EXPLORING METABOLIC AND FUNCTIONAL CHANGES IN STROKE PATIENTS: INSIGHTS FROM A URINARY AND BLOOD-DERIVED METABOLOMIC STUDY

JAMIE NICHOLE PETERSSON
Bachelor of Science, **University of Lethbridge, 2020**

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

MASTER OF SCIENCE

in

NEUROSCIENCE

Department of Neuroscience
University of Lethbridge
LETHBRIDGE, ALBERTA, CANADA

EXPLORING METABOLIC AND FUNCTIONAL CHANGES IN STROKE PATIENTS: INSIGHTS FROM A URINARY AND BLOOD-DERIVED METABOLOMIC STUDY

JAMIE PETERSSON

Date of Defense: August 3, 2023

Dr. Gerlinde Metz	Professor	Ph.D.
Tony Montina	Instructor	M.Sc.
Thesis Co-Supervisors		
Dr. Chantel Debert	Assistant Professor	M.D./M.Sc.
Thesis Examination Committee Member		
Dr. Chelsea Ekstrand	Professor	Ph.D.
Thesis Examination Committee Member		
Dr. Robbin Gibb	Professor	Ph.D.
Chair, Thesis Examination Committee		

DEDICATION

I would like to dedicate my MSc. Thesis to my ever-loving father, who always supported my dreams and aspirations, even if he never fully understood them.

ABSTRACT

The neuropathological sequelae of stroke and subsequent process of recovery are incompletely understood. The manuscript-based thesis presented herein used a clinical pilot population to elucidate changes in cellular metabolic dynamics following stoke, thereby contributing to the existing knowledge base on the pathophysiological mechanisms orchestrating stroke recovery. The investigation explores these biochemical profiles via a nuclear magnetic resonance (NMR)-driven metabolomic profiling approach. The objective of this research was to determine if this powerful technique can (1) uncover the biochemical pathways underlying stroke repair and functional recovery and (2) yield novel biomarkers indicative of diagnosis, prognosis, and treatment efficacy subsequent to stroke. Urine samples and serum samples along with clinical stroke assessments were collected during the acute phase of stroke and the chronic phase, six months later. Employing a 700MHz ^{1}H NMR spectrometer, the metabolomic profiles of the patients were acquired. Following NMR post-processing, a combination of univariate and multivariate statistical analysis, along with biological pathway analysis was conducted. Lastly, clinical correlations illustrated the relationship between significant metabolite concentrations and clinical measures. Ultimately, NMR-based metabolomics provided valuable insights into post-stroke cellular functions and established a foundational framework for future investigations to develop standardized clinical assays, advance personalized neurorehabilitation regimens, and enhance overall quality of life for stroke-afflicted individuals.

PREFACE

The contributions of authors, funding, and review board statement for the following manuscript-based thesis are as follows:

Jamie Petersson is the primary author of the statistics overview and the 2 studies within this thesis. Elani Bykowski contributed to these studies during sample preparation, data acquisition, and statistical analysis. Tony Montina and Dr. Gerlinde Metz provided revisions. Dr. Chantel Debert's contributions included study design and sample collection. Dr. Chelsea Ekstrand assisted with the development of the linear regression model.

The author would like to thank Eric Paxman, Regan King, Janis Yajure, Mark Piitz, Sean Dukelow, and Chester Ho for assisting with recruitment, collection of samples and clinical measures, and the conceptualization of this project.

This research was funded by The Hotchkiss Brain Institute at the University of Calgary (Chantel Debert, Tony Montina, and Gerlinde Metz), CIHR Project Scheme, grant #363195 (G.M.) and NSERC Discovery Grant, grant #05628 (G.M.).

The studies herein were conducted in accordance with the Declaration of Helsinki. The present protocols were reviewed and approved by the University of Calgary Conjoint Health Research Ethics Board (CHREB; Ethics ID: REB14-1017, approved 29 January 2015) and the University of Lethbridge Human Participant Research Committee (HSRC; Protocol #2015-061, approved 18 August 2015) in accordance to the standards set forth by the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans.

ACKNOWLEDGEMENTS

First, I would like to thank Elani Bykowski for her guidance and assistance with sample preparation and analysis. I extend my gratitude to my supervisors, Dr. Gerlinde Metz and Tony Montina, for their patience, mentorship, and unwavering support throughout my thesis and the many years prior working under them. I also acknowledge Michael Opyr and Vincent Weiler for their technical expertise. I would like to thank all the students in the Metz/Montina lab during my research at the University of Lethbridge. Lastly, I am grateful for the endless loving encouragement of my loving partner, Hayes, and my entire family.

TABLE OF CONTENTS

Dedication iii

Abstract iv

Preface v

Acknowledgements vi

Table of Contents vii

List of Tables viii

List of Figures ix

List of Abbreviations x

Chapter 1: Introduction 1

Chapter 2: From Raw Data to Metabolite Discoveries: Pre-Processing, Statistical Analysis, and Clinical Interpretation 19

Chapter 3: Unraveling Metabolic Changes Following Stroke: Insights from a Urinary Metabolomics Pilot Analysis 39

Chapter 4: Pilot Study Reveals Candidate Metabolic Biomarkers and Pathways in Stroke Patients 81

Chapter 5: Conclusions and Future Directions 134

LIST OF TABLES

Table 1. Patient Characteristics for Stroke Urine Study. 64
Table 2. Significant Metabolites for Stroke Urine Study..... 67
Table 3. Significant Spearman Correlations for Stroke Urine Study..... 71
Table 4. Patient Characteristics for Stroke Blood Study 110
Table 5. Significant Metabolites for Stroke Blood Study. 113
Table 6. Significant Spearman Correlations for Stroke Urine Study..... 118

LIST OF FIGURES

Figure 1. Research Workflow for Urine & Blood Studies..... 34
Figure 2. Orthogonal Projections to Latent Structures Discriminant Analysis for Stroke Urine Study. 68
Figure 3. Receiver Operator Characteristic Curve for Stroke Urine Study. 69
Figure 4. Pathway Topology Analysis for Stroke Urine Study 70
Figure 5. Spearman Correlation Scatterplots for Stroke Urine Study..... 72
Figure 6. Orthogonal Projections to Latent Structures Discriminant Analysis for Stroke Blood Study. 115
Figure 7. Receiver Operator Characteristic Curve for Stroke Blood Study..... 116
Figure 8. Pathway Topology Analysis for Stroke Blood Study..... 117
Figure 9. Spearman Correlation Scatterplots for Stroke Blood Study. 120

LIST OF ABBREVIATIONS

2-HB	2-Hydroxybutyric Acid
ATP	Adenosine 5′-Triphosphate
AUC	Area-Under-the-Curve
AUROC	Area Under the Receiver Operating Characteristic
BCAA	Branched-Chain Amino Acids
BCKA	Branched-Chain α-Keto Acids
BMI	Body Mass Index
BMR	Binary Matrix Resampling
CD206	Cluster of Differentiation 206
CI	Confidence Interval
CMSA	Chedoke-McMaster Stroke Assessment
CNS	Central Nervous System
CSF	Cerebrospinal Fluid
CT	Computed Tomography
CV	Cross-Validation
DAB	Dynamic Adaptive Binning
DNA	Deoxyribonucleic Acid
DV	Dependent Variable
FIM	Functional Independence Measure
GC-MS	Gas Chromatography-Mass Spectrometry
GLM	General Linear Model
HMDB	Human Metabolome Databse
HPLC	High-Performance Liquid Chromatography
HVA	Homovanillate
IV	Independent Variable
KEGG	Kyoto Encyclopedia of Genes and Genomes
KMV	3-Methyl-2-Oxovaleric Acid
LC-MS	Liquid Chromatography-Mass Spectrometry
LMM	Linear Mixed Model
MBL	Mannose-Binding Lectin
MMR	Macrophage Mannose Receptor
MRI	Magnetic Resonance Imaging
MS	Mass Spectroscopy
MSM	Dimethyl Sulfone
MSUD	Maple Syrup Urine Disease
NHISS	National Institutes of Health Stroke Scale
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser Effect Spectroscopy
RNA	Ribonucleic Acid

RNS	Reactive Nitrogen Species
ROC	Receiver Operator Characteristic
ROS	Reactive Oxygen Species
RSPA	Recursive Segment-Wise Peak Alignment
ST	Stroke
tPA	Tissue Plasminogen Activator
PC	Principal Component
PCA	Principal Component Analysis
PLS	Partial Least Squares
PLS-DA	Partial Least Squares Discriminant Analysis
PLS-LDA	Partial Least Squares-Linear Discriminant Analysis
OLS	Ordinary Least Squares
OPLS-DA	Orthogonal Projections to Latent Structures Discriminant Analysis
VIAVC	Variable Importance Analysis Based on Random Variable Combination
TBO	Triple Resonance Broadband
TSP	Trimethylsilyl Propanoic Acid
UCAN	University of Calgary Advising Network
UV	Unit Variance
VIP	Variable Importance in Projection
VMA	4-Hydroxy-3-Methoxymandelate

CHAPTER 1: INTRODUCTION

1.1 STROKE

Stroke is a medical condition in which a disturbance in vasculature causes damage to the central nervous system (CNS), resulting in neurological deficits. The main types of stroke are generally categorized as either ischemic or hemorrhagic. The most common type of stroke is caused by ischemia due to the obstruction of blood vessels, such as by a clot, reducing blood flow to specific parts of the brain. Hemorrhagic stroke is caused by bleeding and often results from a blood vessel leak or a ruptured aneurysm. More specifically, stroke caused by intracerebral hemorrhage occurs when blood collects in the brain parenchyma or ventricular system and is not due to trauma, whereas stroke caused by subarachnoid hemorrhage is caused by blood leaking into the subarachnoid space and is not caused by trauma. Whether due to the occlusion of blood vessels or blood leakage, stroke results in cell death at the affected area and may extend around the stroke site (Sacco et al., 2013).

Stroke is the second leading cause of death and a leading contributor to disability worldwide (Roth et al., 2020). Hence, it is of paramount importance to identify the stroke type to both treat stroke promptly and minimize brain damage, as well as selecting rehabilitation regimens to minimize disability. Depending on stroke location, severity, and other individual differences, the effects of stroke can vary greatly between individuals. A stroke can cause loss or impairments of speech, memory, vision, and limb movements contralateral to the side of cerebral infarction (Langhorne et al., 2011). Rehabilitation is currently the primary approach to promote recovery of these disabilities in stroke patients. This includes motor therapy, cognitive rehabilitation, speech therapy, and visual therapy (Langhorne et al., 2011). Without proper rehabilitation methods, the stroke patient may suffer from long-term post-stroke disabilities that could have been prevented.

Nevertheless, recovery from stroke is usually incomplete and 40% of patients suffer from moderate to severe impairments and another 10% are left with severe permanent disabilities (Ontario Stroke Network, n.d.).

Multiple clinical assessments can be used to quantify a patient's stroke severity, post-stroke physical impairments and disability. The present thesis, however, will focus on the following three assessments: The National Institutes of Health Stroke Scale (NIHSS), the Functional Independence Measure (FIM), and the Chedoke-McMaster Stroke Assessment (CMSA). The NIHSS (Brott et al., 1989) can be used to measure stroke severity and has been used to predict stroke patient outcomes (Muir et al., 1996). It assess the patient's level of consciousness, visual integrity, eye movements, facial movements, arm and leg movements, coordination, sensation, neglect, speech, and language (Kwah & Diong, 2014). The scores on the 15 items are recorded on a scale from 0-2, 0-3, or 0-4 and summed for a total possible score between 0-42, with a score of 42 indicating the most severe impairments (Kwah & Diong, 2014). The FIM (Keith et al., 1987) is used to measure the level of disability of the stroke patients by assessing the motor control, self-care, communication, and social and cognitive skills. The scores on 18 items are recorded on a scale from 1-7 and summed for a total possible score of 126, with a score of 126 corresponding to complete functional independence (Mackintosh, 2009). The CMSA (Gowland et al., 1993) measures physical impairments and disabilities on a 7-point scale with 1 being paralysis and 7 being normal movement. The CMSA has two inventories: the impairment inventory that classifies patients based on motor recovery, and the disability inventory that measures the patients change in disability (Gowland et al., 1993). The present study focuses on the impairment inventory, and more specifically, the hand and arm dimensions of the assessment.

Currently, imaging techniques such as computed tomography (CT) and magnetic resonance imaging (MRI) are central in confirming the incidence of stroke, however, no imaging

technique is able to advise on a patient's therapy or provide accurate prognosis (Sacco et al., 2013). CT angiogram is the most commonly used technique and a cornerstone for stroke diagnostic imaging because of its greater availability and relatively lower cost (Sacco et al., 2013). MRI is the preferred imaging technique for diagnosis because it has much higher diagnostic sensitivity. Nevertheless, MRI is less commonly used because it is costly and less widely available (Timpone et al., 2020). Both of these imaging methods are costly, have limited availability, and can take several hours longer than the time required for treatment. Since stroke needs to be treated rapidly, there is an urgent need for a better diagnostic system (Qureshi et al., 2017). It is important to note that NMR is also costly, has limited availability, and could take a comparable amount of time if implemented in clinical practice. The value of this research lies in the NMR-based metabolomic biomarkers ability to be translated into inexpensive standardized commercial assays that could even be used outside the hospital setting (Dagonnier et al., 2021). Moreover, rehabilitation from stroke has proven to be complicated due to the heterogeneity of stroke and its effects (Campbell & Khatri, 2020). It is imperative that patients receive a personalized treatment regimen to maximize their personal potential for recovery. One promising technique that has potential to aid in diagnosis and treatment monitoring is the use of fluid biomarkers determined through metabolomics.

1.2 STROKE METABOLOMICS

Stroke is associated with characteristic pathogenic and biochemical changes. During an ischemic stroke, neurons are deprived of oxygen and energy and are unable to maintain their normal homeostatic processes (Murphy et al., 2008). The ensuing acute responses further aggravate the injury, including oxidative and nitrative stress, excitotoxicity, inflammation, and apoptosis. During excitotoxicity, the excitatory neurotransmitter, glutamate, is released into the

extracellular space and exacerbates tissue damage. The ensuing rise in glutamate depolarizes the tissue surrounding the infarct, further increasing metabolic activity and complicating the recovery from the already depleted oxygen supply. Glutamate activates the nitric oxide synthase pathway and results in the production of free radicals, including reactive oxygen species (ROS) and reactive nitrogen species (RNS), and leading to oxidative stress (Khoshnam et al., 2017; Lai et al., 2014). Oxidative and nitrative stress occur when toxic free radicals, such as ROS and RNS, are produced in excess of normal levels. The overabundance of free radical causes deoxyribonucleic acid (DNA) damage, alters cellular signaling mechanisms, and contributes to apoptosis (Allen & Bayraktutan, 2009). This, along with the accumulation of lactic acid from energy depletion, initiates inflammation. The neuro-inflammatory response begins with recruitment of inflammatory cells, such as microglial cells, as the brain's response to protect neurons after injury. The microglia produce many protective substances, but their over-activation releases cytotoxic substances, such as ROS and nitric oxide and results in further inflammation (Khoshnam et al., 2017; Lucas et al., 2009). Other immune cells, such as astrocytes, along with inflammatory factors, including cytokines, also contribute to brain inflammation. While their initial response exacerbates the infarction, they later play an important role in recovery (Khoshnam et al., 2017; Swanson et al., 2004). For example, cytokines promote glial scar formation and signal for tissue remodeling and immune cells generate neuroprotective substances and engulf cellular debris (Doyle et al., 2008). Finally, apoptosis is triggered by mitochondrial signaling pathways and cell surface death receptors (Fuchs & Steller, 2015; Khoshnam et al., 2017). Opposing pathways can either activate or inhibit the activity of caspases, a family of cysteine proteases involved in apoptosis. Their activity leads to nuclear DNA damage, cell fragmentation, and ultimately cell death. The cells undergoing apoptosis have the ability to alter surrounding cells by activating cell proliferation or further cell death (Fuchs & Steller, 2015). In the penumbra around the ischemic core after stroke,

this process regulating cell survival and death determines the fate of each cell (Uzdensky, 2019). Thus, excitotoxicity, oxidative and nitrative stress, and inflammation all contribute to cell death during stroke and continue to have an impact during recovery.

Metabolomics is the systematic study of metabolic products, otherwise known as metabolites, in living systems (Nicholson et al., 1999). Examples of metabolites include sugars, hormones, alcohols, amino acids, and organic acids (Vignoli et al., 2019). These metabolites can be found in many different types of biological samples, but the present study focuses on metabolites in urine and blood. Human metabolism and its products are unique and dynamic and can change due to small alterations in the internal and external environment (Vignoli et al., 2019). Metabolomics offers a snapshot into the condition of multiple organs and the biological system as a whole (Nicholson et al., 1999). Additionally, metabolomics can provide valuable information about an individual's overall health because metabolites are downstream from other biological systems, such as the genome, the epigenome, the transcriptome, and the proteome. Since metabolism is perturbed by abnormal cellular processes, metabolomics provides an integrative view of biochemical processes occurring in an organism (Nicholson & Lindon, 2008). This powerful tool that can provide insights into biochemical fingerprints and cellular function after neurological injury. There are numerous neurochemical insults that accompany stroke and monitoring and quantifying metabolites after stroke has the potential to identify these neurochemical events. For these reasons, metabolomics is well equipped to detect and monitor neuropathophysiological states such as stroke.

Several techniques are used to study metabolomics due to the large diversity in metabolites. Due to its vastness, not one single method is sufficient to measure the entire range of the metabolome. The two main technologies used to quantify metabolites are nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) (Emwas, 2015). NMR is highly

reproducible, non-destructive to the sample, non-biased, requires little sample preparation and analysis (Emwas, 2015; Nicholson & Lindon, 2008; Wishart, 2008), and can identify a plethora of metabolites simultaneously involved in numerous biological pathways (Nicholson et al., 1999). However, one disadvantage of NMR compared to MS is its limited sensitivity. NMR and MS are able to quantify metabolites at the micromolar and nanomolar level, respectively. In addition, MS can detect ions that have no protons or carbon, including metal ions. The most common MS technique is liquid chromatography-mass spectrometry (LC-MS) and it combines high-performance liquid chromatography (HPLC) and MS. LC-MS is subject to limitations as well, including more laborious and time-consuming sample preparation and analysis, sample destruction, and less reliable identification. Gas chromatography-mass spectrometry (GC-MS) is another technique that combines the methods of gas chromatography with MS. It is advantageous because it is sensitive, highly reproducible, and has high separation efficiency. However, GC-MS is only able to detect volatile compounds and much like LC-MS, the sample preparation is time-consuming and complicated due to possible sample degradation and results in sample destruction. Using complementary MS methods, such as LC-MS and GC-MS together, can greatly increase the number of detectible metabolites. In comparing these techniques, GC-MS is able to detect 179 and 99 compounds in urine and blood, respectively, with 89 and 70 being unique to GC-MS, again respectively (Bouatra et al., 2013; Psychogios et al., 2011). Conversely, LC-MS can identify 127 and 98 compounds in blood and urine, respectively, with 98 and 78 being unique to LC-MS, again respectively (Bouatra et al., 2013; Psychogios et al., 2011). Lastly, NMR has the greatest number of unique (108) and detectable (209) metabolites in human urine when compared to other techniques (Bouatra et al., 2013), and detects additional unique (20) and detectable (49) compounds in serum (Psychogios et al., 2011). For the reasons outlined above, NMR has been selected as the analytical tool of choice for the present study.

1.3 NMR-BASED METABOLOMICS

In a biological sample, each individual metabolite that contains at least one hydrogen atom has a biochemical signature that is detectable by ^{1}H NMR. ^{1}H NMR detects hydrogen atoms in each metabolite to generate its individual "chemical shift" fingerprint, which is displayed as a single peak or multiple peaks on the NMR spectrum. Thus, the NMR spectrum as a whole is the composition of fingerprints that each molecule in the biological sample gives off (Nicholson & Lindon, 2008). Each individual compound is easily identifiable because no metabolite will have the same number of peaks, peak intensities, chemical shifts, spin couplings, or line shapes (Wishart 2008).

Metabolomics is amenable to a wide array of biological samples, including urine, feces, blood, saliva, tissue, and cerebrospinal fluid (CSF). In the present study, the focus will be on urine and blood. Urine is the amber-coloured liquid by-product of the kidneys, produced when filtering toxins from blood. Urine contains water soluble waste products, including electrolytes, proteins, hormones, and metabolites. Many of the compounds that are undetectable in blood are easily detectible in urine because the kidneys remove these wastes from the blood such that they are highly concentrated in urine. Given that, urine is a suitable biofluid for clinical diagnostics and patient monitoring (Bouatra et al., 2013). Further, analyzing both urine and blood is complementary and can provide a more comprehensive picture of the biological processes occurring in the body. Urine also requires minimal sample preparation and data collection and is fast and facile (Bouatra et al., 2013). Blood, however, requires filtration to remove the high molecular weight components (Wishart, 2008), making sample preparation slightly more laborious but still much less relative to MS based techniques. The present study focuses on serum, which is the supernatant fluid after the blood has clotted and it therefore lacks blood cells and clotting

proteins. As primary carrier of small molecules, such as nutrients and wastes, blood serum circulates through the entire body and can give insight into its physiological response to illness or injury (Psychogios et al., 2011). Metabolomic analysis of urine and blood is simple, rapid and non-invasive or minimally invasive (Vignoli et al., 2019), making this technique desirable in a clinical setting. Metabolomics offers an advantageous avenue in the discovery of stroke biomarkers because metabolites are small enough to cross the blood-brain barrier and can provide insight into the metabolic dysfunction occurring in the brain during stroke. One benefit of urinary metabolomics is that urine is not subject to homeostatic regulation, like blood (Qureshi et al., 2017). It is important to note that urinary metabolites are significantly influenced by the environment whereas blood is much more stable (Vignoli et al., 2019). Thus, blood can provide information about the current metabolic state of the body whereas urine gives insights into the metabolites resulting from previous physiological processes that are being excreted. For this reason, large differences are often found across the urinary and blood metabolomes even though these metabolomes can also be partially congruent. Urine and serum were chosen for the present study due to their ease of collection, the information they hold about the body, and their complementary nature.

Biomarkers are molecules found in body fluids that identify normal or pathological states, monitor treatment efficacy, and predict future disease outcomes (Atkinson et al., 2001). Biomarkers are practical because they can differentiate between ischemic and hemorrhagic stroke and rule out other causes of neuronal damage. Stroke therapies are time-sensitive, hence a rapid, precise biomarker could save lives and greatly reduce disability. As mentioned earlier, imaging techniques are not always readily available to diagnose stroke. At present, no urinary or serum biomarkers are routinely used in a clinical setting to diagnose or monitor stroke. Regardless, it is beneficial to develop an alternative diagnostic technique that can confirm stroke diagnosis and

prognosis when other techniques are unavailable or provide a false negative diagnosis of stroke (Sacco et al., 2013). Our best current predictors of stroke prognosis are age, initial neurological deficit (Donnan et al., 2008) and clinical assessments such as the National Institute of Health Stroke Scale (NIHSS; Muir et al., 1996). Thus, metabolomic biomarkers have potential to fill the need for accurate stroke biomarkers for diagnosis, prognosis, and treatment response.

Although there are no stroke biomarkers currently being used clinically, there is a multitude of biomarker signatures that have displayed potential in the literature. The large majority of current stroke biomarkers are markers of diagnosis, such as circular RNAs (Zhang et al., 2023), microRNAs (Eyileten et al., 2018), glial fibrillary acidic protein, activated protein C- protein C inhibitor complex, retinol binding protein 4 (Misra et al., 2017), S100B, brain natriuretic peptide (Monbailliu et al., 2017), amino acids, LysoPC(16:0), LysoPC(18:2), uric acid, citrate and palmitic acid (Ke et al., 2019), C-reactive protein, alpha2-plasmin inhibitor, plasminogen activator inhibitor-1, P-selectin, adiponectin, leptin, homocysteine, asymmetric dimethylarginine, lipoprotein (a), haptoglobin, and serum amyloid A (Andone et al., 2021). Furthermore, various prognostic indicators exist, including cortisol, copeptin, procalcitonin, adipocyte fatty acid–binding protein, mannose-binding lectin (Montellano et al., 2021), pro-inflammatory cytokines, Von Willebrand factor (Huang et al., 2023), thrombin antithrombin, E-selectin, resistin, lipoprotein-associated phospholipase A2, d-dimers, matrix malloproteinases, and endostatin, vascular endothelial growth factor (Andone et al., 2021). On the other hand, there are very few markers for treatment efficacy. To name a few, alpha2-plasmin inhibitor, thrombin antithrombin, and d-dimers are notable for monitoring treatments (Andone et al., 2021). Despite the extensive pool of biomarker candidates, rigorous clinical validation studies are imperative for their practical implementation. Additionally, the majority of the aforementioned biosignatures stem from proteomic investigations. While proteomics may yield an effective panel of biomarkers, embracing

other ‘omics’ techniques is crucial for a comprehensive understanding of the overall stroke pathophysiology and recovery processes. This thesis aims to contribute to the evolving landscape of knowledge in this field.

Metabolites assume critical roles within interconnected biological networks, each occupying a distinct position in a metabolic pathway that collectively serves a broader function. The interplay among these compounds, enzymes, and genes in these networks forms the foundation of metabolomic analysis (Gao et al., 2010). Much like metabolites, pathways are susceptible to perturbations arising from abnormal conditions (Johnson et al., 2015). Numerous metabolomic pathways have been associated with stroke, with several pathways consistently being identified as significantly altered in existing literature. Among them are aminoacyl-tRNA biosynthesis (Jia et al., 2021; Ke et al., 2019; Sidorov et al., 2020; Qi et al., 2022), nitrogen metabolism (Ke et al., 2019; Sidorov et al., 2020), glyoxylate and dicarboxylate metabolism (Jia et al., 2021; Ke et al., 2019), glycerophospholipid metabolism (Qi et al., 2022; Chen et al., 2023), butanoate metabolism (Chen et al., 2023; Jia et al., 2021), synthesis and degradation of ketone bodies (Jia et al., 2021; Ke et al., 2019) and amino acid synthesis and metabolism (Jia et al., 2021; Ke et al., 2019; Sidorov et al., 2020), such as phenylalanine metabolism, and phenylalanine, tyrosine, and tryptophan biosynthesis (Jia et al., 2021; Ke et al., 2019). These pathways play pivotal roles in energy metabolism, protein synthesis, oxidative stress, cellular homeostasis, and neurotransmitter synthesis in the aftermath of stroke. Through an exploration of their foundational mechanisms, as undertaken in this thesis, metabolomic pathways offer insights into the intricate processes governing stroke and subsequent recovery.

The use of metabolites taken from human biofluids is a valuable and non-invasive method to monitor pathological and physiological changes that occur due to changes in health and disease. NMR-based metabolomics of urine and serum is amenable to the creation of personalized

treatments for patients, and its ensuing biomarkers can be used to detect and monitor disease, and observe efficacy of treatment programs (Emwas et al., 2013). NMR-based metabolomics has proven to be an effective tool in biomarker discovery for many neurological conditions, including stroke (Qureshi et al., 2017), spinal cord injury (Bykowski et al., 2021a; Bykowski et al., 2023), traumatic brain injury (Bykowski et al., 2021b), and sport-related concussion (Paxman et al., 2017; Wanner et al., 2021). Hence, the metabolic profiles of stroke patients can be used as a reliable, efficient, and readily available indicators to diagnose stroke, predict treatment outcomes, and monitor treatment efficacy for each individual patient, thereby enabling clinicians to provide the best possible personalized treatment for all patients and optimizing neurorehabilitation outcomes.

1.4 PERSONALIZED MEDICINE AND NMR-BASED METABOLOMICS

Stroke affects each individual patient differently due to genetics, lifestyle factors, brain organization, stroke severity, and stroke location. The current techniques in stroke rehabilitation are generic and derived from population averages. Testing each treatment on a single patient the therapy that produces the most effective results is costly and time-consuming (Koen et al., 2016). Furthermore, different patients rarely respond to treatments with the same timeline, efficacy, and outcomes. These and other inter-individual differences highlight the necessity for personalized medicine approaches in stroke diagnosis and management. Personalized medicine is the utilization of cellular and molecular information, often through biomarkers from ‘omics’ sciences, to diagnose, treat, monitor, and prevent disease. NMR-based metabolomics can be used to generate personalized treatment regimens by analyzing a patient’s metabolome and finding biomarkers that confirm stroke diagnosis and predict future outcomes. Recovery and gradual neurological improvement can be monitored by taking clinically accessible samples during the treatment process and monitoring biomarker levels (Koen et al., 2016). Metabolite biomarkers are excellent

tools for personalized medicine because they are fast, reproducible, cost-effective, non-invasive, and have good quantitative accuracy (Xia et al., 2013). Biomarkers could potentially be assessed preclinically to predict treatment response and select optimal treatment program. It is most practical to implement a panel of biomarkers consisting of 1-10 of the most significantly altered metabolites because, in a hospital setting, it is not feasible to create an assay consisting of a larger number of metabolites to diagnose and monitor stroke (Xia et al., 2013). Determining a biomarker or panel of biomarkers, or biosignatures, for stroke could guide the development of personalized medicine by revealing the underlying cellular mechanisms responsible for individual differences in stroke prognosis, recovery, and stroke itself (Koen et al., 2016). To our knowledge, the present research is the first to determine NMR-based metabolomic biomarkers indicative of triage, prognosis, and treatment success in patients with stroke.

1.5 RESEARCH OBJECTIVES AND HYPOTHESES

The purpose of this study is to determine if ^{1}H NMR metabolomic profiling can produce novel biomarkers indicative of recovery processes and treatment efficacy in stroke. In a personalized medicine approach, metabolic profiles can be used to predict, monitor, and improve a patient's neurorehabilitation progress. Ultimately, the discovery of metabolomic biomarkers of neurorehabilitation, along with a better understanding of the repair processes, will allow for a more personalized and effective treatment program for patients.

1.5.1 Objectives

The main objective of this research is to identify potential metabolomic biomarkers that provide information about stroke severity, prognosis, and recovery. For this MSc thesis, the research will address the following specific objectives:

(1) Identify distinct differences in the urinary and blood metabolomic profiles in stroke patients during the acute and chronic phase post-lesion;

(2) Pinpoint which metabolites are leading to the observed differences and correlating them to clinical measures of extent of recovery;

(3) Discover which biological pathways are involved in the repair and functional recovery process following stroke.

1.5.2 Hypotheses

It is hypothesized that the metabolites identified in both urine and blood will vary from the time initial samples are collected to when final samples are collected, post-rehabilitation. It is expected that recovery processes occurring following stroke will be reflected by the metabolic profile. Significantly altered metabolites are predicted to correlate strongly to clinical measures of function; thus, revealing involvement in stroke injury and recovery processes. Further, this information will be used to determine which specific metabolites are most reliable to denote injury severity, potential recovery and treatment efficacy. It is anticipated that significant biological pathways will uncover the interactions within a biological system during recovery from stroke. This can provide insights into which biomolecules and cellular processes are involved in stroke repair and recovery. In summation, the overarching hypothesis of this thesis are that (1) alterations in the urinary and blood metabolomic profiles after stroke will reveal distinct recovery and repair processes in the acute phase of stroke compared to the chronic phase; and that (2) metabolite changes will correlate to clinical measures of function and reflect the injury severity and extent of recovery.

REFERENCES

Allen, C. L., & Bayraktutan, U. (2009). Oxidative stress and its role in the pathogenesis of ischaemic stroke. *International Journal of Stroke, 4*(6), 461-470. https://doi.org/10.1111/j.1747-4949.2009.00387.x

Andone, S., Bajko, Z., Motataianu, A., Mosora, O., & Balasa, R. (2021). The role of biomarkers in atherothrombotic stroke-A systematic review. *International Journal of Molecular Sciences, 22*(16), 9032. https://doi.org/10.3390/ijms22169032

Atkinson, A. J., Colburn, W. A., DeGruttola, V. G., DeMets, D. L., Downing, G. J., Hoth, D. F., Oates, J. A., Peck, C. C., Schooley, R. T., Spilker, B. A., Woodcock, J., Zeger, S. L., & Biomarkers Definitions Working Group. (2001). Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology and Therapeutics, 69*(3), 89-95. https://doi.org/10.1067/mcp.2001.113989

Bouatra, S., Aziat, F., Mandal, R., Guo, A. C., Wilson, M. R., Knox, C., Bjorndahl, T. C., Krishnamurthy, R., Saleem, F., Liu, P., Dame, Z. T., Poelzer, J., Huynh, J., Yallou, F. S., Psychogios, N., Dong, E., Bogumil, R., Roehring, C., & Wishart, D. S. (2013). The human urine metabolome. *PloS One, 8*(9), e73076. https://doi.org/10.1371/journal.pone.0073076

Brott, T., Adams, J., H P, Olinger, C. P., Marler, J. R., Barsan, W. G., Biller, J., . . . Hertzberg, V. (1989). Measurements of acute cerebral infarction: A clinical examination scale. *Stroke, 20*(7), 864-870.

Bykowski, E. A., Petersson, J. N., Dukelow, S., Ho, C., Debert, C. T., Montina, T., & Metz, G. A. S. (2021a). Urinary biomarkers indicative of recovery from spinal cord injury: A pilot study. *IBRO Neuroscience Reports, 10*(1), 178-185. https://doi.org/10.1016/j.ibneur.2021.02.007

Bykowski, E. A., Petersson, J. N., Dukelow, S., Ho, C., Debert, C. T., Montina, T., & Metz, G. A. S. (2021b). Urinary metabolomic signatures as indicators of injury severity following traumatic brain injury: A pilot study. *IBRO Neuroscience Reports, 11*, 200-206. https://doi.org/10.1016/j.ibneur.2021.10.003

Bykowski, E. A., Petersson, J. N., Dukelow, S., Ho, C., Debert, C. T., Montina, T., & Metz, G. A. S. (2023). Identification of serum metabolites as prognostic biomarkers following spinal cord injury: A pilot study. *Metabolites, 13*(5), 605. https://doi.org/10.3390/metabo13050605

Campbell, B. C. V. & Khatri, P. (2020). Stroke. *The Lancet (British Edition), 396*(10244), 129. https://doi.org/10.1016/S0140-6736(20)31179-X

Chen, Q., Zhou, T., Yuan, J., Xiong, X., Liu, X., Qiu, Z., Hu, L., Lu, H., He, Q., Liu, C., & Yang, Q. (2023). Metabolomics profiling to characterize cerebral ischemia-reperfusion injury in mice. *Frontiers in Pharmacology, 14*, 1091616-1091616. https://doi.org/10.3389/fphar.2023.1091616

Dagonnier, M., Donnan, G. A., Davis, S. M., Dewey, H. M., & Howells, D. W. (2021). Acute stroke biomarkers: Are we there yet? *Frontiers in Neurology, 12*, 619721-619721. https://doi.org/10.3389/fneur.2021.619721

Donnan, G. A., Fisher, M., Macleod, M., & Davis, S. M. (2008). Stroke. *Lancet, the, 371(*9624), 1612-1623. doi:10.1016/S0140-6736(08)60694-7

Doyle, K. P., Simon, R. P., & Stenzel-Poore, M. P. (2008). Mechanisms of ischemic brain damage. *Neuropharmacology, 55*(3), 310-318. https://doi.org/10.1016/j.neuropharm.2008.01.005

Emwas, A. M. (2015). The strengths and weaknesses of NMR spectroscopy and mass spectrometry with particular focus on metabolomics research. *Methods in Molecular Biology (Clifton, N.J.)*, 1277, 161.

Emwas, A. M., Salek, R. M., Griffin, J. L., & Merzaban, J. (2013). NMR-based metabolomics in human disease diagnosis: Applications, limitations, and recommendations. *Metabolomics, 9*(5), 1048-1072. https://doi.org/10.1007/s11306-013-0524-y

Eyileten, C., Wicik, Z., De Rosa, S., Mirowska-Guzel, D., Soplinska, A., Indolfi, C., Jastrzebska-Kurkowska, I., Czlonkowska, A., & Postula, M. (2018). MicroRNAs as diagnostic and prognostic biomarkers in ischemic stroke-A comprehensive review and bioinformatic analysis. *Cells (Basel, Switzerland), 7*(12), 249. https://doi.org/10.3390/cells7120249

Fuchs, Y., & Steller, H. (2015). Live to die another way: Modes of programmed cell death and the signals emanating from dying cells. *Nature Reviews. Molecular Cell Biology, 16*(6), 329-344. https://doi.org/10.1038/nrm3999

Gao, J., Tarcea, V. G., Karnovsky, A., Mirel, B. R., Weymouth, T. E., Beecher, C. W., Cavalcoli, J. D., Athey, B. D., Omenn, G. S., Burant, C. F., & Jagadish, H. V. (2010). Metscape: A cytoscape plug-in for visualizing and interpreting metabolomic data in the context of human metabolic networks. *Bioinformatics, 26*(7), 971-973. https://doi.org/10.1093/bioinformatics/btq048

Gowland, C., Stratford, P., Ward, M., Moreland, J., Torresin, W., Vanhullenaar, S., Sanford, J., Barreca, S., Vanspall, B., & Plews, N. (1993). Measuring physical impairment and disability with the chedoke-mcmaster stroke assessment. *Stroke (1970), 24*(1), 58-63. https://doi.org/10.1161/01.STR.24.1.58

Huang, Y., Wang, Z., Huang, Z., & Liu, Z. (2023). Biomarkers and the outcomes of ischemic stroke. *Frontiers in Molecular Neuroscience, 16*, 1171101-1171101. https://doi.org/10.3389/fnmol.2023.1171101

Jia, J., Zhang, H., Liang, X., Dai, Y., Liu, L., Tan, K., Ma, R., Luo, J., Ding, Y., & Ke, C. (2021). Application of metabolomics to the discovery of biomarkers for ischemic stroke in the murine model: A comparison with the clinical results. *Molecular Neurobiology, 58*(12), 6415-6426. https://doi.org/10.1007/s12035-021-02535-2

Johnson, C. H., Ivanisevic, J., Benton, H. P., & Siuzdak, G. (2015). Bioinformatics: The next frontier of metabolomics. *Analytical Chemistry (Washington), 87*(1), 147-156. https://doi.org/10.1021/ac5040693

Ke, C., Pan, C., Zhang, Y., Zhu, X., & Zhang, Y. (2019). Metabolomics facilitates the discovery of metabolic biomarkers and pathways for ischemic stroke: A systematic review. *Metabolomics, 15*(12), 152-21. https://doi.org/10.1007/s11306-019-1615-1

Keith, R. A., Granger, C. V., Hamilton, B. B., & Sherwin, F. S. (1987). The functional independence measure: A new tool for rehabilitation. *Advances in Clinical Rehabilitation, 1*, 6-18.

