APPLICATION OF SYSTEMS BIOLOGY ANALYSIS TO HEPATIC INJURY FOLLOWING HEMORRHAGIC SHOCK

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

CHARLES E. DETERMAN JR.

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DR. GREGORY J. BEILMAN, ADVISOR

MAY 2014

© Charles E. Determan Jr. 2014

Acknowledgements

Firstly, I would like to offer my sincerest gratitude to Dr. Greg Beilman for his guidance and strong support in all of my endeavors. I have been deeply fortunate to have an advisor that encouraged me to pursue my individual interests and find ways to apply them in new and exciting ways. It is thanks to his unwavering support that my desire to continue expanding my scientific horizons beyond individual topics continues to this day. I also want to express my thanks to the entire lab staff that I have had the pleasure of working alongside. This body of work only became possible as a result of their extensive work to develop the hemorrhagic shock model and begin metabolomics investigations. I wish to thank Dan Lexcen for his prior contributions upon which this work is built upon as well as his insight into the doctoral process. I also owe a great thanks to Elizabeth Lusczek for our wonderful discussions on systems biology and extensively on various analytical techniques to continue exploring this grand area of study. I would like to also thank Nancy Witowski for our numerous discussions which greatly expanded my knowledge of physiology and biology as a whole proving invaluable in the interpretation of my datasets. In addition, all of the studies herein were not completed in a single lab and I would therefore like to express my thanks to the University of Minnesota's Nuclear Magnetic Resonance Facility, the BioMedical Genomics Center, and the Minnesota Supercomputing Institute for their invaluable assistance in providing the resources for acquiring and analyzing this data.

I would also like to thank my committee collectively: Greg Beilman, Lester Drewes, Marshall Hampton, and Gyorgy Simon. Each of you has assisted me from specific analytical questions to general guidance and I am deeply grateful for your

thoughtful consideration of this body of work. I look forward to continued professional relationships that may continue to benefit us all.

To my family, I thank you so much for your support in my educational pursuits. Although I am thankful to all members of my family for numerous reasons I wish to thank my parents specifically. To my mother, thank you for teaching me to always be myself, value the simple things in life, and pursue all that makes one happy. To my father, thank you for all of our conversations over the years. You have taught me the beauty of pursuing knowledge and I thoroughly enjoy exploring the diverse topics of life that are beyond my own expertise with you.

Lastly, I wish to extend my deepest gratitude and love to my wife Brittany. I have been so fortunate to have such a supportive partner in my life. Through all the good times and the few bad, you have continuously supported me in pursuit of my doctorate. I now eagerly look forward towards the next chapter of our life. All that I have is made infinitely more special because I am able to share it with you.

Abstract

Introduction: This dissertation is focused on the metabolomic and transcriptomic changes that occur as a result of carbohydrate prefeeding during hemorrhagic shock and trauma within the liver of a porcine model. The risk of trauma and hemorrhagic shock continues to be an important issue in both military and civilian sectors. As such, we explored the impact of a prior fed state upon the overall response to hemorrhagic shock and resuscitation. The primary hypotheses were that changes in metabolism at the metabolomic and transcriptomic levels would be dependent upon the fed state. In addition, this thesis explores a more comprehensive analysis of metabolomics datasets to standardize analysis and improve overall consistency.

Materials and Methods: Algorithm comparison was accomplished using six commonly applied methods to three synthetic datasets, of different sample sizes, and three openly accessible published datasets. This comparison also incorporated metrics to measure consistency of identified features (i.e. stability) to provide further confidence in results. Metabolomics analysis was accomplished with nuclear magnetic resonance spectroscopy (NMR) and Chenomx software to profile and quantify metabolites in liver extracts. The metabolome was subsequently analyzed with partial least squares discriminant analysis (PLS-DA). Transcriptomics analysis was conducted using next-generation sequencing (NGS) technology to employ RNA-sequencing (RNA-seq) on mRNA extracts from liver biopsies. The RNA-seq data was analyzed using typical processing techniques to generate a count matrix and subsequently analyzed with the Bioconductor package EdgeR.

Results: The comparison of algorithms showed that the best algorithm is associated with differently structured datasets (e.g. number of features, number of groups, sample size, etc.). Analysis of the liver metabolome revealed changes in carbon energy sources, amino acid metabolism, oxidative stress, and membrane maintenance. Transcriptomic analysis revealed changes in carbohydrate metabolism, cytokine inflammation, cholesterol synthesis and apoptosis. In addition, there is evidence of increased cytoskeleton reorganization which may correspond to a shrunken, catabolic state which provides and anti-inflammatory condition to mitigate cellular damage.

Conclusion: The response to hemorrhagic shock and resuscitation is altered with respect to a fasted or carbohydrate prefed state. Metabolomics and transcriptomic analyses suggest altered metabolic pathways as a result of fed state. Altered carbohydrate metabolism was readily identified thereby confirming both methods were successful. Additionally, indications of membrane maintenance that follow cytoskeletal remodeling and cellular shrinkage are potentially reflected by 3-Hydroxyisovalerate and sn-Glycero-3-phosphocholine. These results provide further evidence for pre-conditioning (e.g. altered diet) and hypertonic resuscitation methods to possibly improve patient outcome. Further research is required in alternative prefeeding substrates (e.g. protein, lipid, etc.) as well as improving the integration of different systems level datasets to understand more thoroughly the systemic effects of hemorrhagic shock and resuscitation.

Table of Contents

List of Tables

Appendix C

List of Figures

I. INTRODUCTION

Hemorrhagic shock and traumatic injury information

Worldwide, trauma is a leading cause of death for those between 5 and 45 years irrespective of demographics[1,2]. Of these deaths, up to 50% are due to hemorrhagic shock. Patients injured during active military duty are also commonly admitted with hemorrhagic injuries[3]. Furthermore, survival of the initial insult can lead to further complications including multiple organ dysfunction syndrome (MODS), which remains a predominant cause of late deaths[4]. Although treatments have improved and trauma care has continued to keep pace with an aging population, the overall survival of trauma patients remains constant[5].

Briefly, hemorrhagic shock is a life-threatening condition resulting from inadequate tissue perfusion, followed by activation of the inflammatory cascade and changes in metabolic processes. The reduced blood flow to peripheral tissues results in hypoxia and subsequent reperfusion injury[6]. This decreased volume and available oxygen leads to an extensive metabolic cascade. In brief, the lack of oxygen slows/stops oxidative phosphorylation via the electron transport chain and shifts the cells a glycolytic state[7]. Without oxygen as the final electron acceptor, the mitochondria balance membrane potential via reversal of ATP synthase thereby using ATP to translocate protons across the inner membrane[8]. This decrease in ATP leads to conversion of adenine nucleotides to hypoxanthine and xanthine, substrates for xanthine oxidase which serves as another ROS generator[9,10]. Inflammatory pathways are also activated which are promoted by inducible nitric oxide synthase (iNOS). Furthermore, the large amounts of

oxygen returned upon resuscitation result in further ROS via enzymes such as NADPH oxidase[11] and tissue reperfusion damage.

Metabolic component of hemorrhagic shock

There are many systemic changes in response to hemorrhagic shock; however, this dissertation focuses on the metabolic response of the liver. The inadequate tissue perfusion following hemorrhagic shock leads to decreased oxygen availability to the mitochondria resulting in a switch towards anaerobic metabolism. The liver serves an important function as a regulator of metabolism during stressed states. Initially, the shift towards anaerobic metabolism stimulates the liver to increase glycogenolysis and process elevated lactate produced in the peripheral tissues. The liver also provides a major site of detoxification and production of alternate metabolic fuel sources including amino acids and lipids.

Of particular interest, the impact of fed state on the hepatic response to hemorrhagic shock is evaluated. Previously, the effect of different fed states on liver metabolic processes following injury and hemorrhagic shock were not well known. To elucidate these processes a systems level analysis via metabolomics and transcriptomics provide a broad exploration of potential responses.

Metabolomics

Metabolomics, briefly defined, is a study of all the low molecular weight metabolites that collectively make up the metabolome. This technique allows a comprehensive analysis of the changes within the metabolome and their respective correlations with a physiologic response[12]. One commonly used technique is via nuclear

magnetic resonance spectroscopy (NMR) and has been successfully applied towards studies of a diverse set of medical studies including pancreatitis, cancer, and diabetes[13,14].

Transcriptomics

Transcriptomics is the study of all the RNA that collectively makes up the transcriptome. The transcriptome provides a global analysis of the biochemical state of the organism. In contrast to metabolomics which provides a picture of the metabolic end products, transcriptomics presents what the metabolic response is able to do. The expressed RNA corresponds to proteins that may be actively translated to express a particular phenotype of biochemical pathway. In this manner, one can evaluate the 'state' of the organism at a given time or condition.

The application of metabolomics and transcriptomics to the study of hemorrhagic shock may provide novel insights into the complex metabolic responses that can direct future efforts to improve quality of care and reduced mortality.

Purpose of the dissertation

This dissertation focuses on the systems level analysis of the hepatic response to hemorrhagic shock and trauma. It is composed of four chapters of background material and three chapters of experimental results. The second chapter provides information regarding known physiologic and metabolic responses of hemorrhagic shock. As the primary focus of this dissertation is on the liver it is addressed in Chapter 3. Chapters 4 and 5 provide concise overviews of metabolomics and transcriptomics including common methods and analyses. Given the lack of an objective means to determine the most appropriate method to analyze metabolomics data, the first study presented in Chapter 6 consists of the development of a program to apply systematically multiple accepted

methods for the analysis of metabolomics data sets and evaluate performance and stability of algorithm performance. Chapter 7 presents the analysis of the liver metabolome in response to hemorrhagic shock and trauma. Although any 'omics' level analysis is powerful, each is inherently limited to its position in the biological cascade from genes to metabolites. To acquire a more holistic perspective a transcriptomic analysis is subsequently provided in Chapter 8. In conclusion, Chapter 9 provides an initial integration of these two levels of biological expression followed by suggested future directions.

CHAPTER 2

Hemorrhagic Shock

Introduction

Hemorrhagic shock often follows traumatic injury characterized by inadequate perfusion leading to hypovolemic and hypoxic conditions. This results in several complex metabolic and physiologic processes that take place systemically to help the body recover. This chapter will consist of a brief overview of these metabolic and physiologic changes.

In response to acute hypovolemia, multiple endocrine compounds are released including glucagon, cortisol and adrenocorticotropic hormone (ACTH). In order to maintain normal blood pressure angiotensin II and vasopressin are also released to elicit vasoconstriction and decrease fluid excretion. As a consequence of reduced blood flow and vasoconstriction, tissues begin to experience hypoxia and must convert to anaerobic energy states. Glucagon helps to provide the increased glucose demand of anaerobic glycolysis by inducing hyperglycemia via hepatic glycogenolysis. However, if the duration of shock is extended the glycogen stores become depleted resulting in a subsequent hypoglycemic state leading to mitochondrial dysfunction and overall decompensation.

In an effort to restore circulation, reperfusion methods are employed using crystalloids (often first choice), colloids, hypertonic saline, and blood transfusions. However, this effort to reestablish blood flow and oxygen supply enhances injury and aggravates damage at cellular level commonly referred to as ischemia-reperfusion injury. Injured cells remain in the hypoxic state and may be unable to restore fully functional mitochondria. This hypoxic condition stimulates the transcription factor nuclear factor kappa beta (NF-κB). Once released from the inhibitory protein IκB, NF-κB translocates

to the nucleus where it stimulates production of pro-inflammatory cytokines (e.g. TNF-α, IL-1β, and IL-6), chemokines and adhesion molecules[15]. Evidence suggests that the pro-inflammatory cytokines, reactive oxygen species, and deranged Ca^{2+} homeostasis elicits cell death. The mechanism of cell death however is also up for debate as indication of necrosis and apoptosis are present. In response, a combination theory was developed called "necrapoptosis" whereby there are common mechanisms that lead to both forms of cell death[16].

Ischemia-reperfusion has been extensively studied over the years with many improvements in treatment that are applicable to hemorrhagic shock. One particular discovery that has not yet found application is pre-conditioning. Pre-conditioning is the process whereby an organ is exposed to short periods of ischemia. The applications are more easily seen in liver surgery and transplants but not directly applicable for trauma induced hemorrhagic shock. However, this suggests that organs may be stimulated into a state that is more prepared to survive a hypoxic episode. One way this pre-conditioning may be administered is a different fed state as characterized by previous studies that suggest increased carbohydrates improves survival in rats[17]. However, recent evidence has shown that mouse models poorly mimic human trauma, burns and endotoxemia[18]. As such, the following investigations focus on a porcine model of trauma and hemorrhagic shock to reflect more accurately the human response.

Animal Preparation

All of our experiments have the same preparation and administration of the shock protocol (Appendix A). The experimental protocol was approved by the University of Minnesota Animal Use Committee and was conducted in accordance with established

guidelines for the treatment of laboratory animals. Male Yorkshire-crossbreed pigs (Manthei Hog Farm, LLC, Elk River, MN) weighing 15-20 kg were fasted overnight prior to surgery but are allowed water *ad libitum*. Two experimental groups were utilized: Carbohydrate Prefed (CPF, n = 32) and Fasted (FS, n = 32). CPF animals were given 7cc/kg bolus of Karo Syrup® (mixture of sugars including ~ 15% glucose, maltose, fructose and sucrose) diluted with water 1 hour prior to induction of anesthesia.

The animals were then anesthetized with an intramuscular dose of telazol (Wyeth Animal Health, Madison, NJ). Anesthesia was maintained by an i.v. infusion of propofol (2-9 mg/kg, AstraZeneca Pharmaceuticals, Wilmington, England) and 60% inhaled nitrous oxide throughout the experiment. Upon sedation, the pigs were orally intubated and ventilated (Servo 900, Siemens-Alema AB, Sweden) to maintain a PO2 of 70-120 torr and a PCO2 of 35-45 torr (SERVO Ventilator 900C, Siemens, Malvern, PA). Peripheral intravenous lines were placed in the surgically exposed right femoral artery and right jugular vein. A catheter was placed in the right femoral artery for continuous measurement of blood pressure and blood sampling. An introducer (7 French Avanti, Cordis Corporation, Miami Lakes, FL) was placed into the right jugular vein and a Swan-Ganz catheter (5 French, Edwards Lifesciences, Irvine, CA) was placed for measurements. To evaluate the status of the animal, we monitored pulmonary artery pressure, pulmonary wedge pressure, cardiac output, and mixed arterial blood sampling. The measurements are useful in gauging the extent of and severity of the hemorrhage. Animals underwent a midline laparotomy, splenectomy, and a Foley catheter was placed in the urinary bladder via stab cystostomy. Under stressful conditions, the spleen is capable of contracting, releasing stored blood, and changing the intravascular volume.

Because the spleen can sequester 22% of the circulating red cell volume and a similar amount of plasma[19], a splenectomy is imperative to minimize the effect that the spleen has on hemostasis. The inferior vena cava (IVC) was cannulated for the controlled hemorrhage. After surgical preparation the animals entered a stabilization period until plasma lactate levels reach a value of 2.0 mmol/L or less (Instrumentation Laboratory Co., Lexington, ME).

Polytrauma

As most cases of hemorrhagic shock are caused by trauma we chose to utilize a model of polytrauma. Upon stabilization from surgical preparation, the animals undergo a simulated pulmonary contusion from a captive bolt device to create a blunt percussive injury. Hemorrhagic shock was induced by removal of blood via IVC catheter into acid-citrate-dextrose (ACD) blood bags to obtain a systolic blood pressure (SBP) of 45 to 55mm Hg (typically ~40% of total blood volume). A liver crush injury was induced using a modified Holcomb clamp technique[20] in the liver parenchyma.

Resuscitation:

Fluid therapy is ideally administered as soon as possible to compensate for hypovolemia. However, certain situations do not facilitate immediate full resuscitation such as injuries suffered by soldiers. A limited resuscitation can be administered during transportation to additional medical assistance. To simulate such a situation, after 45 minutes of shock animals received lactated Ringer's fluid given as 20 cc/kg intravenous (IV) boluses to maintain a systolic blood pressure greater than 80 mmHg for one hour of limited resuscitation. Afterwards the animals underwent a full resuscitation protocol for the following 24 hours. This resuscitation included fluid, shed blood, and ventilator

support in a protocolized fashion. Animals were resuscitated to standard clinical endpoints of systolic blood pressure (90 mmHg), urine output (1 cm3/kg/h), and hemoglobin (≥6 g/dL) for 20 h. Throughout the shock and resuscitation periods, animals are maintained on the ventilator and received an adjusted-dose of propofol and nitrous oxide to maintain an appropriate level of sedation and comfort as evaluated by clinical examination and a bispectral index (BIS) monitor. After the resuscitation period, surviving animals were extubated and sent to recovery and subsequently euthanized.

Sample Collection

At several time points throughout the experiment, liver biopsies were taken from the periphery of the liver ranging in weight from 0.2 to 0.6 grams. Biopsies were flash frozen in liquid nitrogen and stored at -80°C until preparation for further analysis. Biopsies were taken at the following timepoints: baseline after the animal stabilized from instrumentation (B), 45 minutes after hemorrhage (S45), 2, 8, and 20 hours after full resuscitation (FR2, FR8, FR20). This full protocol is presented diagrammatically in Appendix A. In this manner, an overall response including prior to and throughout the insult would be measured.

CHAPTER 3

The Liver

Introduction

In Chapter 2, the physiologic and metabolic responses to hemorrhagic shock were discussed in a systemic context. The majority of this dissertation focuses on metabolites and genes expressed in the liver. As such, this chapter will provide a general overview of the response of liver to hemorrhagic shock and trauma. The liver is a complex organ with a wide range of functions that are critical for survival. To date the numerous and intricate functions of the liver cannot be replicated for any definite length of time in the event of liver failure. To my knowledge, there are no major differences between human and porcine liver beyond the number of lobes (humans = 4, pigs = 5). As such, this chapter and subsequent discussions assume the porcine liver displays the same functions and metabolic pathways as humans.

Both lipid synthesis and degradation processes take place in the liver. This includes cholesterol synthesis, lipogenesis and lipoprotein production. Bile salts are also produced via CYPs-mediated oxidation of cholesterol. These bile salts are an important component of bile to help emulsify lipids and fat-soluble vitamins (e.g. Vitamin K) for digestion. Most of the synthesized bile is stored in the gall bladder for later use.

Among the most well studied and important functions of the liver is glucose regulation. When glucose levels are high and the body does not require further energy production, glucose is typically converting into glycogen. The liver serves as the primary glucose reservoir for most peripheral tissues and likewise a primary location for glycogenolysis and gluconeogenesis during periods of starvation or physiologic stress. The regulation of these processes consists of multiple different stimuli. Glycogenic stimuli include insulin, glucocorticoids, parasympathetic impulses and gluconeogenic precursors (e.g. fructose, amino acids). Glycogenolysis is primarily regulated by

glucagon, noroadrenaline and ATP but also stimulated by adenosine, nucleotides and nitric oxide [21].

Detoxification of both endogenous molecules (e.g. ammonia, steroid hormones) and exogenous compounds (i.e. antibiotics) is perhaps one of the most important functions of the liver. The process of detoxification is typically described in two phases known as "Transformation/Modification" and "Conjugation"[22]; however a third phase is added known as "Transport". In essence, the enzymes of phase I, mostly of the cytochrome P450 (CYPs) family, begin detoxification of lipid soluble compounds by converting them into water soluble forms for phase II detoxification. These CYPs require oxygen in order to function resulting in ROS byproducts. These free radicals are mediated by antioxidants such as glutathione and vitamin C; however, excessive production can damage liver cells. This response may also be dependent upon diet as alternate diets in mice have shown an altered antioxidant response in the liver[23]. Phase II enzymes further increase solubility and decrease toxicity by conjugating a secondary molecule. Unlike phase I, phase II enzymes consist of several families including UDP-glucouronlytransferases (UGTs), Glutathione S-transferases (GSTs), Sulfotransferases (SULTs) and amino acid conjugating enzymes[24]. Phase III enzymes consist of transporters belonging to the ABC transporter family. Once the toxin conjugates have been formed these transporters move them into the bile for excretion[25].

Response to Hemorrhagic Shock and Resuscitation

Hemorrhagic shock induces the release of glucagon, catecholamines and glucocorticoids which rapidly increase liver glycogenolysis and gluconeogenesis causing hyperglycemia. This is typical of many other forms of trauma including sepsis and burns

where energy expenditure is increased resulting in a hypermetabolic state. The hyperglycemic state is followed by a rapid increase in insulin to absorb the available glucose in peripheral tissues as well as suppress hepatic gluconeogenesis and glycogenolysis. However, both peripheral tissues and the liver develop insulin resistance which contributes to extended hyperglycemia. Although the mechanisms for insulin resistance are not fully understood, evidence suggests altered insulin regulating pathways such as the PI3-kinase-Akt pathway[26]. This hyperglycemic state does abate during late stages of shock when hepatic glycogen is exhausted leading to a subsequent hypoglycemic state[27].

In addition to altered glucose regulation, evidence suggests a relationship between mitochondrial dysfunction is common following trauma and hemorrhagic shock[28] and correlated with irreversible pathologic damage[29–31]. Decreased hepatic mitochondrial function results in decreased ATP and subsequently increased degradation products. These degradation products require removal and may overwhelm the capacity of the hypoxic liver resulting in liver failure and toxin buildup. Treatments such as the addition of pyruvate and antioxidants have been suggested as methods to improve outcome[32–34]. In addition, previous studies have suggested the use of insulin and glucose may improve outcome in rats[35] however there are no studies that apply this induced carbohydrate state in larger mammals. This is explored in Chapters 7 and 8.

