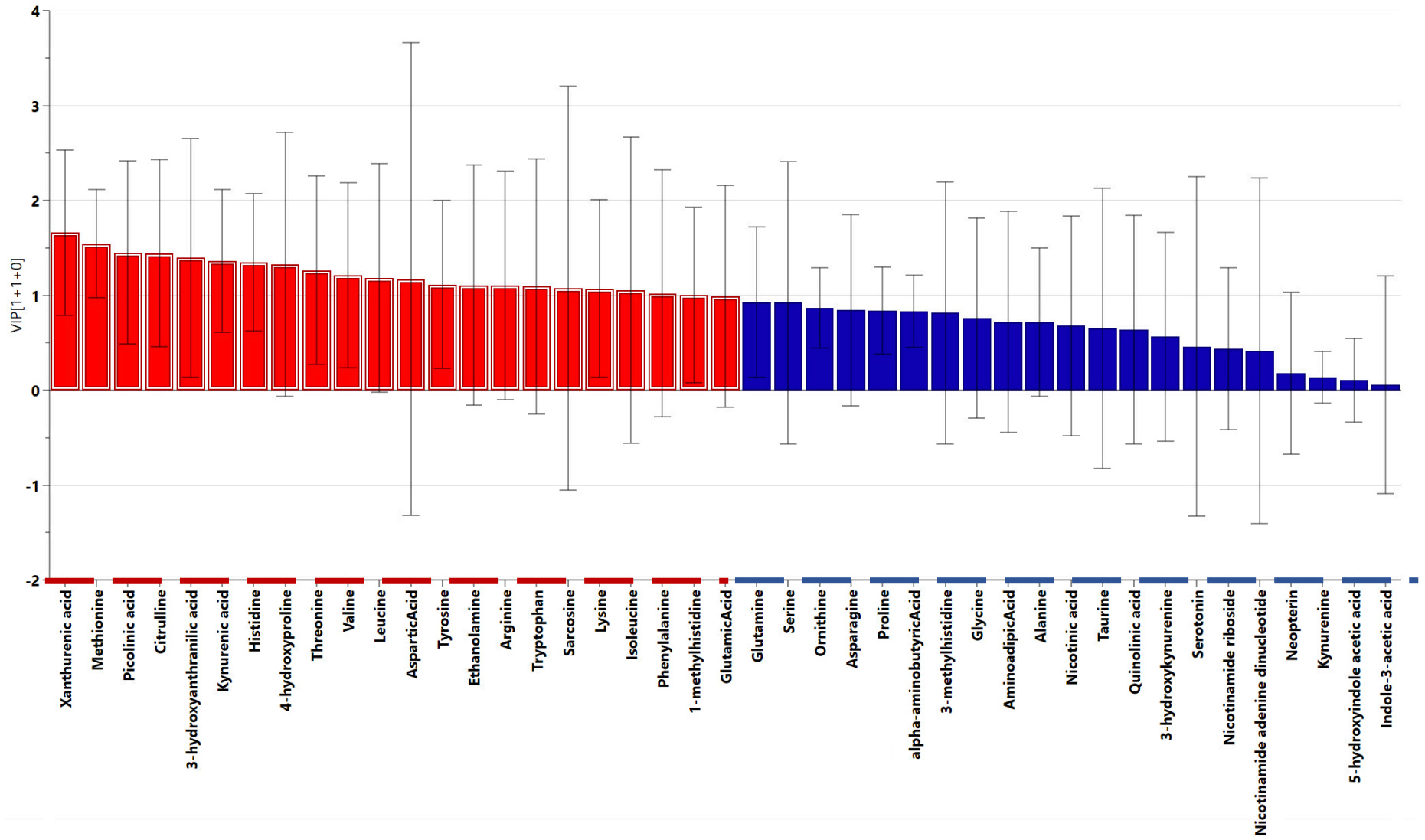



**Figure 3.27: OPLS-DA model based on sex differences- Variable importance of Projection (VIP) of all quantified biogenic amine and tryptophan metabolites (figure 21A).** Arbitrary threshold of 1 applied, showing the most discriminant metabolites in the OPLS-DA model.

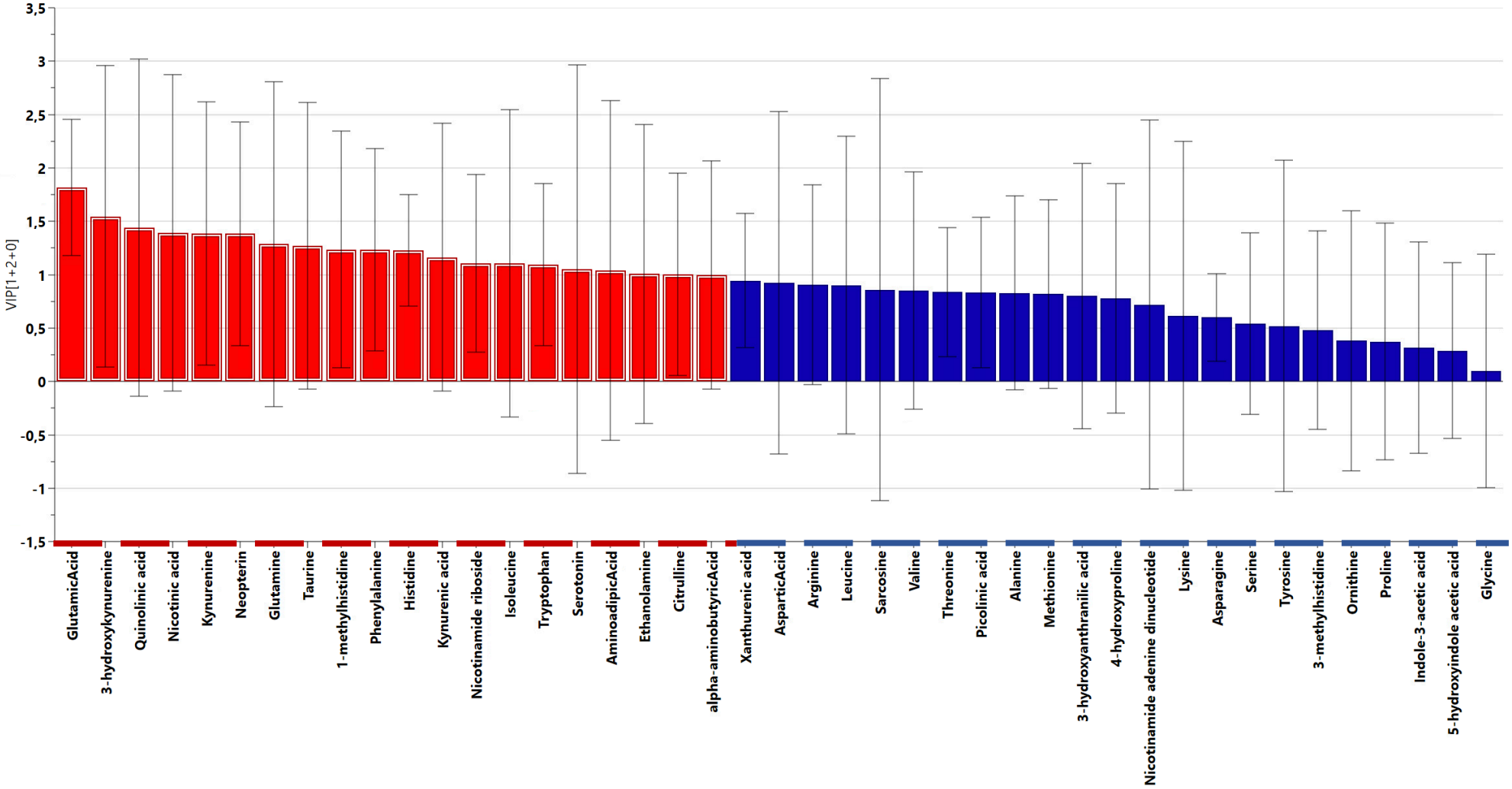



**Figure 3.28: OPLS-DA model based on body mass index (BMI) - Variable importance of Projection (VIP) of all quantified biogenic amine and tryptophan metabolites (figure 21C).** Arbitrary threshold of 1 applied, showing the most discriminant metabolites in the OPLS-DA model.

| Metabolite | *SARS-CoV-2 Positive* | | *Sex differences* | *Body Mass Index* |
|---|---|---|---|---|
| 1-methylhistidine | ↑ | | ✓ | ✓ |
| 3-methylhistidine | ↓ | | | |
| 4-hydroxyproline | ↑ | | ✓ | |
| Alanine | ↑ | | | |
| alpha-aminobutyricAcid | ↑ | | | ✓ |
| AminoadipicAcid | ↑ | | | ✓ |
| Arginine | ↓ | ✓ | ✓ | |
| Asparagine | ↓ | | | |
| AsparticAcid | ↑ | | ✓ | |
| Citrulline | ↓ | ✓ | ✓ | ✓ |
| Ethanolamine | ↑ | | ✓ | ✓ |
| GlutamicAcid | ↑ | ✓ | ✓ | ✓ |
| Glutamine | ↑ | ✓ | | ✓ |
| Glycine | ↑ | | | |
| Histidine | ↓ | ✓ | ✓ | ✓ |
| Isoleucine | ↓ | | ✓ | ✓ |
| Leucine | ↑ | | ✓ | |
| Lysine | ↑ | | ✓ | |
| Methionine | ↓ | | ✓ | |
| Ornithine | ↑ | | | |
| Phenylalanine | ↑ | ✓ | ✓ | ✓ |
| Proline | ↑ | | | |
| Sarcosine | ↑ | | ✓ | |
| Serine | ↓ | | | |
| Taurine | ↑ | ✓ | | ✓ |
| Threonine | ↑ | ✓ | ✓ | |
| Tyrosine | ↓ | | ✓ | |
| Valine | ↓ | | ✓ | |
| Tryptophan | ↓ | ✓ | ✓ | ✓ |
| 3-hydroxykynurenine | ↑ | ✓ | | ✓ |
| 3-hydroxyanthranilic acid | ↓ | | ✓ | |
| Kynurenic acid | ↑ | | ✓ | ✓ |
| Nicotinamide riboside | ↑ | | | ✓ |
| Quinolinic acid | ↑ | ✓ | | ✓ |
| Nicotinic acid | ↑ | ✓ | | ✓ |
| Picolinic acid | ↓ | | ✓ | |
| Xanthurenic acid | ↓ | | ✓ | |
| Kynurenine | ↑ | ✓ | | |
| Indole-3-acetic acid | ↓ | | | |
| Serotonin | ↓ | | | ✓ |
| 5-hydroxyindole acetic acid | ↑ | | | |
| Nicotinamide adenine dinucleotide | ↓ | | | |
| Neopterin | ↑ | ✓ | | ✓ |

**Table 3.3: Comparisons between variable importance loadings from OPLS-DA models of biogenic amines and tryptophan pathway metabolites.** Each column refers to a specific OPLS-DA model and whether the metabolite is specific in the VIP analysis. SARS-CoV-2 positive column also notes if the metabolite is elevated in the disease cohort in comparison to healthy controls. Arbitrary threshold of 1 applied to VIP analysis.

### 3.3.8 Influence of sex differences and BMI on biogenic amines and tryptophan pathway metabolite modelling

To investigate the influence of sex and BMI on statistical modelling, OPLS-DA was conducted. OPLS-DA (figure 3.26A) scores plot of SARS-CoV-2 positive (+) patients and healthy control profiles based on sex classification, demonstrates strong clustering of sexes within groups ($R^2X$ 0.332, $R^2Y$ 0.473 $Q^2Y$ 0.24). Loadings plot (figure 3.26B) of OPLS-DA model based on sex differences, presenting the most significant metabolites influencing the model clustering highlighting higher levels of xanthurenic acid, methionine, 3-hydroanthranillic acid, picolinic acid and men and higher levels of phenylalanine, glutamic acid and aspartic acid in women. VIP plot (figure 3.27) states the most substantial metabolite driving the separation between groups is xanthurenic acid, methionine and picolinic acid. OPLS-DA modelling (figure 3.26C) of SARS-CoV-2 positive (+) patients and healthy control profiles based on BMI classification (2= 18.5-24.9, 3= 25.2-29.9, 4= >30) ($R^2X$ 0.443, $R^2Y$ 0.424 $Q^2Y$ 0.126), presenting grouping of BMIs. Due to generally higher BMIs in SARS-CoV-2 infection, poor clustering according to BMI within groups is observed. The most substantial contributors to this model (figure 3.26D and figure 3.28) are quinolinic acid, kynurenic acid, 3-hydroxykynurenine and picolinic acid, all higher in the group 3 and 4 categories.

Table 3.3 shows a comparison of all 3 OPLS-DA and the most substantial contributors to the OPLS-DA models. Briefly, sex differences are generally more influenced by metabolites involved in muscle catabolism and branched chain amino acids, whereas BMI is more greatly influenced by metabolites involved in inflammatory pathways.

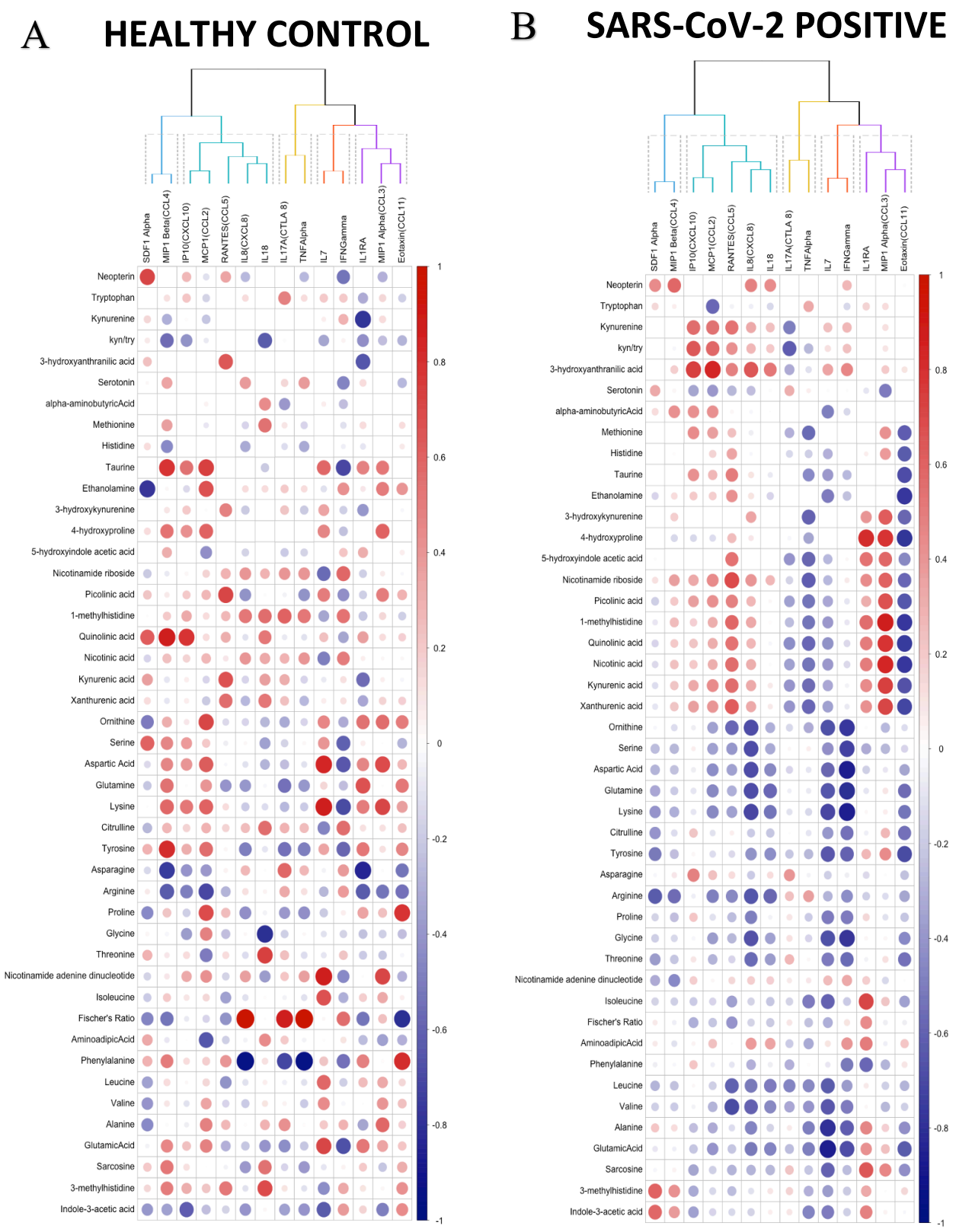



**Figure 3.29: Hierarchically clustered Spearman's correlation of biogenic amines and tryptophan pathway metabolites against measured cytokines.** Correlation analysis of healthy controls (A) and SARS-CoV-2 positive (B) patients shown separated for direct motif comparison of metabolic signatures. Only correlation values with a p-value significance of >0.05 are shown, and blank squares indicate a non-significant result.

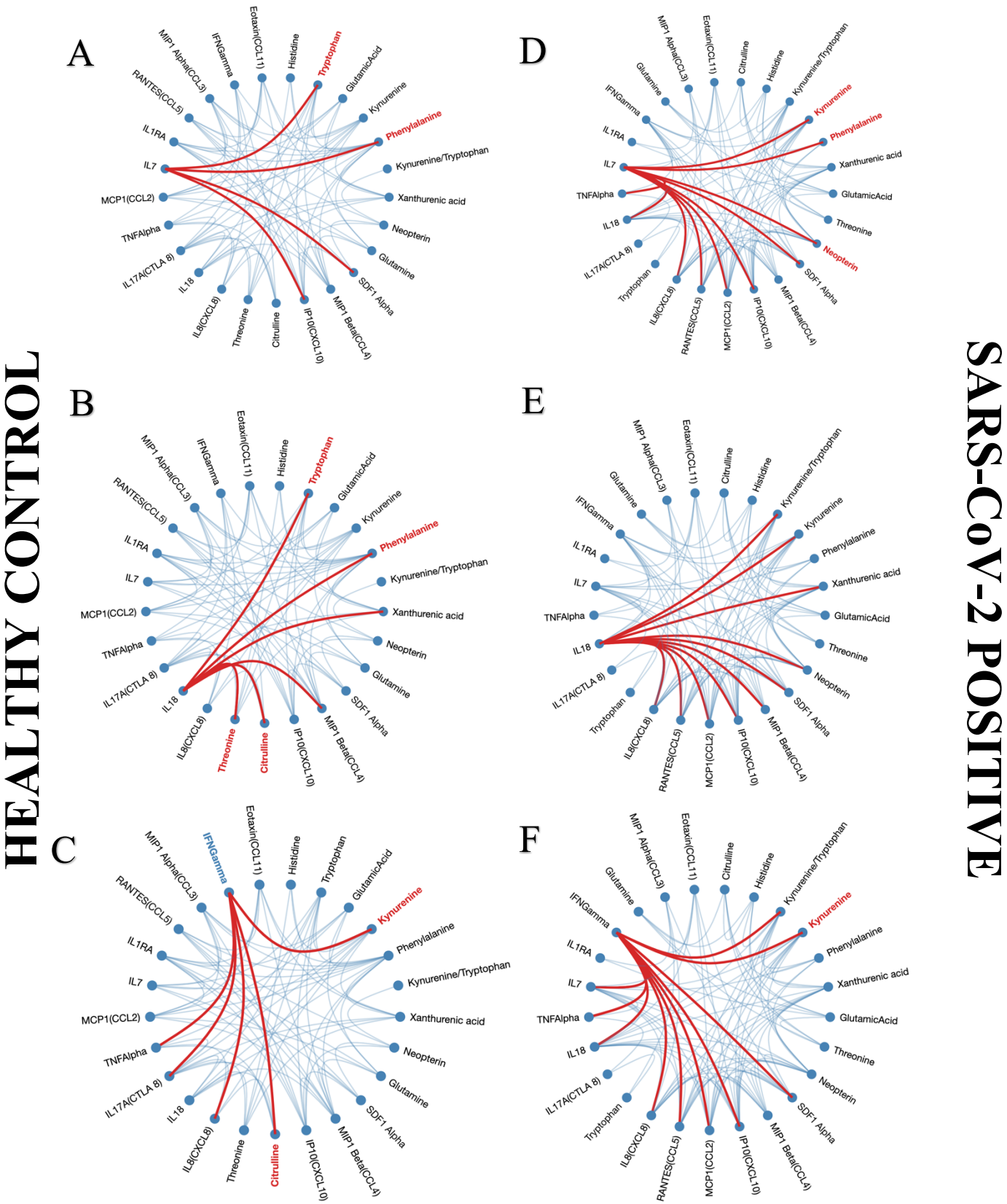



**Figure 3.30: Hierarchical edge clustering of significant metabolites (pvalue >0.05) and array of cytokines.** Red lines states positive correlation between cytokine and metabolite, whereas blue line states negative correlation. Hierarchical maps for healthy controls (A-C) and SARS-CoV-2 patients (D-F).

### 3.3.9 Correlation of cytokines and biogenic amines and tryptophan pathway metabolites

To evaluate the relationship between cytokines and these two distinct metabolic pathways, correlation analysis was conducted. Figure 3.29 shows two distinct correlation maps between healthy controls and SARS-CoV-2 patients. Eotaxin appears to correlate inversely with metabolites in comparison to MIP1-alpha and IL1RA in SARS-CoV-2 patients (figure 24B). Eotaxin is strongly anticorrelated methionine, taurine, ethanolamine and a significant number of tryptophan pathway metabolites. The same tryptophan pathway metabolites are found to be strongly positively correlated with MIP-1 alpha, IL1RA and RANTES. Strong anticorrelation is observed between IL8, IL7 and IFNgamma with a host of amino acids from cluster 3, 4 and 5 along the y axis (figure 3.29B). Interestingly kynurenine and 3-hydroxyanthranillic acid are strongly positively correlated with IP10, MCP1, RANTES, IL18, and IL8 in SARS-CoV-2 infection. It is noteworthy that a more sporadic and generally weaker correlation map is observed in healthy controls (figure 3.29A).

To assess directionality of the association between cytokines and significant metabolites (pvalue <0.05) between healthy controls and SARS-CoV-2 patients, hierarchical edge clustering maps (figure 3.30) were produced, specifically focusing on IL7, IL18 and IFNγ as the cytokines with the highest correlation with the reduced set of significant metabolites, and motifs were compared. IL7 in healthy controls (figure 3.30A) was significantly clustered with metabolites tryptophan (depleted in SARS-CoV-2 infection) and phenylalanine (elevated in SARS-CoV-2 infection), and cytokines IP10 and SDF1 alpha. In SARS-CoV-2 infection (figure 3.30D), IL7 correlates with kynurenine, phenylalanine, neopterin, in addition to an array of proinflammatory cytokines such as MCP1 and IP10. IL-18 in healthy

controls (figure 3.30B) correlates with metabolites tryptophan, phenylalanine, xanthurenic acid, citrulline and threonine and cytokine MIP1 beta. In SARS-CoV-2 infection (figure 3.30E) IL18 correlates with the kynurenine:tryptophan ratio, kynurenine, xanthurenic acid, and neopterin, in addition to the same array of cytokines in figure 3.30D. Pro-inflammatory IFNγ in healthy controls (figure 3.30C) correlates with kynurenine, citrulline and cytokines IL8, IL17A and TNFalpha. In comparison, in SARS-CoV-2 patients (figure 3.30F), IFNγ correlates with metabolites kynurenine and the kynurenine:tryptophan ratio, in addition to the same array of pro-inflammatory cytokines seen in figures 3.30D and 3.30E, showing a strong pro-inflammatory component in the SARS-CoV-2 signature in comparison to the normal immune state of healthy controls.

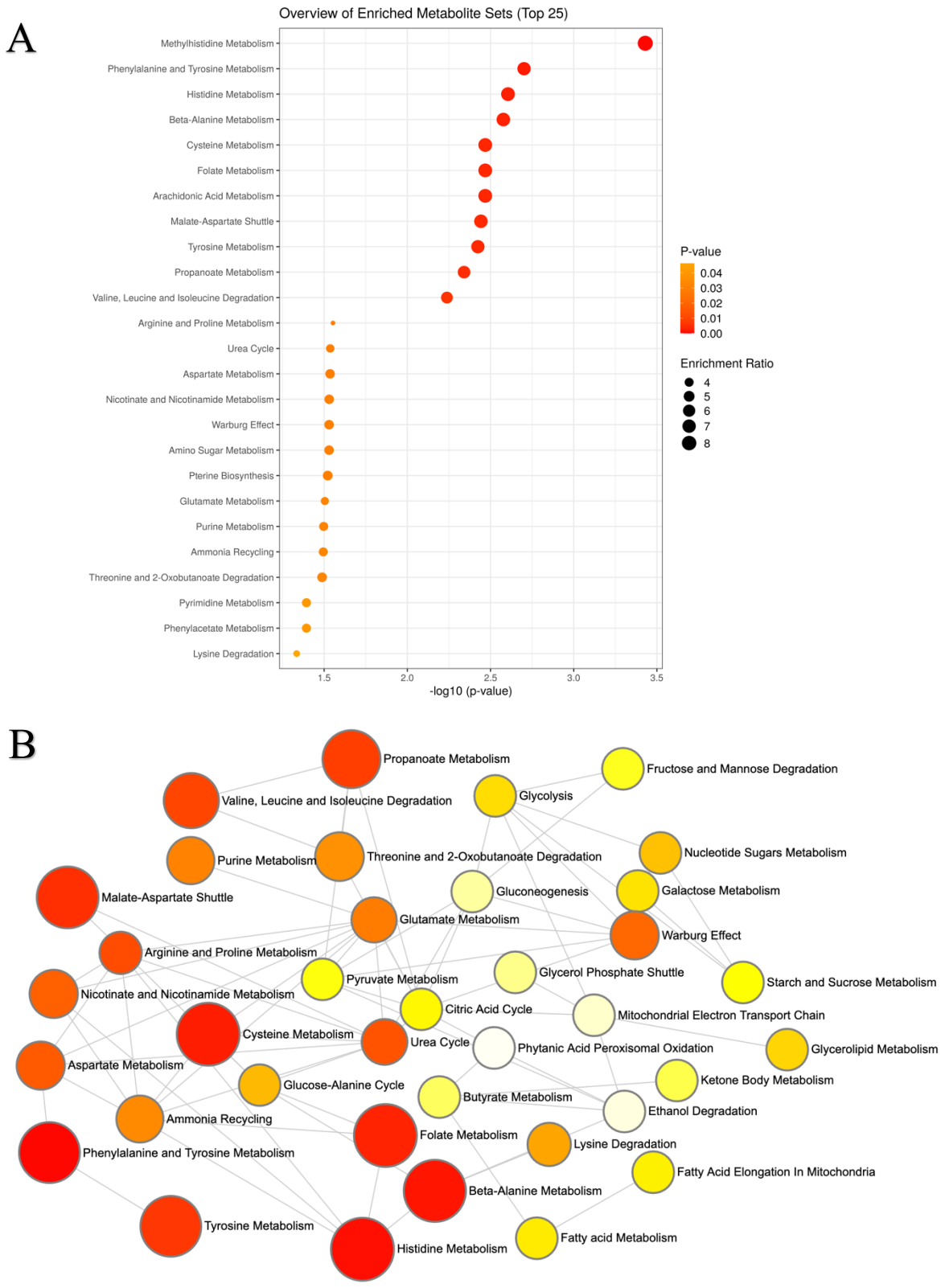



**Figure 3.31: Metabolite set enrichment analysis (MSEA) and network analysis.** A) Associated pathways and networks associated with the abundance of metabolites in SARS-CoV-2 patients compared to healthy controls. B) Connections between each pathway network and associations related to MSEA of metabolite abundance and proportional significance between SARS-CoV-2 patients and healthy controls.

### 3.3.10 Metabolite set enrichment analysis

Metabolite set enrichment analysis (figure 3.31A) was carried out of the full metabolite set to assess the relationship between metabolite concentration, differences between groups and pathway mapping. The most significant pathways associated with differences in SARS-CoV-2 infection and healthy controls are methylhistidine and histidine metabolism, associated with significantly depleted histidine and Fischer's ratio in SARS-CoV-2 infection, phenylalanine and tyrosine metabolism, associated with a significantly increased concentration of phenylalanine and phenylalanine to tyrosine ration. Figure 3.31B visualises the mapping and importance of each pathway within the metabolite set enrichment, with many liver metabolism associated pathways strongly correlated.

## 3.4 Discussion

### 3.4.1 Impact of SARS-CoV-2 infection on clinical pathology markers

The clinical course and severity of SARS-CoV-2 infection can be classified by specific biochemical alterations. Findings from previous studies in community acquired pneumonia (CAP) and chronic obstructive pulmonary disease, have suggested that D-dimer is elevated in more severe cases in these lung diseases [123]. For those with severe disease phenotypes, predictors of mortality have been found to be elevated blood urea nitrogen, C-reactive protein (CRP), blood creatinine [124], all elevated parameters we see within this cohort (figure 3.4). Elevations in D-dimer and liver function tests (albumin, bilirubin, AST, ALT, ALP) in SARS-CoV-2 infection have also been reported to be associated with poorer outcomes, but within this hospitalised cohort, D-dimer and liver function tests are within the upper limit of normal (ULN) or depleted (figure 3.4) suggesting a milder disease phenotype. During initial inflammatory stages, IL-6 induces the liver to produce a significant number of acute phase proteins, including CRP, serum amyloid A (SAA), fibrinogen, haptoglobin, and a1-antitrypsin [125]. The results from biochemical testing in figure 3.4, suggests a heterogenous cohort, with some pathology signatures more consistent with mild disease, as opposed to severe disease [126].

### 3.4.2 SARS-CoV-2 patient plasma NMR metabolic signatures

Statistical modelling of 1D (figure 3.10), and CPMG (figure 3.13) spectral NMR data indicated that the SARS-CoV-2 virus elicited the detection of a specific signature for SARS-CoV-2

patients in comparison to healthy counterparts. Similar patterns from PCA score plots of standard 1D, CPMG and DIRE spectral data were observed, and presented two clear group clusters, with some overlap of SARS-CoV-2 patients with healthy controls. Following OPLS-DA modelling, one healthy control repeatedly clustered with SARS-CoV-2 patients and was found to be positive for SARS-CoV-2 antibodies, supporting the robustness of the modelling.

OPLS-DA modelling of CPMG spectra data revealed a key discriminatory metabolite to drive differences between SARS-CoV-2 patients healthy controls were attributed to a-1-*N*-actyl glycoprotein signals, glycoprotein A (GlycA) and glycoprotein B (GlycB) found to be significantly elevated, and another signal termed the Supramolecular Phospholipid Composite (SpC) found to be depleted in the inflammatory response in SARS-CoV-2 infection reported by Lodge et al [127], which results in the increased sensitivity of GlycA:GlycB and SpC components in relation to the disease signature. GlycA is a composite signal of *N*-acetyl signals from 5 proteins: a-1-acid-glycoprotein, a-1-antichymotrypsin, a-1- antitrypsin, haptoglobin and transferrin. The a-1-acid-glycoprotein is the strongest component of the GlycA signal, and weighs with the highest correlation, although inter-individual differences have been reported with regards to the weighting of each individual protein in the signal [128]. Several biological functions have attributed to the a-1-acid glycoprotein, such as immune-modulation via a macrophage-release inhibitory factor that prevents IL1 activation of thymocyte and lymphocyte proliferation [129, 130] and inhibition of platelet aggregation [131]. Acute phase inflammation has been associated with 2-5-fold increases in plasma GlycA signals [132], and GlycB has been found to increase in inflammatory conditions such as diabetes and obesity [133]. GlycA and GlycB have been found to correlated with CRP plasma

levels; the primary clinical biomarker for inflammation and have been suggested to be more accurate measures of inflammation [133].

In addition to elevated GlycA and GlycB, higher concentrations of alanine, lactate, and pyruvate were detected in SARS-CoV-2 patients in comparison to healthy controls (figures 3.3 and 3.11C). The lactate:pyruvate ratio was found to be significantly increased in SARS-CoV-2 infection compared to healthy controls (pvalue <0.0001), and has been found to be an indicator of impaired mitochondrial function, specifically respiration as it is indirectly associated with the NADH:NAD+ redox state in the cytoplasm. Plasma glycoproteins are primarily synthesized and released systemically in the liver, where they are glycosylated and assist with the solubilization of hydrophobic compounds in blood [134]. *N*-acetyl signals from glycosylated amino sugar residues such as a-1-*N*-actyl glycoprotein have been reported to be elevated in the acute phase of inflammatory diseases such as obesity, diabetes, cardiovascular disease and systemic autoimmune disease [128, 135-137,].

### 3.4.3 Application of DIRE pulse sequences to investigate SARS-CoV-2 infection

An amended DIRE (DIffusion and Relaxation Editing) pulse sequence [127] was applied to investigate SARS-CoV-2 infection due to the ability of the sequence to distinguish signals from molecules with slow translational diffusion but with increased segmental motion. These molecules include plasma glycoproteins and molecules within lipoprotein subfractions. Distinct NMR signatures were detected between SARS-CoV-2 patients and healthy controls, with elevated GlycA and GlycB, and decreased intensity of a cluster of choline head group signals (termed SpC, and identity confirmed by hetero- and homonuclear NMR experiments)

in SARS-CoV-2 infection. OPLS-DA modelling of SARS-CoV-2 patients and healthy controls (figure 3.13B) presents a very strong model ($R^2X$ 0.122, $R^2Y$ 0.975, $Q^2$ 0.682), with two clearly defined groups, with the loadings (figure 3.13C) presenting the most discriminatory metabolites to be two triglyceride peaks, GlycA, GlycB, SpC and branched chain sugars associated with glycoproteins to be highly correlated within the model projection.

DIRE spectra give a clear and significant signature for SARS-CoV-2 infection, the major components of this signature are elevated GlycA and GlycB and depleted SpC signal in SARS-CoV-2 infection compared to healthy controls. The depleted SpC signal comes partially from a molecule with the same diffusion constant as GlycA and GlycB, identified as linoleoyl-phosphatidylcholine [127]. Lysophosphatidylcholines, in blood, are known to bind to and be carried by a-1-acid glycoprotein, the primary contributor to the GlycA signal, and thus have the same molecular diffusion properties and would be detected at the same signal. Another component of the SpC is viably an HDL-phospholipid, based on the correlations between SpC and the array of lipoproteins, with strong correlation found between HDL-phospholipids and SPC in both healthy controls and SARS-CoV-2 patients (figure 3.19). Figure 3.19 further presents the correlations between SpC and GlycA and SpC and GlycB against lipoproteins, with results suggesting that these components originate from different contributions from the LDL and HDL glycoprotein bound phospholipids in each signal, which is significantly reduced in SARS-CoV-2 infection. Weaker correlations are observed with HDL1 (phospholipids) and HDL2 (cholesterols) as these subfractions are depleted in SARS-CoV-2 infection (figure 3.19B), and thus suggests that a significant proportion of the SpC signal is present in HDL3 and HDL4. Correlation is observed between SpC and LDLs, but again this is weaker than the correlations between SpC and HDLs. From these results, we can infer that

SpC markers are from HDL3 and HDL4. Strong correlation of SpC and proinflammatory cytokines IL1RA, IL17A and IL10 ratios (figure 3.22) suggests a strong immune component in SARS-CoV-2 infection. Overexpression or under-expression of IL1RA has been found to have a significant effect on lipoprotein metabolism, affecting cholesterol metabolism [139], an effect we see in SARS-CoV-2 patients. As a-1-acid glycoprotein is significantly elevated in SARS-CoV-2 infection due to inflammation, this makes GlycA, GlyB and SpC signals very sensitive to the presence of the disease.

### 3.4.4 Obesity and sex differences in SARS-CoV infection

The role of obesity in SARS-CoV-2 infection has highlighted the global obesity endemic, representing an unprecedented risk factor, with obese patients twice as likely to suffer from severe SARS-CoV-2 disease than healthy weight counterparts [140]. A very weak model was produced when modelling obesity between SARS-CoV-2 patients and healthy controls, primarily due to the limited sample number. In this instance, to investigate the role of obesity in SARS-CoV-2 infection, obesity was modelled within the group, looking at two groups with BMI <30 and BMI >30(figures 3.12 and 3.16). Loadings from OPLS-DA modelling of standard 1D (figure 3.12) spectra of these groups, presented the most discriminatory metabolites to be elevated valine, lactate, GlycA, choline species, glucose and formate. Discriminatory modelling of DIRE spectra (figure 3.16) further reiterates these findings, reducing these components down, highlighting elevated triglycerides, GlycA and GlycB, SpC and sugar chains associated with glycoproteins. Both models suggest that obesity presents an elevated SARS-CoV-2 infection metabolically, with key SARS-CoV-2 sentinel metabolites to be further elevated in patients with a BMI >30. Several studies have implicated triglycerides with SARS-

CoV-2 infection [141,142], with high correlation to diabetes, severe disease and poorer outcomes. Additionally, studies have indicated that obesity may play a role in prolonged viral shedding in infected patients, subsequently increasing disease transmission in the population [143]. Obesity is thought to contribute to viral propagation through the upregulation of ACE2 receptor and thus promoting entry of the SARS-CoV-2 virion into an amalgamation of cell types such as bronchial epithelial cells [143]. The metabolic signature found within these patients, presents obesity to be a severe metabolic implication, with a heightened SARS-CoV-2 response.

A sex-related difference in SARS-CoV-2 infection has been noted since the start of the pandemic, with higher degrees of mortality and severity found in men [144]. OPLS-DA analysis presents a significant albeit weak model (figure 3.11B) from standard 1D NMR spectra, with a clear clustering of males and females, but notably, a specific cluster of SARS-CoV-2 infected males in the upper quadrant, with separation driven by higher levels of VLDL1, LDL1, VLDL2, and LDL2 lipoproteins in women, and higher glucose in men. Altered lipid profiles have been documented with SARS-CoV-2 infection, with lower proportions of HDL cholesterols found in men($p$value<0.004) with poorer outcomes and severe disease phenotypes. Weaker modelling found from DIRE spectra (figure 3.15B) in this instance but defines increased lipoproteins in females. Again, the infection exerts a clear influence in the PCA plot indicating that sex is not a significant confounder in the analysis.

### 3.4.5 Lipoprotein changes in SARS-CoV-2 disease

OPLS-DA modelling of lipoprotein subfractions ($R^2$=0.906, $Q^2$, 0.885) (figure 3.17C) shows clear separation of healthy controls and SARS-CoV-2 patients with the exception of single healthy control which sits in the SARS-CoV-2 cluster. The model itself shows less discrimination in comparison to resolution of 1D NMR spectra (figure 3.10B), which can be attributed to the reduced number of variables. Nonetheless, a high degree of model robustness was found with an area under of the curve of 1. The most discriminant lipoprotein subfractions driving the OPLS-DA model separation were higher concentrations of VLDL cholesterols and phospholipids in SARS-CoV-2 patients, in addition to LDL triglycerides and its subfractions.

The signature detected using lipoprotein analysis suggests a deep systemic metabolic aberration, with similar patterns attributed to the SARS-CoV-2 group also found in relation to type 1 diabetes [145]. One study demonstrated that type 1 diabetes was associated with a high concentration of total cholesterol, LDL triglycerides, VLDL subfractions, LDL and HDL subfractions [146]. An increased risk of cardiovascular disease in diabetics has also been reported to be associated with higher concentrations of LDL subfractions (LDL5 and LDL6) [147]. Furthermore, clinical observations have found that small and medium LDLs are cardiovascular risk predictor in type 1 diabetes; with LDL particle concentrations associated with vascular wall thickening [148]. A characteristic of type 1 diabetes and cardiovascular disease is increased liver fat content [149]. A strong correlation between liver fat content and higher VLDL concentrations was found in this study in SARS-CoV-2 disease, consistent with metabolic markers of liver dysfunction such as elevated tyrosine, phenylalanine,

glutamine:glutamate ratio and the Fischer's ratio [150]. It is noteworthy that abnormalities in cardiac structure and function have been noted in COVID-19 patients, up to 3 months post-diagnosis and recovery [151]. Aberrations in lipoprotein profiles, specifically elevations in LDL triglycerides, have been found to be associated with vascular artery calcification and thickening [152]. Similar lipoprotein elevations are seen in the SARS-CoV-2 signature (figure 3.22), with the model also influenced by elevations in LDL, LDL5 and LDL6 cholesterol and phospholipid apolipoprotein B.

### 3.4.6 Cytokine Storm in relation to lipoprotein metabolic profile

To investigate the association between inflammatory profiles and lipoprotein changes, a panel of 14 cytokines, found to be above the limit of detect, was correlated with lipoprotein subfractions. Figure 3.22 shows strong correlation between lipoprotein subfractions and hierarchically clustered cytokines, which presented 4 distinct cytokine clusters. Cluster 1 (C1) presents MIP-1B, SDFa, IL-1a and IL-11, correlating with depleted LDL particles and depleted VLDL particles. MIP-1B was found to significantly elevated in SARS-CoV-2 infection (figure 3.20) in comparison to healthy controls (pvalue <0.01). C1 cytokines are likely a representation of inflammation driven by innate immune cells; consistent monocyte and macrophage activation. Cluster 2 (C2) comprises of IL-19 and IL-10 and ratios of IL-10 to IL-6 and IP-10; generally weaker correlation with lipoproteins were observed. Cluster 3a (C3a), IL-8 and MCP-1 were weakly anticorrelated with the majority of lipoproteins, but strongly positively correlated with LDL1 and LDL2, cholesterol and triglyceride subfractions. This finding corroborates with the correlation from C1, with this cluster also reflecting inflammation mediated innate immunity, implicating a core function by activated neutrophils

and monocytes/macrophages. IP-10 and RANTES (cluster 3c, C3c) shows the strongest correlation with the lipoprotein panel, (correlated with L1–L4 LDL subfractions and anticorrelated with HDL, VLDL, and L6 LDL parameters). Both cytokines are significantly elevated in SARS-CoV-2 infection (pvalue<0.01). RANTES is also strongly correlated with GlycA, a marker for systemic inflammation, and has been suggested to be a more robust marker of inflammation than CRP or IL-6[152]. Previous studies have found that amplified levels of circulating pro-inflammatory cytokines, such as Interferon γ and IL-6, and chemokines MCP1 and IP-10 are associated with pulmonary inflammation and respiratory distress in the aetiology of SARS-CoV-2[153].

The functional capacity of RANTES and IP-10 in mediating inflammation and cytotoxic T cell response for viral load clearance, and their correlation to GlycA suggests that an elevation in these cytokines is a reflection of disease severity [153). Studies have further shown in the pathogenesis of SARS-CoV-2, inflammatory markers, such as IL-6, did not change significantly, but some like IP-10, decreased in patients who entered remission, but remained elevated in poor patient outcomes [154]. Cluster 4b (C4b) consists of IL-7 and IL-1RA, both significantly elevated in SARS-CoV-2 patients and TNF-α. TNF-α and IL-1RA are both anti-inflammatory and increase elimination of the infected cells [155]. However, given the substantial “cytokine storm” that ensues following SARS-CoV-2 infection, concentrations of TNF-α and IL-1RA may be insufficient to alleviate and counteract the level of proinflammatory cytokines released in response to the virus. IL-1RA is found to be strongly correlated with HDL and VLDL lipoprotein particles; HDL has been proposed to be an anti-inflammatory lipoprotein with protective effects against oxidized lipids [ and has been found to reduce T cell activation [156], and thus

the correlation of HDL with IL-1RA supports the implication that IL-1RA reduces the inflammatory cascade.

### 3.4.7 Biogenic amines and tryptophan metabolism in relation to SARS-CoV-2 infection

Aberrations in lipid metabolism and alterations in amine concentrations have been reported in response to SARS-CoV-2 infection, with similar reports from SARS-1 and MERS outbreaks, with further reports of a sustained chronic change [154]. 13 out. of 43 quantified metabolites were found to be significant when comparing SARS-CoV-2 infected patients to healthy controls (figure 3.23). Discriminatory analysis using OPLS-DA (figure 3.24C), presents two clear clusters, driven by depleted histidine and tryptophan and elevated neopterin, quinolinic acid, glutamic acid, kynurenine and taurine (figure 3.25) in SARS-CoV-2 patients. This signature is consistent with reports of inflammation seen with cardiovascular disease and diabetes [158]. Quinolinic acid has been suggested to act as an agonist for glutamate receptors and neuronal stimulation; however, upregulation of this mechanism is thought to lead to neuronal damage indirectly through neurotoxicity [159). Additionally, quinolinic acid and kynurenic acid are thought to regulate glutamic acid function, with elevated quinolinic acid resulting in glutamic acid release in cells [160]. A significant glutamic acid release may be responsible for the significantly lower glutamine/glutamic acid ratio seen in SARS-CoV-2 infection. Glutamic acid and glutamine play pivotal roles in the homeostasis in immune cell function, including T-cells and astrocytes within the central nervous system (CNS) [161] where it acts as an immunomodulator, upregulating glutamate surface receptors [162]. The glutamine/glutamic ratio is considered an important diagnostic marker, with glutamine consumption rates increased by immune cells when placed under catabolic conditions such

as sepsis [163]. A crucial finding with this data is the depletion of tryptophan, and the clinical implication and association with depression, a key long-COVID symptom found in nearly 1 in 5 SARS-CoV-2 recovery cases [152], suggesting this finding may be associated with long term implications.

Sex differences in the amino acid and tryptophan pathway metabolites were also modelled to determine if there was any sex bias, but a very weak model was produced, with no clear clustering of disease and healthy within the sex clusters. The key metabolites influencing the OPLS-DA. were depleted xanthurenic acid, methionine, picolinic acid and citrulline in SARS-CoV-2 infection (figure 3.26A), suggesting that the SARS-CoV-2 infection model in figure 3.23C is not influenced by sex. BMI was also assessed to investigate the influence on the model. Clear clustering of BMI groups was observed, with most SARS-CoV-2 patients seen in the higher BMI groups. The model was driven by elevated glutamic acid, 3-hydroxykynurenine, quinolinic acid and nicotinic acid (figure 3.26C) in SARS-CoV-2 infection. These markers suggest an enhanced inflammatory signature for SARS-CoV-2 patients with elevated BMIs, consistent with findings from standard 1D NMR spectra BMI modelling.

### 3.4.8 Immuno-metabolic signatures of biogenic amines and tryptophan pathway metabolites in SARS-CoV-2 infection

To investigate the inflammatory metabolic signature further, cytokines were correlated with quantified metabolites, with hierarchical clustering presenting 5 clusters with the metabolites (figure 3.29). The first cluster of cytokines (C1) shows very weak correlation of SDF1 alpha and MIP1-beta, which only strongly positively correlate to neopterin, 3-methylhistidine and

indole-3-acetic acid, and negatively correlate to arginine. MIP1-beta is a chemoattractant, significantly elevated in SARS-CoV-2 infection, and has been found to be implicated in the neuroinflammatory pathway [161] and thus the results from this correlation analysis suggest an underlying systemic metabolic response response. The second cluster of cytokines (C2) includes IP10, MCP1, RANTES, IL8 and IL18. Interestingly these inflammatory cytokines. associated with cytotoxic T cell activation, are strongly positively correlated with kynurenine, kynurenine/tryptophan ratio and 3-hydroxyanthranillic acid, and thus suggests these cytokines are a reflection of the severity of respiratory disease. In cluster 3 (C3), TNF-a is strongly negatively correlated with the majority of tryptophan pathway metabolites suggesting downregulation of TNF-a, and thus suggesting an inhibition of its anti-inflammatory functionality.

In cluster 4 (C4), Interferon gamma (IFNg) shows strong negative correlation to a battery of amino acids and specifically branched chain amino acids. Branched chain amino acids, valine, leucine and isoleucine are essential amino acids and involved in several neurological processes but also in immune modulation, modulating the production IFNg [164]. This correlation suggests a complex interaction between metabolites and immune modulation, further exemplified by the hierarchical clustering maps (figure 3.30C and 3.30F), presenting IFNg as highly associated with a several pro-inflammatory cytokines and also kynurenine. The most enriched pathway highlighted from metabolite set enrichment analysis (MSEA) is methylhistidine metabolism, driven by the significantly depleted histidine plasma concentrations in SARS-CoV-2 infection (figure 3.31A). A significantly depleted Fischer's ratio in SARS-COV-2 infection further implicates the aberrations in liver function, compared to normal. Histidine has been found to augment the suppression of hepatic glucose produce via

central insulin action [165]. Glucose intolerance in type 2 diabetes is related to enhanced hepatic glucose production via gluconeogenesis. Insulin supresses hepatic gluconeogenic gene expression through neuro insulin receptors and through direct interaction with hepatic insulin receptors [166]. In the absence of sufficient histidine, Kimura et al suggest that there is a significant dysregulation in insulin function, resulting in elevated hepatic glucose production.

In the final cluster of cytokines (C5), IL1RA, MIP1-alpha show strong correlations, correlating antagonistically Eotaxin. Strong positive correlation is seen between IL1RA and MIP1-a with the majority of tryptophan pathway metabolites, whereas strong negative correlation is observed with Eotaxin for the same group. MIP1-alpha has previously been found to be implicated with the tryptophan pathway, with picolinic acid identified to be involved in macrophage activation and modulates the expression of chemokines in macrophages [167]. Interestingly, Eotaxin has been found to be increased in neurodegenerative disorders [168] and is also found to be increased in SARS-CoV-2 infection, with the strong correlation with tryptophan metabolites suggesting an underlying immune-metabolic mechanism.

These findings suggest that the metabolic signature found in biogenic amine and tryptophan pathway metabolites, is a complex multiorgan mosaic disease, and further presents the complexity of the disease.

## 3.5 Conclusions

The findings from this study suggest that there is a complex, multi-organ response to SARS-CoV-2 infection, and although primarily a respiratory disease, it has been found to have profound systemic metabolic effects, with signatures similar to diabetes and cardiovascular disease. Results from the DIRE spectra provide a novel orthogonal approach to diagnostics for SARS-CoV-2 infection, and possibly post-active infection as well. Whilst the cytokine storm that is characteristic of severe SARS-CoV-2 infection is attributed to an uncontrolled over-production of soluble markers of inflammation, it is difficult to capture the peak of this storm in more mild disease given the short half-life of cytokines and thus more time-points would be imperative for future studies. Further profiling of systemic changes longitudinally would prove to be advantageous for the prediction of future risk factors.

# CHAPTER FOUR: UK- PRIMARY BILIARY CHOLANGITIS (UK-PBC) NESTED COHORT STUDY

The aim of this work was to apply metabolic profiling methods to comprehensively investigate the molecular mechanism driving Primary Biliary Cholangitis (PBC), and also to profile differences between treatment responders and non-responders. Furthermore, the aim of this project is to integrate multi-omic data for in depth disease screening and classification.

# 4.1 INTRODUCTION

### 4.1.1: Introduction

Primary Biliary Cholangitis (PBC, formerly known as Primary Biliary Cirrhosis) is a chronic autoimmune cholestatic liver disease, characterised by progressive bile duct damage as a result of decreased bile flow due to impaired secretion by hepatocytes or through bile flow obstruction. The pathogenesis of PBC, is defined as a prototypical autoimmune disease, with the presentation of anti-mitochondrial antibodies, resulting in inflammation localized to small intrahepatic bile ductules. The prolonged inflammation and accumulation of inflammatory infiltrates results in the progressive destruction of cholangioctyes, causing a spectrum of cholangitis and biliary cirrhosis. Typically, PBC manifests with a slow progression rate, with incrementally increasing cholestasis and gradual development of liver fibrosis, cumulatively resulting in cirrhosis, hypertension and liver failure [169].

The presentation and onset of PBC in patients greatly varies, given the multifactorial nature of the aetiology and complicated pathogenesis, which has yet to be fully elucidated. Evidence has suggested that a crucial factor is the complex interplay between genetic susceptibility and environmental factors, resulting in a mosaic of disease pathogenesis [170].

Detection and diagnosis of PBC is often made in later stages of the disease process, based on broad presentation of symptoms; biliary damage, pruritus and fatigue, whilst liver biochemistry may remain normal during the earlier stages of disease progression. High

morbidity and mortality remain a major issue with the disease, with approximately 10% of all liver transplants in North America and Europe attributed to PBC [171]. PBC not only has a markedly negative impact on patient quality of life, but also has a high economic impact with the cost of PBC hospitalisations in the US estimated to be $69-115 million annually [172]. Hepatocellular carcinomas are amongst the most common malignancies of the liver, with between approximately 70-90% of all cases diagnosed against a background of chronic liver disease and cirrhosis [173]. The increased incidence and frequency of hepatocellular carcinoma, in PBC patients is well documented, with PBC patients up to 16% more likely to develop hepatocellular carcinomas. In addition, hepatocellular carcinomas are amongst the main causes of death in patients with chronic liver cirrhosis globally. In light of the significant risk of hepatocellular carcinoma development in PBC, PBC patients are often considered for earlier transplantation, increasing the global economic and healthcare burden of this disease [174]. Several studies have indicated an increased risk of extrahepatic malignancies such as stomach or pancreatic cancer, in comparison to hepatocellular carcinomas, but with no association to the stage of liver disease [175].

### 4.1.2: Diagnosis, Ursodeoxycholic Acid and Treatment

The biochemical diagnostic criterion for PBC specifies recurring elevated alkaline phosphatase (ALP), and gamma-glutamyl transferase (GGT), decreased serum albumin and elevated bilirubin. Elevated levels of Bilirubin have long been defined as a significant prognostic marker for clinical outcomes in PBC, however the change in bilirubin levels is often unremarkable in the early stages of the disease, thus making it unviable as a criterion for early intervention [176]. ALP is used as a measure for patient response to

ursodeoxycholic acid (UDCA), the frontline choice of treatment for PBC; in the absence of cirrhosis, Serum ALP concentrations correlate with the severity of inflammation and reduction in intrahepatic bile ducts [187]. Patients with lower levels of ALP have been found to have stronger positive outcomes and are less likely to develop progressive liver disease [177].

The gold standard for treatment remains life-long drug therapy with ursodeoxycholic acid (UDCA) in combination with further drug therapy to treat the amalgamation of symptoms that present with the disease and the progression of liver disease. UDCA is predominantly hydrophilic, dihydroxy bile acid, and is present in humans in a concentration of approximately 3% of the total bile acid pool [175]. This small proportion of endogenous UDCA in human biliary bile acids (3α,7β-dihydroxy5β-cholan-24-oic acid) is considered to be attributed to the gut microbiome epimerization of the 7-hydroxyl group of chenodeoxycholic acid (CDCA), a bile acid synthesized directly from cholesterol [175]. Whilst the full mechanism of metabolism of UDCA is yet to be elucidated, what is understood is that UDCA is predominantly metabolised to its glycine conjugated and taurine conjugated forms [175]. Obeticholic acid (OCA) is a recently approved drug for use in PBC treatment, used either in combination with UDCA in adults who have shown inadequate responses to UCDA, or as a singular therapy in patients intolerant of UCDA [179].

Both UDCA and OCA can be synthesized and administered therapeutically and have complementary mechanisms of action. OCA is a farnesoid X receptor (FXR) agonist, acting to decrease bile acid secretion, which subsequently reduces accumulation of toxic bile acids that lead to cholestatic biliary damage. Simultaneously, OCA also increases bile acid

secretion by increasing expression of secretogogues such as VIP receptor-1, inducing further choleretic effects such as increased bile acid flow, a mechanism that UDCA also shares, increasing bile acid clearance in the liver and decreasing inflammation through FXR regulation and limiting pro-inflammatory responses to damage-associated molecular pumps (DAMPs)[180].

Failure to respond adequately to UDCA is associated with increased risk of mortality due to inevitable progression to end-stage liver disease, finally requiring liver transplantation. The therapeutic response to UDCA is typically measured in terms of liver biochemistry but several classifications such as the 'Paris' or 'Toronto' definitions exist, each outlaying different biochemical stratifications, but here we specifically focus on the Paris definition which is more widely used in the UK. The Paris I definition of UDCA response, is based on the following criteria: total bilirubin ≤17μmol/L, aspartate transaminase (AST) <2 times the upper limit of normal (ULN) and alkaline phosphatase (ALP) <3×ULN after 12 months of treatment with UDCA [181].

UDCA is a choleretic and hydrophilic endogenous secondary bile acid, with multiple effects, but is primarily used to replace some of the more toxic bile acids, leading to reduced indirect choleretic damage. UDCA induces the secretion of bile acids from hepatocytes, thereby reducing further hepatocyte damage and apoptosis, and the likelihood of subsequent inflammation and fibrosis [182]. In addition, the UDCA activates the AE2 transporter pathway; the stimulation of AE2 expression by glucocorticoid receptors stimulates biliary bicarbonate secretion and in turn reduces bile toxicity. This results in the enrichment and expansion of bile acid availability and increasing the life cycle of bile acids.

The immunomodulatory impact of UDCA has also been observed, assumed to be attributed to the UDCA-induced activation of the glucocorticoid receptor [182].

Responsiveness to UDCA is an important prognostic marker in Primary Biliary Cholangitis, with UDCA responsive patients found to have improved health outcomes comparable to the general population. However, predicting patient response to UDCA is limited. Studies have found insufficient response to UDCA in 30% of North American and European PBC patients, which incrementally rises to 32-44.7% in Chinese patients [183]. Despite UDCA response being a marker for patient outcomes, very little is defined or well established regarding the mechanism of action of the secondary bile acid and variable patient response. Further investigations are required to fully elucidate the mechanism of response in order to optimise treatment [182]. An argument for inconsistent UDCA response, is that it is the manifestation of materially different bile acid pools in individuals, alternatively variable modification of UDCA itself by the gut microbiome during entero-hepatic recirculation has been proposed as a mechanism [184].

Bile acid sequestrants such as cholestrymine and colestipol are often prescribed to treat a common yet highly problematic complication of PBC, pruritus, that is estimated to affect 60-70% of patients. Bile acid sequestrants typically act as bile acid-binding resins, highly positively charged molecules that bind to the negatively charged bile acids in the intestine and inhibit lipid solubilizing activity and consequently block cholesterol absorption [185]. In addition, bile acid sequestrants inhibit the reabsorption of bile acids and thus increases bile acid synthesis, the mechanism of which in turn sequesters cholesterol synthesis in the liver, contributing to a reduction in global serum cholesterol levels [185]. However, bile acid

sequestrants have been found to have limited effectiveness and tolerability in practice. The pathogenesis of pruritus in PBC is still not well understood, and whilst several putative pruritogens have been proposed, treatment remains limited and ineffective with often hepatotoxic side effects in their own right, and liver transplant may be indicated even in the absence of advanced liver failure [182]. The heavy burden of cholestatic pruritus for PBC patients, indicates a pressing need to investigate potential pruritogens to improve treatment and quality of life.

### 4.1.3: Biochemistry of Primary Biliary Cholangitis

Consistently elevated levels of serum alkaline phosphatase (ALP) are typically detected in PBC patients. In addition, treatment response and failure of UDCA is universally defined by ALP levels as a key prognostic variable, identifying those with greatest risk of disease progression with detected ALP levels 1.67x the upper limit of normal as currently defined by the Paris II definition [181]. Characteristics of biliary damage include the loss of bile ducts, elevated bilirubin levels, progression of fibrosis, and ultimately can lead to portal hypertension and end-stage liver disease [182]. The origin of ALP elevation in hepatobiliary disease has yet to be fully elucidated but it is currently believed that the rise in serum ALP is a result of increased de novo synthesis in the liver followed by release into systemic circulation [183]. This is further aggravated by the accumulation of retained bile acids, which in turn induce the production of ALP, and promote leakage into circulation [184].

Another significant clinical marker of liver disease and PBC is the aspartate transaminase (AST) to alanine aminotransferase (ALT) ratio, an increase in which has been indicated as a

predictor of cirrhosis [185]. Currently, the mechanism of action is associated with the elevation in this ratio, but it has been suggested that sinusoidal clearance of AST decreases in cirrhotic patients, resulting in an accumulation of AST and the high AST/ALT ratio in PBC patients is the result of reduced hepatocyte accumulation [186].

Increased biliary constituents in serum are detected in PBC patients. Bilirubin levels are a crucial marker of liver function and prognostic factor for disease progression in liver disease, with poor prognosis defined as 2 x the upper limit of normal (<1000U/L) (Paris II definition) after 1 year of UDCA treatment [181]. Elevated levels of bilirubin correlate with hepatic disease severity, although higher concentrations are atypical of early-stage PBC. However, significant inter-patient variation in bilirubin levels in PBC has been reported [182]. In PBC, serum bilirubin elevation is predominantly its conjugated form, suggesting that the normal hepatocytic activity of glycosylating enzymes is preserved to produce conjugated bilirubin. Accumulation and elevation of conjugated bilirubin may occur as a result of obstructed bile flow, compromised bile duct integrity or reduced bile production due to malfunctioning activity of bile efflux transporters [183]. In advanced PBC, haemolysis is the result of excess bile acids, which in turn, may further contribute to elevation of serum bilirubin levels. As strong detergents, bile acids can solubilize the cytoplasm membranes of erythrocytes and other blood cells causing them to haemolyse, resulting in an amalgamation of haematological complications in end-stage PBC [183].