Khoshnam, S. E., Winlow, W., Farzaneh, M., Farbood, Y., & Moghaddam, H. F. (2017). Pathogenic mechanisms following ischemic stroke. *Neurological Sciences, 38*(7), 1167-1186. https://doi.org/10.1007/s10072-017-2938-1

Koen, N., Du Preez, I., & Loots, D. T. (2016). Metabolomics and personalized medicine. In R. Donev (Ed.), *Advances in protein chemistry and structural biology* (pp. 53-78). Elsevier. https://doi.org/10.1016/bs.apcsb.2015.09.003

Kwah, L. K., & Diong, J. (2014). National institutes of health stroke scale (NIHSS). *Journal of Physiotherapy, 60*(1), 61-61. doi:10.1016/j.jphys.2013.12.012

Lai, T. W., Zhang, S., & Wang, Y. T. (2014). Excitotoxicity and stroke: Identifying novel targets for neuroprotection. *Progress in Neurobiology, 115*, 157-188. https://doi.org/10.1016/j.pneurobio.2013.11.006

Lucas, S., Rothwell, N. J., & Gibson, R. M. (2009). The role of inflammation in CNS injury and disease: The role of inflammation in CNS. *British Journal of Pharmacology, 147*(S1), S232-S240. https://doi.org/10.1038/sj.bjp.0706400

Mackintosh, S. (2009). Functional independence measure. *Australian Journal of Physiotherapy, 55*(1), 65-65. doi:10.1016/S0004-9514(09)70066-2

Misra, S., Kumar, A., Kumar, P., Yadav, A. K., Mohania, D., Pandit, A. K., Prasad, K., & Vibha, D. (2017). Blood-based protein biomarkers for stroke differentiation: A systematic review. Proteomics. *Clinical Applications, 11*(9-10), 1700007-n/a. https://doi.org/10.1002/prca.201700007

Monbailliu, T., Goossens, J., & Hachimi-Idrissi, S. (2017). Blood protein biomarkers as diagnostic tool for ischemic stroke: A systematic review. *Biomarkers in Medicine, 11*(6), 503-512. https://doi.org/10.2217/bmm-2016-0232

Montellano, F. A., Ungethuem, K., Ramiro, L., Nacu, A., Hellwig, S., Fluri, F., Whiteley, W. N., Bustamante, A., Montaner, J., & Heuschmann, P. U. (2021). Role of blood-based biomarkers in ischemic stroke prognosis A systematic review. *Stroke (1970), 52*(2), 543-551. https://doi.org/10.1161/STROKEAHA.120.029232

Muir, K., Weir, C., Murray, G., Povey, C., & Lees, K. (1996). Comparison of neurological scales and scoring systems for acute stroke prognosis. *Stroke (1970), 27*(10), 1817-1820. https://doi.org/10.1161/01.STR.27.10.1817

Murphy, T. H., Li, P., Betts, K., & Liu, R. (2008). Two-photon imaging of stroke onset in vivo reveals that NMDA-receptor independent ischemic depolarization is the major cause of rapid reversible damage to dendrites and spines. *The Journal of Neuroscience, 28*(7), 1756-1772. https://doi.org/10.1523/JNEUROSCI.5128-07.2008

Nicholson, J. K., & Lindon, J. C. (2008). Metabonomics. *Nature (London), 455*(7216), 1054-1056. https://doi.org/10.1038/4551054a

Nicholson, J. K., Lindon, J. C., & 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), 1181-1189. https://doi.org/10.1080/004982599238047

Ontario Stroke Network. (n.d.) *Fact sheet: Stroke statistics*. http://www.ontariostrokenetwork.ca/pdf/Final_Fact_Sheet_Stroke_Stats_3.pdf

Paxman, E. J., Wanner, Z., Boora, J., Douglas, K., Metz, G., Montina, T., & Debert, C. (2017). Metabolomic biomarkers of concussion. *British Journal of Sports Medicine, 51*(11), A4-A4. https://doi.org/10.1136/bjsports-2016-097270.9

Psychogios, N., Hau, D. D., Peng, J., Guo, A. C., Mandal, R., Bouatra, S., Sinelnikov, I., Krishnamurthy, R., Eisner, R., Gautam, B., Young, N., Xia, J., Knox, C., Dong, E., Huang, P., Hollander, Z., Pedersen, T. L., Smith, S. R., Bamforth, F., . . . Wishart, D. S. (2011). The human serum metabolome. *PloS One, 6*(2), e16957-e16957. https://doi.org/10.1371/journal.pone.0016957

Qi, B., Zhang, Y., Xu, B., Zhang, Y., Fei, G., Lin, L., & Li, Q. (2022). Metabolomic characterization of acute ischemic stroke facilitates metabolomic biomarker discovery. Applied *Biochemistry and Biotechnology, 194*(11), 5443-5455. https://doi.org/10.1007/s12010-022-04024-1

Qureshi, M. I., Vorkas, P. A., Coupland, A. P., Jenkins, I. H., Holmes, E., & Davies, A. H. (2017). Lessons from metabonomics on the neurobiology of stroke. *SAGE Publications*. https://doi.org/10.1177/1073858416673327

Roth, G. A., Mensah, G. A., Johnson, C. O., Addolorato, G., Ammirati, E., Baddour, L. M., Barengo, N. C., Beaton, A. Z., Benjamin, E. J., Benziger, C. P., Bonny, A., Brauer, M., Brodmann, M., Cahill, T. J., Carapetis, J., Catapano, A. L., Chugh, S. S., Cooper, L. T., Coresh, J., . . . GBD-NHLBI-JACC Global Burden of Cardiovascular Diseases Writing Group. (2020). Global burden of cardiovascular diseases and risk factors, 1990-2019 update from the GBD 2019 study. *Journal of the American College of Cardiology, 76*(25), 2982-3021. https://doi.org/10.1016/j.jacc.2020.11.010

Sacco, R. L., Kasner, S. E., Broderick, J. P., Caplan, L. R., Connors, J. J. (., Culebras, A., Elkind, M. S. V., George, M. G., Hamdan, A. D., Higashida, R. T., Hoh, B. L., Janis, L. S., Kase, C. S., Kleindorfer, D. O., Lee, J., Moseley, M. E., Peterson, E. D., Turan, T. N., Valderrama, A. L., . . . Council on Cardiovascular and Stroke Nursing. (2013). An updated definition of stroke for the 21st century: A statement for healthcare professionals from the american heart Association/American stroke association. *Stroke (1970), 44*(7), 2064-2089. https://doi.org/10.1161/STR.0b013e318296aeca

Sidorov, E., Bejar, C., Xu, C., Ray, B., Reddivari, L., Chainakul, J., Vanamala, J. K. P., & Sanghera, D. K. (2020). Potential metabolite biomarkers for acute versus chronic stage of ischemic stroke: A pilot study. *Journal of Stroke and Cerebrovascular Diseases, 29*(4), 104618-104618. https://doi.org/10.1016/j.jstrokecerebrovasdis.2019.104618

Swanson, R., Ying, W., & Kauppinen, T. (2004). Astrocyte influences on ischemic neuronal death. *Current Molecular Medicine, 4*(2), 193-205. https://doi.org/10.2174/1566524043479185

Timpone, V. M., Jensen, A., Poisson, S. N., & Trivedi, P. S. (2020). Compliance with imaging guidelines for workup of transient ischemic attack evidence from the nationwide emergency department sample. *Stroke (1970), 51*(8), 2563-2567. https://doi.org/10.1161/STROKEAHA.120.029858

Uzdensky, A. B. (2019). Apoptosis regulation in the penumbra after ischemic stroke: Expression of pro- and antiapoptotic proteins. *Apoptosis (London), 24*(9-10), 687-702. https://doi.org/10.1007/s10495-019-01556-6

Vignoli, A., Ghini, V., Meoni, G., Licari, C., Takis, P. G., Tenori, L., Turano, P., & Luchinat, C. (2019). High-Throughput metabolomics by 1D NMR. *Angewandte Chemie (International Ed.), 58*(4), 968-994. https://doi.org/10.1002/anie.201804736

Wanner, Z. R., Southam, C. G., Sanghavi, P., Boora, N. S., Paxman, E. J., Dukelow, S. P., Benson, B. W., Montina, T., Metz, G. A. S., & Debert, C. T. (2021). Alterations in urine metabolomics following sport-related concussion: A 1H NMR-based analysis. *Frontiers in Neurology, 12*, 645829-645829. https://doi.org/10.3389/fneur.2021.645829

Wishart, D. S. (2008). Quantitative metabolomics using NMR. *TrAC, Trends in Analytical Chemistry (Regular Ed.), 27*(3), 228-237. https://doi.org/10.1016/j.trac.2007.12.001

Xia, J., Broadhurst, D. I., Wilson, M., & Wishart, D. S. (2013). Translational biomarker discovery in clinical metabolomics: An introductory tutorial. *Metabolomics, 9*(2), 280-299. https://doi.org/10.1007/s11306-012-0482-9

Zhang, X., Wan, M., Min, X., Chu, G., Luo, Y., Han, Z., Li, W., Xu, R., Luo, J., Li, W., Yang, Y., Ma, Y., Jiao, L., & Wang, T. (2023). Circular RNA as biomarkers for acute ischemic stroke: A systematic review and meta-analysis. *CNS Neuroscience & Therapeutics, 29*(8), 2086-2100. https://doi.org/10.1111/cns.14220

CHAPTER 2: FROM RAW DATA TO METABOLITE DISCOVERIES: PRE-PROCESSING, STATISTICAL ANALYSIS, AND CLINICAL INTERPRETATION

In this chapter, the steps that are taken to prepare the data for processing and the subsequent statistical methods are outlined. First, the data are pre-processed using data reduction, normalization, log transformation, and Pareto scaling. A linear regression model is applied to the metabolomics data to remove the effects of covariates. Then, univariate statistics and a machine learning technique called VIAVC is used to determine significant metabolites. Multivariate tests, including principal component analysis (PCA), partial least squares discriminant analysis (PLS-DA), and orthogonal projections to latent structures discriminant analysis (OPLS-DA) are conducted. Double 10-fold cross-validation and permutation testing are performed on PLS-DA and OPLS-DA to validate the results. Subsequently, ROCs and AUCs are produced. In relation to the overall methods, these statistical steps follow sample preparation and NMR data acquisition. After data pre-processing and statistical analysis, the data are ready to be used to identify significant metabolites and run pathway analysis. Finally, clinical correlations can be conducted using the significant metabolites to observe their relationship with clinical assessment scores. This research workflow is illustrated in Figure 1.

2.1 DATA PRE-PROCESSING

2.1.1 Data Reduction

Data reduction is used to go from thousands of data points down to hundreds while still maintaining the same quality of information. In NMR-based metabolomics, this technique is done by creating discrete bins. The process of binning or bucketing is a common NMR-based metabolomics technique that involves the division of the NMR spectrum into hundreds of fixed or variable-sized sections by identifying peaks and peak patterns. Several methods can be used to

separate spectra into bins. One method is based on uniform binning, which splits the spectrum into bins of fixed size, however, this is not ideal for NMR spectra since peaks are not equal size and uniform bins can separate potentially significant peaks into two bins (Anderson et al., 2011). Dynamic binning is another method that smooths the spectra into a composite waveform that is used to dynamically bin the spectra (Davis et al. 2007). A limitation of dynamic binning is that adjacent peaks can be misinterpreted as misaligned peaks and subsequently removed when creating the composite spectrum. The method used in this paper is dynamic adaptive binning (DAB) followed by manual correction. DAB is a MATLAB-based binning algorithm that determines bin edges with dynamic programming (Anderson et al., 2011). Spectral peak positions can be displaced due to ionic strength, sample pH, and composition (Spraul et al., 1994). Binning can help reduce the impact that these shifts have on the spectra and later identification of metabolites. Additionally, binning reduces the dimensionality of the data thereby increasing the power of later statistics, including PLS-DA and OPLS-DA (Anderson et al., 2011).

Peak alignment is an important factor influencing the accuracy of binning algorithms (Anderson et al., 2011). Urine has many overlapping metabolite species and is especially variable in regards to pH, ionic strength, and other factors that influence chemical shifts. In the present study, recursive segment-wise peak alignment (RSPA) is used to correct major and minor peak positions in complex spectra, specifically urine. Peak alignment is needed to ensure that the data used in later statistics generates valid and meaningful results (Veselkov et al., 2009).

2.1.2 Normalization

Data normalization is required to properly and objectively compare samples in a dataset. It is especially important for urinary metabolomics since there is often large variations in metabolite concentrations across samples depending on each individual person's fluid intake before sample collection (Craig et al., 2006; Xia et al., 2013). The data tables typically produced in NMR-based

metabolomics have samples (each spectrum) in the rows and metabolite bins in the columns. When normalizing, the total integral density across all metabolite bins (i.e., the sum across all columns) of each spectrum (row) is set to the same value, therefore the sum across all metabolites for each sample is normalized (Craig et al., 2006; Emwas et al., 2013).

2.1.3 Data Transformation

Parametric statistical tests are used under the assumption that the data are normally distributed. Often with clinical and metabolomic data, the data are non-normal and using parametric statistics in this case leads to false interpretations. For this reason, it is essential to transform the data before making any statistical inferences (Xia et al., 2013). Log transformation of the data minimizes variance and makes it much more effective during multivariate analysis (Emwas et al., 2013). Transformation is performed for the columns, or normalized metabolite concentrations, in the metabolomics data tables (Emwas et al., 2018). The present study makes use of a logarithmic transformation, which takes the value x and replaces it with $\log_{10}(x)$. This reduces extremely high values and transforms the data to be normally distributed (Xia et al., 2013).

2.1.4 Data Scaling

Data scaling is an essential pre-processing step when using machine learning algorithms. Contrarily to data normalization, which occurs across each row (or sample), scaling adjusts each column or spectral bin (or metabolite) of metabolomics data tables. Scaling alters the metabolite features to fit within a specified scale, which often consists of mean-centering and scaling. Mean-centering subtracts the mean from each individual data point in the column while scaling consists of dividing each data point by the standard deviation (Craig et al., 2006; Xia et al., 2013). After mean centering and scaling, each bin (or column) can be represented with respect to a deviation from a common mean and standard deviation allowing for metabolites with large differences in concentration to be more easily compared (Xia et al., 2013). Scaling to the unit variance (UV) is

the most common method and involves simply dividing by the standard deviation. UV scaling shrinks long variables and stretches short variables so that each variable can influence data analysis equally. This technique is useful if there is no information about the importance of variables, but ideally extraneous or noisy variables should be weighted down and pertinent variables are weighted up so they contribute more to the multivariate modelling (Eriksson et al., 2013). The present study uses Pareto scaling, where the individual spectral bins are mean-centered and divided by the square root of the standard deviation of the data points in the column (Craig et al., 2006; Eriksson et al., 2013). This technique is a middle ground between no scaling and UV scaling and is ideal for multivariate modelling of spectroscopic data (Eriksson et al., 2013).

2.2 UNIVARIATE & BIVARIATE STATISTICS

2.2.1 Statistical Tests and Correction

As mentioned earlier, metabolite concentrations are not typically normally distributed and a *log* transformation does not guarantee parametric data. Therefore, the Wilcoxon-Mann-Whitney test (Mann & Whitney, 1947) can be applied to compare the distribution of metabolites across the comparison groups. Alternatively, the Wilcoxon signed ranks test (Ridgman, 1989) can be used for a paired study design. Bonferroni correction (Bonferroni, 1935) is often used when multiple metabolites are analyzed together to avoid false positives (Vignoli et al., 2019). Bonferroni correction is a useful method to account for multiple comparisons, which is essential in metabolomics because hundreds of metabolites are being measured simultaneously for each individual subject. This technique increases the confidence level relative to the number of comparisons being used (Xia et al., 2013).

2.2.2 Receiver Operator Characteristic (ROC) Curve

ROC graphs are used to visualize and evaluate classifier performance (Fawcett, 2006) and they can illustrate the diagnostic ability of a biomarker. ROCs are univariate, because they can be conducted on a single bin, but they are also multivariate when applied to multiple bins. Sensitivity is the true positive rate, or the proportion of positive tests for a disease correctly classified as positive. Specificity is the true negative rate, or the probability of healthy subjects correctly classified as negative (Xia et al., 2013). ROC curves plot the sensitivity versus 1-specificity, which is the false positive rate or the proportion of positive tests that are incorrectly classified and should be negative. Sensitivity and 1-specificity are between 0-1, with better sensitivity being closer to 1 and better 1-specificity being closer to 0. Thus, the diagonal of the ROC graph represents random guessing during classification and any point in the lower right triangle is said to perform worse than random guessing (Fawcett, 2006). When ROC curve analysis is being conducted for a biomarker model, the analysis is performed on the test sets produced by cross-validation (CV). Using the different subsets of metabolites, this outputs multiple ROC curves, which are averaged into one smooth curve (Xia et al., 2013). Q^2 measures the model fit for the regression and is calculated using the difference between the predicted variables and the models true output variables (Debik et al., 2022). Area-under-the-curve (AUC) is a single metric that summarizes the ROC and is the probability that a randomly chosen positive diagnosis will be ranked higher than a randomly chosen negative test (Xia et al., 2013). In other words, the AUC measures accuracy of classifying subjects. A perfect classifier has an AUC of 1.0, an AUC of 0.5 means subjects are being classified at random, and an AUC of less than 0.5 means classification is worse than random. Often confidence intervals (CI) set at a confidence level of 95% are calculated alongside ROCs. A CI estimates the range of values in the sample population and indicates the extent of standard error in a given sample. Another standard measure for a classification task is predictive accuracy, which

is a method to assess the efficacy of a biomarker by determining the percentage of subjects correctly classified (Xia et al., 2013).

2.2.3 Correlations

The Spearman rank correlation (Spearman, 1904) is the nonparametric equivalent to the Pearson's correlation. The Spearman's correlation evaluates the monotonic relationship between two continuous or ordinal variables, whereas the Pearson's correlation determines the linear relationship between two continuous variables. In metabolomics, most studies utilize correlation analysis such as metabolomic correlation network analysis to investigate relationships between metabolites and their concentrations (Steuer, 2006). Similarly to the present work, several studies have correlated metabolite levels to clinical scores (Bent et al., 2018; Dong et al., 2021). Here, the Spearman rank correlation was chosen for comparing metabolite levels (continuous) to clinical scores (ordinal). Three sets of correlations were calculated to evaluate the ability of significant metabolites to determine triage potential, prognostic potential, and treatment response. For triage correlations, the initial metabolite concentration was correlated to the NIHSS, FIM, and CMSA. For prognostic correlations, the initial metabolite concentration was correlated to the percent difference in initial to 6-month FIM and CMSA scores. For treatment efficacy correlations, the percent difference in initial to 6-month metabolite concentrations was correlated to the percent difference in FIM and CMSA scores. A Bonferroni-Holm correction threshold value was used to determine significance.

2.3 MULTIVARIATE STATISTICS

Multivariate statistical analysis produces plots of metabolic activity using the intensities of peaks in the spectrum as its coordinates. This can then be used to visualize the data by clustering the data points (Nicholson & Lindon, 2008). Multivariate analysis is necessary to draw meaning

from complicated NMR data containing a large number of metabolites (Emwas et al., 2013). Principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA) are two multivariate methods that reduce dimensionality of the data set. PCA is an unsupervised statistical technique that excludes outliers and clusters the data into groups based only on the measured variance. Similarly, PLS-DA is a supervised method that discriminates between groups and separates them based on a particular biological effect or known groupings (Emwas et al., 2013). These methods use a 2D scatterplot to capture the patterns of variance in the data (Xia et al., 2013). PCAs and PLS-DAs produce variable importance in projection (VIP) scores created by univariate hypothesis testing on each individual metabolite, either parametric (*t*-test) or non-parametric (Mann-Whitney U test). R^2 and Q^2 are performance measures produced by PLS-DA models. R^2 is the coefficient of determination, which chooses the optimal model structure and indicates the goodness of fit or how much the variance is explained by the model. Q^2 is the cross-validated R^2, which protects against model over-fitting and estimates the goodness of prediction or the predictive ability of the model. An R^2 and Q^2 value of one represents an optimal model structure and predictive ability, respectively. Although PCA and PLS-DA (pattern recognition techniques) can be used for biomarker discovery, they primarily highlight variable differences across the groups and machine learning algorithms are more ideal to build predictive models. These algorithms model the relationship between a dependent variable and one or more independent variables (Xia et al., 2013).

2.3.1 Unsupervised Analysis

Unsupervised data analysis approaches help to summarize, visualize, and discover trends in the data through data reduction or data clustering (Vignoli et al., 2019). PCA is a linear transformation that attempts to preserve as much of the variance in the original data as possible (Worley & Powers, 2013). For principal component analysis (PCA), variation patterns in the data

are used to build the model and no labelling information is used (Debik et al., 2022). The data are visualized in a lower-dimensional space by calculating the weighted sum of metabolite concentrations in the sample. This is done to explain the variance in the data, which is displayed in the scores plot figure by revealing differences across the groups, clustering of the data, and outliers (Debik et al., 2022). During PCA, the matrix containing metabolite concentration data is used to create the principal components (PCs), where PC1 explains the highest variation in the data, and PC2, PC3, etc. describe less and less variation. The PCs are the coordinates used to create points on the scatter plot and are used to compare samples based on their entire metabolome (Koen et al., 2016). PCA is often used in longitudinal studies to visualize trends across repeated measures (Debik et al., 2022). It is important to note that for this unsupervised technique, group differences are only revealed when within group variation is adequately less than between group variation. For this reason, metabolomic studies often use supervised methods that rely on the class membership of each sample (Worley & Powers, 2013).

2.3.2 Supervised Analysis

In supervised learning, training and testing sets of samples are used to generate models that can be used to make predictions about new, unclassified samples using knowledge obtained from pre-existing samples. This can be done through projection and data reduction (e.g., PLS-DA) or machine learning (e.g., OPLS-DA) (Vignoli et al., 2019). In contrast to unsupervised models, supervised methods use class, or group, labels to build the model (Debik et al., 2022). Partial least squares (PLS) decomposes the data using linear combinations of the original variables, called latent variables, to create a multivariate regression model. The data are projected to a lower-dimensional space to visualize grouping of the samples. Partial least squares discriminant analysis (PLS-DA) is a variant of PLS and uses PLS regression for classification purposes and biomarker selection (Debik et al., 2022). It uses double cross validation to optimize model complexity and

assess PLS-DA model quality (Szymanska et al., 2012). Variable importance to the projection (VIP) scores for PLS-DA are produced using the weighted sum of squares of the loading coefficients, which represent the importance of variation in each metabolite to the observed group separation in the principal components. The weighted sum of squares of the loadings coefficients provides a rank of potentially useful biomarkers, which are the metabolites with the highest VIP scores that are most characteristic of their group (Debik et al., 2022; Koen et al., 2016). Orthogonal projection to latent structures discriminant analysis (OPLS-DA) is an extension of PLS-DA that removes variation from the model, thereby reducing its complexity and making PLS-DA easier to interpret. In OPLS-DA, the variation in X is separated into two parts, one linearly related (predictive) to Y and the other orthogonal (uncorrelated) to Y. OPLS-DA results in a rotation that displays the between-group difference on the x-axis and the within-group difference on the y-axis (Debik et al., 2022; Wiklund et al., 2008).

2.4 PERFORMANCE EVALUATION

2.4.1 Cross-Validation

Cross-validation (CV) is used to assess prediction performance of the biomarkers and prevent over-fitting of the model. CV splits the data into a training set, a validation set, and a test set. The model is developed and optimized using the validation and training sets and model performance is validated using the test set (Westerhuis et al., 2008). Multiple iterations of CV are performed using different sets to test the model and create an optimal model. CV is useful for data sets with small and heterogeneous sample populations since it evaluates the generalizability of the model's predictive ability (Xia et al., 2013). Double cross-validation has two nested loops: CV1 and CV2. CV1 optimizes the PLS-DA model complexity and CV2 tests performance of the final model. CV2 is split into a test set, which is set aside and a rest set, which is used in the CV1 single

cross validation. The rest set going into CV1 is further split into a training set and a validation set. The training set is used to develop classification models (Szymanska et al., 2012). The training and validation sets are used to create an optimal model which is used in CV2 to predict samples in the test set. This CV1 process is repeated until all rest set samples have been in the validation set only once. Further, the CV2 process is repeated until all samples have been included in the test set only once. A prediction error is taken and stored for each model. Finally, the prediction error for all iterations is calculated and the lowest prediction error is selected for each, and an average is taken of the final group of prediction errors, which is used to assess model quality (Szymanska et al., 2012; Westerhuis et al., 2008). Validating the model is an important step when using supervised multivariate analysis in metabolomics because they are prone to overfitting. In addition, receiver operator characteristic (ROC) curves can be used in the multivariate case to determine CV model performance and predictive ability of a set of biomarkers and produces a curve for each test set, which can be averaged to produce a single smoothed curve (Xia et al., 2013). For classification problems, the area under the receiver operating characteristic curve (AUROC) and Q2 statistics are used to illustrate separation between the groups and validate PLS-DA models (Debik et al., 2022; Szymanska et al., 2012). These techniques are further explained in the univariate statistics section above.

2.4.2 Permutation Testing

After validation, multivariate models should endure permutation testing to assess significance (Debik et al., 2022). Permutation testing is another form of model validation that tests significance based on if the model classifies the samples into their respective groups at random or not. A new classification model is formed by randomly permuting the labels of the samples. After permuting multiple models, a null distribution (referring to the null hypothesis) is created and compared to the correctly assigned model to create a *p*-value. A *p*-value <0.05 indicates that there

is less than 5% chance that the correctly assigned model will be produced at random, meaning the model is significantly different from random guessing (Xia et al., 2013).

2.5 MACHINE LEARNING

Machine learning is a data analysis technique in which the computer uses algorithms to identify patterns in the data and learn without human instruction (Debik et al., 2022). Machine learning is required to complete multivariate statistics on the matrix of subjects by metabolites and get a single score that can be used to compute other statistics, such as the ROC curve. The technique also spits out the importance of each metabolite in regard to its contribution to the model's performance. The importance can be determined by the model's loadings vector from regression-based methods (PLS/linear regression). These statistical modelling algorithms have potential to cause model overfitting, meaning that it over exaggerates the predictive ability of the model and its biomarkers. Fortunately, double cross-validation (CV) can be used to prevent over-fitting of the model (Xia et al., 2013), which is used in the present study.

2.5.1 Variable Importance Analysis based on random Variable Combination (VIAVC)

VIAVC is a statistical technique that reveals the importance of metabolites in a dataset and takes into account the synergistic effects among a subset of variables. This machine learning algorithm was implemented into the present study using a MATLAB code provided by Yun et al. (2015). VIAVC uses binary matrix resampling (BMR), which is a technique that generates a population of different variable combinations (subsets) and ensures each variable is sampled with the same probability. Specifically, it randomly samples thousands of combinations of half the variables and takes each individual variable, removes it from the model to discover if it performs better or worse with the variable included, and repeats this for each variable. 10-fold CV from partial least squares-linear discriminant analysis (PLS-LDA) is used on each classifier model from

the BMR sub-sets. Variable importance is determined using the percent increase or decrease of AUC when the variable or metabolite is included and excluded from the PLS-LDA model and by its interaction with a subset of other variables. Based on this, it defines 4 types of variables: strongly informative, weakly informative, uninformative, and interfering. Many iterations are performed where uninformative and interfering variables are removed and informative while weakly informative variables are retained until no further uninformative and interfering variables exist. Finally, a paired *t*-test reveals differences in distributions among the two groups with and without the variable. The *t*-test ranks the final informative metabolites based on *p*-value, which is Bonferroni-Holm corrected. An average of the models with the variable included and excluded is calculated. VIAVC outputs an F-ranked best subset, which is formed using 10-fold CV, where all uninformative and interfering variables are removed. Another output is a subset of variables with the best predictive accuracy, called the VIAVC best subset, ranking informative metabolites based on the *t*-test *p*-value. To determine the best variable subset, the same algorithms is employed but this time using 10-fold double cross validation (10-fold DCV) to determine each variables predictive ability (Yun et al., 2015). In summary, VIAVC is a robust technique to determine significant metabolites because it employs BMR in PLS-LDA modeling, the ROC and AUC, and is double ten-fold cross-validated.

In the metabolomics literature, both univariate and multivariate statistical methods are used to provide information about the metabolites. It is recommended to use both techniques because they output different results and provide complementary information. Univariate analysis provides information on independent concentration changes in a single metabolite and multivariate gives insights into the simultaneous relationship between metabolites (Saccenti et al., 2014). As for univariate tests used in this thesis, paired *t*-test and Wilcoxon signed rank test are commonly used in metabolomics where each individual bin is tested for parametricity using the Shapiro-Wilk test

and a paired *t*-test is used for parametric bins and a Wilcoxon signed rank test is executed for non-parametric bins. Often, metabolomics studies also use multivariate supervised machine learning algorithms to classify samples and train a model of metabolomics data (Vignoli et al., 2019). VIAVC was chosen as the machine learning algorithm in this thesis because it is useful when drawing information from small datasets and provides a more rigorous set of significant metabolites. Although VIAVC has not received the uptake in the field that it deserves, several studies from our group have published results using this combined univariate and multivariate machine learning approach (Bykowski et al., 2021a; Bykowski et al., 2021b; Bykowski et al., 2023; Heynen et al., 2021; Stroud et al., 2022; Wanner et al., 2021). A couple of reasons why significance may be found in univariate tests and not multivariate analysis is because uninformative variables reduce the statistical power, and thus mask information and covariates are poorly estimated, leading to poor model performance (Saccenti et al., 2014). Since VIAVC goes through many iterations removing uninformative variables and interfering variables and uses linear discriminant analysis, it does not possess these same hindrances as many other multivariate tests. Thus, using VIAVC alongside the traditional univariate analyses provides complementary information about metabolite interactions and produces more robust results.

2.6 LINEAR REGRESSION IN METABOLOMICS

Simple linear regression or multiple linear regression are general linear models (GLMs) that use one or more independent variables (IV) to model a relationship with a dependent variable (DV). In metabolomics, regression is used to determine if metabolites have a statistically significant relationship to one or multiple covariates or to detect significant differences in metabolite concentrations across groups (Debik et al., 2022). For example, linear regression has previously been used to investigate the relationship between metabolic concentration and age, with body mass index (BMI) as a covariate (Yu et al., 2012). Simple linear regression uses one IV and

one DV. Since other variables are not included in the regression, no variables are being controlled for in simple linear regression. However, in multiple linear regression there are multiple IVs, and the ones that are not being investigated for their effect on the DV are confounding factors that need to be controlled for. The output reflects the effect of the IV being investigated on the DV adjusted for the covariates (Debik et al., 2022). Often, researchers compare the results from a simple linear regression and multiple linear regression to determine how much the confounders alter the relationship (Pourhoseingholi et al., 2012).

When finding the line of best fit during a linear regression, residuals are calculated. For example, during a linear regression looking at the relationship between height and weight, the residual reflects the parameter height adjusted for weight (Kohler & Kreuter, 2012). Given that, a linear regression model could be used to regress out the effects of confounding factors as a data pre-processing step. This is possible by subtracting the residuals of the confounds in the multiple linear regression from the IV in question and the DV. For instance, Yu et al. (2012) ran a linear regression model with metabolite concentration as the DV and BMI as the IV and used the residuals to plot metabolite concentrations to adjust for BMI. Similarly, Linear mixed models (LMMs) are frequently used in longitudinal metabolomics studies to adjust for dependencies between repeated samples (Debik et al., 2022) and it has been used as a data pre-processing step (Wanichthanarak et al., 2019). LMMs, like linear regression, are GLMs but they use fixed and random effects. Fixed effects are variables that are expected to affect the DV and random effects are factors we are trying to control that account for random variability (Debik et al., 2022). Importantly, ordinary least squares (OLS) regression is often used in clinical magnetic resonance imaging (MRI) studies prior to analysis by using the residuals of the confound regression model (Dinga et al., 2020). This method is precisely how the regression model was used in the present

study. In summation, a linear regression model can be applied as a data pre-processing step to adjust for variation in the data by regressing out the effects of confounding factors.

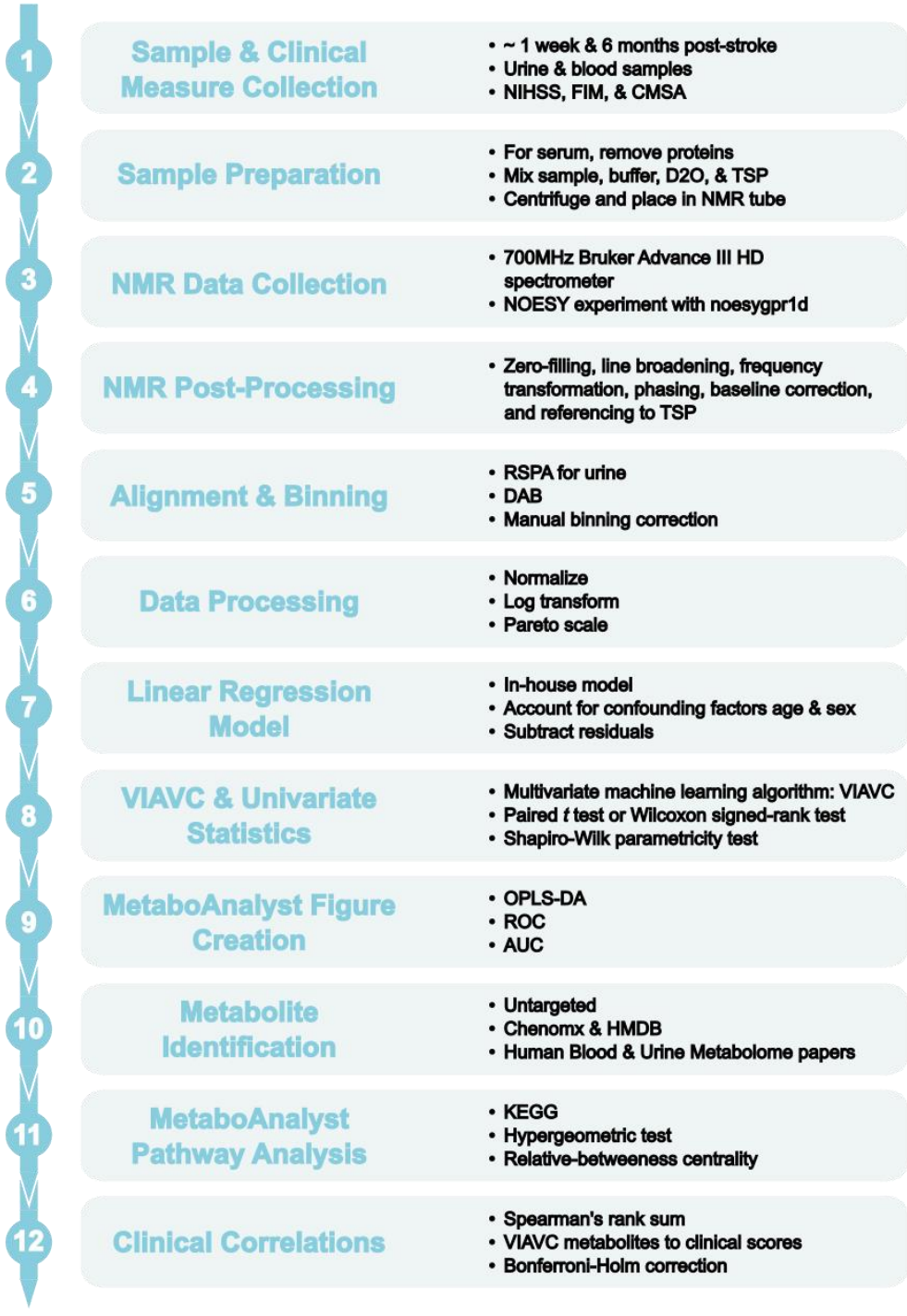

Figure 1. Visual outlining the research workflow and bioinformatics pipeline used throughout this manuscript-based thesis.

REFERENCES

Anderson, P. E., Mahle, D. A., Doom, T. E., Reo, N. V., DelRaso, N. J., & Raymer, M. L. (2011). Dynamic adaptive binning: An improved quantification technique for NMR spectroscopic data. *Metabolomics, 7*(2), 179-190. https://doi.org/10.1007/s11306-010-0242-7

Bent, S., Lawton, B., Warren, T., Widjaja, F., Dang, K., Fahey, J. W., Cornblatt, B., Kinchen, J. M., Delucchi, K., & Hendren, R. L. (2018). Identification of urinary metabolites that correlate with clinical improvements in children with autism treated with sulforaphane from broccoli. *Molecular Autism, 9*(1), 35-35. https://doi.org/10.1186/s13229-018-0218-4

Bonferroni C. E. *Il Calcolo Delle Assicurazioni Su Gruppi Di Test*. In studi Onore Del Professore Salvatore Ortu Carboni. Rome, Italy. 1935a; 13-60.