Several pathways exist to generate ROS which under normal conditions are mediated by endogenous mechanisms. However, ischemia-reperfusion injuries can result in ROS levels that exceeds the liver capacity by stimulation of Kupffer cells, the primary source of vascular reactive oxygen species during the reperfusion period[36]. Furthermore,

macrophages, which facilitate removal of apoptotic bodies, and pro-inflammatory cytokines (e.g. TNF-α, IL-1, interferon-γ) also generate and induce more ROS respectively[37,38]. As a result, antioxidant strategies have been pursued to address the oxygen stress of hepatocytes. These studies have included stimulation of increased glutathione, superoxide dismutase and heat shock proteins in ischemia-reperfusion models[39]. Strategies to remove the inflammatory oxidant stress via depletion of Kupffer cells has also been shown to effective[40].

Despite the advancements in understanding the molecular mechanisms of the liver following ischemia and reperfusion much is left to be discovered. Furthermore, most of the previous research has been applied to small animal models which may not accurately reflect human physiology. Of interest herein, there have been few investigations on the impact of diet on the response of the liver to hemorrhagic shock and trauma. To explore this question, two systems level approaches, metabolomics (Chapter 4) and transcriptomics (Chapter 5), will be employed to acquire a more holistic picture of potentially altered metabolic pathways.

CHAPTER 4

The Metabolomics Process

Introduction

The complex metabolic changes and the physiology of the liver were discussed in the previous chapter. In this chapter, it will be presented that metabolomics provides an ideal tool towards investigating the hepatic response to hemorrhagic shock and the impact of a carbohydrate fed state. This chapter will provide a concise overview of the field of metabolomics and how one can identify and quantify metabolites from collected tissues. This method will be applied to the porcine hemorrhagic shock and traumatic injury in Chapter 7.

In essence, the objective of any 'omic' science is the global, non-targeted quantification of the biomolecules comprising a particular level of organization from cell and tissue, to an entire organism. This effort initially began with identifying the genetic structure of organisms (genomics) and subsequently identifying the corresponding gene transcripts (transcriptomics – Chapter 5). The study of the translated protein products has also continued to be developed with potential for post-translational modifications. In the end, the molecular cascade leads to the regulation of small molecule metabolites. The high-throughput profiling of all of these metabolites within a single sample constitutes the study of metabolomics[12]. The quantitative analysis of a large array of metabolic constituents (metabolome) provides a holistic, or 'systems level', view of the biochemical state of an organism. This potentially allows for the exploration of altered metabolic pathways within a given physiologic state. This makes metabolomics an ideal method for exploring the complex changes elicited by hemorrhagic shock and trauma.

Metabolomic experiments utilize various analytical methods such as gas or liquid chromatography, mass spectrometry (MS), or nuclear magnetic resonance spectroscopy

(NMR). Each of these methods separately or in combination is capable of quantifying metabolites in biological samples. While each method possesses advantages and disadvantages, the ultimate product is an extremely large and complex dataset wherein the number of features (e.g. metabolites) often far outnumbers the samples. These types of datasets are commonly referred to as 'high-dimensional'. These datasets require the use of multivariate analysis techniques such as partial least squares discriminant analysis (PLSDA) and random forest (RF). Despite the availability of such methods, there is currently no recommendation on when to use which method. This problem will be addressed in Chapter 6.

Our metabolomics experiment utilizes 1H-NMR and was designed to evaluate if a carbohydrate fed state alters the hepatic response to hemorrhagic shock and trauma. This subsequently allows an insight into metabolic processes, as a consequence of a fed state, that are associated with hemorrhagic shock and its complications. This chapter discusses the overall metabolomic process including NMR, metabolite quantification, and statistical analysis.

NMR Metabolomics

Proton NMR spectroscopy is particularly well suited for untargeted metabolomic analysis as the large number of small molecules (100's to 1000's) within a biological sample can be quantified simultaneously[12]. NMR spectroscopy has be used successfully to analyze the metabolome of multiple biofluids and tissues including urine[13,41,42], serum[43,44], and tissue extracts[45,46]. Additionally, with advances in solid state NMR analysis, metabolomics is now being applied on intact tissues to study various medical problems including breast cancer[47], skeletal muscle disease[48,49] and pancreatic cancer[50,51].

While mass spectrometry (MS) is known to be more sensitive than NMR spectroscopy[52,53], NMR is an accepted metabolomics platform with some distinct advantages over MS. NMR can be more cost-effective and easier to automate than MS, making it preferable for high-throughput studies[53]. The required sample preparation for MS analysis can cause metabolite loss, while NMR can measure intact biological samples in both liquid and solid states with minimal sample preparation. Finally, NMR is non-selective and capable of simultaneously detecting a wide range of small molecules, whereas MS may require more than one sample preparation or platform to cover the range of small molecules in a biological sample[52,54,55].

Sample collection and preparation

The first step to quality NMR acquisition is proper sample collection. In our experiments, liver is the tissue of choice and the remainder of the chapter will be focused tissue metabolite extraction. Immediately after collection, the degradation process has begun. It is imperative to slow the tissue degradation process as much as possible to get an accurate picture of the metabolic state. Tissue biopsies are typically flash frozen in liquid nitrogen and stored at -80°C until future metabolite extraction.

There are also multiple metabolite extraction techniques each with advantages and disadvantages as well as target metabolite group[45]. The specific method chosen is dependent upon the goals of the experiment (e.g. profile as many metabolites as possible, extract only hydrophobic metabolites, etc.). Once a method has been decided, the precise and consistent handling of the tissue is important to yield the most consistent and clear spectra. Multiple factors may affect the quality of the NMR spectra including the pH, salinity, temperature, and concentration of metal ions in the sample. These changes

affect the sensitivity of NMR, reduce the quality of data, and increase the time required to identify and quantify the chemical signatures in the spectra.

NMR spectroscopy theory

NMR is based upon exploiting the inherent magnetic properties of nuclear particles. By placing the nuclei in a strong magnetic field the orientation is fixed along the same axis as the applied field. These nuclei are then pulsed with a radio frequency of the same frequency as the specific nuclei precession (i.e. rotation rate about the axis). This causes the magnetization vector of the nuclei to turn orthogonal to the primary axis. Because of the constant magnetic field the nuclei begin to transition back towards primary axis while continuing to precess. This results in the magnetic vector spiraling around the primary axis thereby inducing an electrical signal that is received by the radio frequency coils to be subsequently interpreted by a computer resulting in the observed spectra (Figure 1).

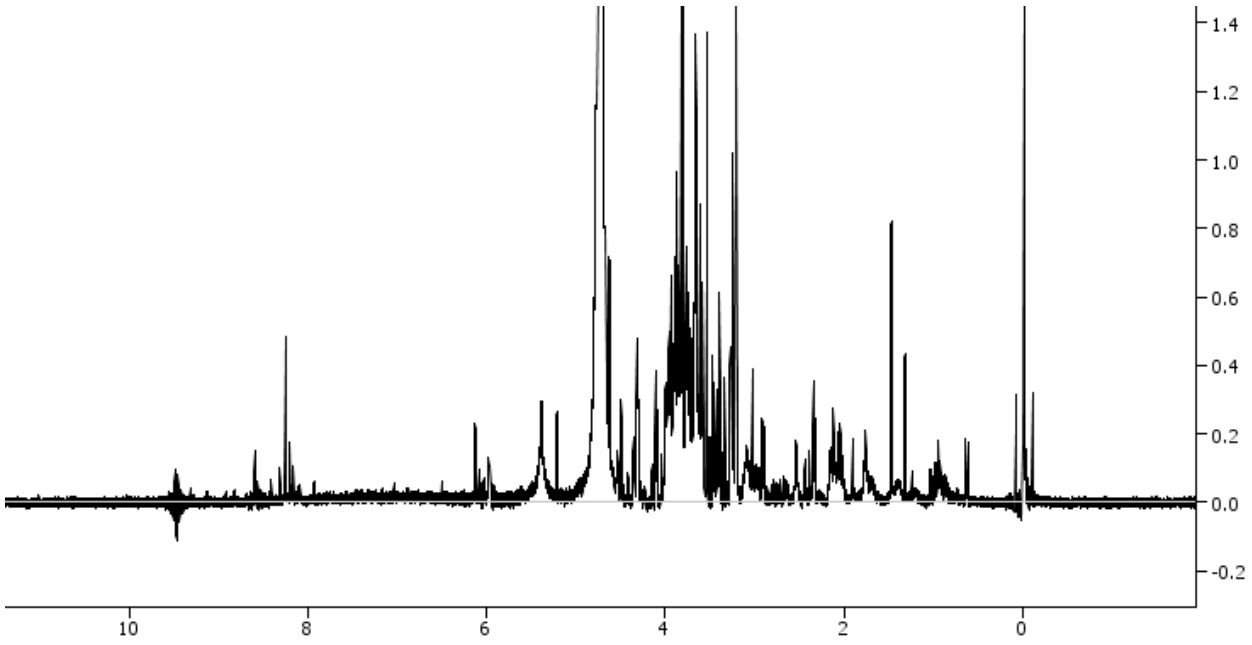

Figure 1 – Representative ^{1}H-NMR spectra

Each of the elemental nuclei possesses its own resonant frequency determined by the applied magnetic field of the NMR. This resonant frequency may be further modified by other surrounding nuclei which alter the magnetic field applied (i.e. shielding). This is what allows different compounds to possess different resonant frequencies. To interpret these frequencies, the chemical shift of a compound is calculated as a reference to a known internal standard. The chemical shift, in parts per million (ppm), is set at 0 for the internal standard. The observed chemical shift varies between compounds facilitating identification (i.e. compound fingerprinting). In this manner, many different molecules can be identified simultaneously without bias, provided the compounds all contain the specific nuclei being analyzed (e.g. ^{1}H, ^{13}C, etc.). This provides an advantage of mass spectrometry whereby different ionization properties of compounds can affect detection and quantification.

Quantification

Several techniques are utilized to preprocess the data in preparation for large statistical analysis. With raw NMR spectra it is necessary to align, phase and correct the baseline in order to optimize the spectra for accurate quantification of the NMR chemical signals. An internal standard of known concentration must also have been added to the sample prior to NMR acquisition. The resulting preprocessed spectra are then imported in Chenomx NMR Suite 7.5 (Alberta, Canada). Once imported, Chenomx can conduct multiple types of analysis including binning and targeted profiling on the imported spectra[56]. We elected to utilize targeted profiling which quantifies individual metabolites relative to the internal standard (Figure 2). Compound identifications are based on the

human metabolome database (HMDB), spiking experiments, published literature, and Chenomx NMR suite version 7.5 chemical signatures.

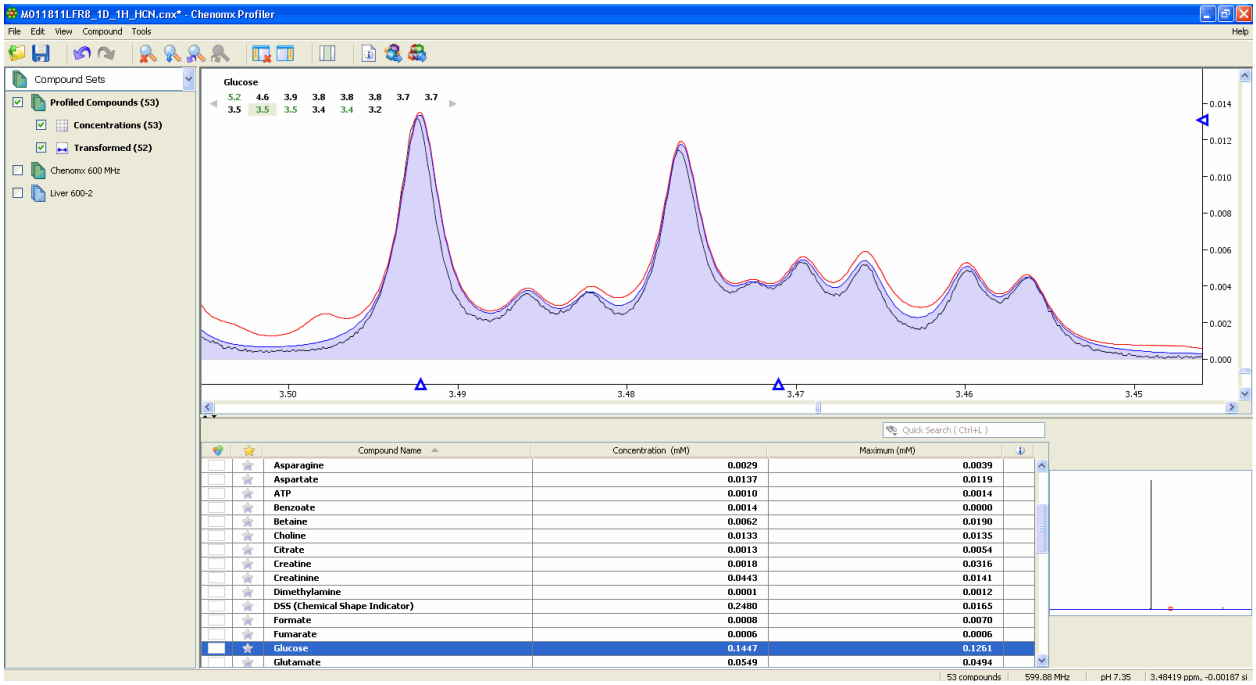

Figure 2 – Example of quantitative profiling a peak with Chemomx.

Statistical Analysis

Data Pre-treatment

Biochemical processes have a large amount of variation. Some changes may need to be very large to elicit a response whereas small changes may also result in physiologic changes. To account for these variations, data pre-treatment can be used to adjust the data to emphasize the important biological implications. One common pre-treatment method to accomplish this is scaling the data. In our experiments we utilize the Autoscaling method (i.e. unit variance scaling) which gives equal importance to all metabolites despite several orders of magnitude difference in their concentrations. There are other scaling methods such as Pareto and Vast but the Autoscaling method has been shown to perform better for explorative analyses[57].

Data Analysis

There are many ways to analyze metabolomics datasets including univariate and multivariate statistical techniques. Univariate analyses may be conducted with t-tests, analysis of variance (ANOVA), and/or regression. However, repeatedly applying a univariate test rapidly increases the Type I error rate leading to potential false positives. Corrections such as Bonferroni and Benjamini-Hochberg False Discovery Rate decrease Type I error but also increase Type II errors increasing the risk of overlooking information. Multivariate statistics, on the other hand, were developed to analyze multiple variables simultaneously thereby omitting the need for subsequent corrections. Multivariate methods are diverse and the field continues to develop alongside the expansion of 'omics' investigations. As a consequence, there are several methods that are utilized by different labs. Among the most commonly used methods are partial least squares discriminant analysis (PLSDA), random forest (RF), gradient boosting machines (GBM), support vector machines (SVM), regularized generalized linear models (GLMNET), and predication analysis of microarrays (PAM). Collectively, these are 'supervised' methods, wherein the classification information (e.g. treatment, control, etc.) is incorporated into the development of the classification model. Although each method is an effective classification algorithm previous comparisons of algorithms in gene expression experiments report that different algorithms provide improved accuracy[58]. However, an extrapolation of these prior conclusions to metabolomics data would be premature as there is no single method that is established to evaluate algorithms applied to metabolomics data. This problem is addressed in Chapter 6 with the development of a novel, open-source program for the systematic comparison of algorithm performance.

CHAPTER 5

The Transcriptomics Process

Introduction

The complex metabolic changes and the physiology of the liver and the potential application of metabolomics have been discussed in the prior chapter. In this chapter, it will be presented that transcriptomics provides another excellent tool towards investigating the impact of a carbohydrate fed state on the hepatic response to hemorrhagic shock. This chapter will discuss the field of transcriptomics, experimental design and bioinformatics analysis. This method will be applied to the porcine hemorrhagic shock and traumatic injury model in Chapter 8.

Transcriptomics experiments utilize either microarrays or next generation sequencers. Each of these methods is capable of quantifying RNA transcripts in biological samples; however, microarrays are beginning to be replaced by RNA-sequencing (RNA-seq) for multiple reasons. Microarrays require a genomic sequence of the organisms and the transcripts to be physically plated. Although high-throughput, this has several limitations including a need for higher amounts of purified RNA, limited ability to distinguish between isoforms, and high background noise. Furthermore, the expression level determined by RNA-seq of each RNA unit is measured by the number of sequenced fragments that map to the transcript, which is expected to correlate directly with its abundance level. As a consequence, the expression signal of a transcript measured by RNA-seq is limited by the sequencing depth and is dependent on the expression levels of other transcripts in contrast to microarrays probe intensities which are independent of each other[59]. This, as well as other technical differences, has motivated a growing number of statistical algorithms that implement a variety of approaches for normalization and differential expression (DE) detection[60].

Next generation sequencing (NGS) technology improvements and reduced costs has facilitated the analysis of gene expression levels and is currently being applied to complex medical investigations such as Alzheimer's Diesease[61]. RNA-seq is a powerful method whereby the identification of gene isoforms, translocation events, nucleotide variations and post transcriptional base modifications become possible in a high-throughput manner[59]. This potentially allows for the exploration of altered physiologic pathways within a given pathologic state. This makes transcriptomics a powerful technique for exploring the complex changes elicited by hemorrhagic shock and trauma.

These 'high-dimensional' datasets require further processing in order to use the RNA sequence reads. If the model organism has a sequenced genome the reads may be aligned and mapped to a reference genome otherwise the reads must be assembled *de novo*. Once the reads have been identified subsequent multivariate analysis can be applied for DE. There are multiple different tools to analyze these datasets including commercial and open-source that are continually being improved and compared[60,62,63]. Among the most commonly used is the open-source edgeR bioconductor package[64]. This package uses empirical Bayes estimation and exact tests based on the negative binomial distribution, which has been shown to best approximate RNA-seq distributions. It also has the capacity to create design matrices to incorporate multiple factors including time into the model.

Transcriptomics via RNA-sequencing

RNA Sequencing Platforms

At the start of an RNA-seq experiment a sequencing platform must be selected. There are several next generation sequences (NGS) available[65]. Most platforms now use

sequencing-by-synthesis technology but differ in the use of a polymerase or ligase. Platforms also differ by either sequencing single molecules or multiple identical copies of a molecule (i.e. ensemble). Single molecule sequencing (SMS) platforms avoid the amplification via PCR thereby providing a direct reflection of RNA expression but possess an inherently high error rate. The non-SMS platforms, such as Illumina and SOLiD, have a very low sequencing error rate but have higher numbers of mismatches. Additionally, non-SMS platforms offer increased 'sequencing depth' thereby facilitating detection of low expressed transcripts[66].

These different platforms provide data that varies and changes how it can be interpreted. Reads generated must be assembled in order to identify transcripts. Longer reads facilitate accurate assembly. Roche 454 and PacBio provide long reads (>500 nt), however, the paired end sequencing provided by Illumina allows increased nucleotide length as well (few hundred nt)[66]. For our experiments, we have selected the Illumina platform utilizing the paired end sequencing. This provided a trade-off wherein a very low error rate could be maintained while having medium length reads relative to long reads of Roche 454 and PacBio. This facilitates the exploratory nature of our experiment to investigate possible perturbations to hemorrhagic shock as a result of fed state. The remainder of this section will focus on preparation for the Illumina platform.

Sample Preparation

The first step to quality RNA-sequencing is proper sample collection. In our experiments, liver is the tissue of choice and the remainder of the chapter will be focused tissue metabolite extraction. Immediately after collection, the degradation process has begun. It is imperative to slow the tissue degradation process as much as possible to get

an accurate picture of the metabolic state. Tissue biopsies are typically flash frozen in liquid nitrogen and stored at -80°C until future metabolite extraction.

There are a few different extraction methods for RNA purification each with advantages and disadvantages. The two most common methods are the acid guanidinium thiocyanate-phenol-chloroform (AGPC) and the silica column based method. Although each method is acceptable for several experiments, the specific experimental requirements may suggest a specific method. If speed and ease of use is desired, the silica column method is often employed whereas if higher purity is required the AGPC method is used. For our experiment, the rapid silica based column, provided by Qiagen (Qiagen, Chatsworth, CA), provided satisfactory RNA quality and facilitated rapid analysis of our samples.

Library Preparation

In order to sequence the extracted RNA it must be converted into molecules that can be sequenced by the specific sequencer. For most technologies, including Illumina, a complementary DNA (cDNA) library is generated. This method facilitates the adaptation of previously developed genomic tools to sequence DNA molecules and has been widely successful in transcriptomic studies[67]. First mRNA is further purified by exploiting the poly-A tails using oligo-dT beads. The Illumina platform requires shorter reads, so the purified mRNA is fragmented to suitable lengths. These fragments are subsequently converted into cDNA by a reverse transcriptase. In this manner, cDNA contains all expressed genes but omits non-coding DNA (ncDNA) sequences such as introns, enhancers, or other regulatory elements. Adapters are subsequently bound and ligated to cDNA to allow sequencing.

Sequencing by Synthesis (SBS)

Sequencing by synthesis is the most common method for RNA-seq and is employed by Illumina. In brief, the fragmented DNA is bound by oligonucleotide ligation adaptors. These oligos facilitate hybridization to the surface of the 'flow cell', which is lined with the complementary oligos for the adapters. The DNA molecules are amplified by bridge amplification. A sequencing primer is subsequently bound to the opposite end of the molecules. Sequencing proceeds with nucleotides that were developed to emit a specific color when excited by a laser which can be subsequently recorded to identify the bases. This proceeds for a user-defined number of cycles resulting in large volumes of data which must be assembled and quantified.

Data Analysis

Mapping/Alignment

As previously mentioned, if the model organism has a previously sequenced genome the transcriptome can be mapped upon it. This 'reference-based' (i.e. *ab initio*) mapping requires an alignment program to map cDNA reads to a genome. There are several programs, commercial and open source, that are available to accomplish this analysis including BLAT[68], TopHat[69], SpliceMap[70], MapSplice[71], GSNAP[72], and QPALMA[73]. We selected TopHat for read mapping as it is a very fast mapper and one of the most commonly used open-source alignment programs[66].