### 4.1.4: Disease Pathogenesis, Autoimmunity and Genetic Susceptibility

The aetiology of PBC remains to be fully elucidated. Autoimmunity in PBC is characterised through the presentation and loss of tolerance to a cascade of ubiquitous mitochondrial autoantigens, the E2 subunits of the oxo-dehydrogenase pathway of oxidative phosphorylation [170]. There are four autoreactive mitochondrial antigens targeted by anti-mitochondrial antibodies in PBC: the pyruvate dehydrogenase E2 complex (PDC-E2), E3-binding protein(E3-BP), ketoglutaric acid dehydrogenase E2 complex (OGDC-E2) and branched-chain 2-oxo-acid dehydrogenase E2 complex (BCKD-E2). Amongst the four autoantigens, there is substantial homology, all participating in oxidative phosphorylation and sharing lipoic acid moieties [187]. Most commonly antibodies react to PDC-E2, either singularly in some patients or in addition to other target antigens. PBC is suggested to be the only disease in which autoreactive T cells and B cells responding to PDC-E2 are detected. In response to autoreactive T and B cells, evidence suggests that apoptosis is the most important mechanism driving biliary epithelial cell loss. Furthermore, defective clearance of apoptotic bodies released from damaged biliary epithelial cells has been proposed as another source of autoantigens and further autoimmunity [188]. Biliary epithelial cell stress responses include necrosis, apoptosis, senescence and autophagy. Biliary epithelial cell survival from stress is associated with the production of DAMPs, chemokines, cytokines and growth factors [188]. This interplay promotes the de-differentiation of healthy biliary epithelial cells, which lose E- cadherin but increase mesenchymal antigens (S100A4, fibronectin, vimentin and α-smooth muscle actin (α-SMA). A positive association between p21- expressing and S100A4-expressing biliary epithelial cell during T-cell-led portal tract inflammation (AD10) has been found, suggesting complementary processes. Investigations

in individual differences in the nature or degree of the biliary epithelial cell response to stress could prove to improve understanding UDCA non-response and reveal potential therapeutic targets [188].

The manifestation of this subtype of antimitochondrial autoantigen, has been found to be associated with bile duct damage, which results in the development of lymphatic cholangitis, progressing to biliary fibrosis [188]. Thyroid hormone metabolism is an important factor in maintaining normal hepatic function and bilirubin metabolism. A significant proportion of PBC patients present with thyroid hormone imbalances and this is frequently associated with an elevation of alanine aminotransferase and alkaline phosphatase in hyperthyroidism and aspartate aminotransferases in hypothyroidism [189].

Genetic variation is one of the key components in disease susceptibility. Genome Wide Association Studies (GWAS) identified a strong association between PBC and NF-κB14 in the liver. Murine studies found that when deficient in this gene, PBC-like fibrosis developed with a strong ductular reaction (AD17) [190]. The final stages of the complex biliary epithelial cell biology of PBC, may therefore be accelerated fibrosis, with the implication that non-response to UDCA may reflect the severity of fibrosis progression. Familial studies showed the prevalence of PBC to be 0.72% in first-degree relatives and 1.2% in offspring of PBC patients. The highest risk observed in daughters of affected females, with a point mutation relative risk for this subset, reaching 87% in some studies [190]. A large study recorded that 20% of sisters, 15% of mothers and 10% of daughters of PBC patients tested positive for anti-mitochondrial antibodies, presenting with or without symptoms [191]. Whilst the

presence of anti-microbial antibodies does not always result in a PBC diagnosis, the existence indicates a predisposition, particularly in the event of familial history [191].

Studies have revealed that having a first-degree relative, a history of urinary tract infections, smoking history and use of hormone replacement therapies are associated with the risk of developing the disease. In addition, other autoimmune diseases are found in approximately 30% of patients [191]. A history of cholestasis or pruritus during pregnancy associated with PBC implies a crucial hormonal and gender link, that suggests early disease may be present for decades prior to presentation with abnormal liver biochemistry and symptoms [191] The genetic basis of PBC is also implied by the much higher disease concordance rates in monozygotic twins than in dizygotic twins [171], and the prevalence of PBC being 100 times higher in some aboriginal populations [192]. PBC patients and their families, typically present with other autoimmune diseases, such as celiac disease, thyroid disease and Sjogren's disease, indicating a shared genetic predisposition [171]. However, in comparison to other autoimmune diseases, there is little association between PBC and the presence of any particular major-histocompatibility complex alleles [171]. A clear gender imbalance is characteristic with PBC; with the ratio of women to men affected as high as 10 to 1, but with the same disease manifestation in both sexes. Recent studies have suggested that a reason for this could be the higher incidence of X chromosome monosomy in lymphoid cells [190].

### 4.1.5: Epidemiology

In the UK, approximately 15,000 people are affected by the rare disease, but globally 1 in 1000 women over the age of 40 years are living with the consequences of progressive

chronic biliary inflammation, which has a significant impact on quality of life [170]. Onset is typically observed after 40 years of age, but incidence has been recorded in individuals as young as 15 years, although rare.

Investigations in the geo-epidemiology of PBC has revealed crucial associations that underpin the pathophysiology of PBC, in particular the impact of several environmental conditions that may increase disease risk and susceptibility, such as high levels of cadmium in soil. A higher incidence of PBC has been recorded in previously industrialised regions in the UK. A notable study in 1980 by Triger et al. reported the initial connection between PBC and environmental factors. PBC prevalence was found to be approximately ten-fold higher in a community receiving water from a particular reservoir in comparison to the general population [193]. Exposure to cadmium is believed to be another potential contributing factor in these regions, with severe structural liver damage and increased IgA deposits found in patients with high levels of chronic exposure [194].

Some xenobiotic associations have been found between household chemicals and PBC, such as acetone nail polish remover usage and increased disease incidence [194]. Studies have also found an increased incidence of PBC near superfund “rubbish-tip” sites identified in a large-scale epidemiological study in Newcastle, England [197]. The exposure to unidentified environmental agents could prove to be key in understanding the geospatial clustering of the disease and the underlying environmental links.

### 4.1.6: UK-PBC Nested Cohort Study

The UK-PBC Nested Cohort Study is a nationwide consortium project. This work utilises samples obtained as part of an MRC-funded initiative, aimed to develop a unique multi-modal research platform to study the mechanism of UDCA non-response and PBC disease stratification, with an objective of developing a framework for optimal patient treatment. The broad hypotheses of the study are that UDCA response is associated with a distinct molecular and cellular phenotype and that the associations may highlight potential biomarkers for patient stratification or identify molecular pathways relevant to therapeutic targets.

The UK-PBC initiative began in 2007 as an effort to establish a large patient cohort for high-throughput genetic studies. The study has since gone to establish comprehensive large scale genomic, transcriptomic, proteomic and metabonomic studies of PBC. Participants were recruited via the UK-PBC consortium, a research network consisting of 150 NHS trustees across England, with patients identified by local research nurses funded through the National Institute for Health Research (NIHR) Comprehensive Research Network (CRN). Comprehensive medical records and validated questionnaires characterising symptoms were collected and blood was provided by participants for DNA extraction.

As part of the cohort, matched urine, serum and stool samples were collected at designated research centres, collected for 470 PBC patients and 101 control participants recruited through the NIHR bioresource, that were age and sex matched. Of the 470 PBC patients, 406 were on UDCA therapy, 276 were defined as UDCA responders based on an ALP threshold of

1.67 X Upper limit of Normal (ULN). A multiplatform strategy of analysis utilising both NMR spectroscopy and UPLC-MS was used to generate metabolic phenotypes for patients.

### 4.1.7: Metabolic Profiling and Primary Biliary Cholangitis

Metabonomics is a comparatively new technology in the omics hemisphere but has proven to be a powerful and unique tool in the definition of global metabolic profiling for disease and health. The combination of statistical tools and different spectroscopy technologies facilitates the detection of different fields of analytes and metabolites, as well as the detection of xenobiotic compounds [196]. Metabolic profiling has been successfully used in defining signatures in liver diseases such as hepatocellular carcinoma [197].

In the current study, NMR spectroscopy was used for global metabolic profiling and semi-targeted UPLC-MS was used for bile acid profiling and quantification in urine, serum and stool samples. Given the established understanding of PBC and liver disease, circulating bile acids detected in serum are instrumental in understanding the mechanisms underlying biliary damage, and also variability in patient response to treatment. Analysis of stool samples will aid understanding of the impact of the microbiome on UDCA response, and further understanding of the impact of gut microbiota on PBC disease severity. These analyses will be combined to create a mosaic portrait for PBC and inconsistent UDCA response, to increase understanding of bile acid pools in individuals, and modification of UDCA by the gut microbiome during entero-hepatic recirculation.

# 4.2 METHODS

### 4.2.1: UK-PBC Nested Cohort Study Design

The UK-PBC Nested Cohort was recruited from 18 different research centres located across the United Kingdom, with a final sampling number of >600 patients (REC reference: 14/NW/1146). Recruited PBC patients were assigned to cohorts on the basis of specific criteria (Table 4.1), with the primary definer being the use of UDCA for treatment (Cohort A), and PBC patients who were treatment-naïve, i.e. those who were not on UDCA (Cohort B). Samples from age and sex-matched healthy controls were recruited via the NIHR Bioresource, with samples collected from participants at a research centre using standardised methods. Sample collection and processing was concentrated in each of these centres, using standardised protocols and methods to ensure consistent sample collection and handling. Matching serum, urine and faecal samples were collected from each participant for metabonomic studies at an appointed research visit, with urine and faecal samples immediately stored at -80°C until profiling, and serum samples were centrifuged and stored at -80 °C until profiling.

**Table 4.1:** UK-PBC study cohort outline and criteria

| Cohort | Criteria |
|---|---|
| **A** | Patients with diagnosed PBC, incipient PBC or PBC/Autoimmune Hepatitis who **ARE** receiving treatment with UDCA |
| **B** | Patients with diagnosed PBC, incipient PBC or PBC/Autoimmune Hepatitis who **ARE NOT** receiving treatment with UDCA (treatment naïve) |
| **C** | Patients with liver diseases other than PBC |
| **E** | Healthy age and sex matched controls from NIHR Bioresource |
| **F** | Healthy best friend &/or same sex sibling controls |

Informed consent was obtained from all participating patients and patients deemed unsuitable for the criteria of the study were excluded and informed. Clinical research nurses collected comprehensive medical records and validated questionnaires characterising symptoms, in addition to socio-geographical and alcohol consumption data.

### 4.2.2: $^{1}$H NMR data acquisition methods

$^{1}$H NMR data acquisition of urine, serum and faecal samples were undertaken according to previously published and peer -reviewed protocols and conducted by Dr Alexandros Pechliavanis at Imperial College London [45]. For the purpose of this thesis, only urine samples were analysed using $^{1}$H NMR spectroscopy and faecal samples were analysed using 16S rRNA. Samples were thawed at room temperature and randomised. Urine samples were centrifuged at 18,000 rpm at 4°C for 5 minutes, Urine samples were then diluted in NMR urine buffer (1.5M KH2PO4, 2mM NaN3, 1% 3-trimethylsilyl-2,2,3,3-2$H_4$-propionic acid (TSP)in 99.9% $D_2O$, pH 7.4) and then aliquoted into 3 mm NMR tubes. Due to the limited sample volume, no quality controls were prepared, aside from sample blanks to ensure integrity of the experimental run.

NMR data acquisition was acquired using a Bruker Avance II DRX 600MHz spectrometer operating with a BBI probe at 600MHz at 300K and an automatic sample handling SampleJet unit (Brukler Biospin GmBH, Germany). To ensure high quality data acquisition, spectral acquisition was supervised, and samples that did not meet the quality thresholds were rerun a maximum of three times. The primary thresholds for data quality were that the residual water resonance should not affect the spectra outside of δ 4.6-4.8 and have an observed spectral concentration of less than 10mmol. In addition, after applying

exponential line broadening of 0.3 Hz prior to Fourier transformation, the line width of alanine at δ 1.5 should be less than 1.4 Hz. NMR data were processed using Topspin v3.2.3 software with Icon NMR (Bruker Biospin GmBH, Germany), with automatic baseline-correction, phasing, and referencing to trimethylsilylpropanoic acid (TSP) at 0 ppm. Prior to analysis, the spectrometer was calibrated to ensure good water suppression and good shim quality (magnetic field homogeneity). For 1D experiments, a total of 32 scans were acquired.

### 4.2.3: Data Preprocessing

Raw data were imported into MATLAB(v2014a) and pre-processed using in-house software. Regions corresponding to TSP (<0.09 ppm), water (4.7-4.9 ppm), urea (5.63-5.99 ppm) and noise region with no resonances (>9.9 ppm) were removed from the data. Spectra were calibrated using an automated calibration script and to account for different dilutions, normalised using probabilistic quotient normalisation (PQN)[49]. The spectra were then manually examined for chemical shift variation and aligned using the recursive segment-wise peak alignment (RSPA) algorithm [50].

### 4.2.4: Statistical Analysis

Following spectral preprocessing and normalisation, data were exported into SIMCA 15.0.1 (Umetrics, Umea, Sweden) for multivariate analysis. Initial Principal Component Analysis (PCA) was carried out to identify extreme outliers and inspect sample variability. Extreme outliers that were observed due to poor data acquisition as a result of insufficient sample

volume, were removed from further analysis. Following removal of outliers, PCA was performed to examine clustering between disease and control groups according to the biochemical threshold of ALP 1.6x ULN. Supervised partial least square discriminant analysis (PLS-DA) was performed to investigate metabolites that varied between disease groups. Statistical significance and robustness of PLS-DA models were then validated by permutation testing of the Y matrix for 1000 permutations. The obtained p-values were adjusted using Benjamini-Hochberg (BH) correction [48] with a false discovery rate (FDR) threshold of 0.05.

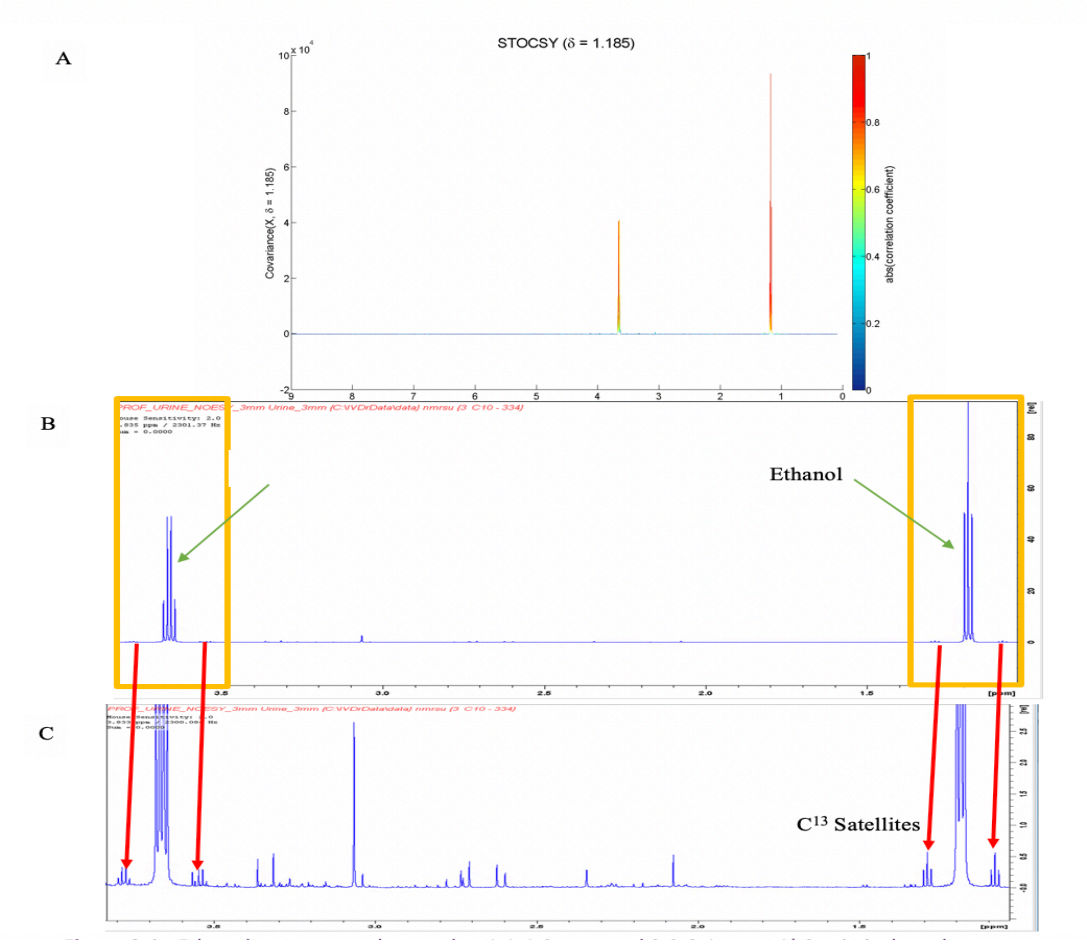



**Figure 4.1:** Ethanol resonances detected at 1.1-1.3 ppm and 2.9-3.1 ppm. A) Statistical total correlation spectroscopy (STOCSY) of assigned ethanol resonance (1.195 ppm). B) Very high concentration of assigned ethanol. C) Magnification of spectra in Fig. 1B, highlighting $C^{13}$ satellite abundance.

A high proportion of the samples were found to contain non-physiological concentrations of ethanol (Figure 4.1) suggesting 'contamination' of the samples. This contamination possibly arose from the collection tubes since sample blanks prepared in the laboratory using the same chemical buffer and reference standards, but no sample, did not contain these high concentrations of ethanol.

In most cases of alcohol consumption, ethanol was present at an estimated 130 mM concentration assuming urinary creatinine to be present at approximately 10mM [40]. However, some samples were contained as much as 320 mM of ethanol. The high concentration of ethanol was further corroborated by the fact that many spectra showed samples with $C^{13}$ satellites of the ethanol $CH_3$ peak which was clearly visible at 1.29ppm and 1.07 ppm (Figure 4.1). This level of ethanol is not consistent with physiological concentrations, even after heavy drinking which is typically considered to be ±65.1 -86.8 mmol [198].

In addition, a smaller subset of the samples containing credible physiological amount of ethanol, which were accompanied by evidence of ethanol metabolites such as ethanol glucuronide, ethyl glycoside or ethanol sulphate, which would be expected if the ethanol originated from ingestion and was subsequently metabolised (Figure 4.2).

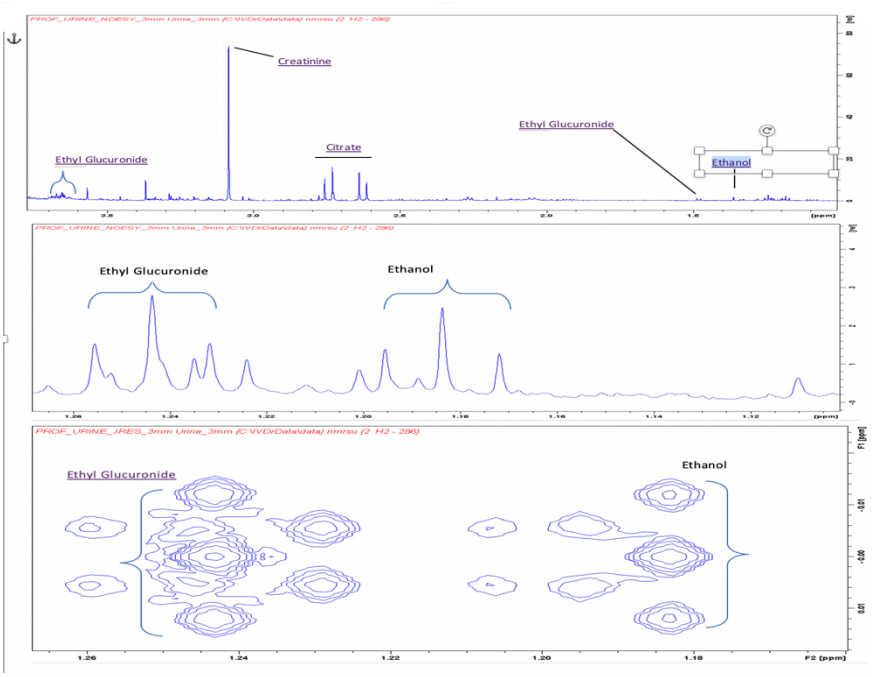



**Figure 4.2:** Ethanol-related resonances detected at 1.1-1.3 ppm and 2.9-3.1 ppm. A) Skyline plot of ethanol regions). B) Magnification of ethyl glucuronide and ethanol C) Magnification of JRES in Fig. 2.1, ethyl glucuronide and ethanol.

The number of samples containing physiological versus non-physiological levels of ethanol was calculated and compared with the self-reported data on alcohol consumption (Table 4.2).

**Table 4.2:** Number of samples containing ethanol and ethanol glucuronide signals

| **Cohort** | **Total no. samples in Cohort** | **No. samples containing ethanol (Percentage of Cohort)** | **Number of participants declaring alcohol consumption (%of Cohort, % of samples)** | **No. samples containing ethanol glucuronide (Percentage)** |
|---|---|---|---|---|
| A (PBC treated) | 406 | 326 (80.3) | 167(41.1, 51.2) | 46 (11.3) |
| B (PBC treatment naïve) | 64 | 50 (78.1) | 31(48.4, 78.1) | 9 (14.1) |
| C (Liver disease) | 3 | 2 (66.7) | 1(33.3, 50) | 2 (66.7) |
| E control | 101 | 61 (60.3) | 78(77.2, 100) | 11 (10.8) |
| F control | 3 | 3 (100) | 1(33.3, 33.3) | 0 (0) |
| Total | 577 | 442 (76.6) | 276(65.4, 47.8) | 68 (11.7) |

Since an average of 76.6% of the samples contained ethanol, of which only 11.7 contained the ethanol metabolite ethanol glucuronide and 47.8% of the participants recording alcohol consumptions, it was strongly suggestive of contamination rather than being an endogenous phenomenon. Given the high percentage of 'contaminated samples, it was not feasible to model the spectra without removal of the ethanol spectral region. Subsequently, new PCA and PLS-DA models were recalculated after removal of the regions δ 1.0-1.34 and δ 3.655 to 3.828.

Statistical Total Correlation Spectroscopy (STOCSY) [20] was carried out to identify and confirm signals from the same compound and elucidate the molecular structure. Metabolite assignments were made using an in-house database and using the Human Metabolome Database (http://www.hmdb.ca). For selected metabolites, showing a trend in differentiating participants with PBC from controls (phenylacetylglutamine (PAG), Hippurate and 4-cresylsulphate (paracresol sulphate)), peaks were fitted using a method reported by Loo et al. [191]). Briefly, this method uses a multiblock pattern recognition approach wherein a reference pattern from the dataset is used to locate the area where the correct spectral features are present in the sample. Individual PCA models are calculated for each spectral region, for blocks of 0.5 ppm width. The reference pattern is obtained for identified features of interest by applying STOCSY, which involves the computation of correlation matrices between the intensity of all composite points for a subset of samples containing the specific features of interest identified through PCA modelling, thereby reducing statistical noise from confounders.

After removal of the spectra failing to meet the criteria required by the COMPASS pipeline, for example spectra where the relevant regions were confounded by drug resonances, hippurate was quantified in 382 spectra, PAG in 267 spectra and 4-cresylsulphate in 309 spectra. Differences in concentrations between the PBC (Cohorts A and B) and control (Cohort E) and between responders and non-responders in Cohort A were calculated using Wilcoxon rank sum testing (corrected for multiple testing, Benjamini-Hochberg). Since hippurate is derived from bacteria in the proximal colon whereas protein putrefaction, which is responsible for generating PAG and 4-cresylsulphate is generated mainly in the distal colon [200], the following ratios were calculated: hippurate: 4cresylsulphate; hippurate: PAG and hipurate: (PAG + 4 cresyl sulphate).

#### 4.2.5: 16s rRNA sequencing and bioinformatics analysis

16s rRNA was carried out by Dr. Julie MacDonald, Imperial College London, on a miSeq platform using faecal sample preparation methods as previously described in Mullish et al., (2019) [201]. Briefly, DNA was extracted from 250 mg of faecal samples using PowerLyzer PowerSoil DNA Isolation Kit (Mo Bio, Carlsbad, CA, USA) as per manufacturer's instructions, with the addition of a bead-beating step. Sample libraries amplifying the V1–V2 region of the 16S rRNA gene were prepared as per Illumina's 16S Metagenomic Sequencing Library Preparation Protocol.

Samples were sequenced in a randomised order, generated in two runs on an Illumina MiSeq platform using the MiSeq Reagent Kit v3 (Illumina Inc, San Diego, USA) with paired-

end 300-bp chemistry. Data were demultiplexed, barcodes removed and FASTQ files generated using Illumina software BCL Convert.

For annotation, the “DADA2” R-package [202] was used amplicon sequence variants (ASV) generation. For each experimental batch, a quality filtering step was implemented for during primer sequence removal to avoid introducing bias [202]. ASVs from each batch were merged and chimeras removed. Taxonomy was assigned with the use of the IdTaxa function from DECIPHER R-package [203] and using the Genome Taxonomy Database release 95 (GTDB_r95) as a reference database to match to, provided in the package website (http://www2.decipher.codes/Downloads.html). To assess the potential of contaminant sequences isContaminant from decontam R-package (204) was implemented into the workflow. Twelve contaminant sequences were detected and removed, two of them commonly found in DNA reagent kits. For statistical analyses, features were filtered to remove ASVs with a zero count in more than 90 % of samples, using a geometric Bayesian multiplicative (GBM) replacement method [205] and data were centered-log-ratio (clr)-transformed. Data were then correlated against quantified NMR metabolites, using Spearman’s rank, and visualised using heatmaps which were generated for each cohort and response group.

# 4.3 RESULTS

### 4.3.1: Untargeted $^{1}$H NMR metabonomics on UK-PBC Nested Cohort Study Urine samples

Metabolic profiling of 577 human urine samples from the UK-PBC Nested Cohort was undertaken using untargeted $^{1}$H NMR spectroscopy on the spectral dataset after removal of the ethanol resonances and demonstrated some degree of separation of the control cohort (E) from the PBC cohorts (Figure 4.3A) but this was not significant as the $Q^2Y$ model statistic was negative. From the loadings plot the main differentiating metabolites were PAG, hippurate and the bile acid region (δ 0.59-0.75). No differentiation according to whether the participants were responders or non-responders was evident (Figure 4.3B).

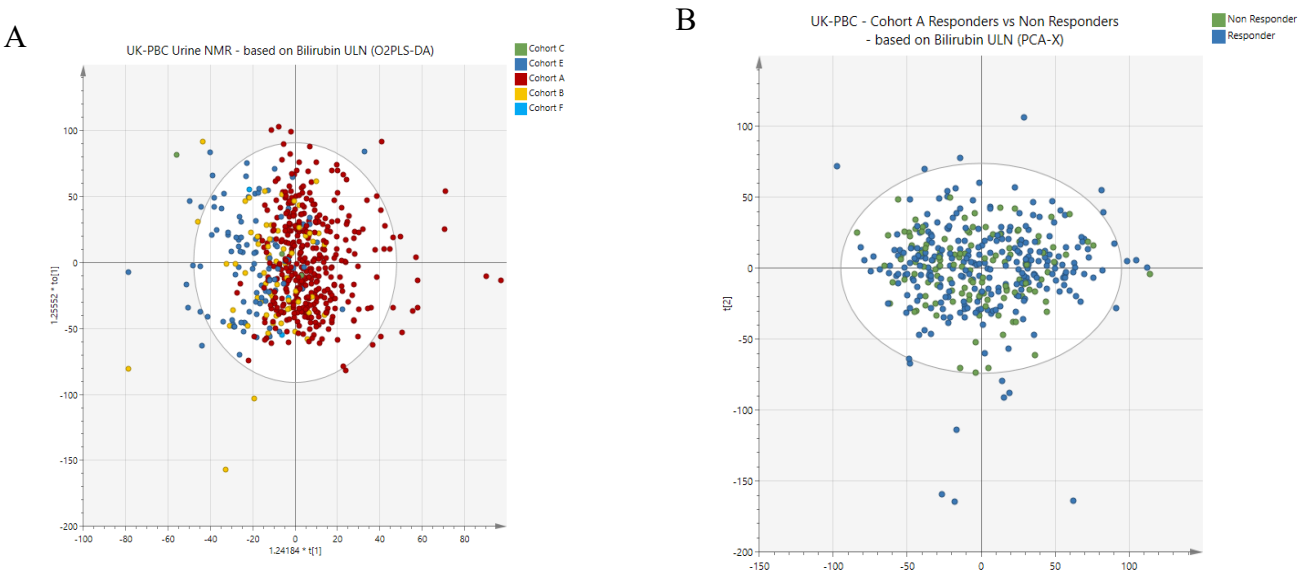
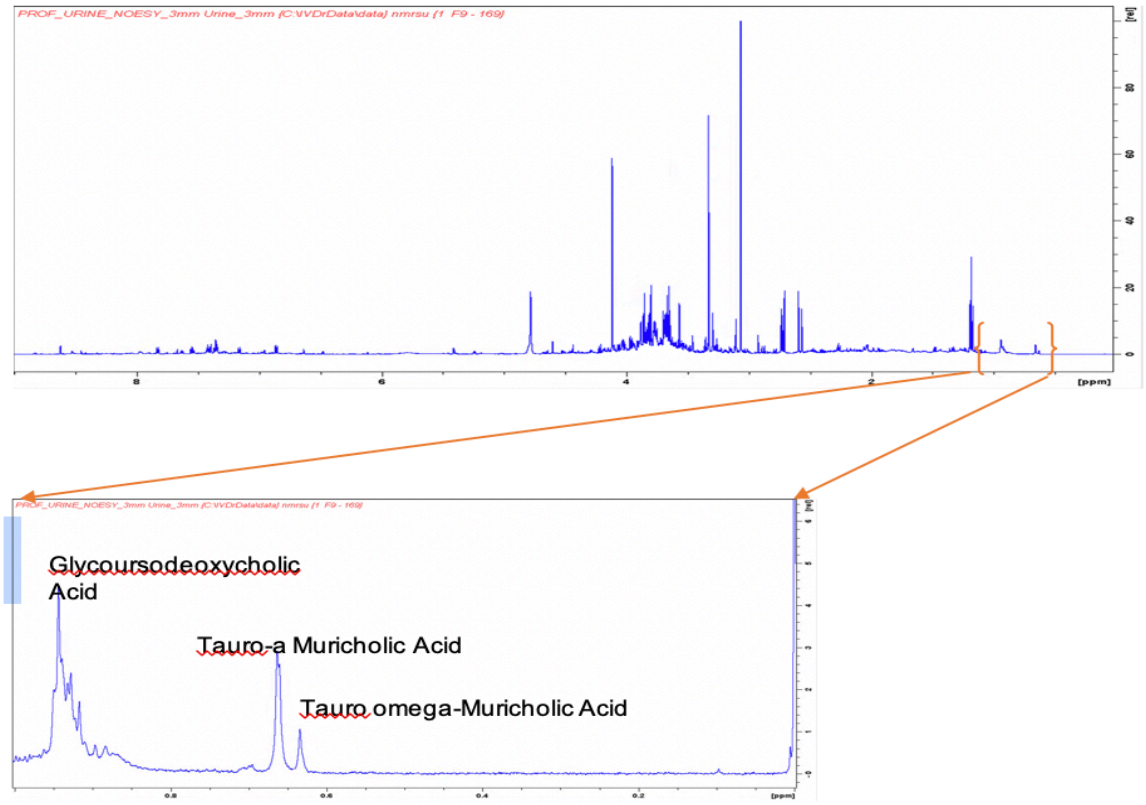



**Figure 4.3: Urine NMR multivariate analysis following removal of detected Ethanol resonances and regression analysis using Bilirubin as a regressor.** A) Orthogonal 2 Partial Least Squares Discriminant Analysis (O2PLS-DA) on all cohorts within the UK-PBC Nested Cohort based on a Bilirubin ULN of +/- 1.2 mg/dL. $R^2X$ = 0.007, $R^2Y$ = 0.05, $Q^2Y$ = -0.323 B) PCA on Cohort A UDCA response based on a Bilirubin ULN +/- 1.2 mg/dL.

The data were compromised by positional variation for several peaks, indicative of pH differences or molecular binding flux, and thus deconvolution of the data by standard methods were not sufficient to clear clustering of patient metabolomes. Since Cohorts C and F contained only 3 samples each and were collected at different clinics from the other cohorts, the analysis was focussed on cohorts A, B and E to provide sufficient statistical power for testing.

With respect to the bile acids, again a marked difference in bile acid excretion across the cohort was observed (Figure 4.4), mainly in the participants with PBC, as would be expected. Although Cohort A were treated with ursodeoxycholic acid (UDCA), other bile acids were equally as prominent as metabolites of UDCA such as glycoursodeoxycholic acid, the main urinary metabolite of UDCA and no significant PLS-DA model identifying differentiation of responders and non-responders in cohort A was found. PCA and OPLS-DA models run on a smaller region containing the C18 axial bile acid signals indicated that there was some degree of separation according to cohort and between responders and non-responders. As before, however, the models were weak or not significant (Figure 4.5) reflecting the massive variation observed throughout the whole study. Given the spectral complexity and the generally low urinary concentrations of the bile acids, a mass spectrometry-based method was developed to quantify the bile acids and is described in appendix 2 and yielded a similarly complex picture. The analysis of this data, however, is beyond the scope of the current thesis.

A




B

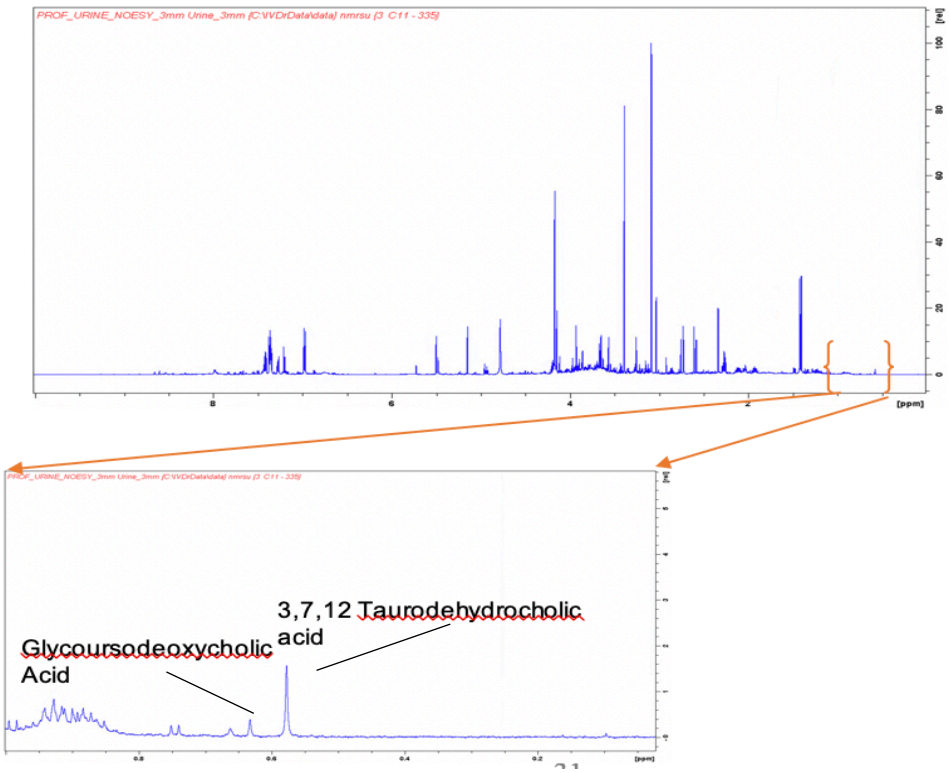



**Figure 4.4: Exemplar spectra with magnified region containing the C18 axial signals from bile acids.** A) UDCA Responder. B) UDCA Non-Responder.

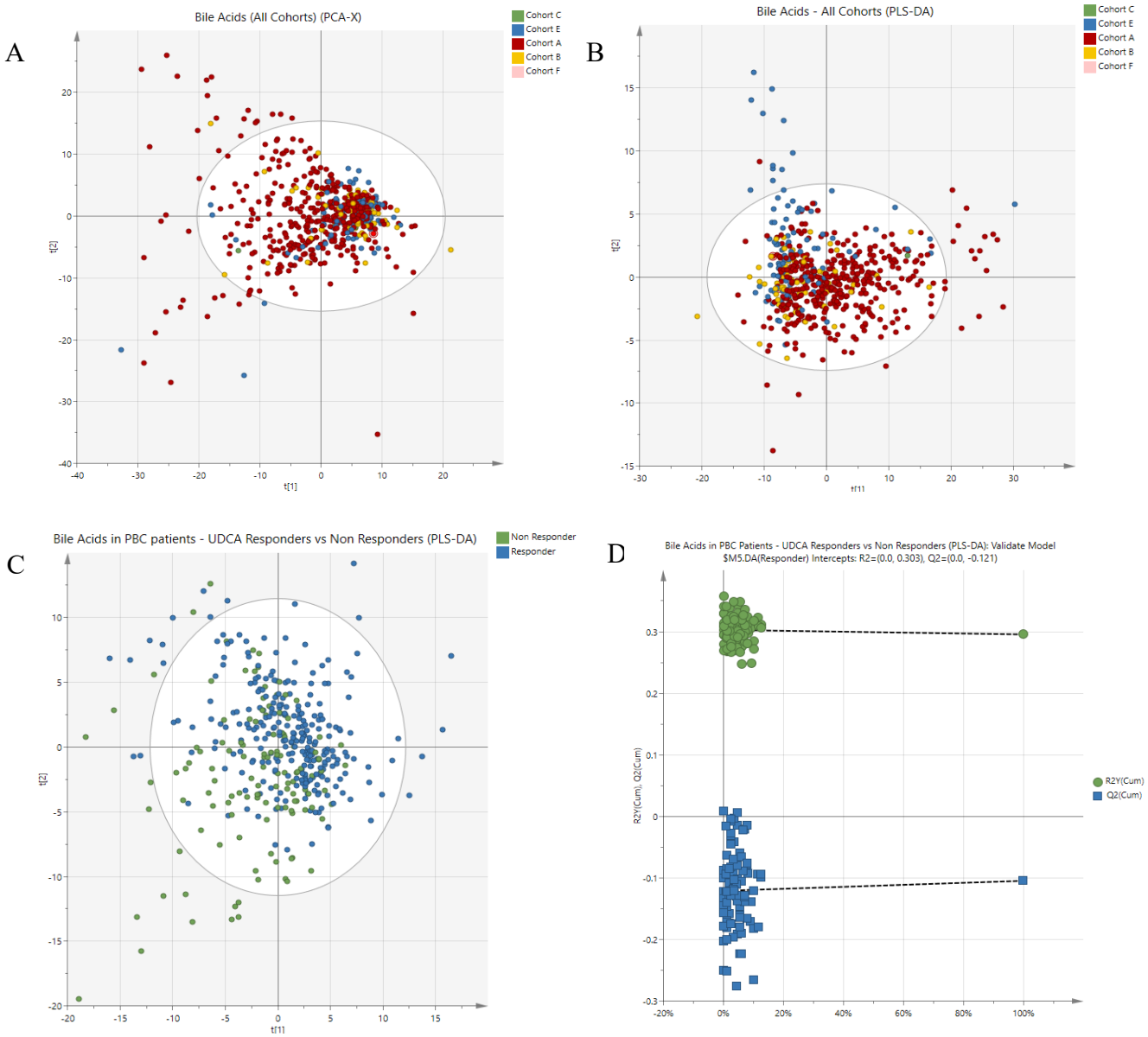



**Figure 4.5: Bile Acid region- supervised and unsupervised analysis.** A) PCA of all cohorts based on Bile Acid defined region. B) Partial Least Squares Discriminant Analysis (PLS-DA) on all cohorts based on an ALP ULN +/- 1.6µkat/L. C) Partial Least Squares Discriminant Analysis (PLS-DA) on UDCA response within Cohort A based on an ALP ULN +/- 1.6µkat/L. D) Permutation testing for model validation of PLS-DA (C).

The biological complexity of the study was compounded by variation in the condition of the samples. The attempt to overcome the issue of ethanol contamination was partially successful but other inherent technical problems in the data prevented a robust analysis.

For example, a high level of chemical shift variation was observed following examinations of the loadings for certain peaks. An example of this problem is illustrated for creatinine (Figure 4.6) and is likely to reflect either disparity in the pH of samples or large differences in osmolality driven by the overwhelming concentration of the ethanol signal.

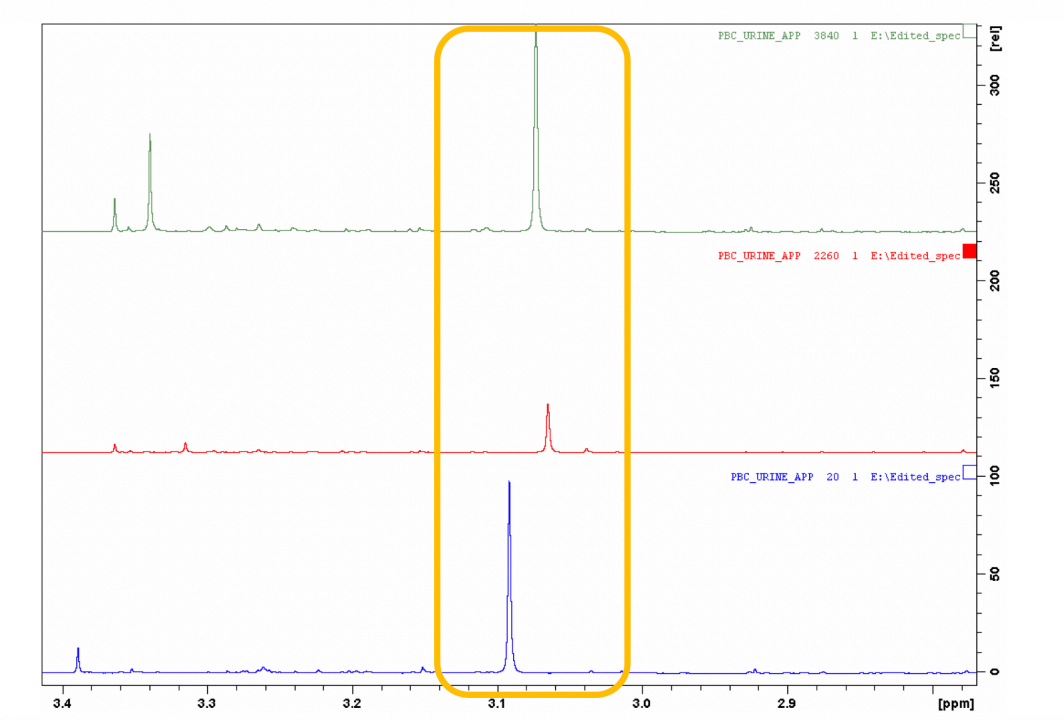



**Figure 4.6:** Example of highly variable chemical shifts observed in urine NMR spectra. Shift amongst 3 urine samples shown for assigned creatinine.

In order to simplify the spectral data and to minimize the artefacts in modelling caused by chemical shift, a multiblock approach was used to focus on sub-regions of the NMR urine spectra and PCA was performed on spectral regions of 0.5ppm. Some suggestion of clustering for a subset of the control group (Cohort E) was identified in the PCA scores plots for the regions the regions 7.0-7.5 and 7.5-8.0 ppm (Figure 4.7). As can be seen the cluster of control samples on the far left of the scores plot, indicated in blue contained higher levels of hippurate than the rest of the spectra, whereas samples on the right-hand side of the scores plot contained more phenylacetylglutamine (PAG) and relatively less hippurate. Since

the remainder of the control and PBC group were heavily overlapped, with no visible disease-related clustering in the first three components, no supervised models were calculated.

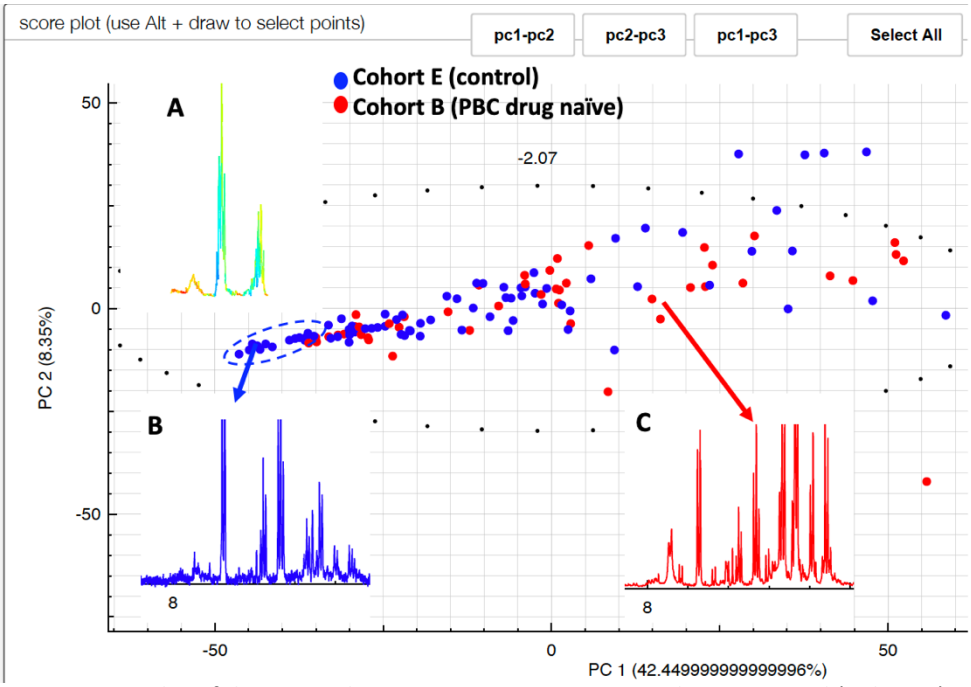



**Figure 4.7:** PCA scores plot of the spectral region 7.5-8.0 ppm generated using control (Cohort E) and drug naïve PBC (cohort B) urine samples showing clustering of a subset of the control samples that contain relatively higher levels of hippurate.

For cohort A, multiblock PCA models indicated some patterns within three main regions: 6.5-7.0 ppm, 7.0-7.5 ppm and 7.5-8.0ppm, corresponding to variance driven by a subset of samples mainly comprised of responders to UDCA. The three regions were combined and remodelled using PCA (Figure 4.8). The loadings indicated that the main difference was driven by higher levels on PAG and hippurate in these samples. However, similar to the cohort comparison, the majority of the samples overlapped with no group differentiation based on response and no further models were generated. The identification of the three main gut-microbial metabolites observed in urine is shown in Figure 4.11.

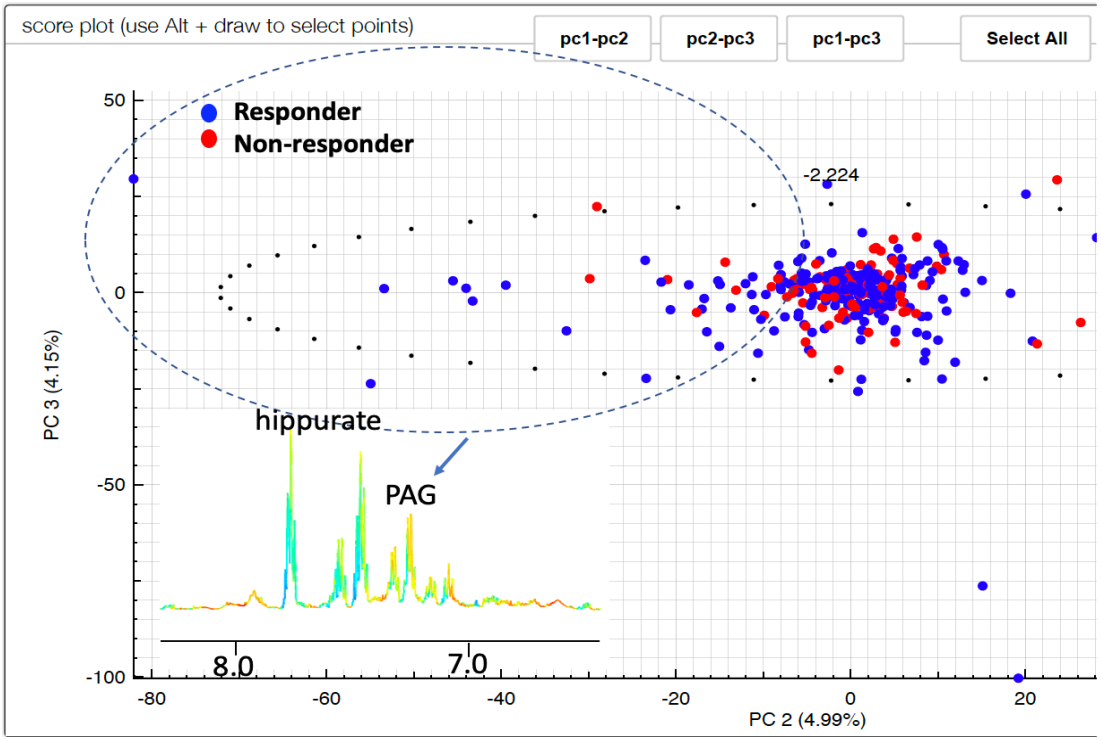



**Figure 4.8:** PCA scores plot of the spectral region 6.5-8.0 ppm generated from the Cohort A containing urine samples from patients with PBC treated with UDCA. Patients who responded positively to therapy are denoted in blue and non-responders are plotted in red.

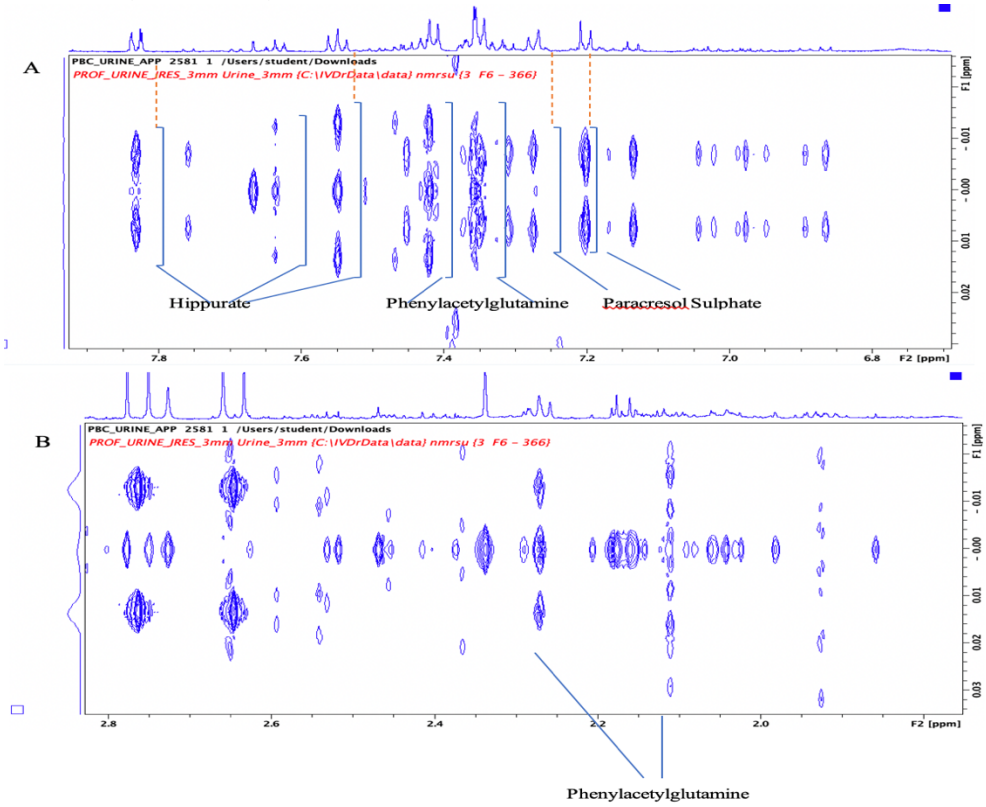



**Figure 4.9:** Exemplar $^{1}$H J-Resolved NMR urine spectrum showing assignment for hippurate, PAG and 4-cresylsulphate.

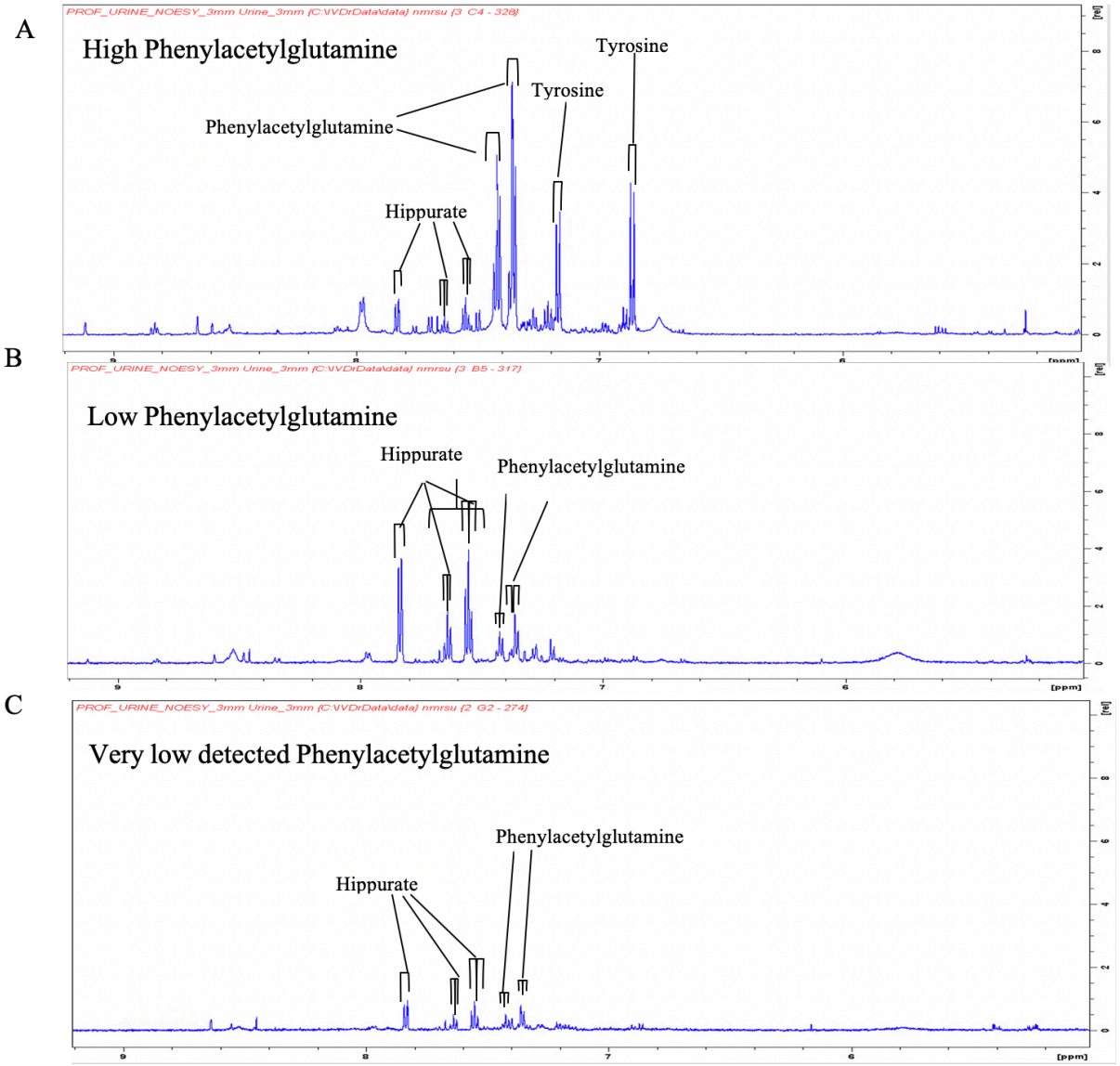



**Figure 4.10: Exemplar spectra showing variable levels of phenylacetylglutamine from the aromatic (high frequency) region $\delta$5.5-9.0.** A) Exemplar spectra containing high phenylacetylglutamine. B) Exemplar spectra containing low phenylacetylglutamine. C) Exemplar spectra containing very low phenylacetylglutamine.

The variance of PAG across the study set was extremely high (Figure 4.10) with some samples containing almost none and others containing levels amounting to 5-10 mM (as calculated relative to the TSP concentration).

### 4.3.2: Targeted $^{1}$H NMR profiling of gut microbial metabolites indicated as different between the PBC and control participants.

Based on the indication of their potential trend in differentiating between responders and non-responders in PBC patients treated with UDCA in the multiblock PCA analysis, the relative estimated concentrations for hippurate, PAG, and 4-cresylsulphate were calculated by the COMPASS tool as reported by Loo et al [199]. COMPASS calculates likelihood of a sample containing a given compound based on the pattern of signals defined by a reference spectrum and assigns an arbitrary concentration value if the metabolite is deemed to be present. For each spectrum a code is given according to the correlation coefficient. For 4-cresylsulphate: green = cc>6.0; amber = cc<0.6>0.5; red= cc<0.5. An example of spectra falling into each of the three categories is given for cresol (Figure 4.11). As can be seen, the bottom trace shows a green code where there is a high degree of confidence in the metabolite assignment and the arbitrary intensity is around 120 units, whereas the amber trace shows a set of signals with an authentic pattern but a lower intensity. The red trace is confounded by the presence of signals from other molecules and if 4-cresyl sulphate is present, it cannot be detected or measured with certainty. Therefore, only spectra containing green or amber (all confirmed via visual inspection) signals were included for quantification. Using this system, the extracted statistics (Table 4.3) indicated that hippurate was best fitted with 83.7 of the spectra returning integral values, whereas substantial peak overlap and chemical shift impeded the fitting of both PAG and 4-cresylsulphate, for which only 55.5 and 64.7% of samples respectively were integrated successfully.

**Table 4.3: Number and percentage of samples quantified for the three selected gut microbial metabolites**

| Cohort | Number (percentage) of samples with hippurate | Number (percentage) of samples with phenylacetylglutamine | Number (percentage) of samples with 4-cresyl sulphate |
|---|---|---|---|
| **A** | 270 (83.9%) | 197 (61.2%) | 224 (69.6) |
| **B** | 40 (78.4%) | 22 (43.1%) | 27 (52.9%) |
| **E** | 72 (88.9%) | 48 (59.3%) | 58 (71.6%) |

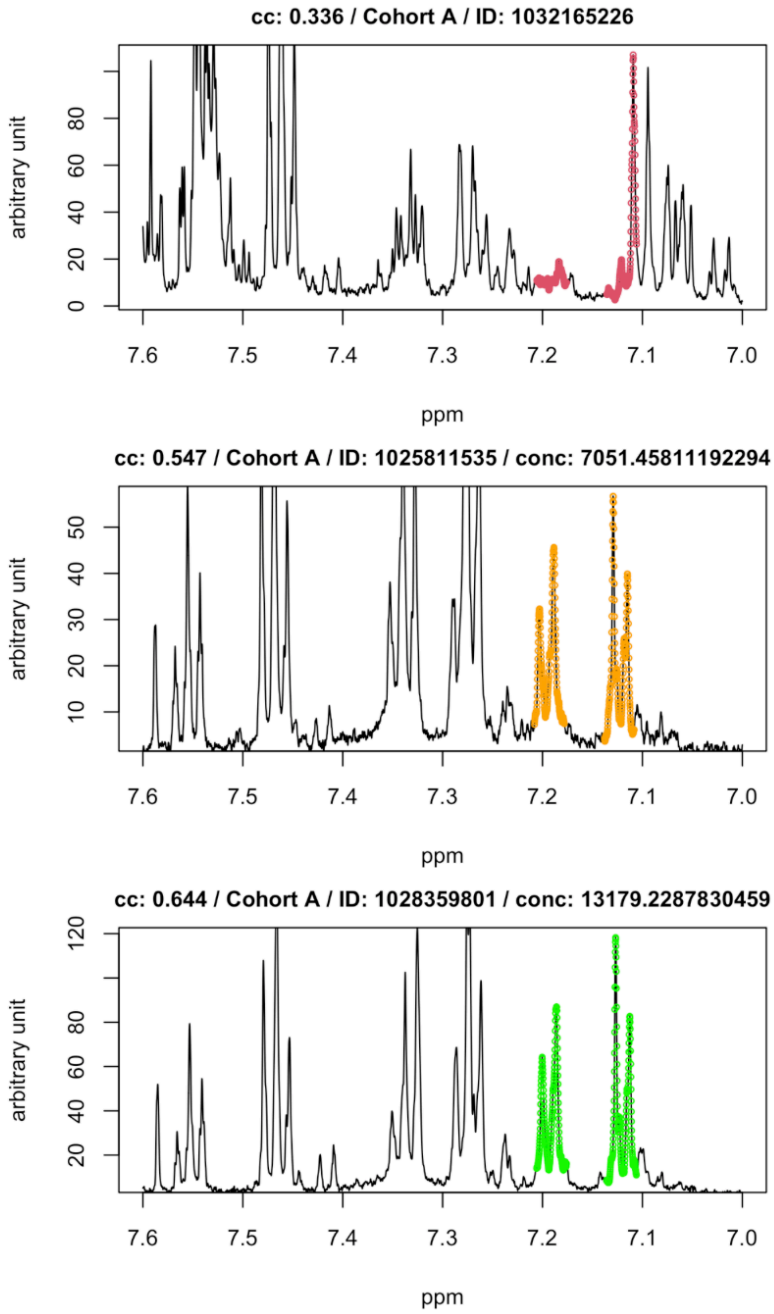



**Figure 4.11:** Signals relating to 4-cresylsulphate coded according to the degree of certainty of their presence in the spectrum based on the correlation coefficient with an authentic reference pattern.

The Wilcoxon Rank sum index integration of the area by cross-correlation (CC) with the reference pattern for PBC cohorts (Cohort A and B) compared to healthy controls (Cohort E) is shown in Figure 4.12. None of the comparisons were found to be statistically significant.