Bykowski, E. A., Petersson, J. N., Dukelow, S., Ho, C., Debert, C. T., Montina, T., & Metz, G. A. S. (2021a). Urinary biomarkers indicative of recovery from spinal cord injury: A pilot study. *IBRO Neuroscience Reports, 10*(1), 178-185. https://doi.org/10.1016/j.ibneur.2021.02.007

Bykowski, E. A., Petersson, J. N., Dukelow, S., Ho, C., Debert, C. T., Montina, T., & Metz, G. A. S. (2021b). Urinary metabolomic signatures as indicators of injury severity following traumatic brain injury: A pilot study. *IBRO Neuroscience Reports, 11*, 200-206. https://doi.org/10.1016/j.ibneur.2021.10.003

Bykowski, E. A., Petersson, J. N., Dukelow, S., Ho, C., Debert, C. T., Montina, T., & Metz, G. A. S. (2023). Identification of serum metabolites as prognostic biomarkers following spinal cord injury: A pilot study. *Metabolites, 13*(5), 605. https://doi.org/10.3390/metabo13050605

Craig, A., Cloarec, O., Holmes, E., Nicholson, J. K., & Lindon, J. C. (2006). Scaling and normalization effects in NMR spectroscopic metabonomic data sets. *Analytical Chemistry (Washington), 7*8(7), 2262-2267. https://doi.org/10.1021/ac0519312

Davis, R. A., Charlton, A. J., Godward, J., Jones, S. A., Harrison, M., & Wilson, J. C. (2007). Adaptive binning: An improved binning method for metabolomics data using the undecimated wavelet transform. *Chemometrics and Intelligent Laboratory Systems, 85*(1), 144-154. https://doi.org/10.1016/j.chemolab.2006.08.014

Debik, J., Sangermani, M., Wang, F., Madssen, T. S., & Giskeødegård, G. F. (2022). Multivariate analysis of NMR-based metabolomic data. *NMR in Biomedicine, 35*(2), e4638-n/a. https://doi.org/10.1002/nbm.4638

Dinga, R., Schmaal, L., Penninx, B. W., Veltman, D. J., & Marquand, A. F. (2020). Controlling for effects of confounding variables on machine learning predictions. *BioRxiv*. https://doi.org/10.1101/2020.08.17.255034

Dong, M., Hu, L., Wei, Y., & Chen, G. (2021). Metabolomics profiling reveals altered lipid metabolism and identifies a panel of lipid metabolites as biomarkers for Parkinson's disease related anxiety disorder. *Neuroscience Letters, 745*, 135626-135626. https://doi.org/10.1016/j.neulet.2021.135626

Emwas, A. M., Salek, R. M., Griffin, J. L., & Merzaban, J. (2013). NMR-based metabolomics in human disease diagnosis: Applications, limitations, and recommendations. *Metabolomics, 9*(5), 1048-1072. https://doi.org/10.1007/s11306-013-0524-y

Emwas, A. H., Saccenti, E., Gao, X., McKay, R. T., Martins dos Santos, Vitor A.P, Roy, R., & Wishart, D. S. (2018). Recommended strategies for spectral processing and post-processing of 1D 1H-NMR data of biofluids with a particular focus on urine. *Metabolomics, 14*(3), 1-23. https://doi.org/10.1007/s11306-018-1321-4

Eriksson, L., Byrne, T., Johansson, E., Trygg, J., & Vikström, C. (2013). *Multi-and megavariate data analysis basic principles and applications* (Vol. 1). Umetrics Academy.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861-874. https://doi.org/10.1016/j.patrec.2005.10.010

Heynen, J. P., McHugh, R. R., Boora, N. S., Simcock, G., Kildea, S., Austin, M., Laplante, D. P., King, S., Montina, T., & Metz, G. A. S. (2023). Urinary 1H NMR metabolomic analysis of prenatal maternal stress due to a natural disaster reveals metabolic risk factors for non-communicable diseases: The QF2011 queensland flood study. *Metabolites, 13*(4), 579. https://doi.org/10.3390/metabo13040579

Koen, N., Du Preez, I., & Loots, D. T. (2016). Metabolomics and personalized medicine. In R. Donev (Ed.), *Advances in protein chemistry and structural biology* (pp. 53-78). Elsevier. https://doi.org/10.1016/bs.apcsb.2015.09.003

Kohler, U., & F. Kreuter. (3rd Ed.). (2012). *Data Analysis Using Stata*. College Station, TX: Stata Press.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics, 18*(1), 50-60. https://doi.org/10.1214/aoms/1177730491

Nicholson, J. K., & Lindon, J. C. (2008). Metabonomics. *Nature (London), 455*(7216), 1054-1056. https://doi.org/10.1038/4551054a

Pourhoseingholi, M. A., Baghestani, A. R., & Vahedi, M. (2012). How to control confounding effects by statistical analysis. *Gastroenterology and Hepatology from Bed to Bench, 5*(2), 79-83.

Ridgman, W. (1989). Statistical Methods, 8th edn, by G. W. Snedecor & W. G. Cochran. 503 pp. Ames: Iowa State University Press. *The Journal of Agricultural Science, 115*(1), 153-153. doi:10.1017/S0021859600074104

Saccenti, E., Hoefsloot, H. C. J., Smilde, A. K., Westerhuis, J. A., & Heniks, M. M. W. B. (2014). Reflections on univariate and multivariate analysis of metabolomics data. *Metabolomics, 10*(3), 361-374. https://doi.org/10.1007/s11306-013-0598-6

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology, 15*(1), 72-101. https://doi.org/10.2307/1412159

Spraul, M., Neidig, P., Klauck, U., Kessler, P., Holmes, E., Nicholson, J., Sweatman, B., Salman, Farrant, R., Rahr, E., Beddell, C., & Lindon, J. (1994). Automatic reduction of nmr

spectroscopic data for statistical and pattern-recognition classification of samples. *Journal of Pharmaceutical and Biomedical Analysis, 12*(10), 1215-1225. https://doi.org/10.1016/0731-7085(94)00073-5

Steuer, R. (2006). On the analysis and interpretation of correlations in metabolomic data. *Briefings in Bioinformatics, 7*(2), 151-158. https://doi.org/10.1093/bib/bbl009

Stroud, J. E., Gale, M. S., Zwart, S. R., Heer, M., Smith, S. M., Montina, T., & Metz, G. A. S. (2022). Longitudinal metabolomic profiles reveal sex-specific adjustments to long-duration spaceflight and return to earth. *Cellular and Molecular Life Sciences : CMLS, 79*(11), 578-578. https://doi.org/10.1007/s00018-022-04566-x

Szymanska, E., Saccenti, E., Smilde, A. K., & Westerhuis, J. A. (2012). Double-check: Validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics, 8*(Suppl 1), S3-S16. https://doi.org/10.1007/s11306-011-0330-3

Veselkov, K. A., Lindon, J. C., Ebbels, T. M. D., Crockford, D., Volynkin, V. V., Holmes, E., Davies, D. B., & Nicholson, J. K. (2009). Recursive segment-wise peak alignment of biological 1H NMR spectra for improved metabolic biomarker recovery. *Analytical Chemistry (Washington), 81*(1), 56-66. https://doi.org/10.1021/ac8011544

Vignoli, A., Ghini, V., Meoni, G., Licari, C., Takis, P. G., Tenori, L., Turano, P., & Luchinat, C. (2019). High-Throughput metabolomics by 1D NMR. *Angewandte Chemie (International Ed.), 58*(4), 968-994. https://doi.org/10.1002/anie.201804736

Wanichthanarak, K., Jeamsripong, S., Pornputtapong, N., & Khoomrung, S. (2019). Accounting for biological variation with linear mixed-effects modelling improves the quality of clinical metabolomics data. *Computational and Structural Biotechnology Journal, 17*, 611-618. https://doi.org/10.1016/j.csbj.2019.04.009

Wanner, Z. R., Southam, C. G., Sanghavi, P., Boora, N. S., Paxman, E. J., Dukelow, S. P., Benson, B. W., Montina, T., Metz, G. A. S., & Debert, C. T. (2021). Alterations in urine metabolomics following sport-related concussion: A 1H NMR-based analysis. *Frontiers in Neurology, 12*, 645829-645829. https://doi.org/10.3389/fneur.2021.645829

Westerhuis, J. A., Hoefsloot, H. C. J., Smit, S., Vis, D. J., Smilde, A. K., van Velzen, Ewoud J. J, van Duijnhoven, John P. M, & van Dorsten, F. A. (2008). Assessment of PLSDA cross validation. *Metabolomics, 4*(1), 81-89. https://doi.org/10.1007/s11306-007-0099-6

Wiklund, S., Johansson, E., Sjöström, L., Mellerowicz, E. J., Edlund, U., Shockcor, J. P., Gottfries, J., Moritz, T., & Trygg, J. (2008). Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models. *Analytical Chemistry (Washington), 80*(1), 115-122. https://doi.org/10.1021/ac0713510

Worley, B., & Powers, R. (2013). Multivariate analysis in metabolomics. *Current Metabolomics, 1*(1), 92-107. https://doi.org/10.2174/2213235X11301010092

Xia, J., Broadhurst, D. I., Wilson, M., & Wishart, D. S. (2013). Translational biomarker discovery in clinical metabolomics: An introductory tutorial. *Metabolomics, 9*(2), 280-299. https://doi.org/10.1007/s11306-012-0482-9

Yu, Z., Zhai, G., Singmann, P., He, Y., Xu, T., Prehn, C., Römisch-Margl, W., Lattka, E., Gieger, C., Soranzo, N., Heinrich, J., Standl, M., Thiering, E., Mittelstraß, K., Wichmann, H., Peters, A., Suhre, K., Li, Y., Adamski, J., . . . Wang-Sattler, R. (2012). Human serum metabolic profiles are age dependent. *Aging Cell, 11*(6), 960-967. https://doi.org/10.1111/j.1474-9726.2012.00865.x

Yun, Y., Liang, F., Deng, B., Lai, G., Vicente Gonçalves, C. M., Lu, H., Yan, J., Huang, X., Yi, L., & Liang, Y. (2015). Informative metabolites identification by variable importance analysis based on random variable combination. *Metabolomics, 11*(6), 1539-1551. https://doi.org/10.1007/s11306-015-0803-x

CHAPTER 3: UNRAVELING METABOLIC CHANGES FOLLOWING STROKE: INSIGHTS FROM A URINARY METABOLOMICS PILOT ANALYSIS

Author Contributions: Jamie N. Petersson spearheaded sample preparation, analysis, and original draft writing. Elani A. Bykowski assisted with sample preparation and analysis. Sean Dukelow, Chester Ho, Tony Montina, Gerlinde A. S. Metz, and Chantel T. Debert participated in conceptualization and research methodology. Chantel T. Debert, Tony Montina, and Gerlinde A. S. Metz supported this work through supervision and acquisition of funding. All authors contributed to manuscript editing.

3.1 INTRODUCTION

Stroke is a serious medical condition in which the central nervous system (CNS) is damaged due to abnormalities in the vasculature, and results in a wide range of neurological deficits. Worldwide, stroke is a major contributor to long-term disability and is ranked as the second leading cause of death (Roth et al., 2020). Stroke can be broadly classified into two categories: ischemic and hemorrhagic. Ischemic stroke occurs when blood flow in the brain is obstructed, often by a clot, while hemorrhagic stroke is caused by bleeding in the brain and can result from a blood vessel leak or ruptured aneurysm. Both types of stroke result in cell death which usually spreads to other areas around the stroke site (Sacco et al., 2013). Early identification of stroke type is crucial for timely treatment and hence, minimizing brain damage. Imaging techniques such as computed tomography (CT) and magnetic resonance imaging (MRI) are used to confirm the incidence of stroke. CT is the mainstay of imaging for stroke diagnosis and is the most commonly used technique. Both techniques have limitations, for example, they can be costly, have limited accessibility, and are time-consuming (Sacco et al., 2013; Timpone et al., 2020). Rehabilitation is the primary approach to promote recovery of disabilities caused by stroke, but

the treatment outcomes can vary greatly between individuals due to the heterogeneity of stroke lesion sites, extent and its functional consequences (Campbell & Khatri, 2020). In order to reduce the patient's risk of suffering from preventable post-stroke disabilities, it is imperative to use a personalized treatment plan that is the most effective for prevention and long-term recovery of each patient. One promising technique that has potential to aid in diagnosis and treatment monitoring is the use of fluid biomarkers determined using metabolomics.

Metabolomics, the study of metabolic products or metabolites in living systems, offers a snapshot into the dynamic cellular processes of organs and the overall health of an individual (Nicholson et al., 1999). Examples of metabolites include sugars, hormones, alcohols, amino acids, and organic acids, which can be found in various biological samples such as urine and blood (Vignoli et al., 2019). The present study focuses on urinary metabolomics partly because of the ease and celerity of sample collection, sample preparation and data collection. Urine is a valuable biofluid produced by the kidneys when filtering and removing toxins from blood, making many metabolites highly concentrated and therefore easily detectible. For this reason, urine is ideal for clinical diagnostics and patient monitoring (Bouatra et al., 2013). This amber-coloured liquid contains biomolecules that reflect the physiological state of an individual, including electrolytes, proteins, hormones, and of course, metabolites. These metabolites are downstream from other biological systems such as the (epi)genome, transcriptome and proteome, and can provide valuable insights into current health states. Additionally, since metabolism is perturbed by abnormal cellular processes, metabolomics provides an integrative view of biochemical processes occurring in an organism (Nicholson & Lindon, 2008). Further, metabolites are small enough to traverse the blood-brain barrier, giving them the unique ability to provide insight about metabolic perturbations in the brain during and following stroke (Qureshi et al., 2017), making metabolomics a powerful tool for detecting and monitoring neuropathophysiological states such as stroke.

Nuclear magnetic resonance (NMR) spectroscopy is a technique used to quantify metabolites by detecting the unique biochemical signature of each individual metabolite in a biological sample. The technology works by detecting hydrogen atoms in each metabolite, which generates a unique "chemical shift" fingerprint that can be observed as a single peak or multiple peaks on the NMR spectrum. The NMR spectrum of a biological sample is therefore a composite of the fingerprints of all the molecules present in the sample. NMR is highly reproducible, non-destructive to the sample, non-biased, requires little sample preparation and analysis (Emwas, 2015; Nicholson & Lindon, 2008; Wishart, 2008), and can simultaneously identify a plethora of metabolites involved in various biological processes (Nicholson et al., 1999). When compared to other metabolite quantification techniques, NMR has the greatest number of unique (108) and detectable (209) metabolites in human urine (Bouatra et al., 2013), and detects additional unique (20) and detectable (49) compounds in serum (Psychogios et al., 2011). Similar to MRI and CT, the high cost, restricted accessibility, and equally long clinical time demands of NMR warrant consideration. The significance of this study stems from the capacity of NMR-driven metabolomic biomarkers to transform into cost-effective, rapid assays for biomarker detection that could find utility in a pre-hospital setting (Dagonnier et al., 2021).

Biomarkers, which are indicators of physiological processes and conditions, offer the ability to distinguish between normal and pathological states, evaluate therapeutic efficacy, and prognosticate future disease outcomes (Atkinson et al., 2001). Urinary and serum biomarkers are not routinely used in a clinical setting to diagnose or monitor stroke. Given the time-sensitive nature of stroke therapies, the development of a rapid and accurate biomarker could improve outcomes significantly. Imaging techniques are not always readily available, can sometimes provide a false negative diagnosis, and are not able to provide an accurate prognosis or advise a patient’s therapy. Thus, it is advantageous to have an alternative technique at the ready that can

confirm stroke diagnosis (Sacco et al., 2013) and monitor patients through their treatment. The use of NMR-based metabolomics can aid in the development of personalized treatment for patients, and its resultant biomarkers can be used to detect and monitor disease, and evaluate the effectiveness of treatment programs (Emwas et al., 2013). This powerful tool has previously been used in the development of biomarkers for various neurological conditions, including stroke (Qureshi et al., 2017), spinal cord injury (Bykowski et al., 2021a; Bykowski et al., 2023), traumatic brain injury (Bykowski et al., 2021b), and sport-related concussion (Paxman et al., 2017, Wanner et al., 2021). This NMR-based metabolomics approach is therefore well-equipped to gain information about diagnosis, prognosis, and ongoing treatment in order to better inform clinicians about each individual stroke patient's optimal neurorehabilitation regimens.

Each individual stroke patient is differentially affected by the brain damage and, consequently, they rarely respond to treatments with the same timeline, efficacy, and outcomes. Inter-individual disparities such as these underscores the indispensability of personalized medicine strategies in both the diagnosis and management of stroke. Personalized medicine employs cellular and molecular information, often derived from 'omics' sciences, to diagnose, treat, monitor, and prevent disease. NMR-based metabolomics can create personalized treatment regimens by analyzing a patient's metabolome for biomarkers that confirm stroke diagnosis and predict outcomes. Clinically accessible samples can be taken during treatment to monitor biomarker levels and gauge recovery and neurological progress (Koen et al., 2016). Metabolite biomarkers represent a highly favorable approach for personalized medicine due to their rapidity, reproducibility, cost-effectiveness, non-invasiveness, and notable quantitative precision (Xia et al., 2013). To facilitate the diagnosis and monitoring of stroke in a hospital setting, it is practical to utilize a biomarker panel composed of the 1-10 most significantly altered metabolites, as larger assays are not feasible (Xia et al., 2013). Such a panel of biomarkers, if identified, is more robust than any single marker

and can effectively inform personalized medicine. A biomarker panel may elucidate the underlying cellular mechanisms contributing to individual variability in stroke prognosis, treatment efficacy, and the condition itself (Koen et al., 2016), thus also revealing potential therapeutic targets. As far as we are aware, the current investigation represents the first attempt to identify NMR-based metabolomic biomarkers that could serve as indicators of triage, prognosis, and extent of recovery in stroke patients.

The present proof-of-principle study seeks to use ^{1}H NMR metabolomic profiling combined with univariate statistics and multivariate machine learning to identify novel urinary biomarkers indicative of recovery processes and treatment response in stroke. The study was designed to evaluate the metabolomic profile initially following stroke and 6 months thereafter, with the aim of identifying the specific metabolites responsible for observed differences and the underlying biochemical pathways that contribute to symptoms and recovery. It is hypothesized that there will be significant changes in the urinary metabolic profile following stroke, which will correlate with clinical measures of function and reveal involvement in injury and recovery processes. These altered metabolites will aid in determining injury severity, potential recovery, and treatment efficacy while uncovering significant biological pathways involved in the repair and recovery process after stroke.

3.2 MATERIALS AND METHODS

3.2.1 Stroke Patient Characteristics and Sample Collection

The present study was part of the project called Understanding Neurological Recovery: the role of resting state fMRI, biomarkers, and robotics after traumatic brain injury, stroke, and spinal cord injury (UCAN), aimed at following patient's recovery after stroke up until around 6 months post-lesion. Participants with stroke were recruited from the inpatient stroke unit at the Foothills

Medical Centre, which has a campus from the University of Calgary, where they were approached by their circle of care to participate in the study. After obtaining consent to contact, a researcher from the UCAN team provided the patients with detailed information about the study and obtained their informed consent. A total of 19 stroke patients were recruited for the study, of which 10 provided two urine samples, including 7 patients with ischemic stroke and 3 with hemorrhagic stroke (average age 61.5±13.2 years, Table 1). Fasting morning urine samples, acquired between 6:00am-8:00am, were collected at two time points: initially after stroke within 2-11 days (median=5, interquartile range=2), and again at 6 months post-injury, within 101-242 days (median=203, interquartile range=29.5). For sample collection, patients used an antiseptic alcohol wipe to clean the urethral opening and filled the collection cup approximately halfway with a midstream urine sample. Urine samples were stored at -80°C and sent to the University of Lethbridge for further processing. Pairwise analyses were performed within-subjects to best control for confounding variables such as diet, lifestyle, body mass index, medical history, and acute versus chronic drug treatments.

3.2.2 Clinical Assessments

In this study, initial clinical assessments were taken within 4-32 days (median=14.5, interquartile range=12) and 6-month follow up assessments were completed 98-242 days (median=203, interquartile range=29.5) after stroke. The National Institutes of Health Stroke Scale (NIHSS; Brott et al., 1989) was used at the initial time-point to measure stroke severity using 15 items, including coordination, vision, speech, and sensory and motor symptoms. All items were recorded on a scale from 0-2, 0-3, or 0-4 and summed for a total possible score between 0-42, with a score of 42 indicating the most severe impairment. We also utilized the Chedoke-McMaster Stroke Assessment (CMSA; Gowland et al., 1993) to evaluate physical impairments and disabilities following stroke. The CMSA was conducted within a few days of injury and at 6-month

follow-up, using a 7-point scale to measure impairments and disabilities, with 1 representing paralysis and 7 representing normal movement. The CMSA includes two inventories: the impairment inventory, which categorizes patients based on their motor recovery, and the disability inventory, which measures changes in disability. This study focused on the hand and arm dimensions of the impairment inventory because the upper extremity is largely found to be impaired in patients with mild and moderate stroke severity. Additionally, we used the Functional Independence Measure (FIM; Keith et al., 1987) to assess the level of disability in stroke patients based on motor control, self-care, communication, social skills, and cognitive skills. The FIM was administered both at baseline and at the 6-month follow-up. Functional scores on 18 items were recorded on a 7-point scale, with a total possible score of 126, indicating complete functional independence.

3.2.3 NMR Sample Preparation, Data Acquisition, and Processing

Urine samples were removed from the -80°C freezer and thawed at room temperature. To ensure a consistent pH and to minimize positional noise within the datasets generated by NMR analysis (Gil et al., 2016; Smelter et al., 2017), urine samples were mixed with a buffer prepared in-house. For concocting the metabolomics buffer, dibasic potassium phosphate (K_2HPO_4) and monobasic potassium phosphate (KH_2PO_4) were mixed in a ratio of 4:1 with a concentration of 0.625 M in deionized H_2O. To inhibit microbial growth and to correct for the effect of divalent cations on chemical shift, 3.75mM NaN_3 and 0.375mM potassium fluoride (KF), respectively, were added to the buffer. The buffer solution's pH was adjusted to 7.4 through the titration of 500µL aliquots of 3M HCl and 3M NaOH. Additionally, D_2O containing 0.05% by weight trimethylsilyl propanoic acid (TSP) was used as a chemical shift reference.

Each sample was prepared by adding 160µL of the prepared buffer, 40µL of D_2O containing 0.02709 % weight/volume TSP, and 400µL of urine. Samples were then mixed and

pipetted into a 2mL microfuge tube, centrifuged at 12000rpm for 5 minutes at 4°C, and 550µL of the supernatant was transferred to a 5mm NMR tube for analysis. Using an NMR pH meter, the sample pH was checked to confirm that the sample complied with a pH of 7.4±0.05. This procedure was performed to ensure that the sample was not contaminated and to verify its suitability for the subsequent NMR analysis. The samples were analyzed immediately on an NMR spectrometer.

In this study, we employed a 700 MHz Bruker Avance III HD NMR spectrometer equipped with a room temperature triple resonance broad band observe (TBO) probe to acquire the NMR data. Prior to data acquisition, three-dimensional and one-dimensional shimming experiments were conducted to correct for any inhomogeneities in the static magnetic field. We performed a one-dimensional Bruker ^{1}H nuclear overhauser effect spectroscopy (NOESY) experiment with water suppression pulse sequence 'noesygpr1d' with a mixing time of 10ms to collect the NMR data. The NOESY pulse sequence produces a spectrum that displays signals of metabolites and high molecular weight molecules (Vignoli et al., 2019). The experiment was completed with 128 scans with a total size of 128k data points. The spectra were pre-processed using Bruker Topspin software (version 3.2, patch level 6) by zero filling to 256k points, line-broadening to 0.3Hz, Fourier transformed, and referenced to the TSP peak (0.00δ). Additionally, automatic phasing and baseline correction were applied to consistently enhance the quality of the spectra. Spectra were then exported as ASCII files to be undergo further processing and statistical analysis in MATLAB.

Recursive segment-wise peak alignment (Veselkov et al., 2009) was applied to all spectra in MATLAB to correct for major and minor peak position variations in urine, which ensures accurate identification of metabolites and generation of valid results. Subsequently, dynamic adaptive binning was applied in MATLAB (Anderson et al., 2011) and the bins underwent manual inspection and correction for errors. The regions corresponding to water and urea were removed from the bins to eliminate any interference from these highly variable regions.

Metabolites were identified using a combination of resources, including Chenomx 8.2 NMR Suite, the Human Metabolome Database (HMBD), and the Human Urine Metabolome paper. These resources contained a list of NMR-derived urinary metabolites and their concentrations, which enabled us to identify the metabolites in our samples (Bouatra et al., 2013; Wishart et al., 2018).

3.2.4 Statistical Analysis

Multivariate statistical analysis was utilized to assess whether the urinary metabolite profiles could distinguish between the 1-month and 6-month post-injury samples. Prior to modeling, the data were normalized to the total metabolome (excluding water and urea), pareto-scaled, which involved mean-centering and dividing by the square root of each variable's standard deviation, and log-transformed (Craig et al., 2006; Wiklund et al., 2008; van den Burg et al., 2006; Box & Cox, 1964). A total of 505 bins were created and included in the analysis.

Following this, we employed a linear regression model to remove the confounding effects of age and sex on the metabolomics data. The model was implemented in Python programming language using the widely used StatsModels package, which is an open-source statistical software package (Python Software Foundation, https://www.python.org/). Linear regression models, including simple and multiple linear regression, are commonly used in metabolomics to determine the relationship between metabolites and covariates or to compare metabolite concentrations across groups (Debik et al., 2022). Residuals calculated during linear regression can be used to regress out the effects of confounding factors as a data pre-processing step. Linear mixed models (LMMs) are frequently used in longitudinal metabolomics studies to adjust for dependencies between repeated samples (Debik et al., 2022; Wanichthanarak et al., 2019). Importantly, ordinary least squares (OLS) regression is often used in clinical MRI studies prior to analysis by using the residuals of the confound regression model (Dinga et al., 2020). This method is precisely how the

regression model was used in the present study. The use of the linear regression model allowed us to account for the variability in the data due to age and sex, thereby enhancing the accuracy and precision of our metabolomics analysis.

Along with Variable Importance Analysis based on random Variable Combination (VIAVC), univariate statistical tests, such as paired *t*-test or paired Wilcoxon signed rank test, were employed to analyze the parametric and non-parametric data, respectively. The Shapiro-Wilk test was used to determine if the data for each bin was parametric or not (Goodpaster et al., 2010). VIAVC is a machine learning algorithm implemented in MATLAB that identifies important metabolites in a dataset by using binary matrix resampling (BMR) to generate subsets of variables with equal probability (Yun et al., 2015). These subsets are then used to perform Partial Least Squares-Linear Discriminant Analysis (PLS-LDA) with 10-fold (F-ranked variables) cross validation (CV) and double 10-fold (best subset) cross-validation (DCV) to rank the importance of each metabolite. Based on the metabolite's interaction with other variables, VIAVC categorizes it as strongly informative, weakly informative, uninformative, or interfering. Through many iterations of removing uninformative and interfering variables, VIAVC outputs an F-ranked subset and a VIAVC best subset of informative metabolites, both ranked by Bonferroni-Holm corrected *p*-value. VIAVC algorithms identified the significant metabolites using the Receiver Operator Characteristic (ROC) test and subsequent Area-Under-the-Curve (AUC) analysis (Fawcett, 2006). The F-ranked best subset is created using 10-fold CV, followed by the best subset, which employs 10-fold DCV and permutation testing using 2000 permutations to determine each variable's predictive ability and assess the model robustness (Szymanska et al., 2012). Overall, VIAVC is a reliable technique to identify significant metabolites due to its thorough approach to modeling and validation.

The statistical analysis was performed using the MetaboAnalystR version 2.0.4 package running inside R version 3.5.3 (Pang et al., 2020). The pathway topology analysis was performed using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database for Homo sapiens as the pathway library, and the hypergeometric test was employed for over-representation analysis. Furthermore, relative-betweenness centrality was used for topology analysis (Wishart et al., 2018; Xia & Wishart, 2010). Only the metabolites that were found to be significantly altered in this study were included in the pathway topology analysis. Receiver operator characteristic (ROC) curves and area under the curve (AUC) are widely used methods to evaluate classifier performance and assess the diagnostic ability of a biomarker (Fawcett, 2006). Using MetaboAnalystR, ROC curves were created to visualize and compare the sensitivity and specificity of the biomarkers, while AUC was used as a single metric to summarize the accuracy of the classification. We used ROC and AUC analysis in our study to assess the diagnostic ability of the urinary biomarkers for stroke, and to compare the performance of different biomarkers in our dataset.

To assess the separation between the study groups, we utilized a combination of unsupervised and supervised multivariate methods. First, we conducted a principal components analysis (PCA) to assess the degree of separation between the groups without supervision. In addition, a supervised orthogonal projection to latent structures discriminant analysis (OPLS-DA) was performed, which allowed visualization of the between-group separation as a function of within-group separation (Wiklund et al., 2008). The combination of these methods provided a comprehensive and robust approach for analyzing the data and assessing the separation between the groups.

To evaluate the relationship between the normalized urinary metabolite concentrations and the patients' clinical assessment scores, Spearman rank-order correlations were calculated. Specifically, we computed 3 sets of correlations: first, between initial NIHSS, CMSA, and FIM

scores and initial metabolite concentrations; second, the percent difference of the patients' CMSA and FIM scores and the initial concentrations; third, the percent difference of the patients CMSA and FIM scores and the percent difference of the normalized urinary metabolite concentrations. In order to control for multiple comparisons, we used the Bonferroni correction method to adjust the significance threshold for our analysis. The Bonferroni corrected *p*-value was obtained by dividing alpha< 0.05 by the number of VIAVC best subset bins tested in this study ($n = 8$), giving a threshold of alpha=0.00625. This approach allowed us to obtain a rigorous and clinically relevant set of metabolites, as previously described by Goodpaster et al. (2010). The percent differences for the CMSA and FIM scores at the two different time points were computed as follows:

(6 Months Score-Initial Score)/((6 Months Score + Initial Score)/2)×100%

Given the small sample size, the statistical analysis in this study was performed with a high level of rigor to ensure the accuracy and validity of the results. The use of a comprehensive methodology that combines multivariate and univariate statistical tests, a linear regression model, machine learning algorithms, OPLS-DA, ROC and AUC, pathway topology analysis, and correlation analysis provides a robust approach for the analysis of complex biological data.

3.3 RESULTS

3.3.1 Stroke Patient Characteristics

Table 1 present's clinical data on patients with stroke collected initially and at 6-months follow up for various clinical parameters and highlights differences in age, sex, stroke type, medications, and co-morbidities among patients. The table also summarizes information about the impact of stroke on impairments, disabilities, and functional independence, as evidenced by changes in NIHSS, FIM, CMSA-arm, and CMSA-hand scores. NIHSS scores (average 5.56±3.64) demonstrate 5 stroke cases with mild stroke severity, with NIHSS scores 4 or less and 4 cases with

NIHSS scores from 5 to 15, indicating moderate stroke severity. The majority of participants exhibited an improvement in clinical scores for FIM, CMSA-arm, and CMSA-hand, with an average improvement of 22.67±14.13, 1.10±1.29, and 0.70±0.67, respectively. This improvement in clinical scores indicates recovery from stroke, with recovery being defined in this paper as the change in metabolite concentration and clinical parameters from initial collection to 6-months post-stroke.

3.3.2 Metabolomic Analysis of Urine Samples

To determine significant metabolic changes in urine, data analysis used VIAVC and paired *t*-test or Wilcoxon signed rank tests. The results of this analysis are presented in Table 2. Out of the 505 bins analyzed, 8 were significant by VIAVC, 8 were significant via paired *t*-test or Wilcoxon signed rank, and 1 bin containing pseudouridine was significant by both.

OPLS-DA scores plots (Figure 2) reveal excellent group separation when comparing urine samples for acute stroke to 6 months post-stroke samples. Panel A was performed on a subset of 8 metabolites significant by VIAVC best subset, while panel B was created using a subset of 15 metabolites significant by paired *t*-test or VIAVC best subset. The VIAVC best subset was also used in the creation of the ROC and associated AUC in Figure 3 and illustrates high sensitivity, specificity, and robustness of the model, with an AUC of 0.929, a 95% confidence interval of 0.778-1, and predictive accuracy of 85%.

The urinary metabolites identified as significant based on VIAVC best subset or paired *t*-test demonstrated a large impact and significance on their respective pathways, as evidenced by Figure 4. Notably, phenylalanine metabolism (p=0.002) displayed the highest significance and the metabolites within the phenylalanine, tyrosine, and tryptophan biosynthesis (p=0.031) pathway resulted in the highest pathway impact. The tyrosine metabolism (p=0.003), purine metabolism (p=0.011), and glycerophospholipid metabolism (p=0.030) pathways were also found to be

significantly altered from initial stroke to 6 months post-stroke. Pathway analysis is used to provide insight into the metabolic changes occurring after stroke and highlight potential targets for future therapeutic interventions or biomarkers for stroke diagnosis and treatment.

3.3.3 Correlation of Metabolomic Signatures Linked to Motor Recovery

Spearman correlations between significant metabolites and the CMSA-hand scores were conducted and the significant results are listed in Table 3 and further visualized in Figure 5. Pseudouridine, the urinary metabolite with the topmost VIAVC best subset significance, illustrated a strong correlation to the CMSA-hand scores in two separate correlation analyses: the initial metabolite concentration correlated to the percent difference in CMSA-hand scores, and the percent difference in metabolite concentration correlated to the percent difference in CMSA-hand scores. The former displays a positive correlation and the latter yields a negative correlation. None of the correlations passed the Bonferroni corrected threshold of alpha=0.00625. The significant spearman correlation values for pseudouridine and CMSA-hand scores suggest a potential relationship between the urinary metabolite and motor recovery following stroke.

3.4 DISCUSSION

The present study provides novel insights into metabolic and functional changes that occur following a stroke. Specifically, this proof-of-principal study pursued a characterization of metabolites and their associated metabolic pathways linked to stroke over time, revealing a close relationship between pseudouridine and motor recovery. Metabolites that were part of the VIAVC best subset displayed a high level of sensitivity and specificity along with a predictive accuracy of 85%, thereby demonstrating potential in serving as a biomarker or biomarker panel for distinguishing between acute stroke and 6 months post-stroke (Figure 2 and Figure 3). Furthermore, the pathways that were most significantly altered between the two time points

include phenylalanine metabolism, tyrosine metabolism, purine metabolism, glycerophospholipid metabolism, and phenylalanine, tyrosine, and tryptophan biosynthesis (Figure 4). Additionally, significant correlations were observed between the urinary metabolite pseudouridine and CMSA-hand scores, which may be indicative of motor recovery following stroke (Table 3 and Figure 5). Thus, urinary pseudouridine levels may have predictive value for CMSA-hand scores. These findings provide new avenues for understanding the underlying biological mechanisms that contribute to post-stroke recovery and may lead to the development of targeted therapeutic interventions and the improvement of clinical management of stroke patients. In the following discussion, we will examine these findings in greater detail and discuss their implications for future research and clinical practice.

3.4.1 Clinical Translation and Classification of Stroke Metabolites

OPLS-DA scores plots provide an effective means for visualizing the separation of the supervised groups and the significance of the metabolite subsets in predicting post-stroke outcomes. OPLS-DA scores plots seen in Figure 2 demonstrate that there are significant differences in the metabolomic profiles of individuals immediately after stroke compared to those at 6 months post-stroke, and that these differences can be detected using multivariate statistical analysis. For both panels A and B, cross validation and permutation measures indicate that the models have good predictive ability and that the discrimination between the groups is statistically significant, with the VIAVC best subset metabolites providing the best classification. The subsets of metabolites used for the analysis were determined using different statistical methods, suggesting that the differences between the two groups are robust and not dependent of the specific method used for analysis. The ROC and AUC in Figure 3 highlights the diagnostic accuracy of this metabolite panel for distinguishing between initial and 6 month post-stroke samples. An AUC of 1.0 represents perfect classification, while an AUC of 0.5 represents random guessing and an AUC

of less than 0.5 represents worse than random classification (Fawcett, 2005). The ROC curve has an AUC of 0.929 and demonstrates a high level of sensitivity and specificity, with a 95% confidence interval of 0.778-1 indicating strong predictive ability of the model and that the results are statistically significant. The predictive accuracy of the model was 85%, indicating that the metabolite panel can correctly classify 85% of samples into their respective groups. These findings suggest that the metabolites identified through the VIAVC best subset analysis can serve as potential biomarkers for distinguishing between acute stroke and post-stroke samples.

3.4.2 VIAVC-based Metabolomic Signatures as Reliable Biomarkers

The metabolites significantly identified by VIAVC revealed a robust profile throughout the results. Several noteworthy metabolites infer increased demand of nucleotide building blocks for deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) synthesis and repair following stroke. Notably, pseudouridine is a modified nucleoside involved in RNA stability and translation (Li et al., 2016). Its downregulation (p=4.55E-23) in urine after stroke may reflect changes in RNA metabolism after injury and during recovery. Similarly, adenosine, a purine nucleoside that increases under stress (Fredholm, 2007), and inosine, a modified adenosine in RNA, were significantly downregulated (p=7.82E-16 and p=1.59E-16, respectively), indicating that levels were raised initially in response to stress or stroke and then lowered during the recovery process. Previous research has demonstrated that inosine promotes axon growth following stroke, suggesting that the metabolites that are altered here may denote alterations in gene expression (Chen et al., 2002; Zai et al., 2009). Further, the deoxyribonucleoside deoxyinosine in DNA (compared to inosine, its equivalent in RNA) was significantly upregulated (p=3.41E-13). Given that adenosine and inosine can be converted into deoxyinosine, their individual regulations may indicate that adenosine and inosine levels dropped as they were being used for DNA synthesis and repair and therefore broken down into deoxyinosine. Overall, there were multiple metabolites

displaying significant regulation changes that are related to gene expression and alterations in DNA and RNA synthesis and repair.

On the other hand, several significant metabolites indicated dysregulation of neurotransmitters following stroke. For instance, several of the identified metabolites are involved in catecholamine synthesis and metabolism. 4-hydroxy-3-methoxymandelate, otherwise known as vanillylmandelic acid (VMA), and homovanillate or homovanillic acid (HVA) are both metabolites involved in the metabolism of catecholamines and were found to be significantly upregulated (p=8.19E-19 and p=7.57E-16, respectively). Both of these catecholamine metabolites are major metabolite end-products, with VMA being the final product of norepinephrine and epinephrine metabolism and HVA being the urinary end-product of dopamine metabolism (Eisenhofer et al., 2004). The upregulation of VMA and HVA in urine after stroke may indicate increased activation of the central nervous system leading to increased breakdown of catecholamines and these metabolites have previously been identified as prognostic biomarkers for stroke (Bonifačić et al., 2017). Interestingly the precursor for catecholamine synthesis, tyrosine (Molinoff & Axelrod, 1971), was significantly downregulated (p=0.028) in stroke patients. Downregulation of tyrosine could lead to decreased production of catecholamines, leading to lower levels of tyrosine being used for their synthesis. These results, which have been corroborated by previous research, illustrate that during stress or stroke, more tyrosine is needed and catecholamines become depleted (Ormstad et al., 2013; Palkovits et al., 1975). Additionally, the levels of the neurotransmitter acetylcholine were reduced over the course of this study (p=0.021). Ethanolamine is a potential precursor for acetylcholine synthesis (Haidar et al., 1994) and its levels were found to be elevated over the course of recovery (p= 5.08E-15). Moreover, acetylcholine and ethanolamine are also involved in glycerophospholipid metabolism, as discussed below. Overall, the metabolomics analysis revealed significant alterations in several metabolites involved in

various physiological processes in stroke patients, suggesting their potential role in the pathophysiology of stroke.

3.4.3 Metabolic Pathways Involved in Stroke Recovery Processes

Phenylalanine Pathways

Phenylalanine metabolism (p=0.002) was significantly altered between the acute phase of stroke and 6 months post-stroke. Previous studies indicate that phenylalanine metabolism is up-regulated in the acute phase after stroke (Sidorov et al., 2020). Feeding into this metabolic pathway and the tyrosine metabolism pathway is the phenylalanine, tyrosine, and tryptophan biosynthesis pathway (p=0.031), which was also found to be significantly altered. Previous studies indicated that phenylalanine, tyrosine, and tryptophan biosynthesis is significantly affected in acute ischemic stroke (Qi et al., 2022). Accumulating evidence links both of these phenylalanine pathways to alterations in post-stroke depression (Chen et al., 2021; Liu et al., 2022). Phenylalanine exhibits multiple metabolic pathways within the body, including its incorporation into proteins, conversion into phenylpyruvic acid, and conversion into tyrosine. High levels of phenylalanine and its derivatives, including phenylacetic acid, exacerbates oxidative stress, elicits lipid peroxidation (Fernandes et al., 2010), and impairs synaptogenesis in rodents and humans (Sanayama et al., 2011). Oxidative stress is a well-known contributing factor to stroke and phenylalanine may play a role in its progression. Interestingly, during lipid peroxidation, reactive oxygen species (ROS) attack and cause damage to glycerophospholipids. Furthermore, lipid peroxidation can have a cascading effect when its products interact with proteins, DNA, and other lipids, leading to additional oxidative damage and disruption of cellular functions. Even alone, phenylalanine and its metabolites can damage DNA by attacking purine and pyrimidine bases and deoxyribose sugar (Sitta et al., 2009). Damage to glycerophopholipids and DNA alterations are further discussed below. It is imperative to consider that the discussed studies focused on high levels of

phenylalanine akin to those in phenylketonuria, and different phenylalanine levels may yield different effects. Regardless, the significant alteration in phenylalanine metabolism, its metabolites, and its association with the biosynthesis pathway highlight the potential role of phenylalanine in stroke progression.