Assembly

Once the reads have been mapped it becomes necessary to assemble them into transcription units. This process is known as transcriptome assembly or reconstruction and is a difficult computational task. Difficulties include the wide range of gene

expression levels (several orders of magnitude), genes originating from mature mRNA and incompletely spliced RNA complicate transcript identification, and short reads make it challenging to determine the correct isoform of a given read[74]. When a reference genome is available, genome-guided methods are often used; however, *de novo* methods such as Velvet[75] and ABySS[76] are also available for non-model organisms. Of the genome-guided methods we selected Cufflinks[77], an assembly program developed by the same group which developed TopHat. Cufflinks does not require extensive amounts of RAM and has improved detection of lower expressed transcripts than similar programs like Scripture[78].

Estimating Expression Levels

Quantifying expressing of transcripts is often the goal of transcriptomics experiments. There are multiple levels of variation that impact the expression of reads including fragmentation (resulting in longer transcripts generating more reads) and fluctuations across samples. As such, RNA-seq data requires normalization of read counts in order to interpret the data appropriately. One common metric is the reads per kilobase of transcript per million mapped reads (RPKM) which normalizes a transcript's count by the length and number of mapped reads in the samples[79]. With respect to pair-end sequencing, the similar fragments per kilobase of transcript per million mapped reads (FPKM) accounts for dependency between paired-end reads and is often the metric of choice[77]. However, when comparing the same genes between samples the effect of gene length cancel out as the same gene will be affected the same way. Furthermore, this standard scaling approach can bias estimates of differential expression in comparison to more general normalization procedures[80] to account for different library sizes. Raw

counts can be estimated with the HTSeq python script (http://www.huber.embl.de/users/anders/HTSeq/) and subsequently normalized.

Differential Expression

Once reads have been quantified it is important to determine what is different between conditions. The count-based data of RNA-seq can be modeled according to the Poisson distribution[81]. However, biological variability is always a concern in high-dimensional studies and this is not readily accounted for using the Poisson distribution. As a result, further methods have been developed that facilitate analysis of transcriptomics data sets with a small number of replicates. These include Cuffdiff[77] as well as R/Bioconductor packages DESeq[82] and EdgeR[64]. We have selected edgeR for differential expression because of the design matrix capability to incorporate repeated samples in addition to group comparisons (see Chapter 8).

Bioinformatics Analysis

Once a list of interesting genes has been determined the goal shifts towards biological interpretation. This is still a difficult and continuously developing area of bioinformatics. Approaches can be divided into two general categories, functional and/or pathway analysis. Functional analysis can be broadly defined as the study of overall functions. This can include such broad categories as cellular processes to specific functions such as actin-myosin contraction. Pathway analysis is sometimes used loosely when analyzing datasets; however, in keeping with the original idea pathway analysis is the study of up or down-regulated biological pathways such as glycolysis.

In essence, most enrichment analysis methods utilize cumulated biological knowledge in a public database such as Gene Ontology[83] or the Kyoto Encyclopedia of

Genes and Genomes (KEGG)[84]. Dozens of methods, commercial and open-source, have been developed for both functional and pathway analysis including, but not limited to DAVID[85], GSEA[86], Ingenuity Pathway Analysis (Ingenuity Systems), and Onto-Express[87]. The details of each method exceed the scope of this thesis and the interested reading is referred to these comprehensive reviews[88,89]. We have selected DAVID (Database for Annotation, Visualization and Integrated Discovery) because the analysis provides both functional annotations of each gene (Functional Annotation Table) and of each list of genes (Functional Annotation Clustering) to provide functional information. Additionally, the Functional Annotation Clustering tool consolidates redundancies in gene ontology categories. In this way, DAVID provides a comprehensive exploratory analysis of our data (see Chapter 8).

CHAPTER 6

Optimal Algorithm for Metabolomics Feature Selection and Classification varies by Dataset

ABSTRACT

Metabolomics, the unbiased and systematic identification and quantification of all metabolites in a biological system, is increasingly applied towards identification of biomarkers for disease diagnosis, prognosis and risk prediction. Applications of metabolomics extend across the health spectrum including Alzheimer's, multiple forms of cancer, diabetes, and trauma. Despite the continued interest in metabolomics there are numerous techniques for analyzing metabolomics datasets with the intent to classify group membership (i.e. Control or Treated). These include Partial Least Squares Discriminant Analysis, Support Vector Machines, Random Forest, Regularized Generalized Linear Models, and Prediction Analysis for Microarrays. Each classification algorithm is dependent upon different assumptions and can potentially lead to alternate conclusions. This project seeks to conduct an in depth comparison of algorithm performance on both simulated and real datasets to determine which algorithms perform best with supporting theoretical justification. Three simulated datasets were generated to validate algorithm performance and mimic 'real' metabolomics data: (1) independent null dataset (no correlation, no discriminatory variables), (2) correlated null (no discriminating variables), (3) correlated discriminatory. This comparison is also applied to 3 open-access datasets including two NMR and one MS dataset. Performance will be evaluated based on the Robustness-Performance-Trade-off (RPT) incorporating a balance between model accuracy and feature selection stability. We also provide a free, open-source R Bioconductor package (*OmicsMarkeR*) that conducts the analyses herein. The proposed work provides an important advancement in metabolomics analysis and helps

alleviate the confusion of potentially paradoxical analyses thereby leading to improved exploration of disease states and identification of clinically important biomarkers.

INTRODUCTION

Metabolomics, similar to the other two common 'omics' approaches (i.e. transcriptomics and proteomics), is defined as the unbiased and systematic identification and quantification of all metabolites in a biological system. Such data is most commonly acquired via Nuclear Magnetic Resonance spectroscopy (NMR) or Mass Spectrometry (MS). The growing field of metabolomics has been increasingly applied towards identification of biomarkers for disease diagnosis, prognosis and risk prediction. Applications of metabolomics extend across the health spectrum including Alzheimer's[90], multiple forms of cancer[91–93], diabetes[94], and trauma[95].

Following the initial pre-processing (e.g. peak picking, deconvolution, integration, etc.), the dataset must ultimately be analyzed to typically classify two or more classes/conditions in addition to identifying the most important metabolites for the discrimination (e.g. biomarker studies). The availability and use of multivariate approaches is rapidly becoming critical with decreased cost and increased access to high-throughput metabolomics platforms including NMR and MS resulting in "large *p*, small *n*" problems (i.e. many more variables than samples). The common univariate tests become grossly underpowered to assess every feature and require a secondary model if classification is desired. The restrictive assumptions of univariate tests (e.g. normality) are typically avoided with more sophisticated multivariate and machine learning algorithms.

But despite the continued interest in metabolomics there is no standard statistical approach resulting in the use of numerous techniques that vary between experiments often with little or no justification. Common methods include Partial Least Squares

Discriminant Analysis (PLSDA), Lasso and Elastic-Net Regularized Generalized Linear Models (GLMNET), Support Vector Machines (SVM), Random Forests (RF), Gradient Boosting Machines (GBM), and Prediction Analysis for Microarrays (PAM). Although each method is an effective classification algorithm, previous comparisons of algorithms in gene expression experiments report that different algorithms provide improved accuracy[58].

Limited algorithm comparisons in metabolomics studies (i.e. comparing two or three methods) often measure performance solely on accuracy and neglect feature selection stability. Even though an analysis reports high accuracy, repeating the biomarker discovery procedures can result in different feature subsets even within the same datasets[96–98]. Therefore, to determine which algorithms perform optimally must be evaluated with regards to both feature selection stability and overall classification accuracy.

In this work, we evaluate the 6 aforementioned classification algorithms performance and stability on both *in silico* and experimentally acquired datasets. Metabolomics datasets are inherently multivariate with both independent and multicollinear variables in addition to possessing a mix of Gaussian and non-Gaussian distributions. To evaluate algorithm performance on such datasets it is necessary to generate standardized datasets that mimic true metabolomics data and possess known results as a benchmark. Furthermore, an application to previously acquired datasets from multiple platforms with previous results is desired.

METHODS

Datasets

Although metabolomics has been more commonly used there is no standard dataset with known results for use in evaluating current or developing algorithms. This necessitates the production of a simulated datasets that accurately mimic typical metabolomics datasets from both NMR and MS. This requires the perturbation of normality in multiple variables and inclusion of multicollinearity as is typical of metabolite distributions and relationships. It is also necessary to determine the performance of algorithms when examining the null condition wherein there is no difference between conditions. Therefore, we propose to generate three simulated datasets (null independent, null correlated, and correlated discriminatory) to analyze algorithm performance that may also be used by others for further performance evaluations. This will be repeated twice for matrices of both NMR and MS scale as the number of resolvable features between the two techniques can be an order of magnitude as NMR typically can resolve 50-75 metabolites whereas MS can resolve 100's to 1000's of metabolites[99]. Although exceedingly large datasets are possible with *in silico* data we chose to select sample sizes that may more accurately reflect empirical datasets given the limits from costs and/or sample availability. Low and high sample sizes were set at 25 and 50 samples per group respectively. Although these are still high for many applications, this allows the use of leave k-fold out cross-validation. For very small datasets one may use leave-one-out cross-validation or the option to forgo validation, if appropriate parameters are known, is available (*optimize* = FALSE).

Simulated Datasets

1. ***Null***: Simulated data will be generated with the *create.random.matrix* function as described in Wongravee et. al[100] with the following noted modifications. The initial

dataset will be generated with random numbers from a normal distribution, each consisting of N samples (*nsamp,* 25/50 per group) and Y variables (*nvar*, NMR = 50, MS = 1000). The normal distribution will be perturbed and samples assigned to groups of equal numbers. We chose to use an alternate numbers of samples and variables to represent more accurately numbers more commonly seen in the literature.

2. ***Null Correlated***: To mimic 'real' metabolomics datasets, correlations will be induced for the second dataset with the *create.corr.matrix* function. Blocks of variables of size b will be randomly assigned and have values replaced with correlated values derived from the first column. We elected to incorporate blocks of size 1 as the smaller metabolomics datasets in NMR possess independent variables. This induced correlation will be perturbed to more accurately represent real data. Derived correlation coefficients will be compared to real metabolomics datasets to validate the method.
3. ***Discriminatory***: To facilitate discriminatory analysis, *D* variables (NMR = 10, MS = 20) will be randomly induced to be discriminatory with the *create.discr.matrix* function. A discriminatory index (*l*) will be selected and for each variable D a random number between –l and l will be added to one group and subtracted from the other.

Real Datasets

NMR datasets include a binary (i.e. 2 groups) urine dataset analyzing cachexia[101] and multi-class (i.e. 4 groups) rumen fluid dataset investigating the impact of altered diets of cows[102]. The MS explores potential biomarkers of Hepatocellular Carcinoma in serum

samples[103] and was accessed from the open source metabolomics data repository Metabolights[104] accession number MTBLS19.

Classifier and Feature Selection Algorithms

1. Partial Least Squares Discriminant Analysis (PLSDA)

Partial Least Squares Discriminant Analysis is a dimension reduction technique analogous to principal component analysis. The algorithm focuses on maximizing the variance of the dependent variables explained by the independent variables[105]. It is robust to multicollinearity, missing data, and skewed distributions[106]. Feature selection is accomplished by weighting the sum of squares of the variable weights known as variable importance of projection (VIP). Feature selection is accomplished when features are ranked with respect to VIP scores to denote important variables. Model derived features will be selected as those with a VIP score >= 1.0. PLSDA is commonly used in metabolomics investigations including multiple forms of cancer, cardiac ischemia, parkinson's disease and asthma[43,93,107–111]. This technique has also been previously implemented in our lab investigating hemorrhagic shock[112]. It is readily available in the R package DiscriMiner[113].

2. Regularized General Linear Model (GLMNET)

Generalized linear models are a more flexible form of linear regression that allows the response variables to have non-parametric distributions. To avoid the risk of overfitting data in multiple linear regression a regularization method can be applied. We have selected to use the elastic-net penalty, which is a compromise between the LASSO and Ridge shrinkage methods and has been shown to outperform LASSO[114]. In brief, elastic net is a weighted average of the lasso and ridge solutions determined the weighted

parameter lamda (λ). This facilitates analysis of data with collinearity and facilitates internal feature selection. Important features are identified as those with non-zero coefficients. These coefficients will also be ranked for subset feature selection. Although it is less common than other techniques it has been used in recent metabolomics studies[115,116]. GLMNET is readily available in the R package glmnet[117].

3. Random Forest (RF)

Random forest is a machine learning algorithm that uses a combination of tree predictors such that each tree depends on the values of a random vector samples independently and the distributions are the same for all the trees in the forest. Each tree constructed provides a 'vote' for the best class. This is constructed on a training subset of the data and tested against the remaining test data known as the 'out-of-bag' (OOB) data. The scaled sum of the votes derived from the trained trees determines the final "score"[118]. The feature selection is determined by permuting variables in the OOB and observing increases in error. These scores will be ranked for subset feature selection and those exceeding a score of 1.0 for model derived results. A variable score indicates greater importance to the model. It is robust to noise and outliers and computationally faster than bagging or boosting. Prior studies have reported error rates comparable if not better than other predictors such as logistic regression, linear discriminant analysis, quadratic discriminant analysis (QDA), K-nearest neighbors (KNN), Support Vector Machines (SVM), classification and regression trees (CART) and Naïve Bayes[118–120]. However, consistency of selected feature rankings has been shown to be problematic[121]. It has been used in several metabolomics studies[122–125] and is readily available in the R package randomForest[126].

4. Gradient Boosting Machine (GBM)

Gradient boosting is another machine learning technique applied most commonly to decision trees that produces robust and interpretable procedures for both regression and classification[127]. Unlike the bagging approach (e.g. random forest) earlier trees are considered to compensate for prior weaknesses in subsequent trees. As with random forest, feature selection is determined by permuting variables in the OOB and observing increases in error resulting in a subsequent variable score. Feature selection will be accomplished via ranking scores and those exceeding 1.0 for model derived results. Boosting has become known as one of the most powerful learning ideas in the last twenty years[128] but curiously has never been applied to metabolomics settings. To our knowledge, this will be the first application of boosting to analyze metabolomics data. Freidman's gradient boosting machine algorithm is available in the R package gbm[129].

5. Support Vector Machines (SVM)

Support vector machine is based on the structural risk minimization principle from statistical learning theory[130]. It can be applied to classification problems with the idea of structural risk minimization to find a hypothesis for which has the lowest probability of error. It has been shown to be robust to both noisy data and outliers. Prior comparisons with PLSDA report improved overall accuracy with less features[131] but feature selection consistency is unknown. This classification algorithm is readily available within the R e1071 package[132]. Feature selection will be accomplished via recursive feature elimination (RFE) as detailed by Guyon[133].

6. Prediction Analysis for Microarrays (PAM)

Prediction Analysis for Microarrays is a modified nearest centroid classification method to include centroid shrinkage and contains an embedded feature selection step[134]. In brief, the average value for each variable is divided by the within-class standard deviation to provide class centroids. These class centroids are then shrunk towards zero by a defined threshold to reduce noise and facilitate variable selection. Then a new sample profile is compared to each of the class centroids. The class whose centroid is closest is the predicted class. The internal feature selection is accomplished by identifying features with non-zero coefficients and are subsequently ranked for subset selection. This technique has not been used widely in metabolomics investigations; however, as the name implies it has been successfully been used for classification in gene expression experiments[135,136]. This algorithm is readily available in the R package pamr**.**

Evaluate Stability of Feature Selection Techniques

The high-dimensional datasets of metabolomics often necessitate feature selection techniques to reduce dimensionality to the most important features to facilitate subsequent analysis. Although many approaches rely exclusively on classification accuracy of feature subsets to facilitate biomarker selection this is problematic where several different feature subsets may yield equally optimal results[137]. It is therefore necessary to evaluate the robustness of feature selection techniques applied to metabolomics data to facilitate improved reproducibility and confidence in identified biomarkers. In brief, algorithm robustness will be evaluated via instance (bootstrapped data subsets) and function (alternate algorithms) perturbation and evaluated by the simple and commonly used Jaccard's Index[138]. Other options include the Dice-Sorensen's Index[139,140], Ochiai's Index[141], Percent of Overlapping Features[142], Kuncheva's Index[143],

Spearman Rank Correlation, and Canberra Distance[144]. A comparison of these metrics is beyond the scope of this article.

Two common approaches are applied within instance perturbation to evaluate the robustness of feature selection techqniues: perturbation at the instance level (i.e. removing or adding samples) or at the feature level (i.e. adding noise). We have selected to evaluate robustness of feature selection algorithms by estimating stability following perturbation at the instance level as the number of samples are the most likely problem facing metabolomics investigations.

Single Feature Selection Stability and Classification Performance

For each feature selection algorithm we estimated stability via instance perturbation with the *fs.stability* function. Instance perturbation parameters will bootstrap 90% ($p = 0.9$) of the data 10 times ($k = 10$) thereby creating a training and testing dataset for each iteration. For each training dataset all 6 feature selection algorithms were run simultaneously to provide a list of selected feature rankings. Each iteration tunes the full model (*optimize* = TRUE) with a tuning grid of a specified resolution determining how fine the tuning parameters are optimized (*resolution = 5*). To avoid overfitting, 10-fold cross-validation is utilized (*k.fold* = 10) wherein $1/10^{th}$ of the data is randomly removed and the model evaluated on this test fold. Results were averaged over all 10 folds to provide the confusion matrix for subsequent performance metrics. The optimized models are then used to extract feature subsets of a user specified length (*f*, NMR = 10, MS = 20) or optionally by the model defined cutoff (*model.features = FALSE*). These feature subsets are compared via the Jaccard index. The overall stability is defined as the average over all pairwise similarity comparisons between each of the feature selection

runs. The final model is refit using the extracted feature subset and re-optimized using the initial tuning grid generated. Lastly, this trimmed model is used to predict the initial testing dataset generated at the start of the iteration. The metrics include Accuracy and Area Under the Receiver Operator Curve (AUC-ROC). These values were all extracted with the *performance.metrics* function to compare each algorithms performance. This is repeated for the additional 9 times utilizing the previously optimized parameters for the full model generation (*optimize.resample = FALSE*).

Balance Stability and Classification Performance

In every scientific investigation in which sample size is a limitation (i.e. most studies), researchers must balance power and sensitivity. The same principle is applied to balancing feature selection robustness and classifier performance as both are integral to confident biomarker identification. We will utilize the robustness-performance trade-off (RPT) to balance robustness and classification performance[145]. In brief, the user can specify the parameter β to control the relative importance of robustness versus classification. The default value of $\beta = 1$ which represents equal importance between robustness and classification.

Ensemble Feature Selection Stability and Classification Performance

Ensemble feature selection has been shown to improve stability in gene expression studies[146,147]. Therefore it is important to incorporate such analysis into metabolomics analysis for each algorithm. In essence, ensemble approaches use different data subsets and aggregating the results following feature selection. As described in the 'Single Feature Selection Stability and Classification Performance' section, stability will be evaluated via instance perturbation with the *fs.ensembl.stability* function. For each

subsample a second level of instance perturbation will generate 40 (*bags* = 40) further datasets via bootstrap aggregation (aka. Bagging)[148]. For each bag a separate feature ranking will be performed. The resulting list of selected feature rankings from each bag will be combined via linear aggregation (*aggregation.metrics* = "CLA"). The Jaccard Index will be used to measure similarity and averaged over all pairwise comparisons for an overall measure of stability. Function perturbation, the use of multiple feature selection algorithms, is also conducted by the list of methods chosen within *fs.stability* and *fs.ensembl.stability*. Lastly, in contrast to non-ensemble approaches, there are no model derived runs because all features must be ranked for aggregation methods.

RESULTS

Simulated Data

Binary Classification - Low Samples

Non-ensemble analysis of random and correlated dataset analyses provided generally expected results. Accuracy exceeded 0.700 for SVM, RF, GLMNET and PAM but stability remained low (>=0.47). While RF achieved the highest accuracy (often in excess of 0.900), it had the lowest stability warranting caution in interpreting results. Notably, accuracy was generally higher with the MS-scale dataset where accuracy exceeded 0.900 for the same four algorithms. However the stability of the feature subsets was also lower (Supplementary File – S1).

Analysis of the NMR-scale discriminatory dataset determined PAM as the optimal model with the highest RPT and TP% of identified features (Table 1). GLMNET performed similarly with better accuracy but lower stability and TP%. The model derived analysis also provided PAM with the highest RPT, however, the low sample size

resulted in a conservative trimming of features resulting in many remaining in the model and decreasing the TP%. The highest TP% was reported by SVM which also had the very high accuracy and AUC-ROC; however it is also noted that there is no internal trimming metric for SVM and only the top 10% of features are returned making this a more restricted subset model. This suggests that experiments with lower sample sizes may need to restrict to only a few of the most discriminate features. The MS-scale datasets determined also reported PAM with the highest RPT and stability but the highest TP% was reported by PLSDA. Such a situation supports the value of using multiple algorithms to determine consistent results.

Ensemble analysis of NMR-scale and MS-scale random and correlated datasets again reflected previous analysis with high accuracy levels for SVM, RF, GLMNET and PAM but low stability (Appendix B – Tables 2 & 6). Analysis of NMR-scale discriminatory dataset reported SVM with the highest RPT with PLSDA, GLMNET and PAM performing similarly. The MS-scale analysis was less conclusive with mixed performance among SVM, GLMNET and PAM (Appendix B – Table 10).

Binary Classification – High Samples

Increased sample size had little effect on accuracy and stability of NMR-scale random and correlated datasets but worsened models of MS-scale data (Appendix B – Table 5). Overall performance greatly improved in the NMR-scale discriminatory dataset with accuracy exceeding 0.9 for three algorithms, stability exceeding 0.8, and TP% to 80% (Table 1). Performance in the MS-scale dataset also improved with stability and TP% increasing most notably for GBM and RF respectively (Appendix B – Table 9). Both

datasets provide a circumstance whereby no algorithm performs best and multiple methods are beneficial.

As expected, there was little effect of high samples on the ensemble analysis of the both random and correlated datasets of MS-scale and NMR-scale sizes. The results of the ensemble NMR-scale discriminatory dataset improved RF and SVM but worsened GLMNET and PAM (Table 1). Likewise, the ensemble MS-scale discriminatory dataset also improved performance (Appendix B – Tables 9). However, the best performing algorithms were PLSDA, PAM and GLMNET. These results suggest that ensemble aggregation may improve, worsen, or have no effect on model performance.