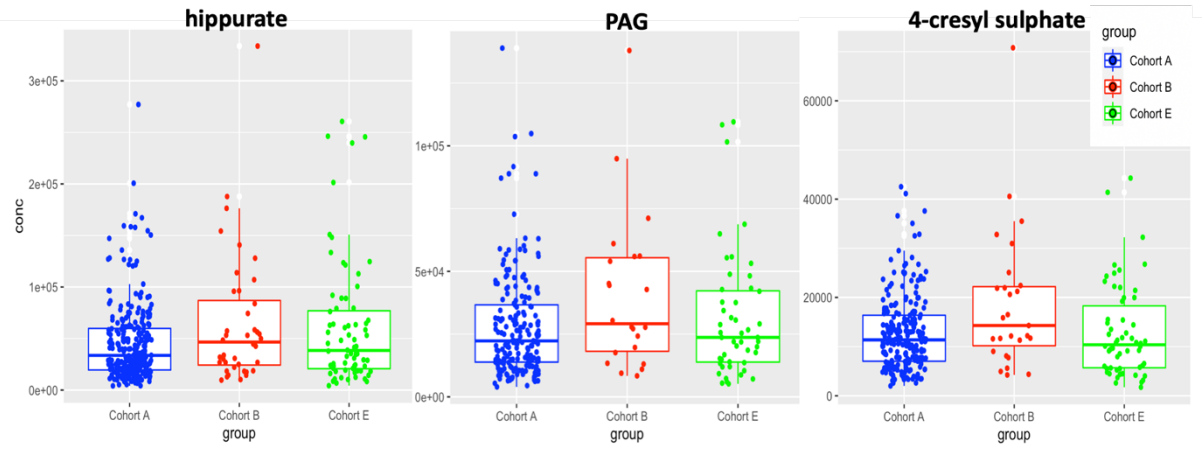



**Figure 4.12: Relative integrals of Hippurate, PAG and 4-cresyl sulphate stratified by cohort. No significant differences were found between PBC disease groups drug naïve (Cohort B) or UDCA treated (Cohort A) and the control group (Cohort E)**

Although no significant difference for any of the three quantified gut microbial metabolites were found for responders and non-responders to UDCA in cohort A. However, consistent with the clustering pattern in the multiblock PCA scores plot (Figure 4.11), a trend towards higher relative concentrations of PAG and Hippurate was noted in the responders (Figure 4.13). Based on the clustering patterns in the multiblock PCA and on the calculated integral distributions, it would suggest that there may be a subset of responders who have a distinct urinary gut microbial profile.

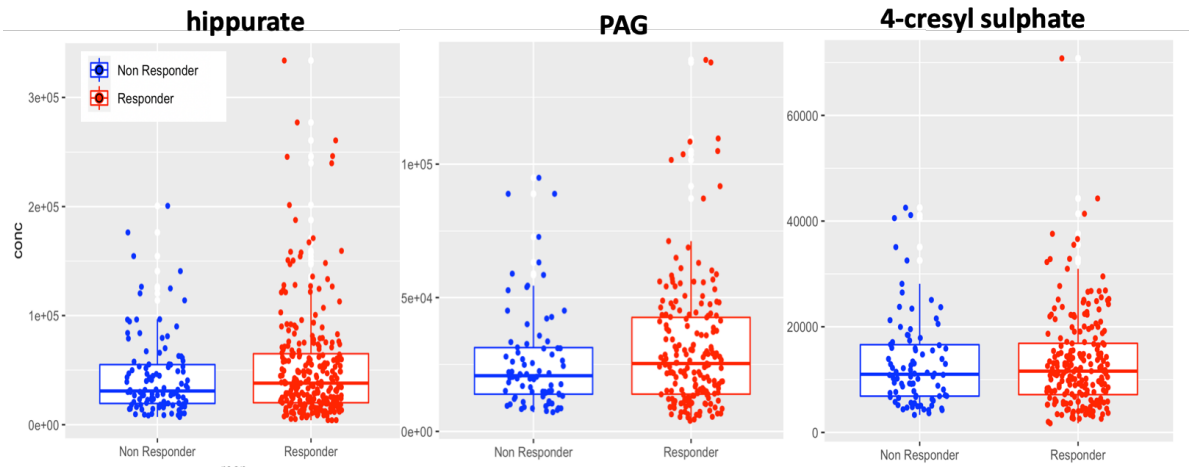



**Figure 4.13: Relative spectral integrals of Hippurate, PAG and 4-cresyl sulphate in Cohort A, stratified by responder / non-responder. No significant differences were found between any group comparison.**

### 4.3.3 Influence of the gut microbiome on urinary gut bacterial metabolites for PBC versus control.

The 16S rRNA derived data were correlated with the urinary gut bacterial metabolites stratified according to cohort (figure 4.14) and responder status (figure 4.15). Of the three sets of metabolite-microbiome ASV correlations, Cohort B consisting of treatment-naive PVC participants was the most distinct and showed the most significant correlations. Of note, strong correlations between several Bacteroidetes and the 4-cresylsulphate and PAG to Hippurate ratio was apparent indicating that Hippurate was driving the association with bacteroides species. In the PBC drug naïve cohort (Cohort B), all three metabolites were found to be strongly inversely associated with Coreobacteriales and some Clostridium classes.

Similarly, the responder and non-responder subsets for Cohort A demonstrated contrasting behaviour in terms of their correlation patterns with the microbial parameters. Specifically, more anti-correlation was observed in non-responders with inverse patterns observed in

responders, between Bacteriodales species and 4-cresylsuphate, PAG and hippurate. Interestingly, strong positive correlation is observed between Clostridium species and the NMR metabolites in non-responders. From general observation, correlation patterns for the Bacteriodales species differs greatly within each response criteria, but with a similar correlation pattern observed for each 3 metabolites. The bacterial profiles were also regressed against the ratios hippurate: 4-cresylsulphate; hippurate: PAG and hipurate: (PAG + 4 cresyl sulphate). The correlation between metabolites is again thought to be driven by hippurate in responders, with strong anti-correlation observed between hippurate and calculated ratios.

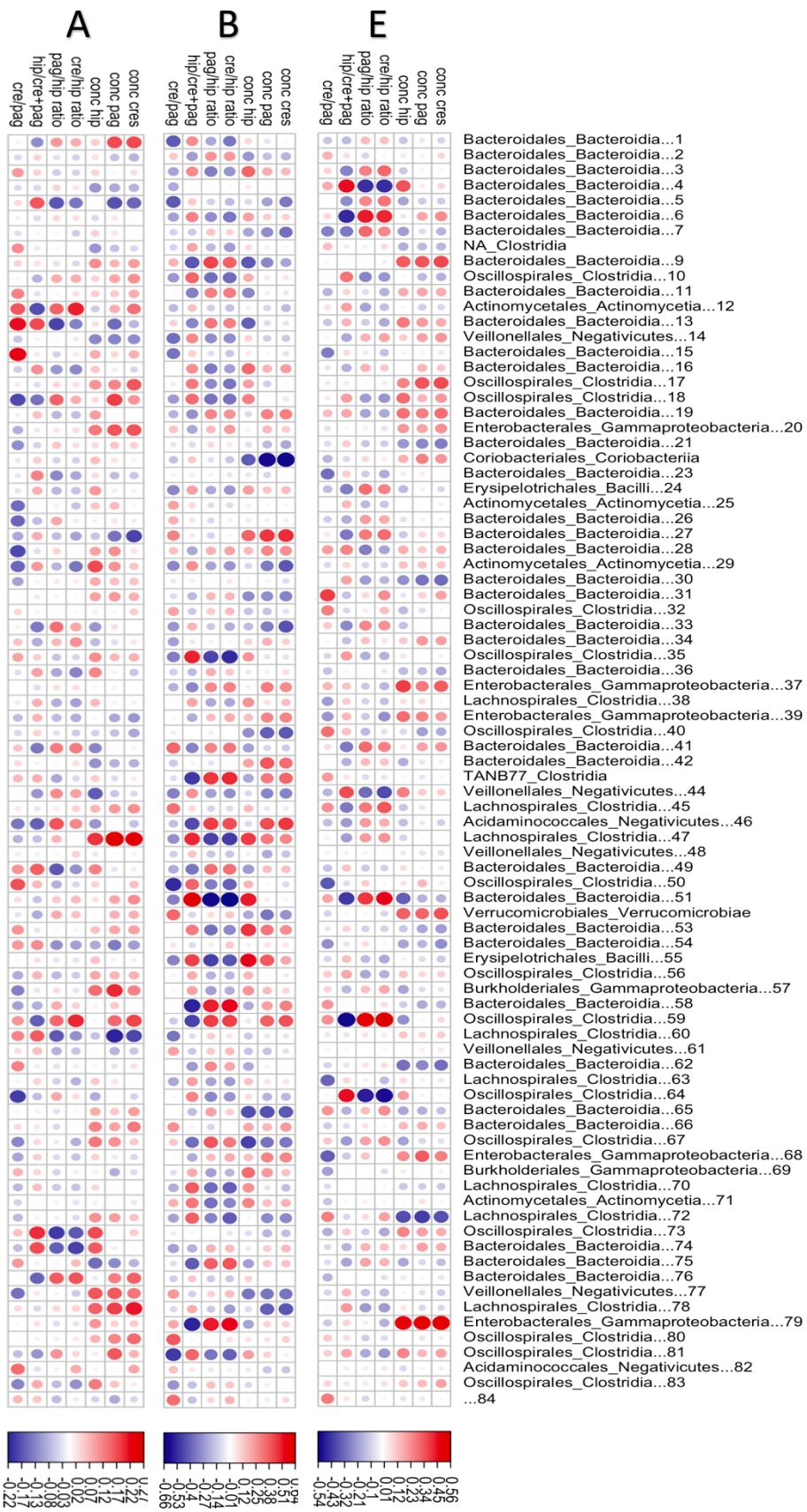



**Figure 4.14:** Correlation matrix between Hippurate, PAG, 4-cresylsulphate and derived ratios and the microbial parameters for each of Cohort A (PBC treated), Cohort B (PBC naïve) and E (controls).

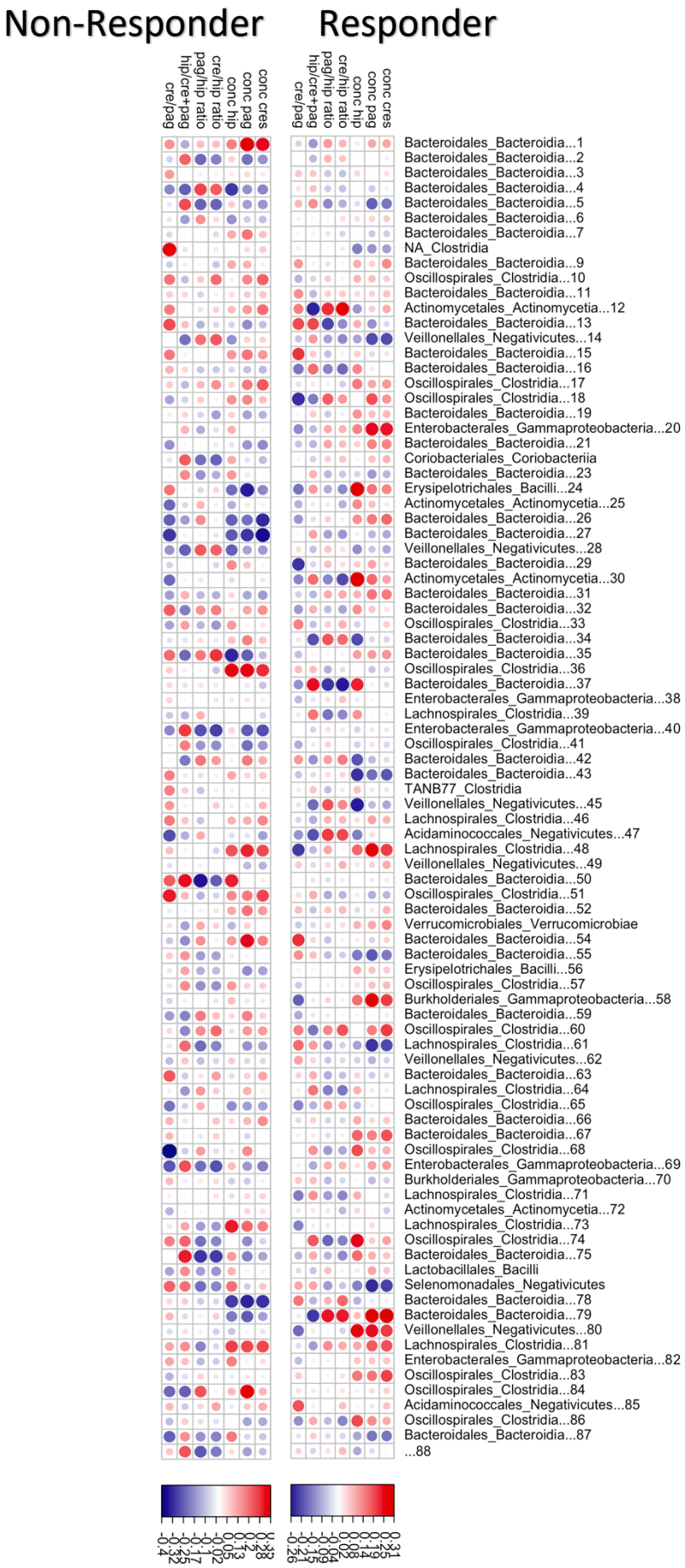



**Figure 4.15:** Correlation matrix based on UDCA response (Cohort A) showing associations between Hippurate, PAG, 4-cresylsulphate and derived ratios and the microbial parameters, stratified according to response to UDCA treatment.

# 4.4 **DISCUSSION**

### 4.4.1 Strategies for mitigating the analytically and biologically confounded data.

Due to extremely high biological and analytical variation, the data were difficult to analyse. The identification of high concentrations of ethanol, assumed to be introduced at the sample collection stage (see results) necessitated the use of modified spectra. There was a high degree of confounding variation due to the wide array of drug metabolites encountered and the substantial variation in signal position made the normal pipeline utilised in metabonomic workflows unusable. Therefore, strategies for overcoming this high degree of confounding variation were introduced to accommodate the flaws in the data. Imperfect data sets, or those poorly adapted for specific analysis types are often encountered in studies that are inherited subsequent to the experimental design stage. Two of the main strategies for analysis of this confounded data used in this study were use of the COMPASS pipeline to overcome the positional shift and identify the correct resonances, and hence metabolites for quantification and subdivision of spectra into smaller regions in order to reduce the impact of variation encountered using a larger spectral range.

A high level of ethanol was detected in 76.6% of all samples (Table 4.2), with no correlation found regarding collection from specific research centres. However, with no ethanol detected in the blank samples, it is plausible that samples were contaminated with a solvent introduced at the sample collection phase, as no ethanol is used in NMR sample preparation negating introduction at this phase. The ethanol is assumed to be exogenous due to the

extremely high concentrations present in some samples, that are incompatible with viable physiological levels, with concentrations as high as 300mMol detected in some samples. A high degree of chemical shift variation of the ethanol resonances, and others, was also observed in analysis, following sample pre-processing. This, in addition to the high level of ethanol, masked any pattern attributable to disease or treatment response in the unsupervised principal components analysis, with ethanol found as the highest discriminant metabolite in any preliminary multivariate analysis models.

For a small number of patients, lower levels of ethanol were observed, in addition to resonances from ethyl glucuronide, a by-product of alcohol metabolism. In this instance, both ethanol and ethyl glucuronide were present in levels approximately 0.5 to 5 mMol, indicating that the presence was likely from alcohol consumption. An additional component to consider is the large discrepancy between patient declaration of alcohol consumption and detection of ethanol resonances, with approximately 25-50% under reporting across all cohorts (table 2), despite the detection of ethyl glucuronide in some instances.

A further element attributing to the failure to observe differences attributable to disease presence or treatment response is the intake of an amalgamation of drugs, with many patients suffering from multiple co-morbidities and excreting metabolites from a broad combination of drugs. Very little pattern was seen in terms of differential drug intake and multivariate analysis.

Despite the analytical limitations of this study, there was some evidence that some of the urinary microbial components and bile acids differed between the healthy state and PBC and that these metabolites were altered between drug naïve and treated participants.

Although no statistically significant differentiation according to disease or response to UDCA therapy was found in the data, there was a clear trend towards subclusters of samples in both comparison that had distinct metabolic profiles driven by the three main observable gut-microbial metabolites found in humans: Hippurate, PAG and 4-cresyl sulphate.

### 4.4.2 Dysregulation of bile acids in PBC patients

Statistical regression modelling, conducted on the bile acid region only (∂ 0.6 – 1.0), identified clustering of disease versus healthy controls (figure 4.3). Bile Acids are crucial regulatory components of liver function and are closely involved with the regulation of glucose, lipid, cholesterol and drug metabolism [198]. In healthy individuals, through the enterohepatic circulation of bile acids, approximately 95% of bile acids are reabsorbed and 5% excreted with faeces and a negligible proportion through urine [206]. The detection of bile acids in urine, is often associated with severe liver dysfunction, with some studies suggesting that urinary bile acid concentrations are more reliable predictors of clinical hepatobiliary disease outcomes due to stability across confounding variables such as BMI and age [206].

Several peaks in the ∂ 0.5 – 1.0 region were identified as bile acids. Through visual inspection of spectral profiles, bile acids were detected in approximately 37.2% of Cohort A

(UDCA treated), with 12.5% found to have very high concentrations of bile acids. Comparatively, only 18.7% of Cohort B (UDCA treatment naïve) samples had bile acids detected, with high concentrations in 6.3% of samples. In healthy controls (Cohort E), bile acids were detected in 8.9% of samples, but with no high concentrations seen. Since the treatment given for PBC to participants in Cohort A was itself a bile acid, UDCA, the spectra were assessed as to whether the observed bile acid signals were from UDCA or any of its known urinary metabolites. Some samples in a subset of the Cohort A participants were found to contain high concentrations of the UDCA metabolite, glycoursodeoxycholic acid. However, interestingly, neither UDCA nor its conjugates were directly observed in all samples in Cohort A and the bile acid signature across participants was variable suggesting that the bile acids in this case were reflective of the disease rather than the therapy. The main bile acids noted were taurine conjugated bile acids including 3,7,12-tauro dehydrocholic acid and tauro-a muricholic acid as confirmed by mass spectrometry analysis (see Appendix 2). Tauro-conjugated bile acids have been shown to vary according to microbial composition, with germ-free animals excreting high levels, manifesting in the heart, plasma and liver concentrations of tauro-conjugated bile acids [207]. Prior studies investigating the impact of alcoholic cirrhosis found similar trends in higher levels of taurine conjugated bile acids, in comparison to the detected glycine-conjugated forms, noting that this increase is associated with dysregulation of homocysteine metabolism, resulting in an increase in hepatic CYP7B1 expression in the bile acid synthesis pathway and taurine concentration [208].

### 4.4.3 PBC-associated disruption of the gut microbiome

Although neither the metabolome nor the microbiome were statistically significantly different between PBC patients and controls, there were trends in both datasets that were indicative of gut microbial disruption in subsets of patients, particularly with respect to response/ non-response to UDCA treatment. For example, multiblock PCA was able to identify a subset of participants in the responder group of Cohort A with higher levels of PAG and Hippurate. Studies have suggested that an altered gut microbiome due to variable xenobiotic interventions could attribute to differential PBC treatment response [209]. Although PAG is predominantly a microbial metabolite, it is also an indicator of ammonia excretion and has found to regulate the phenylalanine/tyrosine pathway in liver dysfunction patients, which provides support the results found in this analysis, with high levels of Phenylacetylglutamine found in UDCA responsive PBC patients [210]. The farnesoid X receptor (FXR) regulates amino acid catabolism and detoxification of ammonia via ureagenesis and glutamine synthesis, with urea cycle dysfunction and hyperammonemia common in patients with cholestatic and chronic liver diseases [211]. Phenylacetylglutamine is involved with ammonia excretion; Phenylacetyl and Phenylbutyrate, through the conjugation to glutamine, provide an alternative pathway for nitrogen disposal via urinary excretion of Phenylacetylglutamine, leading to a net ammonia removal in urine through the prevention of glutamine being re-metabolised to ammonia [212].

Given the associated trend in hippurate perturbation, it is most likely that the higher concentrations of PAG in the responder subset are driven predominantly by gut microbial perturbation. The observed trend of elevated levels of hippurate and PAG in UDCA

responders was further emphasized by the marked difference in the correlation patterns between the 16S rRNA microbial and gut-microbial data for the responder and non-responder groups. Phenylacetylglutamine concentration, when normalised to creatinine and the abundance of several bacterial genera, most specifically several unclassified *Clostridiales* species found to be discriminatory bacterium between PBC patients (Cohort A and B) and healthy controls (C, E and F). Interestingly, the two discriminatory microbial phyla distinguishing UDCA responsiveness are *Firmicutes* and *Bacteroidetes,* phyla which represent up to 99% of the human microbiome. It has been well established that the large variance of both is seen in association with aging, with the *Firmicutes* to *Bacteriodetes* ratio reducing from 10.9 to 0.6, in the middle-aged to the elderly [213]. Studies have suggested that in the presences of a low *Firmicutes* to *Bacteriodetes* species such as *Clostridium* have the opportunity to increase in abundance, which corresponds with the results found in this study, suggesting an abundance of *Clostridium* species associated with the presence of low levels of Phenylacetylglutamine.

Prior studies investigating differences in the gut microbiome of PBC patients have noted significant differences in microbial diversity between patients. The significant impact of UDCA on abundance of differential microbial species in PBC patients was also highlighted, finding that following 6 months of UDCA treatment that the abundance of the species strongly correlated with PBC, differences were reversed [214]. Previous studies have also found that *Bacteriodetes* and *Clostridium* rich microbiomes have been found to alter the synthesis of gut bacterial metabolites [210]. UDCA reduces the concentration of phenylalanine/tyrosine and tryptophan metabolites [210], which may explain the high concentration of cresylsulphate (figure 4.13), a gut microbiome metabolite, in some UDCA

non-responsive PBC patients. However, the extent to which the gut microbiome controls and impacts the synthesis and metabolism of aromatic amino acids, remains unclear and unelucidated.

Findings from Pallister et al., suggest that cresylsulphate and phenylacetylglutamine are potentially toxic uremic by-products, produced via the undigested dietary proteins by colonic bacteria [215]. Studies have shown that higher cresylsulphate and phenylacetylglutamine are strongly linked to early kidney dysfunction [216] and thus could be associated with poor responsiveness in UDCA non-responders, who have a higher 4-cresyl sulphate/phenylacetylglutamine ratio compared to UDCA responders.

### 4.4.4 Analytical Limitations

The NMR Urine analysis was primarily limited by the large concentration of ethanol detected in over 80% of samples. This may have further exacerbated the broad chemical shift variation seen for the majority of detected peaks, and thus negating the use of any automated peak detection algorithms due to the large overlap of peaks which would have unavoidably led to incorrect assignment of metabolites. For this reason, visual inspection and manual metabolite profiling were deemed the most suitable and robust method for analysis. Despite this, the metabolic profiles of PBC vs healthy controls were very apparent on visual inspection, with high concentrations of phenolics and aromatics, an observation which may have been lost with the application of traditional methods of analysis. Through the analytical focus on key associated components of liver disease, such the bile acid region, visual profiling has enabled a more detailed and thorough analysis of individual samples,

enabling the detection of components that may have been overlooked by automated algorithms.

Due to the multiple co-morbidities associated with PBC, a large proportion of patients are prescribed an amalgamation of drugs. Several large drug signals overlapped a significant proportion of the spectra acquired, with paracetamol and ibuprofen identified in a significant proportion of patients and participants. This applied an additional layer of convolution to analysis, with drug peaks masking smaller metabolites that may be significant to the metabolic profile of PBC.

Another element to consider, is the depth of resolution provided from the 1D NMR experimental run, as samples were analysed in 3mm tubes but only at 32 spins per run, a standard resolve for 5mm tubes. For increased sensitivity, and a reduction in the signal to noise ratio, 64 spins per run would have been a more optimal resolution for this experiment and may have resolved some of the shimming issues seen with some samples due to poor sample homogeneity.

## 4.5 Conclusions

Due to the heavy ethanol contamination found in these urine spectra, traditional pattern recognition methods could not be applied. The application of a novel multiblock technique, COMPAS, previously only applied for the detection of exogenous compounds, allowed extraction of key metabolites by comparison of a reference pattern derived spatially from PCA modelling, thereby segregating the spectra according to clean / confounded making relative quantification and associations between class easier to see. Whilst neither the metabolites nor microbiome dataset generated strong associations with PBC disease or treatment in their own right, the integrated analysis generated reproducible patterns in the data corresponding to disease and response to treatment, highlighting significant gut microbiome aberrations between cohorts which are currently under further investigation.

# CHAPTER FIVE: IDENTIFICATION OF A MOLECULAR SIGNATURE REFLECTING PRIOR BURN-INJURY IN CHILDREN

The aim of this project is to investigate the molecular changes in children 3 years post-burn injury using metabolic profiling techniques. Additionally, the aim of this project is to investigate immune changes associated with metabolic changes, through the integration of immune profiling data with metabolic profiling data.

# 5.1 INTRODUCTION

### 5.1.1: Epidemiology

The Lancet Commission on Global Surgery 2030 emphasised that whilst a substantial majority of surgical care is delivered and consumed in high-income countries, nearly 60% of the burden of surgical care is required in low- and middle-income countries (LMICs) where the infrastructure and resources are broadly insufficient [217]. Burn injuries are a significant global health crisis, with non-fatal burns found to be a leading cause of morbidity due to the complex and prolonged healing process, often impacting both the physical and mental health of patients on a chronic and life-long scale [218]. The World Health Organisation (WHO) estimates that 180,000 deaths annually are related to burn injury. Whilst in high-income countries (HIC) the burn death rate has rapidly declined over the past 50 years, WHO states that over 95% burn-related deaths occur in low- and middle- income countries (LMIC), with the majority of injuries being attributed to fire-related burns from open gas stoves and hot oil scalds [219].

Despite a decline in the incidence of hospitalisations related to burn injury in developed countries, children remain a global high-risk group due to sensitivity of skin and risk-taking behaviours. Burns are the fifth most common cause of non-fatal childhood injuries, with significant regional differences in burn rates, where male children under 5 from LMICs twice as likely to die compared to their HIC counterparts. The prevalence of child-death attributed to burn injury is over 7 times higher in LMICs as opposed to HICs, in addition to burn-related injuries being amongst the leading causes for disability adjusted years lost in LMICs [219].

These factors suggest that there is a global pre-requisite for improved burns care for children.

Burn patients not only require acute primary treatment but often need subsequent rehabilitation, reconstruction and long-term anti-scar therapies. These consequences are often accompanied by additional socioeconomic burdens for both burn victims and families(6). Whilst substantial advances have been made in the management and treatment of burn injuries over the past twenty years, morbidity and mortality remain a significant obstacle in the overall improvement of primary treatment, metabolic support, and wound care management [220].

Studies have found that independent of burn severity, patients are at an increased risk for an amalgamation of non-communicable diseases, with typically higher levels of morbidity and mortality. This includes cardiovascular diseases, diabetes, cancer and mental health disorders [221]. Thus, although the primary injury is to the skin, increasing evidence suggests that burn injuries induce a substantial and chronic systemic response to the injury.

### 5.1.2: Pathogenesis and Aetiology

The structure of skin is complex, with a significant and substantial immune and vascular schematic, presenting a multifaceted organ beyond the epidermis (figure 5.1). The current physiological understanding of burn injuries cites three predominant regions of compromisation. The region of coagulation is the first external layer; the tissue damaged at the time of injury; this is surrounded by the zone of stasis; which is characterised by

decreased tissue perfusion. The progression of damage follows the zone of stasis and with the tissue progressing towards necrosis within the first 48 hours following injury. The consequence of such, is that the site of injury expands in area and depth with a broad recovery and immune response [222]. Burn injuries can induce a systemic global response beyond this site of injury, with profound and substantial physiological changes, such as increased muscle and bone catabolism, hepatic steatosis, increased infection susceptibility and multiple organ dysfunction [223].

Whilst prevention of these injuries and the implementation of safety measures and regulations are key to aversion, the role of medical treatment must also be considered. The complex treatment of burn injuries encompasses a broad spectrum, largely dependent on severity of burn and location; with treatments such as early excision and skin grafting, although effective, also carrying high infection risk to the patient [224].

One of the most crucial properties of improved surgical interventions is the reduction of scarring and the reduction in the loss of healthy tissue in the removal of debrided tissue(10). Surgical debridement is the removal  remains the gold standard for the excision of burn tissue, but still carries significant risk, with common side effects associated with the method including incomplete and uneven debridement leading to multiple excisions, prolonged surgical operative time and increased loss of blood [224]. However, the method is the most effective in decreasing the bacterial load that can accrue as a result of the opportunistic environment the necrotic tissue provides (skin permeability), which can often lead to reduced healing time, physiological impairment and functional impairments [224, 225]. Currently the excision procedure is determined via the operating surgeon, with dermal

preservation considered as the prime objective, achieved by layer-by-layer removal of debrided tissue until healthy tissue is reached. Whilst this method has been found to improve overall morbidity and mortality, the outcome is very much operator dependent, with no current protocols available for the determination of how much tissue should be excised [224, 225].

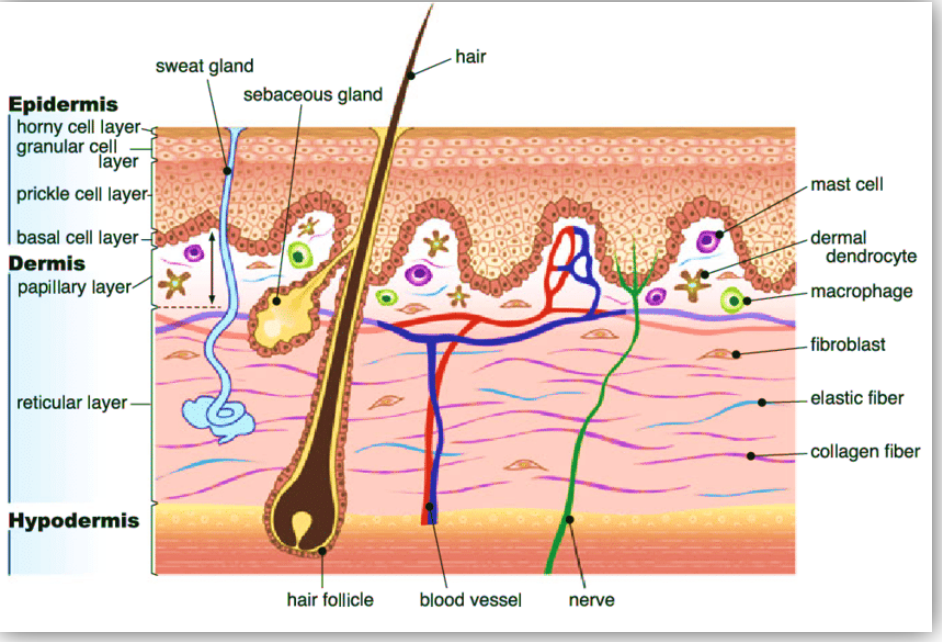



**Figure 5.1: Structural framework with skin and underlying immune and vascular properties. Image source:** Reproduced from: Kawasumi A, Sagawa N, Hayashi S, Yokoyama H, Tamura K. Wound healing in mammals and amphibians: toward limb regeneration in mammals. New Perspectives in Regeneration. 2012:33-49.

### 5.1.3: Differential metabolic rates in children

Adults with a thermal burn injury of 25% total body surface area (TBSA), have a predicted metabolic rate of between 118-210%, and in children, this is thought to reach up to 200% [226] . In a study of 977 paediatric burns patients with burns >30% TBSA, it was found that profound metabolic changes from body composition to cardiac and organ function showed

markedly different alterations for up to 3 years post burn, as well as demonstrating significantly different and a prolonged hypermetabolism. A sustained hyperinflammatory state was also found with significantly elevated levels of cortisol cytokines and acute phase proteins, typically first seen in the first 48 hours post-injury, sustained for up to 3 years [226]. Abnormal insulin sensitivity has been well documented in paediatric burns patients, in addition to aberrant levels of urinary cortisol and catecholamines, serum IL-7, IL-10, IL-12, MIP-1b and MCP-1 have been found to be significantly increased for 3 years post-burn [227]. Whilst insulin sensitivity has been shown to be abnormal, insulinogenic indices display physiological levels, indicating normal pancreatic beta-cell function and a more complex underlying mechanism with regards to insulin sensitivity in paediatric burns patients [228]. It is thought that this is driven by augmented hepatic glucose production, through the increased adipose tissue lipolysis and skeletal muscle proteolysis. As a result, gluconeogenic substrates are also increased, however the full mechanism has yet to be elucidated [228].

The ability of the body to autonomously maintain thermoregulation is mediated by metabolic activity and sweating. This ability is depleted in severe burn patients, and thus the post-burn hypermetabolic response attempts to sequester and compensate for the water and heat loss. It is estimated that 4 litres of water per square meter of burn injury is lost per day, in comparison to 2.5 litres in a healthy control [229]. The physiological response to mediate severe burn injury raises core and skin temperatures 2°C above normal, driven by ATP consumption and substrate oxidation. Studies have shown that the hypermetabolic state can be reduced by increasing ambient temperatures to 33 °C, a temperature considered to be thermally neutral. It has been found that patients who do not reach

hypermetabolic state are usually septic or are compromised through exhaustion in order to maintain the required altered homeostatic body temperature [230].

### 5.1.4 Hypermetabolic global systemic response to Burn Injury

Skin is the largest organ of the human body, encompassing ~1.5-2.0 square meters for the average adult, providing protection against trauma, infectious organisms, radiation and contributing to global homeostasis. The epidermis, which is avascular and composed primarily of keratinocytes, is the primary defence as the outmost layer of the skin. Beneath the epidermis, lies the dermal layer, which functions as both the mechanical strength layer and primary metabolic layer, attributed to the vascular structure that penetrates the organ from underneath. The hypodermis, beneath the dermal layer, is composed of fatty connective tissue and areolar tissues in addition to homing nerves and larger blood vessels that supply the deeper layers. Immune-competent cells are found in all layers of skin and aid the protection boundaries supplied by skin [231].

The chronic multisystemic effects of burn injuries are part of an uncharted aetiology of burn trauma associated hypermetabolism, with the metabolic and systemic inflammatory changes associated with burn injuries having been found to be persistent, long after injury [232]. These systemic changes throughout body have been observed in both severe (>20% total body surface area, TBSA) and non-severe burns. It has been established that in burn induced hypermetabolism, resting energy expenditure is predicted to be increased two-fold in comparison to normal healthy rates, with increased heart rates, increased breathing, body temperature, oxygen consumption, carbon dioxide production, glucose expenditure,

glycogenolysis, lipolysis and proteolysis [233]. As a result of burn hypermetabolism, an overall leaner body mass has been observed in burns patients due to the increased metabolic output [233]. Genomic studies have revealed that an altered white blood cell (WBC) genome exists in burn patients in comparison to other traumas, such as blunt trauma and sepsis, and whilst aberrant gene expression was seen to return to normal within one-month post injury for blunt trauma, for burns patients, expression remained aberrant well beyond a year [234].

### 5.1.5: Trauma and systemic inflammatory responses

Persistent inflammation following burn injury, not only induces a site localised response, but a global systemic response impacting a broad range of systems, from cardiovascular to hepatic. The hypermetabolic state achieved as a response to burn injury is prolonged and persevering, with the severity of the state linked to the longevity and duration of exposure to elevated levels of catecholamines, cortisol and glucagons. A 10-50-fold increase of plasma catecholamine and corticosteroid levels found in burn patients, lasting up to 9 months post-injury, inducing increased myocardial oxygen consumption and increased resting energy expenditure [235].

The prolonged elevated levels of these hormones propagate the significant changes in metabolic rates, growth and physiology seen in burn patients (23). Proinflammatory and anti-inflammatory cytokines are mediators of the subsequent immune changes following burn trauma; crucial in the orchestration of the acute phase response, and ultimately driving towards the restoration of homeostasis. Previous studies have found that despite

the pro-recovery aim of this mechanism, the sustained and increased release of pro-inflammatory cytokines, such as interleukin-1(IL-1), interleukin-6(IL-6) or tumour necrosis factor-α (TNF-α) can contribute to multiple organ failure, hypermetabolism, morbidity and mortality [236]. Anti-inflammatory response is dependent on T lymphocytes of helper Th-2 and mediating cytokines such as interleukin-2(IL-2), interleukin-4(IL-4), interleukin-10(IL-10) and interferon- γ (IFN- γ) work to suppress and counteract pro-inflammatory cytokines. The systemic inflammatory response is driven by macrophages which produces proinflammatory mediators such as IL-6 and TNF-α [237]. Burn injury has been found to result in the prolonged and substantive hypermetabolism that involves the increased production of proinflammatory cytokines, in addition to the increased production of reactive oxygen species (ROS), such as superoxide anion, hydroxyl radical, hydrogen peroxide and reactive nitrogen species such as nitric oxide and peroxynitrite [238]. One study found that hydrogen sulfide was significantly increased in the liver post burn injury. These reactive components have been found to aggravate the systemic inflammatory response and increase the damage to multiple organs [239]. Conversely, increased free radical production results in the impairment of antioxidant mechanisms, as evidenced by the depleted levels of glutathione, catalase, a-tocopherol and ascorbic acid levels. Nitric oxide is suggested to be involved in peripheral vasodilation, and the promotion of transcription and translation of several inflammatory cytokines following burn injury [240]

TNF-α is involved in several important processes, but of note in burn injury, in cellular apoptotic activity. TNF-α also has the capacity to activate neutrophils and macrophages for

antimicrobial defence purposes, and thus further has the capacity to induce secretion of proinflammatory mediators, IL-1 and IL-6. IL-6 has been found to be systematically elevated post-burn, with animal studies finding that IL-6 levels were elevated proportionately to area size of burn injury [241].

Lipid mediators, such as eicosanoids are key molecules in the regulation of the inflammatory profile and promoting homeostasis following burn trauma. The dysregulation of such molecules has been suggested to be involved in chronic inflammation and tissue damage. Eicosanoids have been found to be crucial in supporting antibiotic treatment by increasing mucosal production of bactericidal peptides and enhancing bacterial clearance [242]. Thus, these lipid mediators may prove to be beneficial components in determining patient outcomes and recovery, certainly in the intermediate phase after the burn injury.

Skeletal muscle is the primary site of peripheral glucose usage, in addition to the functional capacity to store surplus glucose and fatty acids within glycogen and triacylglycerol deposits, skeletal muscle plays a vital established role in metabolic regulation [243]. Following severe burn trauma, skeletal muscle is used as an endogenous amino acid store, to supplement exogenous nutritional sources, providing raw materials for more crucial functions including for the synthesis of acute phase proteins and the deposition of new skin. Burn injury patients often become cachectic (loss of muscle mass) due to the combination of muscle amino acid metabolism and ubiquitin/proteasome-mediated protein degradation and/or apoptosis-mediated muscle mass loss [243]. Skeletal muscle mitochondrial dysfunction has been documented in severe burns. It has been found that such mitochondrial dysfunction contributes to an increased metabolic rate within burn patients, with burn patients demonstrating more uncoupled skeletal muscle mitochondria, indicating a source for

increased heat production within skeletal muscle [244]. Mitochondrial dysfunction in relation to burn injury has also further been documented in metabolic studies; α-ketoisovalerate a marker of mitochondrial membrane function, was found to be significantly increased post-burn injury, further supporting the argument for altered mitochondrial function following burn trauma [249].

Burn related mitochondrial dysfunction has been further documented in cardiac tissue, with mitochondria comprising approximately 35% of cardiomyocyte volume in the heart [245]. Animal model studies have revealed a 3-fold accumulation of cytosolic cytochrome-C in burn injury compared to healthy controls in the first 24 hours following injury, suggested to be a means of reducing burn-induced oxidative stress, with lipid peroxidation in cardiac mitochondria found to be increased 30-50% [246]. Burn injury associated cardiac stress is further mediated by an increased catecholamine response; with increased myocardial oxygen consumption and increased cardiac output [247]. There is substantial evidence that suggests that cardiac damage following burn injury, is modulated by immune dysfunction and inflammatory mediators. Macrophage migration inhibitor factor (MIF) is an inflammatory cytokine and a neuroendocrine hormone; and a functional catalytic enzyme, which has been suggested to be released by cardiomyocytes and the skin following burn injury. Murine studies have found that in response to severe burn injury, MIF was a critical mediator of late and prolonged cardiac dysfunction, with anti-MIF treated burn injury mice found to display rapid restoration of normal cardiac function following treatment [248].

Hepatic insult following burn trauma, has been suggested to present immediately and remain elevated for 4-6 weeks, with hepatic injury markers, aspartate aminotransferase

(AST) and alanine aminotransferase (ALT), increased immediately after burn injury [250]. Substantive liver damage has been documented following burn injury; with liver sizes and weights found to be increased by 40-50%, 1 year post injury. Whilst much is unknown about the mechanism, it has been suggested this is attributed to the immune dysregulation and long-term oedema formation. Following initial burn injury, liver damage is a result of hepatic oedema formation, hypoperfusion, proinflammatory cytokines with the release of hepatic enzymes, associated with caspase activation in an attempt to maintain homeostasis following trauma [251]. Following burn injury, serum and hepatic concentration of proinflammatory cytokines, IL-1α/β, IL-6, and TNF-α, are increased and contribute to the ongoing hepatic insult [252].

Each systemic aberration, is in some measure associated to a state of chronic immune dysregulation, suggesting that the impact of burn injury is both systemically prolific and prolonged.

**5.1.6: Previous burn injury studies and related pathologies**

Burn injuries have long been established to induce a global multisystemic effect, far beyond the scope of the initial site of injury. Whilst severe and moderate burns have been well established to induce chronic inflammation, hypermetabolism, muscle wasting and insulin resistance, the mechanism behind these chronic aberrations has yet to be fully understood [253]. However, what is very much apparent, is the long-term multi-systemic effects of burn trauma, both non-severe and severe, which has been well documented. Much of our current knowledge indicates an underlying chronic inflammatory response; the mechanism

of which has yet to be fully understood but which has a substantial impact on multiple organs and systems. Table 1 summarises key studies investigating the systemic effects following burn injury. Several studies have highlighted the significant neurological complication following burn injury associated with inflammation [254, 255, 258]. Chronic inflammation following burn injury has been found to be strongly associated with poorer outcomes and increased morbidity and mortality as a result, with a specific immune signature detected in burn patients [262]. Whilst there is an association between burn injury and diabetes onset [221], studies have further highlighted the complexity of link, with insulin treatment found to improve burn healing [265].

**Table 5.1:** *Summary of burn injury and systemic damage studies*

| *Target* | *Comments* | *Reference* |
|---|---|---|
| Neurological | Endogenous Serotonin pathway contributes to the burn healing process<br>- Simulation of serotonin was found to significantly accelerate cell migration resulting in improved healing<br>- Disruptors of endogenous serotonin stimulation found to impact reepithelization, bigger wound size and high alpha smooth muscle actin- a strong indication of a prolonged differentiation state | **Sadiq et al. (2018) [254]** |
| Neurological | inflammatory mediators and/or growth factors at the site of a burn trigger the synthesis of α1-adrenoceptors on resident epidermal cells and nerve fibres, and an influx of α1-adrenoceptor labelled cells. The heightened expression of α1-adrenoceptors in injured tissue could shape inflammatory and wound healing responses<br>- Simulation of a1-adrenoreceptors evokes inflammatory cytokine production, boosting neurogenic inflammation and pain, and influences cellular migration and proliferation. | **Drummond et al. (2015) [255]** |
| Neurological | Mental health burden injury post-burn is significant and not limited to post-traumatic stress disorder (PTSD)<br>- Patients who screened positive for depression and anxiety, without PTSD were more likely to exhibit chronic in comparison to counterparts with no mental health disorders. | **Herrera-Escobar et al. (2020) [256]** |
| Neurological | Burn survivors had significantly increased mortality related to trauma in comparison to uninjured age and sex matched controls, with five-year mortality found to be significantly greater (11% in burn survivors vs 4% in controls). | **Mason et al. (2019) [257]** |
| Neurological | Burn injury patients are at increased risk of peripheral nervous system | **Duke et al. (2019) [258]** |

| | | |
|---|---|---|
| Chronic Inflammation and Immune Response | CCL5, RANTES and EGF found to be differentially regulated in response to burn injury in young vs aged patients | **Farinas et al. (2018) [259]** |
| Chronic Inflammation and Immune Response | Burn wound inflammatory response found to be regulated by two different T-cell subtypes.<br>- Burn wounds contain significantly greater numbers of T-cells due to the profound infiltration of αβ T-cells inflammation.<br>- Infiltrating αβ T-cells are likely to sequester inflammation.<br>- γδ T-cells activated with elevated CD4 and CD69 expression | **Rani and Schwacha (2017) [260]** |
| Chronic Inflammation and Immune Response | Burn injury patients found to have twice as many hospital admissions and 3.5x the number of days spent in hospital in comparison to uninjured counterparts<br>- Higher rates of infectious disease admissions were found for severe and minor burn injury patients<br>Burn injury patients had 1.75x greater rate of mortality caused by infectious disease during the 5 year period post discharge in comparison to uninjured counterparts | **Duke et al. (2017) [261]** |
| Chronic Inflammation and Immune Response | Differential profiles found for wound injuries with acute cytokine induction found to be faster and different in profile excision injury in comparison to burn injury, with changes to systemic monocyte and neutrophil levels.<br>- Differences in immune profiles between burn and excision patients noted up day 84 post injury. | **Valvis et al. (2015) [262]** |
| Chronic Inflammation and Immune Response | Murine studies found a unique dissemination of bacteria in the Burn wound treatment<br>- Increased polymorphonuclear neutrophil granulocytes dominated inflammation in the group of mines with infected burn wound compared with uninfected burn would group | **Calum, Høiby and Moser (2014) [263]** |
| Chronic Inflammation and Immune Response | Skeletal muscle has the capacity to store surplus glucose and fatty acids within glycogen and triacylglycerol depots respectively, where glycogen and fatty acids can be mobilized during prolonged periods of caloric restriction or heightened metabolic demand (e.g., exercise), to be catabolized in order to maintain cellular ATP availability. Amino acids, on the other hand, are not generally considered to be stored in such a manner within skeletal muscle, | **Porter, Hurren, Herndon and Børshein (2013) [264]** |

| | | |
|---|---|---|
| | i.e., in a temporary pool independent of structural proteins and cellular organelles.<br>- Subsequently, in response to severe thermal trauma, skeletal muscle assumes the role of an amino acid reserve where muscle protein breakdown and amino acid release from skeletal muscle serves to buffer plasma amino acid concentrations.<br>- It is suggested that aggressive nutritional interventions may not cause the shift in the abundance of amino acids, suggesting that skeletal muscle becomes an essential store of protein in patients suffering from severe burn trauma | |
| Metabolic (diabetes) | Insulin treatment post burn injury found to reverse and improve elastic-collagen rearrangement and reepithelization | **Baptista et al. (2018) [265]** |
| Metabolic (diabetes) | Pre-existing Diabetes Mellitus significantly increases the risk of developing hospital acquired infections following burn injury | **Knowlin et al. [266]** |

### 5.1.7: Metabolic profiling and Burn Injury

Whilst much of current literature on biomarker identification for burn injuries focuses on cytokines, metabolic studies to date have highlighted several important biomarkers, ranging from hepatic and urea cycle dysfunction to mitochondrial defects. Table 5.2 summarises current metabolic profiling studies of burn injury. Generally, elevated metabolites such as alanine, histidine and urea cycle derivatives were found to be elevated in burn injury patients.

**Table 5.2:** *Summary of metabonomic elucidated biomarkers associated with burn injury and trauma*

| **Technique & Biofluid, Species** | **Number of participants** | **Biomarkers** | **Reference** |
|---|---|---|---|
| NMR, Plasma, human | 21 severe burn, 3 control | Increased α-ketoisovaleric acid, 3-methylhistidine, and β-hydroxybutyric acid | Zhang et al. (2014) [267] |
| NMR, plasma, murine | 3 burn, 3 control | Increased intramyocellular lipids(IMCL) and free fatty lipids (FFAs) | Tziki et al. (2008) [268] |
| NMR, plasma, rats | 10 burn, 10 control, 10 treated | Increased alanine, histidine, leucine, valine, 3-methylhistidine, creatine, glutamic acid and α-ketoglutarate, α-ketoisovalerate | Yang et al. (2019) [269] |
| NMR, blood, porcine | 15 burn (blood drawn before burn for baseline) | Increased histamine, alanine, phenylalanine, and tyrosine. Increasing 2-hydroxybutyrate, 3-hydroxybutyrate, acetoacetate, and isovalerate | Hendrikson et al. (2019) [270] |
| UPLC-Q-TOF-MS, skin biopsy and serum, murine | 9 burn (blood drawn before burn for baseline) | Skin: glycerophospholipid, ascorbate and aldarate, and phenylalanine metabolism | Alkhalil et al. (2019) [271] |

| | | | |
|---|---|---|---|
| | | Serum: arachidoinc acid, starch and sucrose, and galactose metabolism | |
| UPLC-Q-TOF-MS, plasma,, rats | 40 samples | Increased hypoxanthine, indoxyl sufate, glucuronic acid, gluconic acid, proline, uracil, nitrotyrosine, uric acid, and trihydroxy cholanoic acid in severe cases | Liu et al. (2010) [272] |
| UPLC-MS, skin biopsy, rats | 4 rats | Increased niacinamide and uric acid | Friston et al. (2019) [273] |
| GCMS, urine, rat and human | 10 rats, 15 paediatric human, | Increased 3-methylhistidine | Beffa et al. (2006) [274] |

**5.1.8: Paediatric Burns Study**

The Paediatric Burns Biobanking pilot study recruited 33 children at least 3 years post non-severe burn injury, aged 0-4 years of age at the original time of injury and trauma presentation. All recruited patients had no history of pre-existing illness and no medication at time of biobanking. A cohort of 22 non-injured age and sex matched healthy controls were recruited from the general population, with no prior medical history of note.

Previous work on this cohort, has revealed long-term immune dysfunction, in addition to a diminished vaccine response amongst a subset of patients. Circulating cytokine and vaccine antibody levels were analysed and quantified using multiplex immunoassays and cell profiles measured against a panel of 40 metal-conjugated antibodies and mass cytometry. Preliminary analysis (Johnston et al., 2020) demonstrated elevated levels of TNF- α (1.31-fold), IL-2 (1.18-fold), IL-7 (1.63-fold), and IFN- γ (1.18) 3 years post injury in the paediatric

burn injury patients in comparison to healthy controls. In addition, the analysis also found a significantly reduced immune response in burn patients with regard to the antibody response to the diphtheria, tetanus and pertussis vaccine antigens. Mass cytometry analysis revealed an altered expression of T-cell, B-cell and dendritic cell markers. The results from the study indicated a significantly depressed vaccine response seen in a subset of patients following a non-severe burn, in addition to chronic immune dysfunction 3 years post burn [275].

Metabolic phenotyping is a powerful investigative strategy, that enables the interrogation of the interactions between genotype, environment and lifestyle from a molecular latitude using spectroscopic techniques and biological specimens such as serum, plasma and urine [1]. Metabonomics has been pivotal in understanding the underlying metabolic changes in an amalgamation of diseases, ranging from non-communicable diseases such as diabetes, heart disease and cancers to communicable disease such as SARS-CoV-2, revealing disease specific metabolic signatures and biomarkers [276]. In order to stratify and evaluate patient outcomes, metabolic profiling could prove to be a crucial clinical tool in understanding the requirements and treatment of burn injury patients through the identification of outcome-specific signatures and biomarkers. Through the prediction of patient outcomes, the most appropriate medical interventions can then be recommended for the best possible clinical outcomes.

# 5.2 METHODS

### 5.2.1 Sample Collection

Paediatric burn patients were recruited for the WA Paediatric Burns Biobank pilot study from 2 separate hospital sites; 3 years post-burn injury, aged 0-4 at the time of original injury presentation. Age and sex matched healthy controls were recruited from a pool of participants. All samples were collected and obtained following informed consent of a parent or guardian and collection was conducted with ethical approval from the Child and Adolescent Health Service WA (2015219EP;1111EP;768EP). All patients had no prior pre-existing illness and were not receiving treatment or medical interventions of any kind.

### 5.2.2 Multiplex Cytokine Assay

Cytokine profiling data were generated by Mr. Blair Johnson (University of Western Australia). Simultaneous analysis of cytokine and chemokine levels was conducted using a custom Milliplex MAP human high sensitivity T cell panel multiplex bead assay (MERCK). Quantified cytokines were Tumour necrosis factor- alpha (TNF-α), Interleukin-8 (IL-8), Interleukin 7 (IL-7), Interleukin-6 (IL-6), Interleukin-5 (IL-5), Interleukin-2 (IL-2), Interleukin-1beta (IL-1β), Interleukin-17A (IL-17A), Interleukin-13 (IL-13), Interleukin-12 p70 (IL-12(p70)), Interleukin-10 (IL-10), Interferon-gamma (IFN-γ) and Granulocyte macrophage colony stimulating factor (GM-CSF). Plasma samples were filtered using 0.45 µm syringe filters (Nalgene) and a 50 µl was used for the assay. Standards were prepared for each cytokine and plated in duplicate and the assay performed as per manufacturer's (MERCK)

instructions, and then analysed on a Luminex 200 instrument. Each sample was assayed in duplicate and the mean value for each cytokine/patient was used for statistical analysis.

### 5.2.3 Diptheria-Tetanus-Acellular Pertussis Multiplex Immunoassay

For assessment of vaccine responsiveness, a custom multiplex immune assay was carried out by Mr. Blair Johnson (Univeristy of Western Australia). Total methods can be found as documented by Johnson et al., (2020) but methods are summarised below. Total IgG concentrations against vaccine antigens pertussis toxin (PT), pertactin (PRN), filamentous hemagglutinin (FHA), fimbriae 2/3 (FIM 2/3), tetanus toxin (TT), and diphtheria toxoid (DT) were measured using an in-house multiplex bead-based immunoassay.

The multiplex immunoassay was carried out as per van Gageldonk et al. [277] with minor modifications. In brief, Bio-Plex® COOH-microspheres (6.25 × 106) were conjugated with optimized concentrations of antigen in 1 x PBS pH 7.2 (Life Technologies, AUS) as follows: PT 10 µg/ml, PRN 75 µg/mL, FHA 25 µg/ml, FIM 2/3 100 µg/ml, TT 100 µg/mL, and DT 100 µg/ml. Samples were diluted in PBS containing 3% bovine serum albumin (BSA) and 0.05% Tween 20 (Sigma-Aldrich). MultiScreen Filter Plates (Merck) were pre-wet with 50 µl PBS containing 0.05% Tween 20 (PBS.T) and the liquid removed by vacuum manifold (2–5 mmHg). Diluted plasma samples were mixed with bead-mix (25 µl; PBS containing 4000 beads/region) in individual wells and incubated on a plate shaker (500 rpm) in protective photo-sensitive conditions for 30 min. Plates were washed twice with 100 µl of 0.05% PBS.T before the addition of 100 µl 1:200 RPE-conjugated goat-anti human IgG Fc secondary antibody (Jackson ImmunoResearch Laboratories Inc.) and incubated as above for a further

30-min. Following washing, the beads were resuspended in 125 µl 0.05% PBS.T and read using a bioplex-200 machine. Antigen-specific IgG concentrations (mIU/mL) were determined using a 5-PL linear curve generated with Bioplex Manager software version 5.0.

### 5.2.4 Mass Cytometry data acquisition

Mass cytometry data were generated by Mr. Blair Johnson (University of Western Australia). For Peripheral Blood Mononuclear Cell Immunophenotyping by mass cytometry, a custom method was validated by Johnson et al., (2020). Total methods can be found as documented by Johnson et al., (2020) and methods are summarised below.

Cryopreserved PBMCs were rapidly thawed then transferred into warm RPMI + 10% heat-inactivated fetal calf serum (FCS) (Bovogen, French origin) + Pierce universal nuclease (Thermofisher, Australia). Following staining of approximately two million live PBMCs; 100 µL of the first surface antibody cocktail containing an individual barcoding reagent (to distinguish controls and patients) was added. Samples were then incubated at room temperature for 30 min. Cells were washed, then patient and control pairs were combined into a single tube and washed again. The remainder of the surface staining antibodies were added, 100 µL per sample, and incubated on ice for 30 min. Cells were washed twice, then permeabilized with FoxP3 Fix/Perm buffer (Thermofisher) and incubated with the intracellular antibodies on ice for 30 min. Cells were fixed overnight at 4°C in 4% PFA containing 0.125 µM DNA intercalator. Once washed, cells were then resuspended at 8.5 × 105 cells/mL in a 1:10 suspension of EQ beads and Maxpar water prior to data acquisition on a Fluidigm Helios mass cytometer.

Data were normalized using CyTOF Software (v6.7.1014, normalization passport EQ-P13H2303_ver2) and gated using Flowjo (v10.4.2).

To remove artifacts, data were then cleaned by applying DNA specific gating and criteria for identification and removal of debris and non-cellular events. Patient and control samples were debarcoded and the exported files were further cleaned to remove cell debris based on high cisplatin staining. T-cells (CD3+ CD19–), B-cells (CD19+ CD3–), and other cell lineages (CD3– CD19–) were gated and exported as individual files for use in downstream analysis. The CAPX data analysis pipeline (v2.5, Sydney Cytometry) (27) was then used to down-sample, transform (arcsine), cluster (FlowSOM), and perform dimensionality reduction (tSNE) on the 3 gated populations for each sample. In order to maximise the number of clusters without the introduction of false positives that were biologically similar, B-cells, T-cells, and “other cells” were analyzed with a range of cluster numbers between 10 and 90 to ensure cluster separation on meaningful markers.

### 5.2.5 Biogenic amines and amino acids analysis

All metabolic profiling experiments and assays were performed at the Australian National Phenome Centre. The quantification of biogenic amines and amino acids was conducted according to a previously reported method [51] with minor modifications. In brief, 28 amino acids were quantified from 10 mL of plasma. Sample extraction was completed using a Biomek i5 sample automation system (Beckman Coulter, Mount Waverley, Victoria 3149, Australia). Samples were diluted 1:1 with water, 20 mL of stable isotope labelled (SIL) internal standards in water was added, and protein precipitation was performed by the addition of

90 mL of methanol. After mixing and centrifugation 10 mL of the supernatant was transferred into a Waters 700 mL 96-well plate for derivatization with AccQTag™ reagent (Waters Corp., Milford, MA, USA). The derivatized samples were then diluted 1:4 (v/v) with water prior to LC-MS analysis.

LC-MS analysis was performed using a Waters Acquity UPLC® (Waters Corp., Milford, MA, USA, coupled to a Bruker Impact II QToF-MS (Bruker Daltonics, Bremen, Germany). Resulting raw data files were processed for peak integrations and the calculation of metabolite concentrations using MassLynx software.

### 5.2.6 Tryptophan metabolic pathway analysis

LC-MS analysis was performed using a Waters Acquity UPLC® (Waters Corp., Milford, MA, USA) coupled to a Waters Xevo TQ-S MS (Waters Corp., Wilmslow, UK). Peak integration and calculation of resultant quantification data were performed using the TargetLynx package in MassLynx v4.2 (Waters Corp., Milford, MA, USA).

The quantification of tryptophan and 14 products of its metabolic pathway was performed as previously reported [52]. In brief, tryptophan metabolites were quantified from 50 µL of plasma. Sample extraction was completed at the Australian National Phenome Centre using a Biomek i5 sample automation system. SIL internal standards (20 µL) were added to all samples prior to protein precipitation via the addition of 250 µL of methanol containing 2 mM ammonium formate. After mixing, samples were transferred to a Phenomenex PHREE™ phospholipid removal solid phase extraction plate (Phenomenex, NSW, Aus). PHREE

plates were then washed with an additional 150 µL of methanol containing 2 mM ammonium formate. Eluent collection plates were dried using a SpeedVac vacuum concentrator (Thermo Fisher, Massachusetts, USA). Dried extracts were re-suspended in 100 µL of water with 0.1% formic acid prior to LC-MS analysis.

### 5.2.7 NMR data acquisition

$^{1}$H NMR spectroscopy of plasma was completed according to a previously published method [45]. briefly described as follows: Blood samples were centrifuged at 13,000g for 10 min at 4°C. The plasma supernatant was mixed with 75 mM pH 7.4 sodium phosphate, buffer in 1:1 ratio, and 600 µL were transferred into a Bruker SampleJet™ 5mm NMR tube. For quality control, a pooled study sample of 50 µL supernatant from all individual study samples was made and transferred into a 5 mm SampleJet™ NMR tube using the prior method.

NMR analysis was performed on a 600 MHz Bruker Avance III HD spectrometer, with a BBI probe and fitted with an automated Bruker SampleJet™ robot cooling system set to 5°C. NMR analytical calibration was run prior to the experimental run using the protocol described by Dona et al [45]. Experiments were completed according to the Bruker *in vitro* Diagnostics research (IVDr) methods [46].. For each sample, three experiments were run in tandem with a total analysis time of 12.5 minutes; $^{1}$H 1D experiment with solvent pre-saturation [47] (32 scans, 98304 data points, spectral width of 18028.85 Hz), a 1D Carr-Purcell-Meiboom-Gill (CPMG) spin-echo experiment (32 scans, 73728 data points, spectral width of 12019.23Hz) and lastly a diffusion edited relaxation (DIRE) experiment (2 scans with 40 $t_1$ increments). All

acquired data were processed in automation using Bruker Topspin™ 3.6.2, and ICON™ NMR for phasing, baseline correction and TSP calibration (δ=0).