Tyrosine Metabolism

The present stroke study reveals that tyrosine metabolism was potentially disturbed (p=0.003). Prior research corroborates this finding, showing that tyrosine metabolism is impacted during cerebral infarction (Kong et al., 2022). As discussed previously, tyrosine is the precursor for catecholamine synthesis (Molinoff & Axelrod, 1971) and is significantly downregulated over the course of this study. During stress, there is an increased demand for tyrosine, which could lead to depletion of tyrosine and hence decreased production of catecholamines. HVA and VMA are end-products of catecholamine metabolism and their upregulation seen here could be due to increased breakdown of catecholamines after stroke. The metabolite tyrosine is frequently found to be significantly altered in stroke (Ormstad et al., 2013; Sidorov et al., 2020) and post-stroke depression (Chen et al., 2021; Liu et al., 2022; Xie et al., 2020). While Ormstad et al. (2013) argue that tyrosine levels decrease after stroke, other findings suggest that the concentration of tyrosine is increased in the acute phase (Sidorov et al., 2020). This discrepancy could be due to differences in methodology or sample collection times. More research is needed to decipher the complexity of the tyrosine and catecholamine response after stroke.

Purine Metabolism

Purine metabolism is another pathway that was potentially changed after stroke (p=0.011). Earlier investigations implicated purine metabolism in ischemic stroke (Ke et al., 2019; Sun et al., 2017; Yu et al., 2021). For instance, purine metabolites, such as inosine and adenosine, have proven to be useful in determining stroke severity irrespective of stroke type (Dale et al., 2019).

Researchers have proposed that the neurotoxic effects of oxidative stress and inflammation from stroke are offset by intermediates of purine metabolism (Abbracchio & Burnstock, 1998; Ke et al., 2019). Purines such as adenosine 5′-triphosphate (ATP), adenosine, and inosine are rapidly released following injury such as stroke to minimize damage and support healing (Abbracchio & Burnstock, 1998; Bell et al., 1998) and during these hypoxic conditions, ATP is metabolized into adenosine, and further, inosine (Barsotti & Ipata, 2004). Microglial cells are recruited to the injury site where they are activated by purines to aid repair by releasing inflammatory cytokines (Burnstock, 2016). Adenosine is neuroprotective after stroke owing to its numerous abilities, including controlling blood flow, maintaining homeostasis, regulating the immune response, reducing respiration, preventing breakdown of triacylglycerols, and inhibiting the production of neurotoxic pro-inflammatory cytokines (Fredholm, 2007; Fredholm et al., 2011; Mayne et al., 2001; Williams-Karnesky & Stenzel-Poore, 2009). Similarly, inosine has also displayed anti-inflammatory and neuroprotective effects after stroke (Haskó et al., 2004). Inosine has shown potential in neural repair after CNS injury (Benowitz et al., 1999) and stroke through its promotion of axon growth and alteration of gene expression (Chen et al., 2002; Zai et al., 2009). Moreover, ATP is released after injury as a co-transmitter with acetylcholine (Potter & White, 1980) and catecholamines (Poelchen et al., 2001), and both of these neurotransmitters and their associated metabolites were also significantly altered and are discussed in this section. It is important to note adenosine and inosine are involved in a wide variety of cellular processes, such as being precursors to DNA and RNA as previously discussed. These metabolites may simultaneously be used to synthesize and repair damaged DNA caused by the brain injury. Taking all of this into consideration, the purines are diverse metabolites that play numerous roles in the pathophysiological response to brain injury.

Glycerophospholipid Metabolism

Glycerophospholipid metabolism (p=0.030) is another possible pathway that is perturbed following stroke, as confirmed by other studies investigating stroke (Chen et al., 2023; Farooqui et al., 2000; Ke et al., 2019; Liu et al., 2021; Sun et al., 2017; Qi et al., 2022) and post-stroke depression (Jiang et al., 2022). Glycerophospholipids play diverse roles in neural membranes, providing structure, maintaining homeostasis, and facilitating ion permeability. Upon injury, membrane glycerophopsholipids may become damaged, and the subsequent neuroinflammatory response can lead to the breakdown of glycerophospholipds into lipid mediators that regulate neuroinflammation, oxidative stress, neural cell proliferation, differentiation, and apoptosis (Farooqui et al., 2007). Ethanolamine is a precursor for acetylcholine (Haidar et al., 1994) and phospholipids and previous research strengthens our claim that it is increased following stroke (Smart et al., 1994). Studies have shown that dendrite growth is dependent on the glycerophospholipid phosphatidylethanolamine, which is derived from ethanolamine (Meltzer et al., 2017). Since neural expansion and reformation is needed after the damage caused by stroke, this could potentially be a mechanism promoting recovery. Further, compelling evidence indicated that the α-7 nicotinic acetylcholine receptor (α-7 nAChR) plays a role in neuroprotection by reducing neuroinflammation and oxidative stress in stroke (Han et al., 2014). Activation of the α-7 nAChR by acetylcholine or other agonists has demonstrated beneficial effects in experimental models of stroke through the activation of downstream signalling mechanisms that can mitigate the damaging effects of stroke. Given that acetylcholine activates many other receptors and has a complex impact on the brain, further research is needed to fully understand the neuroprotective role of acetylcholine and its potential therapeutic applications in stroke. All encompassing, the perturbation of glycerophospholipid metabolism is evident following stroke and the potential neuroprotective role of ethanolamine and acetylcholine warrant further investigation to deepen the understanding of their therapeutic applications in stroke.

3.4.4. The Relationship of Metabolic Biomarkers to Clinical Parameters

Metabolic Signatures to Predict Stroke Patient Outcomes

Pseudouridine is a modified nucleoside that is a component of RNA and its metabolite is associated with RNA stability and translation (Li et al., 2016). Post-transcriptionally regulating RNA expression involves modifications after transcription and synthesis. Among these modifications, pseudouridine stands out as the most prevalent post-transcriptional alteration (Maden, 1990). Decreased levels of pseudouridine in urine may suggest impaired RNA synthesis or degradation, which may affect the cells' ability to respond to injury and recover. Pseudouridine has been reliably established as a biomarker for post-stroke depression (Zhang & Zhang, 2015) and ischemic stroke, even after accounting for risk factors (Ament et al., 2022). In this study, the urinary metabolite pseudouridine showcases potential in prognosticating stroke patient outcomes through a significant Spearman correlation ($R = -0.735$, $p=0.015$). The pseudouridine concentration during acute stroke injury was correlated to the CMSA-hand score percent difference. This correlation resulted in a negative relationship, indicating that decreased levels of urinary pseudouridine were linked with greater improvement in CMSA-hand scores. This preliminary finding illustrates that pseudouridine, a prevalent post-transcriptional modification in RNA, holds promise as a prognostic indicator for stroke, with lower levels correlating with improved stroke patient outcomes.

Metabolic Signatures to Monitor Stroke Patient Treatment Efficacy

Pseudouridine proved to be significant once again, displaying promise for guiding stroke recovery due to a significant Spearman correlation ($R = 0.703$, $p=0.023$). In the previous section, the involvement of pseudouridine in RNA synthesis and degradation was outlined, along with its association with stroke (Ament et al., 2022). In the case presented here, the percent difference in pseudouridine concentration from initial sample collection to 6-month sample collection was

correlated to the percent difference in CMSA-hand scores, taken at the same time-points. The correlation resulted in a positive relationship, suggesting that a greater increase in pseudouridine concentration initially to the 6 months' time-point is associated with a significant improvement in CMSA-hand scores. Thus, monitoring pseudouridine levels not only shows potential for stroke prognosis, but also in guiding stroke patient's treatment efficacy. These findings highlight the clinical significance of pseudouridines potential as a valuable marker for assessing therapeutic effectiveness and predicting functional recovery following stroke.

3.5 CONCLUSION AND LIMITATIONS

In conclusion, this pilot study provides valuable insights into the pathophysiological changes that occur following stroke and proposes pseudouridine as a marker for stroke prognosis and intervention success. Specifically, the results demonstrate that metabolomic analysis can identify specific metabolites and pathways that are significantly altered following a stroke, thereby shedding light on the underlying biological mechanisms involved in post-stroke recovery. The study identified a panel of metabolites that showed high sensitivity and specificity in distinguishing between acute stroke and post-stroke samples, indicating their potential as biomarkers for stroke diagnosis. Furthermore, the study revealed significant correlations between urinary pseudouridine levels and CMSA-hand scores, thereby emphasizing pseudouridine as a valuable tool in stroke management.

The findings highlight the importance of metabolic pathways in stroke recovery processes. The study identified significant alterations in phenylalanine metabolism, tyrosine metabolism, purine metabolism, glycerophospholipid metabolism, and phenylalanine, tyrosine, and tryptophan biosynthesis. These pathways are involved in various physiological processes, such as gene expression, neurotransmitter synthesis and metabolism, oxidative stress, and neuroinflammation.

The dysregulation of these pathways may contribute to the pathophysiology of stroke and provide potential targets for therapeutic interventions.

Moreover, the study provides evidence for the reliability and clinical translation of metabolomic signatures as biomarkers for stroke. The OPLS-DA scores plots demonstrated the ability to distinguish between acute stroke and 6 months post-stroke samples using multivariate statistical analysis. The VIAVC metabolite panel showed a high level of diagnostic accuracy, sensitivity, and specificity in classifying the samples. In addition, the urinary metabolite pseudouridine demonstrated clinical importance as a biomarker prognosticating stroke outcomes and monitoring treatment success. These findings have implications for improving stroke diagnosis and clinical management, as well as for the development of targeted therapeutic interventions.

Overall, this comprehensive analysis of metabolomic changes in stroke patients contributes to the understanding of post-stroke recovery mechanisms and opens new avenues for future research and clinical practice. The identified metabolites and pathways provide potential targets for therapeutic interventions and biomarker development, ultimately aiming to improve the outcomes and quality of life for stroke patients. Further research is needed to validate these findings and explore their applicability in larger patient populations.

There are several limitations with this research that must be addressed. First, the limited sample size and diversity in confounding factors such as differences in diet, exercise, mobility, body mass index, injury level, acute versus chronic drug treatment, and medical history need to be considered. In addition to this, a more diverse sample group with an equal number of females would be beneficial to this research. Another limitation to the present investigation is the lack of a baseline or control group such as a musculoskeletal injury group that allows comparison to pathophysiological changes after injury in general. In the absence of a musculoskeletal injury

control group, discerning alterations specifically attributed to stroke from those stemming from general injury becomes challenging. Future studies would benefit from taking these factors into consideration when developing the study design. The longitudinal design of this study helps mitigate many of these confounds by pairwise analysis of the samples, ensuring patients are being compared to themselves only. Moreover, the linear regression model accounted for variations in age and sex to guarantee that these variables did not sully the results. Regardless, these results have important implications for stroke research and treatment and warrant further investigation into the role of urinary metabolites in post-stroke metabolic changes and recovery. Overall, the pilot study sheds light on the metabolomic changes associated with stroke and their potential role in the pathophysiology of stroke, as well as the potential of these metabolites as biomarkers for stroke diagnosis and treatment.

Table 1. Stroke patient clinical and demographic characteristics of n=10 stroke patients indicating the stroke type, vascular territory, side of the brain affected by stroke, the days between stroke and both initial and 6-month sample collection, age, sex, medications, co-morbidities, and various stroke scale scores. These clinical assessments include initial NIHSS scores along with initial and 6-month FIM, CMSA-arm, and CMSA-hand scores.

Patient Code	Stroke Type	Vascular Territory	Affected Side	Sex	Age	Collection Days Post-ST		Medications	Co-Morbidities	NIHSS	FIM		CMSA-Arm		CMSA-Hand	
						Initial	6 Month				Initial	6 Month	Initial	6 Month	Initial	6 Month
ST-01	Ischemic	MCA	Left	M	79	2	242	ASA, Atorvastatin, Clopidogrel, Docusate Sodium, Perindopril	A Fib, Acute Renal Failure, Motor Cycle Accident 1952 - Knocked out; Right Collar Bone Fracture, Left Leg injury d/t combine accident as a teen	NaN	109	123	5	7	5	6
ST-03	Ischemic	LACUNAR	Right	M	37	5	221	ASA, Synthroid, Rosuvastatin, HCTZ, Felodipine	Hypothyriodism, HTN, Smoker (1/2 ppd x 5 years), EtOH	11	92	120	1	3	1	2
ST-05	Ischemic	MCA	Left	M	47	6	206	ASA, Crestor	HTN, Smoker (30 per day)	1	116	125	7	7	5	7
ST-06	Ischemic	MCA	Left	M	64	4	101	Warfarin, Solatol, Metoprolol, Atorvastatin, Levothyroxine, Vit D, Calcium, Magnesium,	A Fib, Valve Disease - Dilated Cardiac Myopathy 1990, irregular heart rate in 2002, 2016, Cardiac	3	105	115	6	5	5	5

								Benzaclin Pump	MRI done, no clot found.							
ST-08	Ischemic	MCA/ACA	Left	M	62	4	200		Diabetes, Smoker, Increased Cholesterol	2	106	124	5	7	6	7
ST-09	HEM	MCA	Left	F	61	5	191	Gravol, Gabapentin, Lovenox, Lantus, Norvasc, Pantoprazole, Restorlax, Aidactone, Vit D, Ativan, Advair	HTN, Diabetes, Asthma, Barrett's Esophagus, Obstructive Sleep Apnea, Chronic Neck and Low Back Pain - Recurring PRP Treatment	9	80	124	4	7	5	6
ST-10	Ischemic	Thalamus	Left	M	72	6	189	ASA, Diamicron, Metoprolol, Fosinopril, gliclazide, plavix, metformin, atorvastatin, Vit D, Lantus, Humulin, Drug Study Rivanrobran vs Placebo)	HTN, Diabetes, Hyperlipidemia, Ischemic heart disease, CABG (Aug 2015), Post CABG enrolled in COMPESS, TIA on 2017-01-11 sent to stroke prevention clinic on 2017-01-13 and then admitted to acute stroke unit)	4	96	122	4	6	5	6
ST-16	HEM	MCA	Left	M	62	8	213	Nicotine, Amlodipine, Baclofen, Acetominaphen	HTN, Hyperlipidemia, Smoker, Prostate Cancer, Degenerative Changes spine, Chronic Sinusitis	10	70	115	4	5	5	5
ST-17	Ischemic	MCA + ICA	Left	M	53	11	221			6	NaN	126	6	6	6	6

ST-19	HEM	PCA	Left	F	78	4	170	Amlodipine, Enoxaparin, Hydrochlorothiazide, Nitro patch Polythylene Glycol	HTN, Hyperlipidemia, Cleft Palate	4	113	123	7	7	7	7

Abbreviations: HEM = hemorrhagic, MCA = middle cerebral artery, ACA = anterior cerebral artery, PCA = posterior cerebral artery, ICA = internal carotid artery, ASA = acetylsalicylic acid, HCTZ = Hydrochlorothiazide, PRP = platelet-rich plasma, CABG = coronary artery bypass graft, HTN = hypertension, TIA = transient ischemic attack

Table 2. Comprehensive summary of the significant urinary stroke metabolites arranged in descending order of VIAVC significance, followed by significance of the paired *t* or Wilcoxon signed rank test. The metabolites are identified by chemical shift in parts per million (ppm) along with their corresponding regulation. Metabolites with more than one resonance peak identified as significant are denoted as metabolite.1, metabolite.2, … metabolite.n.

Metabolite	Chemical Shift (ppm)	VIAVC *p*-Value	Paired *t*/ Wilcoxon *p*-Value	Regulation
Pseudouridine.1	4.299	4.55E-23	0.0020 (W)	Down
4-Hydroxy-3-Methoxymandelate	3.887	8.19E-19	0.3567	Up
Inosine	8.219	1.59E-16	0.1496	Down
Homovanillate	3.874	7.57E-16	0.5282	Up
Adenosine	4.315	7.82E-16	0.269	Down
2-Aminobutyrate	0.982	1.27E-15	0.3256	Down
Ethanolamine	3.156	5.08E-15	0.7229	Up
Deoxyinosine	6.494	3.41E-13	0.6182	Up
Phenylacetic acid.1	7.328	Not Sig.	0.0124	Down
Phenylacetic acid.2	7.312	Not Sig.	0.0132	Down
Acetylcholine	2.157	Not Sig.	0.021	Down
L-Tyrosine	6.907	Not Sig.	0.0281	Down
Anserine	7.138	Not Sig.	0.0362	Down
Pseudouridine.2	4.29	Not Sig.	0.0137 (W)	Down
Alanine	1.493	Not Sig.	0.0371 (W)	Up

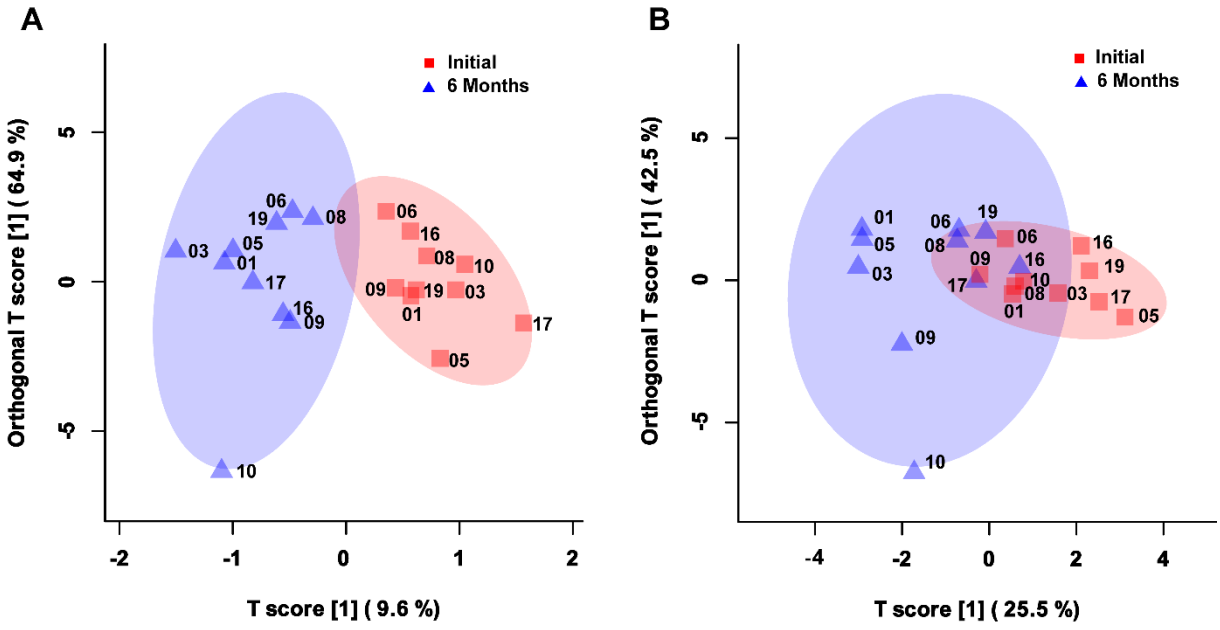

Figure 2. Orthogonal projections to latent structures discriminant analysis (OPLS-DA) scores plots illustrating supervised group separation initially after stroke (red squares) and at 6-months post-stroke (indigo triangles). The x-axis represents the predictive variation between groups and the y-axis displays the orthogonal variation within groups. (A) This scores plot uses the subset of metabolites determined to be significant by VIAVC best subset analysis, consisting of 8 of the 505 bins. The cross-validation and permutation measures were Q^2=0.775 and p<5e-04 for the former, and R^2Y=0.836 and p<5e-04 for the latter. (B) This scores plot was created using the subset of metabolites determined to be significant by paired *t*-test and VIAVC best subset analysis, consisting of 15 of the 505 bins. The cross-validation and permutation measures are Q^2=0.353 and p=0.009 for the former, and R^2Y=0.582 and p=0.01 for the latter.

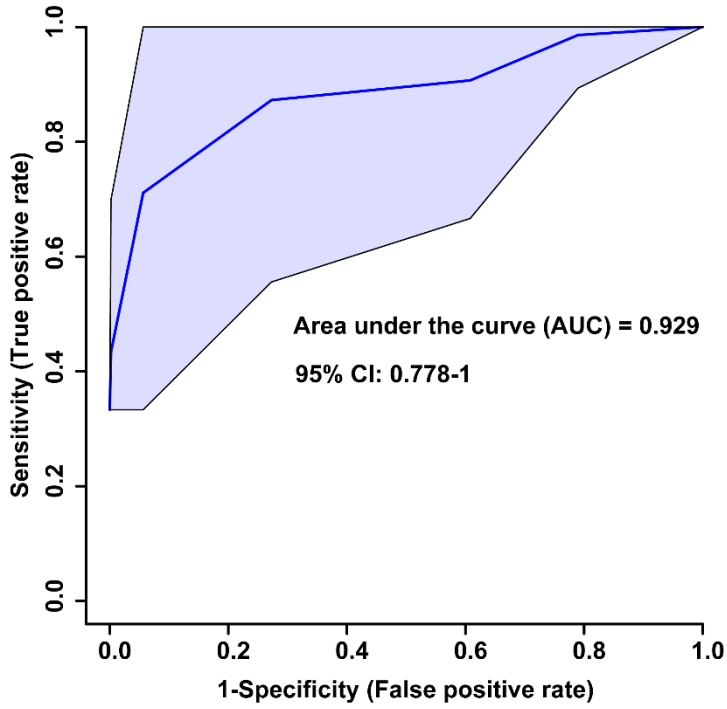

Figure 3. Receiver operator characteristic (ROC) curve displaying a high level of sensitivity and specificity for the separation of initial stroke and 6-months post-stroke samples. Robustness of the model is indicated by the area under the curve (AUC) of 0.929, a 95% confidence interval of 0.778-1 and a predictive accuracy of 85%. This figure was generated using the set of metabolites determined to be significant by VIAVC best subset analysis, which comprises of 8 of the 505 bins.

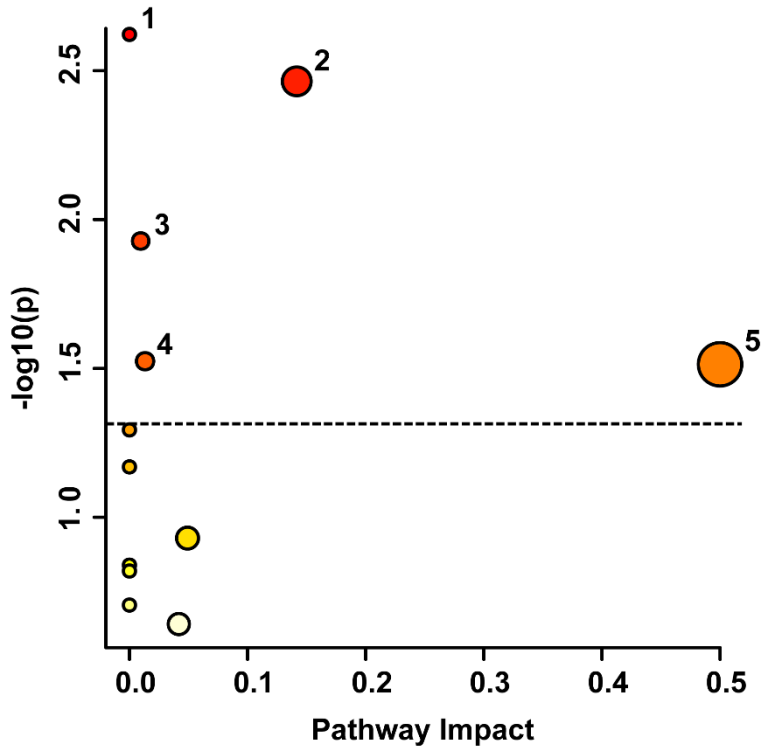

Figure 4. Metabolomic pathway topology analysis for the comparison of initial stroke to 6 months post-stroke using metabolites identified as significant by either VIAVC best subset analysis and a paired *t*-test, corresponding to 15 of the 505 bins, to generate this figure. The figure displays a visualization of metabolic pathways, with the y-axis indicating *p*-values for each pathway, and the x-axis representing the pathway impact, which reflects the extent significant metabolites are able to alter pathway function. The circles represent each pathway, with circle colour indicating degree of significance, and circle size indicating pathway impact. The labelled pathways with $p<0.05$ (significance threshold indicated by the dotted line) and highest pathway impact are highlighted and numbered as follows: 1. Phenylalanine Metabolism ($p=0.002$), 2. Tyrosine Metabolism ($p=0.003$), 3. Purine Metabolism ($p=0.011$), 4. Glycerophospholipid Metabolism ($p=0.030$), and 5. Phenylalanine, Tyrosine and Tryptophan Biosynthesis ($p=0.031$).

Table 3. Significant spearman correlation values for the urinary metabolite pseudouridine and the Chedoke-McMaster Stroke Assessment of the hand (CMSA-hand) scores. The correlation coefficients (R) and associated *p*-values are present for two distinct correlation analyses: (1) initial metabolite concentration correlated to the percent difference in the CMSA-hand scores and (2) the percent difference in pseudouridine concentration correlated to the percent difference in the CMSA-hand scores. For the first comparison, a negative relationship indicates that lower levels of pseudouridine correlate to a greater percent difference, or improvement, in CMSA-hand scores. For the second comparison, a positive relationship reveals that a greater increase in metabolite concentration from the initial time point to 6-months post-stroke correlates to a larger percent difference, or improvement in CMSA-hand scores.

Metabolite	Correlation Values
Initial Metabolite Concentration to Percent Difference CMSA-hand	
Pseudouridine	R = −0.735, p = 0.015
Percent Difference Metabolite Concentration to Percent Difference CMSA-hand	
Pseudouridine	R = 0.703, p = 0.023

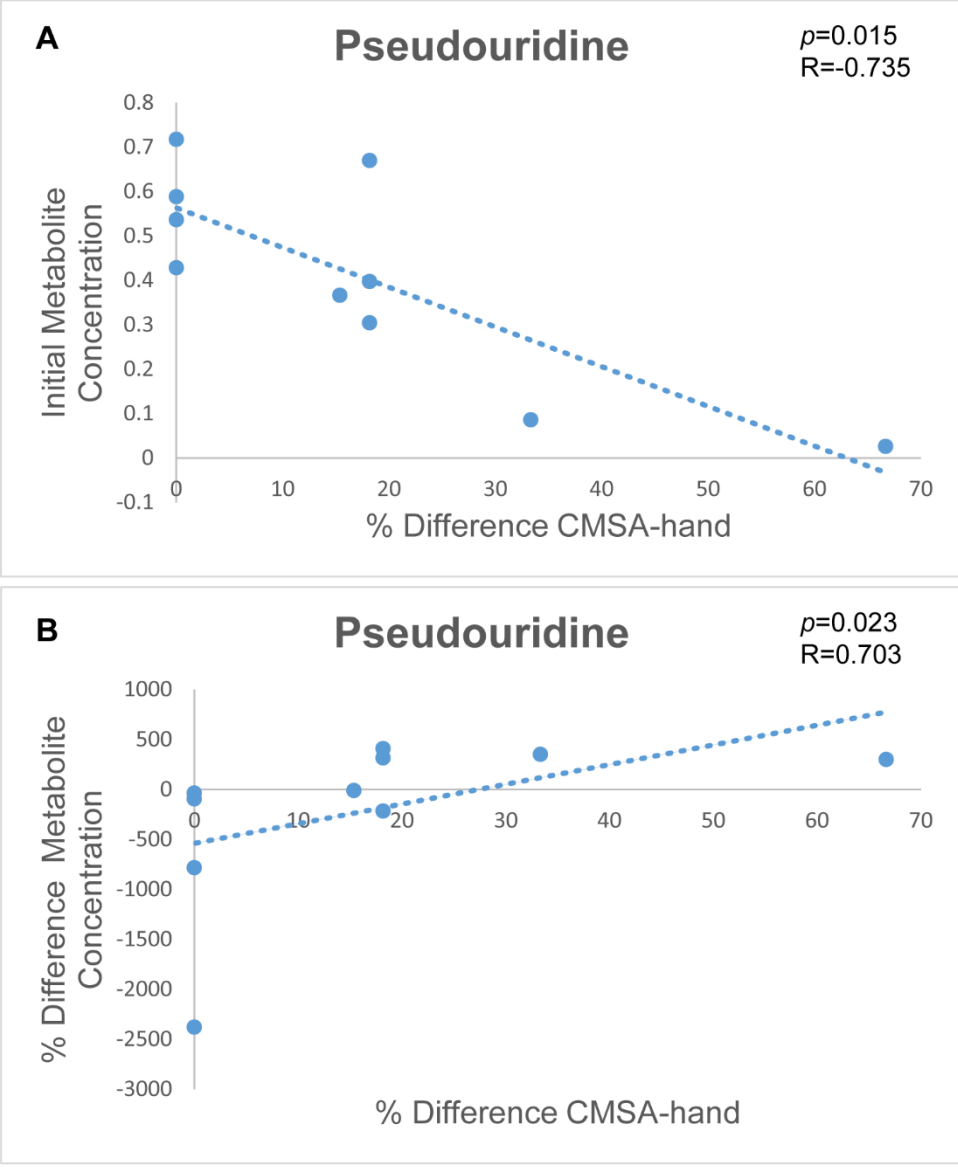

Figure 5. Scatterplots for visualization of significant spearman correlations and corresponding correlation coefficients (R) and *p*-values for two distinct correlation analyses: (A) initial pseudouridine concentration correlated to the percent difference in the CMSA-hand scores and (B) the percent difference in pseudouridine concentration correlated to the percent difference in the CMSA-hand scores. The x-axis illustrates the percent difference in CMSA-hand scores, while the y-axis provides (A) the initial pseudouridine concentration or (B) the percent difference in pseudouridine concentration.

REFERENCES

Abbracchio M., P., & Burnstock, G. (1998). Purinergic signalling: Pathophysiological roles. Japanese Journal of Pharmacology, 78(2), 113-145. https://doi.org/10.1254/jjp.78.113

Ament, Z., Patki, A., Chaudhary, N., Bhave, V. M., Garcia Guarniz, A., Gao, Y., Gerszten, R. E., Correa, A., Judd, S. E., Cushman, M., Long, D. L., Irvin, M. R., & Kimberly, W. T. (2022). Nucleosides associated with incident ischemic stroke in the REGARDS and JHS cohorts. *Neurology*, *98*(21), e2097. https://doi.org/10.1212/WNL.0000000000200262

Anderson, P., Mahle, D., Doom, T., Reo, N., DelRaso, N., & Raymer, M. (2011). Dynamic Adaptive Binning: an Improved Quantification Technique for NMR Spectroscopic Data. *Metabolomics, 7*(2): 179-190. https://link.springer.com/article/10.1007/s11306-010-0242-7 Accessed January 15, 2020.

Atkinson, A. J., Colburn, W. A., DeGruttola, V. G., DeMets, D. L., Downing, G. J., Hoth, D. F., Oates, J. A., Peck, C. C., Schooley, R. T., Spilker, B. A., Woodcock, J., Zeger, S. L., & Biomarkers Definitions Working Group. (2001). Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clinical Pharmacology and Therapeutics, 69*(3), 89-95. https://doi.org/10.1067/mcp.2001.113989

Barsotti, C., & Ipata, P. L. (2004). Metabolic regulation of ATP breakdown and of adenosine production in rat brain extracts. *The International Journal of Biochemistry & Cell Biology, 36*(11), 2214-2225. https://doi.org/10.1016/j.biocel.2004.04.015

Bell, M., Kochanek, P., Carcillo, J., Mi, Z., Schiding, J., Wisniewski, Clark, R., Dixon, C., Marion, D., & Jackson, E. (1998). Interstitial adenosine, inosine, and hypoxanthine are increased after experimental traumatic brain injury in the rat. *Journal of Neurotrauma, 15*(3), 163-170. https://doi.org/10.1089/neu.1998.15.163

Benowitz, L. I., Goldberg, D. E., Madsen, J. R., Soni, D., & Irwin, N. (1999). Inosine stimulates extensive axon collateral growth in the rat corticospinal tract after injury. Proceedings of the National Academy of Sciences - PNAS, 96(23), 13486-13490. https://doi.org/10.1073/pnas.96.23.13486

Bonifačić, D., Aralica, M., Sotošek Tokmadžić, V., Rački, V., Tuškan-Mohar, L., & Kučić, N. (2017). Values of vanillylmandelic acid and homovanillic acid in the urine as potential prognostic biomarkers in ischaemic stroke patients. *Biomarkers, 22*(8), 790-797. https://doi.org/10.1080/1354750X.2017.1351001

Bouatra, S., Aziat, F., Mandal, R., Guo, A.C., Wilson, M.R., Knox, C.,…Wishart, D.S. (2013). The Human Urine Metabolome. *PLoS ONE, 8*(9), 1-28. Doi: 10.1371/journal.pone.0073076

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B, Methodological, 26*(2), 211-252. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x

Brott, T., Adams, J., H P, Olinger, C. P., Marler, J. R., Barsan, W. G., Biller, J., . . . Hertzberg, V. (1989). Measurements of acute cerebral infarction: A clinical examination scale. *Stroke, 20*(7), 864-870.

Burnstock, G. (2016). An introduction to the roles of purinergic signalling in neurodegeneration, neuroprotection and neuroregeneration. *Neuropharmacology, 104*, 4-17. https://doi.org/10.1016/j.neuropharm.2015.05.031

Bykowski, E. A., Petersson, J. N., Dukelow, S., Ho, C., Debert, C. T., Montina, T., & Metz, G. A. S. (2021a). Urinary biomarkers indicative of recovery from spinal cord injury: A pilot study. *IBRO Neuroscience Reports, 10*(1), 178-185. https://doi.org/10.1016/j.ibneur.2021.02.007

Bykowski, E. A., Petersson, J. N., Dukelow, S., Ho, C., Debert, C. T., Montina, T., & Metz, G. A. S. (2021b). Urinary metabolomic signatures as indicators of injury severity following traumatic brain injury: A pilot study. *IBRO Neuroscience Reports, 11*, 200-206. https://doi.org/10.1016/j.ibneur.2021.10.003

Bykowski, E. A., Petersson, J. N., Dukelow, S., Ho, C., Debert, C. T., Montina, T., & Metz, G. A. S. (2023). Identification of serum metabolites as prognostic biomarkers following spinal cord injury: A pilot study. *Metabolites, 13*(5), 605. https://doi.org/10.3390/metabo13050605

Campbell, B. C. V. & Khatri, P. (2020). Stroke. *The Lancet (British Edition), 396*(10244), 129. https://doi.org/10.1016/S0140-6736(20)31179-X

Chen, P., Goldberg, D. E., Kolb, B., Lanser, M., & Benowitz, L. I. (2002). Inosine induces axonal rewiring and improves behavioral outcome after stroke. *Proceedings of the National Academy of Sciences - PNAS, 99*(13), 9031-9036. https://doi.org/10.1073/pnas.132076299

Chen, J., Lv, Y., Li, X., Xiong, J., Liang, H., Xie, L., Wan, C., Chen, Y., Wang, H., Liu, P., & Zheng, H. (2021). Urinary metabolite signatures for predicting elderly stroke survivors with depression. *Neuropsychiatric Disease and Treatment, 17*, 925-933. https://doi.org/10.2147/NDT.S299835

Chen, Q., Zhou, T., Yuan, J., Xiong, X., Liu, X., Qiu, Z., Hu, L., Lu, H., He, Q., Liu, C., & Yang, Q. (2023). Metabolomics profiling to characterize cerebral ischemia-reperfusion injury in mice. *Frontiers in Pharmacology, 14*, 1091616-1091616. https://doi.org/10.3389/fphar.2023.1091616

Craig, A., Cloarec, O., Holmes, E., Nicholson, J. K., & Lindon, J. C. (2006). Scaling and normalization effects in NMR spectroscopic metabonomic data sets. *Analytical Chemistry (Washington), 7*8(7), 2262-2267. https://doi.org/10.1021/ac0519312

Dagonnier, M., Donnan, G. A., Davis, S. M., Dewey, H. M., & Howells, D. W. (2021). Acute stroke biomarkers: Are we there yet? *Frontiers in Neurology, 12*, 619721-619721. https://doi.org/10.3389/fneur.2021.619721

Dale, N., Tian, F., Sagoo, R., Phillips, N., Imray, C., & Roffe, C. (2019). Point-of-care measurements reveal release of purines into venous blood of stroke patients. *Purinergic Signalling, 15*(2), 237-246. https://doi.org/10.1007/s11302-019-09647-4

Debik, J., Sangermani, M., Wang, F., Madssen, T. S., & Giskeødegård, G. F. (2022). Multivariate analysis of NMR-based metabolomic data. *NMR in Biomedicine, 35*(2), e4638-n/a. https://doi.org/10.1002/nbm.4638

Dinga, R., Schmaal, L., Penninx, B. W., Veltman, D. J., & Marquand, A. F. (2020). Controlling for effects of confounding variables on machine learning predictions. *BioRxiv*. https://doi.org/10.1101/2020.08.17.255034

Eisenhofer, G., Kopin, I., & Goldstein, D. (2004). Catecholamine metabolism: A contemporary view with implications for physiology and medicine. *Pharmacological Reviews, 56*(3), 331-349. https://doi.org/10.1124/pr.56.3.1

Emwas, A. M. (2015). The strengths and weaknesses of NMR spectroscopy and mass spectrometry with particular focus on metabolomics research. *Methods in Molecular Biology (Clifton, N.J.)*, 1277, 161.