Multiclass Classification - Low-Samples

NMR-scale random and correlated datasets generally had low accuracy and stability whereas MS-scale had four algorithms consistently had accuracy >= 0.700 but stability still remained low (Appendix B – Table 8). Analysis of the NMR-scale discriminatory dataset determined SVM as the best performing algorithm with the highest RPT whereas PAM and RF had the highest stability and accuracy respectively. The model derived results report GLMNET and PAM among the best; however, as with the binary classification problems the low sample size resulted in untrimmed features wherein nearly all were retained resulting in a decreased TP%. Curiously, SVM reported the highest TP% (64%) again suggesting that the lower sample size may restrict to only a few of the most discriminating features (Table 2). The MS-scale dataset had high predictive accuracy with SVM, GLMNET and PAM but very low with PLSDA and GBM. Curiously, PLSDA had the highest TP% but in addition to a fair AUC-ROC (0.721) and

stability (0.53) suggesting a potential use in comparison to better classifying algorithms for feature selection (Appendix B – Table 12).

Ensemble analysis of random and correlated datasets had little effect on performance of both NMR-scale and MS-scale (Appendix B – Table 8). Analysis of the NMR-scale discriminatory dataset reported GLMNET with the highest RPT but PLSDA with the highest TP% (Table 2). This reflects the single run analysis whereby the improved stability of ensemble methods provided improved classification and stability of GLMNET. The MS-scale analysis reported SVM as the optimum algorithm with the highest RPT and accuracy but PAM reported the highest TP% (Appendix B – Table 12).

Multiclass Classification - High Samples

NMR-scale random and correlated datasets continued to provide expected results, whereby most models had poor classification. In contrast, MS-scale data provided high accuracy but continued to provide very poor stability (Appendix B – Table 7). Analysis of the NMR-scale discriminatory dataset determined GLMNET as the best performing algorithm despite lower accuracy (Table 2). Model derived results determined GLMNET and PAM as the best models however they did not successfully extract any discriminating features resulting in a depleted TP% whereas SVM reported the highest TP% at 60%.

Ensemble analysis of random and correlated datasets again reflected previous analysis with low accuracy and stability for both NMR-scale and MS-scale data. Analysis of the discriminatory dataset did not significantly improve performance for the NMR-scale dataset. This is also reflected in the MS-scale data where only RF and PLSDA stability improved slightly (Appendix B – Table 11).

			Method	RPT	Accuracy	AUC-ROC	Stability	# Features	TP %
Binary	High Sample	Subset	PLSDA	0.923	0.982	0.551	0.87	10	80.0%
			GBM	0.748	0.710	0.730	0.79	10	76.0%
			SVM	0.742	1.000	1.000	0.59	10	61.0%
			RF	0.529	1.000	1.000	0.36	10	59.0%
			GLMNET	0.942	1.000	1.000	0.89	10	80.0%
			PAM	0.936	1.000	1.000	0.88	10	80.0%
		Model Derived	PLSDA	0.931	0.965	0.698	0.90	10	80.0%
			GBM	0.552	0.480	0.705	0.65	14	55.0%
			SVM	0.770	0.990	1.000	0.63	5*	68.0%
			RF	0.549	0.990	1.000	0.38	17	38.8%
			GLMNET	0.927	0.980	0.984	0.88	8	92.5%
			PAM	0.860	0.990	0.985	0.76	44	22.5%
		Ensemble	PLSDA	0.928	0.982	0.551	0.88	10	80.0%
			GBM	0.562	0.430	0.773	0.81	10	80.0%
			SVM	0.773	1.000	1.000	0.63	10	76.0%
			RF	0.765	1.000	1.000	0.62	10	76.0%
			GLMNET	0.817	1.000	1.000	0.69	10	66.0%
			PAM	0.930	1.000	1.000	0.87	10	80.0%
	Low Sample	Subset	PLSDA	0.690	0.875	0.705	0.57	10	61.0%
			GBM	0.395	0.300	0.669	0.58	10	56.0%
			SVM	0.667	1.000	1.000	0.50	10	48.0%
			RF	0.359	0.975	1.000	0.22	10	48.0%
			GLMNET	0.801	0.975	1.000	0.68	10	68.0%
			PAM	0.806	0.900	0.831	0.73	10	78.0%
		Model Derived	PLSDA	0.792	0.895	0.597	0.71	14	50.0%
			GBM	0.585	0.525	0.744	0.66	20	36.5%
			SVM	0.802	1.000	1.000	0.67	5*	68.0%
			RF	0.406	0.925	1.000	0.26	17	25.9%
			GLMNET	0.830	1.000	1.000	0.71	32	28.4%
			PAM	0.925	1.000	1.000	0.86	48	20.8%
		Ensemble	PLSDA	0.694	0.863	0.856	0.58	10	66.0%
			GBM	0.455	0.350	0.750	0.65	10	62.0%
			SVM	0.788	1.000	1.000	0.65	10	70.0%
			RF	0.606	0.975	1.000	0.44	10	56.0%
			GLMNET	0.765	0.975	1.000	0.63	10	57.0%
			PAM	0.748	0.900	0.831	0.64	10	72.0%

Table 1 – Results from NMR-scale Binary Classification Simulations. RPT – Robustness-Performance Trade-off, AUC-ROC – Area under the Receiver Operator Curve, TP % - Percent true positive identifications. *SVM doesn't have internal cutoff so defaults to top 10%

			Method	RPT	Accuracy	AUC-ROC	Stability	# Features	TP %
Multi-class	High Sample	Subset	PLSDA	0.432	0.380	0.622	0.50	10	49.0%
			GBM	0.398	0.340	0.691	0.48	10	46.0%
			SVM	0.531	0.565	0.846	0.50	10	42.0%
			RF	0.257	0.900	0.966	0.15	10	30.0%
			GLMNET	0.615	0.575	0.837	0.66	10	52.0%
			PAM	0.586	0.520	0.670	0.67	10	53.0%
		Model Derived	PLSDA	0.441	0.373	0.678	0.54	12	44.2%
			GBM	0.926	0.895	0.823	0.96	50	19.6%
			SVM	0.559	0.485	0.803	0.66	5	60.0%
			RF	0.326	0.875	0.940	0.20	19	16.3%
			GLMNET	0.786	0.675	0.986	0.94	50	20.0%
			PAM	0.780	0.640	0.730	1.00	50	20.0%
		Ensemble	PLSDA	0.456	0.380	0.623	0.57	10	54.0%
			GBM	0.344	0.265	0.757	0.49	10	44.0%
			SVM	0.467	0.510	0.820	0.43	10	39.0%
			RF	0.299	0.890	0.950	0.18	10	35.0%
			GLMNET	0.538	0.495	0.806	0.59	10	34.0%
			PAM	0.557	0.520	0.670	0.60	10	52.0%
	Low Sample	Subset	PLSDA	0.485	0.394	0.815	0.63	10	58.0%
			GBM	0.397	0.325	0.666	0.51	10	38.0%
			SVM	0.723	0.888	0.904	0.61	10	52.0%
			RF	0.356	0.938	0.967	0.22	10	39.0%
			GLMNET	0.673	0.688	0.875	0.66	10	52.0%
			PAM	0.715	0.675	0.761	0.76	10	53.0%
		Model Derived	PLSDA	0.501	0.425	0.661	0.61	13	40.8%
			GBM	0.517	0.375	0.739	0.83	47	20.4%
			SVM	0.752	0.800	0.888	0.71	5	64.0%
			RF	0.365	0.888	0.954	0.23	17	21.2%
			GLMNET	0.871	0.863	0.938	0.88	49	19.4%
			PAM	0.900	0.825	0.838	0.99	50	20.0%
		Ensemble	PLSDA	0.485	0.394	0.815	0.63	10	62.0%
			GBM	0.388	0.300	0.665	0.55	10	38.0%
			SVM	0.654	0.788	0.917	0.56	10	53.0%
			RF	0.389	0.875	0.996	0.25	10	39.0%
			GLMNET	0.751	0.763	0.913	0.74	10	47.0%
			PAM	0.746	0.775	0.802	0.72	10	53.0%

Table 2 – Results from NMR-scale Multi-class Classification Simulations. RPT – Robustness-Performance Trade-off, AUC-ROC – Area under the Receiver Operator Curve, TP % - Percent true positive identifications. *SVM doesn't have internal cutoff so defaults to top 10%

REAL DATASETS

Eisner – Urine analysis of Cachexia via NMR

Non-ensemble analysis determined PAM and GLMNET as the best overall models according the RPT (0.711, 0.705). Important features were extracted with the *feature.table* function. Adipate, Glucose, 3-Hydroxyisovalerate were identified in all subsamples by both models. PAM also identified creatine and succinate consistently (Table 3). GLMNET identified leucine, quinolinate, and valine as important features (Table 4). Seven of these metabolites were within the top 10 metabolites identified by Eisner et. al.[101] with creatine being the 11th. Furthermore, the other 3 metabolites in the top 10 (myo-inositol, betaine, and *N*,*N*-dimethylglycine) were also identified in the top 10 by GLMNET and PAM. Random forest provided the best classification accuracy at the expense of stability. SVM performed similarly well with respect to classification accuracy (Appendix B – Table 13), however, stability was also quite low (0.31).

Ensemble methods improved stability of GBM, SVM, and RF; however, it noticeably decreased GLMNET stability from 0.64 to 0.49 (Appendix B – Table 16). Performance decreased slightly for PLSDA but improved for PAM; both proved the best overall models. Overall 8 of the 9 identified metabolites (frequency >= 0.9) by PLSDA and PAM were in the top 10 identified by Eisner et.al. (Appendix B – Tables 17 & 18). For this particular dataset the ensemble approach does not appear necessary as neither model performance nor feature stability was significantly improved. Irrespective, the applied methods provide further validation and support for the classification and metabolites selected.

PAM Feature Table		
features	consistency	frequency
Adipate	10	1
Creatine	10	1
Glucose	10	1
Succinate	10	1
X3.Hydroxyisovalerate	10	1
myo.Inositol	9	0.9
Betaine	7	0.7
Glutamine	7	0.7
Quinolinate	6	0.6
cis.Aconitate	6	0.6
Acetate	5	0.5
N.N.Dimethylglycine	5	0.5
Lysine	3	0.3
Leucine	2	0.2

Table 3 – Feature table of PAM analysis consisting of consistency (i.e. number of times the feature identified as important) and frequency (i.e. percentage of iterations feature identified).

GLMNET Feature Table		
features	consistency	frequency
Adipate	10	1
Glucose	10	1
Leucine	10	1
Quinolinate	10	1
Valine	10	1
X3.Hydroxyisovalerate	10	1
myo.Inositol	9	0.9
Succinate	6	0.6
Betaine	4	0.4
Glutamine	4	0.4
Lysine	4	0.4
Creatine	3	0.3
N.N.Dimethylglycine	3	0.3
Acetate	3	0.3
Alanine	2	0.2
Formate	1	0.1
Xylose	1	0.1

Table 4 – Feature table of GLMNET analysis consisting of consistency (i.e. number of times the feature identified as important) and frequency (i.e. percentage of iterations feature identified).

Ametaj – Analysis of Rumen Metabolism via NMR

The PAM and GLMNET models performed very well with the Ametaj dataset with RPT values of 0.850 and 0.806 respectively. Although PLSDA did not have a high RPT (0.497), the stability was very high (0.77) and was therefore evaluated for consistency with PAM and GLMNET. Glucose, endotoxin, and methylamine were consistently identified by PAM, GLMNET and PLSDA (Appendix B – Tables 20-22). Glucose and endotoxin were expected and methylamine was the first statistically significant metabolite discussed by Ametaj et.al.[102]. In addition, uracil, acetate, fumarate, and lactate were also consistently identified by at least two models. All of these metabolites are identified by Ametaj et.al. except for lactate which is reported as non-significant but was approaching significance (0.149). This suggests a potential power issue although the authors' comment on the conversion of lactate to proprionate in ruminates is well supported.

Ensemble methods improved stability of PLSDA, GBM, SVM, and RF; however, it slightly decreased GLMNET and PAM stability by 0.02 and 0.07 respectively (Appendix B – Table 23). This general improvement is expected with smaller sample sizes as variability often greater. The stability of all models was very high with all ≥ 0.67 except for RF. Stability of PLSDA was highest; however, classification accuracy was poor (0.438). Accuracy increased slightly with PAM and GLMNET but decreased in SVM to match GLMNET. These three proved to be the best overall models. Although the stability was high, identified features did vary between models creating what we refer to as a 'hierarchy of confidence' whereby the greatest confidence would be placed in the most consistently identified features within and across algorithms. Endotoxin, glucose,

and methylamine were once more identified by four models (PAM, SVM, PLSDA and GBM). Alanine was also identified by 3 models (PAM, GLMNET, PLSDA). In addition, acetate, 3-PP, and uracil were also identified by at least two models consistently. All of these metabolites were identified as important by Ametaj et.al. Lastly, although most results were consistent, ferulate was also identified by PAM and SVM which was not discussed previously. It is also apparent that given the small sample size and greater variability within the data, this analysis benefits from ensemble methods.

Xiao – Serum Analysis of Hepatocellular Carcinoma via MS

Within the negative mode comparisons PAM mostly performed better than all other algorithms applied with respect to RPT (Appendix B – Tables 28-35). There was only one exception where PAM and GLMNET had almost identical RPT values (0.679, 0.680). However, PAM stability was consistently the highest suggesting it to be the ideal method to explore this particular dataset. Additionally, it should be noted that PAM did not have very high accuracy compared to other methods such as GLMNET and SVM. Random Forest also had very high accuracy but exceptionally low stability making it a good tool for analyses that require high classification accuracy but do not require further feature identification. The higher accuracy makes GLMNET a close second to the PAM approach. An ensemble analysis was performed, however only RF improved stability significantly rendering the ensemble analysis of little help. Identified features were largely consistent with previous analyses (Supplementary File 7), however, each comparison identified at least an additional 15 metabolites not mentioned in the published manuscript.

Within the positive mode comparisons GLMNET performed slightly better than PAM in all except the 1 comparison; however, PAM consistently maintained the highest stability demonstrating that even with very high-dimensional datasets no single algorithm dominates. As with the negative mode, an ensemble analysis was performed but only RF improved stability significantly making ensemble aggregation unnecessary. As with the NMR studies the two models provided further support to mutually identified features (Supplementary File 8). These features included GDCA, Oleoylcarnitine, GCDCA, L-N_2-(2-Carboxyethyl) arginine, Tetracosahexaenoic acid, Palmitoyl carnintine and Linoelaidyl carnitine which support the author's interpretations of bile and fatty acid metabolism. In addition, 48 metabolites were consistently identified by PAM but were not mentioned in the paper however it is possible they were identified as the authors do not present all identified metabolites.

DISCUSSION

The results presented focus on determining if a single classification algorithm performs better than other commonly used algorithms in the field of metabolomics. Although there is strong support for several algorithms, there are additional papers stating one algorithm outperforms another. This is expected according to the concept of No Free Lunch Theorem[149] which in essence states that there is no single model that is appropriate for all problems. Our analyses support this theory and suggest a comparative approach to evaluate algorithm performance as well as to provide additional support via multiple algorithms for future conclusions. As such, the Bioconductor package OmicsMarkeR was created to facilitate the rapid comparison of algorithm performances on individual datasets.

The binary and multivariate simulation results demonstrate the variation in optimal models. Depending on the users approach to a dataset (i.e. feature subset or model derived), number of samples relative to features, the number of groups, or application of ensemble methods a different algorithm may be more appropriate. Within the simulation analyses PLSDA, SVM, PAM, RF, and GLMNET all proved to be optimal algorithms for different situations. Curiously, the reduction in features generally decreased GBM performance but feature selection was relatively consistent with other methods. Random Forest, in most cases, did not prove to be the optimal algorithm but consistently had among the highest accuracy but very low stability; which could be partially mitigated by an ensemble approach. This low stability is expected given the algorithms search encompasses interactions between variables resulting in a larger search space. If a user wished to weigh accuracy more than stability, the RPT value can be easily recalculated with the *RPT* function by increasing the beta parameter (e.g. beta = 1.5 increases the weight of accuracy by 50%).

The Eisner cachexia NMR dataset provided an example of the value behind using multiple algorithms. In the event that algorithms have similar overall performance, analysis of features selected by both may provide further support for identified features. Additionally, this dataset demonstrated the circumstance whereby ensemble approaches are superfluous. Consistent with published results, the Eisner dataset did reflect the high accuracy of SVM but stability was low. The GLMNET and PAM analyses provided respectable accuracy and high stability. The ability to assess stability of identified features for classification models is valuable if rapid and clinically applicable tests are to be developed.

The Ametaj cow diet NMR dataset also benefited from multiple algorithms adding further support to identified features. Furthermore, the multi-class experimental design and low number of samples likely resulted in higher variability. As such the ensemble approach was beneficial by improving most algorithm performances.

The Xiao Hepatocellular Carcinoma MS dataset was analyzed in four separate comparisons within each mode (i.e. positive or negative) following the manuscripts analysis. The negative mode results suggest that PAM is generally the optimum algorithm with respect to RPT values. The positive mode results suggest that GLMNET is the optimum algorithm. However, the goal behind the study was the identification of biomarkers. As such, the consistently higher stability of PAM is noteworthy. A potential extension would be to use the features identified by PAM and build the GLMENT model (or other method) for class prediction to improve accuracy and/or AUC-ROC.

As an example, we provide a very concise sample summary of the final positive comparison which could be used as a general guideline for reporting results: *Six commonly applied metabolomics algorithms were tuned and cross-validated with the OmicsMarkeR bioconductor package via 10-fold cross-validation (k.fold = 10). Feature selection was accomplished with default model parameters. After comparison of overall performance Prediction Analysis of Microarrays (PAM) provided the highest Robustness-Performance-Trade-off (RPT) balancing both classification accuracy (73%) and stability of identified metabolites (Jaccard = 0.74) reflecting good prediction and reproducible metabolite identification. Metabolites identified in all data perturbation runs (k = 10) include GDCA, Oleoylcarnitine, GCDCA, L-N_2-(2-Carboxyethyl) arginine, Tetracosahexaenoic acid, Palmitoyl carnintine and Linoelaidyl carnitine. Forty-eight*

(48) additional metabolites were consistently identified that may be pursued via tandom mass spectrometry (MS/MS).

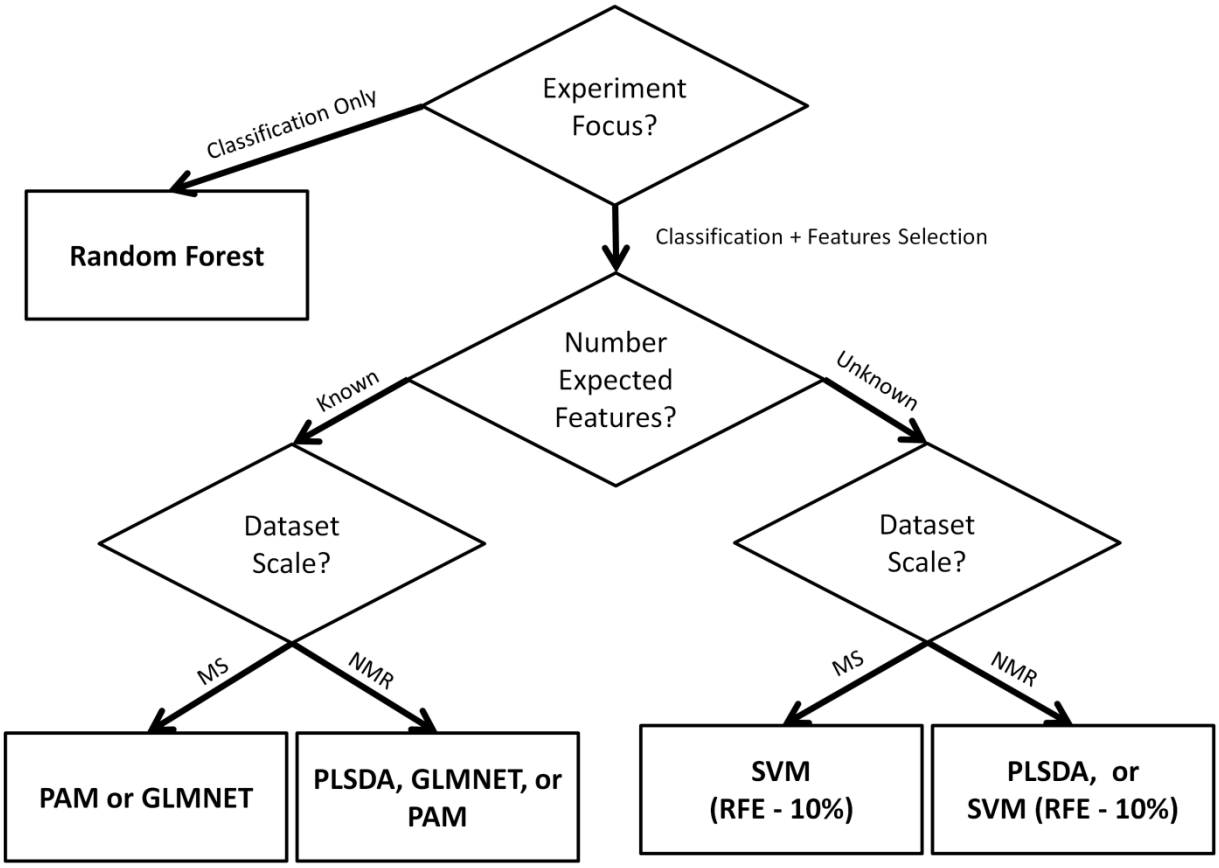

Figure 1. General guideline for which algorithms to apply to a given dataset. The reader is still encouraged to compare multiple algorithms and employ ensemble analysis given the variability in biological datasets.