112 lipoprotein fractions were quantified for each sample, using the Bruker IVDr Lipoprotein Subclass Analysis (B.I.-LISA ™) method [46]. This is completed by mathematically interrogating and quantifying the -$CH_2$ (1.25 ppm) and -$CH_3$ (0.8 ppm) peaks of the 1D spectrum after normalization to the Bruker QuantRef ™ manager within Topspin™ using a PLS-2 regression model. The various lipoprotein subclasses included different fractions of intermediate-density lipoprotein (IDL, density 1.006–1.019 kg/L), very low-density lipoprotein (VLDL, 0.950–1.006 kg/L), low-density lipoprotein (LDL, density 1.09–1.63 kg/L), and high-density lipoprotein (HDL, density 1.063–1.210 kg/L). The LDL sub-fraction was systematized into six density classes (LDL-1 1.019–1.031 kg/L, LDL-2 1.031–1.034 kg/L, LDL-3 1.034–1.037 kg/L, LDL-4 1.037–1.040 kg/L, LDL-5 1.040–1.044 kg/L, LDL-6 1.044–1.063 kg/L), and the HDL sub-fractions into four density classes (HDL-1 1.063–1.100 kg/L, HDL-2 1.100–1.125 kg/L, HDL-3 1.125–1.175 kg/L, and HDL-4 1.175–1.210 kg/L). The 1-acid glycoprotein N-acetyl-glucosamino (N-acetyl) signal integrals were calculated as Glyc A (2.06ppm) and Glyc B (2.10 ppm), as determined from the DIRE spectra.

### 5.2.8 Data Processing and Statistical Analysis of Mass Spectrometry data

Peak integration and calculation of resultant quantification data were performed using the TargetLynx package in MassLynx v4.2 (Waters Corp., Milford, MA, USA). Once data were scaled to account for 1:1 dilution of sample with water and converted to appropriate scales across both assays, data were then modelling using principal components analysis (PCA)

(Simca V.15.0., Umetrics, Sweden) to investigate variation. Scores plots were used to examine the relationship between healthy controls and burn patients. To further interrogate the relationship and differences between healthy controls and burn injury patients, orthogonal partial least-squares discriminant analysis (OPLS-DA) [33]. To test the robustness of the model permutation testing was performed. Subsequent statistical analysis was conducted using R; with "ggplot2" used to visualise log-transformed metabolite concentrations. Statistical significance of individual metabolites was evaluated using Wilcoxon Rank Sum testing, corrected for multiple testing (Benjamini-Hochberg, pvalue<0.01).

Spearman's correlation was carried to interrogate the relationship between each dataset, hierarchically clustered and mapped using ggplot2 visualisations. Ratios of significance for biological interpretation were tested for significance using Mann Whitney U testing, corrected for multiple testing (Benjimini-Hochberg, p value<0.01). Metaboanalyst (V5.0) was used to perform metabolite set enrichment analysis (MSEA) [53], data were log transformed and mean centred, against a reference metabolome extracted from SMPDB (Small Molecule Pathway Database, www.smpdb.ca) comprising of 99 metabolite sets based on normal human metabolic pathways. Differential enrichment was ranked based on p-values (Fisher's exact test).

### 5.2.9 Data Processing and Statistical Analysis of NMR spectral data

Following spectral preprocessing and eretic signal normalisation in *R*, using the "Metabomate v0.5" package, NMR data was exported into SIMCA 14.0.1 (Umetrics, Umea, Sweden) for multivariate analysis. Initial Principal Component Analysis (PCA) was carried out to identify extreme outliers and inspect sample variability. Supervised orthogonal partial least square discriminant analysis (OPLS-DA) was performed to investigate and discriminate metabolites between groups. Statistical significance and robustness of OPLS-DA models were then validated by permutation testing of the Y matrix for 1000 permutations. Permutation testing comprises of the repeated randomisation of class labels in the dataset, with classified models then recalculated using this new permutated dataset. Permutation testing evaluates the null hypothesis; if the null hypothesis is accepted, then no difference will be found in the quality-of-fit value between the original model dataset and the permutated set, and the given classification model is significant. The obtained p-values were adjusted using Benjamini-Hochberg (BH) correction [48] with a false discovery rate (FDR) threshold of 0.05.

Spearman's correlation was carried to interrogate the relationship between each dataset, hierarchically clustered and mapped using ggplot2 and corrplot visualisations. Ratios of significance for biological interpretation were tested for significance using Mann Whitney U testing, corrected for multiple testing (Benjimini-Hochberg, pvalue<0.01). Hierarchical edge clustering was used in conjuction with spearman's correlations, to assess the association between significant lipoproteins, small molecules and cytokine and visualised using R package, "edgebundleR".

# 5.3 RESULTS

### 5.3.1 Univariate analysis of targeted metabolites

All 46 metabolites using the biogenic amine assay and the tryptophan pathway assay were quantified and log transformed (figure 5.2 and 5.3). Univariate analysis identified a total of 12 out of 46 metabolites found to be significantly different between prior burn injury and healthy matched control groups (figure 5.2). Ornithine, Lysine, Alanine, Serine, Proline, Cystine, Asparagine, Phenylalanine, Arginine, Tyrosine, 3-Methylhistidine and Quinolinic acid were found to be significantly elevated (p value <0.05) in burn injury patients in comparison to healthy controls. A further total of 27 of 46 metabolites appeared to be elevated in the burn injury group in comparison with healthy controls (figure 5.3), with 7 metabolites appearing depleted in comparison (glutamic acid, citrulline, beta-aminoisobutyric acid, ethanolamine, taurine, nicotinic acid, serotonin), but these were not found to be significant when adjusted for multiple testing.

### 5.3.2 Multivariate analysis of targeted metabolites

Multivariate analysis was carried out to further assess and investigate the relationship between burn injury and healthy controls. PCA scores plots showed partial separation of burn injury patients compared to healthy controls, with a subset of burn injury patients segregating in the first principal component (figure 5.4A). Orthogonal partial least-squares analysis (OPLS-DA) was then used to further interpret the model and to reduce model complexity. OPLS- DA scores plots revealed clear separation of burn injury and non-burn injury controls groups ($R^2X$ 0.33, $R^2Y$ 484, $Q^2Y$ 0.161) (figure 5.4B) when discriminating on class differences. Figure 5.4C presents the same model but coloured to investigate any clustering of vaccine response. Vaccine response in

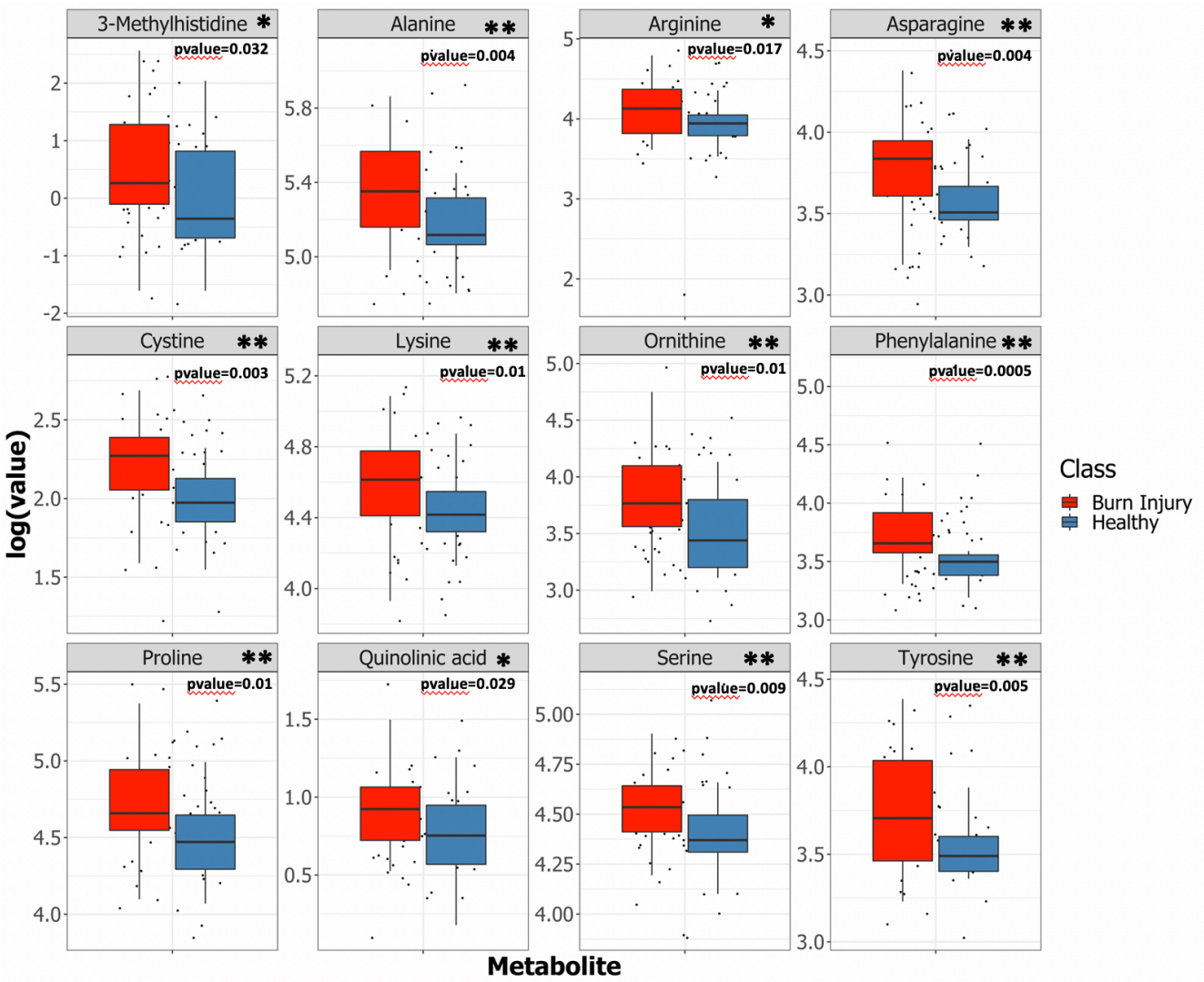



**Figure 5.2: Significant quantified metabolites**. 12 out of 46 quantified metabolites were found to be statistically significant, coloured burn (red), healthy controls (blue), pvalue * <0.05** <0.01, corrected for multiple testing using Benjamini-Hochberg.

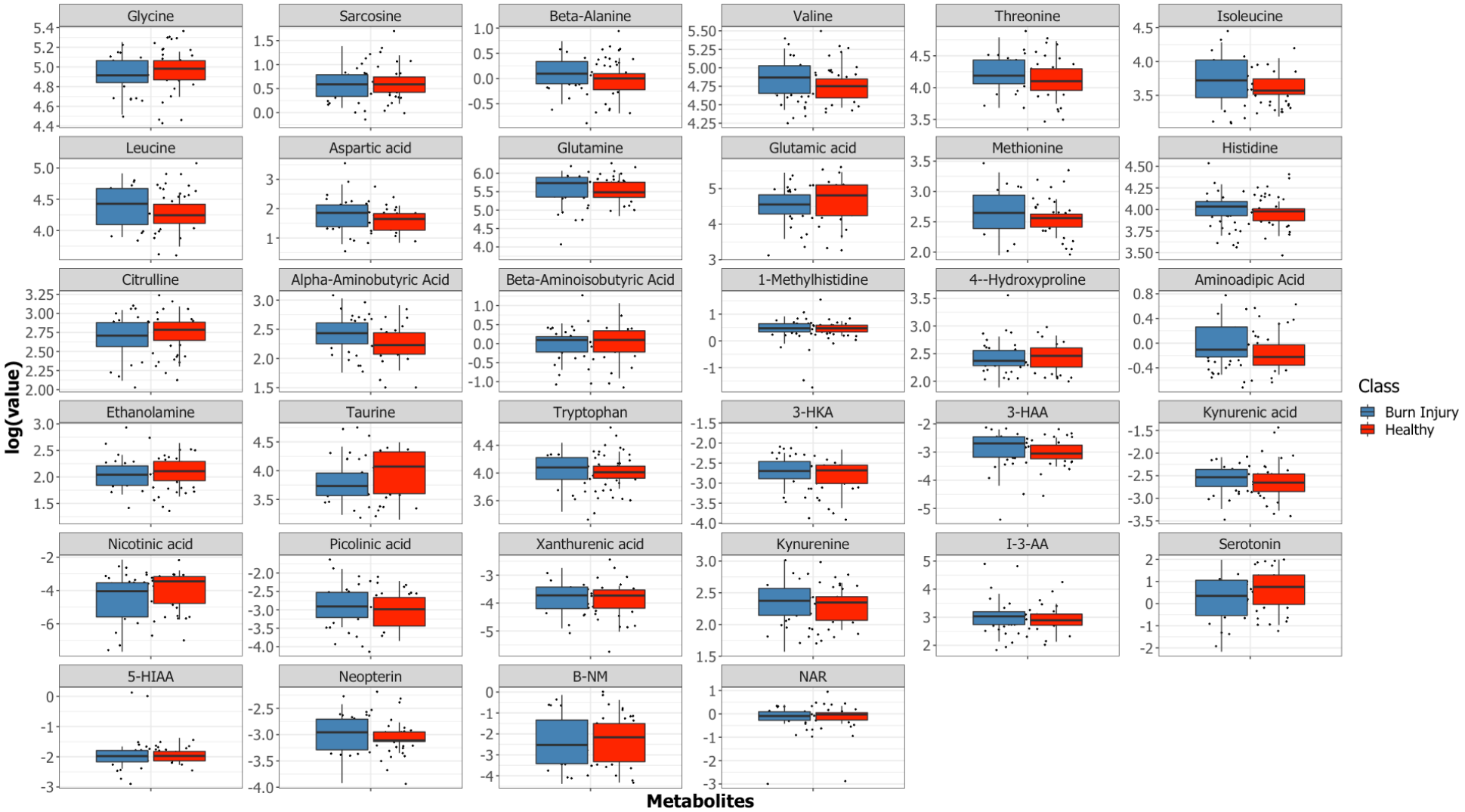



**Figure 5.3. Non-signficant quantified metabolites.** 34 out of 46 quantified metabolites were found to be statistically non- significant, coloured burn (red), corrected for multiple testing using Benjamini-Hochberg. 3-HKA, 3-hydroxynurenine; 3-HAA, 3 hydroxyanthranillic acid.

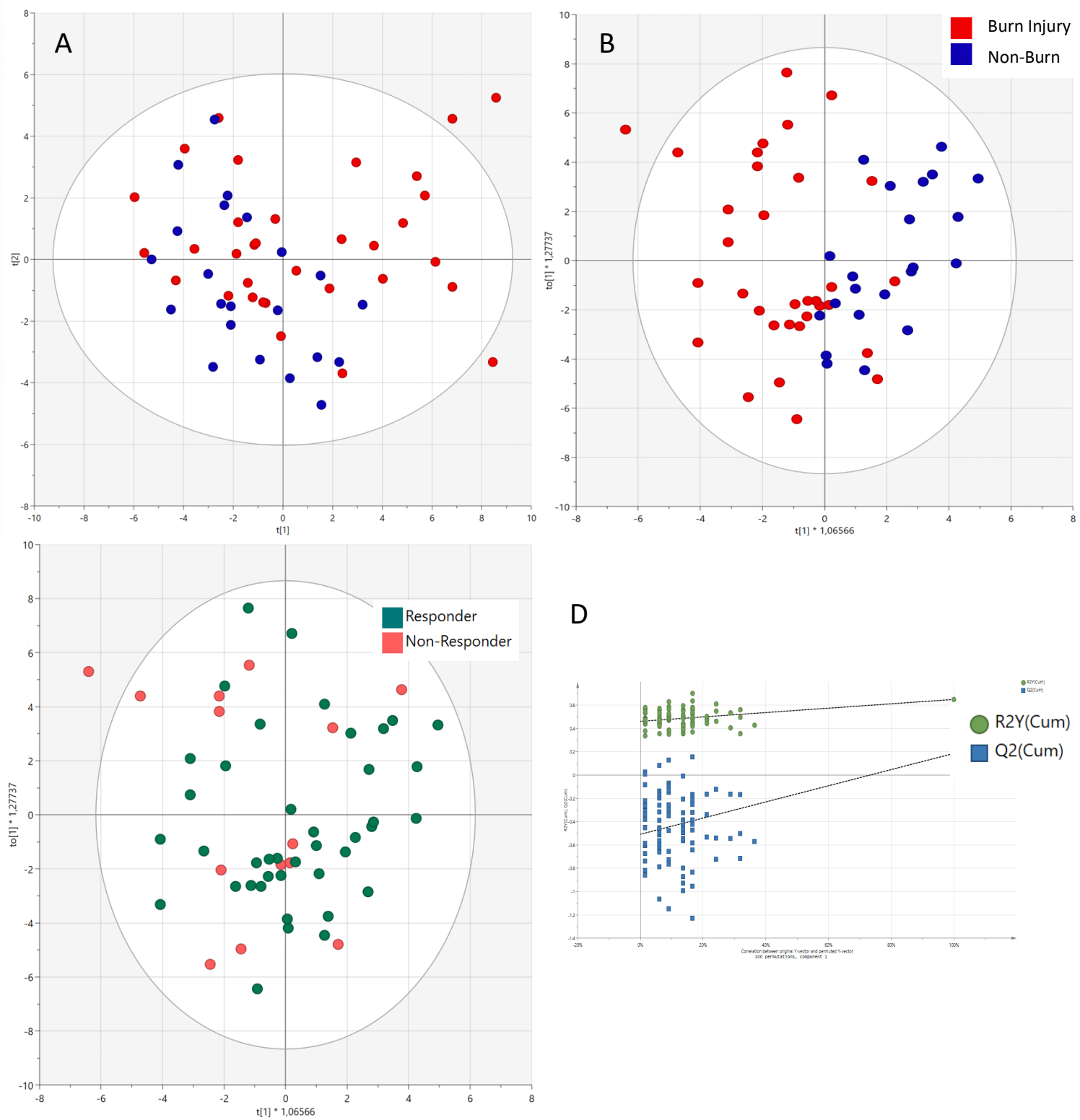



**Figure 5.4. Multivariate analysis of quantified biogenic amine and tryptophan pathway metabolites**. A) PCA based on all metabolites. ($R^2X$ 0.517, $Q^2Y$ 0.211). B) OPLS-DA regression for class (burn injury or non-burn control) discrimination on based on metabolic differences. ($R^2X$ 0.33, $R^2Y$ 484, $Q^2Y$ 0.161). C) Same OPLS-DA model as B but coloured based on vaccine response (responder or non-responder). 14 of 33 burn patients were found to have a significantly diminished vaccine response, whereas, with the exception of one child, all of the non-burn injury children (n=21) responded to the vaccine D) Permutation testing on O-PLSDA model (B) based on 100 random permutations.

burn injury patients was previously found to have significant differences in responsiveness in burn as opposed to non-burn injury controls (pvalue <0.001). Figure 5.4C depicts where vaccine non-responders lie in the model, which is shown to be in the left two quartiles: 14 out of 33 burn injury patients were classed as vaccine non-responders with a significantly lower antibody titre response to vaccination as reported previously by Johnson et al (2020) [275].

For model cross validation, from 100 random permutations, permutation testing estimated good model confidence (figure 5.4D). The variable importance of projection (VIP) plot (figure 5.4) shows the most significant discriminatory metabolites in the model, presenting 22 metabolites out of 46 to be class discriminatory for prior burn injury: Lysine, Alanine, Tyrosine, Asparagine, Proline, Leucine, Methionine, Serine, Taurine, Aspartic acid, Ornithine, Arginine, Aminoadipic acid, Isoleucine, Histidine, Valine, Serotonin, Ethanolamine, 3-Methylhistidine, Threonine, Cystine and Quinolinic acid. All 22 metabolites were elevated in burn injury patients in comparison to healthy controls (figure 5.2 and 5.4). Figure 5.6 compares results from univariate analysis and supervised multivariate analysis, showing a large overlap in results.

When applying an arbitrary predictive score threshold of 1.0 to projections in the VIP, the nine most model significant metabolites were Lysine, Alanine, Tyrosine, Asparagine, Proline, Leucine, Methionine, Serine and Taurine (figure 5.5). Although not statistically significant, the calculated Fischer's ratio, calculated from the ratio of branched chain amino acids against aromatic amino acids, showed a trend towards a higher ratio in prior burn injury in comparison to the control group, which together with elevated taurine levels may be

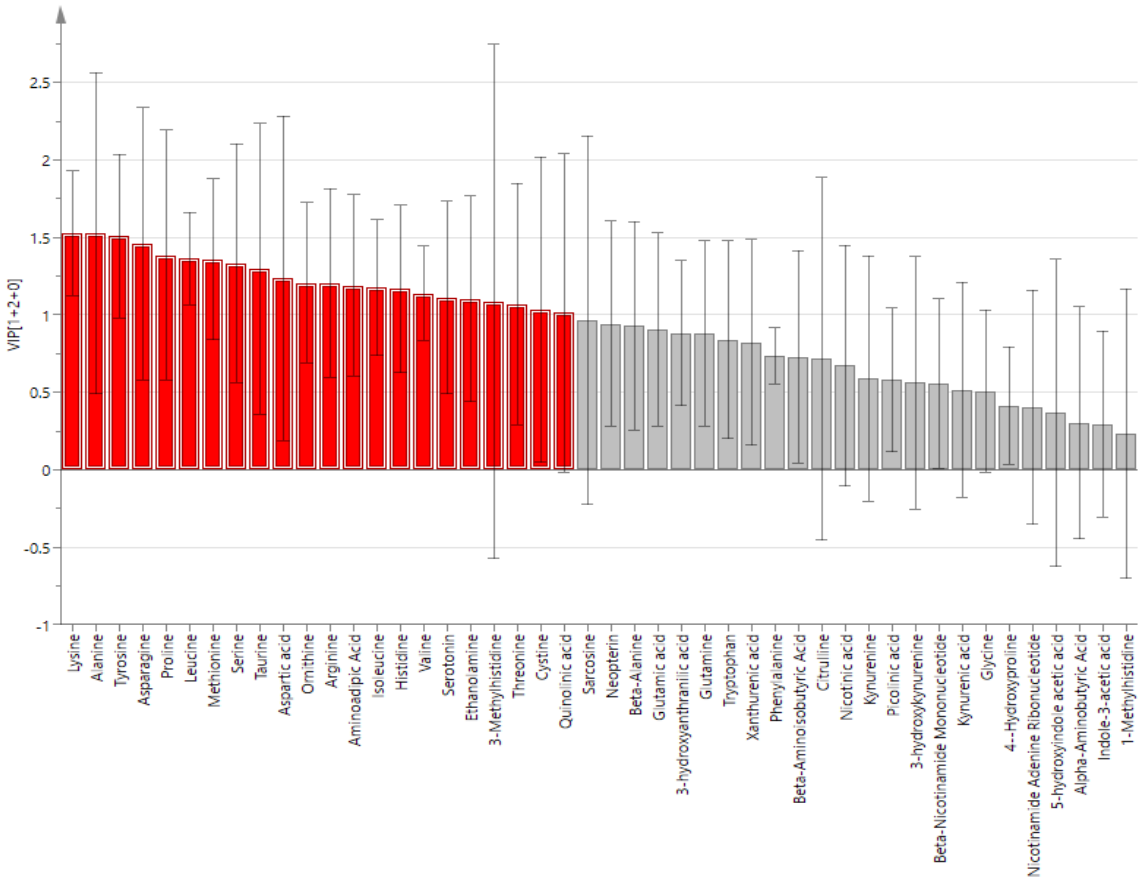



**Figure 5.5:** Ranked variable importance of projection (VIP) plot based OPLS-DA model for burn injury and healthy control; listing the most significant variables contributing to the model.

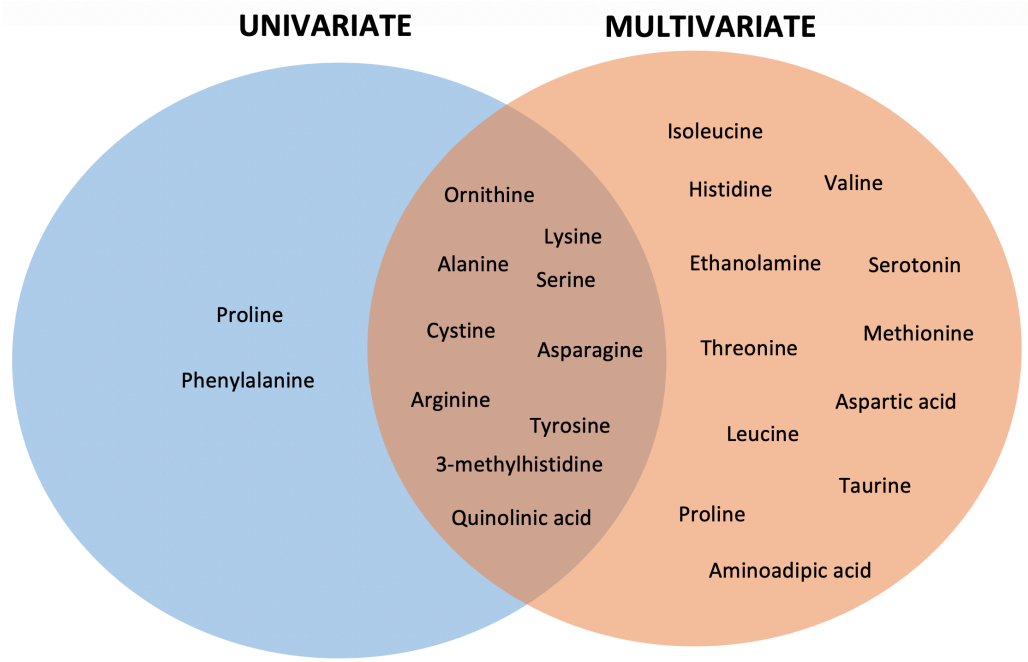



**Figure 5.6: Venn diagram comparing significant metabolites from univariate analysis and supervised multivariate analysis between burn injury and non-burn controls.**

indicative of decreased liver function (table 5.3). A further marker of liver function is the relatively low (phenylalanine/tyrosine) in burn injury patients. Interestingly, neuroendocrine function was estimated to be depleted in burn injury patients in comparison to healthy controls, with a lower ratio of serotonin to 5-hydroxyindoleacetic acid. Neurotoxic quinolinic acid was found to be elevated against all parameters in prior burn injury participants.

### 5.3.3 Correlation analysis and data integration

To assess the association between cytokine levels and immune cell prevalence in burn injury patients in comparison to healthy controls, the Spearman's correlation was calculated. Hierarchical clustering of the correlations produced 5 immune cell clusters and 4 clusters within cytokine measurements respectively. It is noteworthy that more sporadic and negative correlations between cytokines and immune cell counts were seen within the healthy control group. In cluster 1 of the immune cell (figure 5.7A), mature dendritic cells with CD38 surface receptors (mDC CD38) were found to be positively correlated with the third cytokine cluster, which primarily comprised of strong inflammation-inducing cytokines such as interleukin (IL)-12, interferon gamma (IFN- γ), IL-17A, IL-1B and IL-5. Conversely, there was strong negative correlation between these two subsets of inflammatory measures detected in healthy controls. In cluster 2, a strong positive correlation across this subset of immune cells and all cytokines was observed, in comparison to the negative correlations seen in healthy controls (figure 5.7B). In contrast to healthy controls, a much more positive correlation between cytokines and immune cells was seen, indicating activation of cytokines and stimulation of immune cells.

To interrogate the association between metabolite concentrations and cytokines levels, correlation analysis was used. Hierarchical clustering defined 5 clusters of metabolites, and a further 4 clusters of cytokine measurements (figure 5.8A) based on the metabolite-cytokine correlation values. Different correlation patterns of branched chain amino acids and pro-inflammatory cytokines were observed for burn injury patients and non-burn injury controls between.

Generally, negative spearman's correlation coefficients were calculated in non-burn controls, between cytokines and metabolites which could be expected considering the normal immune and inactive inflammatory response of the individuals. Strong positive correlation was seen in burn injury patients across clusters 2, 4 and 5. In contrast, strong negative correlation was seen in cluster 5 for the control group. Notably, Indole-3 acetic acid and Glutamine were amongst the few metabolites negatively correlated across the array of cytokines in burn injury patients.

When integrating metabolites and immune cell phenotypes, generally a quenched immune-metabolite correlation was seen in burn injury patients in comparison to healthy controls. Hierarchical clustering revealed 5 different clusters in metabolites and a further 4 clusters in immune cell subtypes. Notably, a strong negative correlation was seen in a subset of metabolites in cluster 1, against CXCR4 differentiated cells in cluster 1, for burn injury patients (figure 5.9A). This correlation was reflected in healthy controls as a strong positive correlation (figure 5.9B). A diminished immune response in burn injury patients was previously reported in burn injury patients from this dataset, and thus it is unsurprising that a quenched response was found in association to metabolite-immune parameter responses.

**Table 5.3: List of all calculated metabolite ratios.** Metabolite ratios were calculated for physiological

0 Interpretation. Due to small sample size of this pilot study- none were found to be statistically significant.

| **Metabolites** | **Ratio** |
|---|---|
| *Glutamine/Glutamate* | |
| *Burn Injury* | 2.64:1 |
| *Healthy* | 2.17:1 |
| *Tryptophan/Kynurenine* | |
| *Burn Injury* | 5.48:1 |
| *Healthy* | 5.53:1 |
| *Serotonin /5-Hydroxyindole acetic acid* | |
| *Burn Injury* | 10.16:1 |
| *Healthy* | 17.4:1 |
| *Phenylalanine/Tyrosine* | |
| *Burn Injury* | 0.97:1 |
| *Healthy* | 1.1:1 |
| *Quinolinic Acid/Kynurenic Acid* | |
| *Burn Injury* | 31.37:1 |
| *Healthy* | 29.54:1 |
| *3-hydroxykynurenine/Kynurenic Acid* | |
| *Burn Injury* | 0.82:1 |
| *Healthy* | 0.89:1 |
| *Kynurenic Acid/Kynurenine* | |
| *Burn Injury* | 0.007:1 |
| *Healthy* | 0.006:1 |
| *Alanine/Valine* | |
| *Burn Injury* | 1.69:1 |
| *Healthy* | 1.51:1 |
| *Alanine/Isoleucine* | |
| *Burn Injury* | 5.03:1 |
| *Healthy* | 4.74:1 |
| *Alanine/Leucine* | |
| *Burn Injury* | 2.59:1 |
| *Healthy* | 2.47:1 |
| *Asparagine/Arginine* | |
| *Burn Injury* | 0.68:1 |
| *Healthy* | 0.69:1 |
| *3-hydroxykynurenine/Kynurenine* | |
| *Burn Injury* | 0.006:1 |
| *Healthy* | 0.006:1 |
| *Quinolinic Acid/Picolinic Acid* | |
| *Burn Injury* | 44::1 |
| *Healthy* | 43.8::1 |
| *Quinolinic Acid/Tryptophan* | |
| *Burn Injury* | 0.05:1 |
| *Healthy* | 0.039:1 |
| *Fischer's Ratio* | |
| *Burn Injury* | 1.95:1 |
| *Healthy* | 2.05:1 |

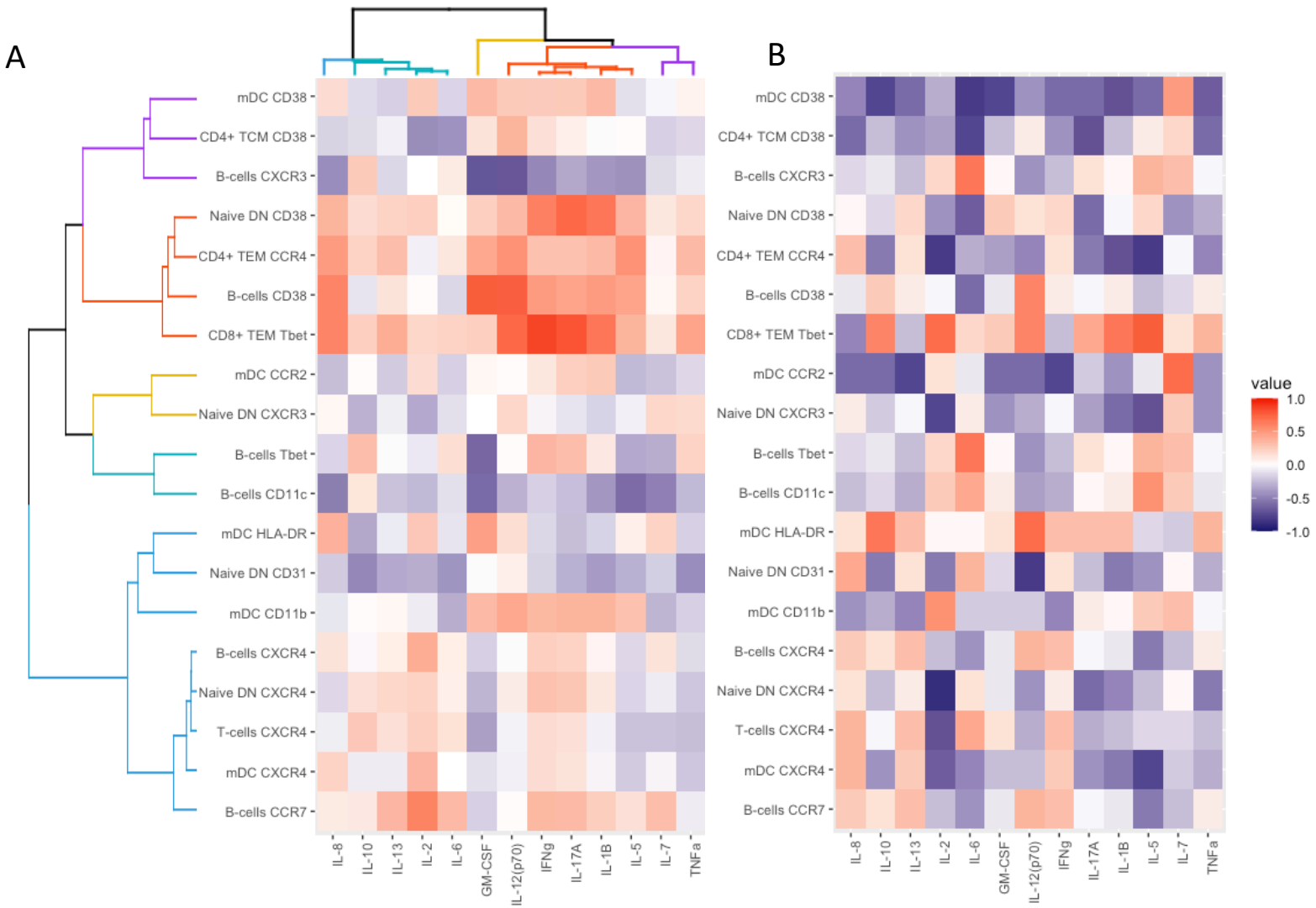



**Figure 5.7:** Correlation heatmaps integrating immune cell phenotyping (mass cytometry) and circulating cytokines. A) Measurements from burn injury patients. B) Measurements from healthy controls. Hierarchical clustering based on clustering from burn injury results and mirrored for direct comparison to non-burn injury age and sex matched controls.

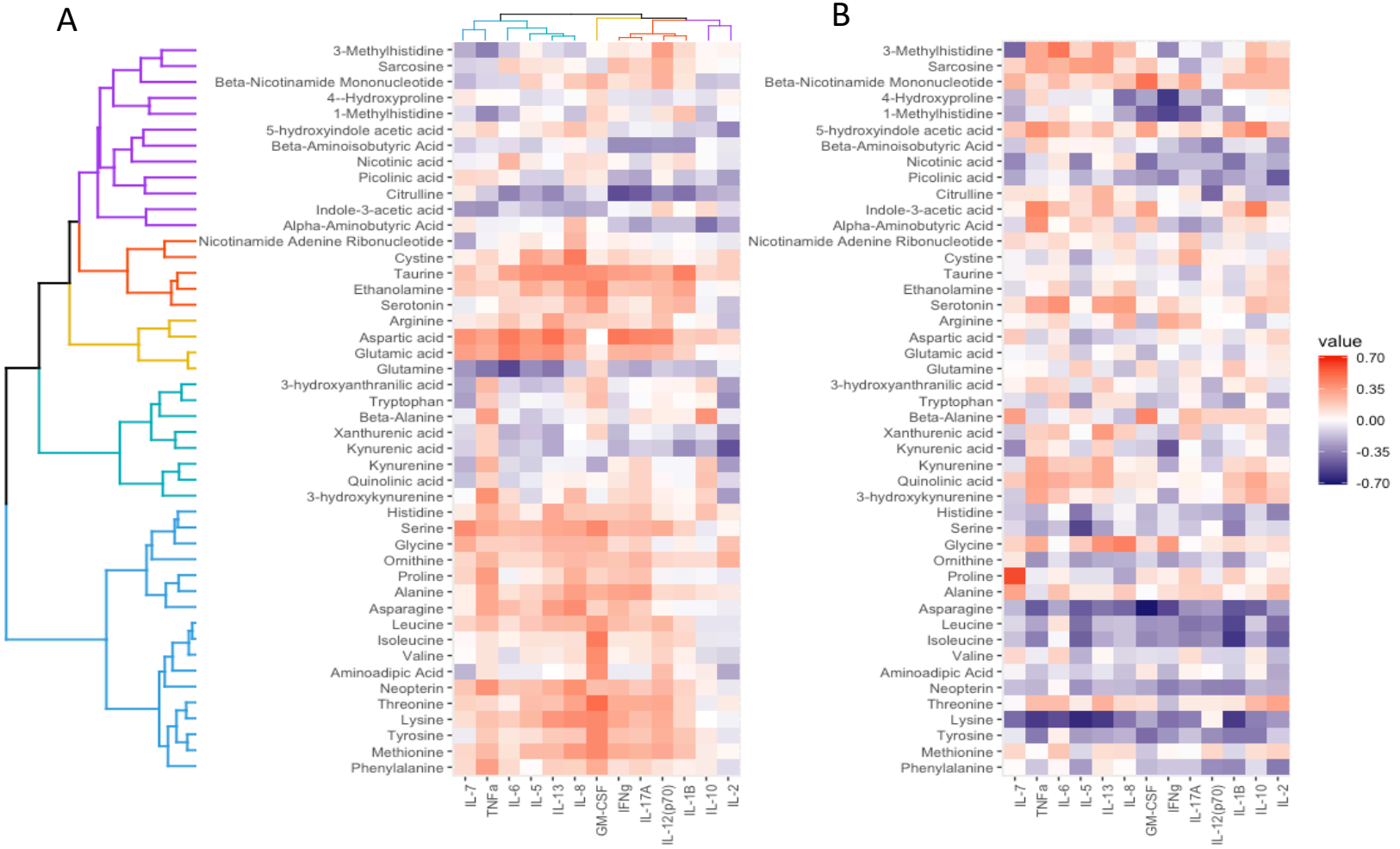



**Figure 5.8:** Correlation heatmaps integrating biogenic amines and tryptophan pathway metabolite concentrations and circulating cytokines. A) Measurements from burn injury patients. B) Measurements from healthy controls. Hierarchical clustering based on clustering from burn injury results and mirrored for direct comparison to healthy age and sex matched controls.

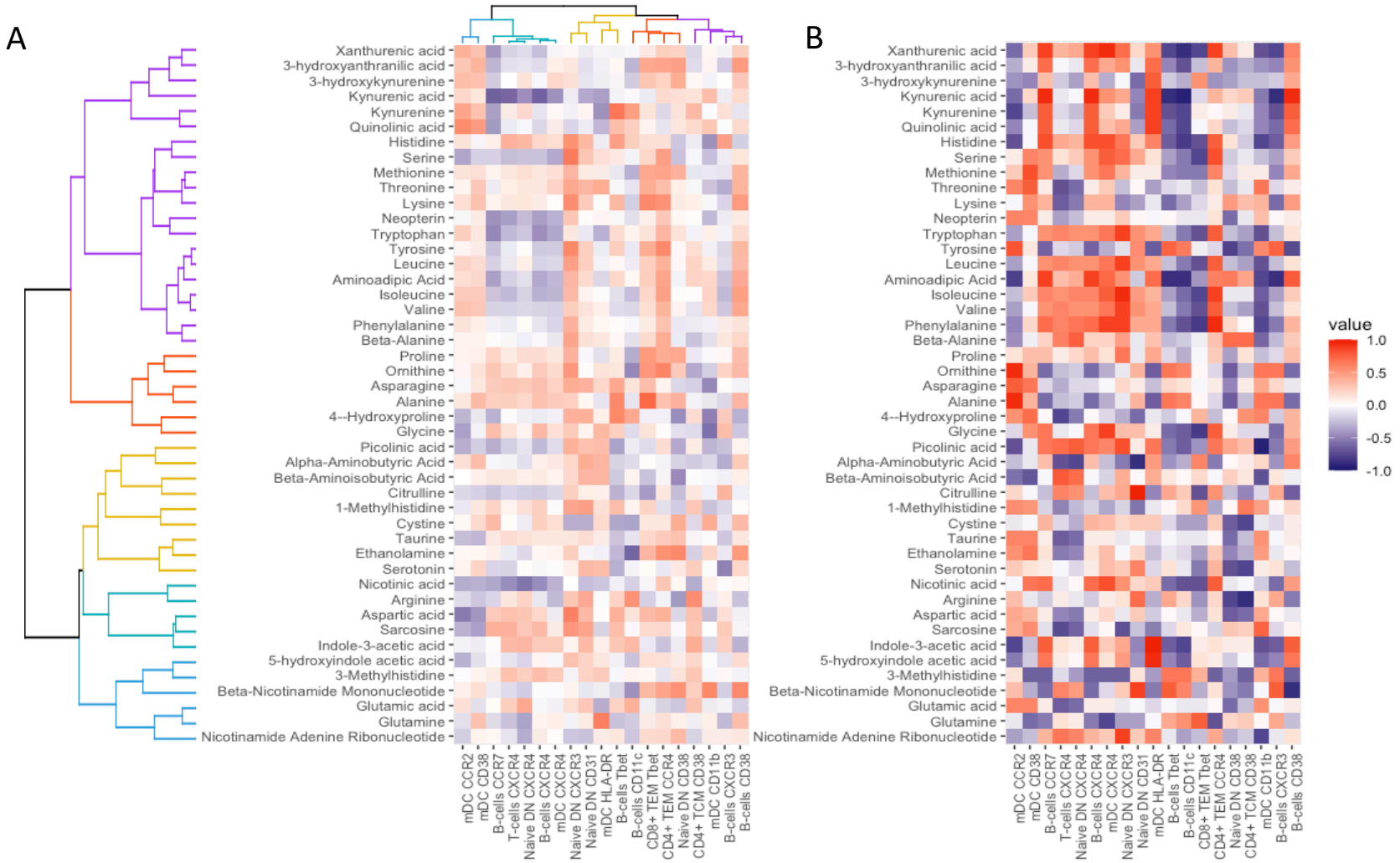



**Figure 5.9:** Correlation heatmaps integrating immune cell phenotyping (mass cytometry) and biogenic amines and tryptophan pathway metabolites. A) Measurements from burn injury patients. B) Measurements from non-burn injury controls. Hierarchical clustering based on clustering from burn injury results and mirrored for direct comparison to healthy age and sex matched controls.

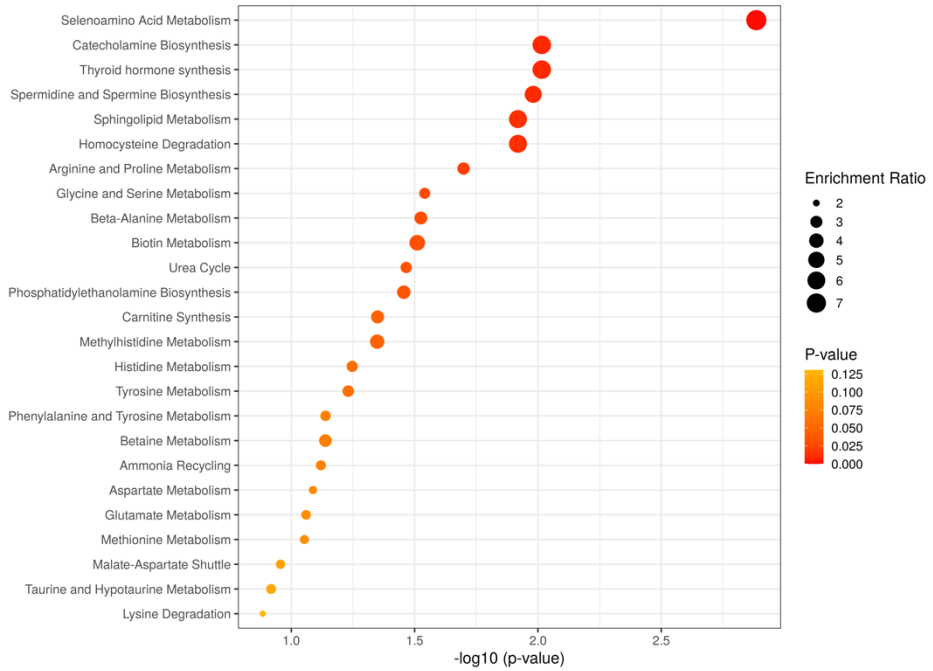



**Figure 5.10: Metabolite set enrichment analysis (MSEA)**- associated pathways and networks associated with the abundance of metabolites in burn injury patients compared to healthy controls.

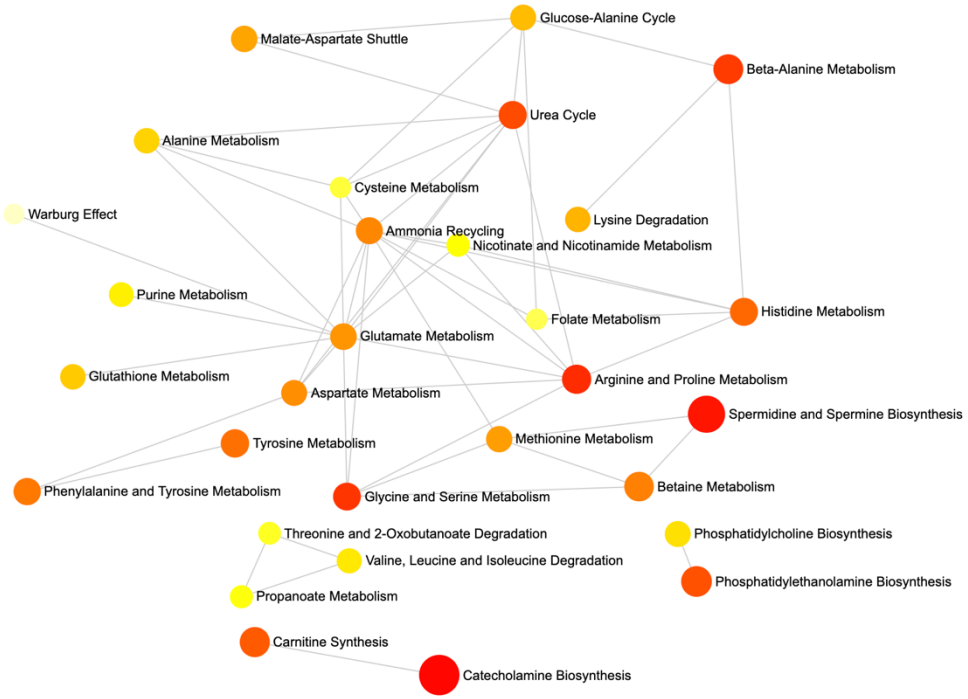



**Figure 5.11: Network analysis of MSEA results**- This figure illustrates the connections between each pathway network and associations related to MSEA of metabolite abundance and proportional significance between burn injury and healthy controls.

### 5.3.4 Metabolite Set Enrichment Analysis

Metabolite Set Enrichment Analysis of all quantified metabolites against the Small Molecule Pathway Database (figure 5.10) showed a multitude of metabolic pathway activation based on differences between burn injury patients and non-burn injury controls. The most significant pathway indicated was selenoamino acid metabolism, with MSEA driven by the high levels of serine and lysine in burn injury patients in comparison to healthy controls, with an enriched Fischer's p value of 0.001. However, in the absence of other selenoamino acid compounds, it is hard to test the validity of this finding. The following subsequent top 4 pathways was catecholamine biosynthesis driven by changes in glutamine and glutamate biosynthesis, thyroid hormone synthesis, spermine and spermidine synthesis and sphingolipid metabolism. Interestingly, all pathways are involved with cellular processes, including cell death and the immune activation. Further network analysis (figure 5.11) presented a complex molecular network, centred around liver metabolism, with differences between burn injury patients and healthy controls suggested to be strongly driven by differences in beta-alanine metabolism and urea cycle dysfunction.

### 5.3.5 Standard 1D, CPMG and DIRE supervised multivariate analysis

3 pulse sequences were used for plasma NMR acquisition, with each sequence providing a different analytical resolve of features. PCA modelling of 1D standard NMR spectra (figure 5.4A) shows a large overlap in burn injury and non-burn controls, with a small cluster of burn injury patients mapping separately. OPLS-DA modelling (figure 5.12B) produces a significant model, with two clear clusters of groups (R2X 0.382, R2Y 0.927, Q2Y 0.516).

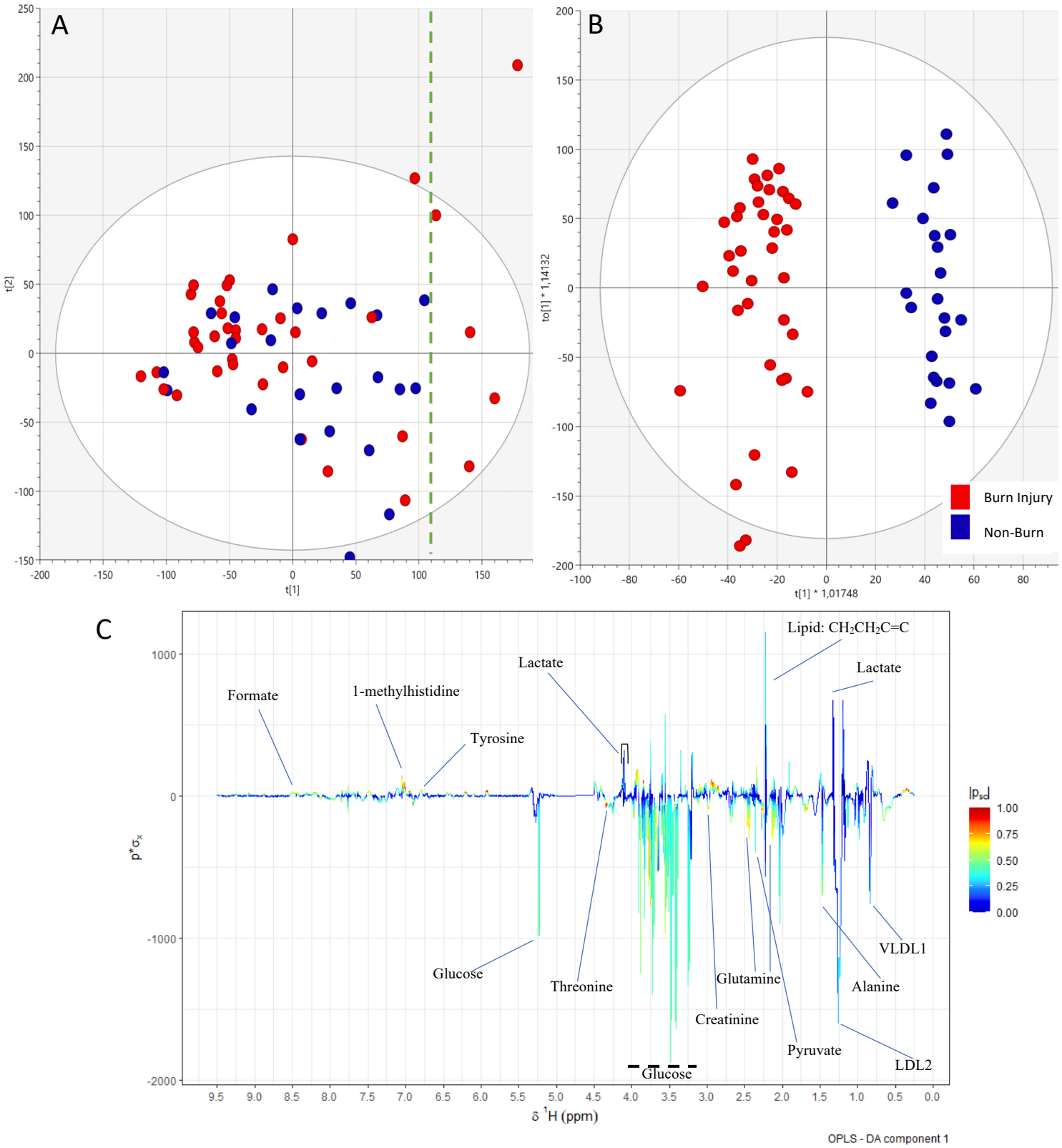



**Figure 5.12: PCA and OPLS-DA of standard 1D NMR spectra between burn injury patients and healthy controls.** A) PCA scores plot of burn injury patients and healthy control profiles $R^2X$ 0.595, $Q^2Y$ 0.34). D) OPLS-DA scores plot of burn injury patients and healthy controls ($R^2X$ 0.382, $R^2Y$ 0.927, $Q^2Y$ 0.516). C) Loadings from OPLS-DA modelling, highlighting the metabolites with the most influence in the model. Colour is equivalent to the absolute value of the correlation loading.

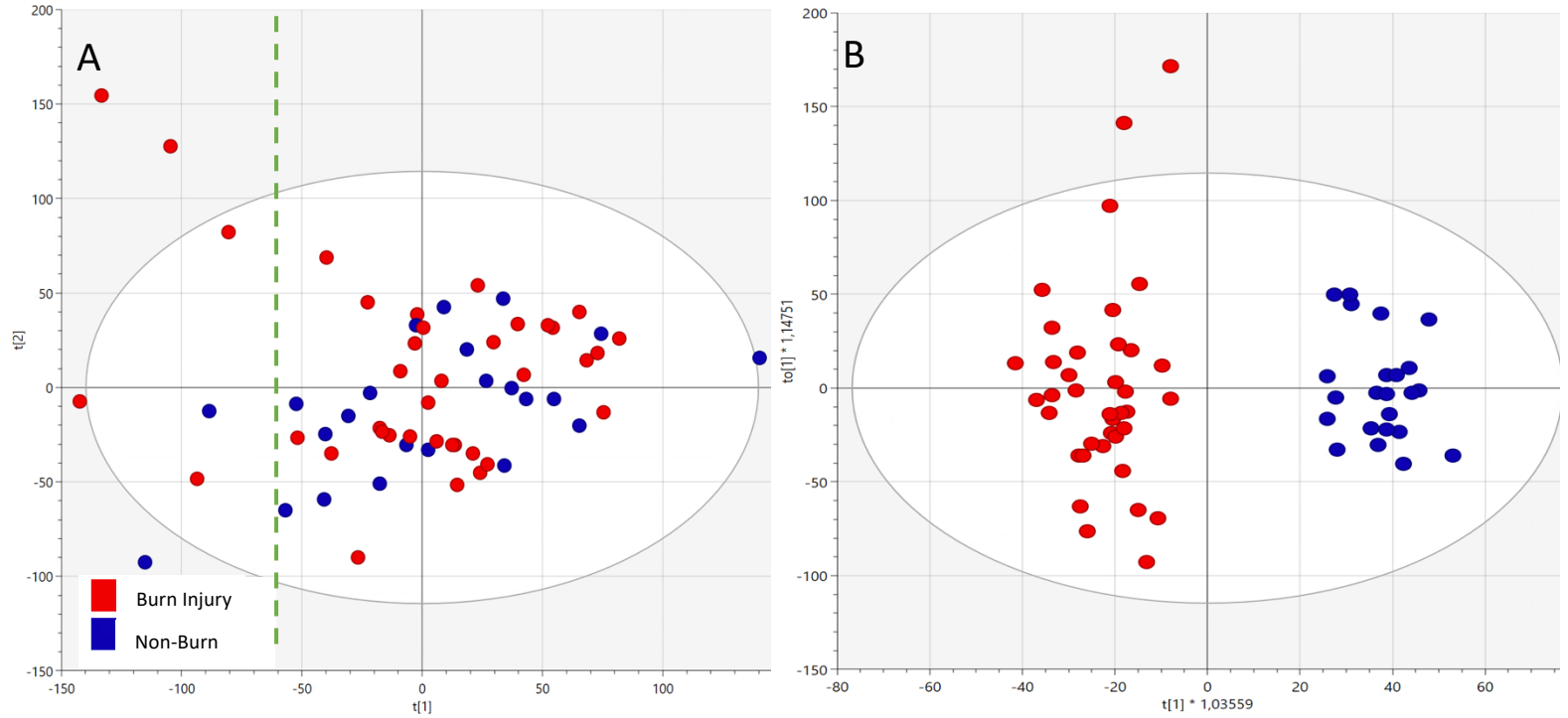



**Figure 5.13: PCA and OPLS-DA of CPMG NMR spectra between burn injury patients and healthy controls.** A) PCA scores plot of burn injury patients and healthy control profiles ($R^2X$ 0.351, $Q^2Y$ 0.116). B) OPLS-DA scores plot of burn injury patients and healthy controls ($R^2X$ 0.092, $R^2Y$ 0.932, $Q^2Y$ 0.462).

Loadings from the OPLS-DA model show discriminatory metabolites between burn injury and non-burn to be LDLs, glutamine, glucose, pyruvate, lactate and alanine. Figure 5.13A shows a PCA model of CPMG NMR spectra, presenting similar findings to the 1D standard NMR spectra model, with a subset of burn injury outliers. OPLS-DA modelling against presented a strong model with clear discrimination of burn injury and non-burn control groups (R2X 0.092, R2Y 0.932, Q2Y 0.462) (figure 5.13B), with a cluster of outliers in the burns group. Figure 5.14A and 5.14B show the aliphatic and aromatic regions from the loadings of the OPLS-DA model, with 8 discriminatory metabolites further identified through STOCSY and validated by in-house metabolite identification databases and HMDB. Figure 5.14A-1 confirms the identification of pyruvate at $\delta$ 2.3629, with high correlation to lactate and alanine. Figure 5.14A-2, shows high correlation of ketone bodies, substances created by the liver during gluconeogenesis, showing high correlation of 3-D-hydroxybutyrate, acetoacetone, acetone and acetate.