Emwas, A. M., Salek, R. M., Griffin, J. L., & Merzaban, J. (2013). NMR-based metabolomics in human disease diagnosis: Applications, limitations, and recommendations. *Metabolomics, 9*(5), 1048-1072. https://doi.org/10.1007/s11306-013-0524-y

Farooqui, A. A., Horrocks, L. A., & Farooqui, T. (2000). *Glycerophospholipids in brain: Their metabolism, incorporation into membranes, functions, and involvement in neurological disorders*. Elsevier Ireland Ltd. https://doi.org/10.1016/S0009-3084(00)00128-6

Farooqui, A. A., Horrocks, L. A., & Farooqui, T. (2007). Modulation of inflammation in brain: A matter of fat. *Journal of Neurochemistry, 101*(3), 577-599. https://doi.org/10.1111/j.1471-4159.2006.04371.x

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861-874. https://doi.org/10.1016/j.patrec.2005.10.010

Fernandes, C. G., Leipnitz, G., Seminotti, B., Amaral, A. U., Zanatta, Â., Vargas, C. R., Dutra Filho, C. S., & Wajner, M. (2010). Experimental evidence that phenylalanine provokes oxidative stress in hippocampus and cerebral cortex of developing rats. *Cellular and Molecular Neurobiology, 30*(2), 317-326. https://doi.org/10.1007/s10571-009-9455-6

Fredholm, B. B. (2007). Adenosine, an endogenous distress signal, modulates tissue damage and repair. *Cell Death and Differentiation, 14*(7), 1315-1323. https://doi.org/10.1038/sj.cdd.4402132

Fredholm, B. B., Johansson, S., & Wang, Y. (2011). Adenosine and the regulation of metabolism and body temperature. In K. Jacobson, & J. Linden (Eds.), *Advances in pharmacology (1990)* (pp. 77-94). Elsevier. https://doi.org/10.1016/B978-0-12-385526-8.00003-5

Gil, R. B., Lehmann, R., Schmitt-Kopplin, P., & Heinzmann, S. S. (2016). 1H NMR-based metabolite profiling workflow to reduce inter-sample chemical shift variations in urine samples for improved biomarker discovery. *Analytical and Bioanalytical Chemistry, 408*(17), 4683-4691. https://doi.org/10.1007/s00216-016-9552-6

Goodpaster, A., Romick-Rosendale, L., & Kennedy, M. (2010). Statistical significance analysis of nuclear magnetic resonance-based metabonomics data. *Analytical Biochemistry, 401*: 134-143.

Gowland, C., Stratford, P., Ward, M., Moreland, J., Torresin, W., Vanhullenaar, S., Sanford, J., Barreca, S., Vanspall, B., & Plews, N. (1993). Measuring physical impairment and

disability with the chedoke-mcmaster stroke assessment. *Stroke (1970), 24*(1), 58-63. https://doi.org/10.1161/01.STR.24.1.58

Haidar, N., Carrara, M., Andriamampandry, C., Kanfer, J., Freysz, L., Dreyfus, H., & Massarelli, R. (1994). Incorporation of [h-3] ethanolamine into acetylcholine by a human cholinergic neuroblastoma clone. *Neurochemical Research, 19*(1), 9-13. https://doi.org/10.1007/BF00966721

Han, Z., Shen, F., He, Y., Degos, V., Camus, M., Maze, M., Young, W. L., & Su, H. (2014). Activation of alpha-7 nicotinic acetylcholine receptor reduces ischemic stroke injury through reduction of pro-inflammatory macrophages and oxidative stress. *PloS One, 9*(8), e105711-e105711. https://doi.org/10.1371/journal.pone.0105711

Haskó, G., Sitkovsky, M. V., & Szabó, C. (2004). Immunomodulatory and neuroprotective effects of inosine. *Trends in Pharmacological Sciences (Regular Ed.), 25*(3), 152-157. https://doi.org/10.1016/j.tips.2004.01.006

Jiang, W., Chen, J., Gong, L., Liu, F., Zhao, H., & Mu, J. (2022). Alteration of glycerophospholipid metabolism in hippocampus of post-stroke depression rats. *Neurochemical Research, 47*(7), 2052-2063. https://doi.org/10.1007/s11064-022-03596-y

Ke, C., Pan, C., Zhang, Y., Zhu, X., & Zhang, Y. (2019). Metabolomics facilitates the discovery of metabolic biomarkers and pathways for ischemic stroke: A systematic review. *Metabolomics, 15*(12), 152-21. https://doi.org/10.1007/s11306-019-1615-1

Keith, R. A., Granger, C. V., Hamilton, B. B., & Sherwin, F. S. (1987). The functional independence measure: A new tool for rehabilitation. *Advances in Clinical Rehabilitation, 1*, 6-18.

Koen, N., Du Preez, I., & Loots, D. T. (2016). Metabolomics and personalized medicine. In R. Donev (Ed.), *Advances in protein chemistry and structural biology* (pp. 53-78). Elsevier. https://doi.org/10.1016/bs.apcsb.2015.09.003

Kong, Y., Feng, Y., Lu, Y., Feng, S., Huang, Z., Wang, Q., Huang, H., Ling, X., Su, Z., & Guo, Y. (2022). Predictive serum biomarkers of patients with cerebral infarction. *Neurological Research (New York), 44*(4), 331-341. https://doi.org/10.1080/01616412.2021.1987055

Li, X., Ma, S., & Yi, C. (2016). Pseudouridine: The fifth RNA nucleotide with renewed interests. *Current Opinion in Chemical Biology, 33*, 108-116. https://doi.org/10.1016/j.cbpa.2016.06.014

Liu, H., Pu, J., Zhou, Q., Yang, L., & Bai, D. (2022). Peripheral blood and urine metabolites and biological functions in post-stroke depression. *Metabolic Brain Disease, 37*(5), 1557-1568. https://doi.org/10.1007/s11011-022-00984-9

Liu, J., Yuan, J., Zhao, J., Zhang, L., Wang, Q., & Wang, G. (2021). Serum metabolomic patterns in young patients with ischemic stroke: A case study. *Metabolomics, 17*(2), 24-24. https://doi.org/10.1007/s11306-021-01774-7

Maden, B. E. (1990). The numerous modified nucleotides in eukaryotic ribosomal RNA. *Progress in Nucleic Acid Research and Molecular Biology, 39*, 241. https://doi.org/10.1016/S0079-6603(08)60629-7

Mayne, M., Fotheringham, J., Yan, H., Power, C., Del Bigio, M. R., Peeling, J., & Geiger, J. D. (2001). Adenosine A2A receptor activation reduces proinflammatory events and decreases cell death following intracerebral hemorrhage. *Annals of Neurology, 49*(6), 727-735. https://doi.org/10.1002/ana.1010

Meltzer, S., Bagley, J. A., Perez, G. L., O'Brien, C. E., DeVault, L., Guo, Y., Jan, L. Y., & Jan, Y. (2017). Phospholipid homeostasis regulates dendrite morphogenesis in drosophila sensory neurons. *Cell Reports (Cambridge), 21*(4), 859-866. https://doi.org/10.1016/j.celrep.2017.09.089

Molinoff, P. B., & Axelrod, J. (1971). Biochemistry of catecholamines. *Annual review of biochemistry*, *40*(1), 465-500.

Nicholson, J. K., & Lindon, J. C. (2008). Metabonomics. *Nature (London), 455*(7216), 1054-1056. https://doi.org/10.1038/4551054a

Nicholson, J. K., Lindon, J. C., & 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), 1181-1189. https://doi.org/10.1080/004982599238047

Ormstad, H., Verkerk, R., Aass, H. C. D., Amthor, K., & Sandvik, L. (2013). Inflammation-induced catabolism of tryptophan and tyrosine in acute ischemic stroke. *Journal of Molecular Neuroscience, 51*(3), 893-902. https://doi.org/10.1007/s12031-013-0097-2

Palkovits, M., Kobayashi, R., Kizer, J., Jacobowitz, D., & Kopin, I. (1975). Effects of stress on catecholamines and tyrosine-hydroxylase activity of individual hypothalamic nuclei. *Neuroendocrinology, 18*(2), 144-153. https://doi.org/10.1159/000122394

Pang, Z., Chong, J., Li, S., & Xia, J. (2020). MetaboAnalystR 3.0: Toward an optimized workflow for global metabolomics. *Metabolites, 10*(5), 186. https://doi.org/10.3390/metabo10050186

Paxman, E. J., Wanner, Z., Boora, J., Douglas, K., Metz, G., Montina, T., & Debert, C. (2017). Metabolomic biomarkers of concussion. *British Journal of Sports Medicine, 51*(11), A4-A4. https://doi.org/10.1136/bjsports-2016-097270.9

Poelchen, W., Sieler, D., Wirkner, K., & Illes, P. (2001). Co-transmitter function of ATP in central catecholaminergic neurons of the rat. *Neuroscience, 102*(3), 593-602. https://doi.org/10.1016/S0306-4522(00)00529-7

Potter, P., & White, T. (1980). Release of adenosine 5'-triphosphate from synaptosomes from different regions of rat-brain. *Neuroscience, 5*(7), 1351-1356. https://doi.org/10.1016/0306-4522(80)90207-9

Psychogios, N., Hau, D. D., Peng, J., Guo, A. C., Mandal, R., Bouatra, S., Sinelnikov, I., Krishnamurthy, R., Eisner, R., Gautam, B., Young, N., Xia, J., Knox, C., Dong, E., Huang, P., Hollander, Z., Pedersen, T. L., Smith, S. R., Bamforth, F., . . . Wishart, D. S. (2011). The human serum metabolome. *PloS One, 6*(2), e16957-e16957. https://doi.org/10.1371/journal.pone.0016957

Qi, B., Zhang, Y., Xu, B., Zhang, Y., Fei, G., Lin, L., & Li, Q. (2022). Metabolomic characterization of acute ischemic stroke facilitates metabolomic biomarker discovery. *Applied Biochemistry and Biotechnology, 194*(11), 5443-5455. https://doi.org/10.1007/s12010-022-04024-1

Qureshi, M. I., Vorkas, P. A., Coupland, A. P., Jenkins, I. H., Holmes, E., & Davies, A. H. (2017). Lessons from metabonomics on the neurobiology of stroke. *SAGE Publications*. https://doi.org/10.1177/1073858416673327

Roth, G. A., Mensah, G. A., Johnson, C. O., Addolorato, G., Ammirati, E., Baddour, L. M., Barengo, N. C., Beaton, A. Z., Benjamin, E. J., Benziger, C. P., Bonny, A., Brauer, M., Brodmann, M., Cahill, T. J., Carapetis, J., Catapano, A. L., Chugh, S. S., Cooper, L. T., Coresh, J., . . . GBD-NHLBI-JACC Global Burden of Cardiovascular Diseases Writing Group. (2020). Global burden of cardiovascular diseases and risk factors, 1990-2019 update from the GBD 2019 study. *Journal of the American College of Cardiology, 76*(25), 2982-3021. https://doi.org/10.1016/j.jacc.2020.11.010

Sacco, R. L., Kasner, S. E., Broderick, J. P., Caplan, L. R., Connors, J. J. (., Culebras, A., Elkind, M. S. V., George, M. G., Hamdan, A. D., Higashida, R. T., Hoh, B. L., Janis, L. S., Kase, C. S., Kleindorfer, D. O., Lee, J., Moseley, M. E., Peterson, E. D., Turan, T. N., Valderrama, A. L., . . . Council on Cardiovascular and Stroke Nursing. (2013). An updated definition of stroke for the 21st century: A statement for healthcare professionals from the american heart Association/American stroke association. *Stroke (1970), 44*(7), 2064-2089. https://doi.org/10.1161/STR.0b013e318296aeca

Sanayama, Y., Nagasaka, H., Takayanagi, M., Ohura, T., Sakamoto, O., Ito, T., Ishige-Wada, M., Usui, H., Yoshino, M., Ohtake, A., Yorifuji, T., Tsukahara, H., Hirayama, S., Miida, T., Fukui, M., & Okano, Y. (2011). Experimental evidence that phenylalanine is strongly associated to oxidative stress in adolescents and adults with phenylketonuria. *Molecular Genetics and Metabolism, 103*(3), 220-225. https://doi.org/10.1016/j.ymgme.2011.03.019

Sidorov, E., Bejar, C., Xu, C., Ray, B., Reddivari, L., Chainakul, J., Vanamala, J. K. P., & Sanghera, D. K. (2020). Potential metabolite biomarkers for acute versus chronic stage of ischemic stroke: A pilot study. *Journal of Stroke and Cerebrovascular Diseases, 29*(4), 104618-104618. https://doi.org/10.1016/j.jstrokecerebrovasdis.2019.104618

Sitta, A., Manfredini, V., Biasi, L., Treméa, R., Schwartz, I. V. D., Wajner, M., & Vargas, C. R. (2009). Evidence that DNA damage is associated to phenylalanine blood levels in leukocytes from phenylketonuric patients. *Mutation Research, 679*(1), 13-16. https://doi.org/10.1016/j.mrgentox.2009.07.013

Smart, S., Fox, G., Allen, K., Swanson, A., Newman, M., Swayne, G., Clark, J., Sales, K., & Williams, S. (1994). Identification of ethanolamine in rat and gerbil brain-tissue extracts by nmr-spectroscopy. *NMR in Biomedicine, 7*(8), 356-365. https://doi.org/10.1002/nbm.1940070806

Smelter, A., Rouchka, E. C., & Moseley, H. N. B. (2017). Detecting and accounting for multiple sources of positional variance in peak list registration analysis and spin system grouping. *Journal of Biomolecular NMR, 68*(4), 281-296. https://doi.org/10.1007/s10858-017-0126-5

Sun, H., Zhao, J., Zhong, D., & Li, G. (2017). Potential serum biomarkers and metabonomic profiling of serum in ischemic stroke patients using UPLC/Q-TOF MS/MS. *PloS One, 12*(12), e0189009-e0189009. https://doi.org/10.1371/journal.pone.0189009

Szymanska, E., Saccenti, E., Smilde, A. K., & Westerhuis, J. A. (2012). Double-check: Validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics, 8*(Suppl 1), S3-S16. https://doi.org/10.1007/s11306-011-0330-3

Timpone, V. M., Jensen, A., Poisson, S. N., & Trivedi, P. S. (2020). Compliance with imaging guidelines for workup of transient ischemic attack evidence from the nationwide emergency department sample. *Stroke (1970), 51*(8), 2563-2567. https://doi.org/10.1161/STROKEAHA.120.029858

van den Berg, Robert A, Hoefsloot, H. C. J., Westerhuis, J. A., Smilde, A. K., & van der Werf, Mariët J. (2006). Centering, scaling, and transformations: Improving the biological information content of metabolomics data. *BMC Genomics, 7*(1), 142-142. https://doi.org/10.1186/1471-2164-7-142

Veselkov, K. A., Lindon, J. C., Ebbels, T. M. D., Crockford, D., Volynkin, V. V., Holmes, E., Davies, D. B., & Nicholson, J. K. (2009). Recursive segment-wise peak alignment of biological 1H NMR spectra for improved metabolic biomarker recovery. *Analytical Chemistry (Washington), 81*(1), 56-66. https://doi.org/10.1021/ac8011544

Vignoli, A., Ghini, V., Meoni, G., Licari, C., Takis, P. G., Tenori, L., Turano, P., & Luchinat, C. (2019). High-Throughput metabolomics by 1D NMR. *Angewandte Chemie (International Ed.), 58*(4), 968-994. https://doi.org/10.1002/anie.201804736

Wanichthanarak, K., Jeamsripong, S., Pornputtapong, N., & Khoomrung, S. (2019). Accounting for biological variation with linear mixed-effects modelling improves the quality of clinical metabolomics data. *Computational and Structural Biotechnology Journal, 17*, 611-618. https://doi.org/10.1016/j.csbj.2019.04.009

Wanner, Z. R., Southam, C. G., Sanghavi, P., Boora, N. S., Paxman, E. J., Dukelow, S. P., Benson, B. W., Montina, T., Metz, G. A. S., & Debert, C. T. (2021). Alterations in urine metabolomics following sport-related concussion: A 1H NMR-based analysis. *Frontiers in Neurology, 12*, 645829-645829. https://doi.org/10.3389/fneur.2021.645829

Wiklund, S., Johansson, E., Sjöström, L., Mellerowicz, E. J., Edlund, U., Shockcor, J. P., Gottfries, J., Moritz, T., & Trygg, J. (2008). Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models. *Analytical Chemistry (Washington), 80*(1), 115-122. https://doi.org/10.1021/ac0713510

Williams-Karnesky, R., & Stenzel-Poore, M. (2009). Adenosine and stroke: Maximizing the therapeutic potential of adenosine as a prophylactic and acute neuroprotectant. *Current Neuropharmacology, 7*(3), 217-227. https://doi.org/10.2174/157015909789152209

Wishart, D. S. (2008). Quantitative metabolomics using NMR. *TrAC, Trends in Analytical Chemistry (Regular Ed.), 27*(3), 228-237. https://doi.org/10.1016/j.trac.2007.12.001

Wishart, D. S., Feunang, Y. D., Marcu, A., Guo, A. C., Liang, K., Vázquez-Fresno, R., Sajed, T., Johnson, D., Li, C., Karu, N., Sayeeda, Z., Lo, E., Assempour, N., Berjanskii, M., Singhal,

S., Arndt, D., Liang, Y., Badran, H., Grant, J., . . . Scalbert, A. (2018). HMDB 4.0: The human metabolome database for 2018. *Nucleic Acids Research, 46*(D1), D608-D617. https://doi.org/10.1093/nar/gkx1089

Xia, J., Broadhurst, D. I., Wilson, M., & Wishart, D. S. (2013). Translational biomarker discovery in clinical metabolomics: An introductory tutorial. *Metabolomics, 9*(2), 280-299. https://doi.org/10.1007/s11306-012-0482-9

Xia, J., & Wishart, D. S. (2010). MetPA: A web-based metabolomics tool for pathway analysis and visualization. *Bioinformatics, 26*(18), 2342-2344. https://doi.org/10.1093/bioinformatics/btq418

Xie, J., Han, Y., Hong, Y., Li, W., Pei, Q., Zhou, X., Zhang, B., & Wang, Y. (2020). Identification of potential metabolite markers for middle-aged patients with post-stroke depression using urine metabolomics. *Neuropsychiatric Disease and Treatment, 16*, 2017-2024. https://doi.org/10.2147/NDT.S271990

Yu, F., Li, X., Feng, X., Wei, M., Luo, Y., Zhao, T., Xiao, B., & Xia, J. (2021). Phenylacetylglutamine, a novel biomarker in acute ischemic stroke. *Frontiers in Cardiovascular Medicine, 8*, 798765-798765. https://doi.org/10.3389/fcvm.2021.798765

Yun, Y., Liang, F., Deng, B., Lai, G., Vicente Gonçalves, C. M., Lu, H., Yan, J., Huang, X., Yi, L., & Liang, Y. (2015). Informative metabolites identification by variable importance analysis based on random variable combination. *Metabolomics, 11*(6), 1539-1551. https://doi.org/10.1007/s11306-015-0803-x

Zai, L., Ferrari, C., Subbaiah, S., Havton, L. A., Coppola, G., Strittmatter, S., Irwin, N., Geschwind, D., & Benowitz, L. I. (2009). Inosine alters gene expression and axonal projections in neurons contralateral to a cortical infarct and improves skilled use of the impaired limb. *The Journal of Neuroscience, 29*(25), 8187-8197. https://doi.org/10.1523/JNEUROSCI.0414-09.2009

Zhang, W., & Zhang, X. (2015). A novel urinary metabolite signature for non-invasive post-stroke depression diagnosis. Cell Biochemistry and Biophysics, 72(3), 661-667. https://doi.org/10.1007/s12013-014-0472-9

CHAPTER 4: PILOT STUDY REVEALS CANDIDATE BLOOD-DERIVED METABOLIC BIOMARKERS AND PATHWAYS IN STROKE PATIENTS

Author Contributions: Jamie N. Petersson spearheaded sample preparation, analysis, and original draft writing. Elani A. Bykowski assisted with sample preparation and analysis. Sean Dukelow, Chester Ho, Tony Montina, Gerlinde A. S. Metz, and Chantel T. Debert participated in conceptualization and research methodology. Chantel T. Debert, Tony Montina, and Gerlinde A. S. Metz supported this work through supervision and acquisition of funding. All authors contributed to manuscript editing.

4.1 INTRODUCTION

With stroke ranking as the second leading cause of death worldwide and a major contributor to disability (Roth et al., 2020), timely identification and personalized treatment are critical in minimizing brain damage. Currently, the most effective treatment for ischemic stroke is tissue plasminogen activator (tPA), but it has a very small 3-hour time window after stroke onset for administration. Further, health professionals have to rapidly diagnose the patient with ischemic stroke and rule out hemorrhagic stroke before administering tPA, because it enhances blood flow and could therefore be fatal to patients with hemorrhagic stroke, who are treated by lowering blood pressure (Campbell & Khatri, 2020). Computed topography (CT) and magnetic resonance imaging (MRI) are most commonly used for stroke diagnosis; however, these techniques have limitations including high cost, low availability, and large time consumption (Sacco et al., 2013; Timpone et al., 2020). Notably, metabolomics has shown promise in differentiating between the two stroke types (Hu et al., 2016; Zhang et al., 2017) and stroke mimics (Tiedt et al., 2020). Along with disease diagnostics, metabolomics also has the advantage of advancing personalized therapies by

determining metabolites and specific metabolic pathways associated with pathophysiology. Personalized medicine is rapidly becoming a more appealing form of treatment, as it has the potential to increase the effectiveness of an individual's recovery. Metabolomics in particular has shown to be a potential precursor for the development of personalized medicine (Holmes et al., 2007).

Through identifying small molecule metabolites, metabolomics involves detecting metabolic alterations in a biological system as a result of pathophysiology (Nicholson et al., 1999). A stroke results in a cascade of neurochemical alterations, many of which may aggravate brain injury and impede recovery. One way of identifying these neurochemical events after stroke is to monitor metabolites, the downstream products of cellular processes. Since metabolism is perturbed by abnormal cellular processes, metabolomics provides an integrative view of the dynamic biochemical processes occurring in an organism (Nicholson & Lindon, 2008). Metabolites encompass a diverse array of compounds that are detectable within biological specimens such as urine and blood (Vignoli et al., 2019). In the context of this study, blood serves as a valuable biofluid due to its accessibility and the comprehensive information it provides. Metabolites are circulated throughout the body in blood. Blood supplies complementary information to urine since the metabolites have not yet been filtered by the kidneys and excreted in urine. Hence, blood provides direct information about which metabolites are being used or produced in the body at that time. Blood contains a rich repertoire of metabolites that reflect the physiological state of an individual, including those originating from the brain (Qureshi et al., 2017). The analysis of blood metabolites through metabolomics can shed light on the metabolic perturbations associated with stroke, providing a unique opportunity to diagnose and monitor the condition. For these reasons, blood, and more specifically serum, the supernatant fluid lacking clotting factors (West, 1985), was chosen as the biological fluid for this study. By detecting and quantifying specific metabolites,

metabolomics can aid in identifying biomarkers that serve as indicators of stroke severity, prognosis, and extent of recovery. The minimally-invasive nature of blood sample collection and the ability to capture a snapshot of an individual's metabolic profile make metabolomics a promising tool in the field of stroke research, enabling personalized medicine strategies and improving clinical decision-making for optimal neurorehabilitation regimens.

By analyzing metabolites in blood serum, which reflect an individual's overall health and offer insights into metabolic perturbations, nuclear magnetic resonance (NMR)-based metabolomics provides a non-destructive, reproducible, and comprehensive approach (Emwas, 2015; Nicholson & Lindon, 2008; Wishart, 2008). This high-throughput technique effectively detects 49 biological compounds in serum, with 20 being unique to NMR (Psychogios et al., 2011). Notably, NMR is similar to MRI and CT in that it incurs significant expenses, is not widely available, and is time-consuming. However, NMR-based biomarkers have potential for conversion into affordable, standardized commercial assays (Dagonnier et al., 2021). This research presents a stepping stone in its development. Furthermore, the development of blood-derived biomarkers can bridge the gap in clinical stroke management, enabling rapid and accurate diagnosis, prognosis, and personalized treatment evaluation. Metabolomics has proven to be amenable to biomarker discovery for diagnostic, prognostic, and treatment success purposes for stroke (Laborde et al., 2012; Qureshi et al., 2017) and other diseases of the central nervous system (CNS) investigated in previous research from our group (Bykowski et al., 2021a; Bykowski et al., 2021b; Bykowski et al., 2023; Paxman et al., 2017; Wanner et al., 2021). In the clinical setting, readily accessible samples can be obtained to monitor biomarkers and assess recovery and neurological progress, offering a minimally-invasive, cost-effective, and precise approach for personalized medicine. The utilization of a biomarker panel consisting of the top 1-10 metabolites that show substantial

alterations can shed light on the biological processes underlying individual variations in prognosis, treatment response, and the condition itself (Koen et al., 2016; Xia et al., 2013).

The present pilot study explores the potential of metabolomics, specifically NMR spectroscopy, as a powerful metabolite-detecting tool for stroke diagnosis and treatment monitoring. With the aim of identifying novel biomarkers for stroke recovery processes and therapeutic efficacy, this research investigated the blood-derived metabolic profile of stroke patients at the time of onset and six months post-stroke, offering valuable insights into injury severity, potential recovery, and underlying biological pathways involved in repair and recovery. To the best of our knowledge, this study made the first attempt to uncover NMR-based metabolomic biomarkers for stroke triage, prognosis, and recovery. Ultimately, this study may lead a more personalized approach to medicine and allow for the development of targeted and individualized treatments for patients. It may also aid in the understanding of the metabolic processes occurring in the injured brain and CNS. The use of NMR spectroscopy along with multivariate machine learning is anticipated to provide a distinct metabolic fingerprint at stroke onset and at 6-months following stroke and output biomarkers for prognostication and treatment monitoring. The pathways underlying these metabolic alterations will be discovered and examined. This thesis chapter outlines the potential of ^{1}H NMR quantitative metabolomic profiling technique in clinical detection of pathophysiological changes pertaining to diagnosis, prognosis, and treatment efficacy.

4.2 MATERIALS AND METHODS

4.2.1 Patient Characteristics and Sample Collection

Study participants were recruited from a study that aimed to investigate the role of resting state fMRI, biomarkers, and robotics in the recovery of patients with traumatic brain injury, stroke,

and spinal cord injury as part of the Understanding Neurological Recovery (UCAN) project. In this study, 19 patients with stroke were recruited from the stroke unit at the Foothills Medical Centre, which is affiliated with the University of Calgary. Patients were approached by their circle of care and provided with detailed information about the study. Following informed consent, a researcher from the UCAN team collected fasting morning blood samples from the patients at two time points: within 2-11 days (median=5, interquartile range=2.25) after stroke, and again at 6 months post-injury, within 101-242 days (median=203, interquartile range=29.25).

Of the 19 stroke patients, 12 provided two blood samples, including 9 patients with ischemic stroke and 3 with hemorrhagic stroke, with an average age of 60.7±13.6 years. To minimize confounding variables such as diet, lifestyle, body mass index, medical history, and acute versus chronic drug treatments, pairwise analyses were performed within-subjects.

After blood collection, whole blood specimens underwent centrifugation, and subsequently, the supernatant fraction containing serum was isolated. The serum is a proteinaceous fluid component of blood that lacks fibrinogens, thereby rendering it incapable of forming clots. The serum samples were then stored at -80°C and sent to the University of Lethbridge for further processing.

4.2.2 Clinical Assessments

In this study, the initial clinical assessments were conducted within a range of 4 to 51 days, with a median of 14.5 days, and an interquartile range of 15.75 days. The 6-month follow-up assessments were completed between 98 to 242 days, with a median of 203 days, and an interquartile range of 32.25 days, after stroke onset.

To measure stroke severity, the National Institutes of Health Stroke Scale (NIHSS; Brott et al., 1989) was used at the initial time-point. The NIHSS consists of 15 items, which assess various functions such as coordination, vision, speech, sensory, and motor involvement. Each item

was scored on a scale from 0-2, 0-3, or 0-4, and the scores were summed to generate a total possible score between 0-42, with higher scores indicating more severe impairment.

Physical impairments and disabilities in stroke patients were evaluated using the Chedoke-McMaster Stroke Assessment (CMSA; Gowland et al., 1993) within a few days of injury and at the 6 month follow-up. The CMSA employs a 7-point scale to measure impairments and disabilities, with 1 representing paralysis and 7 representing normal movement. The CMSA includes two inventories: the impairment inventory, which categorizes patients based on their motor recovery, and the disability inventory, which measures changes in disability. For this study, only the hand and arm dimensions of the impairment inventory were used.

Furthermore, the level of disability in stroke patients based on motor control, self-care, communication, social skills, and cognitive skills was assessed using the Functional Independence Measure (FIM; Keith et al., 1987) at both baseline and 6 month follow-up. The FIM consists of 18 items, and patients' scores were recorded on a 7-point scale, with a total possible score of 126 indicating complete functional independence.

4.2.3 NMR Sample Preparation, Data Acquisition, and Processing

Whole blood samples were obtained and serum was isolated by centrifugation. The serum was stored at -80°C until further use. An in-house metabolomics buffer was prepared with a 4:1 ratio of dibasic potassium phosphate (K_2HPO_4) to monobasic potassium phosphate (KH_2PO_4) and a concentration of 0.625M in deionized H_2O. In order to inhibit microbial growth, 3.75mM NaN_3 was added to the buffer. To correct for the effect of divalent cations on chemical shift, 0.375mM potassium fluoride (KF) was included. The pH of the buffer solution was adjusted to 7.4 through the titration of 500µL aliquots of 3M HCl and 3M NaOH. The addition of KF and D_2O with trimethylsilyl propionic acid (TSP) in the buffer and sample preparation steps corrected for chemical shift variation and served as a chemical shift reference for the ^1H NMR spectroscopy.

These steps were important in reducing variability and enhancing the accuracy of NMR-based metabolomics analysis (Gil et al., 2016; Smelter et al., 2017).

Amicon 0.5mL 3kDa centrifuge filters were used to ensure that water-soluble metabolites were isolated from the protein-rich component, allowing for more accurate and precise metabolomics analysis. Before using the Millipore filters, they underwent ten rounds of washing with deionized water to eliminate any residual glycerol preservative. Following this, 300µL of metabolomics buffer was added using reverse pipetting to each of the Amicon centrifuge filters. Next, 200µL of serum was added and the sample was centrifuged at 14,000g for 30 min at 4°C. Large macromolecules and proteins were collected in the filter and small-molecule metabolites was collected as the filtrate. The filter with large macromolecules was placed upside down in a centrifuge tube and spun again, but this time at 1000g for 2 minutes, resulting in the concentrated solute macromolecules to fall into the centrifuge tube to be stored at -80°C. All serum samples were processed in a containment level two laboratory to minimize the risk of contamination.

For NMR sample preparation, 100µL of phosphate buffer, 120µL of D_2O containing 0.02709% weight/volume TSP, and 380µL of serum filtrate were centrifuged at 12,000rpm for 5 min at 4°C. Insoluble matter was removed by centrifugation before transferring 550µL of the buffered sample into an NMR tube to be loaded into the spectrometer. Prior to spectral acquisition, samples were vortexed to ensure that the serum was mixed thoroughly. The pH of the sample was checked using an NMR pH meter to confirm that it was at a pH of 7.4±0.05. This procedure was performed to ensure that the sample was not contaminated and to verify its suitability for subsequent NMR analysis.

In this study, we utilized a high-end 700 MHz Bruker Avance III HD NMR spectrometer equipped with a room-temperature triple resonance broad band observe (TBO) probe to acquire NMR data from serum samples. Prior to data acquisition, we conducted three-dimensional and

one-dimensional shimming experiments to minimize any inhomogeneities in the static magnetic field. The data were acquired using a one-dimensional ^{1}H nuclear overhauser effect spectroscopy (NOESY) experiment with water suppression pulse sequence 'noesygpr1d' with a mixing time of 10ms, a size of 128k data points, and 128 scans.

The acquired data were processed using state-of-the-art data processing tools to ensure high-quality data analysis. We first conducted zero filling to 256k points in order to increase the resolution of the data, followed by line broadening to 0.3 Hz to enhance the signal-to-noise ratio. Following this, the spectra underwent Fourier transformation and the spectral peaks were calibrated with reference to the TSP peak located at the zero-point on the chemical shift axis (0.00δ). Automatic phase and baseline correction were applied to remove any instrumental artifacts. Bruker TopSpin software (version 3.2, patch level 6) was used to process the raw NMR spectra. Next, the NMR spectra were imported into MATLAB, and dynamic adaptive binning (Anderson et al., 2011) was performed to segment the spectra into 287 bins. The bins were manually inspected and corrected to ensure the accuracy of the analysis (Kiss et al., 2016; Paxman et al., 2018). In order to mitigate any potential confounding effects originating from the regions corresponding to water and urea, we excluded the respective bins from the analysis.

Metabolites were identified using a combination of resources, including Chenomx 8.2 NMR Suite (Chenomx Inc., Edmonton, Alberta, Canada) the Human Metabolome Database (HMBD; Wishart et al., 2018), and the Human Serum Metabolome (Psychogios et al., 2011). The identified metabolites were used to perform metabolic pathway and visualization tests in MetaboanalystR (Chong et al., 2019), a widely used software for metabolomics analysis. Topology analysis, using the Kyoto Encyclopedia of Genes and Genomes (KEGG) and the HMBD libraries (Wishart et al., 2018; Xia & Wishart, 2010), was also conducted to provide insights into the metabolite pathways that were potentially altered following injury and post-recovery.

4.2.4 Statistical Analysis

Multivariate statistical analysis was utilized to determine if there were differences in the serum metabolite profiles between the 1 month and 6 month post-injury samples. The data were normalized, log-transformed, and pareto-scaled before modeling (Craig et al., 2006; Wiklund et al., 2008; van den Burg et al., 2006; Box & Cox, 1964). Pareto scaling involves centering the mean and dividing by the square root of the standard deviation for each variable. This method was employed to ensure data homogenization and to account for potential variable-specific discrepancies in the magnitude of the spectral signal.

To address age and sex confounding effects, we used a linear regression model in Python's StatsModels package (Python Software Foundation, https://www.python.org/). This method can determine metabolite correlations, compare concentrations across groups, and remove confounding effects by processing residuals (Debik et al., 2022). Longitudinal metabolomics studies have previously used linear mixed models (LMMs) to adjust for dependencies (Debik et al., 2022; Wanichthanarak et al., 2019). We followed the ordinary least squares (OLS) regression method, often used in clinical MRI studies (Dinga et al., 2020), to use residuals in eliminating confounding effects, resulting in more accurate and precise metabolomics analysis.

We used various statistical tests to analyze parametric and non-parametric data, including a paired *t*-test or a paired Wilcoxon signed rank test. The Shapiro-Wilk test was used to assess if data for each bin was parametric (Goodpaster et al., 2010). We also used the Variable Importance Analysis based on random Variable Combination (VIAVC), a machine learning algorithm in MATLAB that identifies important metabolites (Yun et al., 2015). VIAVC categorizes metabolites based on their interaction with other variables and outputs an F-ranked subset and a VIAVC best subset of informative metabolites. The algorithm identifies significant metabolites using the Receiver Operator Characteristic (ROC) test and area under the curve (AUC) analysis (Fawcett,

2006). To assess their robustness and determine the best subset, double 10-fold cross-validation and permutation testing using 2000 permutations were used (Szymanska et al., 2012). The F-ranked subset is generated using 10-fold cross-validation. Overall, VIAVC is a reliable technique to identify significant metabolites due to its thorough approach to modeling and validation (Yun et al., 2015).

The study utilized MetaboAnalystR version 2.0.4 package in R version 3.5.3 (Pang et al., 2020) for statistical analysis. For pathway topology analysis, the Kyoto Encyclopedia of Genes and Genomes (KEGG) database was employed, using only significantly altered metabolites. The hypergeometric test was utilized for over-representation analysis and relative-betweenness centrality for topology analysis (Wishart et al., 2018; Xia & Wishart, 2010). To evaluate biomarker diagnostic ability, Receiver operator characteristic (ROC) curves and area under the curve (AUC) were generated using MetaboAnalystR, which helped compare and visualize the sensitivity and specificity of the biomarkers in the dataset (Fawcett, 2006).

To analyze group separation, we employed unsupervised Principal Components Analysis (PCA) and supervised orthogonal projection to latent structures discriminant analysis (OPLS-DA) (Wiklund et al., 2008). OPLS-DA allowed for between-group visualization while considering within-group separation. To illustrate pattern variation, we used unsupervised hierarchical clustering to generate a heat map. This combination of methods provided a comprehensive approach for data analysis and separation assessment.

Finally, Spearman rank-order correlations were used to analyze the relationship between normalized blood-derived metabolite concentrations and patients' clinical assessment scores. Specifically, correlations were conducted between initial NIHSS, FIM, CMSA scores and initial metabolite concentrations along with correlations of percent differences of CMSA and FIM scores and initial and percent difference of normalized blood-derived metabolite concentrations. We

adjusted the significance threshold for multiple comparisons using the Bonferroni correction method. To obtain a Bonferroni-corrected *p*-value, we divided alpha< 0.05 by the number of VIAVC F-best subset tested (n =10; alpha=0.005). This approach identified a rigorous and relevant set of metabolites (Goodpaster et al., 2010). The percent differences of metabolites, CMSA, and FIM scores were determined and are computed as follows:

(6 Months Score-Initial Score)/((6 Months Score + Initial Score)/2)×100%

In summary, the researchers employed a variety of statistical methods to analyze the blood metabolite profiles in the acute stroke and 6 month post-injury samples. To address confounding effects, linear regression models were used, and the data were processed using different techniques, including paired *t*-tests or Wilcoxon signed rank tests, and Shapiro-Wilk tests. VIAVC was used to identify significant metabolites, and pathway topology analysis was conducted using Metaboanalyst. PCA and OPLS-DA were employed using the same platform for group separation assessment. Finally, Spearman correlations were used to analyze the relationship between serum metabolite concentrations and clinical assessment scores.

4.3 RESULTS

4.3.1 Patient Characteristics

Comprehensive clinical and demographic data from 12 stroke patients collected initially and at 6-months follow-up is highlighted in Table 4. It displays differences in age, sex, stroke type, vascular territory, affected hemisphere, blood collection days, medications, and co-morbidities among patients. Additionally, the table provides valuable information on the impact of stroke impairments and disabilities, as evidenced by changes in NIHSS, FIM, CMSA-arm, and CMSA-hand scores. The NIHSS is a standardized tool to measure stroke severity taken during the acute phase of stroke, with scores of 0-4 representing mild stroke severity and scores ranging from 5-15

indicating moderate stroke severity. The average score out of 12 patients was 5.18±3.37 and there were seven patients with mild severity, four patients with moderate severity, and one with unknown severity. FIM and CMSA scores taken initially and at 6 months reflected the functional status and motor impairments of stroke patients. The average improvements for these measures were 20.30±15.28 for FIM, 1.08±1.24 for CMSA-arm, and 0.67±0.65 for CMSA-hand, with the majority of participants showing progress. Improvement in these clinical parameters suggests recovery from stroke, which is defined in this thesis as the pathophysiological, biochemical, and clinical measurement changes from initial collection to 6 months post-stroke.

4.3.2 Metabolomic Analysis

The robust analysis of serum metabolites conducted using VIAVC and paired *t*-test or Wilcoxon signed rank revealed valuable insights into the molecular changes associated with stroke. Table 5 presents the results of this rigorous analysis, encompassing a total of 287 bins and 65 being significant. Notably, the analysis identified 10 metabolites as significant by VIAVC, 59 metabolites as significant through paired *t*-test or Wilcoxon signed rank, and most intriguingly, 4 metabolites demonstrated significance via both methods. These shared significant metabolites are 3-hydroxyisovaleric acid, mannose, glucose, and dimethyl sulfone, which may hold importance in unraveling the underlying metabolic alterations in stroke pathophysiology.