CONCLUSION

Within this paper, we used simulated datasets to demonstrate that algorithm performance varies depending on the investigator's approach to a dataset (i.e. defined number or unknown), number of samples relative to features, the number of groups, and/or if ensemble methods are applied. We also applied this comparative approach to three published, typically designed and freely accessible datasets including two NMR and one MS datasets. Following tuning and cross-validation, the optimal algorithms were used to compare selected features to the respective studies. Results proved very consistent identifying the most discriminating features; however, additional features were

also identified that demonstrate the variation in applied methods and complexity of these large datasets.

This comparative analysis provides a means to objectively choose a particular algorithm in addition to stability metrics to provide the highest confidence in potentially identified biomarkers and direct more focused independent validation. From these results the choice of the 'best' algorithm appears dependent upon the goals of the experiment in addition to the structure of the dataset. If classification is the primary goal, Random Forest is an excellent option. If feature selection is also important there a few options the must be considered such as the number of expected discriminatory features and the datasets scale (i.e. MS or NMR). The diagram in Figure 1 is meant to be a general guide; however, the results herein strongly suggest using multiple algorithms and comparing performance on each unique dataset as well as ensemble methods which have been made far more accessible with the *OmicsMarkeR* package.

In addition, the package contains two permutation functions (*perm.class* and *perm.features*) for assessing model performance and identified features for further evaluation. Current areas of improvement include the addition of further algorithms, improving memory efficiency and easy to access graphics (e.g. scores plots, variable importance plots, etc.). The R package *OmicsMarkeR* for this analysis is publically accessible from the bioconductor platform (www.bioconductor.org).

CHAPTER 7

Fed State Alters the Metabolomic Response to Hemorrhagic Shock and Resuscitation in Porcine Liver

ABSTRACT

Hemorrhagic shock with injury results in alterations of the metabolic state of an organism, which contribute to organ dysfunction and death. Previous investigations have explored the effects of carbohydrate prefeed in murine models but few in clinically relevant large animal models. We performed carbohydrate prefeed in pigs undergoing simulated polytrauma and hemorrhagic shock with resuscitation to determine if carbohydrate prefeeding if the metabolic response to shock is dependent on fed state. Sixty four (64) Yorkshire pigs were divided into two experimental groups: fasted (32) and pre-fed (32). Experimental animals were subjected to a standardized hemorrhagic shock protocol, including pulmonary contusion and liver crush injury. To determine molecular alterations in response to trauma as a result of pre-feeding, liver biopsies were obtained at set timepoints throughout the procedure. Fifty-one (51) metabolites were profiled for each sample via proton nuclear magnetic resonance spectroscopy (^{1}H-NMR). Partial-Least Squared Discriminant Analysis (PLS-DA) was used to examine clustering of the data with respect to fed state. Cross-validated models separated the fed from fasted animals. Metabolites contributing to the separation have known relationships to alternate carbon energy sources, amino acid metabolism, oxidative stress response, and membrane maintenance. In conclusion, we validated an alternate response to shock and resuscitation, dependent upon fed state, through the use of metabolomics.

INTRODUCTION

Hemorrhagic shock is a leading cause of trauma-related mortality in both civilian and military settings. Efforts over the years have significantly improved survival in the military sector[150]; however, hemorrhagic shock remains the most common cause of preventable injury in both the civilian and military sectors[151–154]. Hemorrhagic shock from traumatic injury results in multiple alterations in the metabolic state of an organism many of which are not fully elucidated.

Hemorrhagic shock results in inadequate tissue perfusion leading to decreased oxygen availability to mitochondria resulting in a switch towards anaerobic. The liver serves an important function as a regulator of metabolism during stressed states. Initially, the shift towards anaerobic metabolism stimulates the liver to increase glycogenolysis and process elevated lactate produced in the peripheral tissues. The liver also provides a major site of detoxification and production of alternate metabolic fuel sources including amino acids and lipids. Effects of different fed states on liver metabolic processes following injury and hemorrhagic shock are not well known.

Metabolomics, a high-throughput profiling of all the metabolites within a sample[12], has the ability to identify multiple metabolites simultaneously. This ability potentially permits the elucidation of altered cellular and metabolic pathways in addition to potentially furthering our understanding of the metabolic response to hemorrhagic shock with respect to metabolic state. Practical applications of metabolomics include classification and diagnosis of several forms of cancer and neurodegenerative disorders[14]. In addition, our laboratory has successfully applied metabolomics tools to a clinically

relevant porcine model of hemorrhagic shock in order to identify biomedical pathways and to associate metabolic changes with phase of care[42,46,112].

Of interest herein is the impact of fed state upon the response and recovery from hemorrhagic shock and injury. Prior work from our group reported a significant increase in lung injury in animals receiving a carbohydrate prefeeding immediately prior to injury compared to fasted animals (44% vs. 15%, $p = 0.03$)[155]. This difference in outcome between fed and fasted animals led us to ask the question regarding possible drivers of difference in response. One of the potential mechanisms possible is a difference in metabolism driven by carbohydrate prefeeding. The objective of this study was to determine if the liver responds differently in our clinically relevant model of polytrauma and hemorrhagic shock with respect to fed state with the hypothesis that there would be quantifiable differences in liver metabolites reflecting an altered metabolic response to shock and resuscitation depending upon fed state. The liver is an important regulator during stress and serves as a primary site of gluconeogenesis to meet peripheral glucose requirements thereby making it an ideal location to evaluate alterations in metabolic response. To test this hypothesis, we used our well-established model of hemorrhagic shock and polytrauma comparing the effect of providing a carbohydrate prefeed (CPF) versus a fasted (FS) diet prior to insult. Proton (^{1}H) NMR spectroscopy was used to determine concentrations of metabolites in liver biopsies taken at defined timepoints throughout shock and resuscitation with the goal to identify altered metabolic responses in prefed versus fasted animals so that we may understand potential contributing factors of metabolic state to outcome.

METHODS

Animals

Male Yorkshire-Landrace pigs (15-20 kg) were purchased from Manthei Hog Farm, LLC (Elk River, MN) and housed in Research Animal Resources (RAR) at the University of Minnesota. All studies were approved by the Institutional Animal Care and Use Committee. Pigs were fasted overnight prior to surgery, but were allowed water *ad libitum*.

Animal Preparation and Hemorrhagic Shock Protocol

Sixty four (64) juvenile, male Yorkshire pigs were used in this study. Results from this group of animals have been previously published[42,46,155]. All animals were fasted overnight. Two experimental groups were utilized: Carbohydrate Prefed (CPF, n = 32) and Fasted (FS, n = 32). CPF animals were given 7cc/kg bolus of Karo Syrup® (mixture of sugars including ~ 15% glucose, maltose, fructose and sucrose) diluted with water 1 hour prior to induction of anesthesia. The full experimental polytrauma and shock protocols have been described in detail previously[46,112] and in Chapter 2. Briefly, animals were instrumented and splenectomized. Polytrauma was induced by a captive bolt device to create a blunt percussive injury to the chest and a liver crush injury using a Holcomb clamp technique[20]. Hemorrhagic shock was then induced by withdrawal of blood from the inferior vena cava until a systolic pressure of 45 to 55mmHg was reached for 45 minutes (S45) to simulate delay prior to medical attention. Typically, this resulted in withdrawal of approximately 40% of the pig's blood volume. Blood was placed in an acid-citrate-dextrose bag for later use. Following the shock period, animals received lactated Ringer's fluid given as 20 cc/kg intravenous (IV) boluses to maintain a systolic blood pressure greater than 80 mmHg for one hour of limited resuscitation to simulate

transportation to a medical center; then underwent full resuscitation protocol for the following 24 hours (Appendix A). This resuscitation included fluid, shed blood, and ventilator support in a protocolled fashion. After the resuscitation period, animals were extubated and sent to recovery and subsequently euthanized.

At several time points throughout the experiment, liver biopsies were taken from the periphery of the liver ranging in weight from 0.2 to 0.6 grams. Biopsies were flash frozen in liquid nitrogen and stored at -80°C until preparation for NMR analysis. Biopsies were taken at the following timepoints: baseline after the animal stabilized from instrumentation (B), 45 minutes after hemorrhage (S45), 2, 8, and 20 hours after full resuscitation (FR2, FR8, FR20). Changes induced by hemorrhagic shock polytrauma and 'early' resuscitation were denoted as the difference between S45 and B (S45-B) and between FR2 and S45 (FR2-S45) respectively.

Liver Metabolite Extraction

Stored liver samples were prepared for NMR analysis using a variation of the perchloric acid (PCA) extraction technique[156]. Frozen liver samples were weighed and pulverized into a fine powder using a mortar and pestle in liquid nitrogen, weighed, and kept on ice in eppendorf tubes. Perchloric acid (6%) was added at 5mL/g of tissue and vortexed for 30 seconds. Samples were incubated for 10 minutes and subsequently centrifuged at 12,000g for 10 min at 4°C. The supernatant was collected, pH neutralized with 2M K_2CO_3 to 7.4, incubated on ice for 30 min and centrifuged once more at 12,000g for 10 min at 4°C. The supernatant was collected, frozen and lyophilized in a LABCONCO Freezone 6 Plus freeze-drier (Kansas City, MO).

NMR Spectroscopy

Lyophilized samples were rehydrated with 500µl D_2O and 50µl of internal standard 3mM DSS (Dimethyl-Silapentane-Sulfonate, Sigma-Aldrich, St. Louis, MO). Solution pH was adjusted with DCl and NaOD to 7.4. The final volume was brought to 600µl using D_2O and the sample was transferred into a 5mm tube (Wilmad, Vineland, NJ)

^1H NMR spectra were obtained using a Varian 600 MHz spectrometer with a 5mm HCN triple resonance probe. Spectra were generated from 128 scans with a basic ^1H acquisition protocol consisting of a 45° tip angle, a relaxation delay of 1 sec and an acquisition time of 1.9 sec. All NMR spectra were phase and baseline-corrected and chemical shifts were referenced to the DSS internal standard.

Chenomx software[56] was used to identify and quantify metabolites present in each liver sample. Fine manual phasing and baseline corrections were applied to each spectrum before targeted profiling was performed. Fifty-one (51) metabolites were fit in each liver sample in this study, resulting in a profile containing the concentration of each identified metabolite in millimoles per liter (mM), as determined by comparison to the internal standard.

Statistical Analysis

A multivariate approach was used to analyze each of the timepoints. To determine the response to shock and resuscitation, changes in concentration between baseline and subsequent timepoints were analyzed. All statistical analysis was conducted using the open source R statistical program[157]. For each timepoint/difference the fifty-one profiled metabolites were auto-scaled and mean-centered prior to initial Principal Component Analysis (PCA). Samples that fell outside a 95% Hotelling's ellipse were

considered outliers and removed from further analysis. Datasets, with outliers removed, were subsequently analyzed by Partial Least Squares Discriminant Analysis (PLS-DA), a common discrimination technique utilized in metabolomics[158–160] that has been successfully employed in our previous studies[42,112]. The R packages DiscriMiner[113] and Permute[161] were used collectively to conduct the PLS-DA model, cross-validation, and permutation tests. PLS-DA models were optimized based on the number of misclassifications (NMC) which has been show to be more powerful than other indicators, such as Q^2, at detecting differences between groups[162]. Cross-validation was conducted via cross-model validation, i.e. nested-CV, 2CV,[163] wherein the dataset is randomly split into training and testing datasets (75%, 25%). PLS-DA models were then generated from the training dataset and a leave-10-out internal cross-validation. Optimized models were used to predict the testing set (outer cross-validation). Prediction accuracy is calculated from the resulting confusion matrix. accuracy $\geq$ 85% and R^2 $\geq$0.500 were considered potential models. Model quality assessment was assessed by random permutation of group class with 1000 iterations where a low permutation p-value (>0.05) indicated a strong model. Metabolites were subsequently ranked according to their respective variable importance of projection (VIP) score. The highest VIP scores correspond to which metabolites contribute most to discrimination between groups. The top 10 metabolites represent the primary drivers of the calculated discrimination.

RESULTS

Partial Least Squares Discriminant Analysis (PLS-DA)

PLS-DA allowed discrimination between prefed and fasted state in this hemorrhagic shock model with changes most notable between the fed and fasted state in

the period of time between end of shock (S45) and 2 hours after full resuscitation (FR2). The final model statistics are reported in Table 1. The majority of models reported accuracy $\geq$ 90% and $R^2 \geq 0.500$ (except FR8-FR2); however, not all models passed the permutation tests (Table 1). The model comparing metabolite changes between baseline (B) and S45 was nearly significant (p=0.06) suggesting an alternate response to shock dependent upon fed state. The model evaluating baseline metabolites provided high accuracy (90%) and R^2 (0.586) but demonstrated a non-significant permutation (p = 0.37). The scoreplots generated from the S45-B and FR2-S45 PLSDA models provide clear separation between the groups (Figure 1).

Important Metabolites

The metabolites identified, by VIP scores, during the response to shock (S45-B) primarily reflected differing energy sources such as glycolysis in CPF animals (carbohydrate sugars, lactate) and amino acid metabolism in FS animals (aspartate, asparagine, 3-Hydroxyisovalerate). In addition, some unexpected differences in elevated asparagine and adenosine were observed in CPF animals (Table 2). VIP metabolites identified during the response to resuscitation (FR2-S45) continued to be associated with energy source differences (e.g. branch chain amino acid metabolism). Indications of membrane repair (sn-Glycero-3-phosphocholine) and free radical scavenging (Benzoate) are also reported (Table 3). Lineplots of the highest scoring metabolites are also provided in Figure 2.

Given the high accuracy of each remaining model, selection of important variables was also performed via ranking metabolites by the VIP statistic (Appendix C -

Tables 1-2); however, strong conclusions could not be drawn from models not passing the permutation test.

DISCUSSION

In this article, we identify multiple differences in metabolism during shock and resuscitation associated with pre-injury fed state. This discrimination was most notable during the initial response to shock and resuscitation with subsequent loss of discrimination over time suggesting the larger metabolic effect of shock and injury overcome the smaller metabolic effect of fed state.

Response to Shock (S45-B)

As expected, with a carbohydrate diet, immediately preceeding injury, liver tissue demonstrated a predominant glycolytic metabolism during the response to shock (S45-B) with active glucose, maltose and sucrose utilization followed by a dramatic increase in lactate levels. Elevated lactate levels have been previously shown to be associated with poor outcome[164]. The changes we observed in glucose metabolism compare to recently described measures of intracellular flux analysis of livers under fed and fasted states which reflect depleted glycogen stores in fasted livers[165,166]. These reports suggest that maintaining glucose stores would be beneficial; however, these experiments consisted of animals undergoing burn injury and provided no survival measures. Our data suggests that depleted glucose stores as a result of fasting may me protective following polytrauma and hemorrhagic shock.

Aspartate also increases more rapidly in livers of CPF animals suggesting potentially impaired transport of the Malate-Aspartate shuttle leading to elevated aspartate levels and subsequent breakdown in the Urea cycle. The Malate-Aspartate

shuttle is required for oxidative phosphorylation and therefore impairment may indicate a more severely stressed oxidative state. The urea cycle enzymes may not be functioning as rapidly as in FS animals and this may be a consequence of the carbohydrate metabolic state, leading to a slower conversion to arginine leading to lower arginine and an accumulation of aspartate in CPF animals. Aspartate can also be converted to asparagine, which also increased in CPF, and either recycled back to aspartate or excreted.

VIP metabolites adenosine and arginine have been reported as protective prior to ischemia-reperfusion associated with Nitric Oxide (NO) metabolism[167,168]. Adenosine is a potent vasodilator released as a response to ischemia-reperfusion[169]. In addition, adenosine is the product of AMP degradation and is characteristic of an imbalance in the tissue oxygen supply/demand ratio. Elevated levels of adenosine in CPF suggest a greater imbalance of oxygen. Despite having elevated adenosine levels mean arterial pressure (MAP) is higher in CPF animals. This may be explained as baseline levels of adenosine are essentially identical between FS and CPF animals whereas levels of arginine, another important vasodilator, are much higher in FS at S45 (CPF = 1.23mM, FS = 0.82mM, p-value = 9.6×10^{-3}). Prior literature reports that arginine is produced within the urea cycle, which is potentially upregulated by glucagon to handle increased proteolysis[170]. Our observations of elevated baseline BCAAs and continued decrease of leucine intermediate 3-Hydroxyisovalerate in FS animals during the response to shock adds support to this interpretation.

Xanthine and glutathione were also identified as important metabolites. Xanthine is the product of xanthine oxidoreductase and generates Reactive Oxygen Species (ROS). Initially xanthine levels were higher in FS animals, suggesting increased oxidative stress,

but decreased to levels similar to those in CPF animals during the response to shock. The increasing xanthine levels in CPF animals is consistent with previous *in vitro* evidence idnetifying increased xanthine oxidase activity following prefeeding and ischemia but not following fasting[171]. More recently, however, it has been suggested that ROS production by xanthine oxidase has limited importance in practice as the conversion of xanthine dehydrogenase to xanthine oxidase requires long ischemic periods that are rare in the clinical setting[172]. This suggests that initial conditions (e.g. fed state) potentially precondition the xanthine oxidase activity prior ischemia. Glutathione, an important antioxidant, also increases within CPF animals to modulate ROS. Measurements of ROS species between fed states are needed to determine possible differences in oxidative stress.

Response to Resuscitation (FR2-S45)

During the response to resuscitation (FR2-S45) we found that fumarate declines sharply in the CPF animals. This reduction is potentially a result of succinate dehydrogenase (i.e. Complex II) of the electron transport chain reducing fumarate to succinate during a period of reduced PaO2[173]. This mechanism is supported by our data as succinate levels also increase (+0.05 ± 0.02mM) during the response to resuscitation. Curiously, lysine levels also increase in CPF but decrease in FS animals. One possible explanation is the conversion of lysine to glutamine; an alternative substrate for the Krebs Cycle. Our data reports elevated glutamine levels in CPF animals at FR2 (CPF = 1.44mM, FS = 1.27mM, p-value = 0.09) which inhibit the conversion of glutamine synthase, the final step between lysine and glutamine, resulting in elevated lysine levels[174].

Metabolism in FS animals continues to reflect BCAA degradation, in which valine, leucine, and isoleucine levels are all decreasing but remain higher indicating continued protein breakdown. However, increasing levels of 3-Hydroxyisovalerate (i.e. β-Hydroxy-β-methylbutryate, HMB) in FS animals suggests impaired complete leucine degradation. Prior research on muscle has reported 3-Hydroxyisovalerate as a substrate for HMG-CoA reductase, an important enzyme for cholesterol synthesis to stabilize membrane integrity[175]. To our knowledge there is no literature exploring 3-Hydroxyisovalerate and HMG-CoA in the liver. It is also noted that sn-Glycero-3-phosphocholine (i.e. Glycerophosphocholine, GPC) levels decreased more rapidly in FS relative to CPF animals. GPC is indicative of membrane maintenance and repair[46]. This is supported by a slight increase in Choline levels, potentially being quickly incorporated into membranes.

The hypoxic state following hemorrhagic shock and the subsequent reperfusion results in the production of reactive oxygen species (ROS). Interestingly, our data reports higher benzoate concentration, an oxygen radical scavenger[176,177], in FS animals (Table 3). One potential mechanism is that CPF animals experience greater oxidative stress following anaerobic glycolysis, as noted by prior xanthine levels, where increased antioxidants would prove beneficial; however assays of ROS need to be pursued comparing fed states for definitive conclusions. To our knowledge, no previous research has reported benzoate as a radical scavenger following reperfusion. It is possible that the conversion of benzoate to hippurate for excretion is inhibited resulting in the observed increase, however, further research is needed to determine if this is an alternate antioxidant following reperfusion.

Although this study measured a large number of metabolites, this study is limited by the both the extraction method and tissue utilized. The PCA method only extracts water soluble metabolites and there may be other compounds that could reflect other altered metabolic processes. An investigation using alternate extraction techniques would prove valuable. The liver is only one component of the systemic response to hemorrhagic shock and resuscitation and analysis of other compartments such as serum, urine and muscle would be beneficial.

CONCLUSIONS

The liver metabolic response to shock is initially significantly different depending upon fed state. This difference dissipates over time as the effects of injury overwhelm the effects of prefeeding. Our identification of VIP metabolites validated previously expected changes in metabolism between carbohydrate prefed and fasted animals such as alternate carbon energy sources. In addition, metabolites associated with NO production, branched chain amino acid metabolism, oxidative stress, and membrane maintenance were significantly different between carbohydrate prefed and fasted animals during the response to shock and response to resuscitation. This analysis reveals that fed state has defined implications on the response to hemorrhagic shock and injury and thereby provides some clues about how future therapeutics should be modified. Exploring other pre-feeding methods, such as protein or lipid, would provide further useful information concerning the response to shock.

MODEL	R2	NMC ± SD	ACCURACY (% correct classified	PERMUTATION (p-value)
B	.586	2.8 ± 1.5	95	.37
S45-B	.614	2.8 ± 1.3	95	.06
FR2-S45	.727	2.0 ± 1.1	95	.04
FR8-FR2	.495	4.1 ± 1.2	90	.64
FR20-FR8	.540	4.0 ± 1.3	90	.60

Table 1: Results of PLS-DA model generation for each timepoint/difference.
Abbreviations: NMC (Number of Misclassified), SD (Standard Deviation)

Metabolite	VIP Score	Group	S45 Mean±SE	Difference Mean±SE	Mann-Whitney U
Glucose	3.301	CPF	25.69±2.68	191.18±27.83	$p=5.9x10^{-9}$
		FS	5.62±0.99	-2.27±6.85	
Lactate	2.237	CPF	5.58±0.45	39.81±4.65	$p=8.1x10^{-4}$
		FS	3.42±0.28	18.14±3.44	
Aspartate	2.042	CPF	1.09±0.07	3.02±0.79	$p=4.3x10^{-3}$
		FS	0.82±0.05	-0.21±0.56	
Maltose	1.993	CPF	1.30±0.28	-12.31±3.05	$p=2.1x10^{-5}$
		FS	0.25±0.04	-2.08±0.80	
Sucrose	1.967	CPF	0.76±0.16	-7.96±1.86	$p=2.8x10^{-3}$
		FS	0.14±0.02	-1.72±0.61	
Glutathione	1.412	CPF	1.74±0.17	3.47±1.71	$p=0.168$
		FS	1.38±0.12	-1.68±1.61	
Xanthine	1.391	CPF	0.05±0.01	0.08±0.06	$p=0.109$
		FS	$0.06±4x10^{-3}$	-0.12±0.07	
Asparagine	1.375	CPF	0.35±0.06	1.01±0.47	$p=0.131$
		FS	0.27±0.03	-0.34±0.43	
Adenosine	1.296	CPF	0.28±0.04	1.11±0.31	$p=0.032$
		FS	0.21±0.02	0.34±0.23	
3-Hydroxyisovalerate	1.117	CPF	0.07±0.01	0.08±0.04	$p=0.124$
		FS	0.07±0.01	-0.08±0.07	

Table 2: Mann-Whitney U results of the top 10 VIP (Variable Importance of Projection) metabolites to discriminate Fasted and Carbohydrate Prefed (CPF) samples during the response to shock (S45-B) via PLS-DA. Mean concentrations reported as µM ± SE (standard error).