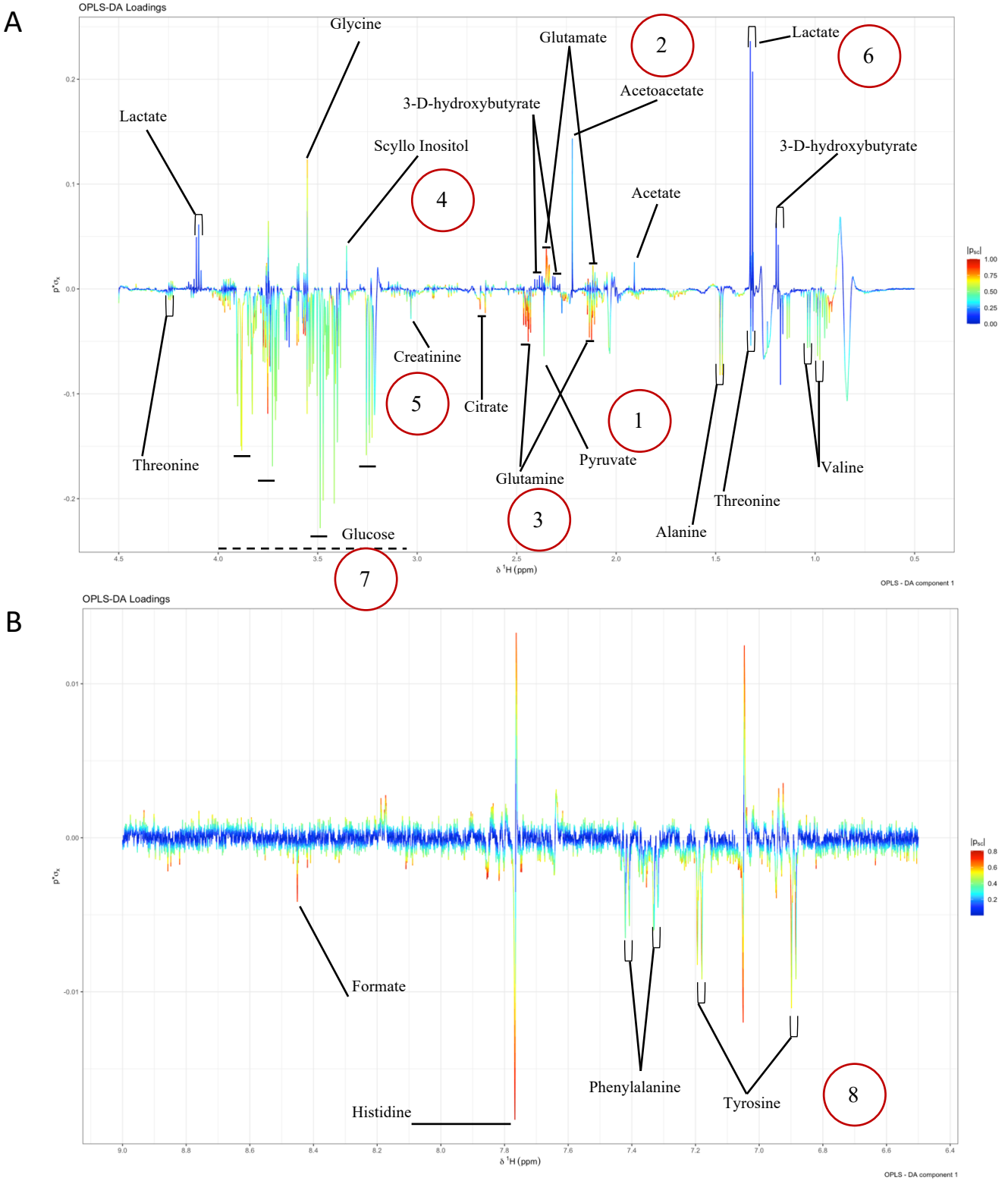
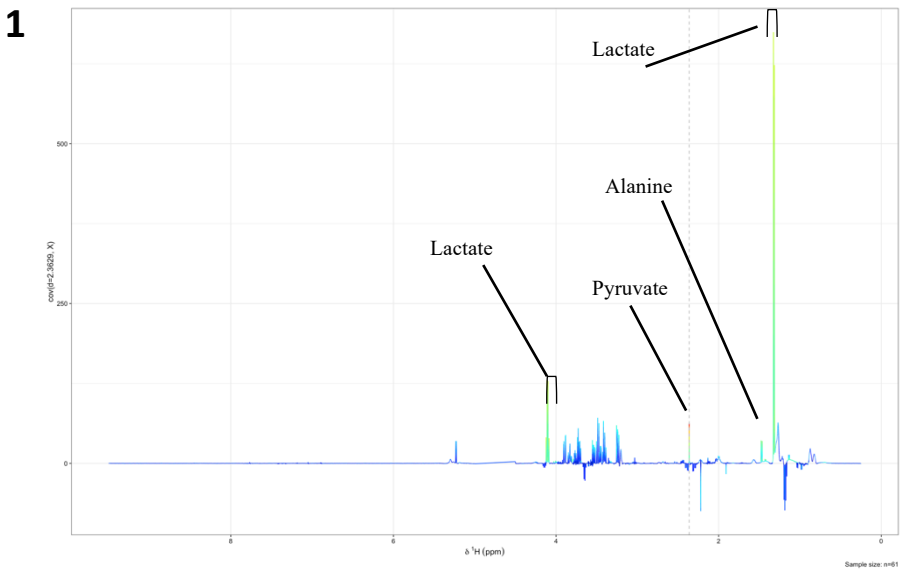
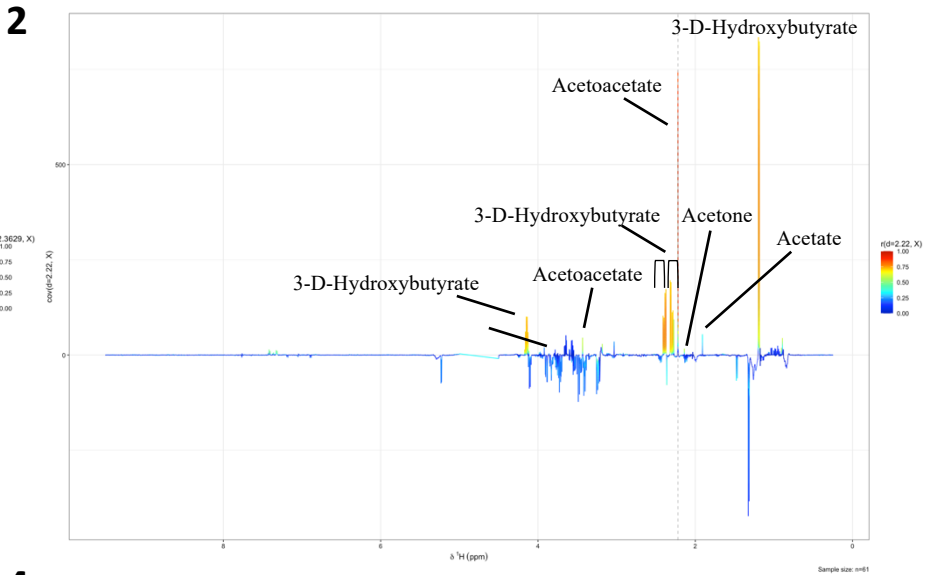
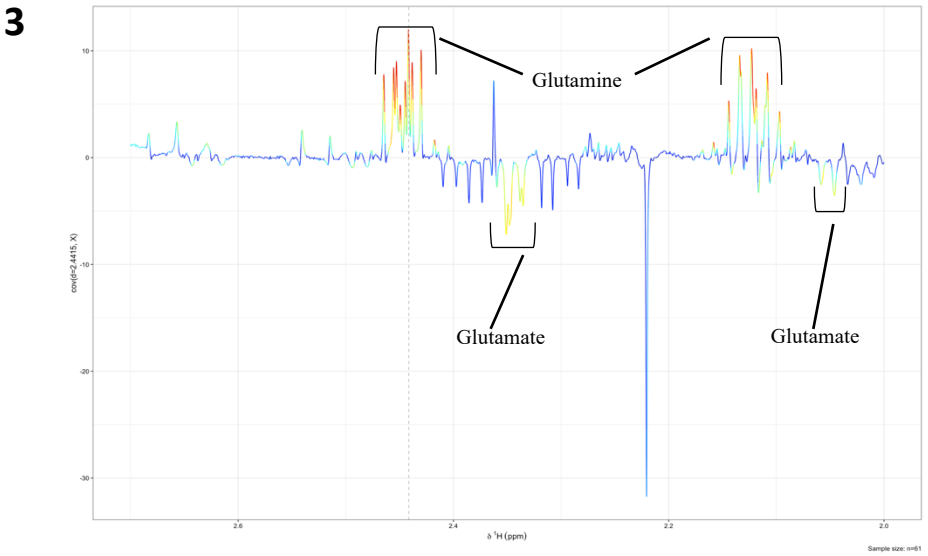
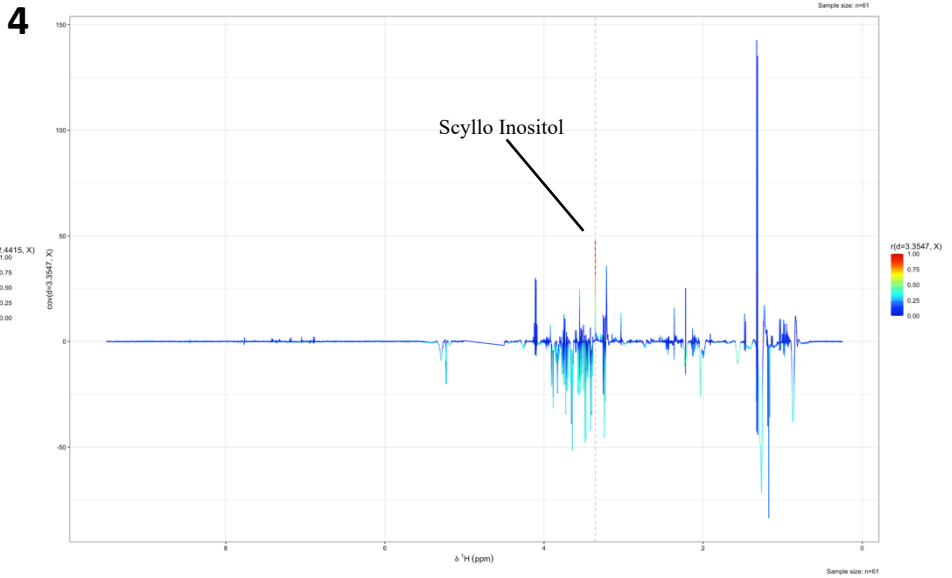



**Figure 5.14:** Loadings from CPMG spectra OPLS-DA model, discriminating burn injury and non-burn controls, highlighting the metabolites with the most influence in the model projection. Colour is equivalent to the absolute value of the correlation loading. A) $\delta$ 0.5-4.5 and B) $\delta$6.5-9.0. Corresponding STOCSYs from labelled regions are found on the following pages.

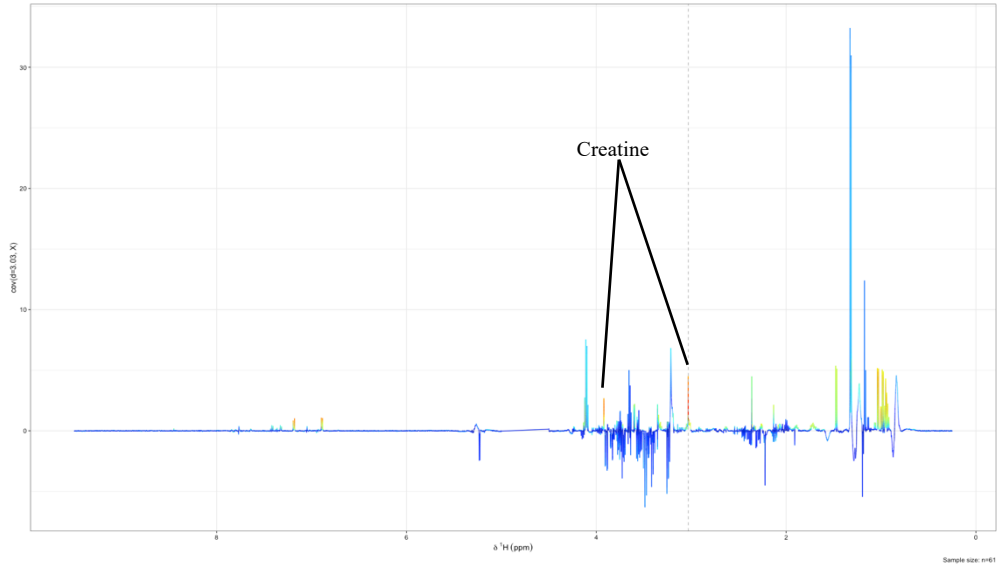

1
Lactate
Alanine
Lactate
Pyruvate


2
3-D-Hydroxybutyrate
Acetoacetate
3-D-Hydroxybutyrate
Acetone
Acetate
3-D-Hydroxybutyrate
Acetoacetate

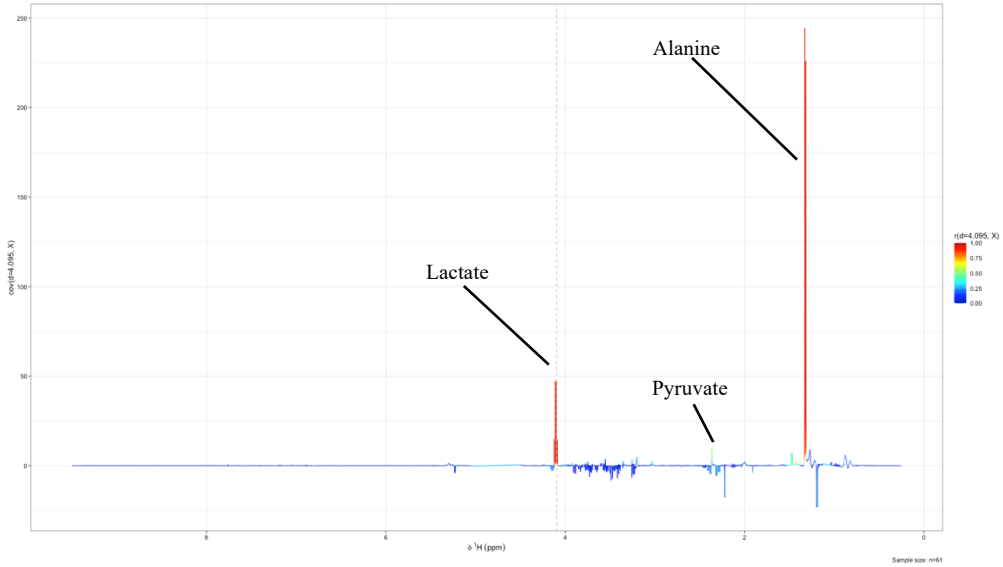

3
Glutamine
Glutamate
Glutamate


4
Scyllo Inositol

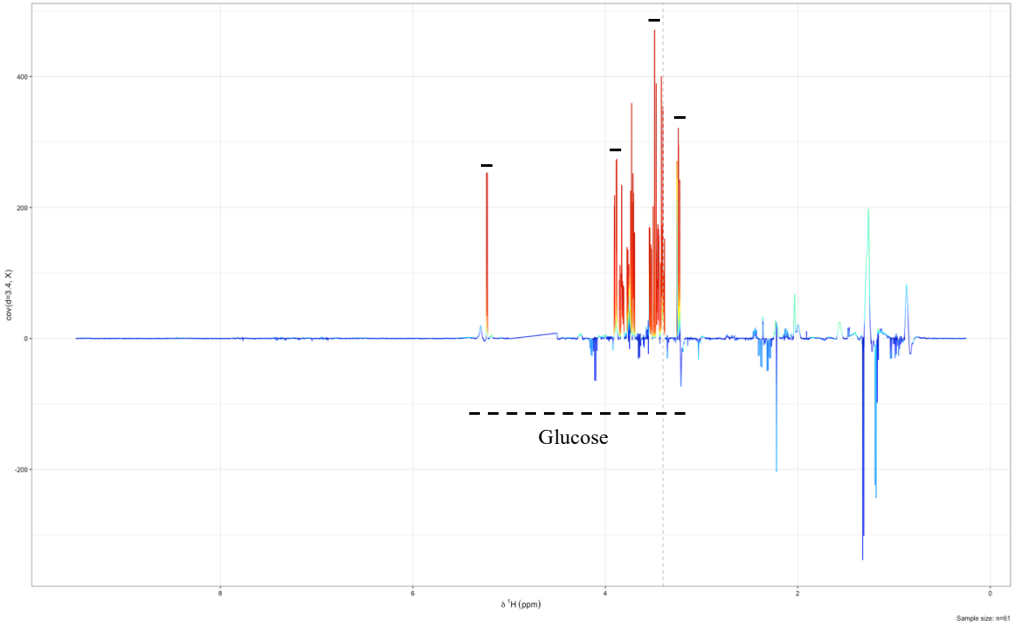
5


Creatine

6


Alanine
Lactate
Pyruvate

7


Glucose

8

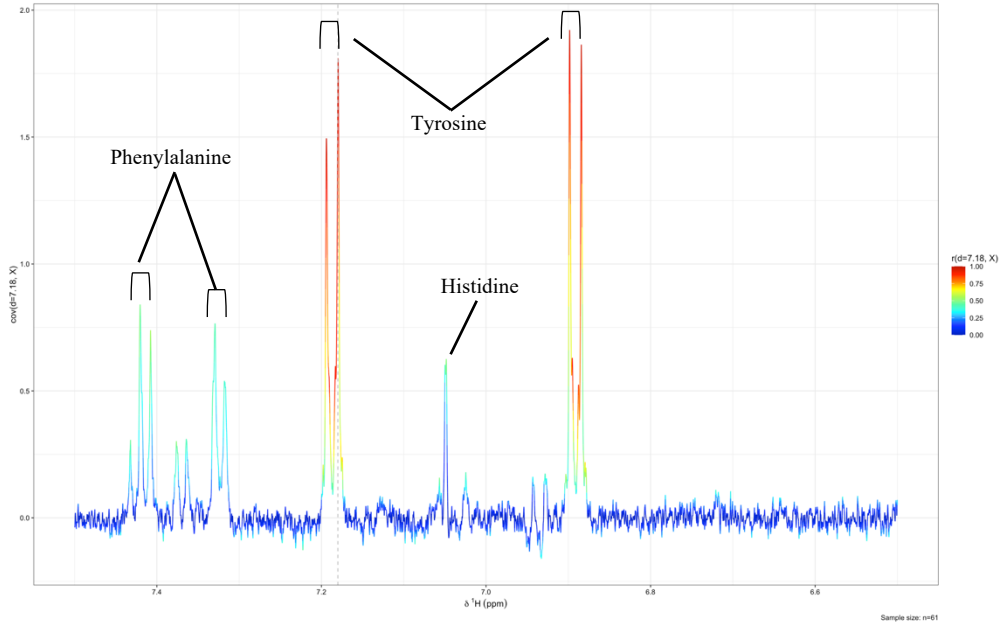

Tyrosine
Phenylalanine
Histidine

Figure 5.14A-3 shows correlation of glutamate at $\delta$2.4415 and glutamine peaks. Scyllo inositol, an isomer of glucose and neuro-signalling molecule, was identified at $\delta$3.3547 (figure 5.14A-4), showing correlation to glucose and lipoproteins. Figure 5.14A-5 identifies creatinine at $\delta$3.03. Figure 5.14A-6 identifies lactate at $\delta$4.095 and corresponding pyruvate and alanine correlation. Multiple peaks attributed to glucose are identified at $\delta$3.4 (figure 5.14A-7). Figure 5.14B-8 identifies tyrosine at $\delta$7.18 and the corresponding correlation to phenylalanine and histidine.

PCA modelling of DIRE data shows very little clustering (figure 5.15A), however OPLS-DA presents a very strong model (figure 5.15B) (R2X 0.15, R2Y 0.989, Q2Y 0.263), with two distinct clusters between burn injury and non-burn. Correlations from loadings (figure 5.15C) show that this is primarily driven by choline species, GlycA and GlycB. Although not a significant OPLS-DA model (figure 5.16B) (R2X 0.111, R2Y 0.949, Q2Y -0.219), when applying class discrimination to vaccine responders and non-responders, the model is driven by branched sugar chains from glycoproteins and choline species (figure 5.16C). All amino acids appear to be elevated such as alanine and lysine, consistent mass spectrometry results.

Integrals from DIRE spectra were calculated (table 5.4) to assess differences between burn injury and non-burn. Elevated GlycA, GlycB, SpC and GlycA+B and depleted SpC+GlycA were measured for burn injury compared to non-burn control counterparts (table 4). Violin plots were used to visualise the distribution of integrals (figure 5.17) and showed a broader spread of results for burn injury patients in comparison to controls, with GlycB found to be the only significant measure significant elevated in burn injury patients (pvalue<0.05) after multiple testing correction.

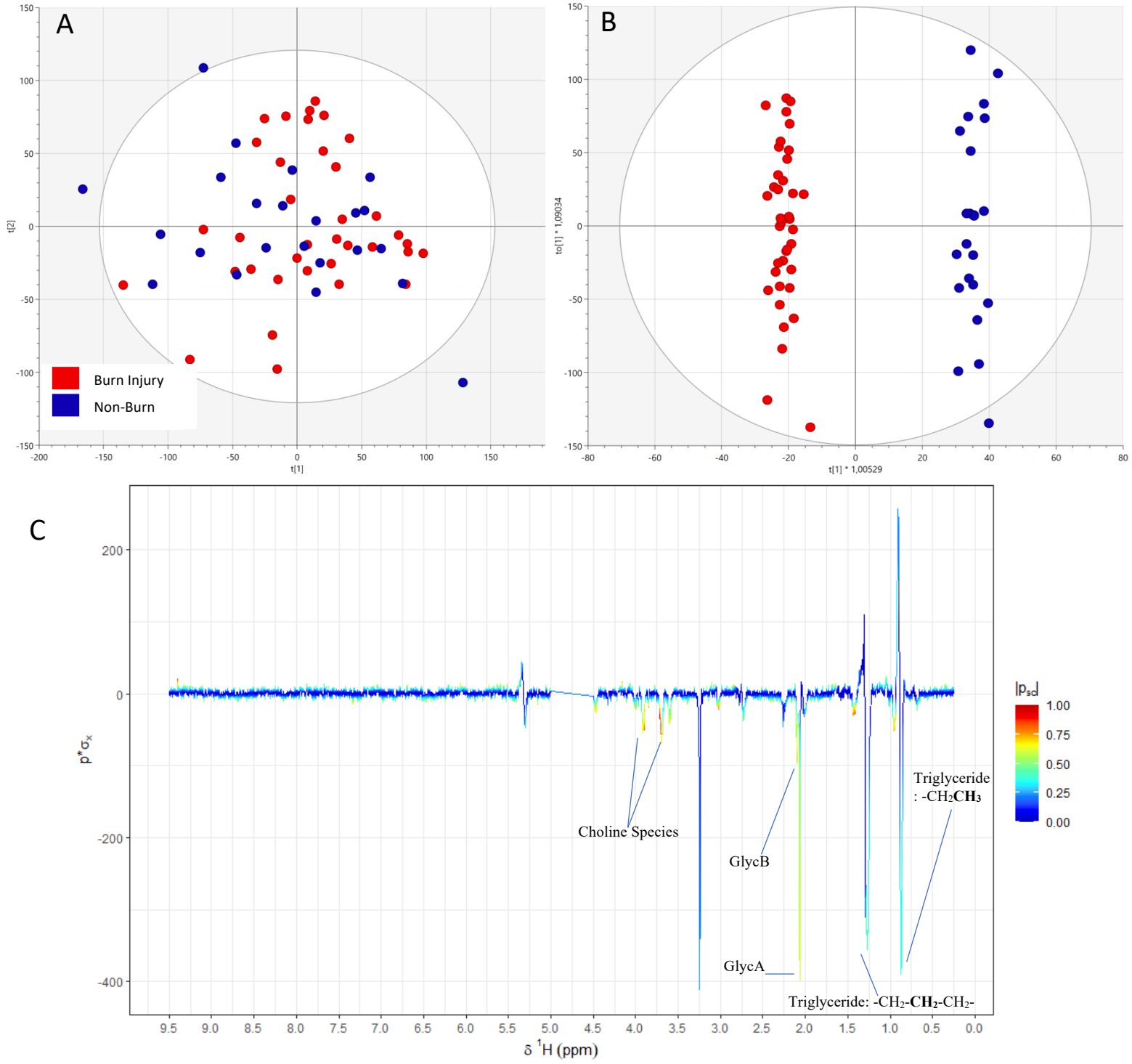



**Figure 5.15: PCA and OPLS-DA of DIRE NMR spectra between burn injury patients and healthy controls.** A) PCA scores plot of burn injury patients and healthy control profiles ($R^2X$ 0.227, $Q^2Y$ 0.0956). D) OPLS-DA scores plot of burn injury patients and healthy controls ($R^2X$ 0.15, $R^2Y$ 0.989, $Q^2Y$ 0.263). C) Loadings from OPLS-DA modelling, highlighting the metabolites with the most influence in the model. Colour is equivalent to the absolute value of the correlation loading.

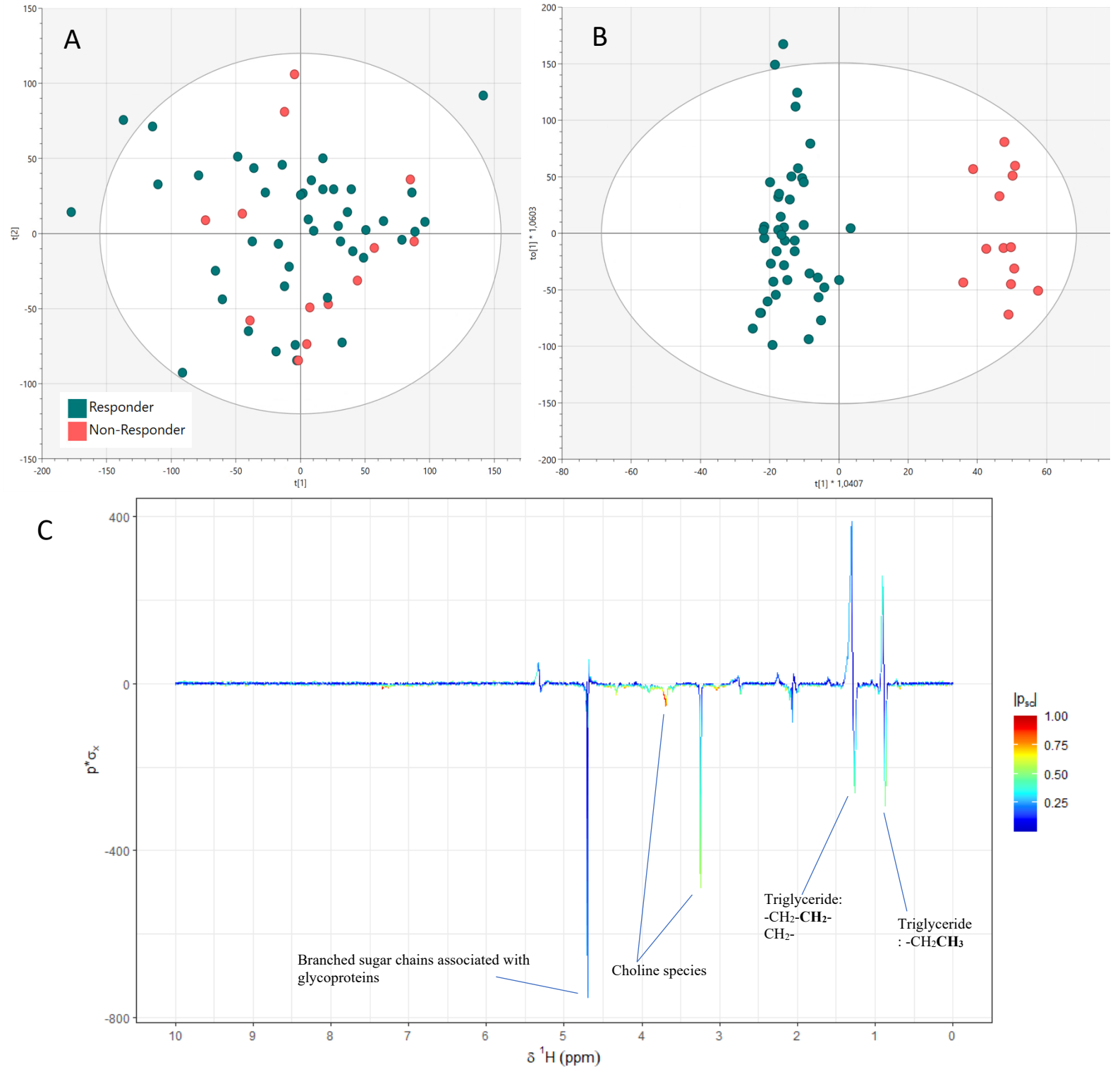



**Figure 5.16: PCA and OPLS-DA of DIRE NMR spectra, discriminating against vaccine response and non-response.** A) PCA scores plot of burn injury patients and healthy control profiles, and their corresponding vaccine response ($R^2Y$ 0.227, $Q^2Y$ 0.09). C) Scores plot of OPLS-DA model discriminating against vaccine response ($R^2X$ 0.111, $R^2Y$ 0.949, $Q^2Y$ -0.219). C) Loadings from OPLS-DA modelling, highlighting the metabolites with the most influence in the model. Colour is equivalent to the absolute value of the correlation loading.

**Table 5.4: Calculated concentrations of DIRE integrals comparing burn injury and non-burn controls.**

| | GlycA | GlycB | Spc | GlycA+GlycB | Spc+GlycA |
|---|---|---|---|---|---|
| | | | | pvalue=0.04 | |
| Burn | 1.53 x $10^7$ [1.20 x $10^7$ – 1.94 x $10^7$] | 3.82 x $10^6$ [2.81 x $10^6$ – 4.79 x $10^6$] | 3.43 x $10^7$ [2.09 x $10^6$ – 4.65 x $10^6$] | 1.90 x $10^7$ [1.49 x $10^7$ – 2.41 x $10^7$] | 2.24 [1.10 – 3.77] |
| Non-Burn | 1.47 x $10^7$ [1.29 x $10^7$ – 1.72 x $10^7$] | 3.39 x $10^6$ [2.96 x $10^6$ – 4.64 x $10^6$] | 3.27 x $10^7$ [2.48 x $10^6$ – 4.61 x $10^6$] | 1.80 x $10^7$ [1.61 x $10^7$ – 2.18 x $10^7$] | 2.27 [1.57 – 3.38] |

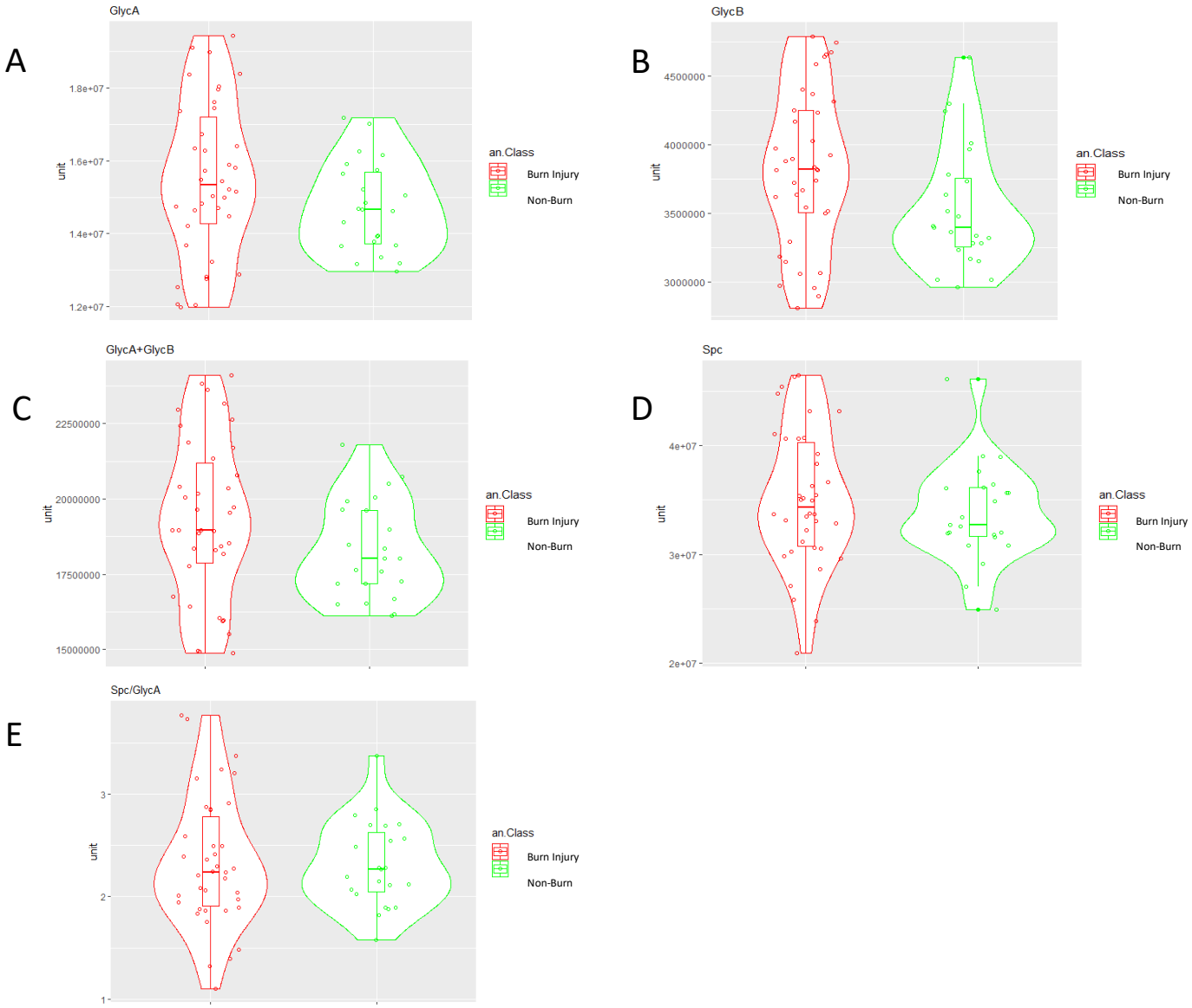



**Figure 5.17: Violin plots showing distribution of calculated DIRE integrals.** Relative concentration of DIRE integrals calculated for burn injury and non-burn controls. A) N-acetyl glycoprotein A (GlycA); B) N-acetyl glycoprotein B (GlycB); C) combined concentration of GlycA and GlycB; D) Supramolecular phospholipid compound (SpC); E) combined concentration of SpC and GlycA. Pvalue <0.05**,** corrected for multiple testing using Benjamini-Hochberg.

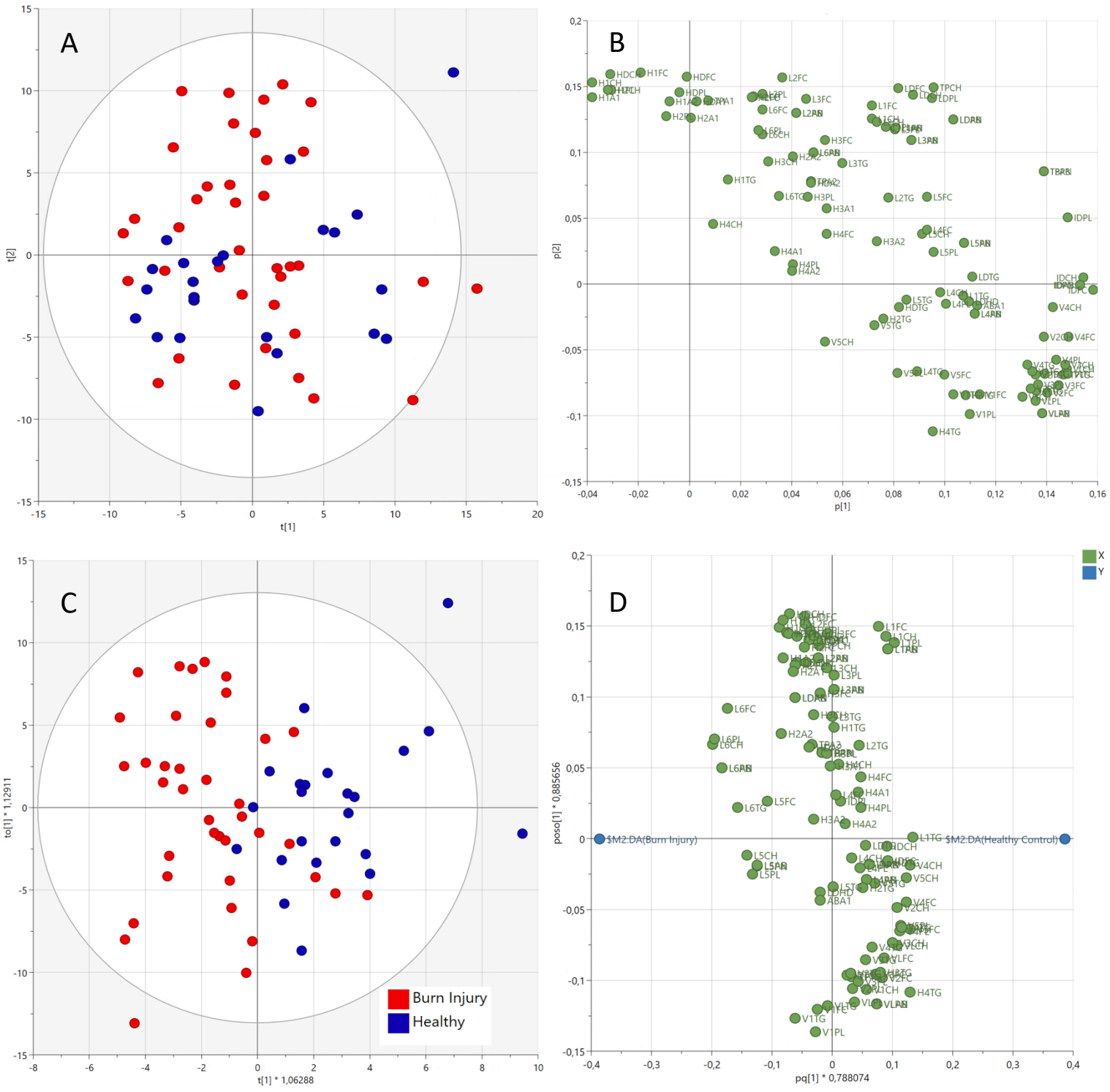



**Figure 5.18: PCA and OPLS-DA of quantified lipoprotein subfractions.** A) PCA scores plot of burn injury patients and healthy control profiles ($R^2Y$ 0.987, $Q^2Y$ 0.933). B) PCA loadings plot of burn injury patients and healthy control profiles. C) Scores plot of OPLS-DA model, presenting the most significant metabolites influencing the model clustering. ($R^2X$ 0.437, $R^2Y$ 0.492, $Q^2Y$ 0.321). D) Loadings plot of OPLS-DA model, presenting the most significant metabolites influencing the model clustering and discrimination between burn injury and healthy controls.

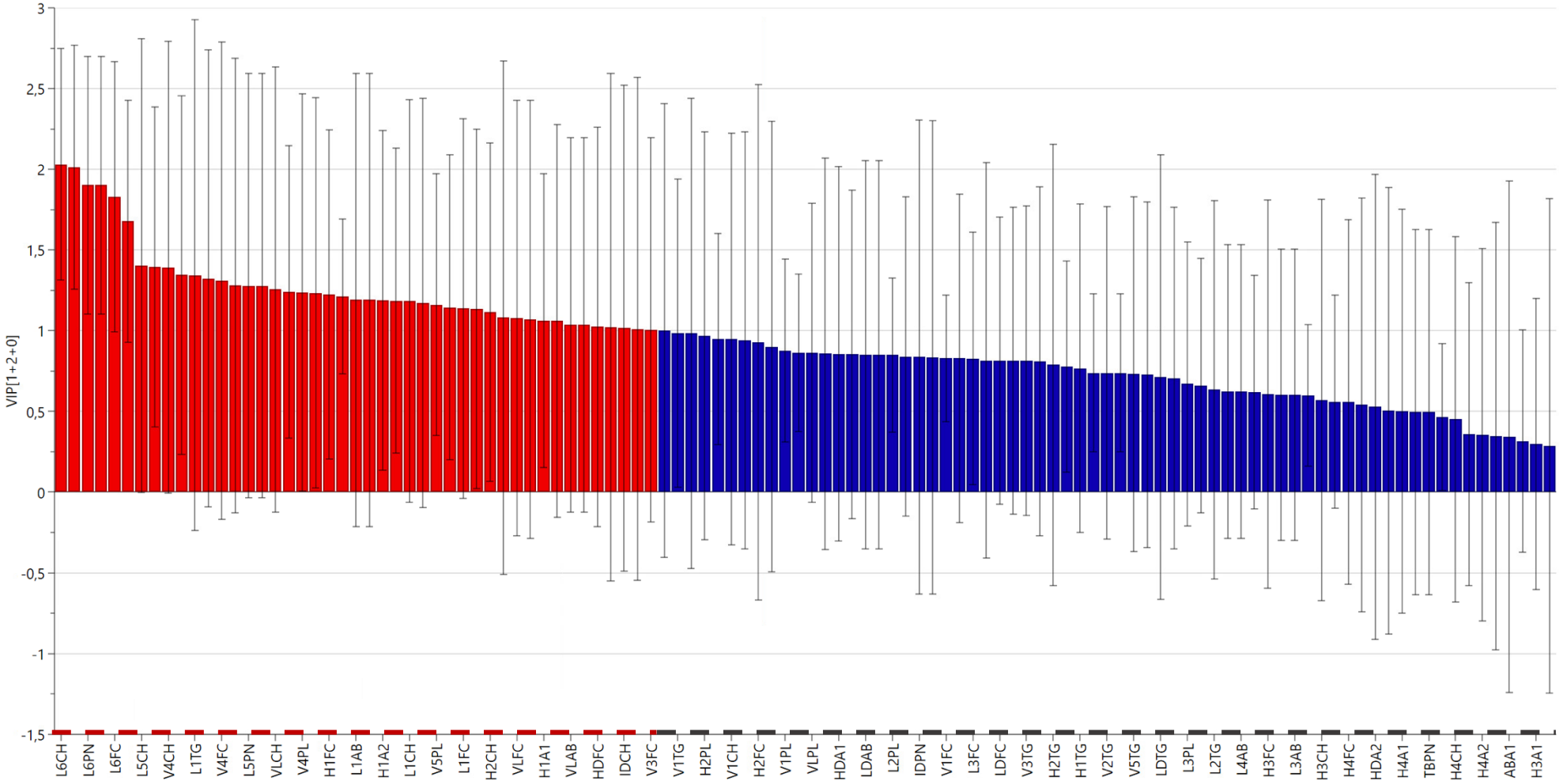



**Figure 5.19: OPLS-DA Variable importance of Projection (VIP):** Arbitrary threshold of 1 applied, showing the most discriminant metabolites in the OPLS-DA model discriminating burn injury and healthy controls. Underlined in red are the lipoprotein parameters the most model substantial and underlined in blue are the lipoprotein parameters with decreasing model influence

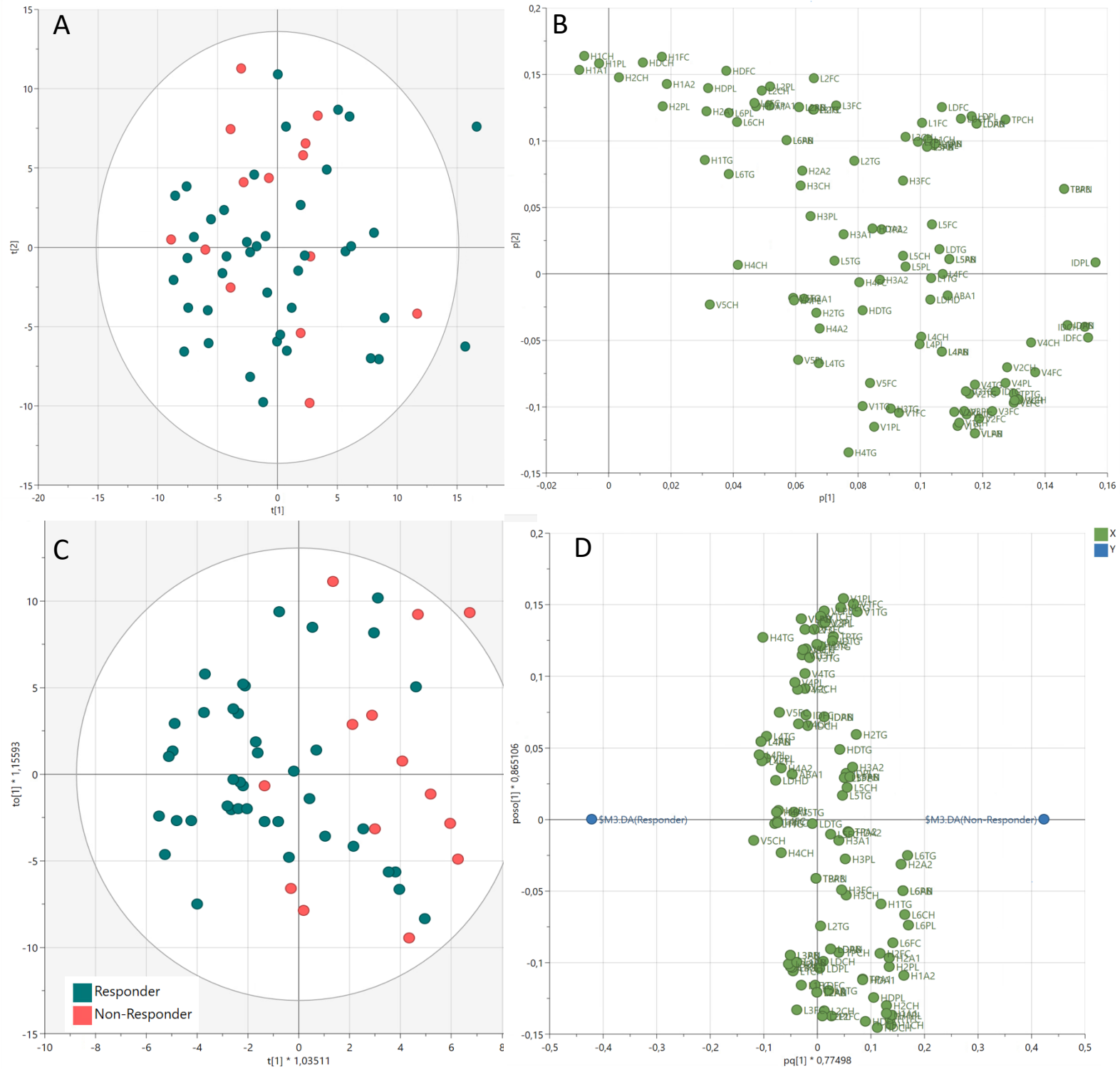



**Figure 5.20: PCA and OPLS-DA of quantified lipoprotein subfractions, discriminating against vaccine response and non-response.** A) PCA scores plot of burn injury patients and healthy control profiles, and their corresponding vaccine response ($R^2X$ 0.992, $Q^2$ 0.94). B) Corresponding PCA loadings plot of vaccine response model C) Scores plot of OPLS-DA model discriminating against vaccine response ($R^2X$ 0.342, $R^2Y$ 0.305, $Q^2$ 0.00606). D) Loadings plot of OPLS-DA model, presenting the most significant metabolites influencing the model clustering and discrimination between vaccine responders and non-responders.

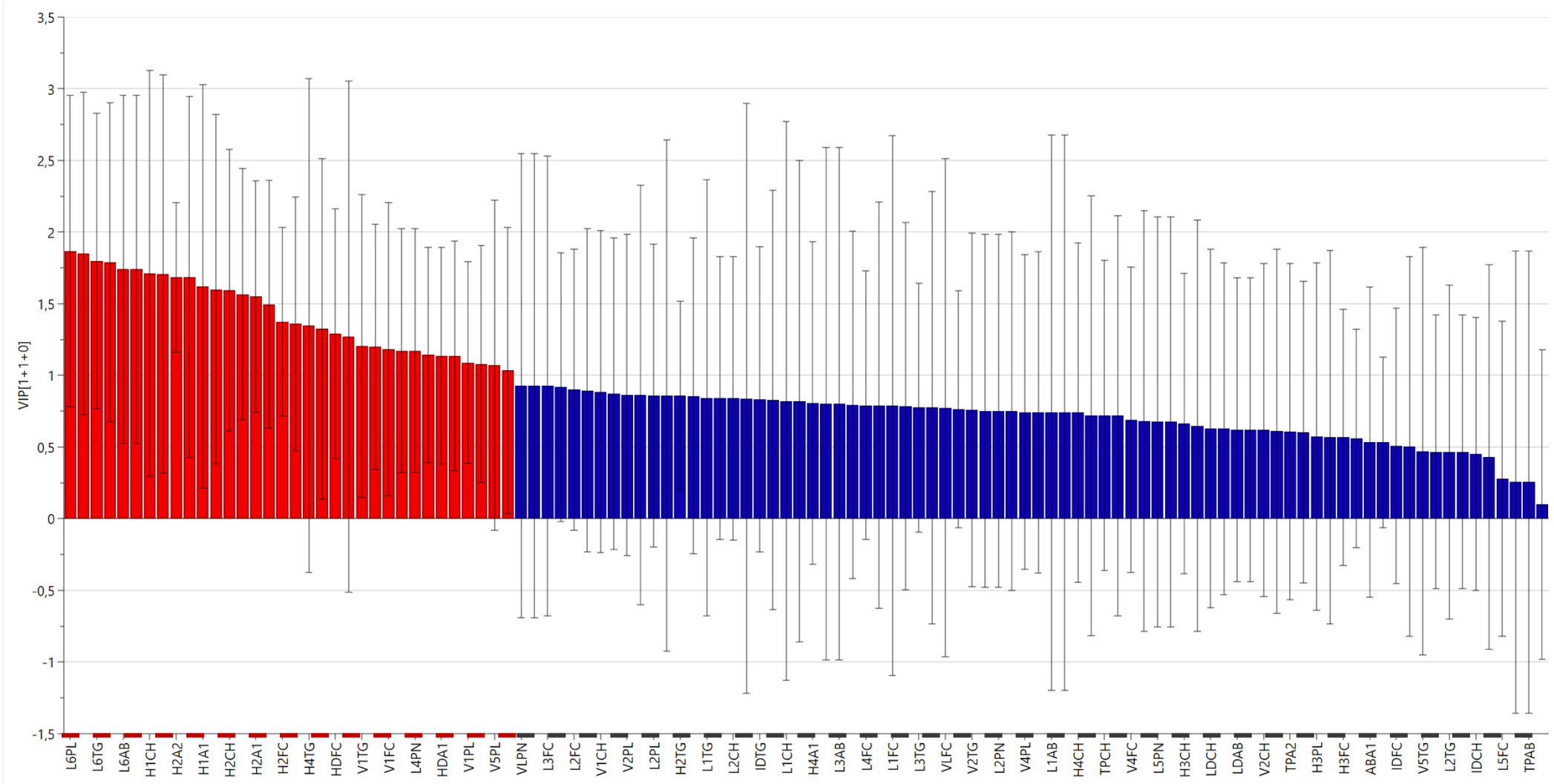



**Figure 5.21: OPLS-DA Variable importance of Projection (VIP):** Arbitrary threshold of 1 applied, showing the most discriminant metabolites in the OPLS-DA model discriminating vaccine response. Underlined in red are the lipoprotein parameters the most model substantial and underlined in blue are the lipoprotein parameters with decreasing model influence

### 5.3.6 Lipoprotein subfraction analysis

Lipoproteins were quantified according to the Bruker iVDR method. PCA modelling of lipoproteins shows a cluster of burn injury patients in the upper first principal component (figure 5.18A). OPLS-DA presents a strong model ($R^2X$ 0.437, $R^2Y$ 0.492, $Q^2Y$ 0.321), with a higher degree of overlap compared to spectral models, which would be expected due to the reduced number of variables (figure 5.18B). Figure 5.19 presents the most important variables in the project; with the first 5 lipoproteins being L6CH, L6PN, L6FC, L5CH, V4CH, and L1TG. The majority of the subfractions driving the separation between burn injury and non-burn controls are attributed to the LDL group.

To investigate what lipoproteins may distinguish vaccine response, similar analyses were carried out, discriminating class on vaccine response and non-response. Little to no patterns and clusters were observed in PCA (figure 5.20A), and a very weak OPLS-DA was derived (figure 5.20C) (R2X 0.342, R2Y 0.305, Q2 0.00606). VIP plot analysis shows that the driving lipoproteins discriminating these two groups are predominantly LDLs and HDLs, with the top 5 discriminatory metabolites being L6PL, L6TG, L6AB, H1CH, and H2A2.

### 5. 3.7 Correlation mapping of DIRE integrals and quantified biogenic amino acids and tryptophan pathway metabolites

Spearman's correlation mapping of cytokines and DIRE integrals (GlycA, GlycB and SpC) presented two distinct profiles for burn injury (figure 5.22A) and non-burn controls (figure 5.22B), with 5 distinct clusters. The first cluster clusters primarily branched chain and

aromatic acid, with a higher degree of positive correlation seen in burn injury patients compared to non-burn controls. The second cluster correlates DIRE integrals, GlycA, GlycB and SpC, showing similar correlation patterns in both burn injury and non-burn. The third cluster correlates tryptophan metabolites, with more positive correlation seen in burn injury correlating picolinic acid, neopterin, beta-alanine, tryptophan, 3-hydroxyanthranillic acid, kynurenic acid and xanthurenic acid. In cluster 4, stronger correlations are observed in non-burn controls, clustering muscle catabolism and urea cycle metabolites such as glutamine and arginine. Weak correlation is observed between both groups in cluster 5.

### 5.3.8 Correlation mapping of cytokines and DIRE integrals

Figure 5.23 presents two distinct cytokine-DIRE correlation maps. Generally, more positive correlations are observed in burn injury patients with strong positive correlation of GlycA and GlycB to all cytokines bar GM-CSF and IL-7. Little correlation is observed between GlycA and cytokines, however interestingly positive correlation is observed between in GM-CSF and GlycA. Stronger correlations are observed for SpC in non-burn, as expected due to SpC being depleted in burn injury patients. Weaker correlations are observed for GlycA/GlycB and SpC/GlycA in non-burn, with strong positive correlation again observed between integrals and cytokines bar GM-CSF and IL-7. In non-burn, IL-6 is observed with very weak correlation against DIRE integrals, whereas in burn injury, IL-6 is strongly positively correlated with GlycA and GlycB, and with very weak correlation to SpC.

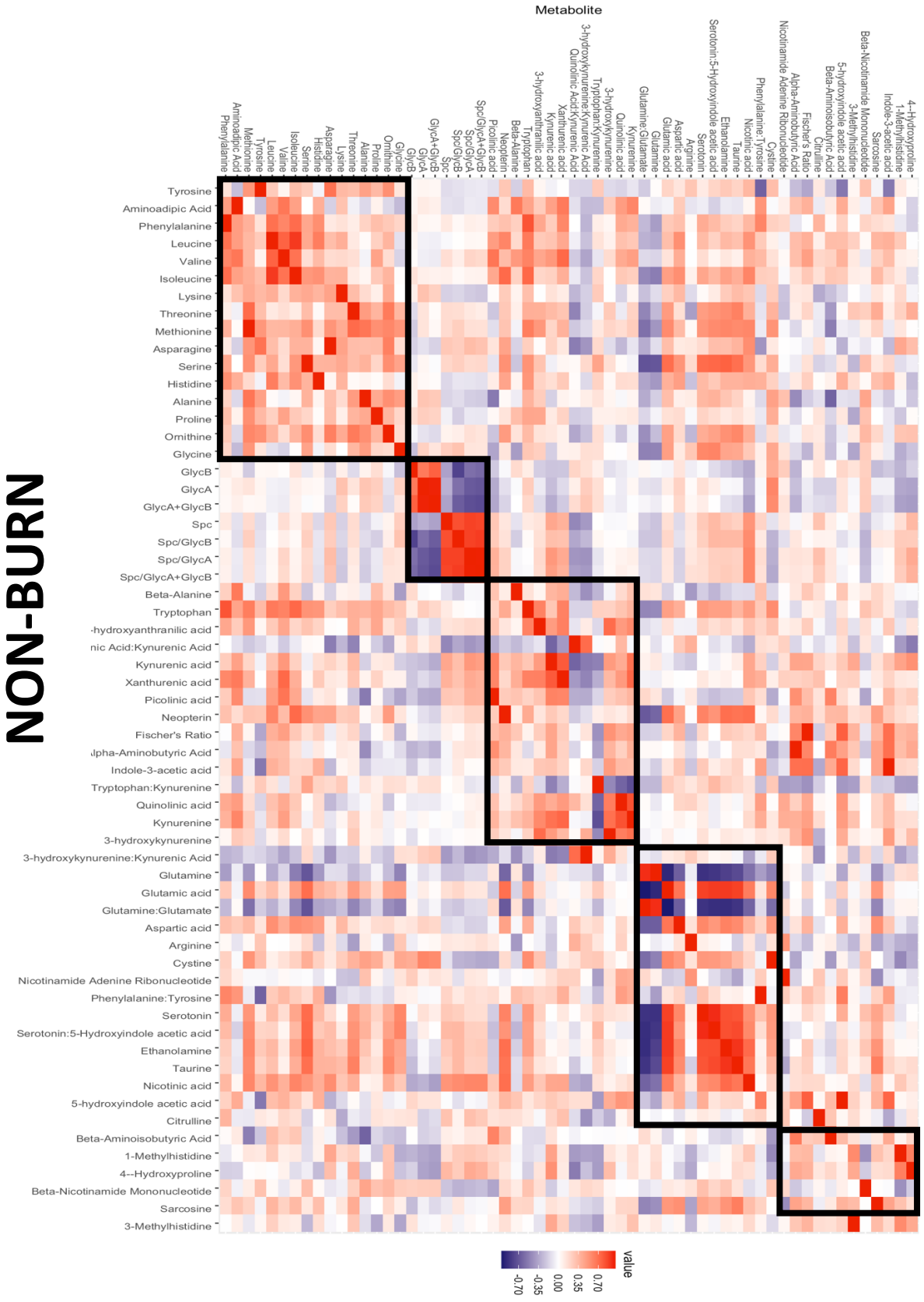



**Figure 5.22A: Metabolite-metabolite spearman's correlations.** Quantified biogenic amine and tryptophan metabolites and DIRE integrals correlated against one another, showing the signature present in non-burn controls. Black boxes highlight 5 distinct clusters. Metabolites are hierarchically clustered according to results derived from burn injury patients, and the subsequent order was applied to non-burn control results for direct motif comparison.

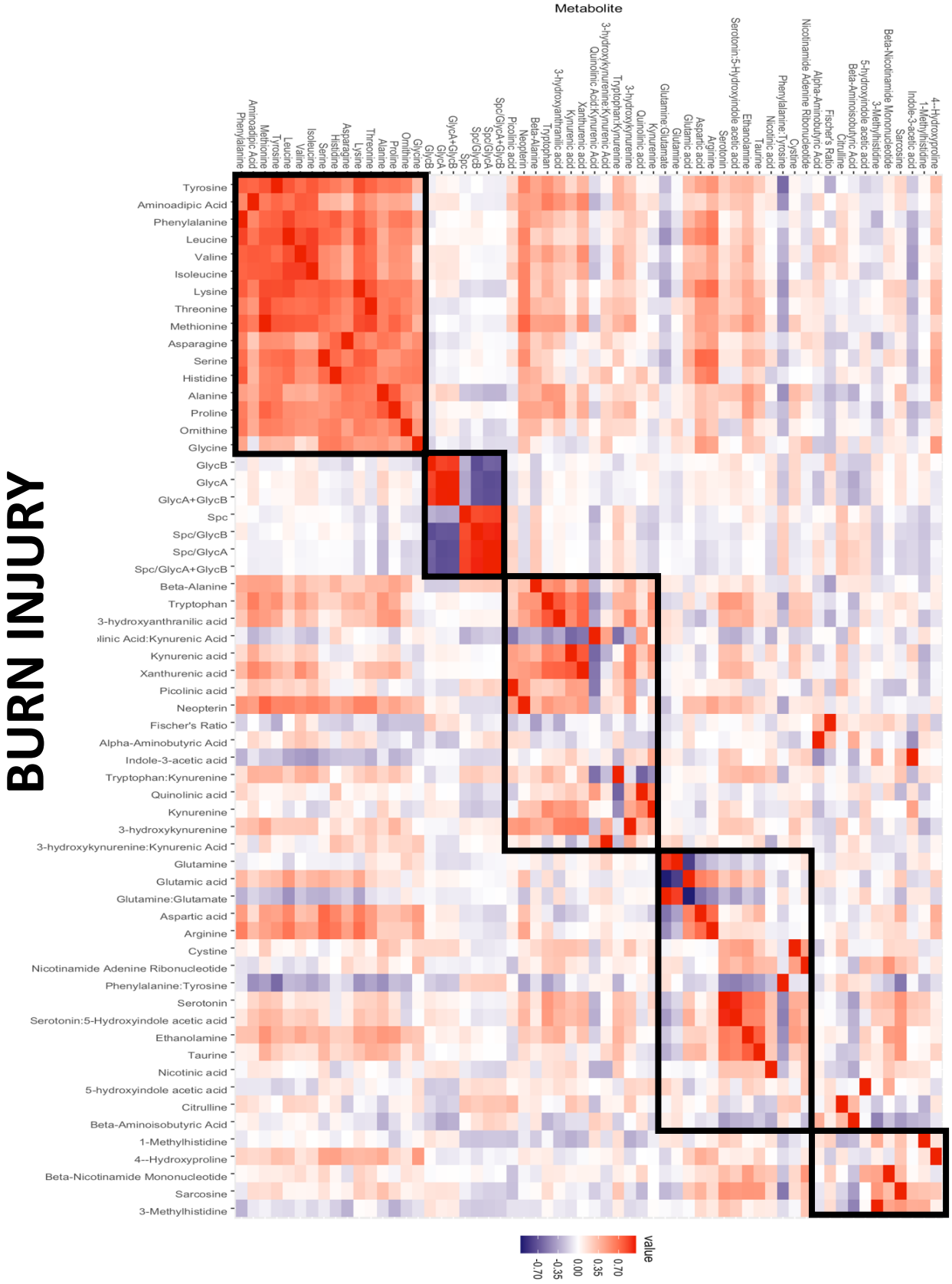



**Figure 5.22B: Metabolite-metabolite spearman's correlations.** Quantified biogenic amine and tryptophan metabolites and DIRE integrals correlated against one another, showing the signature present in burn injury patients. Black boxes highlight 5 distinct clusters. Metabolites are hierarchically clustered according to results derived from burn injury patients, and the subsequent order was applied to non-burn control results for direct motif comparison.

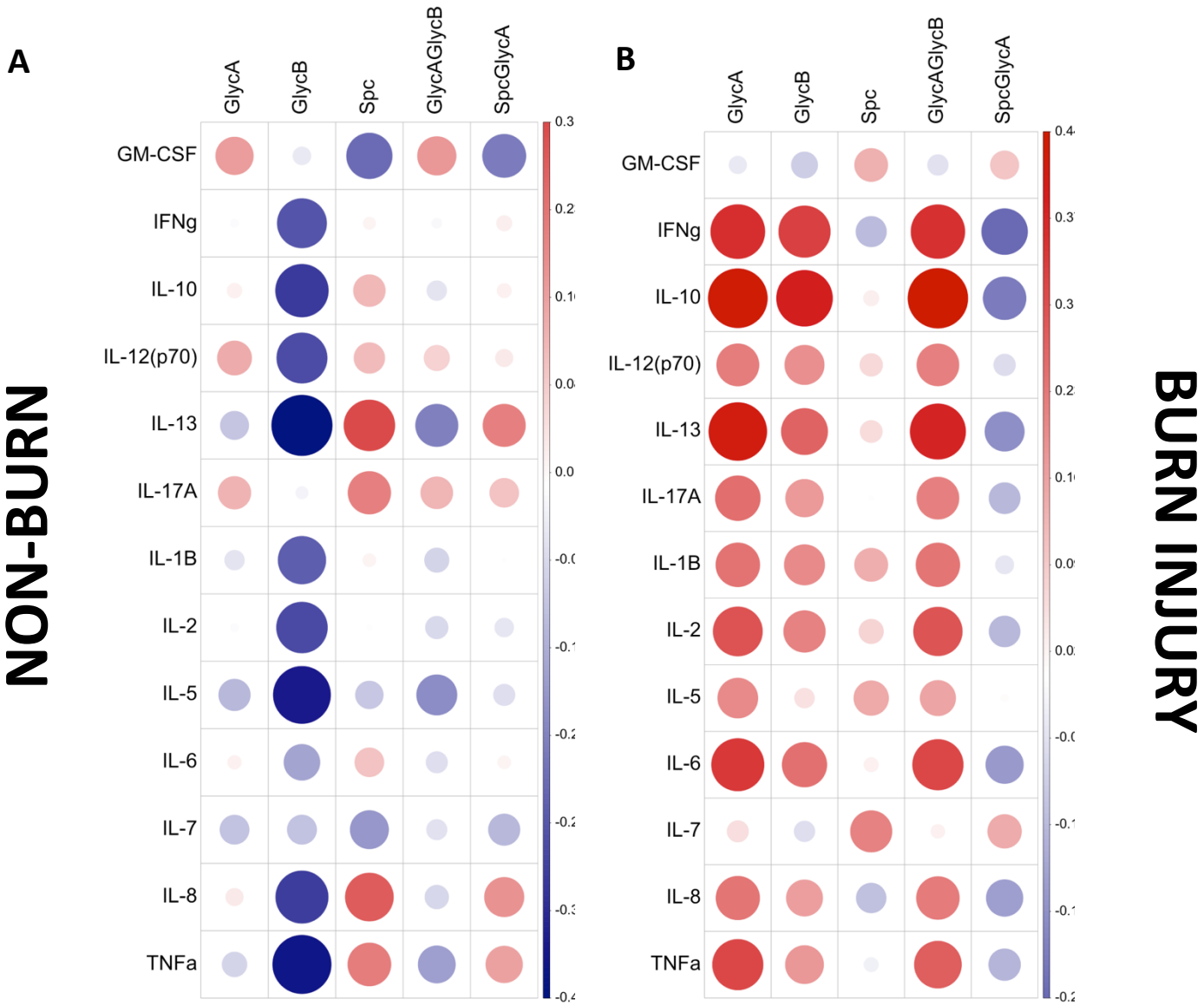



**Figure 5.23: Spearman's correlation between measured cytokines and DIRE quantified integrals.** Correlation analysis for A) non-burn controls and B) burn injury patients. Only significant correlations with a pvalue of <0.05 are shown, with non-significance indicated by a blank box.