Discrimination between the initial blood sample collection and the collection at 6 months is evident after inspection of the OPLS-DA scores plots from Figure 6. Two panels (A and B) are presented, each utilizing different subsets of metabolites determined to be significant by different analysis methods. The OPLS-DA scores plot in Panel A illustrates the separation between the two groups. The x-axis represents the predictive variation between the two groups, while the y-axis displays the orthogonal variation within each group. The model in Panel A was constructed using a subset of 10 metabolites that were determined to be significant by VIAVC best subset analysis

out of the total 287 bins. The cross-validation measure, Q^2, was calculated to be 0.854 (p<5e-04), indicating good predictive accuracy. The model fit measure, R^2Y, was determined to be 0.898 (p<5e-04), indicating a high degree of explained variation. The OPLS-DA scores plot in Panel B further demonstrates the discrimination between the initial blood sample collection and the collection at 6 months post-stroke. This plot was generated using a subset of 65 metabolites that were found to be significant by both paired *t*-test and VIAVC best subset analysis out of the total 287 bins. The cross-validation measure, Q^2, for this model was calculated to be 0.81 (p<5e-04), indicating a reasonably good predictive accuracy. The model fit measure, R^2Y, was determined to be 0.987 (p<5e-04), indicating a high degree of explained variation. Both OPLS-DA models in Figure 6 exhibit strong discrimination, high predictive accuracy, and excellent model fit, underscoring their effectiveness in distinguishing between the initial blood sample collection and the collection at 6 months post-stroke.

The receiver operator characteristic (ROC) curve in Figure 7 exemplifies excellent sensitivity and specificity in discriminating between initial stroke and 6-months post-stroke samples. The utilization of the VIAVC best subset analysis for model construction highlights its remarkable robustness, as demonstrated by the area under the curve (AUC) of 0.982. The 95% confidence interval ranges from 0.889 to 1, affirming the high precision of the model. Notably, the predictive accuracy of the model stands at 85%. The ROC curve in Figure 7 was generated based on a subset of 10 metabolites identified as significant by VIAVC best subset analysis out of the total 287 bins.

The pathway topology analysis presented in Figure 8 offers comprehensive insights into the metabolic alterations associated with stroke. The paired *t*-test and VIAVC best subset analysis identified a subset of significant metabolites that are involved in various metabolic pathways. Among these pathways, aminoacyl-tRNA biosynthesis (p=1.26E-05) exhibited the highest level

of significance, followed by glyoxylate and dicarboxylate metabolism (p=1.91E-04), synthesis and degradation of ketone bodies (p=0.0030), and several others. The synthesis and degradation of ketone bodies metabolic pathway demonstrated the highest impact, suggesting a greater number of metabolites within this pathway exhibits significant alterations between initial stroke and 6 months post-stroke. Among the additional significant pathways identified, the following pathways demonstrated notable significance: valine, leucine, and isoleucine biosynthesis (p=0.0082), glycolysis/gluconeogenesis (p=0.010), alanine, aspartate, and glutamate metabolism (p=0.013), phenylalanine metabolism (p=0.013), glycine, serine, and threonine metabolism (p=0.020), butanoate metabolism (p=0.029), arginine and proline metabolism (p=0.029), and valine, leucine, and isoleucine degradation (p=0.033). These findings highlight the involvement of these pathways in the metabolic changes occurring after stroke.

4.3.3 Correlation of Metabolomic Signatures to Motor Recovery

Table 6 and Figure 9 present the results of the Spearman correlation analysis between blood-based metabolites and clinical parameters in stroke patients. The table summarizes the results and the figure provides a visual of the correlations. Significant associations were observed, shedding light on the relationship between metabolite concentrations and functional outcomes. The Table 6 and Figure 9 unveil the significant findings from multiple correlation analyses, encompassing the relationship between metabolite concentrations and clinical outcomes. Specifically, the significant results demonstrated: (1) the association between initial metabolite concentration and initial FIM scores, (2) the relationship between initial metabolite concentration and percent difference in FIM scores, and (3) the correlation between percent difference in metabolite concentration and percent difference in FIM and CMSA-hand scores. Notably, Alloisoleucine exhibited a robust correlation with clinical measures scores across all three correlation analyses (Panel A, C, and G in Figure 9), indicating its potential as a biomarker for

assessing stroke severity, prognosis, and recovery. Furthermore, 3-methyl-2-oxovaleric acid and 2-hydroxybutyric acid displayed significant correlations with the percent difference in FIM scores (B and D in Figure 9, respectively), suggesting their involvement in functional recovery processes. Strikingly, the negative correlation between dimethyl sulfone and percent difference in FIM scores displayed Bonferroni corrected significance (Panel E in Figure 9) and suggests a potential role of this metabolite in stroke rehabilitation. The observed significant correlation between the percent difference in mannose concentration and percent difference in CMSA-hand scores (Panel F in Figure 9) underscores its involvement in post-stroke functional outcomes. In the first comparison, elevated metabolite levels exhibited a positive association with higher FIM scores, indicating a potential link between increased metabolite concentrations and greater functional independence. Conversely, the second analysis demonstrated an inverse relationship where higher metabolite levels were linked to a negative percent difference, denoting improved clinical measures. In the third comparison, a positive correlation was observed between augmented metabolite concentration from baseline to 6-months post-stroke and a larger percent difference in CMSA-hand scores, signifying enhanced hand motor function. Conversely, a negative relationship was observed between a greater decrease in metabolite concentration over time and a larger improvement (negative percent difference) in FIM scores.

4.4 DISCUSSION

The following sections provide an in-depth discussion of the results presented in this study. First, using VIAVC and OPLS-DA, this pilot work successfully distinguished metabolomic profiles between the acute phase of stroke and the period 6 months post-stroke. Second, pathway topology analysis identified 11 significantly altered pathways following stroke. Among these, three pathways are discussed in detail: aminoacyl tRNA biosynthesis, glyoxylate and

dicarboxylate metabolism, and synthesis and degradation of ketone bodies. Finally, alloisoleucine emerged as a potential comprehensive biomarker for stroke assessment, showing significance diagnosing stroke onset, predicting outcomes, and monitoring treatment efficacy. Four other metabolites, 3-methyl-2-oxovaleric acid, 2-hydroxybutyric acid, dimethyl sulfone, and mannose, have also shown promise in predicting stroke prognosis and treatment success by reflecting processes such as oxidative stress, inflammation, excitotoxicity, and energy deprivation.

4.4.1 Distinguishing Stroke Metabolite Profiles Over Time

The clear separation observed in the OPLS-DA scores plots indicated distinct metabolic profiles between acute stroke and 6 months post-stroke. The discrimination analysis results demonstrated the performance of each panel in terms of Q^2 value, R^2Y value, and *p*-value. Q^2 value represents the predictive accuracy of the model, while R^2Y value indicates the goodness of fit. Both panels showed robust discrimination between the initial blood sample collection and the collection at 6 months post-stroke, as indicated by their high Q^2 and R^2Y values. Hence, the model was able to accurately predict the group membership and explain the observed metabolic variations. The *p*-value, which measures the statistical significance of the discrimination, was less than 5E-04 for both panels, further confirming the significant separation between the two time points and that these results stand irrespective of the subset used for analysis. These results highlight the effectiveness of VIAVC and OPLS-DA in distinguishing the metabolic profiles of blood samples collected initially after stroke and at 6 months post-stroke.

The ROC results presented in Figure 7 highlight the significant diagnostic value of the developed model for discriminating between initial stroke and 6-months post-stroke samples. The high AUC of 0.982 indicated an excellent ability to correctly classify the samples, with a narrow 95% confidence interval (0.889-1) suggesting a reliable estimation. The observed sensitivity and specificity demonstrated the model's capability to accurately differentiate between the two groups.

Moreover, the predictive accuracy of 85% further emphasizes the practical utility of the model given that this analysis displayed exceptional performance when classifying samples as initial stroke or 6 months post-stroke. These findings underscore the importance of the VIAVC best subset analysis in identifying key metabolites that contribute to the discriminative power of the model, offering potential insights into the underlying metabolic alterations associated with stroke progression. The robust performance of the model in the ROC analysis supports its potential as a valuable tool in clinical settings for effective stroke monitoring and patient management. In a clinical setting, OPLS-DA, ROC, and the VIAVC-based metabolites can aid in early stroke diagnosis, risk stratification, monitoring of stroke progression and recovery, and evaluation of the effectiveness of rehabilitation interventions. Moreover, these findings may contribute to the understanding of the underlying mechanisms of stroke and guide the development of targeted therapeutic strategies.

4.4.2 Metabolomic Pathways Analysis and Stroke

Pathway topology analysis boasted 11 distinct metabolic pathways altered following stroke. Figure 8 displays significant changes in aminoacyl-tRNA biosynthesis, glyoxylate and dicarboxylate metabolism, synthesis and degradation of ketone bodies, valine, leucine, and isoleucine biosynthesis, glycolysis/gluconeogenesis, alanine, aspartate, and glutamate metabolism, phenylalanine metabolism, glycine, serine, and threonine metabolism, butanoate metabolism, arginine and proline metabolism, and valine, leucine, and isoleucine degradation. The statistical significance of these pathways suggests their involvement in stroke pathophysiology, reflecting disruptions in energy metabolism, amino acid metabolism, neurotransmission, oxidative stress, inflammation, and cellular homeostasis. Further investigation is necessary to understand the specific roles and mechanisms by which these pathways contribute to stroke development,

progression, and recovery. The following sections will focus on the three most predominant pathways.

Aminoacyl-tRNA Biosynthesis

The process of synthesizing aminoacyl-tRNA encompasses the binding of a transfer RNA molecule to its corresponding amino acid. This essential procedure occurs during translation, facilitating the transportation of the amino acid to the ribosome for the purpose of protein synthesis (Ibba & Söll, 2000). tRNA acetylation is indispensable for the proper functioning and viability of neural cells. After ischemia, protein synthesis halts to prevent the production of toxic unfolded or misfolded proteins. While this pathological process is intended to be protective, it contributes to ischemic cell death and hinders the production of neuroprotective proteins that could aid in the recovery of neurological function (Paschen, 2004; Paschen et al., 2007). Mounting evidence substantiates the role of the aminoacyl-tRNA biosynthesis pathway after ischemic stroke (Chen et al., 2023; Jia et al., 2021; Ke et al., 2019; Qi et al., 2022). In a comparable vein, the research conducted by Sidorov et al. (2020) examined alterations in metabolites during the transition from the acute to the chronic phase of stroke, revealing significant modifications in aminoacyl-tRNA biosynthesis. In light of the information presented, it appears that the ischemia-induced shutdown of tRNA acetylation occurs within hours of stroke onset and may last for days. During the acute phase, aminoacyl-tRNA biosynthesis is upregulated to promote neurological recovery.

Glyoxylate and Dicarboxylate Metabolism

The glyoxylate cycle represents a metabolic route that metabolizes fatty acids into glucose and is an ancillary pathway of the tricarboxylic acid cycle, circumventing its oxidative decarboxylation steps. This pathway has been extensively associated with stress conditions, such as antibiotic and oxidative stress, hypoxia, and starvation (Ahn et al., 2016; Dolan & Welch, 2018; Plancke et al., 2014; Song et al., 2000; Zeng et al., 2022). Further, this pathway exhibits potential

in regulation of respiration and the provision of defence against neurotoxic reactive oxygen species (ROS), generated under oxidative stress, as evidenced by Ahn et al. (2016). Although our investigation did not pinpoint any literature linking glyoxylate and dicarboxylate metabolism with stroke, it is noteworthy that dysregulation in this pathway has been observed in other CNS injuries, such as spinal cord injury (Bykowski et al., 2023) and severe traumatic brain injury (Baker et al., 2018). Nevertheless, there exists a linkage between this metabolic pathway and the prognosis of stroke, particularly in relation to post-stroke depression (Jia et al., 2021; Ke et al., 2019; Liu et al., 2022). Given the information presented herein, it is plausible that the glyoxylate and dicarboxylate metabolism plays a pivotal role in the context of hypoxia and oxidative stress subsequent to a stroke.

Synthesis and Degradation of Ketone Bodies

Ketone bodies are synthesized through the breakdown of fatty acids during times of stress, such as fasting, starvation, and vigorous exercise, serving as an energy supply to organs under conditions of limited glucose availability. The brain predominantly relies on glucose as its primary energy substrate, and any disruption in its supply, as observed in stroke, culminates in detrimental brain damage. Ketones, on the other hand, circulate throughout the body and are distributed to vital regions such as the brain, where they are converted to acetyl-CoA to serve as an alternative energy source during low energy states (Newman & Verdin, 2014). The degradation of ketones yields ATP energy with greater efficiency compared to glucose synthesis and fatty acid oxidation. Hence, it is unsurprising that ketones are a major fuel source for the brain during glucose shortage, such as ischemic conditions. Remarkably, glial cells possess ketogenic capabilities and support neural cell functioning (Guzman & Blazquez, 2004). Further, ketone bodies protect against cell death, prevent free radical and ROS formation, suppress inflammatory factors, reduce oxidative stress, enhance ATP synthesis, and increase cerebral blood flow (Maalouf et al., 2009; White &

Venkatesh, 2011). Overwhelming evidence substantiates the neuroprotective nature of ketone bodies in pathological conditions such as stroke (Maalouf et al., 2009; Prins, 2008; Suzuki et al., 2001; White & Venkatesh, 2011; Xu et al., 2012). The neuroprotective effect of ketone bodies may arise from their ability to impede lactate accumulation, mitigate the presence of toxic free radicals, hinder the apoptotic cascade, and elevate ATP levels (Prins, 2008; Suzuki et al., 2001). The uptake of ketones bodies has exhibited predictive value in determining stroke severity, favorable outcomes, and facilitating functional recovery (Lin et al., 2023; Pikija et al., 2013). Furthermore, the synthesis and degradation of ketone bodies pathway has been implicated in ischemic stroke (Jia et al., 2021; Ke et al., 2019).

In contrast, ketone bodies have exhibited neurotoxic effects, as evidenced by the association between urinary ketones and unfavorable outcomes as well as death following stroke (Wang et al., 2022; You et al., 2021). This adverse impact may arise from high ketone levels, which can generate oxygen free radicals, increase inflammatory markers, accelerate oxidative stress, and induce lipid peroxidation (Jain et al., 1998; Kanikarla-Marie & Jain, 2015; Karavanaki et al., 2012). The conflicting findings observed could attributed variations in individual physiological conditions and levels of ketone bodies. Further investigation is warranted to ascertain whether ketones confer neuroprotective benefits or exert toxic effects subsequent to stroke; however, it is evident that synthesis and degradation of ketone bodies play a central role in this context.

4.4.3 Metabolite Biomarkers for Clinical Translation

Alloisoleucine in the Diagnosis, Prognosis, and Monitoring of Stroke

The findings indicate that the metabolite alloisoleucine serves as a comprehensive biomarker for assessing stroke onset, outcomes, and recovery. The significance was observed consistently across all three of our correlation groups examined: (1) comparing the initial

alloisoleucine concentration with the initial FIM scores (or diagnosis), (2) contrasting the initial alloisoleucine concentration in relation to percent difference in FIM score (or prognosis), and (3) correlating the percent difference in allosioleucine concentration to percent difference in CMSA-hand (or treatment efficacy).

Alloisoleucine, classified as a branched-chain amino acid (BCAA), is a metabolite recognized for its involvement in neurotoxicity. The dysregulation of alloisoleucine, along with other BCAAs, and their corresponding branched-chain α-keto acids (BCKAs) is well documented in maple syrup urine disease (MSUD). In MSUD, a metabolic blockage leads to the accumulation of these metabolites, resulting in damage to the nervous system and manifestation of neurological symptoms (Chuang & Shih, 2001). Although alloisoleucine and the BCKAs are normal constituents of human plasma, they are typically present at low levels (Schadewalt et al., 2000). Importantly, the significance of 3-methyl-2-oxovaleric acid (KMV), an important BCKA, will be further discussed in the next section. Elevated levels of BCKAs and BCAAs, including alloisoleucine, have been associated with a multitude of detrimental effects, encompassing disturbed homeostasis, increased neurotoxicity, neuroinflammation, oxidative stress, DNA damage, free radical production, lipid peroxidation, brain damage, and ensuing cell death (Amaral & Wajner, 2022; Bridi et al., 2003; Hauschild et al., 2019; Zhenyukh et al., 2017). These deleterious effects, commonly observed in MSUD, are also evident in stroke. Thus, it is plausible to hypothesize that dysregulation of BCAAs, such as alloisoleucine, may play a similar role in stroke pathophysiology. Based on findings of Xu et al. (2020), who established a causal association between BCAAs and stroke, it is proposed that BCAAs can serve as diagnostic biomarkers for stroke. This study posits that alloisoleucine is implicated in stroke onset, outcomes, and recovery, thus qualifying it as a potential biomarker for diagnosis, prognosis, and assessment of treatment efficacy in stroke patients.

Metabolic Signatures to Predict Stroke Patient Outcomes

Three metabolites, namely 3-methyl-2-oxovaleric acid, alloisoleucine, and 2-hydroxybutyric acid, have emerged as noteworthy predictors of stroke prognosis. The initial concentration of these metabolites exhibited a discernable relationship with the percent difference in FIM scores. While alloisoleucine has been discussed previously, this section focuses on the negative correlation observed for 3-methyl-2-oxovaleric acid and 2-hydroxybutyrate, shedding light on their prognostic implications.

As previously mentioned, 3-methyl-2-oxovalerate, also known as α-keto-β-methylvalerate (KMV), is a downstream keto-acid derived from the incomplete breakdown of BCAAs, mainly isoleucine. KVM is a known acidogen, neurotoxin, and metabotoxin (Funchal et al., 2004b; Human Metabolome Database [HMDB], n.d.a). Both alloisoleucine and KMV arise from isoleucine, with the formation of KMV in the catabolic pathway of isoleucine and leading to the production of alloisoleucine (Mamer, 2001; Schadewalt et al., 2000). The presence of high levels of this metabolite has been associated with conditions such as MSUD, impaired fasting glucose, and type 2 diabetes (Chuang & Shih, 2001; Menni et al., 2013). Elevated levels of BCKAs, including KMV, have proven to induce cell death through structural and functional damage to brain cells, as well as excitotoxicity by interfering with glutamate uptake in neural cells (Fuchal et al., 2004a; Funchal et al., 2004b). Further, they contribute to the production of free radicals, lipid peroxidation, brain damage, and impairment of protein synthesis, ATP synthesis, and energy production (Bridi et al., 2005; Pessoa-Pureur & Wajner, 2007; Sitta et al., 2014; Sgaravatti et al., 2003). Wang et al. (2019) specifically demonstrated the toxicity of KMV to astrocytes. Altogether, the observed alterations in alloisoleucine and KMV may have prognostic value while indicating that oxidative stress, excitotoxicity, ketosis, and energy deprivation is occurring after stroke.

2-hydroxybutyric acid or α-hydroxybutyrate (2-HB) is a metabolite produced through the catabolism of methionine and threonine or glutathione synthesis. During oxidative stress, there is an upregulation in glutathione synthesis, and the supply for glutathione synthesis becomes limited, resulting in the redirection of this pathway and the subsequent production of 2-HB as a byproduct (Pechlivanis et al., 2010; HMDB, n.d.b). Thus, 2-HB indicates increased glutathione synthesis required during oxidative stress conditions. It has been established a marker for oxidative damage due to insulin resistance, glucose intolerance, lactic acidosis, and ketoacidosis (Cobb et al., 2016; Gall et al., 2010; Landaas & Pettersen, 1975; Sousa et al., 2021). Furthermore, 2-HB is frequently implicated in hypoxic conditions such as intense exercise, where fatty acids and amino acids serve as alternative metabolic fuel sources in place of glucose (Landaas & Pettersen, 1975; Pechlivanis et al., 2010; Rasooli et al., 2021). In regards to stroke, which is characterized by hypoxia and oxidative stress, research-based findings indicate that 2-HB has also been identified as a biomarker for myocardial ischemia, stroke-like episodes in mitochondrial disease, brain death, post-stroke depression, and stroke itself (García-Aguilera et al., 2021; Kolokolova et al., 2010; Laurensen et al., 2017, Liu et al., 2022; Sharma et al., 2021; Xiao et al., 2016). These findings underscore the occurrence of 2-HB in hypoxic and oxidative stress conditions in stroke. Evidently, the multitude of biochemical changes following stroke give rise to alterations in 2-HB, thereby suggesting its potential utility in stroke prognosis.

Metabolic Signatures to Monitor Stroke Patient Treatment Efficacy

Through correlational analysis, a number of biomarkers for treatment monitoring were discovered. In addition to alloisoleucine, mannose and dimethyl sulfone emerged as notable markers. Notably, dimethyl sulfone and mannose were two of the few metabolites determined to be significant by both VIAVC and paired *t*-test or Wilcoxon signed rank test. Specifically, the percent difference in dimethyl sulfone exhibited a negative correlation to the percent difference in

FIM scores and displayed Bonferroni corrected significance. The mannose concentration percent difference demonstrated a positive correlation with the percent difference in CMSA-hand scores. The clinical relevance of these biomarkers in regards to stroke is discussed in this section.

Dimethyl sulfone, also referred to as methylsulfonyl methane (MSM), is a sulfur-based compound that exhibits a diverse range of beneficial effects during metabolic stress. MSM functions as an antioxidant and anti-inflammatory agent (Amirshahrokhi et al., 2011; Ramoutar & Brumaghim, 2010), harnessing these properties to mitigate tissue damage by scavenging free radicals, supporting immune cells, reducing cytokines, vasodialating agents, and lipid peroxidation, as well as providing sulfur for the synthesis of the antioxidant glutathione (Amirshahrokhi & Niapour, 2022; Butawan et al., 2017; Maranon et al., 2008; Nakhostin-Roohi et al., 2011; Parcell, 2002). Consequently, this metabolite has been proposed as a potential treatment for inflammatory and oxidative stress states, such as stroke (Ahn et al., 2015; Amirshahrokhi & Niapour, 2022; Butawan et al., 2017). Further, dimethyl sulfone has demonstrated effectiveness in mitigating oxidative stress and inflammation after exercise (Maranon et al., 2008; Nakhostin-Roohi et al., 2011) and has been implicated in spinal cord injury (Bykowski et al., 2023). While direct literature linking MSM to stroke was not found, it is known that this metabolite is present in the brain (Rosea et al., 20000) and a study by Lin and colleagues (2001) observed one patient with a high MSM concentration and a history of stroke when investigating the compound in individuals with memory loss. Thus, it is reasonable to postulate that dimethyl sulfone is yet another marker for the pathophysiology of stroke, and further, it may offer potential benefits in monitoring stroke recovery.

Mannose is a sugar that can be converted into fucose and glucose, but unlike other sugars, it displays a multitude of neuroprotective effects during inflammatory states. These effects are mediated through mannose-binding lectin (MBL), the macrophage mannose receptor (MMR), also

known as cluster of differentiation 206 (CD206), and mannose itself. Firstly, MBL plays a critical role in activating the complement system within the innate immune system by binding to mannose residues on pathogens (de Vries et al., 2004; Morrison et al., 2011; Walsh et al., 2004). Additionally, it regulates the inflammatory response by post-transcriptional glycosylation of proteins, modulation of inflammation, phagocytosis of cells, and activation of apoptosis (Morrison et al., 2011; Wang et al., 2014). MBL has been implicated in the pathogenesis of ischemic injury, such as stroke (de Vries et al., 2004; Walsh et al., 2005; Wang et al., 2014), and has been proposed as a prognostic marker, indicating stroke severity, functional outcome, and mortality (Zhang et al., 2015). Secondly, MMR is a transmembrane protein expressed in CNS microglia and dendritic cells (Galea et al., 2005) that is upregulated during inflammatory states and is a recognized marker for the promotion of protection, restoration, and recovery by microglial cells (Tang & Le, 2016). During ischemia, MMR is overexpressed in ischemic tissue and is exclusively localized in the ischemic core, indicating the engagement of microglia in tissue repair (Giraldi-Guimarães et al., 2012; Perego et al., 2011). This upregulation is considered a neuroprotective anti-inflammatory mechanism driven by mannose itself and is associated with enhanced functional recovery and reduced neurodegeneration (Giraldi-Guimarães et al., 2012; Xu et al., 2015). Moreover, the mannose receptor CD206 is involved in suppressing inflammation, tissue remodeling, cell migration, would healing, phagocytosis of pathogens, recognition and clearance of toxic molecules, and antigen direction to immune response sites. (Galea et al., 2005; Giraldi-Guimarães et al., 2012; Stein et al., 1992; Wang et al., 2021; Xu et al., 2015). Lastly, mannose itself possesses anti-inflammatory actions and it is a ligand for MMR, playing a key role in the anti-inflammatory mediation provided by the receptor (Xu et al., 2015). This metabolite is also involved in protein glycosylation (Durand & Seta, 2000), which is a process known to occur in response to inflammation (Novokmet et al., 2014).

Among the beneficial effects attributed to mannose, some of the most notable are reduction of oxidative stress, inhibition of phagocytosis, elevation of glutathione levels, attenuation of lipid peroxidation, limitation of pro-inflammatory responses, and activation of anti-inflammatory responses (Shaker et al., 2021; Wang et al., 2021; Zhang et al., 2021). As a result, mannose has been proposed as a therapeutic agent for inflammatory conditions (Xu et al., 2015; Zhang et al., 2021) and has demonstrated improvements in clinical symptoms and reduced mortality, thereby facilitating tissue repair (Wang et al., 2021). Mannose has previously been implicated in insulin resistance (Lee et al., 2016), myocardial infarction (Fortin et al., 2022), and most notably, stroke (Chen et al., 2023; Liu et al., 2017; Sun et al., 2014). Overall, mannose plays a significant role in the neuroinflammatory response subsequent to stroke. Collectively, the involvement of MBL, MMR, and mannose in promoting recovery following stroke positions mannose as a promising candidate for monitoring a patient's response to treatment.

In light of the findings presented, all metabolites determined to be significant markers for diagnosis, prognosis, or treatment success in this section have been closely intertwined with oxidative stress and inflammation. These markers may share a commonality in contributing to similar pathophysiological mechanisms. For instance, an increase in free radicals under ischemic conditions alters the directionality of the branched-chain α-keto acid dehydrogenase complex, which is involved in BCAA metabolism. Interestingly, 2-HB is a substrate for this complex (Cobb et al., 2016); hence, these alterations can contribute to changes in 2-HB and BCKA levels (Gall et al., 2010). This parallels the associations observed in MSUD, where alloisoleucine, KMV, and 2-HB have all been involved, meaning that post-stroke biochemical alterations may resemble MSUD (Chuang & Shih, 2001; Cobb et al., 2016). Remarkably, BCAAs, KMV, 2-HB, and mannose have all previously displayed alterations with intense exercise (Pechlivanis et al., 2010; Rasooli et al., 2021). As seen here, 2-HB, MSM and mannose all contribute to or enhance the synthesis of

glutathione, a vital antioxidant. It is also worth noting that mannose levels gradually increase with worsening glucose perturbations (Fortin et al., 2022), while 2-HB is a marker for glucose intolerance (Cobb et al., 2016; Gall et al., 2010), and KMV has been associated with impaired fasting glucose (Menni et al., 2013). These observations hold particular interest considering that during stroke, amino acids and fatty acids are used as an alternative fuel source in place of glucose. In conclusion, the interplay between alloisoleucine, KMV, 2-HB, MSM, and mannose highlights their potential relevance in stroke-related metabolic alterations and glutathione synthesis. Additionally, the association of these metabolites with glucose disturbances and their similarities to MSUD and exercise-related changes further emphasize their significance in regards to stroke.

4.5 CONCLUSIONS AND LIMITATIONS

To date, techniques used for diagnosis and injury monitoring are costly, time consuming, and have limited availability (Grotta et al., 1999; Sacco et al., 2013; Timpone et al., 2020). There is an urgent need for an efficient, low-cost method to diagnose and monitor stroke. Metabolomics using ^{1}H NMR spectroscopy is anticipated to satisfy this need. The present results answer this need by demonstrating that metabolomic analysis has clinical importance through the discovery of biomarkers that may contribute to stroke diagnosis, prognosis, and therapeutic efficacy. Through translational biomarker discovery, metabolomics has potential to become a new avenue to access neurorehabilitation strategies and improve them for each patient in a personalized manner; hence, optimizing neurorehabilitation outcomes (Xia et al., 2013). Hence, this study has potential to help clinicians assess neurorehabilitation strategies and improve them for each individual patient by using the identified biomarkers that indicate injury repair. The specific metabolites that emerged as potential diagnostic, prognostic, and treatment efficacy biomarkers are alloisoleucine, 3-methyl-2-oxovaleric acid, and 2-hydroxybutyric acid, dimethyl sulfone, and mannose. These biomarkers

offer valuable information about stroke onset, outcomes, and recovery, and may contribute to personalized treatment approaches.

Understanding the abnormal cellular processes underlying the neuropathophysiological sequelae of stroke also holds clinical importance. Pathway analysis revealed altered metabolic pathways associated with energy metabolism, amino acid metabolism, neurotransmission, oxidative stress, inflammation, and cellular homeostasis following stroke. These findings provide insights into the underlying metabolic alterations associated with stroke, although further investigation is needed to understand their specific roles and mechanisms.

This proof-of-principle study demonstrates the effectiveness of metabolomic analysis in distinguishing the metabolic profiles of blood samples collected initially after stroke and at 6 months post-stroke. VIAVC and OPLS-DA techniques successfully separated and discriminated between these time points with high predictive accuracy and statistical significance. The developed model showed excellent diagnostic value, making it a valuable tool for stroke monitoring and patient management.

In conclusion, ^{1}H NMR-based metabolomic analysis holds promise for improving stroke triage, outcome assessment, and management of rehabilitation interventions. It provides a comprehensive understanding of metabolic changes associated with stroke and identifies potential therapeutic targets. These findings have the potential to enhance patient outcomes and contribute to advancements in stroke management in clinical settings.

The study has several limitations that need to be addressed, including the small sample size, limited diversity, and confounding factors. The confounding factors that ought to be considered in future studies are dietary variations, exercise routines, mobility levels, body mass index, injury severity, the distinction between acute and chronic drug treatments, and medical histories. Improving the inclusivity of the sample group, considering factors such as sex and

baseline/control groups, would strengthen future research. The inclusion of a musculoskeletal injury group would significantly enhance the research quality, enabling a comparison to injury in a broader context. The longitudinal design and regression analysis helped mitigate some confounds, but further investigation into the role of serum metabolites in post-stroke metabolic changes and recovery is warranted. In summary, this preliminary study sheds light on the metabolic shifts associated with strokes and their potential as biomarkers for stroke diagnosis and treatment, thereby augmenting our understanding of the underlying pathophysiology of this condition.

Table 4. Clinical profiles of n=12 stroke patients and changes in functional measures. Clinical demographic characteristics include stroke type (HEM=hemorrhagic), vascular territory, hemisphere impacted by stroke, sex, age, the time interval from stroke occurrence to initial and 6-month sample collection, medications, and co-morbidities. Several clinical measures are highlighted, such as NIHSS, FIM, CMSA-Arm, and CMSA-Hand, indicating the impact of stroke and potential recovery over a 6-month period.

Patient Code	Stroke Type	Vascular Territory	Affected Side	Sex	Age	Collection Days Post-ST		Medications	Co-Morbidities	NIHSS	FIM		CMSA-Arm		CMSA-Hand	
						Initial	6 Month				Initial	6 Month	Initial	6 Month	Initial	6 Month
ST-01	Ischemic	MCA	Left	M	79	2	242	ASA, Atorvastatin, Clopidogrel, Docusate Sodium, Perindopril	A Fib, Acute Renal Failure, Motor Cycle Accident 1952 - Knocked out; Right Collar Bone Fracture, Left Leg injury d/t combine accident as a teen	NaN	109	123	5	7	5	6
ST-03	Ischemic	LACUNAR	Right	M	37	5	221	ASA, Synthroid, Rosuvastatin, HCTZ, Felodipine	Hypothyriodism, HTN, Smoker (1/2 ppd x 5 years), EtOH	11	92	120	1	3	1	2
ST-05	Ischemic	MCA	Left	M	47	6	206	ASA, Crestor	HTN, Smoker (30 per day)	1	116	125	7	7	5	7
ST-06	Ischemic	MCA	Left	M	64	4	101	Warfarin, Solatol, Metoprolol, Atorvastatin, Levothyroxine, Vit D, Calcium, Magnesium, Benzaclin Pump	A Fib, Valve Disease - Dilated Cardiac Myopathy 1990, irregular heart rate in 2002, 2016, Cardiac MRI done, no clot found.	3	105	115	6	5	5	5
ST-08	Ischemic	MCA/ ACA	Left	M	62	4	200		Diabetes, Smoker, Increased Cholesterol	2	106	124	5	7	6	7

ST-09	HEM	MCA	Left	F	61	5	191	Gravol, Gabapentin, Lovenox, Lantus, Norvasc, Pantoprazole, Restorlax, Aidactone, Vit D, Ativan, Advair	HTN, Diabetes, Asthma, Barrett's Esophagus, Obstructive Sleep Apnea, Chronic Neck and Low Back Pain - Recurring PRP Treatment	9	80	124	4	7	5	6
ST-10	Ischemic	Thalamus	Left	M	72	6	189	ASA, Diamicron, Metoprolol, Fosinopril, gliclazide, plavix, metformin, atorvastatin, Vit D, Lantus, Humulin, Drug Study Rivanrobran vs Placebo)	HTN, Diabetes, Hyperlipidemia, Ischemic heart disease, CABG (Aug 2015), Post CABG enrolled in COMPESS, TIA on 2017-01-11 sent to stroke prevention clinic on 2017-01-13 and then admitted to acute stroke unit)	4	96	122	4	6	5	6
ST-11	Ischemic	MCA	Left	M	71	2	182	Topamax, Nixium, Elloquis, Vit D, Furosemide	A Fib, Chronic Nerve Pain - Prolapted Bladder - reconstructive surgery leading to chronic pain, Edema in legs, Asthma, Celiac, Seizures since 2005 - controlled with Meds	4	124	123	7	7	6	7
ST-13	Ischemic	MCA	Left	F	42	7	215	ASA, Clopidorel, Sertraline, Amitriptyline	Wolf Parkinsons White	3	NaN	123	5	7	6	6
ST-16	HEM	MCA	Left	M	62	8	213	Nicotine, Amlodipine, Baclofen, Acetominaphen	HTN, Hyperlipidemia, Smoker, Prostate Cancer, Degenerative Changes spine, Chronic Sinusitis	10	70	115	4	5	5	5
ST-17	Ischemic	MCA + ICA	Left	M	53	11	221			6	NaN	126	6	6	6	6

ST-19	HEM	PCA	Left	F	78	4	170	Amlodipine, Enoxaparin, Hydrochlorothiazide, Nitro patch Polythylene Glycol	HTN, Hyperlipidemia, Cleft Palate	4	113	123	7	7	7	7

Abbreviations: HEM = hemorrhagic, MCA = middle cerebral artery, ACA = anterior cerebral artery, PCA = posterior cerebral artery, ICA = internal carotid artery, ASA = acetylsalicylic acid, HCTZ = Hydrochlorothiazide, PRP = platelet-rich plasma, CABG = coronary artery bypass graft, HTN = hypertension, TIA = transient ischemic attack

Table 5. A total of 65 significant blood-derived stroke metabolites determined by VIAVC and paired *t*-test or Wilcoxon signed rank, presenting chemical shifts in ppm and their respective direction of regulation. Metabolites displaying multiple significant resonance peaks are designated as metabolite.1, metabolite.2, and so forth, to indicate their diverse spectral contributions.