Metabolite	VIP Score	Group	FR2 Mean±SE	Mean±SE	Mann-Whitney U
Glucose	3.311	CPF	7.40±0.66	-148.38±24.59	p=9.5x10^{-6}
		FS	5.38±0.31	-4.83±11.85	
3-Hydroxyisovalerate	2.083	CPF	0.07±0.01	-0.02±0.03	p=2.7x10^{-3}
		FS	0.08±0.01	0.13±0.04	
Valine	1.659	CPF	0.23±0.02	-1.15±0.31	p=0.057
		FS	0.30±0.02	-2.40±0.43	
Fumarate	1.658	CPF	0.05±0.01	-0.26±0.04	p=0.011
		FS	0.05±3x10^{-3}	-0.08±0.08	
Benzoate	1.639	CPF	0.15±0.03	-0.03±0.24	p=8.3x10^{-3}
		FS	0.22±0.03	0.81±0.30	
Leucine	1.590	CPF	0.25±0.03	-1.43±0.32	p=0.094
		FS	0.28±0.02	-2.50±0.36	
Choline	1.367	CPF	0.63±0.16	0.66±1.19	p=0.068
		FS	0.47±0.06	-1.94±0.87	
Isoleucine	1.364	CPF	0.14±0.02	-0.63±0.22	p=0.146
		FS	0.18±0.01	-1.19±0.21	
Lysine	1.356	CPF	1.08±0.12	1.12±1.05	p=0.099
		FS	1.03±0.12	-1.60±1.07	
Sn-Glycero-3-phosphocholine	1.390	CPF	5.20±0.48	-26.06±4.84	p=0.080
		FS	3.46±0.42	-39.36±5.37	

Table 3: Mann-Whitney U results of the top 10 VIP (Variable Importance of Projection) metabolites to discriminate Fasted and Carbohydrate Prefed (CPF) samples during the response to resuscitation (FR2-S45) via PLS-DA. Mean concentrations reported as µM ± SE (standard error).

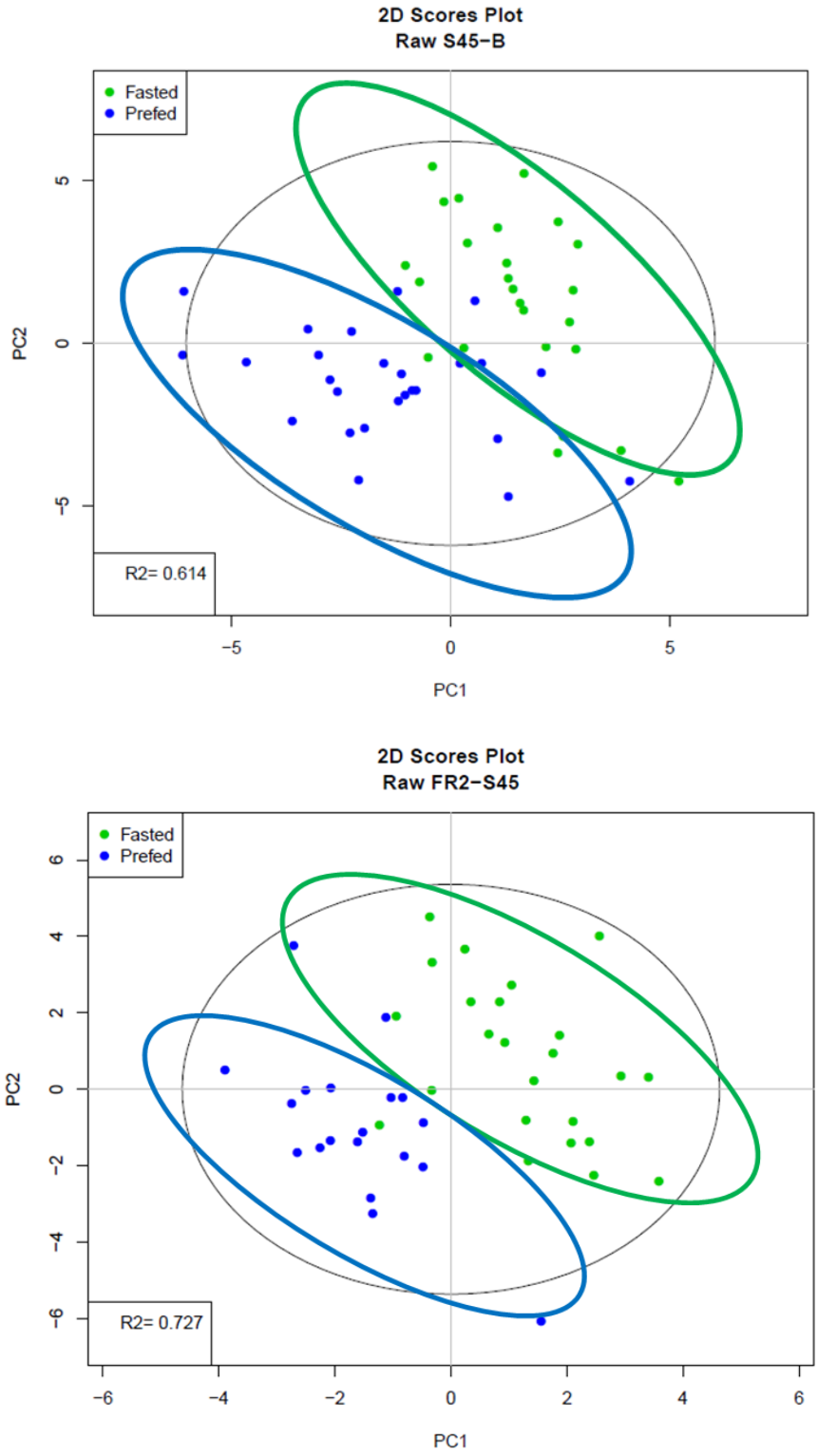

Figure 1 – 2D score plots of response to shock (S45-B) and resuscitation (FR2-S45) PLS-DA models visually representing the discrimination between groups. Green represents Fasted animals and Blue represents Carbohydrate Prefed animals.

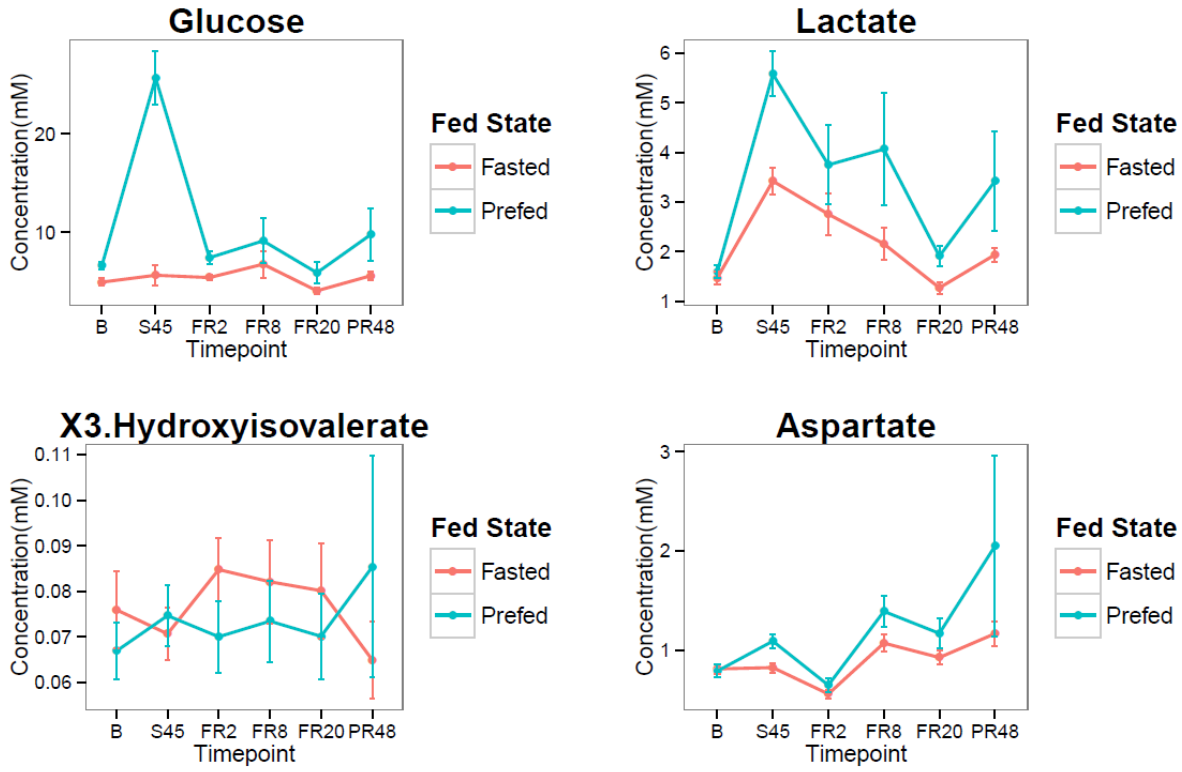

Figure 2 – Lineplots of Glucose, Lactate, 3-Hydroxyisovalerate and Aspartate

CHAPTER 8

Fed State Prior to Hemorrhagic Shock and Polytrauma in a Porcine Model Results in Altered Liver Transcriptomics Response

ABSTRACT

Hemorrhagic shock is a leading cause of trauma-related mortality in both civilian and military settings. Resuscitation often results in reperfusion injury and survivors are susceptible to developing multiple organ failure (MOF). The impact of fed state on the overall response to shock and resuscitation (HS/R) has been explored in some murine models but few clinically relevant large animal models. We have previously used metabolomics to establish that the fed state results in a different metabolic response in the porcine liver following HS/R. In this study, we used our clinically relevant model of hemorrhagic shock and polytrauma and the Illumina HiSeq platform to determine if the liver transcriptomic response is also altered with respect to fed state. Functional analysis of the response to shock and resuscitation confirmed several typical responses including carbohydrate metabolism, cytokine inflammation, decreased cholesterol synthesis, and apoptosis. Our findings also suggest that the fasting state, relative to a carbohydrate prefed state, displays decreased carbohydrate metabolism, increased cytoskeleton reorganization and decreased inflammation in response to hemorrhagic shock and reperfusion. Evidence suggests that this is a consequence of a shrunken, catabolic state of the liver cells which provides an anti-inflammatory condition that partially mitigates hepatocellar damage.

INTRODUCTION

Hemorrhagic shock is a leading cause of trauma-related mortality in both civilian and military settings. Efforts over the years have significantly improved survival in the military sector[150]; however, hemorrhagic shock from traumatic injury results in multiple alterations in the metabolic state of an organism many of which are not fully elucidated. Civilian data estimates between 1400 and 14,000 preventable hemorrhagic trauma deaths occur per year in the United States[151]. Military data has consistently reported hemorrhagic shock as the leading cause of preventable deaths[152–154].

Hemorrhagic shock results in inadequate tissue perfusion leading to decreased oxygen availability to mitochondria. This condition causes a switch towards anaerobic metabolism in addition to a complex inflammatory response. The liver serves an important function as a regulator of metabolism during stressed states. Initially, the shift towards anaerobic metabolism stimulates the liver to increase glycogenolysis and process elevated lactate produced in the peripheral tissues. The liver also provides a major site of detoxification and production of alternate metabolic fuel sources including amino acids and lipids.

Previous research in our laboratory revealed increased acute lung injury and multiple organ failure with a trend toward increased mortality in pigs receiving a carbohydrate prefeed prior to hemorrhagic shock relative to fasted pigs[155]. Our previous metabolomics research has reported an altered response to fed state in the liver and urine as well as phase of care in serum following hemorrhagic shock and resuscitation[42,95,112]. The objective of this study was to determine if this differential response is related to varied genetic expression in the liver with the hypothesis that there would be quantifiable

differences in liver mRNA expression reflecting an altered response to shock and resuscitation with respect to fed state. To test this hypothesis, we used our well-established model of hemorrhagic shock and polytrauma comparing the effect of providing a carbohydrate prefeed (CPF) versus a fasted (FS) diet prior to insult. Extraction of mRNA from liver biopsies and subsequent RNA-Sequencing was used to compare between CPF and FS animals at four timepoints: Baseline (B), and 2, 8 and 20 hours after resuscitation (FR2, FR8, and FR20). Secondly, livers samples were examined over the course of polytrauma, hemorrhagic shock and resuscitation within each group to determine how mRNA expression changed within each group.

MATERIALS AND METHODS

Animal Preparation and Hemorrhagic Shock Protocol

Sixty four (64) juvenile, male Yorkshire pigs were used in this study. All animals were fasted overnight. Two experimental groups were utilized: Carbohydrate Prefed (CPF, n = 32) and Fasted (FS, n = 32). CPF animals were given 7cc/kg bolus of Karo Syrup® (mixture of sugars including ~ 15% glucose, maltose, fructose and sucrose) diluted with water 1 hour prior to induction. The full experimental polytrauma and shock protocols have been described in detail previously[46,112]. Briefly, animals were instrumented and splenectomized. Polytrauma was induced by a captive bolt device to create a blunt percussive injury to the chest and a liver crush injury using a Holcomb clamp technique[20]. Hemorrhagic shock was then induced by withdrawal of blood from the inferior vena cava until a systolic pressure of 45 to 55mmHg was reached for 45 minutes (S45) to simulate time prior to medical help. Typically, this resulted in withdrawal of approximately 40% of the pig's blood volume. Blood was placed in an acid-citrate-

dextrose bag for later use. Following the shock period, animals received lactated Ringer's fluid given as 20 cc/kg intravenous (IV) boluses to maintain a systolic blood pressure greater than 80 mmHg for one hour of limited resuscitation to simulate transportation to a medical center; then underwent full resuscitation protocol for the following 24 hours (Appendix A). This resuscitation included fluid, shed blood, and ventilator support in a protocolled fashion. After the resuscitation period, animals were extubated and sent to recovery and subsequently euthanized.

At several time points throughout the experiment, liver biopsies were taken from the periphery of the liver ranging in weight from 0.2 to 0.6 grams. Biopsies were flash frozen in liquid nitrogen and stored at -80°C until preparation for RNA extraction. Biopsies were taken at the following timepoints: baseline after the animal stabilized from instrumentation (B), 45 minutes after hemorrhage (S45), 2, 8, and 20 hours after full resuscitation (FR2, FR8, FR20). RNA extractions and sequencing were conducted on B, FR2, FR8, and FR20. Analysis was not conducted on the S45 timepoint because of the short time period from baseline to allow RNA changes to take place.

RNA Preparation and Sequencing

RNA Extraction and Quality Assessment. In a previous study, some of the samples from the original 64 animals had been utilized. As such, only those that had all consecutive timepoints were included. Ultimately, only 18 Fasted animals and 5 Prefed animals were analyzed at each timepoint (total n = 92). RNA was purified from liver samples using QIAshredder and Qiagen RNeasy Mini Kits (Qiagen, Chatsworth, CA). Total RNA isolates were quantified using a fluorimetric RiboGreen assay. Total RNA integrity was assessed using capillary electrophoresis, generating an RNA Integrity Number (RIN).

All of the samples were verified as high quality (>1 microgram, RIN = 8+) and thus were converted to Illumina sequencing libraries.

Library Creation. RNA samples were converted to sequencing libraries using Illumina's Truseq RNA Sample Preparation Kit (RS-122-2001). In brief, 1 microgram of total RNA was enriched for mRNA using oligo-dT coated magnetic beads, fragmented, and reverse transcribed in cDNA. The cDNA was fragmented into smaller pieces blunt-ended, and ligated to indexed adaptors and amplified using 15 cycles of PCR. Final library size distribution was validated using capillary electrophoresis and quantified using PicoGreen fluorimetry and qPCR. Libraries were successfully sequenced for all samples.

Cluster generation and sequencing. Truseq libraries were hybridized to a paired-end flow cell and individual fragments were clonally amplified by bridge amplification on the Illumina cBot. Libraries were clustered at a concentration of 12 pM. The flow cell was then loaded on the HiSeq 2000 and sequenced using Illumina's Sequencing by Synthesis (SBS) chemistry. Upon completion of a read, a 7 base pair index was performed for sample identification. Samples were run for 100 cycles with 10 million single reads per sample.

Primary analysis and de-multiplexing. Base call (.bcl) files for each cycle of sequencing were generated by Illumina Real Time Analysis software. The base call files and run folders were then exported to servers maintained at the Minnesota Supercomputing Institute (Minneapolis, MN). Primary analysis and de-muliplexing were performed using Illumina's CASAVA software 1.8.2, resulting in de-multiplexed FASTQ files.

Bioinformatics and Data Analysis

RNA-Seq read sequences produced by the Illumina HiSeq 2000 were aligned with the TopHat software[69] to the NCBI Sscrofa 10.2 reference genome. The BAM files from the TopHat mapping were sorted using SAMtools[178] and raw counts estimated by the Python script HTSeq count (http://www.-huber.embl.de/users/anders/HTSeq/)using the NCBI Sscrofa 10.2 reference genome. Identified contigs were converted to human orthologs with the BiomaRt Bioconductor package[179] which facilitates access to BioMart annotation resources[180]. Gene names used for identification are the official Human Genome Organization (HUGO) Gene Nomenclature Committee (HGNC) designations. The resulting raw counts per gene were used by EdgeR[64] to estimate differential expression (DE). EdgeR (Bioconductor release 3.2.4) uses a pair-wise design to measure differential expression. The analysis is based on a negative binomial model that uses over-dispersion estimates to account for biological variability (i.e., sample to sample differences); this is an alternative to the Poisson estimates of biological variability that are often inappropriate[181]. Genes with less than 10 reads were excluded from the analysis and TMM normalization of the sequenced libraries was performed to remove effects due to differences in library size[182]. The most stringent dispersion method (tag-wise) was used to ensure that differential expression was not due to individual differences. EdgeR generates a log2 fold change for each gene, p values and the Benjamini-Hochberg false discovery rate (FDR) are calculated to statistically test the measured DE. Lastly, genes were filtered to those reporting log2 fold changes ≥ 1 or ≤ -1 (i.e. 2 fold change).

Gene Ontology Analysis

The differentially expressed transcripts calculated between fed states at each time point (e.g. FSvCPF, FR2) and between timepoints within each group (e.g. FR8vB, FS) were

analyzed using the functional annotation tools of DAVID[85,88] and additional literature searches. Lists of DEGs from each comparison were entered into DAVID in addition to the complete list of sequenced genes to serve as background for analysis. DAVID analysis provides both functional annotations of each gene (Functional Annotation Table) and of each list of genes (Functional Annotation Clustering) to provide functional information using gene ontology[83]. The Functional Annotation Clustering tool provides an overall picture of the overrepresented and enriched functions by consolidating redundancies in gene ontology categories. Information on specific genes was obtained using the Functional Annotation Table tool and the scientific literature.

RESULTS

RNA-Sequencing Quality

The RNA-sequencing (RNA-seq) results from the Illumina HiSeq 2000 were output as FASTQ format and mapped to the NCBI Sscrofa 10.2 release of the pig genome (ftp://ftp.ncbi.nih.gov/genbank/genomes/Eukaryotes/vertebrates_mammals/Sus_scrofa/Sscrofa10.2/) using TopHat[69]. One FR8 sample failed to sequence successfully and was omitted from further analysis. For each sample, approximately 11.8 million reads were generated (range, 6.9 to 20.6 million reads). On average 88.8% of the reads generated were mapped uniquely (range, 85.7 – 98.3 percent mapped). Following conversion to human orthologs and Human Gene Nomenclature Committee (HGNC) notation, to facilitate functional analysis, we identified 16,498 genes. The raw data has been made available through the Gene Expression Omnibus database accession number GSE55674.

Differentially Expression

We initially compared gene expression differences between CPF and FS animals at Baseline. Of the 16,498 genes, 68 were identified as differentially expressed genes (DEGs) with 14 upregulated and 54 downregulated in CPF relative to FS. In addition, subsequent analysis comparing each timepoint relative to baseline between CPF and FS identified 13, 29, and 30 DEGs at FR2, FR8 and FR20 versus baseline respectively. At each time point following resuscitation, all DEGs were higher in CPF compared to FS except 2 genes at FR20 versus baseline (CHKA and GCK) (Supplementary File 1).

We then compared gene expression differences between each time point and baseline within each fed state. Within carbohydrate prefed animals 116, 478, and 0 DEGs were identified at FR2, FR8, and FR20 timepoints versus baseline respectively (Supplementary File 2). Observed gene expression changes in fasted animals was far more extensive as 1442, 1460, and 1215 DEGs were identified at FR2, FR8 and FR20 timepoints versus baseline respectively (Supplementary File 3). Table 1 summarizes the DEGs for all comparisons (FDR<0.05).