### 5.3.9 Correlation mapping of lipoproteins and DIRE integrals

Two distinct profiles are observed when mapping lipoproteins and DIRE integrals between non-burn controls and burn injury patients (figure 5.24). Generally, total protein subfractions have stronger correlations in non-burn, with negative correlation observed between total protein fractions and GlycA and GlycB, with the same weak correlation pattern observed in burn injury patients.

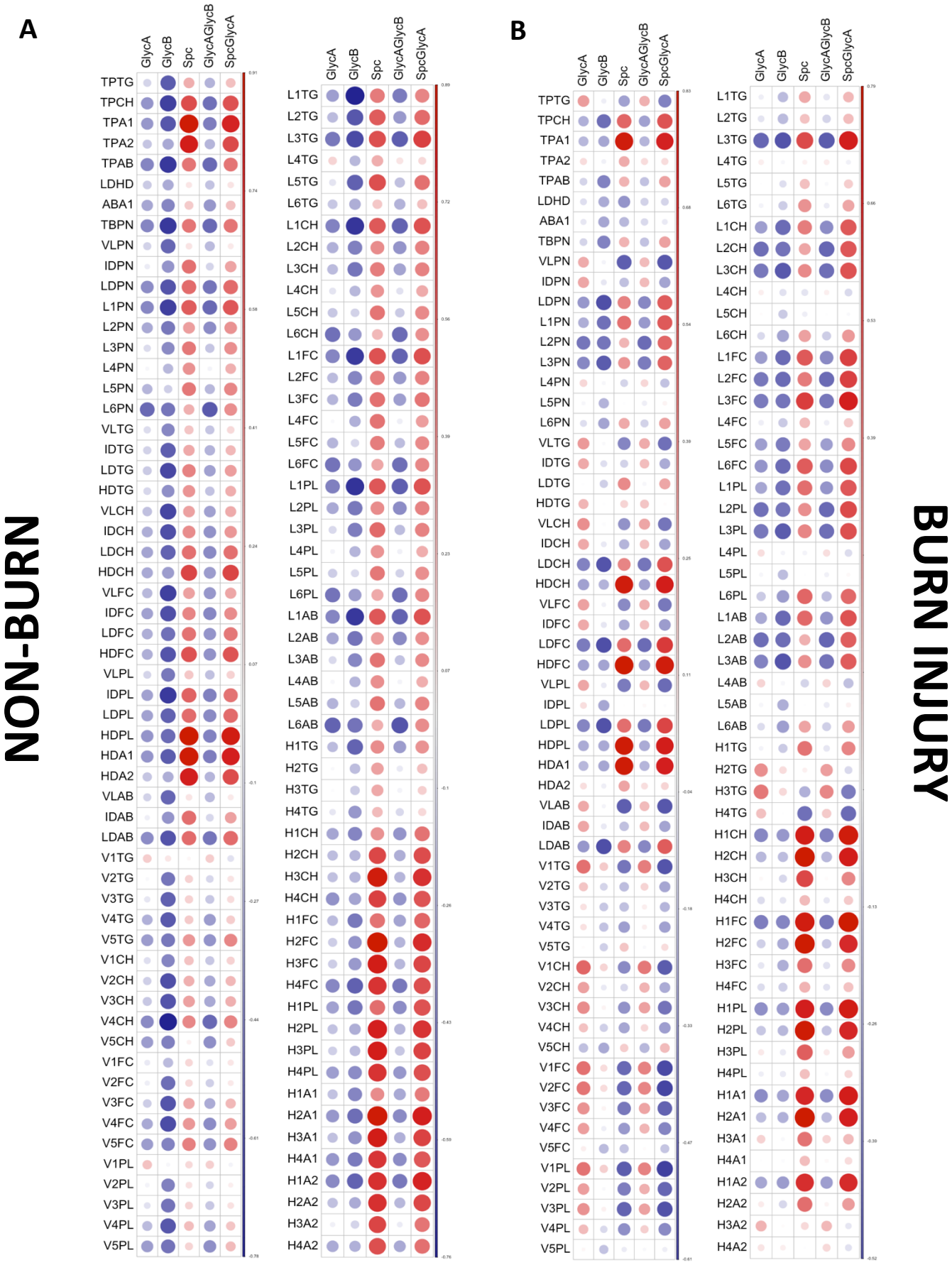



**Figure 5.24: Lipoprotein-DIRE integrals spearman's correlations.** Quantified lipoproteins and DIRE integrals correlated against one another, showing the signature present in A) non-burn controls and B) burn injury patients. Only significant correlations with a pvalue of <0.05 are shown, with non-significance indicated by a blank box.

Positive correlation is observed between total proteins and SpC in non-burn, however overall weaker correlations are observed in burn-injury patients. Similar observations are seen with LDL particle number subfractions, with little correlation observed in larger LDL particle for burn injury comparatively. Weaker correlations are observed with VLDL triglycerides in both non-burn and burn injury, bar V1TG, which exhibits strong correlations for burn injury. Positive correlation is observed in burn injury between GlycA and VLDL cholesterols, free cholesterols and phospholipids, with an inverse correlation observed for non-burn controls. Weak correlation is observed between LDL1, LDL2, LDL4, LDL5 and LDL6 in burn injury compared to non-burn controls. Interestingly, H2TG, H3TG, H3A1 and H3A2 are positively correlated to GlycA in burn injury, an observation absent in non-burn signatures.

Generally, GlycB correlates negatively across the majority of lipoprotein parameters in non-burn controls, with the exception of some VLDL1 species, which are generally weakly correlated across all integrals in non-burn. Stronger correlations are observed between lipoproteins and DIRE integrals in non-burn controls compared to burn injury patients.

### 5.3.10 Correlation mapping of lipoproteins and cytokine

Correlation analysis was used to interrogate the association between lipoprotein concentration and cytokine concentrations. Figure 5.25A and 5.25B show two distinct signatures; with very weak and generally negative correlations observed in burn injury compared to non-burn controls. Ratios LDHD and ABA1 (apolipoprotein A1 : apolipoprotein B100) present inversely, with negative correlations with the array of cytokines, but positive

NON-BURN

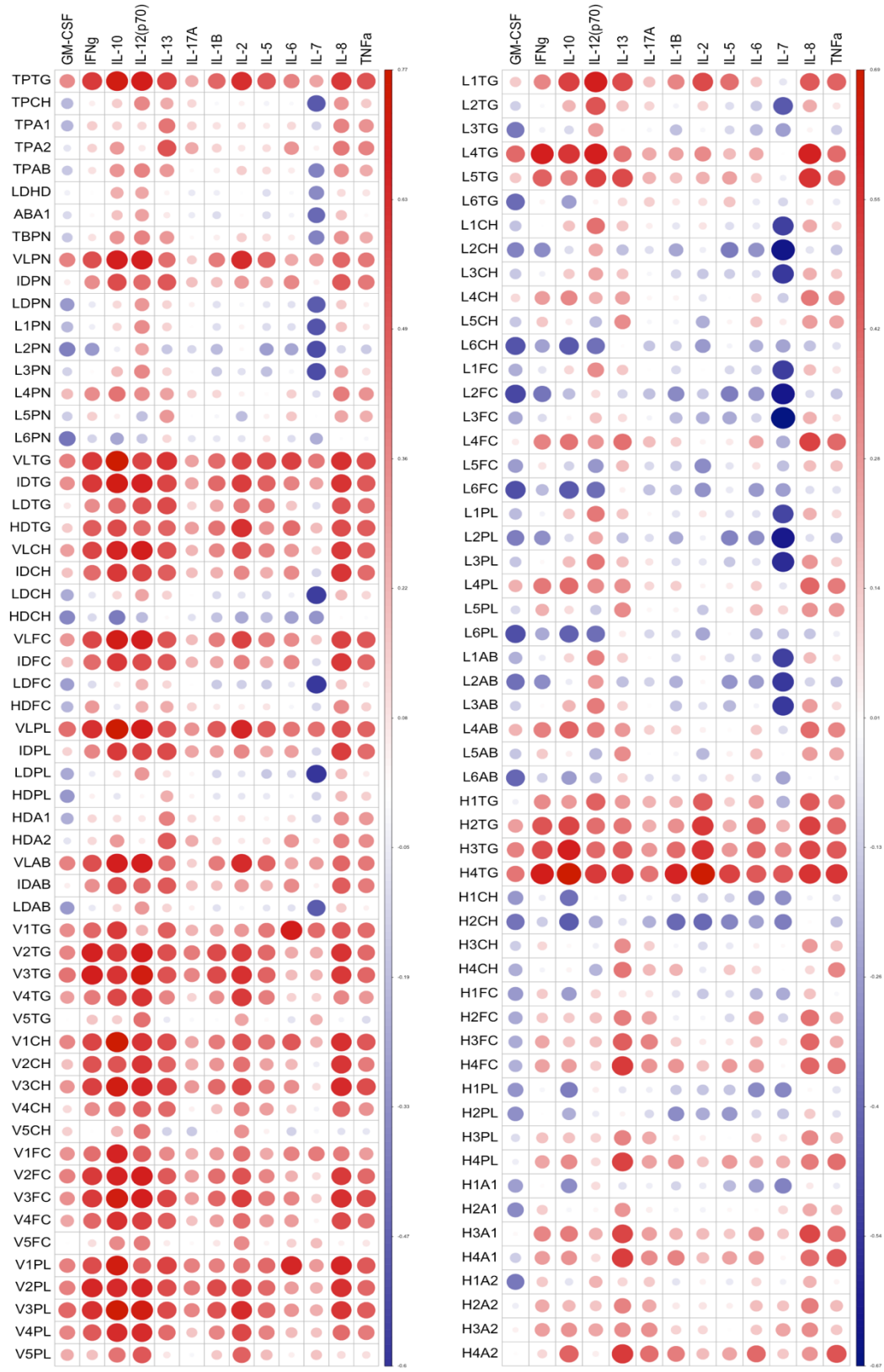



**Figure 5.25A: Lipoprotein - cytokine spearman's correlations.** Quantified lipoproteins and cytokines correlated against one another, showing the signature present in non-burn controls. Only significant correlations with a pvalue of <0.05 are shown, with non-significance indicated by a blank box.

BURN INJURY

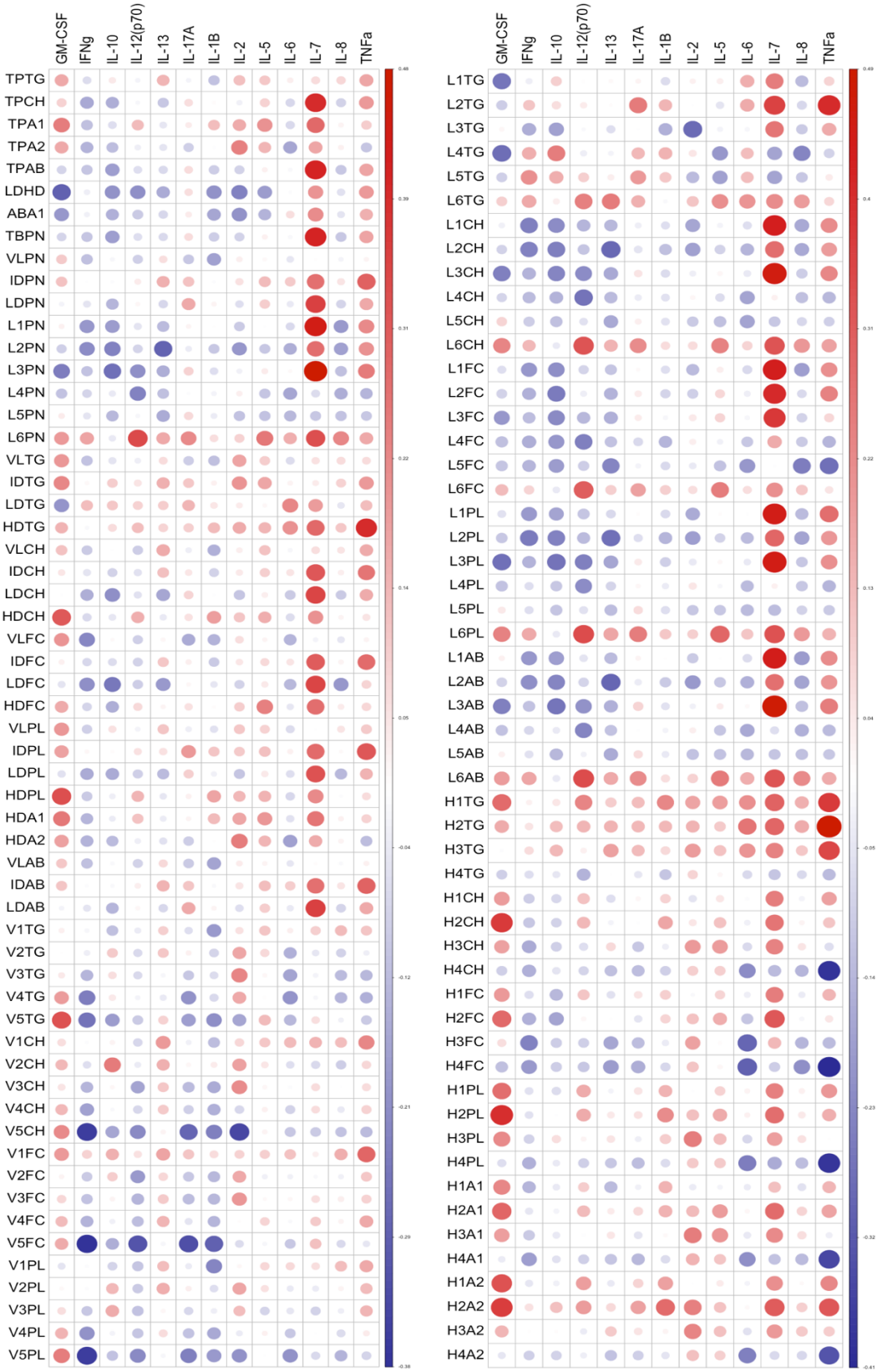



**Figure 5.25B: Lipoprotein - cytokine spearman's correlations.** Quantified lipoproteins and cytokines correlated against one another, showing the signature present in burn injury patients. Only significant correlations with a pvalue of <0.05 are shown, with non-significance indicated by a blank box.

correlation to IL-7 is generally observed in burn injury. Generally, where correlations are significant, negative correlations are observed with IL-7 in non-burn and positive correlation with IL-7 are observed with burn injury. Strong negative correlation of VLDL5 subfractions and IFNgamma, IL-17 and IL-1B in burn injury, with weaker positive correlations observed in non-burn injury. GM-CSF positively correlates with HDL species in burn injury with the exception being HDL4 subfractions which, are negatively correlated with all cytokines in burn injury, but is observed as positively correlated in non-burn controls.

### 5.3.11 Hierarchical correlation maps integrating lipoproteins, small molecules and cytokines

Hierarchical edge clustering (figure 5.26) was used to investigate the correlation signature between significant lipoproteins, significant small molecules (pvalue <0.05) and cytokines. In non-burn (figure 5.26A), alanine correlates to IL- 1B, 2, 5, 6 and 17A, and small molecule glutamic acid. In burn injury (figure 5.26D), elevated alanine maps to creatine, glutamic acid, L5CH, V4CH and cytokines TNFa, GM-CSF, IL-8 and IL-17A. In non-burn controls (figure 5.26B), L6CH maps to other lipoproteins, L6PN and L6FC which could be attributed to composition similarity. In burn injury (figure 5.26E), L6CH maps to L6PN, L6FC, L5CH, V4CH and to cytokines GM-CSF, IL-5,7,8 and 17A, suggesting a stronger link to cytokines in burn injury. In non-burn, creatine maps to glutamic acid and L5CH (figure 5.26C). In burn injury, creatine maps to alanine, IL-2,5,6,7,8,13,17A, TNFa and IFNgamma (figure 5.26F). Generally, more correlation of metabolites with cytokines is observed in burn injury, suggesting that cytokines have a significant contribution to the immuno-metabolic signature observed in burn injury.

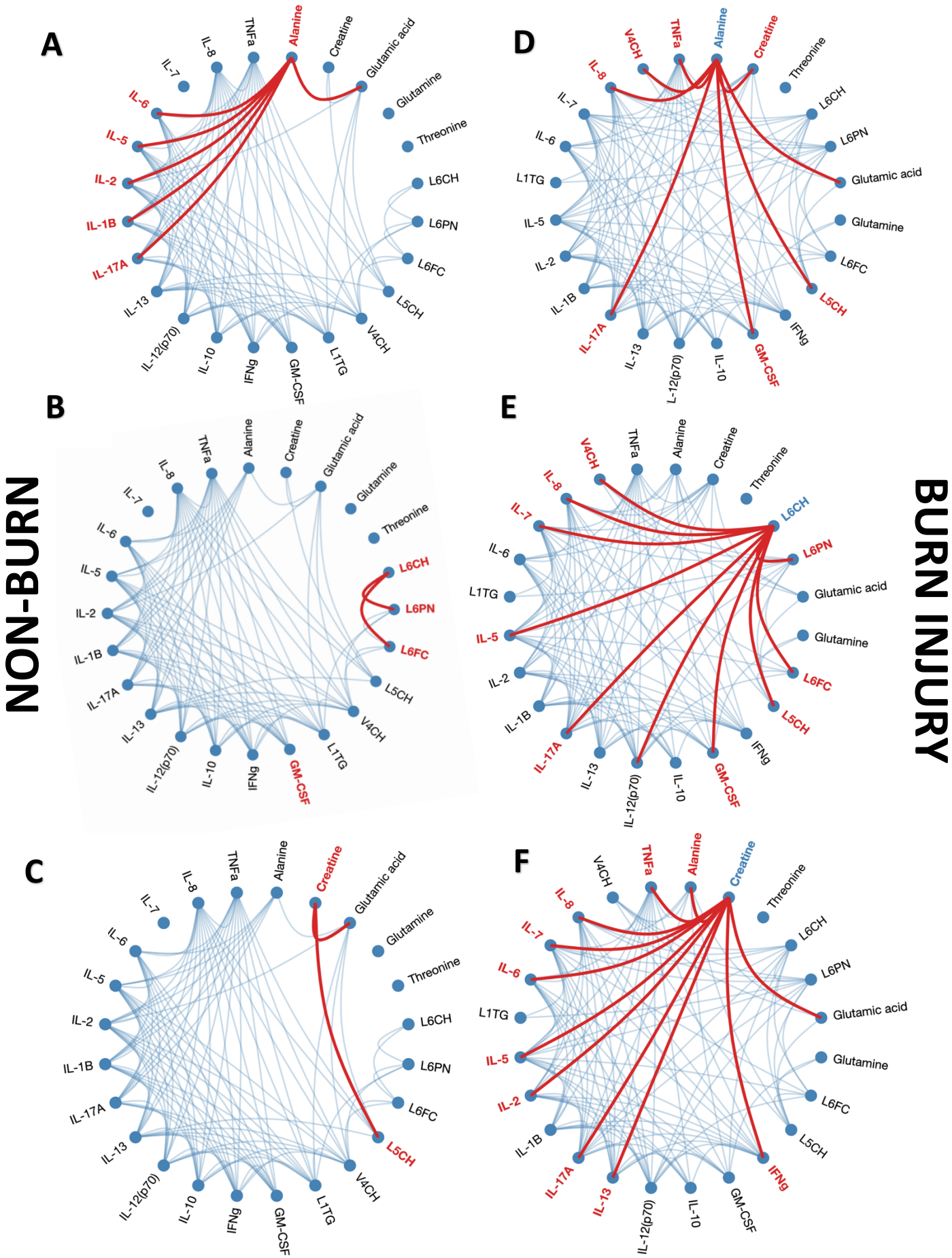



**Figure 5.26: Hierarchical edge correlation between significant lipoproteins, small molecules and cytokines.** Red lines states positive correlation between cytokine and metabolite, whereas blue line states negative correlation. Hierarchical maps for non-burn controls (A-C) and burn injury patients (D-F).

# 5.4 DISCUSSION

### 5.4.1 Targeted metabolic profiling of paediatric burn patients

The hypermetabolic response to severe burn injury is a well-established notion; mediated by up to 50-fold elevations in plasma catecholamines, cortisol and inflammatory cells up to 24 months post injury [278]. In this cohort of acute paediatric burns, arguably a scaled version of this response was observed, with elevated catecholamines and cytokines but decreased inflammatory cells 36 months post-burn injury in paediatric patients. Amongst the most significant differences in metabolites in the prior-burn injury children were Lysine, Alanine, Tyrosine, Asparagine, Proline, Leucine, Methionine, Serine, Taurine, Aspartic acid, and Ornithine. Interestingly, some studies have found that amino acids involved in the urea cycle, such as aspartic acid and ornithine, were associated with inflammatory markers and oxidative stress [279, 280]. High levels of serum serine have been found to be associated with reducing macrophage and neutrophil-mediated inflammatory responses [281], suggesting that an anti-inflammation response is still active 3 years post injury.

Higher levels of Phenylalanine, Methionine, Tyrosine, Lysine, Threonine and Neopterin found in the prior burn injury group were positively correlated across the majority of pro-inflammatory cytokines with exception of IL-10 and IL-2. It is arguable that this trend is seen due to the low levels of IL-10 and IL-2, both of which are produced CD4+ T cells, which is similarly diminished in quantity in burn injury patients.

### 5.4.2 Markers of Urea cycle dysfunction

Increases in catecholamines, glucagon and glucocorticoid production post-burn leads to increased glycogenolysis and protein dissolution in both the liver and skeletal muscle, resulting in increased triglyceride, urea and glucose production [282]. With significantly elevated ornithine and aspartic acid, crucial components of the urea cycle, it is clear that this hypermetabolic response is still ongoing in paediatric patients 3 years post injury.

Phenylalanine is an essential amino acid for humans and is a substrate for phenylalanine-(4)-hydroxylase (PAH), which produces tyrosine, a precursor for the biosynthesis of DOPA and catecholamines dopamine, epinephrine and norepinephrine [283]. Previously, it has been found that serum phenylalanine is positively correlated with poorer outcomes in patients post-trauma, compared to healthy controls [284]. In burn patients, phenylalanine has been found to correlate with clinical course and to be a predictor of increased morbidity and mortality [285].

Elevated levels of plasma serine, phenylalanine and tyrosine have been repeatedly reported in patients with liver dysfunction [286]. Interestingly, it has also previously been found that in acquired liver diseases, either one or both of tyrosine and phenylalanine were found to be elevated [287]. Increased phenylalanine concentrations in serum can be attributed to a reduced conversion rate of phenylalanine by phenylalanine-4-hydroxylase (PAH), which may be the result of increased oxidative stress due to immune activation and inflammation [285].

A disruption in the ornithine cycle branch of the urea cycle is suggested from the results of this study, with elevated Ornithine and Arginine detected in burn injury patients in comparison to healthy controls. The urea cycle is a crucial route through which toxic compounds such as ammonia and surplus nitrogenous compounds are cleared; in addition to the synthesis of Arginine via *de novo synthesis*. Increased Ornithine and Arginine suggests an increase in urea cycling, in addition to the detected increased Alanine, which is used as a nitrogen source in the liver and during muscle catabolism [288]. Interestingly, elevated serum Alanine has been found to promote skeletal muscle atrophy and hyperglycaemia in type 2 diabetes [289]. Furthermore, depleted Glutamate concentrations detected in burn injury patients may be indicative of increased liver function and muscle catabolism, with ammonia converted to Glutamate through transamination in muscle. Decreased Glutamate levels have been detected in patients suffering acute liver injury [290].

### 5.4.2 Neuroactive amino acids

With the suggested urea cycle dysfunction, from prior results, it is imperative to consider the neurological aspect of this disease, due to sensitivity of the blood-brain barrier to urea. Glutamate, Glycine, Serine and glutamine are all neuroactive amino acids, acting as small molecule neurotransmitters and neuromodulators [291].

Serine is one the most elevated amino acids in burn injury patients, in comparison to healthy controls (p value 0.009). Serine is not only crucial for cellular regeneration but also as a neuroactive amino acid, with some studies suggesting that it may also be involved in the extracellular release of Glutamate through a signalling intermediate, Ethanolamine

[292]. Whilst Serine is elevated, both Ethanolamine and Glutamate were found to be reduced in burn injury patients, a trend also detected in patients suffering from depression and mental health disorders [292]. In addition, the decreased Phenylalanine/Tyrosine ratio and Kynurenine/Tryptophan ratios detected in this cohort, are typically consistent with immune cell deactivation [293]. It has been suggested that aberrant expression of phenylalanine maybe involved in the development of neuropsychiatric symptoms, manifesting as depression and anxiety [294]. Both tryptophan and 3-hydroxykynurenine were found to positively correlate with TNF-a in burn injury. TNF-a has been suggested to contribute to the pathogenesis of depression through the activation of a hypothalmo-pituitary-adrenoncortical (HPA) axis, activating neuronal serotonin transporters and the stimulation of the indoleamine 2,3-dioxygenase which subsequently leads to tryptophan depletion [295].

Murine studies have suggested that the accumulation of some amino acids in the brain, such as Tryptophan, Tyrosine, Phenylalanine, Methionine and Histidine, may cause abnormalities in serotonin neurotransmission [296]. Furthermore, a decreased neuroendocrine functionality has been suggested through the reduced Serotonin/5-Hydroxyindoleacetic acid ratio in burn injury patients. Reduced expression of serotonin is thought to be associated with several psychological conditions, including depression [291].

### 5.4.3 Inflammatory cytokines and metabolite associations

Cytokine levels peak immediately post burn, and whilst some studies have suggested that these levels gradually reduce to normal 3-6 months post burn [236]. However, the results

from this study suggest that 36 months post injury, cytokine levels still have not returned to normal (figure 5.9). The chronic inflammatory response seen was associated with a neurotoxic signature, comprising of several metabolites including Quinolinic acid.

A number of metabolites involved in the urea cycle are positively correlated with pro-inflammatory IFN- γ. Recent studies have suggested that increased urea cycle activity may be a response to counter inflammation, suggesting that Arginine is used to prevent inflammation induced NO production, thereby using glucose metabolism as an anti-inflammatory mechanism through the regulation of the urea cycle [297].

### 5.4.4 Immune cell contributions to hypermetabolism

CD38 has historically be considered to be a surface marker used to identify T cells but is also a multifunctional enzyme that utilises nicotinamide adenine dinucleotide (NAD) as a substrate to generate further messengers. The functional role of NAD is predominantly as a regulator of signalling pathways, and for the generation of secondary messengers such as cyclic-ADP-ribose and to regulate NAD dependent deacetyl sirtuins [298]. Both of these pathways have been found to be critical and important to several cellular events and the mechanisms of age and morbidity. CD38 differentiated immune cells have been found to be positively correlated across several inflammatory cytokines such as IL-2, IL-6 and IL-10. It is noteworthy that CD8 is typically seen in acute phase of recovery of some diseases like SARS-CoV-2 [299] but in this cohort it is observed to be present long term with associations with IL-17A and IFNgamma. Studies investigating autoimmune skin diseases such as psoriasis

have found correlation between CD8 T cells and IL-17A [300]. This finding suggests that immune recovery for acute burn continues long after healing.

Generally, negative correlation of indole-3-acetic acid and glutamine against cytokine measurements is observed in burn injury patients, is not entirely unexpected considered that indole-3-acetic acid undergoes metabolism through the glutamine conjugation pathway [301]. This again supports the notion of a disrupted liver metabolism in burn injury patients in response to chronic inflammation.

It is arguable that dysfunctional dendritic cells (DCs) may be responsible for the chronic post-burn inflammatory signal; with many differentiated DCs responsible for the uptake of pro-inflammatory cytokines such as IL-1B and IL-6, as found in this study. Without the abundance of such DC species, there is lack of co-ordination of innate and adaptive immune responses, through the disharmonisation of the detection of the antigen specific CD4+ T cell response [302], which depends on the connection of detection and the uptake of foreign antigens. IL-10 is a key regulatory cytokine limiting and preventing excessive T-cell responses to microbial pathogens to prevent chronic inflammation and tissue damage [303].

### 5.4.5 Metabolic signatures of prior burn injury, hypercatabolism and inflammation

Beyond hypermetabolism, burn injury additionally results in a hypercatabolic condition affecting muscle mass in patients [282]. In addition to higher concentrations of ketone bodies such as 3-D-hydroxybutyrate and acetoacetate being detected in burn injury, higher concentrations of creatine were found. High levels of creatine, and subsequently creatinine

is detected in burn injury patients. Creatinine is produced as a by-product of muscle metabolism, in which creatine is converted in the muscle to creatinine. Elevated serum creatinine has been previously used a predictive marker of burn injury associated with acute kidney injury, with elevated creatinine found to be as high as 50% in some cases [304]. Wallner et al suggest that a biphasic response to burn trauma exists with vast complication in post-burn muscle metabolism, initiated by a catabolic state, resulting in long-term hypercatabolism [305]. Furthermore, it has been postulated that consequences of non-specific adaptive immune response include resorptive bone loss and muscle catabolism. In a randomised study of paediatric burns, those administered with a single dose of bisphosphonate pamidronate within the first ten days following injury, were found to have reduced bone resorption and muscle wastage [306].

A similar lipoprotein signature of elevated HDLs and LDLs was found to a systemic response to chronic cold stress [307]. Similarly, elevated levels of lactate, acetoacetate, creatine, creatinine and choline metabolites but reduced levels of VLDLs, glutamine and glycine were seen as part of a systemic response, consistent to metabolic signature detected in this cohort of acute burn injury. These findings support the notion of a chronic hypermetabolic state following acute injury, driven by an immune-metabolic memory.

The liver plays a crucial role in the regulation of normal ammonia metabolism. Reduced alpha-ketoglutarate is detected in the NMR spectra of burn injury patients, antiphased with glutamate, suggesting a urea cycle dysfunction. Reduced levels of glutamine, as detected by both MS and NMR, suggest that there is an inhibition of detoxification of ammonia. Some studies have suggested that ornithine alpha-ketoglutarate would be an effective treatment

in restoring glutamine levels in muscle, which are significantly depleted (pvalue <0.01) in hypercatabolic patients [308]. It is estimated that one third of total portal blood ammonia is detoxified by glutamine synthesis. Glutamine is essential for immune function as the preferred substrate for energy production in enterocytes and lymphocytes [309].

Urea cycle dysfunction is further suggested through the elevated level of pyruvate comparatively in burn injury patients compared to controls. Pyruvate oxidation is thought to be inhibited by hyperammonemia at the pyruvate and alpha keto-glutarate dehydrogenase steps. Catabolism of branched chain amino acid isoleucine results in the formation of both acetyl-CoA and succinyl CoA for the synthesis of glutamine as TCA cycle intermediates, and thus has the ability to attenuate this dysregulation [310]. However, depleted isoleucine is also noted in burn injury patients suggesting a far deeper systemic aberration.

Elevated alanine in burn injury patients was significant in both MS and NMR (pvalue <0.05). Johansen et al., suggest that alanine may be a marker of neuronal glycolysis as a consequence of hyperammonia. Cell culture studies have found that the vast majority of lactate and alanine was derived from glucose [311], consistent with these results with the detection of elevated glucose in burn injury. It is suggested the pyruvate diverts from the TCA cycle and towards lactate and alanine synthesis with Valbuena et al finding that 60% of pyruvate, lactate and alanine carbons were glucose derived under enhanced glucose metabolism conditions [312].

A significant change in branched chain amino acids in burn injury patients, as highlighted by correlation analysis in figure 5.22B, could be attributed to the diminished level of activity in

burn injury patients following trauma. A diminished level of physical activity in burn patients has been reported by clinicians, consistent with exercise habits and general catabolism seen in obesity [313]. Depletion of branched amino acids may be suggestive of a lack of oxidative stress, with the lack of exercise hindering the production of enzymes [314].

GlycA is suggested to be a robust marker of inflammation, superior to pathology measure such as C-reactive protein (CRP) [132]. Glycosylation patterns of serum proteins are typically expected during the acute phase of injury. Connelly et al., found in children suffering acute Kawasaki disease, elevated GlycA, increased total and small LDL particle numbers and decreased HDL particle numbers (pvalue <0.001) [315]. Similar signatures were observed in this cohort of children, with elevated GlycA and GlycB, depleted SpC, and elevated LDL particle numbers. Studies in infants undergoing acute stress, have found elevated levels of GlycA, positively correlating with inflammation (316], supporting the notion that there is an underlying mechanism of acute stress in this cohort of burn injury children.

## 5.5 Conclusions

An induced state of hypermetabolism post-burn injury has been a well-accepted notion for the past 6 decades. Severe burn injury is typically followed by a substantial hypermetabolic response that persists up to 24 months after injury, mediated by up to 50-fold elevation in plasma catecholamines, cortisol and inflammatory cells [282]. What has not been established, is the chronic profound effect of acute burns. Whilst all paediatric burn injury patients in this study have seemingly healed, the results from this window of analysis all lead to conclusion of a persistent hypermetabolic state 3 years post-burn. Signatures for liver dysfunction and neurotoxicity are amongst the most significant findings, with substantial metabolic aberrations involved with the urea cycle and the production of inflammatory cytokines. A diminished immune cell response in burn injury patients, as also found in previous studies, may be the result of an overwhelmed chronic inflammatory response, leading to an uncoordinated immune response. The results from this study also suggest that there is an inherent need for improved post-injury care, with the systemic effects of burn injury prevalent far beyond the original site of injury. Due to the small sample size of this pilot study, this study requires further follow-up in order to confirm findings in a larger cohort. In addition, further metabonomic assays, such as NMR, would be useful for providing an untargeted metabolic perspective.

# CHAPTER SIX: DISCUSSION

## 6.1 General summary

Each study in this thesis presents a characterised, complex systems disease or condition with markedly distinct signatures, with significant metabolic aberrations from a normal state of metabolism. The three studies were linked by their impact on immune-metabolic interactions and their effects on multiple organs and systems. In order to address these complex conditions, I developed a number of strategies centred around the analysis of multimodal data and integration of that data to generate new insights into disease mechanism or patient stratification. Each data modality was found to present different windows of information, with cytokine profiling and microbiome sequencing enhancing the clinical interpretation of spectroscopic profiling and targeted MS data.

Similarly, different biofluids (plasma, urine, faeces) contributed specific information. Although each study had a main clinical effect, hypothesis or target organ, for example the primary site of injury in PBC is the liver, whereas SARS-CoV-2 infection is generally the lung, all conditions had extensive effects beyond the primary insult or tissue of impact. This was seen most clearly in the paediatric burns study where there is a persistent metabolic 'memory' of burn injury three years post injury in apparently healthy and non-scarred children. The following sections summarize the main findings in the thesis and define the limitations and future avenues of research arising from the studies undertaken here.

### 6.2 Biomarkers of organ dysfunction associated with different diseases

SARS-CoV-2 infection presented a metabolic signature beyond an inflammatory respiratory disease, with inflammatory markers GlycA, GlycB and the phospholipid compound, SpC, found be substantial markers of inflammation [127] all detected by use of a diffusion edited NMR experiment. Standard 1D NMR spectroscopy using the IVDr system for quantification generated information on 112 lipoprotein parameters which were also found to have a specific association with SARS-CoV-2 infection. Mass spectrometry delivered addition, nine metabolites were elevated in plasma and strongly associated with infection (quinolinic acid, glutamic acid, nicotinic acid, aspartic acid, neopterin, kynurenine, phenylalanine, 3-hydroxykynurenine, and taurine; $p < 0.05$), while four plasma metabolites were lower in concentration in infection (tryptophan, histidine, indole-3-acetic acid, and citrulline; $p < 0.05$). These findings support the notion of a state of "phenoconversion". Phenoconversion is a phenomena whereby exposure to a specific disease or drug alters the phenotype of an organism which may be a permanent or reversible change [107]. In this instance, phenoconversion is implicated as a change from a normal state to an aberrant, perturbed pathophysiology, typically associated with a range of metabolic characteristic biomarkers [317]. These findings implicate multiple types of SARS-CoV-2 phenoconversion, with neurological disruption indicated by elevated 3-hydroxykynurenine, quinolinic acid and kynurenine and liver dysfunction indicated by a depleted Fischer's ratio and elevated taurine [318]. A total of eight papers (Appendix 1) were generated from the COVID studies in which my contribution was the correlation of the multimodal spectroscopic data and the extension to integration with clinical and cytokine data. Strong correlation between cytokines and NMR and MS-based metabolic markers was found. Specifically, the

correlation of between LDL and VLDL lipoproteins and inflammatory cytokines implied a signature strongly consistent with that of cardiovascular disease and diabetes, both of which are diseases associated with an inflammatory state [145]. Figure 6.1 summarises the plasma metabolic biomarkers with regard to the systemic effects of SARS-CoV-2 infection.

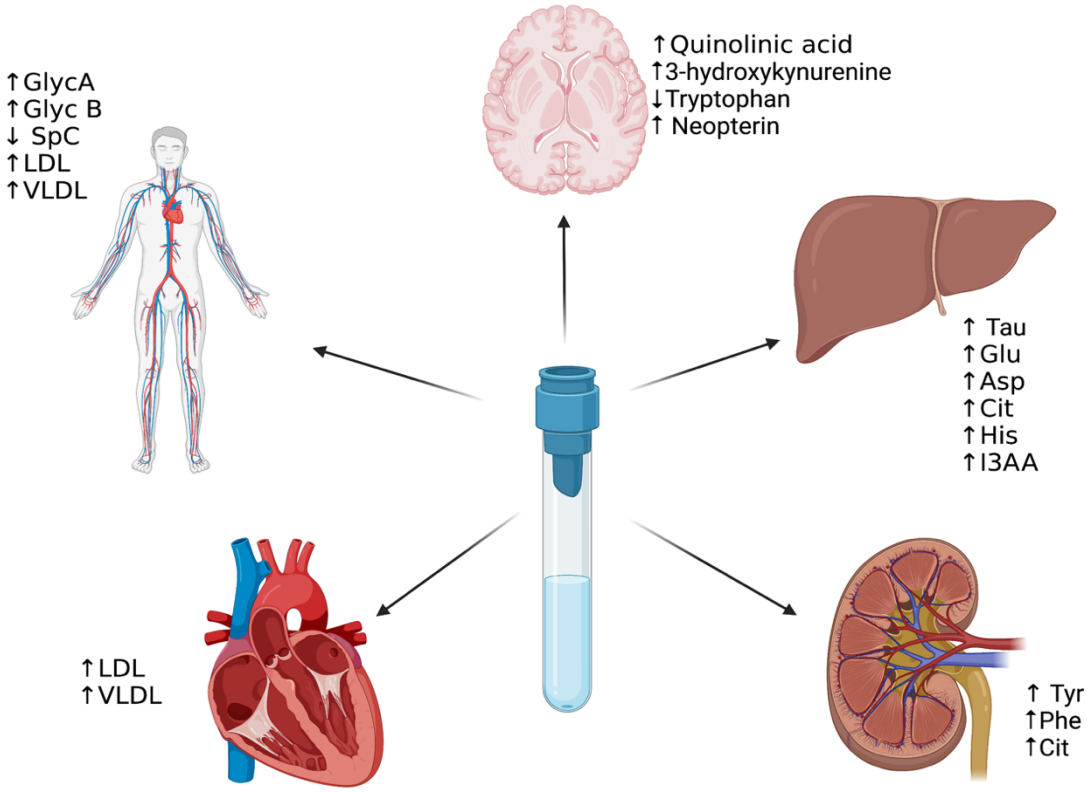



**Figure 6.1: Exemplar systemic-metabolite associations from SARS-CoV-2 infection.**

Urinary markers of PBC were found to be associated with gut microbiome aberrations, with perturbed hippurate found in untreated PBC patients (p value <0.05). A strong correlation was found between hippurate and several gut microbiome species implicated in immunity. Further microbiome associated markers were identified through multivariate modelling and experimentally through the sequential addition of authentic chemical standards. For example, phenylacetylglutamine was found to be different between UDCA responders and

non-responders in PBC disease although this trend was not significant (the lack of robustness reflecting limitations and flaws in experimental design that were beyond the scope of this thesis). The results from this analysis suggest that there is a gut microbiome perturbation in PBC, which can be ameliorated by UDCA treatment but is confounded by the fact that not all patients respond to treatment (figure 6.2).

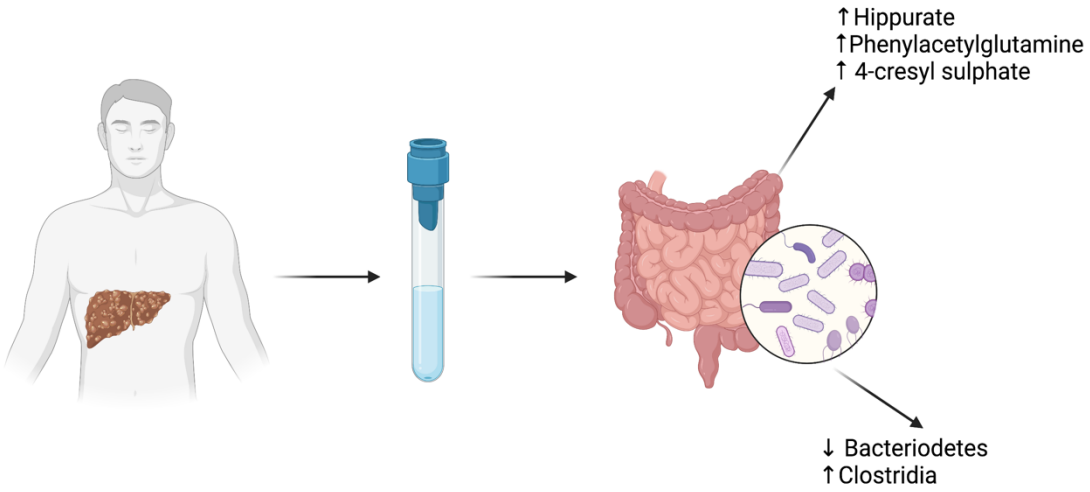



**Figure 6.2: Exemplar metabolite-microbiome biomarkers from PBC UDCA responders.**

Prior acute burn injury in children presented unique findings, with vast systemic changes driven by underlying inflammation still prevalent 3 years after the original insult (figure 6.3). Twelve metabolites were significantly elevated in plasma and strongly associated with impaired systemic and liver function (ornithine, lysine, alanine, serine, proline, cystine, asparagine, phenylalanine, arginine, tyrosine, 3-methylhistidine and quinolinic acid, p value <0.05). Significant elevation of quinolinic acid in children with previous burn injury implicated neurological disruption and would require further investigation given the association prior-burn injury and prevalence of mental health disorders. However, it is worthwhile noting that whilst only quinolinic acid was significant, many other differences

were observed between burn injury and non-burn control counterparts in the tryptophan pathway. Integration of MS metabolites with cytokines and cell data, revealed a distinct signature for each group, with strong correlation between cytokines and key liver function metabolites. From the NMR analysis GlycB, associated with inflammation, was found to be a key discriminator between burn and non-burn participants. LDL lipoprotein species were found to be significantly elevated in burn injury compared to controls and correlated strongly with IL-7. Again, the signature between LDL and cytokines implicates a signature strongly consistent with diabetes, in addition to the elevated levels of glucose detected in burn injury children via NMR profiling.

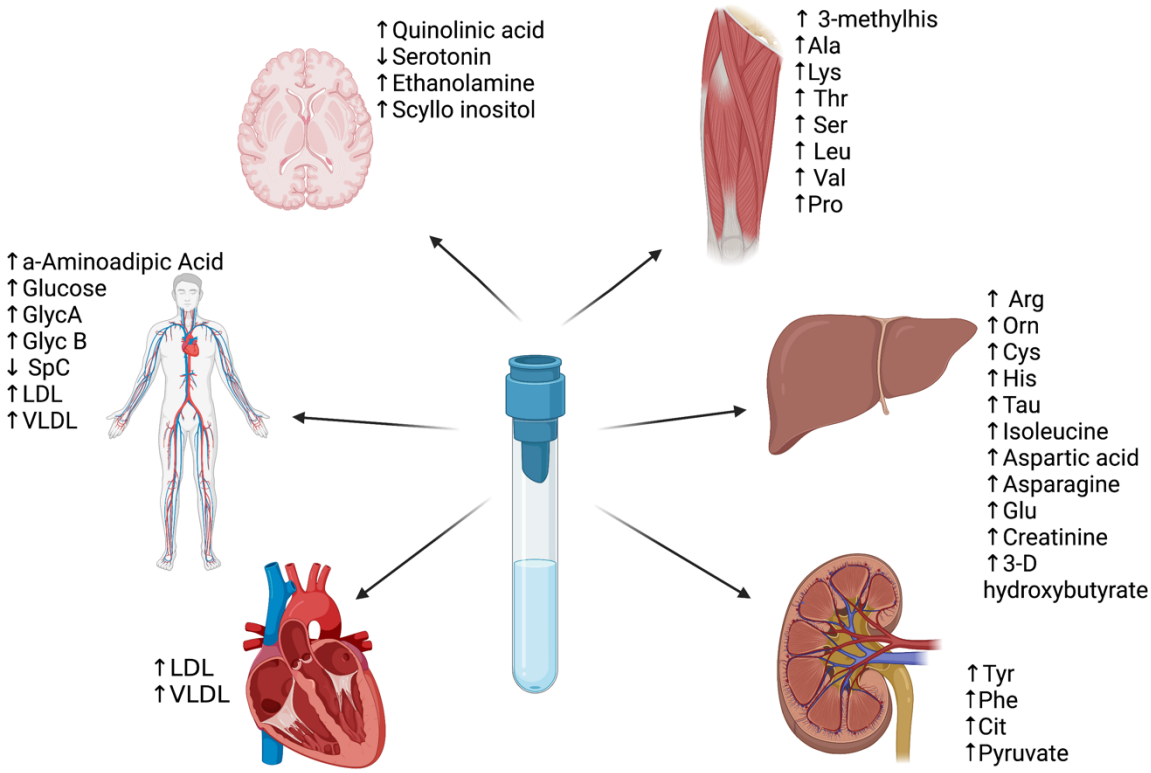



**Figure 6.3: Exemplar systemic-metabolite associations from acute paediatric burn injury.**

Overall, distinct metabolic signatures were detected from each disease, with vast systemic implications seen in SARS-CoV-2 infection and paediatric burn injury (figure 6.1 and figure

6.2). Further profiling of eicosanoids and bile acids may help to further understand the clinical implications of the inflammatory response in these diseases. The results from PBC present a simpler metabolite-microbiome association due to the reduced resolution of the data. Further integration with proteomic studies would be useful to assess the proteins profiles expressed by these different gut microbiome communities. Each study clearly supports the notion for precision medicine; with subsets of patients within each disease group presenting differential metabolic profiles, suggesting that there is a clear pre-requisite to further deconvolute the specific disease state and physiological consequences.

### 6.3 Statistical integration methods

Correlation analysis was used widely throughout the thesis and visualised using a variety of methods such as heatmaps and hierarchical edge clustering. This mode of analysis was useful in interpreting broadly related systemic data and delivered distinct signatures with more depth for biological interpretation. For instance, significant correlation between IL-7 and LDLs were observed in paediatric burn injury, and this subsequently led to highlighting the importance of elevated glucose in paediatric burns, with elevated LDLs and inflammation mirroring the metabolic signature found in diabetes [137]. Onset and an increase of incidence of diabetes in burn patients has been highlighted as a significant consequence of injury and is thought to be attributed to a state of hypermetabolism [289]. Correlation mapping is also a useful tool in comparing signatures between disease and controls, enabling direct comparison of motifs. This in itself is a powerful tool, with easy visual comparison of correlates. In some instances, the pattern of correlation between different datasets for a given disease may be significant, even though the individual datasets

do not show a significant correlation with the disease. This was observed for the PBC dataset where the microbiome-metabolite correlation patterns, although not reaching significance in their own right, displayed substantial differences between the disease and control groups and generated correlation patterns that were biologically coherent. Correlation mapping further extrapolates from global metabolic profiling and metabolites, enabling direct insight into other systemic operators such as cytokines. This is crucial for the translation of metabolic profiling in medical research as metabolites do not work individually, nor do they operate in an isolated system.

Hierarchical edge clustering extrapolates from the traditional heatmap visualisation and restricts findings to those calculated as more closely related and significant. This enables a quicker convolution of results to the reader and presents the most important variables, in comparison to visual heatmap, where it can be difficult to visualise the connections between variables. It further extends the linear relationship between correlation variables to other variable associations, which can be very powerful for biological interpretation. For example, in SARS-CoV-2 infection, the correlation between cytokines was highlighted, in addition to the cytokine-metabolite correlation.

Metabolite set enrichment analysis (MSEA) is a useful tool to investigate potential associations between metabolites in a given set. Despite the same set of metabolites screened for in the enrichment analysis, different pathways were implicated as enriched for SARS-CoV-2 infection and paediatric burn injury. Selenoamino-acid metabolism, catecholamine biosynthesis and thyroid hormone synthesis were calculated as the most enriched pathways in paediatric burn injury, and methylhistidine metabolism, phenylalanine

and tyrosine metabolism and histidine metabolism were selected as the most enriched pathways in SARS-CoV-2 infection. Each pathway is implicated by changes in different metabolites, and thus presents very different metabolite-pathway mapping.

The advantage of using methods such as MSEA, is that it enables physiological interpretation of a set of metabolites and is able to discriminate between disease and control, mapping to specific pathways. However, it is difficult to validate the findings of some pathways, such as selenoamino-acid metabolism, given that targeted analysis is restricted to the window of metabolites screened for, and thus validation of these results was beyond the scope of this thesis. Furthermore, the method is restricted to annotated metabolites in a metabolite set, which is then screened against a database. Both of these methods are restricted by the set of annotated metabolites they contain, and thus do not easily account for untargeted metabolites.

### 6.4 Limitations of the work carried out in this thesis

As with most studies where there is competition for clinical samples and compromise on study design, each of the studies had limitations. Given the small pilot study size of the WA SARS-CoV-2 study and paediatric burn study, it leaves a requirement for an extrapolated and longitudinal studies, to investigate the development of disease and if full “pheno-reversion” of metabolic disease states takes place.

Multivariate analyses using metabolic profiling techniques often do not work well with a restricted sample, as they often highlight inter-sample and individual variation. However

even with larger studies, such as PBC, it can be seen that study results can be heavily confounded by factors such as sample collection, patient medication and alcohol consumption, which in turn can affect the composition of the biofluid analysed, as seen with the chemical shift variation exhibited in urine via NMR analysis. A prevalent theme throughout this work was that all studies are more variable dense, than sample dense, due to the nature of 'omic' data structure, which in turn limits biological interpretation. As seen predominantly with the PBC study, interpretation of results was vastly limited, and the use of pattern recognition was diminished due to the high sample variability. There was also a lack of independent cohorts available for validation, however during the time of this thesis, work is ongoing to build biobanks for reference in each condition.

A dominant issue with clinical studies is limited sample volume. Given that the burns study was carried out on children, this vastly limited the plasma volume that can be collected and thus limits the number of analytical assays. Similar issues were seen with the PBC cohort, whereby analysis was limited by sample volume, reducing the level of resolution seen by NMR, and thus exacerbating the chemical shift variation observed even further. To overcome this, strategies are currently being developed for micro-sampling. This will enable maximisation of information recovery from as little sample as possible and will enhance robust biological interpretation.

### 6.5 Future work arising from this thesis

Future work for the WA SARS-CoV-2 study includes the collection of longitudinal studies. A set of 3- month recovery [319], suggests that some former SARS-CoV-2 achieve

phenoreversion, whereas others do not. Similarly, reversion of some parameters to control levels is apparent for most patients whilst other parameters take longer to recover or may be permanently altered. The collection of longitudinal samples will be more revealing as to the prevalence and ascendancy of sentinel biomarkers. SARS-CoV-2 studies will also be extended epidemiologically, looking at different population groups internationally, which will also enable an insight into the effect of different variants on metabolism. The application of further metabolic profiling techniques such as untargeted MS and MS-based lipidomic assays will also be implemented in the future.

The UK-PBC nested cohort study will be extended with other metabolic profiling methods (paper under review: Appendix 2) and will be integrated with other omic datasets such as proteomics, that have investigated the catabolism of UDCA, with the ultimate aim of targeting these results towards precision medicine therapies. Despite the challenges encountered with the metabonomic data, further work will involve integration with genomic data, which will be useful to investigate a causal link between microbiome changes and further systemic effects.

The Paediatric Burns study is ongoing, will also be complemented by further profiling assays such as eicosanoid profiling and lipid profiling, in addition to currently building a biobank of both children and adults. This will also enable plasma samples to be captured at the time of injury, thus facilitating a more comprehensive metabolic profile. Further work involves the integration of all MS profiling assays, and NMR profiling. Additional omic datasets will hopefully be acquired following funding acquisition, which will be useful to gain further understanding of what changes may be occurring at a genetic level.

### 6.6. Final Conclusions

The work presented in this thesis achieves the development of a new framework for the analysis and integration of multifaceted, multimodal data in order to define complex disease systems. This work further highlights the prerequisite for development of new strategies of integrating data to provide enhanced capacity for biological interpretation, showing that there is merit and purpose in data integration. The work in this thesis was able to deconvolve 3 separate diseases using metabolic profiling techniques, using a similar pipeline for data integration. The signatures that are characteristic of each disease can be used for further downstream inferences; for instance, with SARS-CoV-2 infection, the lipoprotein signature was similar to that seen with cardiovascular disease and diabetes. Each data modality has proven to be useful in the clinical interpretation of the metabonomic data and has strengthened the applicability of the findings to a clinical environment.

Whilst it is currently unlikely that spectroscopic profiling and comprehensive cytokine analysis will be implemented in routine clinical screening due to capital cost of instrumentation and economic value respectively, the integration of these complex datasets will promote the identification of key pathways or effects, which can lead to the development of new laboratory diagnostics, and potentially therapeutic targets.

# CHAPTER SEVEN: REFERENCES


1. Nicholson JK, Lindon JC. Metabonomics. Nature. 2008 Oct;455(7216):1054-6.
2. DeBerardinis RJ, Thompson CB. Cellular metabolism and disease: what do metabolic outliers teach us?. Cell. 2012 Mar 16;148(6):1132-44.
3. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research. 2000 Jan 1;28(1):27-30.
4. Trivedi DK, Hollywood KA, Goodacre R. Metabolomics for the masses: The future of metabolomics in a personalized world. New horizons in translational medicine. 2017 Mar 1;3(6):294-305.
5. Dieterle F, Riefke B, Schlotterbeck G, Ross A, Senn H, Amberg A. NMR and MS methods for metabonomics. InDrug Safety Evaluation 2011 (pp. 385-415). Humana Press.
6. Koek MM, Jellema RH, van der Greef J, Tas AC, Hankemeier T. Quantitative metabolomics based on gas chromatography mass spectrometry: status and perspectives. Metabolomics. 2011 Sep;7(3):307-28.
7. Gika HG, Theodoridis GA, Plumb RS, Wilson ID. Current practice of liquid chromatography–mass spectrometry in metabolomics and metabonomics. Journal of pharmaceutical and biomedical analysis. 2014 Jan 18;87:12-25.
8. Perez de Souza L, Alseekh S, Scossa F, Fernie AR. Ultra-high-performance liquid chromatography high-resolution mass spectrometry variants for metabolomics research. Nature Methods. 2021 May 10.
9. Wen Z, He M, Peng C, Rao Y, Li J, Li Z, Du L, Li Y, Zhou M, Hui O, Feng Y. Metabolomics and 16S rRNA Gene sequencing analyses of changes in the intestinal flora and biomarkers induced by gastrodia-uncaria treatment in a rat model of chronic migraine. Frontiers in pharmacology. 2019 Dec 17;10:1425.
10. Emwas AH, Roy R, McKay RT, Tenori L, Saccenti E, Gowda GA, Raftery D, Alahmari F, Jaremko L, Jaremko M, Wishart DS. NMR spectroscopy for metabolomics research. Metabolites. 2019 Jul;9(7):123.
11. Silverstein RM, Bassler GC. Spectrometric identification of organic compounds. Journal of Chemical Education. 1962 Nov;39(11):546.
12. Dona AC, Kyriakides M, Scott F, Shephard EA, Varshavi D, Veselkov K, Everett JR. A guide to the identification of metabolites in NMR-based metabonomics/metabolomics experiments. Computational and structural biotechnology journal. 2016 Jan 1;14:135-53.
13. Pearce JT, Athersuch TJ, Ebbels TM, Lindon JC, Nicholson JK, Keun HC. Robust algorithms for automated chemical shift calibration of 1D 1H NMR spectra of blood serum. Analytical chemistry. 2008 Sep 15;80(18):7158-62.
14. Weinberger MA, Heggie RM, Holmes HL. Solvent effects in the nuclear magnetic resonance spectra of benzalmalononitriles. Canadian Journal of Chemistry. 1965 Sep 1;43(9):2585-93.
15. Fan TW, Lane AN. Structure-based profiling of metabolites and isotopomers by NMR. Progress in Nuclear Magnetic Resonance Spectroscopy. 2008;2(52):69-117.
16. Herrmann T, Güntert P, Wüthrich K. Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. Journal of biomolecular NMR. 2002 Nov;24(3):171-89.

17. Loria JP, Rance M, Palmer AG. A relaxation-compensated Carr− Purcell− Meiboom− Gill sequence for characterizing chemical exchange by NMR spectroscopy. Journal of the American Chemical Society. 1999 Mar 17;121(10):2331-2.
18. Foxall PJ, Parkinson JA, Sadler IH, Lindon JC, Nicholson JK. Analysis of biological fluids using 600 MHz proton NMR spectroscopy: application of homonuclear two-dimensional J-resolved spectroscopy to urine and blood plasma for spectral simplification and assignment. Journal of pharmaceutical and biomedical analysis. 1993 Jan 1;11(1):21-31.
19. Marion D, Wüthrich K. Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of 1 H-1 H spin-spin coupling constants in proteins. InNMR in structural biology: A collection of papers by Kurt Wüthrich 1995 (pp. 114-121).
20. Cloarec O, Dumas ME, Craig A, Barton RH, Trygg J, Hudson J, Blancher C, Gauguier D, Lindon JC, Holmes E, Nicholson J. Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic 1H NMR data sets. Analytical chemistry. 2005 Mar 1;77(5):1282-9.
21. Keeler J. Understanding NMR spectroscopy. John Wiley & Sons; 2011 Sep 19.
22. Dumez JN, Milani J, Vuichoud B, Bornet A, Lalande-Martin J, Tea I, Yon M, Maucourt M, Deborde C, Moing A, Frydman L. Hyperpolarized NMR of plant and cancer cell extracts at natural abundance. Analyst. 2015;140(17):5860-3.
23. Gowda GN, Raftery D. Can NMR solve some significant challenges in metabolomics?. Journal of Magnetic Resonance. 2015 Nov 1;260:144-60.
24. Moco S, Collino S, Rezzi S, Martin FP. Metabolomics perspectives in pediatric research. Pediatric research. 2013 Apr;73(2):570-6.
25. Wang F, Debik J, Andreassen T, Euceda LR, Haukaas TH, Cannet C, Schäfer H, Bathen TF, Giskeødegård GF. Effect of Repeated Freeze–Thaw Cycles on NMR-Measured Lipoproteins and Metabolites in Biofluids. Journal of proteome research. 2019 Sep 2;18(10):3681-8.
26. Voehler MW, Collier G, Young JK, Stone MP, Germann MW. Performance of cryogenic probes as a function of ionic strength and sample tube geometry. Journal of Magnetic Resonance. 2006 Nov 1;183(1):102-9.
27. Zhang Y, Liu P, Li Y, Zhang AH. Exploration of metabolite signatures using high-throughput mass spectrometry coupled with multivariate data analysis. RSC advances. 2017;7(11):6780-7.
28. Schoeman JC, Harms AC, Van Weeghel M, Berger R, Vreeken RJ, Hankemeier T. Development and application of a UHPLC–MS/MS metabolomics based comprehensive systemic and tissue-specific screening method for inflammatory, oxidative and nitrosative stress. Analytical and bioanalytical chemistry. 2018 Apr;410(10):2551-68.
29. Mazzini FN, Cook F, Gounarides J, Marciano S, Haddad L, Tamaroff AJ, Casciato P, Narvaez A, Mascardi MF, Anders M, Orozco F. Plasma and stool metabolomic biomarkers of non-alcoholic fatty liver disease in Argentina. medRxiv. 2020 Jan 1.
30. Chen L, Zhong F, Zhu J. Bridging Targeted and Untargeted Mass Spectrometry-Based Metabolomics via Hybrid Approaches. Metabolites. 2020 Sep;10(9):348.
31. Wold S, Esbensen K, Geladi P. Principal component analysis. Chemometrics and intelligent laboratory systems. 1987 Aug 1;2(1-3):37-52.