Metabolite	Chemical Shift (ppm)	VIAVC *p*-Value	Paired *t*/ Wilcoxon *p*-Value	Regulation
3-Hydroxyisovaleric acid	1.254	4.21E-23	0.0173	Down
Mannose.1	5.197	1.75E-16	0.0338	Down
Alloisoleucine	0.982	5.30E-16	0.6441	Down
2-hydroxybutyric acid	0.903	9.34E-16	0.2558	Down
Formate	8.464	8.78E-14	0.9202	Up
Glucose.5	3.245	1.89E-11	0.0343	Down
Dimethyl sulfone	3.162	4.34E-10	0.0151	Down
3-Methylhistamine	3.075	1.15E-08	0.0708	Up
3-Methyl-2-Oxovaleric acid	1.097	1.16E-08	0.0625	Up
L-phenylalanine	7.435	3.49E-08	0.7334 (W)	Up
Proline.4	3.336	Not Sig.	< 0.0001	Down
Proline.3	3.342	Not Sig.	0.001	Down
Proline.18	2.006	Not Sig.	0.0019	Down
Proline.19	1.994	Not Sig.	0.0019	Down
Glycerol.2	3.806	Not Sig.	0.003	Up
Proline.20	1.984	Not Sig.	0.0031	Down
Proline.17	2.017	Not Sig.	0.0033	Down
Glycerol.5	3.655	Not Sig.	0.0036	Up
Acetoacetate	3.458	Not Sig.	0.0042	Down
Proline.13	2.054	Not Sig.	0.0044	Down
Alanine.1	1.492	Not Sig.	0.0047	Down
Proline.7	2.356	Not Sig.	0.005	Down
Proline.16	2.027	Not Sig.	0.0055	Down
Proline.6	2.369	Not Sig.	0.0056	Down
Proline.12	2.063	Not Sig.	0.0067	Down
Tryptophan	3.331	Not Sig.	0.0076	Down
Creatine	3.039	Not Sig.	0.0076	Down
Alanine.2	1.479	Not Sig.	0.0077	Down
Glycerol.3	3.802	Not Sig.	0.0083	Up
Proline.10	2.081	Not Sig.	0.009	Down
Mannose.2	3.85	Not Sig.	0.01	Down
Proline.21	1.975	Not Sig.	0.0102	Down
Proline.11	2.072	Not Sig.	0.0112	Down
Proline.15	2.037	Not Sig.	0.0125	Down
Proline.14	2.046	Not Sig.	0.0153	Down
Glycine	3.567	Not Sig.	0.016	Down
Glycerol.4	3.799	Not Sig.	0.019	Up
L-isoleucine	3.683	Not Sig.	0.0202	Up
Pyruvate	2.382	Not Sig.	0.0202	Up
3-Hydroxybutyric acid	1.212	Not Sig.	0.022	Down
Proline.22	1.966	Not Sig.	0.0225	Down
Trimethylamine-N-oxide	3.276	Not Sig.	0.024	Up
Fucose.2	1.237	Not Sig.	0.025	Down
Glucose.2	3.853	Not Sig.	0.0273	Down

Glycerol.1	3.813	Not Sig.	0.028	Up
Glycerol.6	3.647	Not Sig.	0.0359	Up
Glucose.1	5.239	Not Sig.	0.0367	Down
Glucose.4	3.462	Not Sig.	0.039	Down
Kynurenine.1	4.145	Not Sig.	0.0393	Up
Glutamine.1	3.78	Not Sig.	0.0399	Down
Glutamine.2	3.772	Not Sig.	0.0428	Down
Xylitol	3.669	Not Sig.	0.0429	Up
Hippuric acid	3.973	Not Sig.	0.0437	Up
Fucose.1	1.246	Not Sig.	0.0445	Down
N6-Acetyl-L-lysine	1.419	Not Sig.	0.0483	Up
Proline.1	3.352	Not Sig.	0.0010 (W)	Down
Proline.2	3.347	Not Sig.	0.0015 (W)	Down
Proline.8	2.346	Not Sig.	0.0093 (W)	Down
Proline.5	6.652	Not Sig.	0.0161 (W)	Down
Proline.9	2.34	Not Sig.	0.0161 (W)	Down
Acetic acid	1.925	Not Sig.	0.0161 (W)	Down
Caffeine	3.356	Not Sig.	0.0210 (W)	Up
Carnitine	3.233	Not Sig.	0.0210 (W)	Down
Glucose.3	3.766	Not Sig.	0.0425 (W)	Down
Kynurenine.2	3.688	Not Sig.	0.0425 (W)	Up

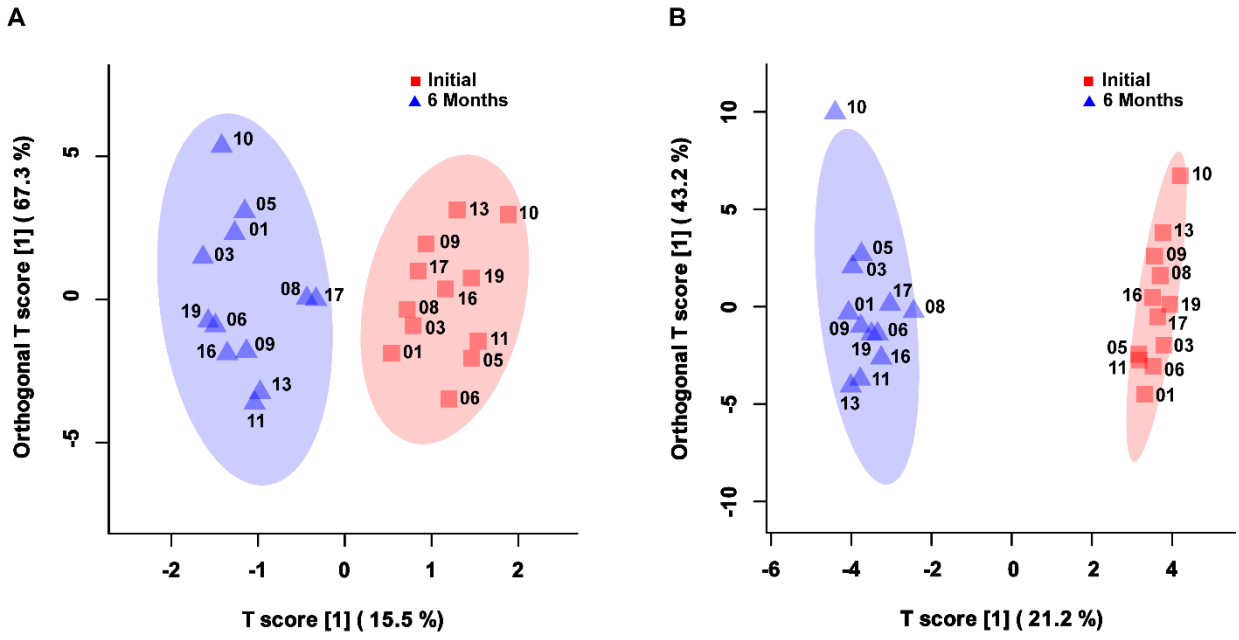

Figure 6. Orthogonal projections to latent structures discriminant analysis (OPLS-DA) scores plots reveal clear discrimination between metabolites from initial blood sample collection (red squares) and the 6 months post-stroke sample collection (indigo triangles). The x-axis showcases the predictive variation between the groups, while the y-axis visually represents the orthogonal variation within each group. Panel A, using a subset of 10 significant metabolites identified by VIAVC best subset analysis, displays Q^2 and R^2Y values of 0.854 (p<5e-04) and 0.898 (p<5e-04), respectively. Panel B, based on a subset of 65 metabolites significant by both paired *t*-test and VIAVC best subset analysis, demonstrates a Q^2 value of 0.81 (p<5e-04) and an R^2Y value of 0.987 (p<5e-04).

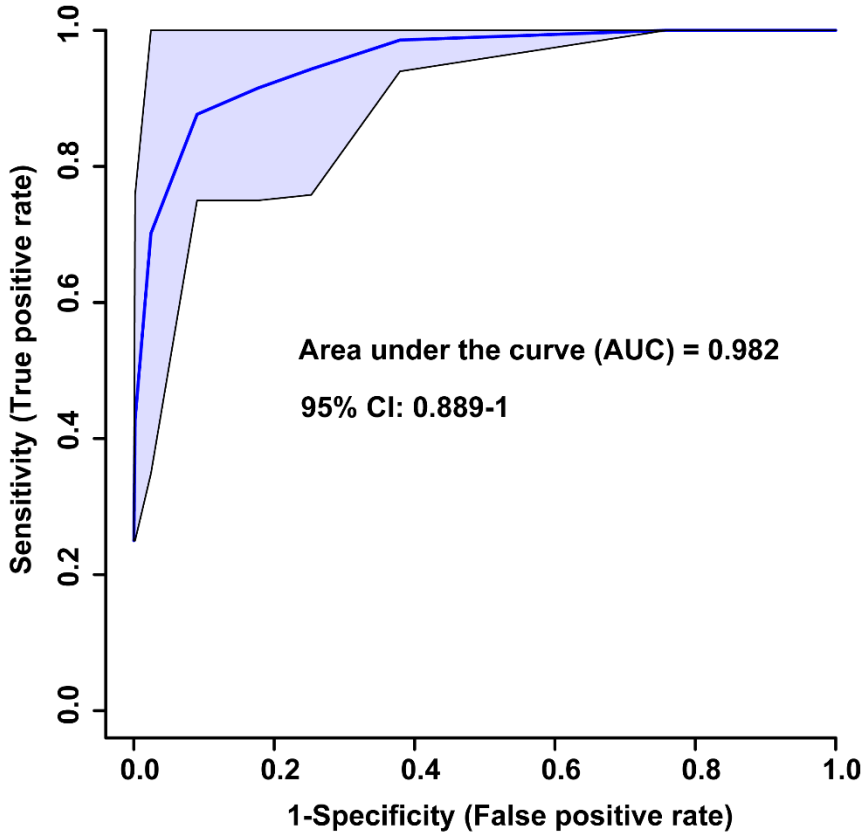

Figure 7. The receiver operator characteristic (ROC) curve, constructed using VIAVC best subset analysis (10 of the 287 total bins), showcases exceptional sensitivity and specificity in distinguishing initial stroke and 6 months post-stroke samples, yielding a robust area under the curve (AUC) of 0.982 with a 95% confidence interval of 0.889-1 and a predictive accuracy of 85%.

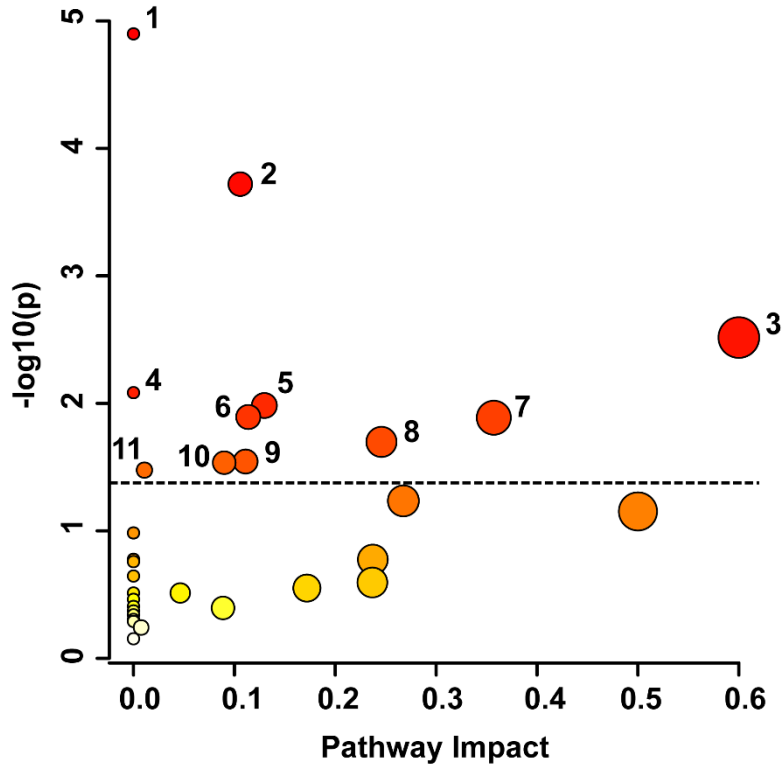

Figure 8. Metabolomic pathway topology analysis reveals the significant biochemical alterations and impact of blood-derived metabolites, identified through paired *t*-test and VIAVC best subset analysis, consisting of 65 of the 287 total bins. The y-axis portrays the *p*-values associated with each pathway and the x-axis illustrates the pathway impact, capturing the number of differential metabolites. Each pathway is represented by circles in the figure, where the colour denotes the level of significance (red represents a low *p*-value), and the size corresponds to the magnitude of pathway impact (larger circles indicate larger impact). Significant pathways surpassing the significance threshold marked by the dotted line at $p<0.05$ are labelled as follows: 1. Aminoacyl-tRNA Biosynthesis (*p*=1.26E-05), 2. Glyoxylate and Dicarboxylate Metabolism (*p*=1.91E-04), 3. Synthesis and Degradation of Ketone Bodies (*p*=0.0030), 4. Valine, Leucine and Isoleucine Biosynthesis (*p*=0.0082), 5. Glycolysis/Gluconeogenesis (*p*=0.010), 6. Alanine, Aspartate and

Glutamate Metabolism (p=0.013), 7. Phenylalanine Metabolism (p=0.013), 8. Glycine, Serine and Threonine Metabolism (p=0.020), 9. Butanoate Metabolism (p=0.029), 10. Arginine and Proline Metabolism (p=0.029), and 11. Valine, Leucine and Isoleucine Degradation (p=0.033).

Table 6. Spearman correlation analysis between blood-based metabolites and clinical parameters, presenting correlation coefficients (R) and corresponding p-values for three distinct analyses: (1) initial metabolite concentration to initial FIM scores, (2) initial metabolite concentration to percent difference in FIM scores, and (3) percent difference in metabolite concentration to percent difference in FIM and CMSA-hand scores. "*" signifies Bonferroni corrected significance.

Metabolite	Correlation Values
Initial Metabolite Concentration to Initial FIM Score	
Alloisoleucine	R = 0.673, p = 0.033
Initial Metabolite Concentration to Percent Difference FIM Score	
3-Methyl-2-Oxovaleric acid	R = -0.661, p = 0.038
Alloisoleucine	R = -0.745, p = 0.013
2-hydroxybutyric acid	R = -0.673, p = 0.033
Percent Difference Metabolite Concentration to Percent Difference Clinical Score	
FIM	
Dimethyl Sulfone	R = -0.842, p = 0.002*
CMSA-hand	
Mannose	R = 0.619, p = 0.032
Alloisoleucine	R = 0.604, p = 0.038

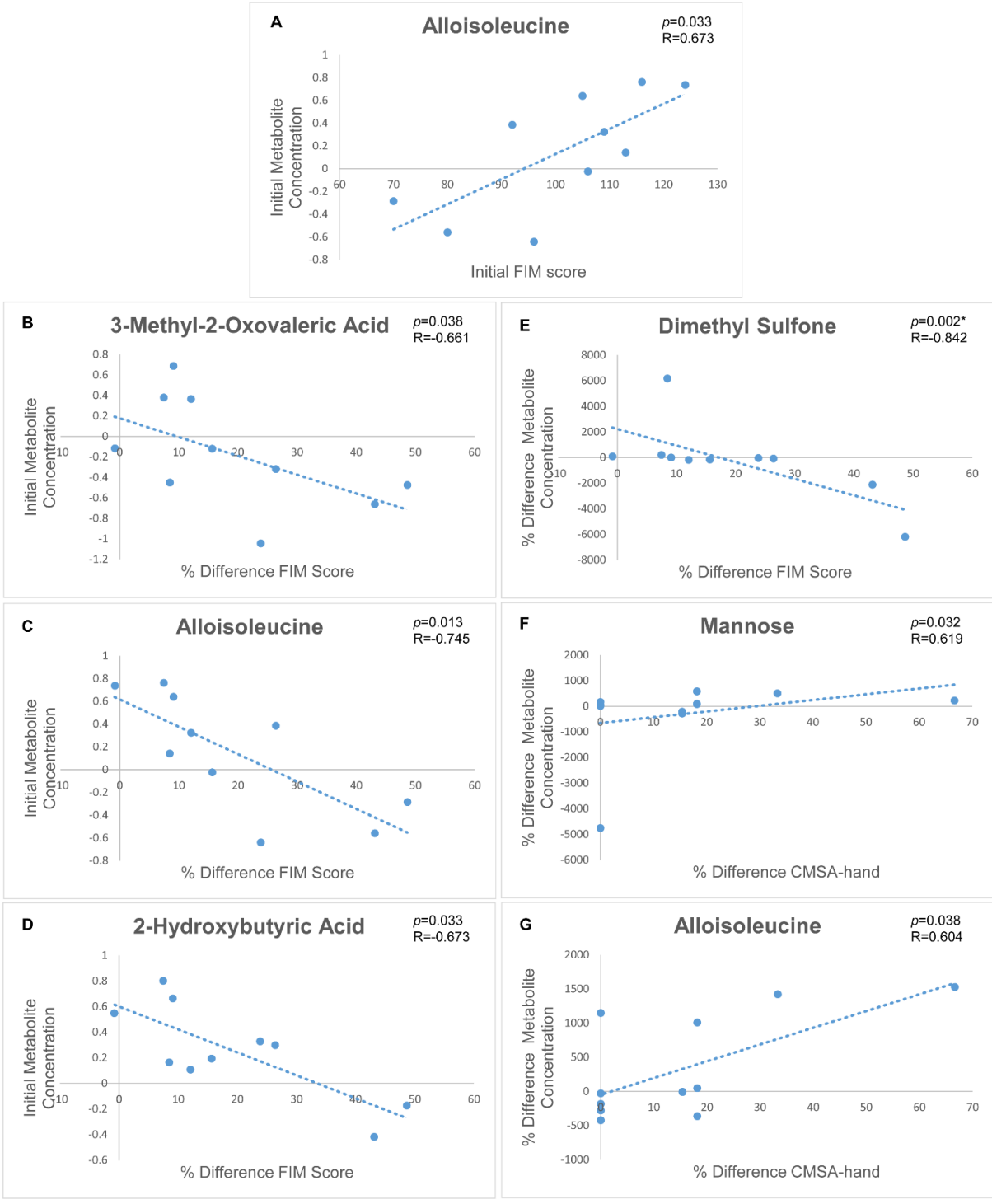

Figure 9. Scatterplots that visually depict the significant Spearman correlations (R) and *p*-values derived from three separate analytical approaches: (A) the relationship between initial metabolite

concentration and initial FIM scores, (B-D) the association between initial metabolite concentration and percent difference in FIM scores, and (E-G) the correlation between percent difference in metabolite concentration and percent difference in FIM and CMSA-hand scores. The x-axis represent the initial or percent difference in clinical scores, while the y-axis presents the initial or percent difference in metabolite concentration, providing a visual depiction of the relationships between the clinical measures and metabolite levels.

REFERENCES

Ahn, H., Kim, J., Lee, M., Kim, Y. J., Cho, Y., & Lee, G. (2015). Methylsulfonylmethane inhibits NLRP3 inflammasome activation. *Cytokine (Philadelphia, Pa.), 71*(2), 223-231. https://doi.org/10.1016/j.cyto.2014.11.001

Ahn, S., Jung, J., Jang, I., Madsen, E. L., & Park, W. (2016). Role of glyoxylate shunt in oxidative stress response. *The Journal of Biological Chemistry, 291*(22), 11928-11938. https://doi.org/10.1074/jbc.M115.708149

Amaral, A. U., & Wajner, M. (2022). Pathophysiology of maple syrup urine disease: Focus on the neurotoxic role of the accumulated branched-chain amino acids and branched-chain α-keto acids. *Neurochemistry International, 157.* https://doi.org/10.1016/j.neuint.2022.105360

Amirshahrokhi, K., Bohlooli, S., & Chinifroush, M. M. (2011). The effect of methylsulfonylmethane on the experimental colitis in the rat. *Toxicology and Applied Pharmacology, 253*(3), 197-202. https://doi.org/10.1016/j.taap.2011.03.017

Amirshahrokhi, K., & Niapour, A. (2022). Methylsulfonylmethane protects against ethanol-induced brain injury in mice through the inhibition of oxidative stress, proinflammatory mediators and apoptotic cell death. *International Immunopharmacology, 106,* 108638-108638. https://doi.org/10.1016/j.intimp.2022.108638

Anderson, P., Mahle, D., Doom, T., Reo, N., DelRaso, N., & Raymer, M. (2011). Dynamic adaptive binning: An improved quantification technique for NMR spectroscopic data. *Metabolomics, 7*(2): 179-190. https://link.springer.com/article/10.1007/s11306-010-0242-7 Accessed January 15, 2020.

Baker, E. W., Henderson, W. M., Kinder, H. A., Hutcheson, J. M., Platt, S. R., & West, F. D. (2018). Scaled traumatic brain injury results in unique metabolomic signatures between gray matter, white matter, and serum in a piglet model. *PloS One, 13*(10), e0206481-e0206481. https://doi.org/10.1371/journal.pone.0206481

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B, Methodological, 26*(2), 211-252. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x

Bridi, R., Araldi, J., Sgarbi, M. B., Testa, C. G., Durigon, K., Wajner, M., & Dutra-Filho, C. S. (2003). Induction of oxidative stress in rat brain by the metabolites accumulating in maple syrup urine disease. *International Journal of Developmental Neuroscience, 21*(6), 327-332. https://doi.org/10.1016/S0736-5748(03)00074-1

Bridi, R., Braun, C., Zorzi, G., Wannmacher, C., Wajner, M., Lissi, E., & Dutra, C. (2005). alpha-keto acids accumulating in maple syrup urine disease stimulate lipid peroxidation and reduce antioxidant defences in cerebral cortex from young rats. *Metabolic Brain Disease, 20*(2), 155-167. https://doi.org/10.1007/s11011-005-4152-8

Brott, T., Adams, J., H P, Olinger, C. P., Marler, J. R., Barsan, W. G., Biller, J., . . . Hertzberg, V. (1989). Measurements of acute cerebral infarction: A clinical examination scale. *Stroke, 20*(7), 864-870.

Butawan, M., Benjamin, R. L., & Bloomer, R. J. (2017). Methylsulfonylmethane: Applications and safety of a novel dietary supplement. *Nutrients, 9*(3), 290. https://doi.org/10.3390/nu9030290

Bykowski, E. A., Petersson, J. N., Dukelow, S., Ho, C., Debert, C. T., Montina, T., & Metz, G. A. S. (2021a). Urinary biomarkers indicative of recovery from spinal cord injury: A pilot study. *IBRO Neuroscience Reports, 10*(1), 178-185. https://doi.org/10.1016/j.ibneur.2021.02.007

Bykowski, E. A., Petersson, J. N., Dukelow, S., Ho, C., Debert, C. T., Montina, T., & Metz, G. A. S. (2021b). Urinary metabolomic signatures as indicators of injury severity following traumatic brain injury: A pilot study. *IBRO Neuroscience Reports, 11*, 200-206. https://doi.org/10.1016/j.ibneur.2021.10.003

Bykowski, E. A., Petersson, J. N., Dukelow, S., Ho, C., Debert, C. T., Montina, T., & Metz, G. A. S. (2023). Identification of serum metabolites as prognostic biomarkers following spinal cord injury: A pilot study. *Metabolites, 13*(5), 605. https://doi.org/10.3390/metabo13050605

Campbell, B. C. V. & Khatri, P. (2020). Stroke. *The Lancet (British Edition), 396*(10244), 129. https://doi.org/10.1016/S0140-6736(20)31179-X

Chen, Q., Zhou, T., Yuan, J., Xiong, X., Liu, X., Qiu, Z., Hu, L., Lu, H., He, Q., Liu, C., & Yang, Q. (2023). Metabolomics profiling to characterize cerebral ischemia-reperfusion injury in mice. *Frontiers in Pharmacology, 14*, 1091616-1091616. https://doi.org/10.3389/fphar.2023.1091616

Chuang, D.T., & Shih, V.E. (2001). Maple syrup urine disease (branched-chain ketoaciduria). In C.R. Scriver, A.L. Beaudet, W.L. Sly, and D. Valle (Eds.), *The Metabolic and Molecular Bases of Inherited Disease* (8th ed., pp.1971–2005). McGraw-Hill, New York.

Cobb, J., Eckhart, A., Motsinger-Reif, A., Carr, B., Groop, L., & Ferrannini, E. (2016). alpha-hydroxybutyric acid is a selective metabolite biomarker of impaired glucose tolerance. *Diabetes Care, 39*(6), 988-995. https://doi.org/10.2337/dc15-2752

Craig, A., Cloarec, O., Holmes, E., Nicholson, J. K., & Lindon, J. C. (2006). Scaling and normalization effects in NMR spectroscopic metabonomic data sets. *Analytical Chemistry (Washington), 7*8(7), 2262-2267. https://doi.org/10.1021/ac0519312

Dagonnier, M., Donnan, G. A., Davis, S. M., Dewey, H. M., & Howells, D. W. (2021). Acute stroke biomarkers: Are we there yet? *Frontiers in Neurology, 12*, 619721-619721. https://doi.org/10.3389/fneur.2021.619721

Debik, J., Sangermani, M., Wang, F., Madssen, T. S., & Giskeødegård, G. F. (2022). Multivariate analysis of NMR-based metabolomic data. *NMR in Biomedicine, 35*(2), e4638-n/a. https://doi.org/10.1002/nbm.4638

de Vries, B., Walter, S. J., Peutz-Kootstra, C. J., Wolfs, T. G. A. M., van Heurn, L. W. E., & Buurman, W. A. (2004). The mannose-binding lectin-pathway is involved in complement activation in the course of renal ischemia-reperfusion injury. *The American Journal of Pathology, 165*(5), 1677-1688. https://doi.org/10.1016/S0002-9440(10)63424-4

Dinga, R., Schmaal, L., Penninx, B. W., Veltman, D. J., & Marquand, A. F. (2020). Controlling for effects of confounding variables on machine learning predictions. *BioRxiv*. https://doi.org/10.1101/2020.08.17.255034

Dolan, S. K., & Welch, M. (2018). The glyoxylate shunt, 60 years on. *Annual Review of Microbiology, 72*(1), 309-330. https://doi.org/10.1146/annurev-micro-090817-062257

Durand, G., & Seta, N. (2000). Protein glycosylation and diseases: Blood and urinary oligosaccharides as markers for diagnosis and therapeutic monitoring. *Clinical Chemistry (Baltimore, Md.), 46*(6), 795-805. https://doi.org/10.1093/clinchem/46.6.795

Emwas, A. M. (2015). The strengths and weaknesses of NMR spectroscopy and mass spectrometry with particular focus on metabolomics research. *Methods in Molecular Biology (Clifton, N.J.)*, 1277, 161.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861-874. https://doi.org/10.1016/j.patrec.2005.10.010

Fortin, E., Ferrannini, G., Campi, B., Mellbin, L., Norhammar, A., Nasman, P., Saba, A., Ferrannini, E., & Ryden, L. (2022). Plasma mannose as a novel marker of myocardial infarction across different glycaemic states: A case control study. *Cardiovascular Diabetology, 21*(1), 1-195. https://doi.org/10.1186/s12933-022-01630-5

Funchal, C., Gottfried, C., Vieira De Almeida, Lúcia Maria, Wajner, M., & Pessoa-Pureur, R. (2004a). Evidence that the branched-chain α-keto acids accumulating in maple syrup urine disease induce morphological alterations and death in cultured astrocytes from rat cerebral cortex. *Glia, 48*(3), 230-240. https://doi.org/10.1002/glia.20072

Funchal, C., Rosa, A., Wajner, M., Wofchuk, S., & Pureur, R. (2004b). Reduction of glutamate uptake into cerebral cortex of developing rats by the branched-chain alpha-keto acids accumulating in maple syrup urine disease. *Neurochemical Research, 29*(4), 747-753. https://doi.org/10.1023/B:NERE.0000018846.66943.30

Galea, I., Palin, K., Newman, T. A., Van Rooijen, N., Perry, V. H., & Boche, D. (2005). Mannose receptor expression specifically reveals perivascular macrophages in normal, injured, and diseased mouse brain. *Glia, 49*(3), 375-384. https://doi.org/10.1002/glia.20124

Gall, W. E., Beebe, K., Lawton, K. A., Adam, K., Mitchell, M. W., Nakhle, P. J., Ryals, J. A., Milburn, M. V., Nannipieri, M., Camastra, S., Natali, A., Ferrannini, E., RISC Study Grp, RISC Study Group, & for the the RISC Study Group. (2010). alpha-hydroxybutyrate is an early biomarker of insulin resistance and glucose intolerance in a nondiabetic population. *PloS One, 5*(5), e10883-e10883. https://doi.org/10.1371/journal.pone.0010883

García-Aguilera, M. E., de San Miguel, Eduardo Rodríguez, Cruz-Pérez, J., Aguirre-Cruz, L., Ramirez-Alfaro, C. M., & Esturau-Escofet, N. (2021). NMR-based metabolomics of human cerebrospinal fluid identifies signature of brain death. *Metabolomics, 17*(5), 40-40. https://doi.org/10.1007/s11306-021-01794-3

Gil, R. B., Lehmann, R., Schmitt-Kopplin, P., & Heinzmann, S. S. (2016). 1H NMR-based metabolite profiling workflow to reduce inter-sample chemical shift variations in urine

samples for improved biomarker discovery. *Analytical and Bioanalytical Chemistry, 408*(17), 4683-4691. https://doi.org/10.1007/s00216-016-9552-6

Giraldi-Guimarães, A., de Freitas, H. T., Coelho, B. d. P., Macedo-Ramos, H., Mendez-Otero, R., Cavalcante, L. A., & Baetas-da-Cruz, W. (2012). Bone marrow mononuclear cells and mannose receptor expression in focal cortical ischemia. *Brain Research, 1452*, 173-184. https://doi.org/10.1016/j.brainres.2012.03.002

Goodpaster, A., Romick-Rosendale, L., & Kennedy, M. (2010). Statistical significance analysis of nuclear magnetic resonance-based metabonomics data. *Analytical Biochemistry, 401*: 134-143.

Gowland, C., Stratford, P., Ward, M., Moreland, J., Torresin, W., Vanhullenaar, S., Sanford, J., Barreca, S., Vanspall, B., & Plews, N. (1993). Measuring physical impairment and disability with the chedoke-mcmaster stroke assessment. *Stroke (1970), 24*(1), 58-63. https://doi.org/10.1161/01.STR.24.1.58

Grotta, J., Chin, D., Lu, M., Patel, S., Levine, Tilley, B., Brott, T., Haley, E., Lyden, P., Kothari, R., Frankel, M., Lewandowski, C., Libman, R., Kwiatkowski, T., Broderick, J., Marler, Corrigan, J., Huff, S., Mitsias, P., . . . NINDS rt-PA Stroke Trial Study Grp. (1999). Agreement and variability in the interpretation of early CT changes in stroke patients qualifying for intravenous rtPA therapy. *Stroke (1970), 30*(8), 1528-1533. https://doi.org/10.1161/01.STR.30.8.1528

Guzman, M., & Blazquez, C. (2004). *Ketone body synthesis in the brain: Possible neuroprotective effects*. Elsevier Ltd. https://doi.org/10.1016/j.plefa.2003.05.001

Hauschild, T. C., Guerreiro, G., Mescka, C. P., Coelho, D. M., Steffens, L., Moura, D. J., Manfredini, V., & Vargas, C. R. (2019). DNA damage induced by alloisoleucine and other metabolites in maple syrup urine disease and protective effect of l-carnitine. *Toxicology in Vitro, 57*, 194-202. https://doi.org/10.1016/j.tiv.2019.03.007

Holmes, E., Ebbels, T. M. D., Nicholson, J. K., Beckonert, O., Keun, H. C., Lindon, J. C., & Bundy, J. (2007). Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nature Protocols, 2*(11), 2692-2703. doi:10.1038/nprot.2007.376

Hu, Z., Zhu, Z., Cao, Y., Wang, L., Sun, X., Dong, J., Fang, Z., Fang, Y., Xu, X., Gao, P., & Hongzhi, S. (2016). Rapid and sensitive differentiating ischemic and hemorrhagic strokes by dried blood spot based direct injection mass spectrometry metabolomics analysis. *Journal of Clinical Laboratory Analysis, 30*(6), 823-830. https://doi.org/10.1002/jcla.21943

Human Metabolome Database (n.d.a). *Showing metabocard for 3-methyl-2-oxovaleric acid (HMDB0000491)*. Retrieved June 26, 2023, from https://hmdb.ca/metabolites/HMDB0000491

Human Metabolome Database (n.d.b). *Showing metabocard for 2-hydroxybutyric acid (HMDB0000491)*. Retrieved June 26, 2023, from https://hmdb.ca/metabolites/HMDB0000008

Ibba, M., & Söll, D. (2000). aminoacyl-trna synthesis. *Annual Review of Biochemistry, 69*(1), 617-650. https://doi.org/10.1146/annurev.biochem.69.1.617

Jain, S. K., Kannan, K., & Lim, G. (1998). Ketosis (acetoacetate) can generate oxygen radicals and cause increased lipid peroxidation and growth inhibition in human endothelial cells. *Free Radical Biology & Medicine, 25*(9), 1083-1088. https://doi.org/10.1016/S0891-5849(98)00140-3

Jia, J., Zhang, H., Liang, X., Dai, Y., Liu, L., Tan, K., Ma, R., Luo, J., Ding, Y., & Ke, C. (2021). Application of metabolomics to the discovery of biomarkers for ischemic stroke in the murine model: A comparison with the clinical results. *Molecular Neurobiology, 58*(12), 6415-6426. https://doi.org/10.1007/s12035-021-02535-2

Kanikarla-Marie, P., & Jain, S. K. (2015). Hyperketonemia (acetoacetate) upregulates NADPH oxidase 4 and elevates oxidative stress, ICAM-1, and monocyte adhesivity in endothelial cells. *Cellular Physiology and Biochemistry, 35*(1), 364-373. https://doi.org/10.1159/000369702

Karavanaki, K., Kakleas, K., Georga, S., Bartzeliotou, A., Mavropoulos, G., Tsouvalas, M., Vogiatzi, A., Papassotiriou, I., & Karayianni, C. (2012). Plasma high sensitivity C-reactive protein and its relationship with cytokine levels in children with newly diagnosed type 1 diabetes and ketoacidosis. *Clinical Biochemistry, 45*(16-17), 1383-1388. https://doi.org/10.1016/j.clinbiochem.2012.05.003

Kiss, D., Ambeskovic, M., Montina, T., & Metz, G. A. S. (2016). Stress transgenerationally programs metabolic pathways linked to altered mental health. *Cellular and Molecular Life Sciences : CMLS, 73*(23), 4547-4557. https://doi.org/10.1007/s00018-016-2272-4

Ke, C., Pan, C., Zhang, Y., Zhu, X., & Zhang, Y. (2019). Metabolomics facilitates the discovery of metabolic biomarkers and pathways for ischemic stroke: A systematic review. *Metabolomics, 15*(12), 1-21. https://doi.org/10.1007/s11306-019-1615-1

Keith, R. A., Granger, C. V., Hamilton, B. B., & Sherwin, F. S. (1987). The functional independence measure: A new tool for rehabilitation. *Advances in Clinical Rehabilitation, 1*, 6-18.

Koen, N., Du Preez, I., & Loots, D. T. (2016). Metabolomics and personalized medicine. In R. Donev (Ed.), *Advances in protein chemistry and structural biology* (pp. 53-78). Elsevier. https://doi.org/10.1016/bs.apcsb.2015.09.003

Kolokolova, T. N., Savel'ev, O. Y., Sergeev, N. M., Shpigun, O. A., Sokolov, K. V., & Skvortsova, V. I. (2010). Nuclear magnetic resonance spectroscopy in solving the analytical problems of medicine: Analysis of cerebrospinal fluid. *Journal of Analytical Chemistry (New York, N.Y.), 65*(10), 1073-1081. https://doi.org/10.1134/S106193481010014X

Laborde, C. M., Mourino-Alvarez, L., Akerstrom, F., Padial, L. R., Vivanco, F., Gil-Dones, F., & Barderas, M. G. (2012). Potential blood biomarkers for stroke. *Expert Review of Proteomics, 9*(4), 437-449. doi:10.1586/EPR.12.33

Landaas, S., & Pettersen, J. (1975). Clinical conditions associated with urinary-excretion of 2-hydroxybutyric acid. *Scandinavian Journal of Clinical and Laboratory Investigation, 35*(3), 259-266. https://doi.org/10.3109/00365517509095738

Laursen, M. R., Hansen, J., Elkjær, C., Stavnager, N., Nielsen, C. B., Pryds, K., Johnsen, J., Nielsen, J. M., Bøtker, H. E., & Johannsen, M. (2017). Untargeted metabolomics reveals a mild impact of remote ischemic conditioning on the plasma metabolome and α-hydroxybutyrate as a possible cardioprotective factor and biomarker of tissue ischemia. *Metabolomics, 13*(6), 67-13. https://doi.org/10.1007/s11306-017-1202-2

Lee, S., Zhang, C., Kilicarslan, M., Piening, B., Bjornson, E., Hallström, B., Groen, A., Ferrannini, E., Laakso, M., Snyder, M., Blüher, M., Uhlen, M., Nielsen, J., Smith, U., Serlie, M., Boren, J., Mardinoglu, A., Sahlgrenska akademin, Institute of Medicine, Department of Molecular and Clinical Medicine, . . . Sahlgrenska Academy. (2016). Integrated network analysis reveals an association between plasma mannose levels and insulin resistance. *Cell Metabolism, 24*(1), 172-184. https://doi.org/10.1016/j.cmet.2016.05.026

Lin, A., Nguy, C. H., Shic, F., & Ross, B. D. (2001). Accumulation of methylsulfonylmethane in the human brain: Identification by multinuclear magnetic resonance spectroscopy. *Toxicology Letters, 123*(2), 169-177. https://doi.org/10.1016/S0378-4274(01)00396-4

Lin, Y., Yang, D., Ni, H., Xu, X., Wu, F., Lin, L., Chen, J., Sun, Y., Huang, Z., Li, S., Jiang, P., Wu, H., Chang, L., Hu, B., Luo, C., Wu, J., & Zhu, D. (2023). Ketone bodies promote stroke recovery via GAT-1-dependent cortical network remodeling. *Cell Reports (Cambridge), 42*(4), 112294-112294. https://doi.org/10.1016/j.celrep.2023.112294

Liu, H., Pu, J., Zhou, Q., Yang, L., & Bai, D. (2022). Peripheral blood and urine metabolites and biological functions in post-stroke depression. *Metabolic Brain Disease, 37*(5), 1557-1568. https://doi.org/10.1007/s11011-022-00984-9

Liu, P., Li, R., Antonov, A. A., Wang, L., Li, W., Hua, Y., ... & Huang, Y. (2017). Discovery of metabolite biomarkers for acute ischemic stroke progression. *Journal of proteome research, 16*(2), 773-779.

Maalouf, M., Rho, J. M., & Mattson, M. P. (2009). The neuroprotective properties of calorie restriction, the ketogenic diet, and ketone bodies. *Brain Research Reviews, 59*(2), 293-315. https://doi.org/10.1016/j.brainresrev.2008.09.002

Mamer, O. A. (2001). *Initial catabolic steps of isoleucine, the R-pathway and the origin of alloisoleucine*. Elsevier B.V. https://doi.org/10.1016/S0378-4347(01)00111-6

Maranon, G., Munoz-Escassi, B., Manley, W., Garcia, C., Cayado, P., Sanchez de la Muela, Mercedes, Olabarri, B., Leon, R., & Vara, E. (2008). The effect of methyl sulphonyl methane supplementation on biomarkers of oxidative stress in sport horses following jumping exercise. *Acta Veterinaria Scandinavica, 50*(1), 45-45. https://doi.org/10.1186/1751-0147-50-45

Menni, C., Fauman, E., Erte, I., Perry, J. R. B., Kastenmueller, G., Shin, S., Petersen, A., Hyde, C., Psatha, M., Ward, K. J., Yuan, W., Milburn, M., Palmer, C. N. A., Frayling, T. M., Trimmer, J., Bell, J. T., Gieger, C., Mohney, R. P., Brosnan, M. J., . . . Spector, T. D.

(2013). Biomarkers for type 2 diabetes and impaired fasting glucose using a nontargeted metabolomics approach. *Diabetes (New York, N.Y.), 62*(12), 4270-4276. https://doi.org/10.2337/db13-0570

Morrison, H., Frye, J., Davis-Gorman, G., Funk, J., McDonagh, P., Stahl, G., & Ritter, L. (2011). The contribution of mannose binding lectin to reperfusion injury after ischemic stroke. *Current Neurovascular Research, 8*(1), 52-63. https://doi.org/10.2174/156720211794520260

Nakhostin-Roohi, B., Barmaki, S., Khoshkhahesh, F., & Bohlooli, S. (2011). Effect of chronic supplementation with methylsulfonylmethane on oxidative stress following acute exercise in untrained healthy men. *Journal of Pharmacy and Pharmacology, 63*(10), 1290-1294. https://doi.org/10.1111/j.2042-7158.2011.01314.x

Newman, J. C., & Verdin, E. (2014). Ketone bodies as signaling metabolites. *Trends in Endocrinology and Metabolism, 25*(1), 42-52. https://doi.org/10.1016/j.tem.2013.09.002

Nicholson, J. K., & Lindon, J. C. (2008). Metabonomics. *Nature (London), 455*(7216), 1054-1056. https://doi.org/10.1038/4551054a

Nicholson, J. K., Lindon, J. C., & 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), 1181-1189. https://doi.org/10.1080/004982599238047

Novokmet, M., Lukic, E., Vuckovic, F., Duric, Z., Keser, T., Rajsl, K., Remondini, D., Castellani, G., Gasparovic, H., Gornik, O., & Lauc, G. (2014). Changes in IgG and total plasma protein glycomes in acute systemic inflammation. *Scientific Reports, 4*(1), 4347-4347. https://doi.org/10.1038/srep04347

Pang, Z., Chong, J., Li, S., & Xia, J. (2020). MetaboAnalystR 3.0: Toward an optimized workflow for global metabolomics. *Metabolites, 10*(5), 186. https://doi.org/10.3390/metabo10050186

Parcell, S. (2002). Sulfur in human nutrition and applications in medicine. *Alternative Medicine Review, 7*(1), 22.