Functional Analysis – Fed State Expression

Analysis of the differentially expressed transcripts using the Database for Annotation, Visualization and Integrated Discovery (DAVID)[85] and accompanying literature searches provided an overall picture of the overrepresented functions between the two fed states at various time points (Supplementary File 4). As expected, comparisons at baseline provide evidence of upregulation of transcripts important for carbohydrate metabolism in CPF animals. Curiously, the majority of the differences were related to contractile proteins associated with the cytoskeleton. Prior literature supports this observation in that cytoskeletal proteins, such as actin, are elevated in

response to glucagon and cellular shrinkage[183–185]. However, these previous studies report even greater actin mRNA expression in response to insulin, which stimulates cellular swelling, suggesting that elevated carbohydrates more strongly induce cytoskeleton remodeling. In contrast to these studies, which were cell based, our in vivo investigation reports higher actin mRNA in fasted animals at baseline. This trend reverses during FR2, FR8 and FR20 where cytoskeletal proteins in CPF exceed FS animals (Figure 1).

Functional Analysis – Responses to Shock and Resuscitation within each Fed State

Carbohydrate Prefed - FR2 v. Baseline

Analysis of each timepoint relative to baseline within CPF animals revealed several responses that are characteristic of hemorrhagic shock and resuscitation. Analysis of FR2vB identified functional responses including increases in processes associated with cell adhesions, hormone responses and cell membrane processes (Supplementary File 5). Exploration of the genes within these categories revealed activation of gene expression related to evidence of wound healing, and apoptosis.

Carbohydrate Prefed - FR8 v. Baseline

Continued resuscitation in CPF animals continued to identify gene expression characteristic of the response to hemorrhagic shock and resuscitation. Functional clusters identified included altered cytokine production (Figure 3), lipid metabolism (Figure 4), oxidation-reduction processes, hormone responses, and peptidase activity (Figure 5) (Supplementary File 5). Genes associated with cytokine production included heat shock 60kDa protein 1 (HSPD1) and myeloid differentiation primary response 88 (MyD88) which both induce pro-inflammatory cytokines[186–188]. Analysis of specific genes within

categories suggested decreased cholesterol synthesis, decreased fatty acid beta-oxidation and apoptosis. All processes are typical following ischemia-reperfusion[189,190]. No genes were found to be significant when comparing FR20 to baseline, possibly a consequence of the small sample size in the CPF group.

Fasted – FR2 v. Baseline

Functional analysis of resuscitation in FS animals returned 37 clusters of similar processes to CPF animals including cell adhesions, wound healing, cytokine inflammation and apoptosis (Supplementary File 6). In contrast to CPF animals, cytoskeleton reorganization appears to be taking place wherein cytoskeletal genes are decreasing (Figure 1). This finding is concordant with previous reports of cytoskeleton destabilization following reperfusion[191].

Fasted – FR8 v. Baseline

Following 8 hours of resuscitation FS, functional analysis identified 45 clusters that were very similar to CPF animals. The processes include cytokine inflammation, decreased cholesterol synthesis, and apoptosis (Supplementary File 6). In addition, cytoskeleton reorganization continues to take place as noted by decreasing cytoskeletal gene expression including actin and multiple elevated myosin heavy and light chains (MYH1, MYH2, MY3, MY6, MY7, MY13, MYL1, MYL2, MYL3) (Figure 6). In addition, glycolysis appears to be downregulated as several glycolytic genes decreased (HK2, PGAM2, ENO3, GPD2, PDK2, PFKM, and PKLR) (Figure 7).

Fasted – FR20 v. Baseline

Lastly, of the 41 clusters identified by functional analysis following 20 hours of resuscitation primary processes continued to include cytoskeleton reorganization in

addition to ion transport and glucose metabolism processes (Supplementary File 6). Genes associated with glycolytic and ion transport processes continue to both be decreased relative to baseline (Figures 7 & 8). Cholesterol and lipid oxidation are also decreased in overall expression (Figure 9).

DISCUSSION

We report that the primary differences between the liver transcriptomic response to hemorrhagic shock between fasted and carbohydrate prefed pigs during the first 24 hours in a polytrauma model of hemorrhagic shock and resuscitation consist of glucose metabolism and cytoskeletal remodeling. Consistent with known changes associated with ischemia and reperfusion, we also found evidence of cytokine activation, apoptosis, and lipid metabolism ontological categories. Additionally, genes associated with increased pro-inflammatory cytokines were elevated in CPF animals relative to FS animals. Our results suggest the anti-inflammatory state of the FS animals may be a result of the fasting metabolic state.

Gene Ontology Profiles

We present the first, to our knowledge, comprehensive exploration of the pig liver transcriptome following a polytrauma model of hemorrhagic shock and resuscitation via RNA-seq. As expected following a carbohydrate prefeed genes associated with carbohydrate metabolism were upregulated as indicated by elevated Glucokinase (GCK) and glycogen phosphorylase (PYGM) in CPF and FS animals respectively reflecting the alternate metabolic state wherein fasted animals are releasing glucose and CPF animals are breaking down available glucose.

The difference in cytoskeleton related genes in FS animals compared to CPF and over time suggests a link between fed state and metabolic response to hemorrhagic shock and resuscitation. Prior *in vitro* investigations have reported that glucagon, which is elevated during fasting periods, causes cellular shrinkage as a result of altered ion fluxes[192,193]. Conversely, elevated insulin results in cellular swelling. Typically, upon swelling, a cell will trigger volume regulatory K^+ efflux. In contrast, following cellular shrinkage, it becomes important for the cell to retain ions. Decreased ion transport is seen in our fasted animals, including several KCN family members (Figure 8), whereby the cells have previously shrunk in response to fasting and possibly don't require additional ion transporters. Additionally, glucagon stimulates the transcription of actin followed by assembly of actin filaments[183–185]. Contrary to these *in vitro* studies, however, our *in vivo* investigation reports increased actin transcription in fasted animals compared to CPF animals. This cellular shrinkage has been previously described as triggering the 'catabolic state' of a cell resulting in a multitude of altered processes[194]. The presence of this cytoskeleton remodeling is further supported in our study by upregulation of contractile proteins including multiple myosins for intracellular transport. In addition, cellular swelling stabilizes actin filaments[195] which helps explain why cytoskeleton mRNA expression is higher in CPF than FS during resuscitation which usually results in actin depolymerization[196].

Previous studies elucidating the impact of cell swelling during ischemia and reperfusion report that cells naturally become hyperosmotic during ischemia and consequently take up additional fluid[197,198]. This increased swelling and resultant cellular edema may compress capillaries thereby reducing or preventing reperfusion of capillary

beds[199–201]. This process results in zones that prevent reperfusion resulting in the 'no-reflow' phenomenon[202]. It is possible that the fasting state displays a shrunken catabolic state that may provide an initial condition to partially mitigate the subsequent swelling that is an unavoidable consequence of ischemia and reperfusion. The benefit of induced cellular shrinkage on hepatocellular injury has been shown previously whereby hypertonic preconditioning (i.e. induced cellular shrinkage) reduces hepatocellular damage following ischemia reperfusion[203]. Furthermore, the use of hypertonic saline for reperfusion has been shown to decrease susceptibility to sepsis, mitigate inflammation, and reduce apoptosis after hemorrhagic shock[204–206] in addition to being attractive for small-volume resuscitation development[153].

The full molecular mechanisms behind the proposed state are not fully elucidated however several components are understood. One direct result of cellular shrinkage is vasodilation potentially mediating the no-reflow condition. Furthermore, cellular shrinkage is a consequence of hypertonic treatments which have been shown mitigate inflammation following shock[207,208]. The hyperosmotic environment causes cellular shrinkage which activates the tyrosine phosphorylation of Janus kinases. These kinases activate the transcription factor STAT3[209] which ultimately stimulates Interleukin 10 (IL-10) production while simultaneously inhibiting pro-inflammatory cytokines (e.g. IL-1β, TNF-α, and intercellular adhesion molecules) by inhibiting the transcription factor NF-κB[210,211] (Figure 10). Although IL-10 was not individually identified, the anti-inflammatory state of the FS is supported as several pro-inflammatory genes are elevated in CPF animals (Figure 8). Further research on the anti-inflammatory properties of cellular shrinkage is required.

Although this study utilized a larger sample size, our groups were unbalanced (18 versus 5). This difference in group size may have partially biased estimates. However, given the quality of the RNA-seq, consistency across timepoints, and agreement with supporting literature supports our results. Lastly, it is important to recall that the liver is only one component of the overall systemic response to hemorrhagic shock and resuscitation. Analysis of other organs and tissues would be should be pursued.

CONCLUSION

Analysis of the liver transcriptome between carbohydrate prefed and fasted pigs following hemorrhagic shock and resuscitation reveals that the first 24 hours results in different changes to the cytoskeleton structure. Our results suggest this both a metabolic response to decreased carbohydrate substrates and structural modifications in response to cellular shrinkage. Further evidence suggests that this shrunken state provides an anti-inflammatory condition that may decrease hepatocellar damage. More work is required to investigate the potential benefits of alternative diets such as the anti-inflammatory effects of increased intake of omega-3 fatty acids.

Comparison	Differentially Expressed Genes
CPFvFS, B	68
CPFvFS, FR2vB	13
CPFvFS, FR8vB	29
CPFvFS, FR20vB	30
FR2vB, CPF	116
FR8vB, CPF	478
FR20vB, CPF	0
FR2vB, FS	1442
FR8vB, FS	1460
FR20vB	1215

Table 1 – Number of differentially expressed genes between CPF and FS animals and between timepoints within groups. Abbreviations: CPF (Carbohydrate Prefed), FS (Fasted), B (Baseline), FR2, FR8 and FR20 (2, 8 and 20 hours after full resuscitation).

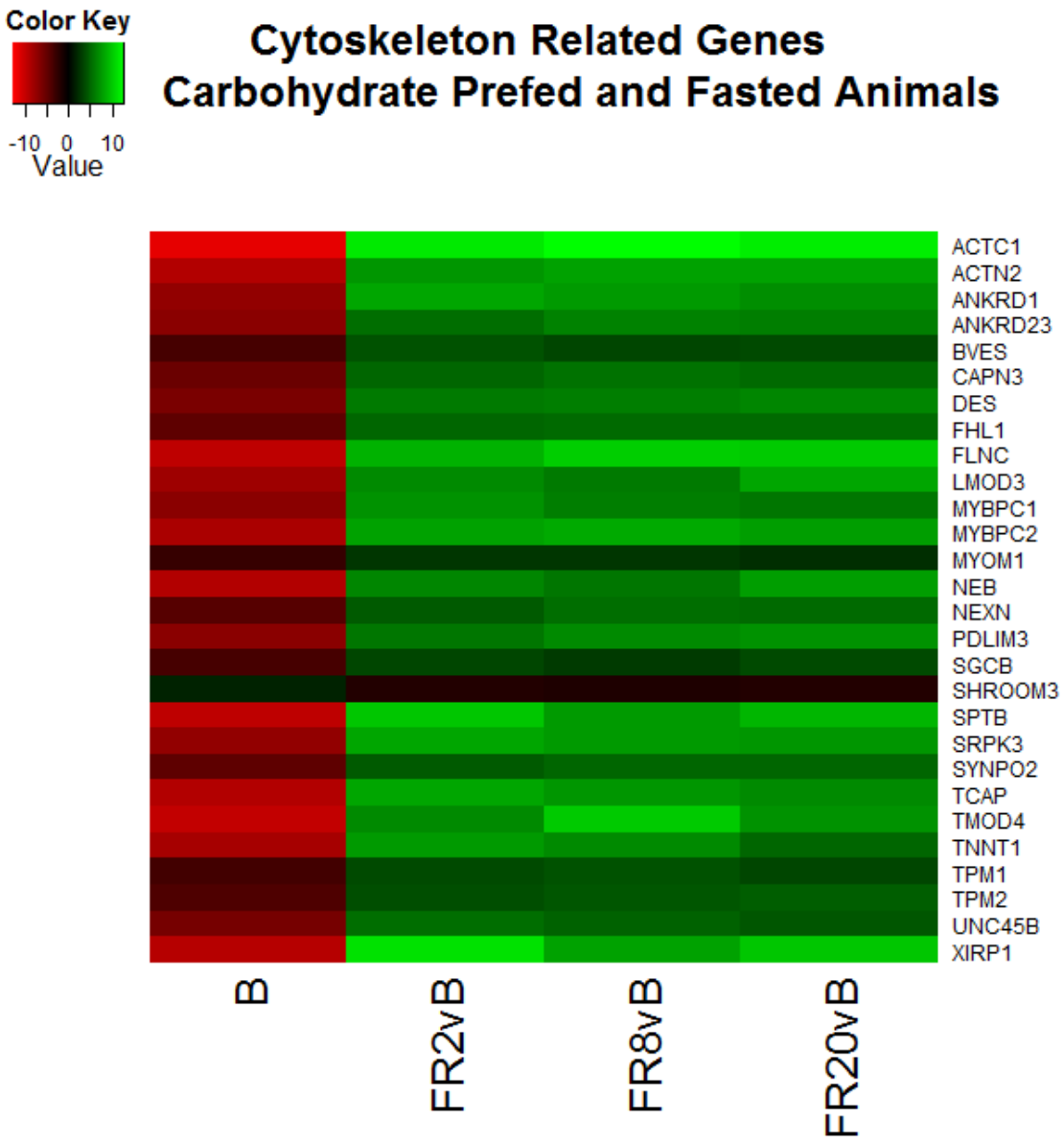

Figure 1 – Heatmap of log2 fold changes of genes associated with cytoskeleton processes between carbohydrate prefed (CPF) and fasted (FS) animals. Rows are differentially expressed genes (DEGs) following RNA sequencing. Columns are denote the respective timepoints Baseline (B), 2 hours full resuscitation change from Baseline (FR2vB), 8 hours full resuscitation change from Baseline (FR8vB), and 20 hours full resuscitation change from Baseline (FR20vB). Green denotes higher concentration or larger changes in mRNA concentration in CPF whereas red denotes the opposite.

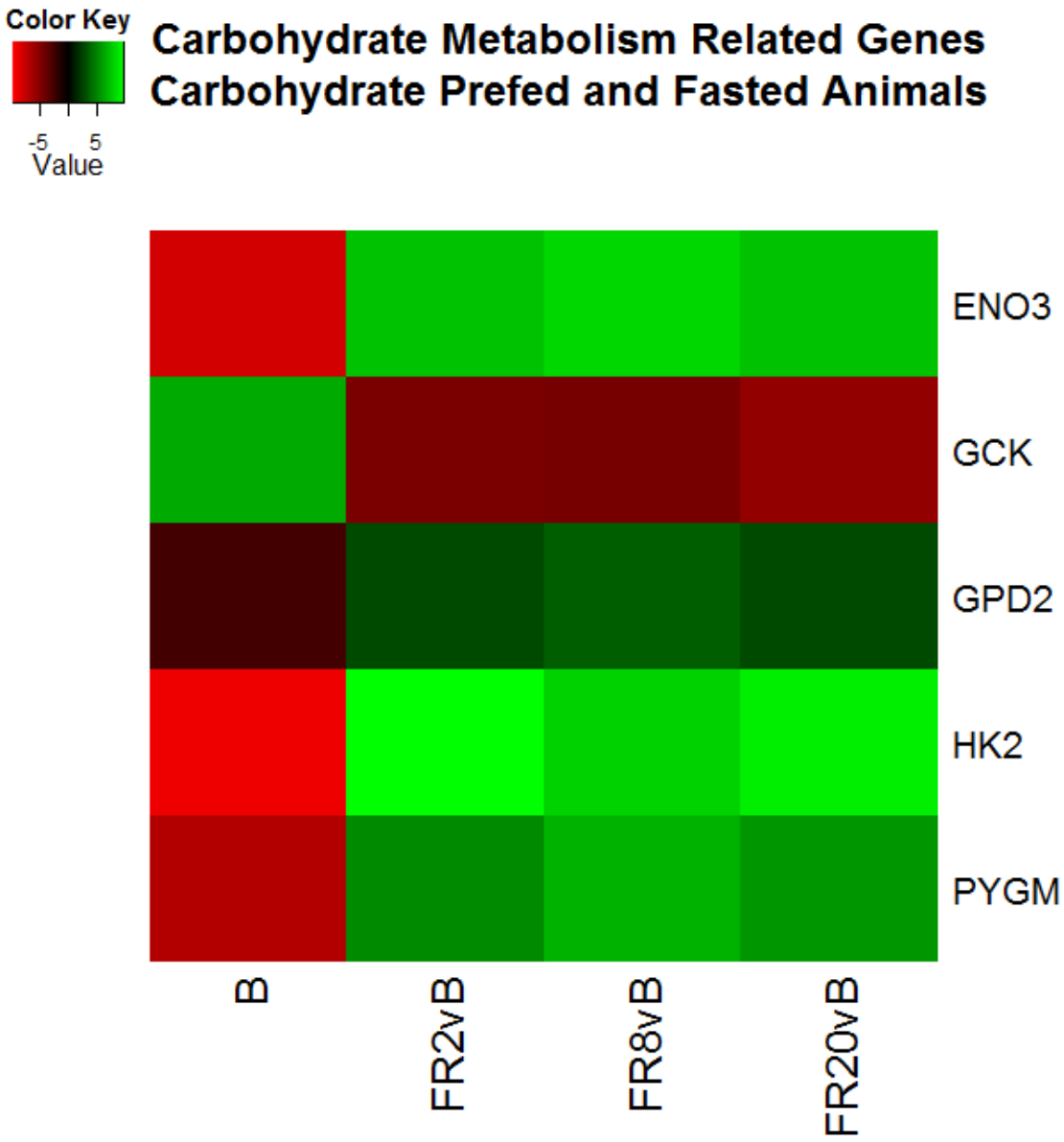

Figure 2 – Heatmap of log2 fold changes of genes associated with carbohydrate metabolism between carbohydrate prefed (CPF) and fasted (FS) animals. Rows are differentially expressed genes (DEGs) following RNA sequencing. Columns are denote the respective timepoints Baseline (B), 2 hours full resuscitation change from Baseline (FR2vB), 8 hours full resuscitation change from Baseline (FR8vB), and 20 hours full resuscitation change from Baseline (FR20vB). Green denotes higher concentration or larger changes in mRNA concentration in CPF whereas red denotes the opposite.

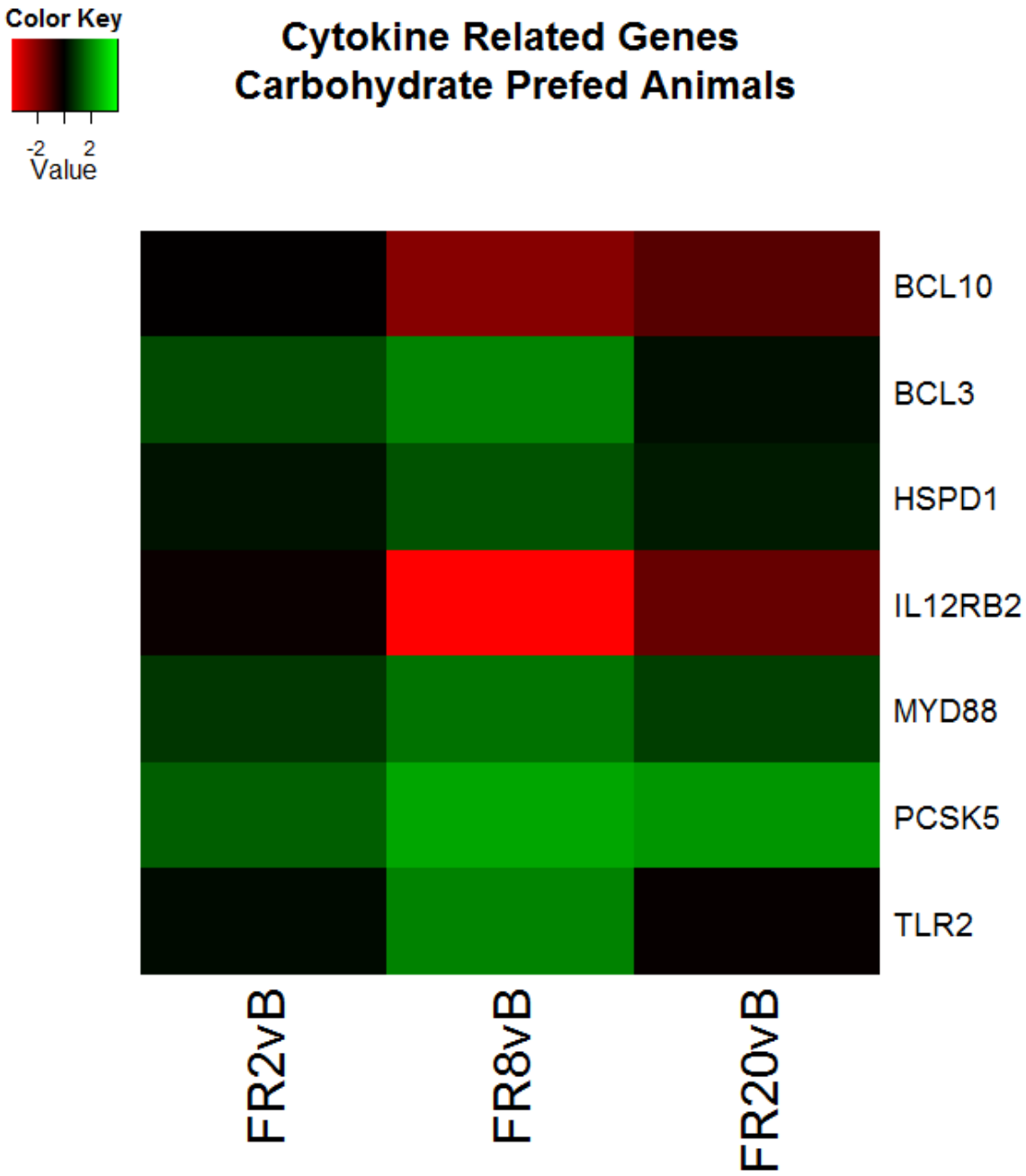

Figure 3 – Heatmap of log2 fold changes of genes associated with cytokine related genes in carbohydrate prefed (CPF) animals at each resuscitation timepoint (2, 8, and 20 hours) relative to Baseline (B). Rows are differentially expressed genes (DEGs) following RNA sequencing. Columns are denote the respective timepoints Baseline (B), 2 hours full resuscitation change from Baseline (FR2vB), 8 hours full resuscitation change from Baseline (FR8vB), and 20 hours full resuscitation change from Baseline (FR20vB). Green denotes increased mRNA expression whereas red denotes the opposite.