32. Wold S, Jonsson J, Sjörström M, Sandberg M, Rännar S. DNA and peptide sequences and chemical processes multivariately modelled by principal component analysis and partial least-squares projections to latent structures. Analytica Chimica Acta. 1993 May 28;277(2):239-53.
33. Galindo-Prieto B, Eriksson L, Trygg J. Variable influence on projection (VIP) for orthogonal projections to latent structures (OPLS). Journal of Chemometrics. 2014 Aug;28(8):623-32.
34. Shomorony I, Cirulli ET, Huang L, Napier LA, Heister RR, Hicks M, Cohen IV, Yu HC, Swisher CL, Schenker-Ahmed NM, Li W. An unsupervised learning approach to identify novel signatures of health and disease from multimodal data. Genome medicine. 2020 Dec;12(1):1-4.
35. Argelaguet R, Velten B, Arnol D, Dietrich S, Zenz T, Marioni JC, Buettner F, Huber W, Stegle O. Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets. Molecular systems biology. 2018 Jun;14(6):e8124.
36. Mukaka MM. A guide to appropriate use of correlation coefficient in medical research. Malawi medical journal. 2012;24(3):69-71.
37. Altman DG. Practical statistics for medical research. CRC press; 1990 Nov 22.
38. Swinscow TD, Campbell MJ. Statistics at square one. London: Bmj; 2002 Apr.
39. Holten D. Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data. IEEE Transactions on visualization and computer graphics. 2006 Nov 20;12(5):741-8.
40. Chowdhury F, Williams A, Johnson P. Validation and comparison of two multiplex technologies, Luminex® and Mesoscale Discovery, for human cytokine profiling. Journal of immunological methods. 2009 Jan 1;340(1):55-64.
41. Hara E. Relationship between obesity, gut microbiome and hepatocellular carcinoma development. Digestive diseases. 2015;33(3):346-5
42. Sartor RB. Genetics and environmental interactions shape the intestinal microbiome to promote inflammatory bowel disease versus mucosal homeostasis. Gastroenterology. 2010 Dec 1;139(6):1816-9.
43. Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. The ISME journal. 2017 Dec;11(12):2639-43.
44. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. Nature methods. 2016 Jul;13(7):581-3.
45. Dona, A. C.; Jimenez, B.; Schafer, H.; Humpfer, E.; Spraul, M.; Lewis, M. R.; Pearce, J. T.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. Precision High-Throughput Proton NMR Spectroscopy of Human Urine, Serum, and Plasma for Large-Scale Metabolic Phenotyping. *Anal Chem* **2014**, *86* (19), 9887–9894. https://doi.org/10.1021/ac5025039.
46. Jiménez, B.; Holmes, E.; Heude, C.; Tolson, R. F.; Harvey, N.; Lodge, S. L.; Chetwynd, A. J.; Cannet, C.; Fang, F.; Pearce, J. T. M.; Lewis, M. R.; Viant, M. R.; Lindon, J. C.; Spraul, M.; Schäfer, H.; Nicholson, J. K. Quantitative Lipoprotein Subclass and Low Molecular Weight Metabolite Analysis in Human Serum and Plasma by 1H NMR Spectroscopy in a Multilaboratory Trial. *Anal. Chem.* **2018**, *90* (20), 11962–11971.
47. Nicholson, J. K.; Foxall, P. J. D.; Spraul, M.; Farrant, R. D.; Lindon, J. C. 750 MHz 1H and 1H-13C NMR Spectroscopy of Human Blood Plasma. *Anal. Chem.* **1995**, *67* (5),

793–811. https://doi.org/10.1021/ac00101a004.
48. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological). 1995 Jan;57(1):289-300
49. Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. Analytical chemistry. 2006 Jul 1;78(13):4281-90.
50. Veselkov KA, Lindon JC, Ebbels TM, Crockford D, Volynkin VV, Holmes E, Davies DB, Nicholson JK. Recursive segment-wise peak alignment of biological 1H NMR spectra for improved metabolic biomarker recovery. Analytical chemistry. 2009 Jan 1;81(1):56-66.
51. Gray N, Lawler NG, Yang R, Morillon AC, Gay MC, Bong SH, Holmes E, Nicholson JK, Whiley L. A simultaneous exploratory and quantitative amino acid and biogenic amine metabolic profiling platform for rapid disease phenotyping via UPLC-QToF-MS. Talanta. 2021 Feb 1;223:121872.
52. Whiley L, Nye LC, Grant I, Andreas N, Chappell KE, Sarafian MH, Misra R, Plumb RS, Lewis MR, Nicholson JK, Holmes E. Ultrahigh-performance liquid chromatography tandem mass spectrometry with electrospray ionization quantification of tryptophan metabolites and markers of gut health in serum and plasma—Application to clinical and epidemiology cohorts. Analytical chemistry. 2019 Mar 8;91(8):5207-16.
53. Chong J, Soufan O, Li C, Caraus I, Li S, Bourque G, Wishart DS, Xia J. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. Nucleic acids research. 2018 Jul 2;46(W1):W486-94.
54. Siu A, Wong YR. Economic impact of SARS: the case of Hong Kong. Asian Economic Papers. 2004;3(1):62-83.
55. Ramadan N, Shaib H. Middle East respiratory syndrome coronavirus (MERS-CoV): A review. Germs. 2019 Mar;9(1):35.
56. Rothan HA, Byrareddy SN. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. Journal of autoimmunity. 2020 May 1;109:102433.
57. Pal M, Berhanu G, Desalegn C, Kandi V. Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2): an update. Cureus. 2020 Mar;12(3).
58. Fehr AR, Perlman S. Coronaviruses: an overview of their replication and pathogenesis. Coronaviruses. 2015:1-23.
59. Ye ZW, Yuan S, Yuen KS, Fung SY, Chan CP, Jin DY. Zoonotic origins of human coronaviruses. International journal of biological sciences. 2020;16(10):1686.
60. Cevik M, Kuppalli K, Kindrachuk J, Peiris M. Virology, transmission, and pathogenesis of SARS-CoV-2. bmj. 2020 Oct 23;371.
61. Gedefaw L, Ullah S, Leung PH, Cai Y, Yip SP, Huang CL. Inflammasome Activation-Induced Hypercoagulopathy: Impact on Cardiovascular Dysfunction Triggered in COVID-19 Patients. Cells. 2021 Apr;10(4):916.
62. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P. A novel coronavirus from patients with pneumonia in China, 2019. New England journal of medicine. 2020 Jan 24.
63. Tang S, Mao Y, Jones RM, Tan Q, Ji JS, Li N, Shen J, Lv Y, Pan L, Ding P, Wang X. Aerosol transmission of SARS-CoV-2? Evidence, prevention and control. Environment international. 2020 Nov;144:106039.

64. Egloff C, Vauloup-Fellous C, Picone O, Mandelbrot L, Roques P. Evidence and possible mechanisms of rare maternal-fetal transmission of SARS-CoV-2. Journal of Clinical Virology. 2020 May 18:104447.
65. Johansen MD, Irving A, Montagutelli X, Tate MD, Rudloff I, Nold MF, Hansbro NG, Kim RY, Donovan C, Liu G, Faiz A. Animal and translational models of SARS-CoV-2 infection and COVID-19. Mucosal immunology. 2020 Nov;13(6):877-91.
66. Zhang R, Li Y, Zhang AL, Wang Y, Molina MJ. Identifying airborne transmission as the dominant route for the spread of COVID-19. Proceedings of the National Academy of Sciences. 2020 Jun 30;117(26):14857-63.
67. Petrosillo N, Viceconte G, Ergonul O, Ippolito G, Petersen E. COVID-19, SARS and MERS: are they closely related?. Clinical Microbiology and Infection. 2020 Jun 1;26(6):729-34.
68. Yang X, Yu Y, Xu J, Shu H, Liu H, Wu Y, Zhang L, Yu Z, Fang M, Yu T, Wang Y. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. The Lancet Respiratory Medicine. 2020 May 1;8(5):475-81.
69. Lin D, Liu L, Zhang M, Hu Y, Yang Q, Guo J, Dai Y, Xu Y, Cai Y, Chen X, Huang K. Evaluations of the serological test in the diagnosis of 2019 novel coronavirus (SARS-CoV-2) infections during the COVID-19 outbreak. European Journal of Clinical Microbiology & Infectious Diseases. 2020 Dec;39(12):2271-7.
70. Munster VJ, Koopmans M, van Doremalen N, van Riel D, de Wit E. A novel coronavirus emerging in China—key questions for impact assessment. New England Journal of Medicine. 2020 Feb 20;382(8):692-4.
71. Sharma A, Tiwari S, Deb MK, Marty JL. Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2): a global pandemic and treatment strategies. International journal of antimicrobial agents. 2020 Aug 1;56(2):106054.
72. Tufan A, Güler AA, Matucci-Cerinic M. COVID-19, immune system response, hyperinflammation and repurposing antirheumatic drugs. Turkish journal of medical sciences. 2020 Apr 21;50(SI-1):620-32.
73. Tay MZ, Poh CM, Rénia L, MacAry PA, Ng LF. The trinity of COVID-19: immunity, inflammation and intervention. Nature Reviews Immunology. 2020 Jun;20(6):363-74.
74. Gemmati D, Bramanti B, Serino ML, Secchiero P, Zauli G, Tisato V. COVID-19 and individual genetic susceptibility/receptivity: role of ACE1/ACE2 genes, immunity, inflammation and coagulation. might the double X-chromosome in females be protective against SARS-CoV-2 compared to the single X-chromosome in males?. International journal of molecular sciences. 2020 Jan;21(10):3474.
75. Chanana N, Palmo T, Sharma K, Kumar R, Graham BB, Pasha Q. Sex-derived attributes contributing to SARS-CoV-2 mortality. American Journal of Physiology-Endocrinology and Metabolism. 2020 May 1;319(3):E562-7.
76. Stawicki SP, Jeanmonod R, Miller AC, Paladino L, Gaieski DF, Yaffee AQ, De Wulf A, Grover J, Papadimos TJ, Bloem C, Galwankar SC. The 2019–2020 novel coronavirus (severe acute respiratory syndrome coronavirus 2) pandemic: A joint american college of academic international medicine-world academic council of emergency medicine multidisciplinary COVID-19 working group consensus paper. Journal of global infectious diseases. 2020 Apr;12(2):47.
77. Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, Cuomo-Dannenburg G, Thompson H, Walker PG, Fu H, Dighe A. Estimates of the severity of coronavirus

disease 2019: a model-based analysis. The Lancet infectious diseases. 2020 Jun 1;20(6):669-77.
78. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, Wang B, Xiang H, Cheng Z, Xiong Y, Zhao Y. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China. Jama. 2020 Mar 17;323(11):1061-9.
79. Shi Y, Wang Y, Shao C, Huang J, Gan J, Huang X, Bucci E, Piacentini M, Ippolito G, Melino G. COVID-19 infection: the perspectives on immune responses.
80. Cevik M, Tate M, Lloyd O, Maraolo AE, Schafers J, Ho A. SARS-CoV-2, SARS-CoV, and MERS-CoV viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and meta-analysis. The Lancet Microbe. 2020 Nov 19.
81. Gibson PG, Qin L, Puah SH. COVID-19 acute respiratory distress syndrome (ARDS): clinical features and differences from typical pre-COVID-19 ARDS. Med J Aust. 2020 Jul 1;213(2):54-6.
82. Li X, Ma X. Acute respiratory failure in COVID-19: is it "typical" ARDS?. Critical Care. 2020 Dec;24:1-5.
83. Tisoncik JR, Korth MJ, Simmons CP, Farrar J, Martin TR, Katze MG. Into the eye of the cytokine storm. Microbiology and Molecular Biology Reviews. 2012 Mar 1;76(1):16-32.
84. Zaim S, Chong JH, Sankaranarayanan V, Harky A. COVID-19 and multi-organ response. Current problems in cardiology. 2020 Apr 28:100618.
85. Chee, Y. J.; Ng, S. J. H.; Yeoh, E. Diabetic Ketoacidosis Precipitated by Covid-19 in a Patient with Newly Diagnosed Diabetes Mellitus. *Diabetes Research and Clinical Practice*. Elsevier Ireland Ltd June 1, 2020, p 108166. https://doi.org/10.1016/j.diabres.2020.108166.
86. Zhang, C.; Shi, L.; Wang, F. S. Liver Injury in COVID-19: Management and Challenges. *The Lancet Gastroenterology and Hepatology*. Elsevier Ltd May 1, 2020, pp 428–430. https://doi.org/10.1016/S2468-1253(20)30057-1.
87. Cai, Q.; Huang, D.; Yu, H.; Zhu, Z.; Xia, Z.; Su, Y.; Li, Z.; Zhou, G.; Gou, J.; Qu, J.; Sun, Y.; Liu, Y.; He, Q.; Chen, J.; Liu, L.; Xu, L. COVID-19: Abnormal Liver Function Tests. *J. Hepatol.* **2020**. https://doi.org/10.1016/j.jhep.2020.04.006.
88. Bompard, F.; Monnier, H.; Saab, I.; Tordjman, M.; Abdoul, H.; Fournier, L.; Sanchez, O.; Lorut, C.; Chassagnon, G.; Revel, M. Pulmonary Embolism in Patients with Covid-19 Pneumonia. *Eur. Respir. J.* **2020**, 2001365. https://doi.org/10.1183/13993003.01365-2020.
89. Ellul, M. A.; Benjamin, L.; Singh, B.; Lant, S.; Michael, B. D.; Easton, A.; Kneen, R.; Defres, S.; Sejvar, J.; Solomon, T.; Mbchb, L.; Michael, B. D.; Phd, E.; Ellul, M. A.; Benjamin, L.; Singh, B.; Lant, S.; Michael, B. D.; Easton, A.; Kneen, R.; Defres, S.; Sejvar, J.; Solomon, T. Neurological Associations of COVID-19. *Lancet Neurol.* **2020**, *0* (0). https://doi.org/10.1016/S1474-4422(20)30221-0.
90. Paterson, R. W.; Brown, R. L.; Benjamin, L.; Nortley, R.; Wiethoff, S.; Bharucha, T.; Jayaseelan, D. L.; Kumar, G.; Raftopoulos, R. E.; Zambreanu, L.; Vivekanandam, V.; Khoo, A.; Geraldes, R.; Chinthapalli, K.; Boyd, E.; Tuzlali, H.; Price, G.; Christofi, G.; Morrow, J.; McNamara, P.; McLoughlin, B.; Lim, S. T.; Mehta, P. R.; Levee, V.; Keddie, S.; Yong, W.; Trip, S. A.; Foulkes, A. J. M.; Hotton, G.; Miller, T. D.; Everitt, A. D.; Carswell, C.; Davies, N. W. S.; Yoong, M.; Attwell, D.; Sreedharan, J.; Silber, E.; Schott, J. M.; Chandratheva, A.; Perry, R. J.; Simister, R.; Checkley, A.; Longley, N.; Farmer, S. F.; Carletti, F.; Houlihan, C.; Thom, M.; Lunn, M. P.; Spillane, J.; Howard, R.; Vincent,

A.; Werring, D. J.; Hoskote, C.; Jäger, H. R.; Manji, H.; Zandi, M. S.; Group, the U. C. L. Q. S. N. H. for N. and N. C.-19 S. The Emerging Spectrum of COVID-19 Neurology: Clinical, Radiological and Laboratory Findings. *Brain* **2020**. https://doi.org/10.1093/brain/awaa240.

91. Ronco, C.; Reis, T.; Husain-Syed, F. Management of Acute Kidney Injury in Patients with COVID-19. *The Lancet Respiratory Medicine*. Lancet Publishing Group 2020. https://doi.org/10.1016/S2213-2600(20)30229-0.
92. Mao, R.; Qiu, Y.; He, J. S.; Tan, J. Y.; Li, X. H.; Liang, J.; Shen, J.; Zhu, L. R.; Chen, Y.; Iacucci, M.; Ng, S. C.; Ghosh, S.; Chen, M. H. Manifestations and Prognosis of Gastrointestinal and Liver Involvement in Patients with COVID-19: A Systematic Review and Meta-Analysis. *Lancet Gastroenterol. Hepatol.* **2020**, *5* (7), 667–678. https://doi.org/10.1016/S2468-1253(20)30126-6.
93. Zheng, Y. Y.; Ma, Y. T.; Zhang, J. Y.; Xie, X. COVID-19 and the Cardiovascular System. *Nature Reviews Cardiology*. Nature Research May 1, 2020, pp 259–260. https://doi.org/10.1038/s41569-020-0360-5.
94. Oxley, T. J.; Mocco, J.; Majidi, S.; Kellner, C. P.; Shoirah, H.; Singh, I. P.; De Leacy, R. A.; Shigematsu, T.; Ladner, T. R.; Yaeger, K. A.; Skliut, M.; Weinberger, J.; Dangayach, N. S.; Bederson, J. B.; Tuhrim, S.; Fifi, J. T. Large-Vessel Stroke as a Presenting Feature of Covid-19 in the Young. *N. Engl. J. Med.* **2020**, *382* (20), e60. https://doi.org/10.1056/NEJMc2009787.
95. Deeks, J. J.; Dinnes, J.; Takwoingi, Y.; Davenport, C.; Spijker, R.; Taylor-Phillips, S.; Adriano, A.; Beese, S.; Dretzke, J.; Ferrante di Ruffano, L.; Harris, I. M.; Price, M. J.; Dittrich, S.; Emperador, D.; Hooft, L.; Leeflang, M. M.; Van den Bruel, A.; Cochrane COVID-19 Diagnostic Test Accuracy Group. Antibody Tests for Identification of Current and Past Infection with SARS-CoV-2. *Cochrane database Syst. Rev.* **2020**, *6* (6), CD013652.
96. Udugama B, Kadhiresan P, Kozlowski HN, Malekjahani A, Osborne M, Li VY, Chen H, Mubareka S, Gubbay JB, Chan WC. Diagnosing COVID-19: the disease and tools for detection. ACS nano. 2020 Mar 30;14(4):3822-35.
97. Boni MF, Lemey P, Jiang X, Lam TT, Perry BW, Castoe TA, Rambaut A, Robertson DL. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. Nature Microbiology. 2020 Nov;5(11):1408-17.
98. Oran DP, Topol EJ. Prevalence of asymptomatic SARS-CoV-2 infection: a narrative review. Annals of internal medicine. 2020 Sep 1;173(5):362-7.
99. Delano MJ, Ward PA. The immune system's role in sepsis progression, resolution, and long-term outcome. Immunological reviews. 2016 Nov;274(1):330-53.
100. Shi Y, Wang Y, Shao C, Huang J, Gan J, Huang X, Bucci E, Piacentini M, Ippolito G, Melino G. COVID-19 infection: the perspectives on immune responses.
101. Ni W, Yang X, Yang D, Bao J, Li R, Xiao Y, Hou C, Wang H, Liu J, Yang D, Xu Y. Role of angiotensin-converting enzyme 2 (ACE2) in COVID-19. Critical Care. 2020 Dec;24(1):1-0.
102. Mogensen TH. Pathogen recognition and inflammatory signaling in innate immune defenses. Clinical microbiology reviews. 2009 Apr 1;22(2):240-73.
103. Attaway AH, Scheraga RG, Bhimraj A, Biehl M, Hatipoğlu U. Severe covid-19 pneumonia: pathogenesis and clinical management. bmj. 2021 Mar 10;372.
104. Bolondi G, Russo E, Gamberini E, Circelli A, Meca MC, Brogi E, Viola L, Bissoni L, Poletti V, Agnoletti V. Iron metabolism and lymphocyte characterisation during

Covid-19 infection in ICU patients: an observational cohort study. World Journal of Emergency Surgery. 2020 Dec;15(1):1-8.
105. Giamarellos-Bourboulis EJ, Netea MG, Rovina N, Akinosoglou K, Antoniadou A, Antonakos N, Damoraki G, Gkavogianni T, Adami ME, Katsaounou P, Ntaganou M. Complex immune dysregulation in COVID-19 patients with severe respiratory failure. Cell host & microbe. 2020 Jun 10;27(6):992-1000.
106. Horby, P.; Lim, W. S.; Emberson, J.; Mafham, M.; Bell, J.; Linsell, L.; Staplin, N.; Brightling, C.; Ustianowski, A.; Elmahi, E.; Prudon, B.; Green, C.; Felton, T.; Chadwick, D.; Rege, K.; Fegan, C.; Chappell, L. C.; Faust, S. N.; Jaki, T.; Jeffery, K.; Montgomery, A.; Rowan, K.; Juszczak, E.; Baillie, J. K.; Haynes, R.; Landray, M. J.; Group, R. C. Dexamethasone for COVID-19-Preliminary Report Effect of Dexamethasone in Hospitalized Patients with COVID-19-Preliminary Report RECOVERY Collaborative Group*. medRxiv 2020, 2020.06.22.20137273.
107. Grein, J.; Ohmagari, N.; Shin, D.; Diaz, G.; Asperges, E.; Castagna, A.; Feldt, T.; Green, G.; Green, M. L.; Lescure, F. X.; Nicastri, E.; Oda, R.; Yo, K.; Quiros-Roldan, E.; Studemeister, A.; Redinski, J.; Ahmed, S.; Bernett, J.; Chelliah, D.; Chen, D.; Chihara, S.; Cohen, S. H.; Cunningham, J.; D'Arminio Monforte, A.; Ismail, S.; Kato, H.; Lapadula, G.; L'Her, E.; Maeno, T.; Majumder, S.; Massari, M.; Mora-Rillo, M.; Mutoh, Y.; Nguyen, D.; Verweij, E.; Zoufaly, A.; Osinusi, A. O.; DeZure, A.; Zhao, Y.; Zhong, L.; Chokkalingam, A.; Elboudwarej, E.; Telep, L.; Timbs, L.; Henne, I.; Sellers, S.; Cao, H.; Tan, S. K.; Winterbourne, L.; Desai, P.; Mera, R.; Gaggar, A.; Myers, R. P.; Brainard, D. M.; Childs, R.; Flanigan, T. Compassionate Use of Remdesivir for Patients with Severe Covid-19. N. Engl. J. Med. 2020, 382 (24), 2327–2336. https://doi.org/10.1056/NEJMoa2007016.
108. Wang, Y.; Zhang, D.; Du, G.; Du, R.; Zhao, J.; Jin, Y.; Fu, S.; Gao, L.; Cheng, Z.; Lu, Q.; Hu, Y.; Luo, G.; Wang, K.; Lu, Y.; Li, H.; Wang, S.; Ruan, S.; Yang, C.; Mei, C.; Wang, Y.; Ding, D.; Wu, F.; Tang, X.; Ye, X.; Ye, Y.; Liu, B.; Yang, J.; Yin, W.; Wang, A.; Fan, G.; Zhou, F.; Liu, Z.; Gu, X.; Xu, J.; Shang, L.; Zhang, Y.; Cao, L.; Guo, T.; Wan, Y.; Qin, H.; Jiang, Y.; Jaki, T.; Hayden, F. G.; Horby, P. W.; Cao, B.; Wang, C. Remdesivir in Adults with Severe COVID-19: A Randomised, Double-Blind, Placebo-Controlled, Multicentre Trial. Lancet 2020, 395 (10236), 1569–1578.
109. Kyriakidis NC, López-Cortés A, González EV, Grimaldos AB, Prado EO. SARS-CoV-2 vaccines strategies: a comprehensive review of phase 3 candidates. npj Vaccines. 2021 Feb 22;6(1):1-7.
110. Gómez CE, Perdiguero B, Esteban M. Emerging sars-cov-2 variants and impact in global vaccination programs against sars-cov-2/covid-19. Vaccines. 2021 Mar;9(3):243.
111. Cele S, Gazy I, Jackson L, Hwa SH, Tegally H, Lustig G, Giandhari J, Pillay S, Wilkinson E, Naidoo Y, Karim F. Escape of SARS-CoV-2 501Y. V2 from neutralization by convalescent plasma. Nature. 2021 May;593(7857):142-6.
112. Nussbaumer-Streit B, Mayr V, Dobrescu AI, Chapman A, Persad E, Klerings I, Wagner G, Siebert U, Ledinger D, Zachariah C, Gartlehner G. Quarantine alone or in combination with other public health measures to control COVID-19: a rapid review. Cochrane Database of Systematic Reviews. 2020(9).
113. Shen, B.; Yi, X.; Sun, Y.; Bi, X.; Du, J.; Zhang, C.; Quan, S.; Zhang, F.; Sun, R.; Qian, L.; Ge, W.; Liu, W.; Liang, S.; Chen, H.; Zhang, Y.; Li, J.; Xu, J.; He, Z.; Chen, B.; Wang, J.; Yan, H.; Zheng, Y.; Wang, D.; Zhu, J.; Kong, Z.; Kang, Z.; Liang, X.; Ding, X.; Ruan, G.;

Xiang, N.; Cai, X.; Gao, H.; Li, L.; Li, S.; Xiao, Q.; Lu, T.; Zhu, Y.; Liu, H.; Chen, H.; Guo, T. Proteomic and Metabolomic Characterization of COVID-19 Patient Sera. Cell 2020.
114. Kimhofer, T., Lodge, S., Whiley, L., Gray, N., Loo, R.L., Lawler, N.G., Nitschke, P., Bong, S.H., Morrison, D.L., Begum, S. and Richards, T., 2020. Integrative modeling of quantitative plasma lipoprotein, metabolic, and amino acid data reveals a multiorgan pathological signature of SARS-CoV-2 infection. Journal of proteome research, 19(11), pp.4442-4454.
115. Meoni G, Ghini V, Maggi L, Vignoli A, Mazzoni A, Salvati L, Capone M, Vanni A, Tenori L, Fontanari P, Lavorini F. Metabolomic/lipidomic profiling of COVID-19 and individual response to tocilizumab. PLoS Pathogens. 2021 Feb 1;17(2):e1009243.
116. Bruzzone C, Bizkarguenaga M, Gil-Redondo R, Diercks T, Arana E, de Vicuña AG, Seco M, Bosch A, Palazón A, San Juan I, Laín A. SARS-CoV-2 infection dysregulates the metabolomic and lipidomic profiles of serum. Iscience. 2020 Oct 23;23(10):101645.
117. Shen B, Yi X, Sun Y, Bi X, Du J, Zhang C, Quan S, Zhang F, Sun R, Qian L, Ge W. Proteomic and metabolomic characterization of COVID-19 patient sera. Cell. 2020 Jul 9;182(1):59-72.
118. Thomas T, Stefanoni D, Reisz JA, Nemkov T, Bertolone L, Francis RO, Hudson KE, Zimring JC, Hansen KC, Hod EA, Spitalnik SL. COVID-19 infection alters kynurenine and fatty acid metabolism, correlating with IL-6 levels and renal status. JCI insight. 2020 Jul 23;5(14).
119. Roberts I, Muelas MW, Taylor JM, Davison AS, Xu Y, Grixti JM, Gotts N, Sorokin A, Goodacre R, Kell DB. Untargeted metabolomics of COVID-19 patient serum reveals potential prognostic markers of both severity and outcome. medRxiv. 2020 Jan 1.
120. Xiao N, Nie M, Pang H, Wang B, Hu J, Meng X, Li K, Ran X, Long Q, Deng H, Chen N. Integrated cytokine and metabolite analysis reveals immunometabolic reprogramming in COVID-19 patients with therapeutic implications. Nature communications. 2021 Mar 12;12(1):1-3.
121. Bylesjö M, Rantalainen M, Cloarec O, Nicholson JK, Holmes E, Trygg J. OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. Journal of Chemometrics: A Journal of the Chemometrics Society. 2006 Aug;20(8-10):341-51.
122. Barter PJ, Nicholls S, Rye KA, Anantharamaiah GM, Navab M, Fogelman AM. Antiinflammatory properties of HDL. Circulation research. 2004 Oct 15;95(8):764-72.
123. Arslan S, Ugurlu S, Bulut G, Akkurt I. The association between plasma D-dimer levels and community-acquired pneumonia. Clinics. 2010;65(6):593-7.
124. Ahnach M, Zbiri S, Nejjari S, Ousti F, Elkettani C. C-reactive protein as an early predictor of COVID-19 severity. Journal of Medical Biochemistry. 2020 Oct 2;39(4):500.
125. Jain S, Gautam V, Naseem S. Acute-phase proteins: As diagnostic tool. Journal of Pharmacy and Bioallied Sciences. 2011 Jan;3(1):118.
126. La Marca A, Capuzzo M, Paglia T, Roli L, Trenti T, Nelson SM. Testing for SARS-CoV-2 (COVID-19): a systematic review and clinical guide to molecular and serological in-vitro diagnostic assays. Reproductive biomedicine online. 2020 Jun 14.
127. Lodge S, Nitschke P, Kimhofer T, Wist J, Bong SH, Loo RL, Masuda R, Begum S, Richards T, Lindon JC, Bermel W. Diffusion and relaxation edited Proton NMR Spectroscopy of plasma reveals a High-Fidelity supramolecular biomarker signature of SARS-CoV-2 infection. Analytical chemistry. 2021 Feb 12;93(8):3976-86.

128. Malo AI, Rull A, Girona J, Domingo P, Fuertes-Martín R, Amigó N, Rodríguez-Borjabad C, Martínez-Micaelo N, Leal M, Peraire J, Correig X. Glycoprotein Profile Assessed by 1H-NMR as a Global Inflammation Marker in Patients with HIV Infection. A Prospective Study. Journal of clinical medicine. 2020 May;9(5):1344.
129. Elg SA, Mayer AR, Carson LF, Twiggs LB, Hill RB, Ramakrishnan S. α-1 acid glycoprotein is an immunosuppressive factor found in ascites from ovarian carcinoma. Cancer: Interdisciplinary International Journal of the American Cancer Society. 1997 Oct 15;80(8):1448-56.
130. Hochepied T, Berger FG, Baumann H, Libert C. α1-Acid glycoprotein: an acute phase protein with inflammatory and immunomodulating properties. Cytokine & growth factor reviews. 2003 Feb 1;14(1):25-34.
131. Gambacorti-Passerini C, Zucchetti M, Russo D, Frapolli R, Verga M, Bungaro S, Tornaghi L, Rossi F, Pioltelli P, Pogliani E, Alberti D. α1 acid glycoprotein binds to imatinib (STI571) and substantially alters its pharmacokinetics in chronic myeloid leukemia patients. Clinical Cancer Research. 2003 Feb 1;9(2):625-32.
132. Ballout RA, Remaley AT. GlycA: A new biomarker for systemic inflammation and cardiovascular disease (CVD) risk assessment. Journal of laboratory and precision medicine. 2020 Apr;5.
133. Lorenzo C, Festa A, Hanley AJ, Rewers MJ, Escalante A, Haffner SM. Novel Protein Glycan–Derived Markers of Systemic Inflammation and C-Reactive Protein in Relation to Glycemia, Insulin Resistance, and Insulin Secretion. Diabetes care. 2017 Mar 1;40(3):375-82.
134. Poeze M, Luiking YC, Breedveld P, Manders S, Deutz NE. Decreased plasma glutamate in early phases of septic shock with acute liver dysfunction is an independent predictor of survival. Clinical Nutrition. 2008 Aug 1;27(4):523-30.
135. Gruppen EG, Riphagen IJ, Connelly MA, Otvos JD, Bakker SJ, Dullaart RP. GlycA, a pro-inflammatory glycoprotein biomarker, and incident cardiovascular disease: relationship with C-reactive protein and renal function. PloS one. 2015 Sep 23;10(9):e0139057.
136. Rodríguez-Carrio J, Alperi-López M, López P, Pérez-Álvarez ÁI, Gil-Serret M, Amigó N, Ulloa C, Benavente L, Ballina-García FJ, Suárez A. GlycA Levels during the Earliest Stages of Rheumatoid Arthritis: Potential Use as a Biomarker of Subclinical Cardiovascular Disease. Journal of clinical medicine. 2020 Aug;9(8):2472.
137. Akinkuolie AO, Pradhan AD, Buring JE, Ridker PM, Mora S. Novel protein glycan side-chain biomarker and risk of incident type 2 diabetes mellitus. Arteriosclerosis, thrombosis, and vascular biology. 2015 Jun;35(6):1544-50.
138. Fuertes-Martín R, Correig X, Vallvé JC, Amigó N. Human Serum/Plasma Glycoprotein Analysis by 1H-NMR, an Emerging Method of Inflammatory Assessment. Journal of clinical medicine. 2020 Feb;9(2):354.
139. Costela-Ruiz VJ, Illescas-Montes R, Puerta-Puerta JM, Ruiz C, Melguizo-Rodríguez L. SARS-CoV-2 infection: The role of cytokines in COVID-19 disease. Cytokine & growth factor reviews. 2020 Jun 2.
140. Mohammad S, Aziz R, Al Mahri S, Malik SS, Haji E, Khan AH, Khatlani TS, Bouchama A. Obesity and COVID-19: what makes obese host so vulnerable?. Immunity & Ageing. 2021 Dec;18(1):1-0.

141. Scalsky RJ, Chen YJ, Desai K, O'Connell JR, Perry JA, Hong CC. Baseline cardiometabolic profiles and SARS-CoV-2 infection in the UK Biobank. Plos one. 2021 Apr 1;16(4):e0248602.
142. Andrade Silva M, da Silva AR, do Amaral MA, Fragas MG, Câmara NO. Metabolic Alterations in SARS-CoV-2 Infection and Its Implication in Kidney Dysfunction. Frontiers in Physiology. 2021 Feb 25;12:147.
143. Yadav R, Aggarwal S, Singh A. SARS-CoV-2-host dynamics: Increased risk of adverse outcomes of COVID-19 in obesity. Diabetes & Metabolic Syndrome: Clinical Research & Reviews. 2020 Sep 1;14(5):1355-60.
144. Stawicki SP, Jeanmonod R, Miller AC, Paladino L, Gaieski DF, Yaffee AQ, De Wulf A, Grover J, Papadimos TJ, Bloem C, Galwankar SC. The 2019–2020 novel coronavirus (severe acute respiratory syndrome coronavirus 2) pandemic: A joint american college of academic international medicine-world academic council of emergency medicine multidisciplinary COVID-19 working group consensus paper. Journal of global infectious diseases. 2020 Apr;12(2):47.
145. Llauradó G, Amigó N, Cano A, Ballesta S, Albert L, Mazarico I, Fernández-Veledo S, Pedro-Botet J, Vendrell J, González-Clemente JM. Specific nuclear magnetic resonance lipoprotein subclass profiles and central arterial stiffness in type 1 diabetes mellitus: a case control study. Journal of clinical medicine. 2019 Nov;8(11):1875.
146. Krauss RM. Lipids and lipoproteins in patients with type 2 diabetes. Diabetes care. 2004 Jun 1;27(6):1496-504.
147. Schofield J, Ho J, Soran H. Cardiovascular risk in type 1 diabetes mellitus. Diabetes Therapy. 2019 Jun;10(3):773-89.
148. Linton MF, Yancey PG, Davies SS, Jerome WG, Linton EF, Song WL, Doran AC, Vickers KC. The role of lipids and lipoproteins in atherosclerosis. Endotext [Internet]. 2019 Jan 3.
149. de Ferranti SD, de Boer IH, Fonseca V, Fox CS, Golden SH, Lavie CJ, Magge SN, Marx N, McGuire DK, Orchard TJ, Zinman B. Type 1 diabetes mellitus and cardiovascular disease: a scientific statement from the American Heart Association and American Diabetes Association [published online ahead of print August 11, 2014]. Circulation. doi.;10.
150. Stepien M, Duarte-Salles T, Fedirko V, Floegel A, Barupal DK, Rinaldi S, Achaintre D, Assi N, Tjønneland A, Overvad K, Bastide N. Alteration of amino acid and biogenic amine metabolism in hepatobiliary cancers: Findings from a prospective cohort study. International journal of cancer. 2016 Jan 15;138(2):348-60.
151. Chung MK, Zidar DA, Bristow MR, Cameron SJ, Chan T, Harding III CV, Kwon DH, Singh T, Tilton JC, Tsai EJ, Tucker NR. COVID-19 and cardiovascular disease: from bench to bedside. Circulation research. 2021 Apr 16;128(8):1214-36.
152. Connelly MA, Otvos JD, Shalaurova I, Playford MP, Mehta NN. GlycA, a novel biomarker of systemic inflammation and cardiovascular disease risk. Journal of translational medicine. 2017 Dec;15(1):1-5.
153. Costela-Ruiz VJ, Illescas-Montes R, Puerta-Puerta JM, Ruiz C, Melguizo-Rodríguez L. SARS-CoV-2 infection: The role of cytokines in COVID-19 disease. Cytokine & growth factor reviews. 2020 Jun 2.

154. Buszko M, Park JH, Verthelyi D, Sen R, Young HA, Rosenberg AS. The dynamic changes in cytokine responses in COVID-19: a snapshot of the current state of knowledge.
155. Brites F, Martin M, Guillas I, Kontush A. Antioxidative activity of high-density lipoprotein (HDL): Mechanistic insights into potential clinical benefit. BBA clinical. 2017 Dec 1;8:66-77.
156. Kaji H. High-density lipoproteins and the immune system. Journal of lipids. 2013 Jan 1;2013.
157. Wu Q, Zhou L, Sun X, Yan Z, Hu C, Wu J, Xu L, Li X, Liu H, Yin P, Li K. Altered lipid metabolism in recovered SARS patients twelve years after infection. Scientific reports. 2017 Aug 22;7(1):1-2.
158. Petrie JR, Guzik TJ, Touyz RM. Diabetes, hypertension, and cardiovascular disease: clinical insights and vascular mechanisms. Canadian Journal of Cardiology. 2018 May 1;34(5):575-84.
159. Ritchie SC, Kettunen J, Brozynska M, Nath AP, Havulinna AS, Männistö S, Perola M, Salomaa V, Ala-Korpela M, Abraham G, Würtz P. Elevated serum alpha-1 antitrypsin is a major component of GlycA-associated risk for future morbidity and mortality. PLoS One. 2019 Oct 23;14(10):e0223692.
160. Schwarcz R. Kynurenines and glutamate: multiple links and therapeutic implications. Advances in Pharmacology. 2016 Jan 1;76:13-37.
161. Schousboe A, Waagepetersen HS. Role of astrocytes in glutamate homeostasis: implications for excitotoxicity. Neurotoxicity research. 2005 Sep;8(3):221-5.
162. Pacheco R, Gallart T, Lluis C, Franco R. Role of glutamate on T-cell mediated immunity. Journal of neuroimmunology. 2007 Apr 1;185(1-2):9-19.
163. Newsholme P. Why is L-glutamine metabolism important to cells of the immune system in health, postinjury, surgery or infection?. The Journal of nutrition. 2001 Sep 1;131(9):2515S-22S.
164. Nicastro, H., Da Luz, C.R., Chaves, D.F.S., Bechara, L.R.G., Voltarelli, V.A., Rogero, M.M. and Lancha, A.H., 2012. Does branched-chain amino acids supplementation modulate skeletal muscle remodeling through inflammation modulation? Possible mechanisms of action. *Journal of nutrition and metabolism*, *2012*.
165. Kimura K, Nakamura Y, Inaba Y, Matsumoto M, Kido Y, Asahara SI, Matsuda T, Watanabe H, Maeda A, Inagaki F, Mukai C. Histidine augments the suppression of hepatic glucose production by central insulin action. Diabetes. 2013 Jul 1;62(7):2266-77.
166. Inoue M, Chang L, Hwang J, Chiang SH, Saltiel AR. The exocyst complex is required for targeting of Glut4 to the plasma membrane by insulin. Nature. 2003 Apr;422(6932):629-33.
167. Bosco MC, Rapisarda A, Massazza S, Melillo G, Young H, Varesio L. The tryptophan catabolite picolinic acid selectively induces the chemokines macrophage inflammatory protein-1α and-1β in macrophages. The Journal of Immunology. 2000 Mar 15;164(6):3283-91.
168. Huber AK, Giles DA, Segal BM, Irani DN. An emerging role for eotaxins in neurodegenerative disease. Clinical Immunology. 2018 Apr 1;189:29-33.
169. Hirschfield, G.M. and Gershwin, M.E., 2013. The immunobiology and pathophysiology of primary biliary cirrhosis. *Annual review of pathology: Mechanisms of disease*, *8*, pp.303-330.

170. Selmi, C., Mayo, M.J., Bach, N., Ishibashi, H., Invernizzi, P., Gish, R.G., Gordon, S.C., Wright, H.I., Zweiban, B., Podda, M. and Gershwin, M.E., 2004. Primary biliary cirrhosis in monozygotic and dizygotic twins: genetics, epigenetics, and environment. *Gastroenterology*, *127*(2), pp.485-492.
171. Montano-Loza, A.J., Hansen, B.E., Corpechot, C., Roccarina, D., Thorburn, D., Trivedi, P., Hirschfield, G., McDowell, P., Poupon, R., Dumortier, J. and Bosch, A., 2019. Factors associated with recurrence of primary biliary cholangitis after liver transplantation and effects on graft and patient survival. *Gastroenterology*, *156*(1), pp.96-107.
172. Kim, W.R., Brown Jr, R.S., Terrault, N.A. and El-Serag, H., 2002. Burden of liver disease in the United States: summary of a workshop. *Hepatology*, *36*(1), pp.227-242.
173. Sanyal, A.J., Yoon, S.K. and Lencioni, R., 2010. The etiology of hepatocellular carcinoma and consequences for treatment. *The oncologist*, *15*(Supplement 4), pp.14-22.
174. Jones, D.E., Metcalf, J.V., Collier, J.D., Bassendine, M.F. and James, O.F., 1997. Hepatocellular carcinoma in primary biliary cirrhosis and its impact on outcomes. *Hepatology*, *26*(5), pp.1138-1142.
175. Liang, Y., Yang, Z. and Zhong, R., 2012. Primary biliary cirrhosis and cancer risk: A systematic review and meta-analysis. *Hepatology*, *56*(4), pp.1409-1417.
176. Reshetnyak, V.I., 2015. Primary biliary cirrhosis: Clinical and laboratory criteria for its diagnosis. *World Journal of Gastroenterology: WJG*, *21*(25), p.7683.
177. Poupon, R., 2015. Liver alkaline phosphatase: a missing link between choleresis and biliary inflammation. *Hepatology*, *61*(6), pp.2080-2090.
178. Hofmann AF. Pharmacology of ursodeoxycholic acid, an enterohepatic drug. Scandinavian Journal of Gastroenterology. 1994 Jan 1;29(sup204):1-5.
179. Bahar, R., Wong, K.A., Liu, C.H. and Bowlus, C.L., 2018. Update on new drugs and those in development for the treatment of primary biliary cholangitis. *Gastroenterology & hepatology*, *14*(3), p.154.
180. Đanić, M., Stanimirov, B., Pavlović, N., Goločorbin-Kon, S., Al-Salami, H., Stankov, K. and Mikov, M., 2018. Pharmacological applications of bile acids and their derivatives in the treatment of metabolic syndrome. *Frontiers in pharmacology*, *9*.
181. Corpechot, C., Chazouillères, O. and Poupon, R., 2011. Early primary biliary cirrhosis: biochemical response to treatment and prediction of long-term outcome. *Journal of hepatology*, *55*(6), pp.1361-1367.
182. Paumgartner, G. and Beuers, U., 2002. Ursodeoxycholic acid in cholestatic liver disease: mechanisms of action and therapeutic use revisited. *Hepatology*, *36*(3), pp.525-531.
183. Cheung, A.C., Lammers, W.J., Perez, C.F.M., van Buuren, H.R., Gulamhusein, A., Trivedi, P.J., Lazaridis, K.N., Ponsioen, C.Y., Floreani, A., Hirschfield, G.M. and Corpechot, C., 2019. Effects of age and sex of response to ursodeoxycholic acid and transplant-free survival in patients with primary biliary cholangitis. *Clinical Gastroenterology and Hepatology*.
184. Sato, K., Meng, F., Fava, G., Glaser, S. and Alpini, G., 2018. Functional roles of gut bacteria imbalance in cholangiopathies. *Liver Research*.
185. Staels, B., Handelsman, Y. and Fonseca, V., 2010. Bile acid sequestrants for lipid and glucose control. *Current diabetes reports*, *10*(1), pp.70-77.

186. Giannini, E.G., Testa, R. and Savarino, V., 2005. Liver enzyme alteration: a guide for clinicians. *Cmaj*, *172*(3), pp.367-379.
187. Leung, P.S., Choi, J., Yang, G., Woo, E., Kenny, T.P. and Gershwin, M.E., 2016. A contemporary perspective on the molecular characteristics of mitochondrial autoantigens and diagnosis in primary biliary cholangitis. *Expert review of molecular diagnostics*, *16*(6), pp.697-705.
188. Lleo, A., Invernizzi, P., Mackay, I.R., Prince, H., Zhong, R.Q. and Gershwin, M.E., 2008. Etiopathogenesis of primary biliary cirrhosis. *World journal of gastroenterology: WJG*, *14*(21), p.3328.
189. Malik, R. and Hodgson, H., 2002. The relationship between the thyroid gland and the liver. *QJM: An International Journal of Medicine*, *95*(9), pp.559-569.
190. Gulamhusein, A.F., Juran, B.D. and Lazaridis, K.N., 2015, November. Genome-wide association studies in primary biliary cirrhosis. In *Seminars in liver disease* (Vol. 35, No. 04, pp. 392-401). Thieme Medical Publishers.
191. Lazaridis, K.N., Juran, B.D., Boe, G.M., Slusser, J.P., De Andrade, M., Homburger, H.A., Ghosh, K., Dickson, E.R., Lindor, K.D. and Petersen, G.M., 2007. Increased prevalence of antimitochondrial antibodies in first-degree relatives of patients with primary biliary cirrhosis. *Hepatology*, *46*(3), pp.785-792.
192. Chew, M. and Bowlus, C.L., 2018. Primary biliary cholangitis: Diagnosis and treatment. *Liver Research*, *2*(2), pp.81-86.
193. Triger, D.R., Berg, P.A. and Rodes, J., 1984. Epidemiology of primary biliary cirrhosis. *Liver*, *4*(3), pp.195-200.
194. Inamine, T. and Schnabl, B., 2018. Immunoglobulin A and liver diseases. *Journal of gastroenterology*, *53*(6), pp.691-700.
195. Smyk, D., Mytilinaiou, M.G., Rigopoulou, E.I. and Bogdanos, D.P., 2010. PBC triggers in water reservoirs, coal mining areas and waste disposal sites: from Newcastle to New York. *Disease markers*, *29*(6), pp.337-344.
196. Nicholson, J.K., Lindon, J.C. and Holmes, E., 1999. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica*, *29*(11), pp.1181-1189.
197. Fitian AI, Cabrera R. Disease monitoring of hepatocellular carcinoma through metabolomics. World journal of hepatology. 2017 Jan 8;9(1):1.
198. Yang Z, Kusumanchi P, Ross RA, Heathers L, Chandler K, Oshodi A, Thoudam T, Li F, Wang L, Liangpunsakul S. Serum metabolomic profiling identifies key metabolic signatures associated with pathogenesis of alcoholic liver disease in humans. Hepatology communications. 2019 Apr;3(4):542-2.
199. Loo RL, Chan Q, Antti H, Li JV, Ashrafian H, Elliott P, Stamler J, Nicholson JK, Holmes E, Wist J. Manuscript Strategy for improved characterisation of human metabolic phenotypes using a COmbined Multiblock Principal components Analysis with Statistical Spectroscopy (COMPASS). Bioinformatics. 2020 Jul 21.
200. Windey K, De Preter V, Verbeke K. Relevance of protein fermentation to gut health. Molecular nutrition & food research. 2012 Jan;56(1):184-96.
201. Mullish BH, Pechlivanis A, Barker GF, Thursz MR, Marchesi JR, McDonald JA. Functional microbiomics: evaluation of gut microbiota-bile acid metabolism interactions in health and disease. Methods. 2018 Oct 1;149:49-58.

202. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. Nature methods. 2016 Jul;13(7):581-3.
203. Wright ES. Using DECIPHER v2. 0 to analyze big biological sequence data in R. R Journal. 2016 Aug 1;8(1).
204. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome. 2018 Dec;6(1):1-4.
205. Martín-Fernández JA, Hron K, Templ M, Filzmoser P, Palarea-Albaladejo J. Bayesian-multiplicative treatment of count zeros in compositional data sets. Statistical Modelling. 2015 Apr;15(2):134-58.
206. Bathena SP, Thakare R, Gautam N, Mukherjee S, Olivera M, Meza J, Alnouti Y. Urinary bile acids as biomarkers for liver diseases I. Stability of the baseline profile in healthy subjects. Toxicological Sciences. 2015 Feb 1;143(2):296-307.
207. Swann JR, Want EJ, Geier FM, Spagou K, Wilson ID, Sidaway JE, Nicholson JK, Holmes E. Systemic gut microbial modulation of bile acid metabolism in host tissue compartments. Proceedings of the National Academy of Sciences. 2011 Mar 15;108(Supplement 1):4523-30.
208. Yang Z, Kusumanchi P, Ross RA, Heathers L, Chandler K, Oshodi A, Thoudam T, Li F, Wang L, Liangpunsakul S. Serum metabolomic profiling identifies key metabolic signatures associated with pathogenesis of alcoholic liver disease in humans. Hepatology communications. 2019 Apr;3(4):542-57.
209. Wang J, Yang G, Dubrovsky AM, Choi J, Leung PS. Xenobiotics and loss of tolerance in primary biliary cholangitis. World journal of gastroenterology. 2016 Jan 7;22(1):338.
210. Yoon S, Ji SC, Yang J, Kim YK, Lee S, Yu KS, Jang IJ, Chung JY, Cho JY. Ursodeoxycholic acid improves liver function via phenylalanine/tyrosine pathway and microbiome remodelling in patients with liver dysfunction. Scientific reports. 2018 Aug 8;8(1):1-1.
211. Massafra V, Milona A, Vos HR, Ramos RJ, Gerrits J, Willemsen EC, Pittol JM, Ijssennagger N, Houweling M, Prinsen HC, Verhoeven-Duif NM. Farnesoid X receptor activation promotes hepatic amino acid catabolism and ammonium clearance in mice. Gastroenterology. 2017 May 1;152(6):1462-76.
212. Kristiansen RG, Rose CF, Ytrebø LM. Glycine and hyperammonemia: potential target for the treatment of hepatic encephalopathy. Metabolic brain disease. 2016 Dec 1;31(6):1269-73.
213. Mariat D, Firmesse O, Levenez F, Guimarães VD, Sokol H, Doré J, Corthier G, Furet JP. The Firmicutes/Bacteroidetes ratio of the human microbiota changes with age. BMC microbiology. 2009 Dec;9(1):123.
214. Tang, R., Wei, Y., Li, Y., Chen, W., Chen, H., Wang, Q., Yang, F., Miao, Q., Xiao, X., Zhang, H. and Lian, M., 2018. Gut microbial profile is altered in primary biliary cholangitis and partially restored after UDCA therapy. *Gut*, *67*(3), pp.534-541.
215. Pallister T, Jackson MA, Martin TC, Zierer J, Jennings A, Mohney RP, MacGregor A, Steves CJ, Cassidy A, Spector TD, Menni C. Hippurate as a metabolomic marker of gut microbiome diversity: Modulation by diet and relationship to metabolic syndrome. Scientific reports. 2017 Oct 20;7(1):1-9.
216. Barrios C, Beaumont M, Pallister T, Villar J, Goodrich JK, Clark A, Pascual J, Ley RE, Spector TD, Bell JT, Menni C. Gut-microbiota-metabolite axis in early renal function decline. PloS one. 2015 Aug 4;10(8):e0134311.

217. Meara JG, Leather AJ, Hagander L, Alkire BC, Alonso N, Ameh EA, Bickler SW, Conteh L, Dare AJ, Davies J, Mérisier ED. Global Surgery 2030: evidence and solutions for achieving health, welfare, and economic development. The lancet. 2015 Aug 8;386(9993):569-624.
218. Smolle C, Cambiaso-Daniel J, Forbes AA, Wurzer P, Hundeshagen G, Branski LK, Huss F, Kamolz LP. Recent trends in burn epidemiology worldwide: a systematic review. Burns. 2017 Mar 1;43(2):249-57.
219. World Health Organisation. Burns. 2018 Available from: https://www.who.int/en/news-room/fact-sheets/detail/burns
220. Rowan MP, Cancio LC, Elster EA, Burmeister DM, Rose LF, Natesan S, et al. Burn wound healing and treatment: Review and advancements [Internet]. Vol. 19, Critical Care. BioMed Central Ltd.; 2015 [cited 2020 Nov 3]. Available from: /pmc/articles/PMC4464872/?report=abstract
221. Barrett LW, Fear VS, Waithman JC, Wood FM, Fear MW. Understanding acute burn injury as a chronic disease. Burns & trauma. 2019 Dec 1;7.
222. Bohr S, Patel SJ, Shen K, Vitalo AG, Brines M, Cerami A, et al. Alternative erythropoietin-mediated signaling prevents secondary microvascular thrombosis and inflammation within cutaneous burns. Proceedings of the National Academy of Sciences of the United States of America. 2013 Feb 26 ; 110(9):3513–8.
223. Auger C, Samadi O, Jeschke MG. The biochemical alterations underlying post-burn hypermetabolism. Vol. 1863, Biochimica et Biophysica Acta - Molecular Basis of Disease. Elsevier B.V.; 2017. p. 2633–44.
224. Edmondson SJ, Ali Jumabhoy I, Murray A. Time to start putting down the knife: A systematic review of burns excision tools of randomised and non-randomised trials [Internet]. Vol. 44, Burns. Elsevier Ltd; 2018. p. 1721–37.
225. Finnerty CC, Jeschke MG, Branski LK, Barret JP, Dziewulski P, Herndon DN. Hypertrophic scarring: the greatest unmet challenge after burn injury [Internet]. Vol. 388, The Lancet. Lancet Publishing Group; 2016 [cited 2020 Nov 10]. p. 1427–36. Available from: /pmc/articles/PMC5380137/?report=abstract
226. Kawasumi A, Sagawa N, Hayashi S, Yokoyama H, Tamura K. Wound healing in mammals and amphibians: toward limb regeneration in mammals. New Perspectives in Regeneration. 2012:33-49.
227. Hart DW, Wolf SE, Herndon DN, Chinkes DL, Lal SO, Obeng MK, Beauford RB, RT RP. Energy expenditure and caloric balance after burn: increased feeding leads to fat rather than lean mass accretion. Annals of surgery. 2002 Jan;235(1):152.
228. Gauglitz GG, Herndon DN, Kulp GA, Meyer III WJ, Jeschke MG. Abnormal insulin sensitivity persists up to three years in pediatric patients post-burn. The Journal of Clinical Endocrinology & Metabolism. 2009 May 1;94(5):1656-64.
229. Robinson LE, van Soeren MH. Insulin resistance and hyperglycemia in critical illness: role of insulin in glycemic control. AACN Advanced Critical Care. 2004 Jan;15(1):45-62.
230. Zawacki BE, Spitzer KW, Mason Jr AD, Johns LA. Does increased evaporative water loss cause hypermetabolism in burned patients?. Annals of surgery. 1970 Feb;171(2):236.
231. Aulick LH, Wilmore DW, Mason Jr AD, Pruitt Jr BA. Influence of the burn wound on peripheral circulation in thermally injured patients. American Journal of Physiology-Heart and Circulatory Physiology. 1977 Oct 1;233(4):H520-6.

232. Stone II R, Natesan S, Kowalczewski CJ, Mangum LH, Clay NE, Clohessy RM, Carlsson AH, Tassin DH, Chan RK, Rizzo JA, Christy RJ. Advancements in regenerative strategies through the continuum of burn care. Frontiers in pharmacology. 2018 Jul 9;9:672.
233. Monsaingeon A, Molimard R. Wound Healing Comparison of Healing Rates of Burn Wounds and of Excisional Wounds. European Surgical Research. 1976;8(4):337-43.
234. Hart DW, Wolf SE, Chinkes DL, Beauford RB, Mlcak RP, Heggers JP, Wolfe RR, Herndon DN. Effects of early excision and aggressive enteral feeding on hypermetabolism, catabolism, and sepsis after severe burn. Journal of Trauma and Acute Care Surgery. 2003 Apr 1;54(4):755-64.
235. Jeschke MG, Gauglitz GG, Kulp GA, Finnerty CC, Williams FN, Kraft R, Suman OE, Mlcak RP, Herndon DN. Long-term persistance of the pathophysiologic response to severe burn injury. PloS one. 2011 Jul 18;6(7):e21245.
236. Nielson CB, Duethman NC, Howard JM, Moncure M, Wood JG. Burns: pathophysiology of systemic complications and current management. Journal of Burn Care & Research. 2017 Jan 1;38(1):e469-81.
237. Kim HS, Kim JH, Yim H, Kim D. Changes in the levels of interleukins 6, 8, and 10, tumor necrosis factor alpha, and granulocyte-colony stimulating factor in Korean burn patients: relation to burn size and postburn time. Ann Lab Med. 2012 Sep 1;32(5):339-44.
238. Rutkowski R, Pancewicz SA, Rutkowski K, Rutkowska J. Reactive oxygen and nitrogen species in inflammatory process. Polski merkuriusz lekarski: organ Polskiego Towarzystwa Lekarskiego. 2007 Aug 1;23(134):131-6.
239. Zhang J, Sio SW, Moochhala S, Bhatia M. Role of hydrogen sulfide in severe burn injury-induced inflammation in mice. Molecular medicine. 2010 Sep;16(9):417-24.
240. Horton JW. Free radicals and lipid peroxidation mediated injury in burn trauma: the role of antioxidant therapy. Toxicology. 2003 Jul 15;189(1-2):75-88.
241. Chaudhry H, Zhou J, Zhong YI, Ali MM, McGuire F, Nagarkatti PS, Nagarkatti M. Role of cytokines as a double-edged sword in sepsis. In Vivo. 2013 Nov 1;27(6):669-84.
242. Dennis EA, Norris PC. Eicosanoid storm in infection and inflammation. Nature Reviews Immunology. 2015 Aug;15(8):511-23.
243. Jensen J, Rustad PI, Kolnes AJ, Lai YC. The role of skeletal muscle glycogen breakdown for regulation of insulin sensitivity by exercise. Frontiers in physiology. 2011 Dec 30;2:112.
244. Porter C, Hurren NM, Herndon DN, Børsheim E. Whole body and skeletal muscle protein turnover in recovery from burns. International journal of burns and trauma. 2013;3(1):9.
245. Merritt EK, Cross JM, Bamman MM. Inflammatory and protein metabolism signaling responses in human skeletal muscle after burn injury. Journal of burn care & research. 2012 Mar 1;33(2):291-7.
246. Porter C, Herndon DN, Bhattarai N, Ogunbileje JO, Szczesny B, Szabo C, Toliver-Kinsky T, Sidossis LS. Differential acute and chronic effects of burn trauma on murine skeletal muscle bioenergetics. Burns. 2016 Feb 1;42(1):112-22.
247. Zhang H, Wang HY, Bassel-Duby R, Maass DL, Johnston WE, Horton JW, Tao W. Role of interleukin-6 in cardiac inflammation and dysfunction after burn complicated by sepsis. American Journal of Physiology-Heart and Circulatory Physiology. 2007 May;292(5):H2408-16.

248. Willis MS, Carlson DL, DiMaio JM, White MD, White DJ, Adams IV GA, Horton JW, Giroir BP. Macrophage migration inhibitory factor mediates late cardiac dysfunction after burn injury. American Journal of Physiology-Heart and Circulatory Physiology. 2005 Feb;288(2):H795-804.
249. Yang G, Zhang Y, Wu D, Wang C, Yang YJ, Fan SJ, Xia L, Wei Y, Peng X. 1H-NMR metabolomics identifies significant changes in hypermetabolism after glutamine administration in burned rats. American journal of translational research. 2019;11(12):7286.
250. Abdullahi A, Barayan D, Vinaik R, Diao L, Yu N, Jeschke MG. Activation of ER stress signalling increases mortality after a major trauma. Journal of Cellular and Molecular Medicine. 2020 Sep;24(17):9764-73.
251. Jeschke MG. The hepatic response to thermal injury: is the liver important for postburn outcomes?. Molecular medicine. 2009 Sep;15(9):337-51.
252. Biffl WL, Moore EE, Moore FA, Peterson VM. Interleukin-6 in the injured patient. Marker of injury or mediator of inflammation?. Annals of surgery. 1996 Nov;224(5):647.
253. Porter C, Tompkins RG, Finnerty CC, Sidossis LS, Suman OE, Herndon DN. The metabolic stress response to burn trauma: current understanding and therapies. The Lancet. 2016 Oct 1;388(10052):1417-26.
254. Sadiq A, Shah A, Jeschke MG, Belo C, Qasim Hayat M, Murad S, Amini-Nik S. The role of serotonin during skin healing in post-thermal injury. International journal of molecular sciences. 2018 Apr;19(4):1034.
255. Drummond PD, Dawson LF, Finch PM, Drummond ES, Wood FM, Fear MW. Up-regulation of cutaneous α1-adrenoceptors after a burn. Burns. 2015 Sep 1;41(6):1227-34.
256. Herrera-Escobar JP, Seshadri AJ, Stanek E, Lu K, Han K, Sanchez S, Kaafarani HM, Salim A, Levy-Carrick NC, Nehra D. Mental Health Burden After Injury: It's About More than Just Posttraumatic Stress Disorder. Annals of surgery. 2020 Jun 4.
257. Mason SA, Nathens AB, Byrne JP, Diong C, Fowler RA, Karanicolas PJ, Moineddin R, Jeschke MG. Increased rate of long-term mortality among burn survivors: a population-based matched cohort study. Annals of surgery. 2019 Jun 1;269(6):1192-9.
258. Duke JM, Randall SM, Fear MW, Boyd JH, Rea S, Wood FM. Burn induced nervous system morbidity among burn and non-burn trauma patients compared with non-injured people. Burns. 2019 Aug 1;45(5):1041-50.
259. Farina JA, Rosique MJ, Rosique RG. Curbing inflammation in burn patients. International journal of inflammation. 2013 May 20;2013.
260. Rani M, Schwacha MG. Aging and the pathogenic response to burn. Aging and disease. 2012 Apr;3(2):171.
261. Duke JM, Randall SM, Wood FM, Boyd JH, Fear MW. Burns and long-term infectious disease morbidity: a population-based study. Burns. 2017 Mar 1;43(2):273-81.
262. Valvis SM, Waithman J, Wood FM, Fear MW, Fear VS. The immune response to skin trauma is dependent on the etiology of injury in a mouse model of burn and excision. Journal of Investigative Dermatology. 2015 Aug 1;135(8):2119-28.
263. Calum H, Høiby N, Moser C. Burn mouse models. InPseudomonas methods and protocols 2014 (pp. 793-802). Humana Press, New York, NY.