Paschen, W. (2004). Endoplasmic reticulum dysfunction in brain pathology: Critical role of protein synthesis. *Current Neurovascular Research, 1*(2), 173-181. https://doi.org/10.2174/1567202043480125

Paschen, W., Proud, C. G., & Mies, G. (2007). Shut-down of translation, a global neuronal stress response: Mechanisms and pathological relevance. *Current Pharmaceutical Design, 13*(18), 1887-1902. https://doi.org/10.2174/138161207780858401

Paxman, E. J., Wanner, Z., Boora, J., Douglas, K., Metz, G., Montina, T., & Debert, C. (2017). Metabolomic biomarkers of concussion. *British Journal of Sports Medicine, 51*(11), A4-A4. https://doi.org/10.1136/bjsports-2016-097270.9

Pechlivanis, A., Kostidis, S., Saraslanidis, P., Petridou, A., Tsalis, G., Mougios, V., Gika, H. G., Mikros, E., & Theodoridis, G. A. (2010). 1H NMR-based metabonomic investigation of

the effect of two different exercise sessions on the metabolic fingerprint of human urine. *Journal of Proteome Research, 9*(12), 6405-6416. https://doi.org/10.1021/pr100684t

Perego, C., Fumagalli, S., & De Simoni, M. (2011). Temporal pattern of expression and colocalization of microglia/macrophage phenotype markers following brain ischemic injury in mice. *Journal of Neuroinflammation, 8*(1), 174-174. https://doi.org/10.1186/1742-2094-8-174

Pessoa-Pureur, R., & Wajner, M. (2007). Cytoskeleton as a potential target in the neuropathology of maple syrup urine disease: Insight from animal studies. *Journal of Inherited Metabolic Disease, 30*(5), 664-672. https://doi.org/10.1007/s10545-007-0562-6

Pikija, S., Trkulja, V., Simundic, A., Vrcek, E., Boskovic, K., & Bacani, S. (2013). Is on-admission capillary blood beta-hydroxybutyrate concentration associated with the acute stroke severity and short-term functional outcome? *Neurological Research (New York), 35*(9), 959-967. https://doi.org/10.1179/1743132813Y.0000000239

Plancke, C., Vigeolas, H., Höhner, R., Roberty, S., Emonds-Alt, B., Larosa, V., Willamme, R., Duby, F., Onga Dhali, D., Thonart, P., Hiligsmann, S., Franck, F., Eppe, G., Cardol, P., Hippler, M., & Remacle, C. (2014). Lack of isocitrate lyase in chlamydomonas leads to changes in carbon metabolism and in the response to oxidative stress under mixotrophic growth. *The Plant Journal : For Cell and Molecular Biology, 77*(3), 404-417. https://doi.org/10.1111/tpj.12392

Prins, M. L. (2008). Cerebral metabolic adaptation and ketone metabolism after brain injury. *Journal of Cerebral Blood Flow & Metabolism, 28*(1), 1-16. https://doi.org/10.1038/sj.jcbfm.9600543

Psychogios, N., Hau, D. D., Peng, J., Guo, A. C., Mandal, R., Bouatra, S., Sinelnikov, I., Krishnamurthy, R., Eisner, R., Gautam, B., Young, N., Xia, J., Knox, C., Dong, E., Huang, P., Hollander, Z., Pedersen, T. L., Smith, S. R., Bamforth, F., . . . Wishart, D. S. (2011). The human serum metabolome. *PloS One, 6*(2), e16957-e16957. https://doi.org/10.1371/journal.pone.0016957

Qi, B., Zhang, Y., Xu, B., Zhang, Y., Fei, G., Lin, L., & Li, Q. (2022). Metabolomic characterization of acute ischemic stroke facilitates metabolomic biomarker discovery. *Applied Biochemistry and Biotechnology, 194*(11), 5443-5455. https://doi.org/10.1007/s12010-022-04024-1

Qureshi, M. I., Vorkas, P. A., Coupland, A. P., Jenkins, I. H., Holmes, E., & Davies, A. H. (2017). Lessons from metabonomics on the neurobiology of stroke. *SAGE Publications*. https://doi.org/10.1177/1073858416673327

Ramoutar, R. R., & Brumaghim, J. L. (2010). Antioxidant and anticancer properties and mechanisms of inorganic selenium, oxo-sulfur, and oxo-selenium compounds. *Cell Biochemistry and Biophysics, 58*(1), 1-23. https://doi.org/10.1007/s12013-010-9088-x

Rasooli, S. A., Fathi, R., Golzar, F. A., & Baghersalimi, M. (2021). The effect of circuit resistance training on plasma levels of amino acids, alpha-hydroxybutyrate, mannose, and urinary levels of glycine conjugated adducts in obese adolescent boys. *Applied Physiology, Nutrition, and Metabolism, 46*(6), 561-570. https://doi.org/10.1139/apnm-2020-0171

Rosea, S. E., Chalk, J. B., Galloway, G. J., & Doddrell, D. M. (2000). Detection of dimethyl sulfone in the human brain by in vivo proton magnetic resonance spectroscopy. *Magnetic Resonance Imaging, 18*(1), 95-98. https://doi.org/10.1016/S0730-725X(99)00110-1

Roth, G. A., Mensah, G. A., Johnson, C. O., Addolorato, G., Ammirati, E., Baddour, L. M., Barengo, N. C., Beaton, A. Z., Benjamin, E. J., Benziger, C. P., Bonny, A., Brauer, M., Brodmann, M., Cahill, T. J., Carapetis, J., Catapano, A. L., Chugh, S. S., Cooper, L. T., Coresh, J., . . . GBD-NHLBI-JACC Global Burden of Cardiovascular Diseases Writing Group. (2020). Global burden of cardiovascular diseases and risk factors, 1990-2019 update from the GBD 2019 study. *Journal of the American College of Cardiology, 76*(25), 2982-3021. https://doi.org/10.1016/j.jacc.2020.11.010

Sacco, R. L., Kasner, S. E., Broderick, J. P., Caplan, L. R., Connors, J. J. (., Culebras, A., Elkind, M. S. V., George, M. G., Hamdan, A. D., Higashida, R. T., Hoh, B. L., Janis, L. S., Kase, C. S., Kleindorfer, D. O., Lee, J., Moseley, M. E., Peterson, E. D., Turan, T. N., Valderrama, A. L., . . . Council on Cardiovascular and Stroke Nursing. (2013). An updated definition of stroke for the 21st century: A statement for healthcare professionals from the american heart Association/American stroke association. *Stroke (1970), 44*(7), 2064-2089. https://doi.org/10.1161/STR.0b013e318296aeca

Schadewaldt, P., Bodner-Leidecker, A., Hammen, H., & Wendel, U. (2000). Formation of L-alloisoleucine in vivo: An L-[C-13]isoleucine study in man. *Pediatric Research, 47*(2), 271-277. https://doi.org/10.1203/00006450-200002000-00020

Sgaravatti, A. M., Rosa, R. B., Schuck, P. F., Ribeiro, C. A. J., Wannmacher, C. M. D., Wyse, A. T. S., Dutra-Filho, C. S., & Wajner, M. (2003). Inhibition of brain energy metabolism by the α-keto acids accumulating in maple syrup urine disease. *Biochimica Et Biophysica Acta, 1639*(3), 232-238. https://doi.org/10.1016/j.bbadis.2003.09.010

Shaker, M. E., Eisa, N. H., Elgaml, A., El-Mesery, A., El-Shafey, M., El-Dosoky, M., El-Mowafy, M., & El-Mesery, M. (2021). Ingestion of mannose ameliorates thioacetamide-induced intrahepatic oxidative stress, inflammation and fibrosis in rats. *Life Sciences, 286*, 120040. https://doi.org/10.1016/j.lfs.2021.120040

Sharma, R., Reinstadler, B., Engelstad, K., Skinner, O. S., Stackowitz, E., Haller, R. G., Clish, C. B., Pierce, K., Walker, M. A., Fryer, R., Oglesbee, D., Mao, X., Shungu, D. C., Khatri, A., Hirano, M., De Vivo, D. C., & Mootha, V. K. (2021). Circulating markers of NADH-reductive stress correlate with mitochondrial disease severity. *The Journal of Clinical Investigation, 131*(2), 1-16. https://doi.org/10.1172/JCI136055

Sidorov, E., Bejar, C., Xu, C., Ray, B., Reddivari, L., Chainakul, J., Vanamala, J. K. P., & Sanghera, D. K. (2020). Potential metabolite biomarkers for acute versus chronic stage of ischemic stroke: A pilot study. *Journal of Stroke and Cerebrovascular Diseases, 29*(4), 104618-104618. https://doi.org/10.1016/j.jstrokecerebrovasdis.2019.104618

Sitta, A., Ribas, G. S., Mescka, C. P., Barschak, A. G., Wajner, M., & Vargas, C. R. (2014). Neurological damage in MSUD: The role of oxidative stress. *Cellular and Molecular Neurobiology, 34*(2), 157-165. https://doi.org/10.1007/s10571-013-0002-0

Smelter, A., Rouchka, E. C., & Moseley, H. N. B. (2017). Detecting and accounting for multiple sources of positional variance in peak list registration analysis and spin system grouping.

Journal of Biomolecular NMR, 68(4), 281-296. https://doi.org/10.1007/s10858-017-0126-5

Song, S. (2000). Can the glyoxylate pathway contribute to fat-induced hepatic insulin resistance? *Medical Hypotheses, 54*(5), 739-747. https://doi.org/10.1054/mehy.1999.0943

Sousa, A. P., Cunha, D. M., Franco, C., Teixeira, C., Gojon, F., Baylina, P., & Fernandes, R. (2021). Which role plays 2-hydroxybutyric acid on insulin resistance? *Metabolites, 11*(12), 835. https://doi.org/10.3390/metabo11120835

Stein, M., Keshav, S., Harris, N., & Gordon, S. (1992). Interleukin 4 potently enhances murine macrophage mannose receptor activity: A marker of alternative immunologic macrophage activation. *The Journal of Experimental Medicine, 176*(1), 287-292. https://doi.org/10.1084/jem.176.1.287

Sun, D., Tiedt, S., Yu, B., Jian, X., Gottesman, R. F., Mosley, T. H., Boerwinkle, E., Dichgans, M., & Fornage, M. (2019). A prospective study of serum metabolites and risk of ischemic stroke. *Neurology, 92*(16), E1890-E1898. https://doi.org/10.1212/WNL.0000000000007279

Suzuki, M., Suzuki, M., Sato, K., Dohi, S., Sato, T., Matsuura, A., & Hiraide, A. (2001). Effect of β-hydroxybutyrate, a cerebral function improving agent, on cerebral hypoxia, anoxia and ischemia in mice and rats. *Japanese Journal of Pharmacology, 87*(2), 143-150. https://doi.org/10.1254/jjp.87.143

Szymanska, E., Saccenti, E., Smilde, A. K., & Westerhuis, J. A. (2012). Double-check: Validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics, 8*(Suppl 1), S3-S16. https://doi.org/10.1007/s11306-011-0330-3

Tang, Y., & Le, W. (2016). Differential roles of M1 and M2 microglia in neurodegenerative diseases. *Molecular Neurobiology, 53*(2), 1181-1194. https://doi.org/10.1007/s12035-014-9070-5

Tiedt, S., Brandmaier, S., Kollmeier, H., Duering, M., Artati, A., Adamski, J., Klein, M., Liebig, T., Holdt, L. M., Teupser, D., Wang-Sattler, R., Schwedhelm, E., Gieger, C., & Dichgans, M. (2020). Circulating metabolites differentiate acute ischemic stroke from stroke mimics. *Annals of Neurology, 88*(4), 736-746. https://doi.org/10.1002/ana.25859

Timpone, V. M., Jensen, A., Poisson, S. N., & Trivedi, P. S. (2020). Compliance with imaging guidelines for workup of transient ischemic attack evidence from the nationwide emergency department sample. *Stroke (1970), 51*(8), 2563-2567. https://doi.org/10.1161/STROKEAHA.120.029858

van den Berg, Robert A, Hoefsloot, H. C. J., Westerhuis, J. A., Smilde, A. K., & van der Werf, Mariët J. (2006). Centering, scaling, and transformations: Improving the biological information content of metabolomics data. *BMC Genomics, 7*(1), 142-142. https://doi.org/10.1186/1471-2164-7-142

Vignoli, A., Ghini, V., Meoni, G., Licari, C., Takis, P. G., Tenori, L., Turano, P., & Luchinat, C. (2019). High-Throughput metabolomics by 1D NMR. *Angewandte Chemie (International Ed.), 58*(4), 968-994. https://doi.org/10.1002/anie.201804736

Walsh, M., Bourcier, T., Takahashi, K., Shi, L., Busche, M., Rother, R., Solomon, S., Ezekowitz, R., & Stahl, G. (2005). Mannose-binding lectin is a regulator of inflammation that accompanies myocardial ischemia and reperfusion injury. *The Journal of Immunology (1950), 175*(1), 541-546. https://doi.org/10.4049/jimmunol.175.1.541

Wang, A., Tian, X., Zuo, Y., Xu, Q., Meng, X., Chen, P., Li, H., & Wang, Y. (2022). Urine ketone bodies and adverse outcomes in patients with acute ischemic stroke or TIA. *Atherosclerosis Plus, 48*, 20-26. https://doi.org/10.1016/j.athplu.2022.03.001

Wang, J., Motlagh, N. J., Wang, C., Wojtkiewicz, G. R., Schmidt, S., Chau, C., Narsimhan, R., Kuellenberg, E. G., Zhu, C., Linnoila, J., Yao, Z., & Chen, J. W. (2021). D-mannose suppresses oxidative response and blocks phagocytosis in experimental neuroinflammation. *Proceedings of the National Academy of Sciences - PNAS, 118*(44), 1. https://doi.org/10.1073/pnas.2107663118

Wang, T., Suzuki, K., Kakisaka, K., Onodera, M., Sawara, K., & Takikawa, Y. (2019). L-carnitine prevents ammonia-induced cytotoxicity and disturbances in intracellular amino acid levels in human astrocytes. *Journal of Gastroenterology and Hepatology, 34*(7), 1249-1255. https://doi.org/10.1111/jgh.14497

Wang, Z., Wang, Z., Sun, Z., Sun, Z., Zhang, L., & Zhang, L. (2014). The relationship between serum mannose-binding lectin levels and acute ischemic stroke risk. *Neurochemical Research, 39*(2), 248-253. https://doi.org/10.1007/s11064-013-1214-x

Wanichthanarak, K., Jeamsripong, S., Pornputtapong, N., & Khoomrung, S. (2019). Accounting for biological variation with linear mixed-effects modelling improves the quality of clinical metabolomics data. *Computational and Structural Biotechnology Journal, 17*, 611-618. https://doi.org/10.1016/j.csbj.2019.04.009

Wanner, Z. R., Southam, C. G., Sanghavi, P., Boora, N. S., Paxman, E. J., Dukelow, S. P., Benson, B. W., Montina, T., Metz, G. A. S., & Debert, C. T. (2021). Alterations in urine metabolomics following sport-related concussion: A 1H NMR-based analysis. *Frontiers in Neurology, 12*, 645829-645829. https://doi.org/10.3389/fneur.2021.645829

White, H., & Venkatesh, B. (2011). Clinical review: Ketones and brain injury. *Critical Care (London, England), 15*(2), 219-219. https://doi.org/10.1186/cc10020

Wiklund, S., Johansson, E., Sjöström, L., Mellerowicz, E. J., Edlund, U., Shockcor, J. P., Gottfries, J., Moritz, T., & Trygg, J. (2008). Visualization of GC/TOF-MS-based metabolomics data for identification of biochemically interesting compounds using OPLS class models. *Analytical Chemistry (Washington), 80*(1), 115-122. https://doi.org/10.1021/ac0713510

Wishart, D. S. (2008). Quantitative metabolomics using NMR. *TrAC, Trends in Analytical Chemistry (Regular Ed.), 27*(3), 228-237. https://doi.org/10.1016/j.trac.2007.12.001

Wishart, D. S., Feunang, Y. D., Marcu, A., Guo, A. C., Liang, K., Vázquez-Fresno, R., Sajed, T., Johnson, D., Li, C., Karu, N., Sayeeda, Z., Lo, E., Assempour, N., Berjanskii, M., Singhal, S., Arndt, D., Liang, Y., Badran, H., Grant, J., . . . Scalbert, A. (2018). HMDB 4.0: The human metabolome database for 2018. *Nucleic Acids Research, 46*(D1), D608-D617. https://doi.org/10.1093/nar/gkx1089

West, J. B. (Ed.). (1985). *Best and Taylor's physiological basis of medical practice* (11th ed., pp 334–336). Baltimore MD, USA: Waverly Press, Inc.

Xia, J., Broadhurst, D. I., Wilson, M., & Wishart, D. S. (2013). Translational biomarker discovery in clinical metabolomics: An introductory tutorial. *Metabolomics, 9*(2), 280-299. https://doi.org/10.1007/s11306-012-0482-9

Xia, J., & Wishart, D. S. (2010). MetPA: A web-based metabolomics tool for pathway analysis and visualization. *Bioinformatics, 26*(18), 2342-2344. https://doi.org/10.1093/bioinformatics/btq418

Xiao, J., Zhang, J., Sun, D., Wang, L., Yu, L., Wu, H., Wang, D., & Qiu, X. (2016). Discriminating poststroke depression from stroke by nuclear magnetic resonance spectroscopy-based metabonomic analysis. *Neuropsychiatric Disease and Treatment, 12*, 1919-1925. https://doi.org/10.2147/NDT.S110613

Xu, H., Wang, X., Geng, G., Xu, X., Liu, L., Zhang, Y., Wang, Z., Wang, L., & Li, Y. (2023). Association of circulating branched-chain amino acids with cardiovascular diseases: A mendelian randomization study. *Nutrients, 15*(7), 1580. https://doi.org/10.3390/nu15071580

Xu, K., LaManna, J. C., & Puchowicz, M. A. (2012). Neuroprotective properties of ketone bodies. *Advances in Experimental Medicine and Biology, 737*, 97-102. https://doi.org/10.1007/978-1-4614-1566-4_15

Xu, X., Zhang, P., Shen, Y., Li, H., Wang, Y., Lu, G., & Zhou, J. (2015). Mannose prevents acute lung injury through mannose receptor pathway and contributes to regulate PPARγ and TGF-beta 1 level. *International Journal of Clinical and Experimental Pathology, 8*(6), 6214-6224.

You, S., Xu, J., Ou, Z., Zhong, C., Han, Q., Chen, J., Cao, Y., & Liu, C. (2021). Prognostic significance of urinary protein and urinary ketone bodies in acute ischemic stroke. *Nutrition, Metabolism, and Cardiovascular Diseases, 31*(11), 3152-3160. https://doi.org/10.1016/j.numecd.2021.07.010

Yun, Y., Liang, F., Deng, B., Lai, G., Vicente Gonçalves, C. M., Lu, H., Yan, J., Huang, X., Yi, L., & Liang, Y. (2015). Informative metabolites identification by variable importance analysis based on random variable combination. *Metabolomics, 11*(6), 1539-1551. https://doi.org/10.1007/s11306-015-0803-x

Zeng, N., Zhang, N., Ma, X., Wang, Y., Zhang, Y., Wang, D., Pu, F., & Li, B. (2022). Transcriptomics integrated with metabolomics: Assessing the central metabolism of different cells after cell differentiation in *Aureobasidium pullulans* NG. *J Fungi (Basel). 8*(8), 882. doi: 10.3390/jof8080882.

Zhang, W., Cheng, H., Gui, Y., Zhan, Q., Li, S., Qiao, W., & Tong, A. (2021). Mannose treatment: A promising novel strategy to suppress inflammation. *Frontiers in Immunology, 12*, 756920-756920. https://doi.org/10.3389/fimmu.2021.756920

Zhang, X., Li, Y., Liang, Y., Sun, P., Wu, X., Song, J., Sun, X., Hong, M., Gao, P., & Deng, D. (2017). distinguishing intracerebral hemorrhage from acute cerebral infarction through

metabolomics. *Revista De Investigacion Clinica, 69*(6), 319-328. https://doi.org/10.24875/RIC.17002348

Zhang, Z., Zhang, Z., Wang, C., Wang, C., Wang, J., Wang, J., Zhang, Z., Zhang, Z., Yang, Y., Yang, Y., Gao, L., Gao, L., Zhang, X., Zhang, X., Chang, T., Chang, T., Gao, G., Gao, G., Li, L., & Li, L. (2015). Prognostic value of mannose-binding lectin: 90-day outcome in patients with acute ischemic stroke. *Molecular Neurobiology, 51*(1), 230-239. https://doi.org/10.1007/s12035-014-8682-0

Zhenyukh, O., Civantos, E., Ruiz-Ortega, M., Sánchez, M. S., Vázquez, C., Peiró, C., Egido, J., & Mas, S. (2017). High concentration of branched-chain amino acids promotes oxidative stress, inflammation and migration of human peripheral blood mononuclear cells via mTORC1 activation. *Free Radical Biology & Medicine, 104*, 165-177. https://doi.org/10.1016/j.freeradbiomed.2017.01.009

CHAPTER 5: CONCLUSIONS AND FUTURE DIRECTIONS

5.1 BRIDGING HYPOTHESES AND RESULTS

The present Master's thesis sought to investigate the clinical application of ^{1}H NMR-based urinary and serum metabolomics for patients with stroke. The main hypothesis posited that the metabolic profiles of stroke patients in both the acute and chronic phases could be effectively distinguished, offering significant practical utility. Several metabolites were expected to emerge with clinical significance as biosignatures for diagnostication, prognostication, and therapeutic efficacy. Additionally, the research aimed to uncover relevant biochemical pathways involved in the neuropathogenesis of stroke.

The studies conducted in Chapter 3 and Chapter 4 of this thesis addressed the aforementioned theories, by focusing on urine and serum, respectively. In both biofluids, the initial metabolomic profiles were distinguishable from the profiles at 6 months post-stroke. Of note, the VIAVC best subset exhibited robust performance in discerning significance between the two groups, achieving an impressive predictive accuracy of 85% in both urine and blood samples. These findings highlight the exceptional translational ability of the work presented in this thesis. Interestingly, apart from alanine, which presented opposing regulations, there were no significant overlapping metabolites between urine and blood. The finding of differential alanine regulation may be evidence of the complementary nature of urine and blood, wherein an increase in the urinary metabolite excretion leads to a corresponding decrease in the blood-based metabolite.

The underlying biochemical changes associated with stroke were assessed using pathway analysis. Although numerous pathways emerged as significant between urine and serum, only phenylalanine metabolism demonstrated significant alterations in both. However, nearly all identified pathways are involved in inflammation and oxidative stress, either in neurotoxic or

neuroprotective capacities. For example, the neuroprotective effect of the synthesis and degradation of ketone bodies was demonstrated through its many remarkable capabilities, such as preventing ROS formation and suppressing inflammatory factors (Maalouf et al., 2009). Thus, the elaborate network of pathways involved in stroke-related injury processes is outlined throughout this thesis.

To ascertain the clinical applicability of the significant metabolites as biomarkers, their correlations with stroke severity scores were examined. Most notably, pseudouridine emerged as a prognostic and treatment efficacy marker of considerable significance in urine. In blood, alloisoleucine was a prominent marker, displaying significance in diagnosis, prognosis, and gauging treatment success. Furthermore, 3-methyl-2-oxovaleric acid and 2-hydroxybutyric acid were identified as potential prognostic signatures, while dimethyl sulfone and mannose showed promise as signatures for rehabilitative monitoring. Hence, pseudourinine in urine and alloisoleucine in blood have the greatest potential to be effective overall biomarkers; however, further investigation is warranted to evaluate the ability of dimethyl sulfone and mannose in blood for monitoring recovery. The results presented position metabolic biomarkers as compelling candidates for the diagnostication of pathological states, prognostication of future disease outcomes, and evaluation of treatment effectiveness.

In conclusion, the hypotheses in question have been successfully validated in a pilot study population, demonstrating their empirical substantiation. The composition of metabolites in urine and blood exhibited notable variations between the two sample collection time-points, indicating distinct metabolomic profiles at the onset of stroke and 6 months post-stroke. The observed changes in the metabolites are strongly correlated with clinical measures of function, providing valuable insights into the impact of stroke injury and subsequent recovery processes. Furthermore, the identification of significantly altered pathways shed light on the intricate interplay within the

biological system during the process of stroke repair and recovery. This comprehensive analysis not only elucidated the biomolecules and cellular processes involved in stroke pathophysiology, but also validated the potential translational ability of this analysis in the clinical context of stroke. Importantly, the results presented in this thesis demand validation through independent datasets or experiments possessing greater statistical power. Nonetheless, the insights generated by this research serves as a formidable foundation to guide future investigations.

5.2 ADDRESSING LIMITATIONS AND FUTURE DIRECTIONS

5.1.1 Challenges and Charting Future Trajectories

Given the preliminary nature of the research elucidated in this thesis, it is imperative to acknowledge several limitations that need to be addressed. This section aims to delineate the challenges associated with the current investigation and the field encompassing it, while at the same time offering potential solutions. Moreover, an assessment of this research's contributions towards resolving these challenges will be exhausted.

The heterogeneity of stroke, including its neurochemical insults and impairments, poses a constant challenge within the already heterogeneous field of metabolomics, where interindividual variations in metabolites are vastly different. While this study accounted for age and sex differences, several other stroke-related factors need to be considered, such as stroke type, vascular territory, stroke severity, and affected side. Regarding stroke type, there is extensive research in the field being conducted on ischemic stroke, but the investigation of biochemical alterations associated with hemorrhagic stroke remains relatively understudied. Therefore, further exploration of hemorrhagic stroke is warranted to discover the similarities and differences between stroke types. Further, distinguishing ischemic stroke and intracerebral hemorrhage remains a challenge, as there is no single biomarker or panel of biomarkers that can differentiate the two (Kamtchum-

Tatuene & Jickling, 2019). The biomarkers in this study are no exception as our biomarkers were tested on patients with both ischemic and hemorrhagic strokes. It would be advantageous to conduct further research to see if the biomarkers presented in this thesis are markers for stroke in general or ischemic stroke, given the predominant presence of ischemic stroke cases in the participant pool. Moreover, this study focused on patients with mild to moderate stroke severity based on NIHSS assessments. Future studies may examine whether there are different biomarkers for mild, moderate, and severe stroke. Our results could represent general markers for mild and moderate stroke, but further research is imperative to clarify the proposed signatures' relevance in severe stroke. In the realm of metabolomics, diagnostic stroke biosignatures have received substantial attention, whereas prognostic markers remain understudied and little to none treatment efficacy markers have been considered. Through this research, we aspire to ignite future investigations into treatment response following stroke to facilitate the implementation of personalized rehabilitation strategies by clinicians.

Drawing upon our experience in the field and the preliminary findings of our pilot studies, we offer several recommendations for future research endeavors to address inherent limitations. First, we advocate for clinical studies recruit a substantial and diverse sample group, with a minimum retention of 50 participants after accounting for potential dropouts. Ideally, the sample size should be made up of hundreds of individuals, maintaining a balanced representation of males and females from diverse ethnic backgrounds (Dunn et al., 2012; Roth & Powers, 2022). Moreover, future research should consider various confounding factors, including but not limited to differences in age, sex, demographic diversity, diet, exercise patterns, mobility, body mass index, stroke type, affected side, vascular territory, injury level, acute versus chronic drug treatment, and medical history. In our own research, we mitigated the influence of these covariates by pairwise analysis of the samples. However, it is important to note that these factors may change

following stoke and influence the patient throughout and following treatment. Nevertheless, to enhance study robustness, the inclusion of a baseline group or a musculoskeletal injury control group is recommended to compare the biochemical changes observed in normative and injury-related contexts. Although it is extremely challenging to collect samples in a clinical setting, it is advised to minimize the temporal range of samples collected post-injury. In the present study, initial sample collection occurred in a range of 2-32 days and 2-11 days and final sample collection occurred within 98-242 days and 101-242 in urine and blood, respectively. To capture the distinct biochemical changes that occur during the hyperacute (hours), acute (days), subacute (months), and chronic (6 months) phases of injury, stringent control of sample collection times is crucial.

The translatability of the selected biomarkers in this study to a clinical stroke setting holds paramount importance in minimizing brain damage and disability. It is imperative to identify a panel of biomarkers that can accommodate individual variations and exhibits robustness in marking stroke onset, outcomes, and recovery. The findings presented herein make significant contributions towards the development of personalized medicine for stroke patients, by discovering robust biomarkers with potential applicability across stroke types, different age groups, and sexes. However, before considering the clinical implementation of these biomarkers, it is essential to conduct comprehensive studies for clinical validation and research assay optimization (Laborde et al., 2012). Thus, after further confirmation of the biosignatures found in this study, they need to be tested preclinically to determine their predictive value in treatment response and the potential of selecting optimal treatments programs in their presence. Furthermore, in-depth research should be undertaken on each biomarker identified in the present investigation to better understand their contributions to the pathogenic and biochemical changes specific to stroke.

The translational ability of the identified biomarkers also relies on their ability to be applied as cost-effective standardized commercial assays for rapid urinalysis or blood testing. Although NMR, similar to CT and MRI, is expensive and not widely available, its efficacy in biomarker discovery is noteworthy due to its ability to detect a wide range of metabolites in biological samples that can be easily translated into the clinical setting. Additionally, unlike MRI and CT, NMR can offer insights into prognosis and treatment efficacy. While other metabolomics techniques also hold utility, NMR stands out by identifying the largest number of metabolites, employing non-destructive sample analysis, and streamlining sample preparation procedures. It is important to note that no single method can capture the entire range of the metabolome. Hence, it is advantageous for studies to use multiple platforms, such as a combination of NMR, GC-MS, and LC-MS, to maximize the total number of metabolites identified, as demonstrated by previous research indicating detection of 295 metabolites in urine (Bouatra et al., 2013) and 168 in serum (Psychogios et al., 2011) using all three techniques. Going a step further, it is recommended that investigations delve into more biological alterations occurring in the body by using multiple 'omics' techniques, including genomics, epigenomics, transcriptomics, proteomics, and metabolomics. Lastly, while urine and blood were selected in this study due to their minimally invasive nature and minimal sample preparation requirements, exploring other biological samples such as cerebrospinal fluid, feces, saliva, hair, and tissue may provide additional insights into systemic changes and potentially uncover novel biomarkers with clinical applications. By incorporating the suggested techniques, a robust all-encompassing study can be conducted with enhanced translational capacity.

Given the results of the present research displaying alterations in ketone bodies, our research group aspires to conduct further research on the ketogenic diet (KD) and its impact on stroke and other CNS injuries. New research is beginning to emerge associating stroke with the

neuroprotective role of ketones and the KD (Arora & Mehta, 2020; Maalouf et al., 2009; Yang et al., 2015). The results and discussion in this thesis corroborate its involvement in stroke and emphasizes the need for further research. Importantly, although large majority of research associates KD with beneficial effects following stroke, several studies have noted ketones can display neurotoxic effects (Kanikarla-Marie & Jain, 2015; Karavanaki et al., 2012). Therefore, it is vital to determine when and at what amounts ketones are neuroprotective compared to neurotoxic, as this may impact the results when treating patients with the KD.

The overarching UCAN study that encompasses this research follows stroke, traumatic brain injury, and spinal cord injury patients throughout recovery (Bykowski et al., 2021a; Bykowski et al., 2021b; Bykowski et al., 2023). Along with the metabolomics data of the current investigation, resting state fMRI and robotics assessments were performed. It would be beneficial to compare the metabolomics findings with those obtained from MRI and robotics to explore potential relationships between these modalities. Furthermore, the investigation of differences across CNS injury types warrants further investigation. By comparing all three injury types, it is possible to examine commonalities across CNS injuries as a whole, which can be contrasted with cerebral injuries.

5.1.2 Novel Statistical Frameworks for Closing Literature Gaps

Chapter 2 delineates the intricate statistical steps taken following sample preparation and NMR data acquisition. Initially, a series of data pre-processing steps are undertaken, including data reduction (binning), normalization, log transformation, pareto scaling, and a linear regression model. Subsequently, a comprehensive array of univariate and multivariate statistical analyses are conducted, encompassing VIAVC, PCA, PLS-DA, OPLS-DA, ROC, and AUC. Finally, the obtained results undergo clinical interpretation, involving the identification of significant metabolites, analysis of metabolic pathways, and establishment of clinical correlations.

Most notably, the utilization of a linear regression model as a pre-processing step to mitigate confounding factors out of the metabolomics data successfully removed age and sex variation from the data. It is worth noting that LMMs have previously been used in metabolomics to adjust for covariates. However, this particular technique is not ideal for all types of metabolomics investigations, particularly longitudinal studies with small sample sizes. The majority of metabolomics studies have disregarded the effects of confounding factors despite the substantial variation observed between individual samples, attributable to characteristics such as age, sex, genetic variations, and underlying health conditions. Moreover, even subtle lifestyle and environmental changes can instigate alterations in the metabolome. The linear regression model developed for this thesis addresses the need for a data pre-processing technique that removes the variation in the data. The robust results presented in this thesis, coupled with the extensive supporting literature, unequivocally confirm the vigor of this approach, further strengthening the statistical analysis conducted.

5.1.3 Concluding Remarks

In conclusion, this thesis provided biochemical fingerprints and insight into cellular functions after neuropathophysiological states such as stroke. The challenges and future trajectories of stroke research in the metabolomics field has been highlighted and our linear regression model is proposed as a data pre-processing step to alleviate some concerns related to confounding factors in the variance observed in the data. The identified biomarkers hold promise for personalized medicine in stroke patients, but extensive validation and research are required. This thesis advocates for inexpensive standardized commercial assays, with NMR as a noteworthy technique for biomarker discovery. Overall, this research contributes to the understanding of stroke and lays the foundation for future studies to advance personalized rehabilitation regimens and improve patient outcomes.

REFERENCES

Arora, N., & Mehta, T. R. (2020). Role of the ketogenic diet in acute neurological diseases. *Clinical Neurology and Neurosurgery, 192*, 105727-105727. https://doi.org/10.1016/j.clineuro.2020.105727

Bouatra, S., Aziat, F., Mandal, R., Guo, A.C., Wilson, M.R., Knox, C.,…Wishart, D.S. (2013). The Human Urine Metabolome. *PLoS ONE, 8*(9), 1-28. Doi: 10.1371/journal.pone.0073076

Bykowski, E. A., Petersson, J. N., Dukelow, S., Ho, C., Debert, C. T., Montina, T., & Metz, G. A. S. (2021a). Urinary biomarkers indicative of recovery from spinal cord injury: A pilot study. *IBRO Neuroscience Reports, 10*(1), 178-185. https://doi.org/10.1016/j.ibneur.2021.02.007

Bykowski, E. A., Petersson, J. N., Dukelow, S., Ho, C., Debert, C. T., Montina, T., & Metz, G. A. S. (2021b). Urinary metabolomic signatures as indicators of injury severity following traumatic brain injury: A pilot study. *IBRO Neuroscience Reports, 11*, 200-206. https://doi.org/10.1016/j.ibneur.2021.10.003

Bykowski, E. A., Petersson, J. N., Dukelow, S., Ho, C., Debert, C. T., Montina, T., & Metz, G. A. S. (2023). Identification of serum metabolites as prognostic biomarkers following spinal cord injury: A pilot study. *Metabolites, 13*(5), 605. https://doi.org/10.3390/metabo13050605

Dunn, W. B., Wilson, I. D., Nicholls, A. W., & Broadhurst, D. (2012). The importance of experimental design and QC samples in large-scale and MS-driven untargeted metabolomic studies of humans. *Bioanalysis, 4*(18), 2249-2264. https://doi.org/10.4155/BIO.12.204

Kamtchum-Tatuene, J., & Jickling, G. C. (2019). Blood biomarkers for stroke diagnosis and management. *Neuromolecular Medicine, 21*(4), 344-368. https://doi.org/10.1007/s12017-019-08530-0

Kanikarla-Marie, P., & Jain, S. K. (2015). Hyperketonemia (acetoacetate) upregulates NADPH oxidase 4 and elevates oxidative stress, ICAM-1, and monocyte adhesivity in endothelial cells. *Cellular Physiology and Biochemistry, 35*(1), 364-373. https://doi.org/10.1159/000369702

Karavanaki, K., Kakleas, K., Georga, S., Bartzeliotou, A., Mavropoulos, G., Tsouvalas, M., Vogiatzi, A., Papassotiriou, I., & Karayianni, C. (2012). Plasma high sensitivity C-reactive protein and its relationship with cytokine levels in children with newly diagnosed type 1 diabetes and ketoacidosis. *Clinical Biochemistry, 45*(16-17), 1383-1388. https://doi.org/10.1016/j.clinbiochem.2012.05.003

Laborde, C. M., Mourino-Alvarez, L., Akerstrom, F., Padial, L. R., Vivanco, F., Gil-Dones, F., & Barderas, M. G. (2012). Potential blood biomarkers for stroke. *Expert Review of Proteomics, 9*(4), 437-449. doi:10.1586/EPR.12.33

Maalouf, M., Rho, J. M., & Mattson, M. P. (2009). The neuroprotective properties of calorie restriction, the ketogenic diet, and ketone bodies. *Brain Research Reviews, 59*(2), 293-315. https://doi.org/10.1016/j.brainresrev.2008.09.002

Psychogios, N., Hau, D. D., Peng, J., Guo, A. C., Mandal, R., Bouatra, S., Sinelnikov, I., Krishnamurthy, R., Eisner, R., Gautam, B., Young, N., Xia, J., Knox, C., Dong, E., Huang, P., Hollander, Z., Pedersen, T. L., Smith, S. R., Bamforth, F., . . . Wishart, D. S. (2011). The human serum metabolome. *PloS One, 6*(2), e16957-e16957. https://doi.org/10.1371/journal.pone.0016957

Roth, H. E., & Powers, R. (2022). Meta-analysis reveals both the promises and the challenges of clinical metabolomics. *Cancers, 14*(16), 3992. https://doi.org/10.3390/cancers14163992

Yang, Q., Guo, M., Wang, X., Zhao, Y., Ding, H., Cui, M., & Dong, Q. (2017). Ischemic preconditioning with ketogenic diet improves brain ischemic tolerance through increased extracellular adenosine levels and hypoxia-inducible factors. *Brain Research, 1667*, 11-18. https://doi.org/10.1016/j.brainres.2017.04.010