264. Porter C, Herndon DN, Sidossis LS, Børsheim E. The impact of severe burns on skeletal muscle mitochondrial function. Burns. 2013 Sep 1;39(6):1039-47.
265. Baptista PV, McCusker MP, Carvalho A, Ferreira DA, Mohan NM, Martins M, Fernandes AR. Nano-strategies to fight multidrug resistant bacteria—"A Battle of the Titans". Frontiers in microbiology. 2018 Jul 2;9:1441.
266. Knowlin L, Strassle PD, Williams FN, Thompson R, Jones S, Weber DJ, van Duin D, Cairns BA, Charles A. Burn injury outcomes in patients with pre-existing diabetic mellitus: risk of hospital-acquired infections and inpatient mortality. Burns. 2018 Mar 1;44(2):272-9.
267. Zhang Y, Cai B, Jiang H, Yan H, Yang H, Peng J, Wang W, Ma S, Wu X, Peng X. Use of 1 h-nuclear magnetic resonance to screen a set of biomarkers for monitoring metabolic disturbances in severe burn patients. Critical care. 2014 Aug;18(4):1-0.
268. Tzika AA, Astrakas LG, Cao H, Mintzopoulos D, Zhang Q, Padfield K, Yu H, Mindrinos MN, Rahme LG, Tompkins RG. Murine intramyocellular lipids quantified by NMR act as metabolic biomarkers in burn trauma. International journal of molecular medicine. 2008 Jun 1;21(6):825-32.
269. Yang G, Zhang Y, Wu D, Wang C, Yang YJ, Fan SJ, Xia L, Wei Y, Peng X. 1H-NMR metabolomics identifies significant changes in hypermetabolism after glutamine administration in burned rats. American journal of translational research. 2019;11(12):7286.
270. Hendrickson C, Linden K, Kreyer S, Beilman G, Scaravilli V, Wendorff D, Necsoiu C, Batchinsky AI, Cancio LC, Chung KK, Lusczek ER. 1H-NMR metabolomics identifies significant changes in metabolism over time in a porcine model of severe burn and smoke inhalation. Metabolites. 2019 Jul;9(7):142.
271. Alkhalil A, Day A, Ball R, Kumar R, Moffatt LT, Shupp JW. 393 Validation of a Rodent Model for Studying Metabolomic Response to Burn Injury. Journal of Burn Care & Research. 2019 Mar 9;40(Supplement_1):S169-70.
272. Liu XR, Zheng XF, Ji SZ, Lv YH, Zheng DY, Xia ZF, Zhang WD. Metabolomic analysis of thermally injured and/or septic rats. Burns. 2010 Nov 1;36(7):992-8.
273. Friston D, Laycock H, Nagy I, Want EJ. Microdialysis workflow for metabotyping superficial pathologies: application to burn injury. Analytical chemistry. 2019 Apr 25;91(10):6541-8.
274. Beffa DC, Carter EA, Lu XM, Yu YM, Prelack K, Sheridan RL, Young VR, Fischman AJ, Tompkins RG. Negative chemical ionization gas chromatography/mass spectrometry to quantify urinary 3-methylhistidine: application to burn injury. Analytical biochemistry. 2006 Aug 1;355(1):95-101.
275. Johnson BZ, McAlister S, McGuire HM, Palanivelu V, Stevenson A, Richmond P, Palmer DJ, Metcalfe J, Prescott SL, Wood FM, Fazekas de St Groth B. Pediatric burn survivors have long-term immune dysfunction with diminished vaccine response. Frontiers in immunology. 2020 Jul 21;11:1481.
276. Blasco H, Bessy C, Plantier L, Lefevre A, Piver E, Bernard L, Marlet J, Stefic K, Benz-de Bretagne I, Cannet P, Lumbu H. The specific metabolome profiling of patients infected by SARS-COV-2 supports the key role of tryptophan-nicotinamide pathway and cytosine metabolism. Scientific reports. 2020 Oct 8;10(1):1-2.
277. van Gageldonk PG, van Schaijk FG, van der Klis FR, Berbers GA. Development and validation of a multiplex immunoassay for the simultaneous determination of serum

antibodies to Bordetella pertussis, diphtheria and tetanus. Journal of immunological methods. 2008 Jun 1;335(1-2):79-89.
278. Wall S, Allorto NL. The hypermetabolic response to burn injury and modulation of this response: an overview. Wound Healing Southern Africa. 2015 Nov 1;8(2).
279. Cao YF, Li J, Zhang Z, Liu J, Sun XY, Feng XF, Luo HH, Yang W, Li SN, Yang X, Fang ZZ. Plasma levels of amino acids related to urea cycle and risk of type 2 diabetes mellitus in Chinese adults. Frontiers in endocrinology. 2019 Feb 18;10:50.
280. Lercher A, Bhattacharya A, Popa AM, Caldera M, Schlapansky MF, Baazim H, Agerer B, Gürtl B, Kosack L, Májek P, Brunner JS. Type I interferon signaling disrupts the hepatic urea cycle and alters systemic metabolism to suppress T cell function. Immunity. 2019 Dec 17;51(6):1074-87.
281. He F, Yin Z, Wu C, Xia Y, Wu M, Li P, Zhang H, Yin Y, Li N, Zhu G, Ren W. L-serine lowers the inflammatory responses during Pasteurella multocida infection. Infection and immunity. 2019 Dec;87(12).
282. Williams FN, Herndon DN, Jeschke MG. The hypermetabolic response to burn injury and interventions to modify this response. Clinics in plastic surgery. 2009 Oct 1;36(4):583-96.
283. Daubner SC, Le T, Wang S. Tyrosine hydroxylase and regulation of dopamine synthesis. Archives of biochemistry and biophysics. 2011 Apr 1;508(1):1-2.
284. Jeschke MG, Chinkes DL, Finnerty CC, Kulp G, Suman OE, Norbury WB, Branski LK, Gauglitz GG, Mlcak RP, Herndon DN. The pathophysiologic response to severe burn injury. Annals of surgery. 2008 Sep;248(3):387.
285. Ploder M, Neurauter G, Spittler A, Schroecksnadel K, Roth E, Fuchs D. Serum phenylalanine in patients post trauma and with sepsis correlate to neopterin concentrations. Amino acids. 2008 Aug;35(2):303-7.
286. Chase RA, Davies M, Trewby PN, Silk DB, Williams R. Plasma amino acid profiles in patients with fulminant hepatic failure treated by repeated polyacrylonitrile membrane hemodialysis. Gastroenterology. 1978 Dec 1;75(6):1033-40.
287. Matthews DE. An overview of phenylalanine and tyrosine kinetics in humans. The Journal of nutrition. 2007 Jun 1;137(6):1549S-55S.
288. Ginguay A, Cynober L, Curis E, Nicolis I. Ornithine aminotransferase, an important glutamate-metabolizing enzyme at the crossroads of multiple metabolic pathways. Biology. 2017 Mar;6(1):18.
289. Okun JG, Rusu PM, Chan AY, Wu Y, Yap YW, Sharkie T, Schumacher J, Schmidt KV, Roberts-Thomson KM, Russell RD, Zota A. Liver alanine catabolism promotes skeletal muscle atrophy and hyperglycaemia in type 2 diabetes. Nature metabolism. 2021 Mar;3(3):394-409.
290. Cruzat V, Macedo Rogero M, Noel Keane K, Curi R, Newsholme P. Glutamine: metabolism and immune function, supplementation and clinical translation. Nutrients. 2018 Nov;10(11):1564.
291. Dalangin R, Kim A, Campbell RE. The Role of Amino Acids in Neurotransmission and Fluorescent Tools for Their Detection. International Journal of Molecular Sciences. 2020 Jan;21(17):6197.
292. Lu YR, Fu XY, Shi LG, Jiang Y, Wu JL, Weng XJ, Wang ZP, Wu XY, Lin Z, Liu WB, Li HC. Decreased plasma neuroactive amino acids and increased nitric oxide levels in melancholic major depressive disorder. BMC psychiatry. 2014 Dec;14(1):1-7.

293. Geisler S, Gostner JM, Becker K, Ueberall F, Fuchs D. Immune activation and inflammation increase the plasma phenylalanine-to-tyrosine ratio. Pteridines. 2013 Jun 1;24(1):27-31.
294. Ashe K, Kelso W, Farrand S, Panetta J, Fazio T, De Jong G, Walterfang M. Psychiatric and cognitive aspects of phenylketonuria: The limitations of diet and promise of new treatments. Frontiers in psychiatry. 2019 Sep 10;10:561.
295. Himmerich H, Berthold-Losleben M, Pollmächer T. The relevance of the TNF-alpha system in psychiatric disorders. Fortschritte der Neurologie-Psychiatrie. 2009 May 4;77(6):334-45.
296. Griffin JW, Bradshaw PC. Amino acid catabolism in Alzheimer's disease brain: friend or foe?. Oxidative medicine and cellular longevity. 2017 Oct;2017.
297. Brocker C, Thompson DC, Vasiliou V. The role of hyperosmotic stress in inflammation and disease. Biomolecular concepts. 2012 Aug 1;3(4):345-64.
298. Zhang N, Sauve AA. Regulatory effects of NAD+ metabolic pathways on sirtuin activity. Progress in molecular biology and translational science. 2018 Jan 1;154:71-104.
299. Luo XH, Zhu Y, Mao J, Du RC. T cell immunobiology and cytokine storm of COVID-19. Scandinavian Journal of Immunology. 2021 Mar;93(3):e12989.
300. Piskin G, De Boer OJ, Van Der Loos CM, Teeling P, Bos JD, Teunissen MB. Overrepresentation of IL-17A and IL-22 producing CD8 T cells in lesional skin suggests their involvement in the pathogenesis of psoriasis. PloS one. 2010 Nov 24;5(11):e14108.
301. Bartel B, LeClere S, Magidin M, Zolman BK. Inputs to the active indole-3-acetic acid pool: de novo synthesis, conjugate hydrolysis, and indole-3-butyric acid b-oxidation. Journal of Plant Growth Regulation. 2001 Sep 1;20(3):198-216.
302. Pennock ND, White JT, Cross EW, Cheney EE, Tamburini BA, Kedl RM. T cell responses: naive to memory and everything in between. Advances in physiology education. 2013 Dec;37(4):273-83.
303. Schülke S. Induction of interleukin-10 producing dendritic cells as a tool to suppress allergen-specific T helper 2 responses. Frontiers in immunology. 2018 Mar 19;9:455.
304. Emami A, Javanmardi F, Rajaee M, Pirbonyeh N, Keshavarzi A, Fotouhi M, Hosseini SM. Predictive biomarkers for acute kidney injury in burn patients. Journal of burn care & research. 2019 Aug 14;40(5):601-5.
305. Wallner C, Huber J, Drysch M, Schmidt SV, Wagner JM, Dadras M, Lehnhardt M, Behr B. Myostatin upregulation in patients in the chronic phase of severe burn injury leads to muscle cell catabolism. European Surgical Research. 2019;60(1-2):86-96.
306. Børsheim E, Herndon DN, Hawkins HK, Suman OE, Cotter M, Klein GL. Pamidronate attenuates muscle loss after pediatric burn injury. Journal of Bone and Mineral Research. 2014 Jun;29(6):1369-72.
307. Constantinou C, Karavia EA, Xepapadaki E, Petropoulou PI, Papakosta E, Karavyraki M, Zvintzou E, Theodoropoulos V, Filou S, Hatziri A, Kalogeropoulou C. Advances in high-density lipoprotein physiology: surprises, overturns, and promises. American Journal of Physiology-Endocrinology and Metabolism. 2016.
308. Blonde-Cynober F, Aussel C, Cynober L. Abnormalities in branched-chain amino acid metabolismin cirrhosis: influence of hormonal and nutritional factors and directions for future research. Clinical Nutrition. 1999 Feb 1;18(1):5-13.

309. Cruzat V, Macedo Rogero M, Noel Keane K, Curi R, Newsholme P. Glutamine: metabolism and immune function, supplementation and clinical translation. Nutrients. 2018 Nov;10(11):1564.
310. Nie C, He T, Zhang W, Zhang G, Ma X. Branched chain amino acids: beyond nutrition metabolism. International journal of molecular sciences. 2018 Apr;19(4):954.
311. DeBerardinis RJ, Mancuso A, Daikhin E, Nissim I, Yudkoff M, Wehrli S, Thompson CB. Beyond aerobic glycolysis: transformed cells can engage in glutamine metabolism that exceeds the requirement for protein and nucleotide synthesis. Proceedings of the National Academy of Sciences. 2007 Dec 4;104(49):19345-50.
312. Valbuena GN, Rizzardini M, Cimini S, Siskos AP, Bendotti C, Cantoni L, Keun HC. Metabolomic analysis reveals increased aerobic glycolysis and amino acid deficit in a cellular model of amyotrophic lateral sclerosis. Molecular neurobiology. 2016 May 1;53(4):2222-40.
313. Porter C, Hardee J, Herndon DN, Suman OE. The role of exercise in the rehabilitation of patients with severe burns. Exercise and sport sciences reviews. 2015 Jan;43(1):34.
314. Elliott P, Posma JM, Chan Q, Garcia-Perez I, Wijeyesekera A, Bictash M, Ebbels TM, Ueshima H, Zhao L, Van Horn L, Daviglus M. Urinary metabolic signatures of human adiposity. Science translational medicine. 2015 Apr 29;7(285):285ra62-.
315. Connelly MA, Shimizu C, Winegar DA, Shalaurova I, Pourfarzib R, Otvos JD, Kanegaye JT, Tremoulet AH, Burns JC. Differences in GlycA and lipoprotein particle parameters may help distinguish acute kawasaki disease from other febrile illnesses in children. BMC pediatrics. 2016 Dec;16(1):1-1.
316. O'Connor M, Ponsonby AL, Collier F, Liu R, Sly PD, Azzopardi P, Lycett K, Goldfeld S, Arnup SJ, Burgner D, Priest N. Exposure to adversity and inflammatory outcomes in mid and late childhood. Brain, Behavior, & Immunity-Health. 2020 Dec 1;9:100146.
317. Shah RR, Smith RL. Addressing phenoconversion: the Achilles' heel of personalized medicine. British journal of clinical pharmacology. 2015 Feb;79(2):222-40.
318. Lawler NG, Gray N, Kimhofer T, Boughton B, Gay M, Yang R, Morillon AC, Chin ST, Ryan M, Begum S, Bong SH. Systemic Perturbations in Amine and Kynurenine Metabolism Associated with Acute SARS-CoV-2 Infection and Inflammatory Cytokine Responses. Journal of proteome research. 2021 Mar 16.
319. Holmes E, Wist J, Masuda R, Lodge S, Nitschke P, Kimhofer T, Loo RL, Begum S, Boughton B, Yang R, Morillon AC. Incomplete Systemic Recovery and Metabolic Phenoreversion in Post-Acute-Phase Nonhospitalized COVID-19 Patients: Implications for Assessment of Post-Acute COVID-19 Syndrome. Journal of Proteome Research. 2021 May 19.
320. Dinarello CA. Historical insights into cytokines. European journal of immunology. 2007 Nov;37(S1):S34-45.
321. Den Besten G, Van Eunen K, Groen AK, Venema K, Reijngoud DJ, Bakker BM. The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. Journal of lipid research. 2013 Sep 1;54(9):2325-40.

*Appendix I : Published Paper I*

# Integrative Modeling of Quantitative Plasma Lipoprotein, Metabolic, and Amino Acid Data Reveals a Multiorgan Pathological Signature of SARS-CoV-2 Infection

in
Journal of Proteome Research 2020, 19, 4442-4454

This is an open access article published under an ACS AuthorChoice License, which permits copying and redistribution of the article or any adaptations for non-commercial purposes.

Journal of proteome research

pubs.acs.org/jpr

Article

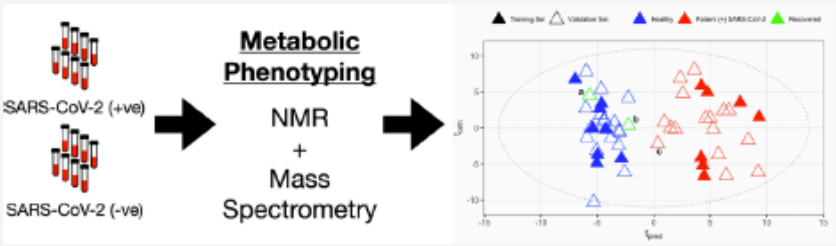
# Integrative Modeling of Quantitative Plasma Lipoprotein, Metabolic, and Amino Acid Data Reveals a Multiorgan Pathological Signature of SARS-CoV-2 Infection

Torben Kimhofer, Samantha Lodge, Luke Whiley, Nicola Gray, Ruey Leng Loo, Nathan G. Lawler, Philipp Nitschke, Sze-How Bong, David L. Morrison, Sofina Begum, Toby Richards, Bu B. Yeap, Chris Smith, Kenneth G. C. Smith, Elaine Holmes,* and Jeremy K. Nicholson*

Cite This: J. Proteome Res. 2020, 19, 4442−4454

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information




**ABSTRACT:** The metabolic effects of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection on human blood plasma were characterized using multiplatform metabolic phenotyping with nuclear magnetic resonance (NMR) spectroscopy and liquid chromatography−mass spectrometry (LC-MS). Quantitative measurements of lipoprotein subfractions, $\alpha$-1-acid glycoprotein, glucose, and biogenic amines were made on samples from symptomatic coronavirus disease 19 (COVID-19) patients who had tested positive for the SARS-CoV-2 virus ($n$ = 17) and from age- and gender-matched controls ($n$ = 25). Data were analyzed using an orthogonal-projections to latent structures (OPLS) method and used to construct an exceptionally strong (AUROC = 1) hybrid NMR-MS model that enabled detailed metabolic discrimination between the groups and their biochemical relationships. Key discriminant metabolites included markers of inflammation including elevated $\alpha$-1-acid glycoprotein and an increased kynurenine/tryptophan ratio. There was also an abnormal lipoprotein, glucose, and amino acid signature consistent with diabetes and coronary artery disease (low total and HDL Apolipoprotein A1, low HDL triglycerides, high LDL and VLDL triglycerides), plus multiple highly significant amino acid markers of liver dysfunction (including the elevated glutamine/glutamate and Fischer's ratios) that present themselves as part of a distinct SARS-CoV-2 infection pattern. A multivariate training-test set model was validated using independent samples from additional SARS-CoV-2 positive patients and controls. The predictive model showed a sensitivity of 100% for SARS-CoV-2 positivity. The breadth of the disturbed pathways indicates a systemic signature of SARS-CoV-2 positivity that includes elements of liver dysfunction, dyslipidemia, diabetes, and coronary heart disease risk that are consistent with recent reports that COVID-19 is a systemic disease affecting multiple organs and systems. Metabolights study reference: MTBLS2014.

**KEYWORDS:** *COVID-19, SARS-CoV-2, lipoproteins, amino acids, NMR spectroscopy, mass spectrometry, metabolic phenotyping, biomarkers, systems model, multiorgan damage, mosaic disease*

## INTRODUCTION

The coronavirus disease 19 (COVID-19) pandemic resulting from severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection has spread to 213 countries, infected >24 million people, and killed >820 thousand people worldwide. The disease continues unchecked in many countries and is still accelerating in some. COVID-19 symptoms range from nearly asymptomatic to mild sore throat and fatigue to severe respiratory distress, multiorgan failure, and, in respiratory cases, death due to an immunological cytokine storm.[1,2]

Recent studies indicate that SARS-CoV-2 can also precipitate a type of new-onset diabetes[3] and liver dysfunction,[4] with up to 77% of SARS-CoV-2 positive patients having abnormal liver

Special Issue: Proteomics in Pandemic Disease

Received: July 10, 2020
Published: August 17, 2020



ACS Publications © 2020 American Chemical Society

*Appendix I : Published Paper II*

# NMR Spectroscopic Windows on the Systemic Effects of SARS-CoV-2 Infection on Plasma Lipoproteins and Metabolites in Relation to Circulating Cytokines

in
Journal of Proteome Research 2021, 20, 1382-1396

This is an open access article published under a Creative Commons Attribution (CC-BY) License, which permits unrestricted use, distribution and reproduction in any medium, provided the author and source are cited.

Journal of proteome research

pubs.acs.org/jpr

Article

# NMR Spectroscopic Windows on the Systemic Effects of SARS-CoV-2 Infection on Plasma Lipoproteins and Metabolites in Relation to Circulating Cytokines

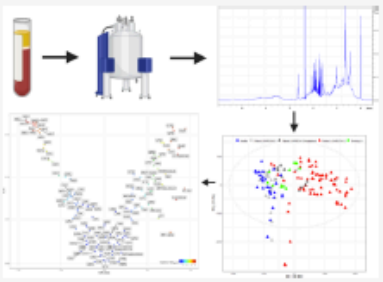
Samantha Lodge, Philipp Nitschke, Torben Kimhofer, Jerome D. Coudert, Sofina Begum, Sze-How Bong, Toby Richards, Dale Edgar, Edward Raby, Manfred Spraul, Hartmut Schaefer, John C. Lindon, Ruey Leng Loo, Elaine Holmes,* and Jeremy K. Nicholson*

Cite This: J. Proteome Res. 2021, 20, 1382−1396

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

**ABSTRACT:** To investigate the systemic metabolic effects of SARS-CoV-2 infection, we analyzed $^1$H NMR spectroscopic data on human blood plasma and co-modeled with multiple plasma cytokines and chemokines (measured in parallel). Thus, 600 MHz $^1$H solvent-suppressed single-pulse, spin-echo, and 2D J-resolved spectra were collected on plasma recorded from SARS-CoV-2 rRT-PCR-positive patients ($n$ = 15, with multiple sampling timepoints) and age-matched healthy controls ($n$ = 34, confirmed rRT-PCR negative), together with patients with COVID-19/influenza-like clinical symptoms who tested SARS-CoV-2 negative ($n$ = 35). We compared the single-pulse NMR spectral data with *in vitro* diagnostic research (IVDr) information on quantitative lipoprotein profiles (112 parameters) extracted from the raw 1D NMR data. All NMR methods gave highly significant discrimination of SARS-CoV-2 positive patients from controls and SARS-CoV-2 negative patients with individual NMR methods, giving different diagnostic information windows on disease-induced phenoconversion. Longitudinal trajectory analysis in selected patients indicated that metabolic recovery was incomplete in individuals without detectable virus in the recovery phase. We observed four plasma cytokine clusters that expressed complex differential statistical relationships with multiple lipoproteins and metabolites. These included the following: cluster 1, comprising MIP-1$\beta$, SDF-1$\alpha$, IL-22, and IL-1$\alpha$, which correlated with multiple increased LDL and VLDL subfractions; cluster 2, including IL-10 and IL-17A, which was only weakly linked to the lipoprotein profile; cluster 3, which included IL-8 and MCP-1 and were inversely correlated with multiple lipoproteins. IL-18, IL-6, and IFN-$\gamma$ together with IP-10 and RANTES exhibited strong positive correlations with LDL1−4 subfractions and negative correlations with multiple HDL subfractions. Collectively, these data show a distinct pattern indicative of a multilevel cellular immune response to SARS CoV-2 infection interacting with the plasma lipoproteome giving a strong and characteristic immunometabolic phenotype of the disease. We observed that some patients in the respiratory recovery phase and testing virus-free were still metabolically highly abnormal, which indicates a new role for these technologies in assessing full systemic recovery.



**KEYWORDS:** *plasma, COVID-19, SARS-CoV-2, NMR spectroscopy, single-pulse, spin-echo, IVDr, metabolic phenotyping, biomarkers, diagnostic modeling, lipoproteins*

## INTRODUCTION

COVID-19 is a complex respiratory and systemic disease caused by infection with the SARS-CoV-2 coronavirus. The virus, both directly and indirectly, causes systemic damage to multiple organs resulting in several new-onset pathologies including diabetes,[1] liver dysfunction,[2] neurological damage, and renal damage,[3,4] as well as possible pulmonary[5] and cardiovascular damage.[6] We have shown that SARS-CoV-2 infection induces a metabolic "phenoconversion" process that gives a characteristic complex blood plasma signature exhibiting multiple pathological effects of the disease and in a complex mosaic of biochemical change.[7] This COVID-19 phenoconversion signature is readily detectable by both nuclear magnetic resonance (NMR) and mass spectrometric techniques, with the spectral profiles providing multiple windows onto the disease sequelae in different organ systems.[7] The general term, phenoconversion, is defined here as follows:

Received: November 2, 2020
Published: January 11, 2021



ACS Publications

© 2021 The Authors. Published by American Chemical Society



https://dx.doi.org/10.1021/acs.jproteome.0c00876
J. Proteome Res. 2021, 20, 1382−1396

*Appendix I : Published Paper III*

# Diffusion and Relaxation Edited Proton NMR Spectroscopy of Plasma Reveals a High-Fidelity Supramolecular Biomarker Signature of SARS-CoV-2 Infection

in
Analytical Chemistry 2021, 93, 3976-3986

analytical chemistry

pubs.acs.org/ac

Article

# Diffusion and Relaxation Edited Proton NMR Spectroscopy of Plasma Reveals a High-Fidelity Supramolecular Biomarker Signature of SARS-CoV-2 Infection

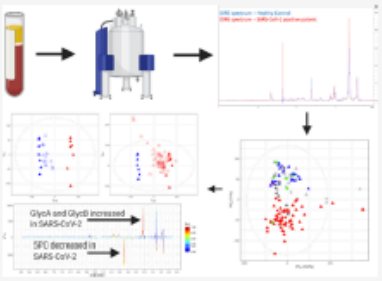
Samantha Lodge, Philipp Nitschke, Torben Kimhofer, Julien Wist, Sze-How Bong, Ruey Leng Loo, Reika Masuda, Sofina Begum, Toby Richards, John C. Lindon, Wolfgang Bermel, Tony Reinsperger, Hartmut Schaefer, Manfred Spraul, Elaine Holmes,* and Jeremy K. Nicholson*

Cite This: *Anal. Chem.* 2021, 93, 3976−3986

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

**ABSTRACT:** We have applied nuclear magnetic resonance spectroscopy based plasma phenotyping to reveal diagnostic molecular signatures of SARS-CoV-2 infection *via* combined diffusional and relaxation editing (DIRE). We compared plasma from healthy age-matched controls ($n$ = 26) with SARS-CoV-2 negative non-hospitalized respiratory patients and hospitalized respiratory patients ($n$ = 23 and 11 respectively) with SARS-CoV-2 rRT-PCR positive respiratory patients ($n$ = 17, with longitudinal sampling time-points). DIRE data were modelled using principal component analysis and orthogonal projections to latent structures discriminant analysis (O-PLS-DA), with statistical cross-validation indices indicating excellent model generalization for the classification of SARS-CoV-2 positivity for all comparator groups (area under the receiver operator characteristic curve = 1). DIRE spectra show biomarker signal combinations conferred by differential concentrations of metabolites with selected molecular mobility properties. These comprise the following: (a) composite $N$-acetyl signals from $\alpha$-1-acid glycoprotein and other glycoproteins (designated GlycA and GlycB) that were elevated in SARS-CoV-2 positive patients [$p = 2.52 \times 10^{-10}$ (GlycA) and $1.25 \times 10^{-9}$ (GlycB) *vs* controls], (b) two diagnostic supramolecular phospholipid composite signals that were identified (SPC-A and SPC-B) from the $-^{+}N-(CH_3)_3$ choline headgroups of lysophosphatidylcholines carried on plasma glycoproteins and from phospholipids in high-density lipoprotein subfractions (SPC-A) together with a phospholipid component of low-density lipoprotein (SPC−B). The integrals of the summed SPC signals ($SPC_{total}$) were reduced in SARS-CoV-2 positive patients relative to both controls ($p = 1.40 \times 10^{-7}$) and SARS-CoV-2 negative patients ($p = 4.52 \times 10^{-8}$) but were not significantly different between controls and SARS-CoV-2 negative patients. The identity of the SPC signal components was determined using one and two dimensional diffusional, relaxation, and statistical spectroscopic experiments. The $SPC_{total}$/GlycA ratios were also significantly different for control *versus* SARS-CoV-2 positive patients ($p = 1.23 \times 10^{-10}$) and for SARS-CoV-2 negatives *versus* positives ($p = 1.60 \times 10^{-9}$). Thus, plasma $SPC_{total}$ and $SPC_{total}$/GlycA are proposed as sensitive molecular markers for SARS-CoV-2 positivity that could effectively augment current COVID-19 diagnostics and may have value in functional assessment of the disease recovery process in patients with long-term symptoms.




## INTRODUCTION

The COVID-19 disease pandemic resulting from SARS-CoV-2 infection has so far resulted in over 101 million cases and 2.4 million deaths worldwide. The range of clinical expression of COVID-19 is extreme, varying from asymptomatic or mild to severe respiratory distress and multiple organ damage, with or without respiratory involvement. There is an unmet need for accurate diagnosis and prediction of disease severity at an early stage so that individual infections can be monitored and managed effectively. There is also a need for new functional markers of patient recovery in COVID-19, especially for classifying the complex systemic complications of the disease. Hence, we applied biofluid metabolic phenotyping to COVID-19 as these approaches have previously given deep insights into a range of pathological processes that can be modelled using multivariate statistical methods to achieve classification of disease subtypes and detection of novel biomarkers.[1,2]

We recently proposed an extension to the *metabolic phenoconversion* concept to explore the systemic shifts in plasma

Received: November 25, 2020
Accepted: January 28, 2021
Published: February 12, 2021



ACS Publications

© 2021 The Authors. Published by American Chemical Society



https://dx.doi.org/10.1021/acs.analchem.0c04952
*Anal. Chem.* 2021, 93, 3976−3986

*Appendix I : Published Paper IV*

# Low Volume in Vitro Diagnostic Proton NMR Spectroscopy of Human Blood Plasma for Lipoprotein and Metabolite Analysis: Application to SARS-CoV-2 Biomarkers

in
Journal of Proteome Research 2021, 20, 1415-1423

This is an open access article published under a Creative Commons Non-Commercial No Derivative Works (CC-BY-NC-ND) Attribution License, which permits copying and redistribution of the article, and creation of adaptations, all for non-commercial purposes.

Journal of proteome research

pubs.acs.org/jpr

Technical Note

# Low Volume in Vitro Diagnostic Proton NMR Spectroscopy of Human Blood Plasma for Lipoprotein and Metabolite Analysis: Application to SARS-CoV-2 Biomarkers

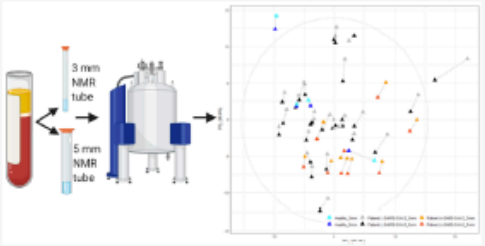
Samantha Lodge, Philipp Nitschke, Ruey Leng Loo, Torben Kimhofer, Sze-How Bong, Toby Richards, Sofina Begum, Manfred Spraul, Hartmut Schaefer, John C. Lindon, Elaine Holmes,* and Jeremy K. Nicholson*

Cite This: J. Proteome Res. 2021, 20, 1415−1423

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

**ABSTRACT:** The utility of low sample volume in vitro diagnostic (IVDr) proton nuclear magnetic resonance ($^1$H NMR) spectroscopic experiments on blood plasma for information recovery from limited availability or high value samples was exemplified using plasma from patients with SARS-CoV-2 infection and normal controls. $^1$H NMR spectra were obtained using solvent-suppressed 1D, spin−echo (CPMG), and 2-dimensional J-resolved (JRES) spectroscopy using both 3 mm outer diameter SampleJet NMR tubes (100 μL plasma) and 5 mm SampleJet NMR tubes (300 μL plasma) under in vitro diagnostic conditions. We noted near identical diagnostic models in both standard and low volume IVDr lipoprotein analysis (measuring 112 lipoprotein parameters) with a comparison of the two tubes yielding $R^2$ values ranging between 0.82 and 0.99 for the 40 paired lipoprotein parameters samples. Lipoprotein measurements for the 3 mm tubes were achieved without time penalty over the 5 mm tubes as defined by biomarker recovery for SARS-CoV-2. Overall, biomarker pattern recovery for the lipoproteins was extremely similar, but there were some small positive offsets in the linear equations for several variables due to small shimming artifacts, but there was minimal degradation of the biological information. For the standard untargeted 1D, CPMG, and JRES NMR experiments on the same samples, the reduced signal-to-noise was more constraining and required greater scanning times to achieve similar differential diagnostic performance (15 min per sample per experiment for 3 mm 1D and CPMG, compared to 4 min for the 5 mm tubes). We conclude that the 3 mm IVDr method is fit-for-purpose for quantitative lipoprotein measurements, allowing the preparation of smaller volumes for high value or limited volume samples that is common in clinical studies. If there are no analytical time constraints, the lower volume experiments are equally informative for untargeted profiling.




**KEYWORDS:** *blood plasma, $^1$H NMR spectroscopy, IVDr, miniaturization, metabolic phenotyping, biomarkers, low volume, 3 mm NMR tubes, SARS-CoV-2, spin−echo, J-resolved*

## INTRODUCTION

High field proton NMR spectroscopy of biofluids coupled with pattern recognition methods have been applied extensively to problems in metabolic biochemistry, toxicology, and experimental diagnostics for more than three decades.[1−4] More recently, data analytic procedures such as curve and regression fitting have been used to extract quantitative lipoprotein and metabolic data from 1-dimensional blood plasma and urine spectra.[5−7] The mainstay of most biofluid NMR studies is the standard 5 mm outer diameter liquid state NMR tubes (with 5 mm NMR probes) used in the vast majority of all high-resolution chemical and biochemical experiments. A range of tube sizes is also widely available, for example 1 mm to 10 mm for high resolution studies and 15 mm 25 mm tubes for wide-bore NMR applications. Each tube size has been optimized for a particular sample volume range and there are sensitivity consequences for using smaller sample volumes at the same sample concentration because the absolute analyte mass in the probe is reduced proportionately.

Of note, the signal-to-noise ratio (S/N) in an NMR experiment is proportional to the absolute analyte mass in the receiver coil; halving this will require a quadruple number of scans to obtain the same S/N in an identical probe with identical

Received: October 14, 2020
Published: January 23, 2021



ACS Publications

© 2021 The Authors. Published by American Chemical Society

*Appendix I : Published Paper V*

# Systemic Perturbations in Amine and Kynurenine Metabolism Associated with Acute SARS-CoV-2 Infection and Inflammatory Cytokine Responses

in
Journal of Proteome Research 2021, 20, 2796-2811

Journal of proteome research

pubs.acs.org/jpr

Article

# Systemic Perturbations in Amine and Kynurenine Metabolism Associated with Acute SARS-CoV-2 Infection and Inflammatory Cytokine Responses

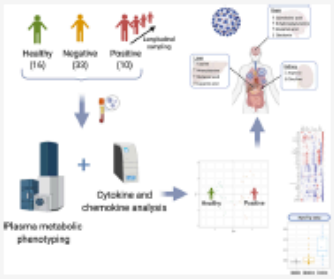
Nathan G. Lawler,* Nicola Gray,* Torben Kimhofer, Berin Boughton, Melvin Gay, Rongchang Yang, Aude-Claire Morillon, Sung-Tong Chin, Monique Ryan, Sofina Begum, Sze How Bong, Jerome D. Coudert, Dale Edgar, Edward Raby, Sven Pettersson, Toby Richards, Elaine Holmes, Luke Whiley,* and Jeremy K. Nicholson*

Cite This: J. Proteome Res. 2021, 20, 2796−2811

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

**ABSTRACT:** We performed quantitative metabolic phenotyping of blood plasma in parallel with cytokine/chemokine analysis from participants who were either SARS-CoV-2 (+) ($n$ = 10) or SARS-CoV-2 (-) ($n$ = 49). SARS-CoV-2 positivity was associated with a unique metabolic phenotype and demonstrated a complex systemic response to infection, including severe perturbations in amino acid and kynurenine metabolic pathways. Nine metabolites were elevated in plasma and strongly associated with infection (quinolinic acid, glutamic acid, nicotinic acid, aspartic acid, neopterin, kynurenine, phenylalanine, 3-hydroxykynurenine, and taurine; $p < 0.05$), while four metabolites were lower in infection (tryptophan, histidine, indole-3-acetic acid, and citrulline; $p < 0.05$). This signature supports a systemic metabolic phenoconversion following infection, indicating possible neurotoxicity and neurological disruption (elevations of 3-hydroxykynurenine and quinolinic acid) and liver dysfunction (reduction in Fischer's ratio and elevation of taurine). Finally, we report correlations between the key metabolite changes observed in the disease with concentrations of proinflammatory cytokines and chemokines showing strong immunometabolic disorder in response to SARS-CoV-2 infection.




**KEYWORDS:** *biogenic amines, tryptophan, kynurenine, cytokines, host response, SARS-CoV-2, COVID-19, metabolic phenotyping, phenoconversion*

## 1. INTRODUCTION

In the current COVID-19 pandemic, over 118 million people have been diagnosed with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection and over 2.7 million people have died to date[1]. The manifestation of SARS-CoV-2 infection ranges from asymptomatic to severe respiratory illness, sometimes with a variety of complications. The virus has been reported to cause lymphopenia and systemic hyperinflammatory responses that result in lung,[2,3] heart,[4] liver,[5,6] and kidney damage[7] and can present in the form of neurological dysfunction,[8−10] diabetes,[11−13] and stroke.[14]

While there has been a rapid response from the scientific community resulting in a surge of SARS-CoV-2/COVID-19 focused research, understanding the exact viral pathogenesis, metabolic interactions, and mechanisms that underpin an individual's disease trajectory remains poor. Understanding sequential mechanisms of SARS-CoV-2 action is critical to developing actionable therapeutic interventions to mitigate adverse outcomes following infection. Furthermore, elucidation of the metabolic host responses will enable the stratification of individuals who are poor responders to infection with resulting poor prognosis. Predicting prognosis post-infection will allow for earlier intervention and increase the potential for personalized treatments and improved recovery rates across the population. As with previous human severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS), there is concern that some of the complex expressions of the disease may have long-term consequences, where the criteria for "recovery" are still ill-defined[15] and are based mainly on reduction in respiratory symptoms. However, detection of systemic biomarkers related to multiple systems gives the potential for measuring systemic recovery as a well as acute illness.

Received: January 20, 2021
Published: March 16, 2021



ACS Publications

© 2021 The Authors. Published by American Chemical Society

*Appendix I : Published Paper VI*

# Quantitative In-Vitro Diagnostic NMR Spectroscopy for Lipoprotein and Metabolite Measurements in Plasma and Serum: Recommendations for Analytical Artifact Minimization with Special Reference to COVID-19/SARS-CoV-2 Samples

in
Journal of Proteome Research 2020, 19, 4428-4441

Made available through a Creative Commons CC-BY-NC-ND License

Journal of proteome research

pubs.acs.org/jpr

Article

# Quantitative In-Vitro Diagnostic NMR Spectroscopy for Lipoprotein and Metabolite Measurements in Plasma and Serum: Recommendations for Analytical Artifact Minimization with Special Reference to COVID-19/SARS-CoV-2 Samples

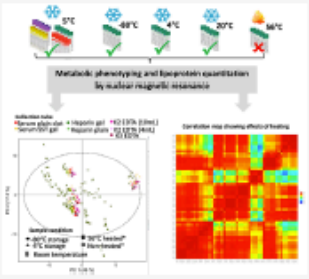
Ruey Leng Loo,[#] Samantha Lodge,[#] Torben Kimhofer, Sze-How Bong, Sofina Begum, Luke Whiley, Nicola Gray, John C. Lindon, Philipp Nitschke, Nathan G. Lawler, Hartmut Schäfer, Manfred Spraul, Toby Richards, Jeremy K. Nicholson,* and Elaine Holmes*

Cite This: J. Proteome Res. 2020, 19, 4428−4441

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

**ABSTRACT:** Quantitative nuclear magnetic resonance (NMR) spectroscopy of blood plasma is widely used to investigate perturbed metabolic processes in human diseases. The reliability of biochemical data derived from these measurements is dependent on the quality of the sample collection and exact preparation and analysis protocols. Here, we describe systematically, the impact of variations in sample collection and preparation on information recovery from quantitative proton ($^1$H) NMR spectroscopy of human blood plasma and serum. The effects of variation of blood collection tube sizes and preservatives, successive freeze−thaw cycles, sample storage at −80 °C, and short-term storage at 4 and 20 °C on the quantitative lipoprotein and metabolite patterns were investigated. Storage of plasma samples at 4 °C for up to 48 h, freezing at −80 °C and blood sample collection tube choice have few and minor effects on quantitative lipoprotein profiles, and even storage at 4 °C for up to 168 h caused little information loss. In contrast, the impact of heat-treatment (56 °C for 30 min), which has been used for inactivation of SARS-CoV-2 and other viruses, that may be required prior to analytical measurements in low level biosecurity facilities induced marked changes in both lipoprotein and low molecular weight metabolite profiles. It was conclusively demonstrated that this heat inactivation procedure degrades lipoproteins and changes metabolic information in complex ways. Plasma from control individuals and SARS-CoV-2 infected patients are differentially altered resulting in the creation of artifactual pseudo-biomarkers and destruction of real biomarkers to the extent that data from heat-treated samples are largely uninterpretable. We also present several simple blood sample handling recommendations for optimal NMR-based biomarker discovery investigations in SARS CoV-2 studies and general clinical biomarker research.



**KEYWORDS:** *lipoproteins, quantitative NMR spectroscopy, metabolic profiling, freeze−thaw, sample preparation, sample storage, information stability COVID-19, SARS-CoV-2, sample heat treatment*

## INTRODUCTION

Advances in clinical diagnosis, prognosis and possible treatment selection are increasingly driven by the use of molecular phenotyping and bioinformatic tools to aid classification of diseases subtypes and to define underlying individual variations in patient biology. Nuclear magnetic resonance (NMR)-based metabolic profiling of biofluids and tissues and linked data modeling offer possibilities for deep patient phenotyping by generation of detailed chemical fingerprints that can be related to clinical end points and response to therapeutic interventions.[1−3] $^1$H NMR-spectroscopy of biofluids including blood plasma and serum have been widely used to investigate systemic biochemistry in health and disease for decades.[1,4−6] Plasma NMR studies, in particular, have proved useful in diabetes,[5,7] obesity,[8,9] rheumatoid arthritis,[10] cancers[11] infectious diseases[12] rare diseases,[13,14] and neurological conditions.[15,16] Despite previous attempts at method and data harmonization, metabolic phenotyping still does not have uniformly applied standards as the analytical platforms and methods used are often locally optimized for each laboratory or research group. Although frameworks for standardizing measurement reporting have been proposed,[17,18] they do not result in directly statistically comparable data across the community, except perhaps for

Special Issue: Proteomics in Pandemic Disease

Received: July 15, 2020

Published: August 27, 2020



ACS Publications © 2020 American Chemical Society



https://dx.doi.org/10.1021/acs.jproteome.0c00537
J. Proteome Res. 2020, 19, 4428−4441

*Appendix I : Published Paper VII*

# Incomplete Systemic Recovery and Metabolic Phenoreversion in Post-Acute-Phase Nonhospitalized COVID-19 Patients: Implications for Assessment of Post-Acute COVID-19 Syndrome

in
Journal of Proteome Research 2021, XX, XXXX-XXXX

Journal of proteome research

pubs.acs.org/jpr

Article

# Incomplete Systemic Recovery and Metabolic Phenoreversion in Post-Acute-Phase Nonhospitalized COVID-19 Patients: Implications for Assessment of Post-Acute COVID-19 Syndrome

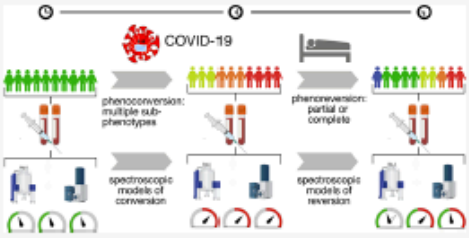
Elaine Holmes,* Julien Wist,* Reika Masuda, Samantha Lodge, Philipp Nitschke, Torben Kimhofer, Ruey Leng Loo, Sofina Begum, Berin Boughton, Rongchang Yang, Aude-Claire Morillon, Sung-Tong Chin, Drew Hall, Monique Ryan, Sze-How Bong, Melvin Gay, Dale W. Edgar, John C. Lindon, Toby Richards, Bu B. Yeap, Sven Pettersson, Manfred Spraul, Hartmut Schaefer, Nathan G. Lawler, Nicola Gray, Luke Whiley, and Jeremy K. Nicholson*

Cite This: https://doi.org/10.1021/acs.jproteome.1c00224

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

**ABSTRACT:** We present a multivariate metabotyping approach to assess the functional recovery of nonhospitalized COVID-19 patients and the possible biochemical sequelae of "Post-Acute COVID-19 Syndrome", colloquially known as long-COVID. Blood samples were taken from patients ca. 3 months after acute COVID-19 infection with further assessment of symptoms at 6 months. Some 57% of the patients had one or more persistent symptoms including respiratory-related symptoms like cough, dyspnea, and rhinorrhea or other nonrespiratory symptoms including chronic fatigue, anosmia, myalgia, or joint pain. Plasma samples were quantitatively analyzed for lipoproteins, glycoproteins, amino acids, biogenic amines, and tryptophan pathway intermediates using Nuclear Magnetic Resonance (NMR) spectroscopy and mass spectrometry. Metabolic data for the follow-up patients ($n = 27$) were compared with controls ($n = 41$) and hospitalized severe acute respiratory syndrome SARS-CoV-2 positive patients ($n = 18$, with multiple time-points). Univariate and multivariate statistics revealed variable patterns of functional recovery with many patients exhibiting residual COVID-19 biomarker signatures. Several parameters were persistently perturbed, e.g., elevated taurine ($p = 3.6 \times 10^{-3}$ versus controls) and reduced glutamine/glutamate ratio ($p = 6.95 \times 10^{-8}$ versus controls), indicative of possible liver and muscle damage and a high energy demand linked to more generalized tissue repair or immune function. Some parameters showed near-complete normalization, e.g., the plasma apolipoprotein B100/A1 ratio was similar to that of healthy controls but significantly lower ($p = 4.2 \times 10^{-3}$) than post-acute COVID-19 patients, reflecting partial reversion of the metabolic phenotype (phenoreversion) toward the healthy metabolic state. Plasma neopterin was normalized in all follow-up patients, indicative of a reduction in the adaptive immune activity that has been previously detected in active SARS-CoV-2 infection. Other systemic inflammatory biomarkers such as GlycA and the kynurenine/tryptophan ratio remained elevated in some, but not all, patients. Correlation analysis, principal component analysis (PCA), and orthogonal-partial least-squares discriminant analysis (O-PLS-DA) showed that the follow-up patients were, as a group, metabolically distinct from controls and partially comapped with the acute-phase patients. Significant systematic metabolic differences between asymptomatic and symptomatic follow-up patients were also observed for multiple metabolites. The overall metabolic variance of the symptomatic patients was significantly greater than that of nonsymptomatic patients for multiple parameters ($\chi^2$ $p = 0.014$). Thus, asymptomatic follow-up patients including those with post-acute COVID-19 Syndrome displayed a spectrum of multiple persistent biochemical pathophysiology, suggesting that the metabolic phenotyping approach may be deployed for multisystem functional assessment of individual post-acute COVID-19 patients.




**KEYWORDS:** *COVID-19, post-acute COVID-19 syndrome, long-COVID syndrome, SARS-CoV-2, plasma, phenoconversion,*
continued...

**Received:** March 19, 2021

ACS Publications

© XXXX The Authors. Published by American Chemical Society

*Appendix II : Paper I UNDER REVIEW*

# Metabolic and Bacterial Phenotypes Associated with Inadequate Response to Ursodeoxycholic Acid in Primary Biliary Cholangitis

in
Hepatology

**Metabolic and bacterial phenotypes associated with inadequate response to ursodeoxycholic acid in primary biliary cholangitis**

Laura Martinez-Gili[1], Alexandros Pechlivanis[1,2], Julie A.K. McDonald[3], Sofina Begum[3], Jonathan Badrock[4], Jessica K. Dyson[5,6], Rebecca Jones[7], Gideon Hirschfield[8], Stephen D. Ryder[9], Richard Sandford[4], Simon Rushbrook[10], Douglas Thorburn[11], Simon D. Taylor-Robinson[12], Mary M.E. Crossey[3], Julian R. Marchesi[3], George Mells[4,13], Elaine Holmes[3,14], David Jones[5,6]

1. Division of Systems Medicine, Department of Metabolism, Digestion and Reproduction, Faculty of Medicine, Imperial College London, London, UK.
2. Biomic_Auth, Bioanalysis and Omics Laboratory, Center for Interdisciplinary Research and Innovation (CIRI-AUTH), Balkan Centre, Thessaloniki, Greece.
3. Division of Digestive Diseases, Department of Metabolism, Digestion and Reproduction, Faculty of Medicine, Imperial College London, London, UK.
4. Academic Department of Medical Genetics, Cambridge University, Cambridge, UK.
5. Liver Unit, Freeman Hospital, Newcastle Upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK.
6. Institute of Translational and Clinical Research, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK.
7. Leeds Liver Unit, St James's University Hospital, Leeds, UK.
8. Centre for Liver and Gastroenterology Research and National Institute for Health Research (NIHR) Birmingham Biomedical Research Centre, Institute of Immunology and Immunotherapy, University of Birmingham, Birmingham, UK.
9. NIHR Biomedical Research Centre at Nottingham University Hospitals NHS Trust and the University of Nottingham, Nottingham, UK.
10. Department of Gastroenterology, Norfolk and Norwich University Hospital, Norwich, UK.
11. University College London Institute of Liver and Digestive Health, UCL Royal Free Campus, Royal Free Hospital, London, UK.
12. Department of Electrical and Electronic Engineering, Faculty of Engineering, Imperial College London, UK.
13. Department of Hepatology, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK.
14. Centre for Computational & Systems Medicine, Health Futures Institute, Murdoch University, Western Australia, 6150, Australia.

**Abstract:**
**BACKGROUND & AIMS:** Primary biliary cholangitis (PBC) is an autoimmune cholestatic liver disease with ursodeoxycholic acid (UDCA) as first-line treatment. An incomplete treatment response means higher risk of progressing to cirrhosis. The mechanism for UDCA non-response is unclear. We characterized the bile acid (BA) and bacterial profiles of PBC patients treated with UDCA to assess whether a different treatment response associates with a different phenotype.
**APPROACH & RESULTS:** Patients from the UK-PBC cohort (n= 454) were treated with UDCA for a minimum of 12-months and their response to therapy assessed. BA from serum, urine and faeces were analysed using Ultra-High-Performance Chromatography-Mass Spectrometry and faecal bacterial composition measured using 16S rRNA gene sequencing. We found distinct stool BA signatures in UDCA non-responders (n= 214), complete-responders (responders with a good prognosis; n= 224) and a newly identified group of partial-responders (responders with a bad prognosis; n= 16), with the two responder sub-groups showing opposite trends. BA in urine were higher in non-responders, and only 12-dehydrocholate was increased in urine of complete-responders. Differences in bacterial composition were found in confounder-matched sub-analyses after adjusting for bacterial load; complete-responders had lower abundance of *Oscillospirales* and a predicted increase in a gene of the 7-dehydroxylation pathway. Partial-responders had differences in bacteria involved in BA deconjugation including *Bacteroides* and *Parabacteroides*.
**CONCLUSION:** UDCA response failure was associated with a specific bacterial-derived secondary bile acid signature. Modest differences in bacterial taxonomy indicate that the phenotype of non-responders does not rely on specific microorganisms but rather on a decreased capacity to generate and excrete oxo- and epimerised secondary bile acids.

*Appendix III : Paper I UNDER REVIEW*

# Systemic Long-Term Metabolic Effects of Acute Non-Severe Paediatric Burn Injury

in
Scientific Reports

**Systemic long-term metabolic effects of acute non-severe paediatric burn injury**

**Sofina Begum[1,2], Blair Z. Johnson[3], Aude-Claire Morillon[2], Rongchang Yang[2], Sze How Bong[2], Luke Whiley[2,4,5], Nicola Gray[2,5], Vanessa S. Fear[6], Jeremy K. Nicholson[2,5,7,8], Elaine Holmes[1,2,7], Fiona M. Wood[3,9] and Mark W. Fear[3]**

1 Department of Metabolism, Digestion and Reproduction, Faculty of Medicine, Imperial College London, Sir Alexander Fleming Building, South Kensington, London, SW7 2AZ, United Kingdom.

2 Australian National Phenome Centre, Computational and Systems Medicine, Health Futures Institute, Murdoch University, Harry Perkins Building, Perth, Western Australia, 6150, Australia.

3 School of Biomedical Sciences, The University of Western Australia, Perth, WA, Australia

4 Perron Institute for Neurological and Translational Science, Nedlands, Western Australia, 6009, Australia.

5 Centre for Computational and Systems Medicine, Health Futures Institute, Murdoch University, Harry Perkins Building, Perth, Western Australia, 6150, Australia.

6 Translational Genetics, Telethon Kids Institute, Perth, Western Australia, Australia

7 Medical School, Faculty of Health and Medical Sciences, University of Western Australia, and Department of Endocrinology and Diabetes, Fiona Stanley Hospital, Harry Perkins Building, Murdoch, Perth, Western Australia, 6150, Australia.

8 Institute of Global Health Innovation, Imperial College London, Level 1, Faculty Building South Kensington Campus, London SW7 2AZ, United Kingdom.

9 Burns Service of Western Australia, WA Department of Health, Perth, Western Australia, 6150, Australia.

**Summary**

A growing body of evidence supports the concept that a global systemic response to thermal trauma extends beyond the site of injury. This provokes an immunosuppressed state that predisposes paediatric patients to poor recovery and chronic systemic effects. Targeted mass spectroscopy metabolic profiling of biogenic amines, tryptophan metabolites and multiplex cytokine assays were performed on serum collected from burn injury patients three years post injury; with age and sex matched healthy controls. A panel of metabolites including urea cycle intermediates, aromatic amino acids and quinolinic acid were present in higher concentrations in children with previous burns. 12 out of 46 measured metabolites (pvalue<0.01) were found to be significantly elevated in children with previous burn injuries. Correlation analysis indicated the presence of multiple cytokine-metabolite associations in the prior burn injury participants that were absent from the controls. These data suggest that there is a metabolic imprint of even non-severe burn trauma, linked to long-term immune dysfunction that may underpin poor long-term health outcomes observed.

**TO THE EDITOR**

Paediatric burn injury remains the most severe of childhood traumas; not only for the acute impact of the injury itself but also the subsequent increased risk of long-term morbidities. This increased risk of morbidity encompasses a range of conditions including cardiovascular disease, diabetes, mental health disorders and even elevated age-adjusted mortality (Duke et al., 2015). Long-term systemic change has previously been reported for severe childhood burns (Jeschke et al., 2011), but increasing evidence suggests even non-severe burns may have long-term effects (Duke et al., 2015). This is significant as in many developed countries non-severe injury accounts for over 90% of paediatric burns (Duke et al., 2011).

In an effort to explore the underlying mechanisms by which non-severe paediatric burns increase the risk of chronic disease, paediatric patients (n=33) aged 0-4 years at the time of original injury were recruited from the Western Australia Paediatric Burns Unit. A blood sample was drawn from each participant 3 years post-burn injury. A cohort of healthy age and sex matched children who had not undergone burn trauma (n=22) were recruited as a control group. All plasma samples were obtained following informed consent of a parent or guardian (ethical approval from the Child and Adolescent Health Service WA (2015219EP; 1111EP;768EP)). All burn injury patients and healthy controls had no prior pre-existing illness and were not receiving treatment or medical interventions at the time of sampling.

We applied a metabolic phenotyping approach (Nicholson et al., 2012) to investigate inherent differences in metabolism between the children with and without prior burns. Validated targeted UPLC-MS (Ultra Performance Liquid Chromatography- Mass Spectrometry) assays were used to quantify 46 biogenic amines, amino acids and tryptophan pathway metabolites (Gray et al., 2021, Whiley et al., 2019) in plasma (methodological details in supplementary material). The burn-injury cohort had significantly different plasma metabolite profiles compared to the non-burn control counterparts, based on an Orthogonal Partial Least-Squares Discriminant Analysis (OPLS-DA) model ($R^2X$=0.40, $Q^2Y$=0.19; figure 1A) with some evidence of differentiation based on the principal component analysis model (figure S1A). The OPLS-DA model indicated that 22

of the 46 measured metabolites contributed to the differential metabolite pattern associated with burn injury based on the variable importance score plot (figure S3 & table S1), all of which were elevated in the burn injury group. Of these metabolites, 12 remained statistically significant between burn injury and healthy groups after Wilcoxon rank sum testing with Bonferroni correction (figure 1B, table S1). Key metabolites included alanine, 3-methylhistidine, asparagine, ornithine, arginine, tyrosine, phenylalanine, cystine, lysine, proline, serine and quinolinic acid. Branched chain amino acids were influential in the OPLS-DA model but did not retain significance after Bonferroni correction.

Consistent with these observations, a significant increase in the turnover of arginine, ornithine, proline and leucine has been reported after burn injury (Yu et al., 2001) and appear to be specific to burn trauma. Alanine and increased levels of plasma aromatic amino acids have also been associated with burn injury but were also elevated after other trauma, suggesting these changes are associated with general stress (Alkhalil et al., 2020). Increased serum concentrations of aromatic amino acids, alanine and lysine have been associated with liver disease (Trico et al., 2021). The Fischer's ratio (branched-chain amino acids to aromatic amino acids) is used as a diagnostic marker for assessing liver metabolism, hepatic functional reserve and the severity of liver dysfunction, with a decreased ratio indicative of dysfunction (Fischer et al., 1975). Here we saw a trend towards reduced Fischer's ratio in the burn injury cohort. Whilst non-significant, the trend is consistent with the previously reported prolonged insult to hepatic function following severe burn injury, associated with a profound hypermetabolic and hyperinflammatory response (Jeschke et al., 2011). Our results are consistent with other studies that have shown similarly long "memory" in the liver and muscle tissues following severe burn injury with lysine, urea cycle and tryptophan metabolism being amongst the most disrupted pathways.

Disturbances in the tryptophan pathway have previously been implicated in severe burn injury in children, with increased kynurenine observed in urine after injury patients (Barlow et al., 1972). This pathway has potentially toxic downstream metabolites, most notably quinolinic acid which is a known neurotoxin and immunosuppressant that has been reported to be elevated in juvenile idiopathic inflammatory myopathy (Rider et al., 2002) where plasma levels were significantly predictive of disease activity. It has also been

previously found to upregulate chemokine production (Guillemin et al., 2003). Interestingly, a subset of the burn-injury cohort (figure S1) with the most abnormal plasma profiles had the highest quinolinic acid concentrations, suggesting further interrogation of how this may impact on long-term health is important.

To interrogate the association between the metabolic and immune changes post-burn correlational analysis was used with previously reported cytokine data from this cohort (Johnson et al., 2020). Hierarchical clustering defined 5 metabolic and 4 cytokine clusters (figure 2A). Metabolite-metabolite and immune-metabolite correlations were generally weak in the control group, with lower concentrations of cytokines in the control group. In contrast strong positive correlation is seen in burn injury patients across clusters 2, 4 and 5(figure 2A). The control group (figure 2B) demonstrates negative correlations between cytokines and metabolites in cluster 1 (containing branched chain and aromatic amino acids), in contrast to positive associations in the burns group. This suggests disruption in liver metabolism post burn injury and highlights the correlation between liver and immune function (further highlighted in figure S4). Interactions between metabolites and cytokines in the burn injury cohort were emphasised through hierarchical edge clustering, linking significant metabolites (pvalue <0.05) with cytokines (figure 2C, 2D), highlighting the difference in immune-metabolite interactions between cohorts. Whereas no metabolite-cytokine correlations above (r=0.8) existed in the control group, the burn-injury group showed a range of correlations including those between the metabolite hubs serine and tyrosine and IFN-γ, GM-CSF, IL-12. This suggests that the perturbation in metabolism post-injury is linked to the changes to immune function.

In summary, these data indicate that a “metabolic memory” of non-severe burn injury persists in children for over three years and that this may be linked to observed immune dysfunction and long-term morbidity.