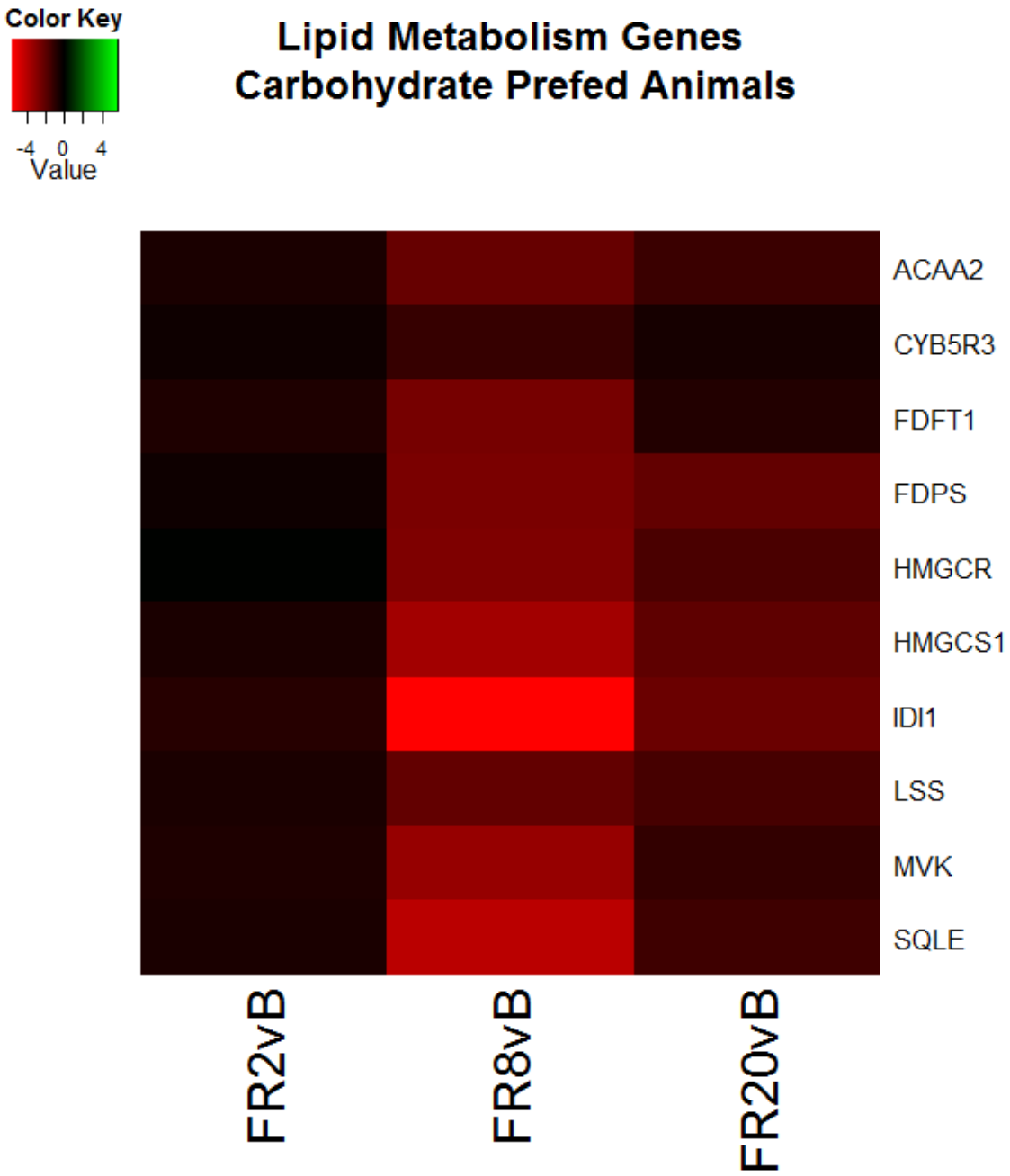

Figure 4 – Heatmap of log2 fold changes of genes associated with lipid metabolism in carbohydrate prefed (CPF) animals at each resuscitation timepoint (2, 8, and 20 hours) relative to Baseline (B). Rows are differentially expressed genes (DEGs) following RNA sequencing. Columns are denote the respective timepoints Baseline (B), 2 hours full resuscitation change from Baseline (FR2vB), 8 hours full resuscitation change from Baseline (FR8vB), and 20 hours full resuscitation change from Baseline (FR20vB). Green denotes increased mRNA expression whereas red denotes the opposite.

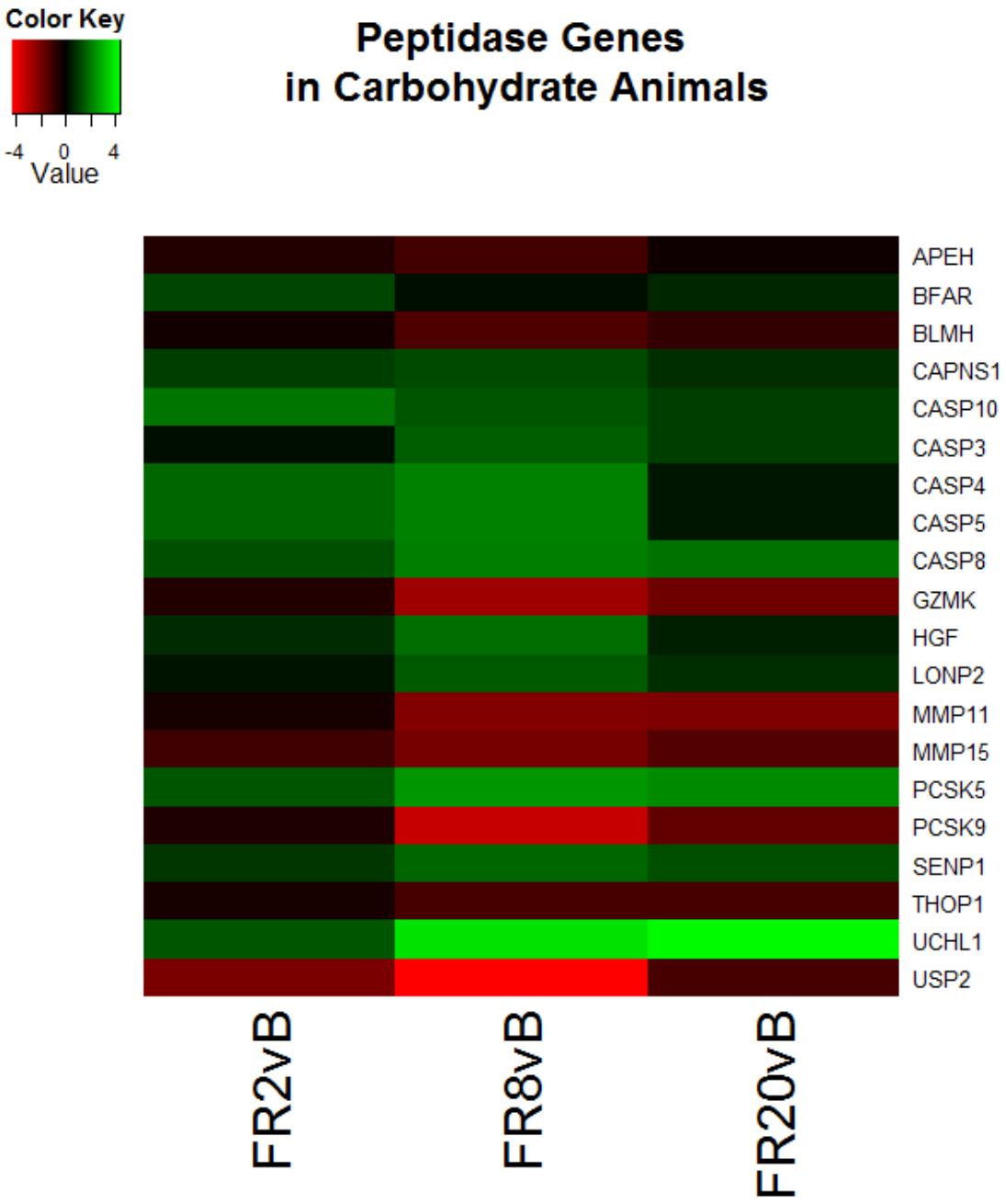

Figure 5 – Heatmap of log2 fold changes of genes associated with peptidase activity in carbohydrate prefed (CPF) animals at each resuscitation timepoint (2, 8, and 20 hours) relative to Baseline (B). Rows are differentially expressed genes (DEGs) following RNA sequencing. Columns are denote the respective timepoints Baseline (B), 2 hours full resuscitation change from Baseline (FR2vB), 8 hours full resuscitation change from Baseline (FR8vB), and 20 hours full resuscitation change from Baseline (FR20vB). Green denotes increased mRNA expression whereas red denotes the opposite.

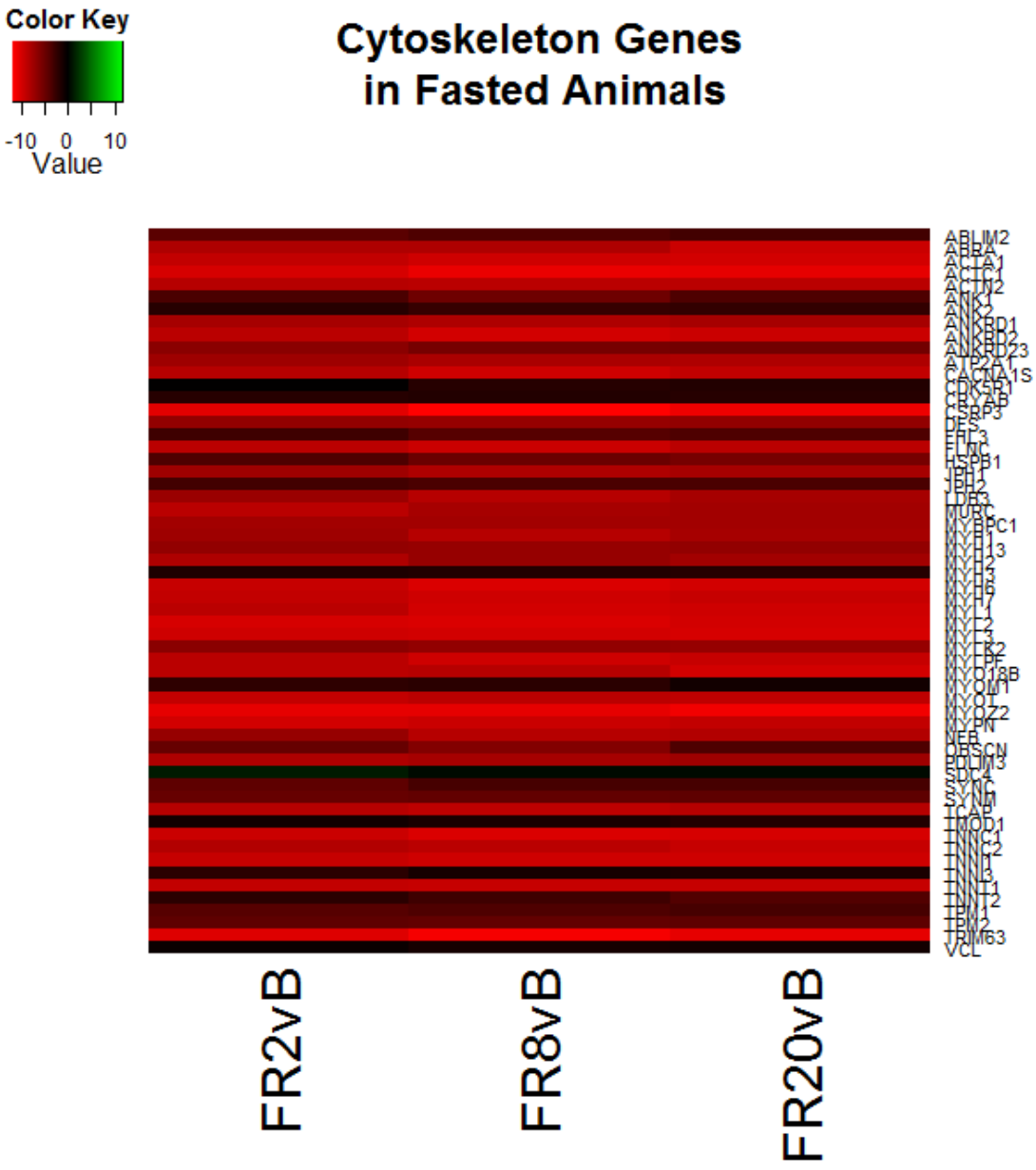

Figure 6 – Heatmap of log2 fold changes of genes associated with cytoskeleton processes in fasted (FS) animals at each resuscitation timepoint (2, 8, and 20 hours) relative to Baseline (B). Rows are differentially expressed genes (DEGs) following RNA sequencing. Columns are denote the respective timepoints Baseline (B), 2 hours full resuscitation change from Baseline (FR2vB), 8 hours full resuscitation change from Baseline (FR8vB), and 20 hours full resuscitation change from Baseline (FR20vB). Green denotes increased mRNA expression whereas red denotes the opposite.

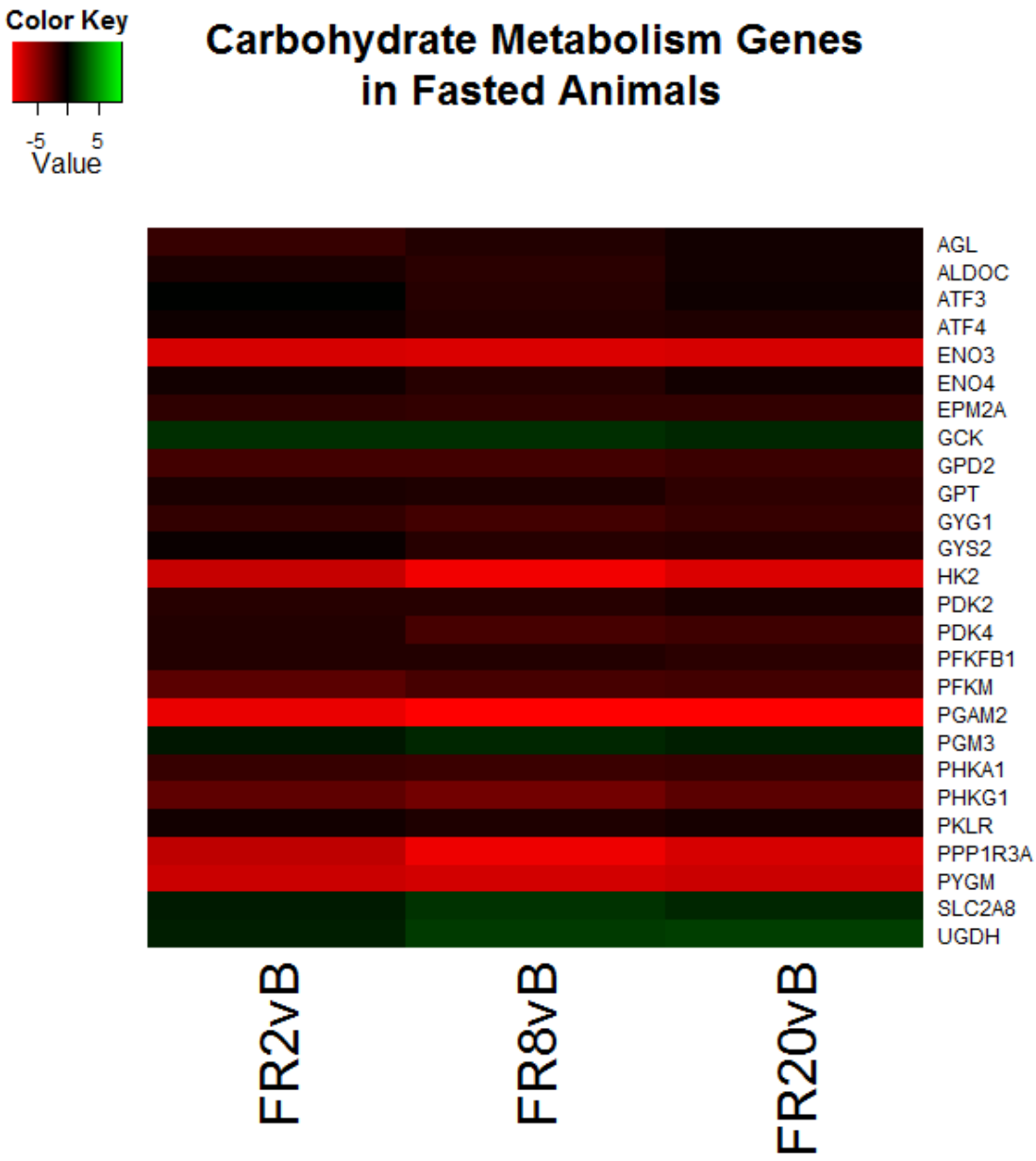

Figure 7 – Heatmap of log2 fold changes of genes associated with carbohydrate metabolism in fasted (FS) animals at each resuscitation timepoint (2, 8, and 20 hours) relative to Baseline (B). Rows are differentially expressed genes (DEGs) following RNA sequencing. Columns are denote the respective timepoints Baseline (B), 2 hours full resuscitation change from Baseline (FR2vB), 8 hours full resuscitation change from Baseline (FR8vB), and 20 hours full resuscitation change from Baseline (FR20vB). Green denotes increased mRNA expression whereas red denotes the opposite.

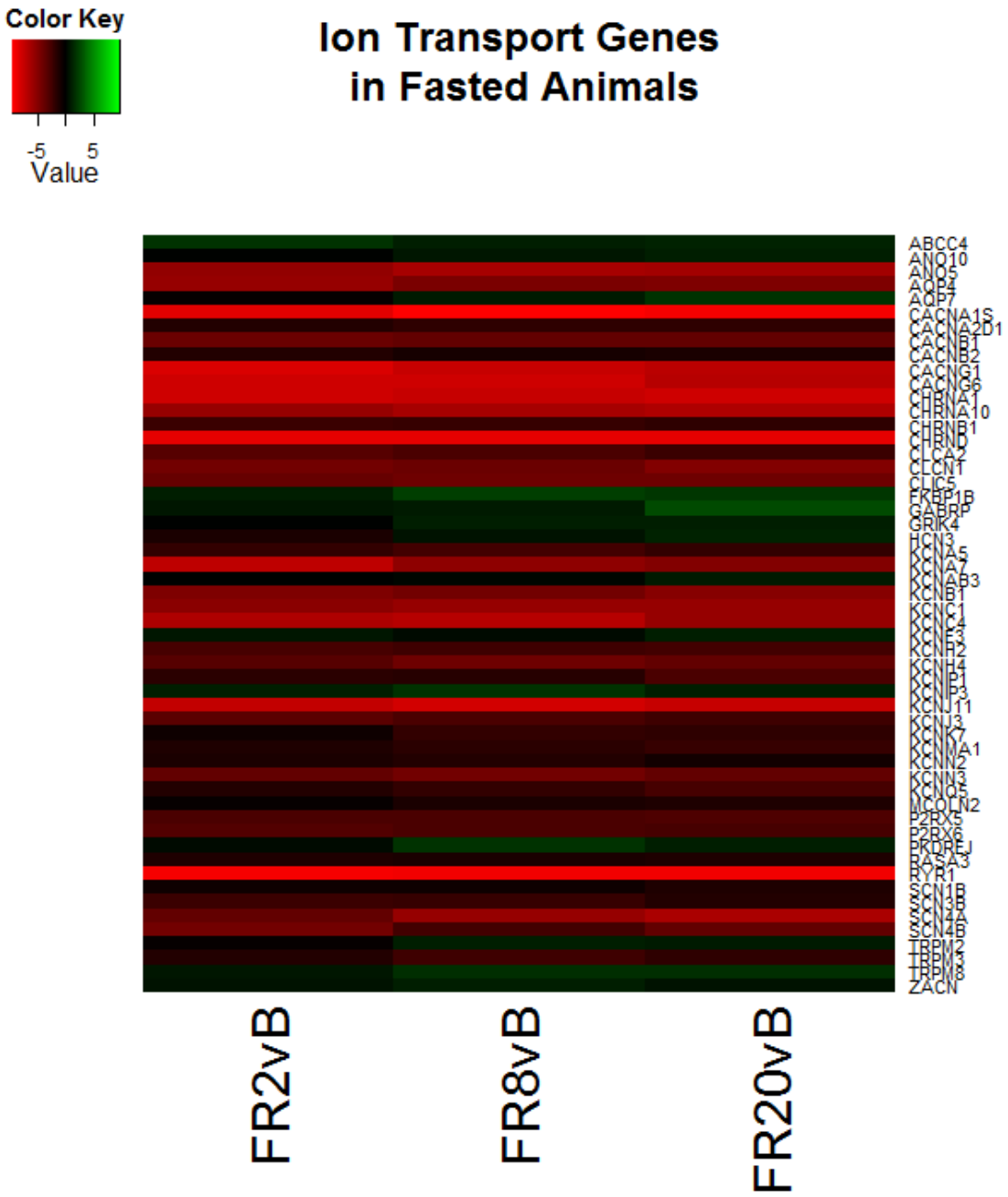

Figure 8 – Heatmap of log2 fold changes of genes associated with ion transport in fasted (FS) animals at each resuscitation timepoint (2, 8, and 20 hours) relative to Baseline (B). Rows are differentially expressed genes (DEGs) following RNA sequencing. Columns are denote the respective timepoints Baseline (B), 2 hours full resuscitation change from Baseline (FR2vB), 8 hours full resuscitation change from Baseline (FR8vB), and 20 hours full resuscitation change from Baseline (FR20vB). Green denotes increased mRNA expression whereas red denotes the opposite.

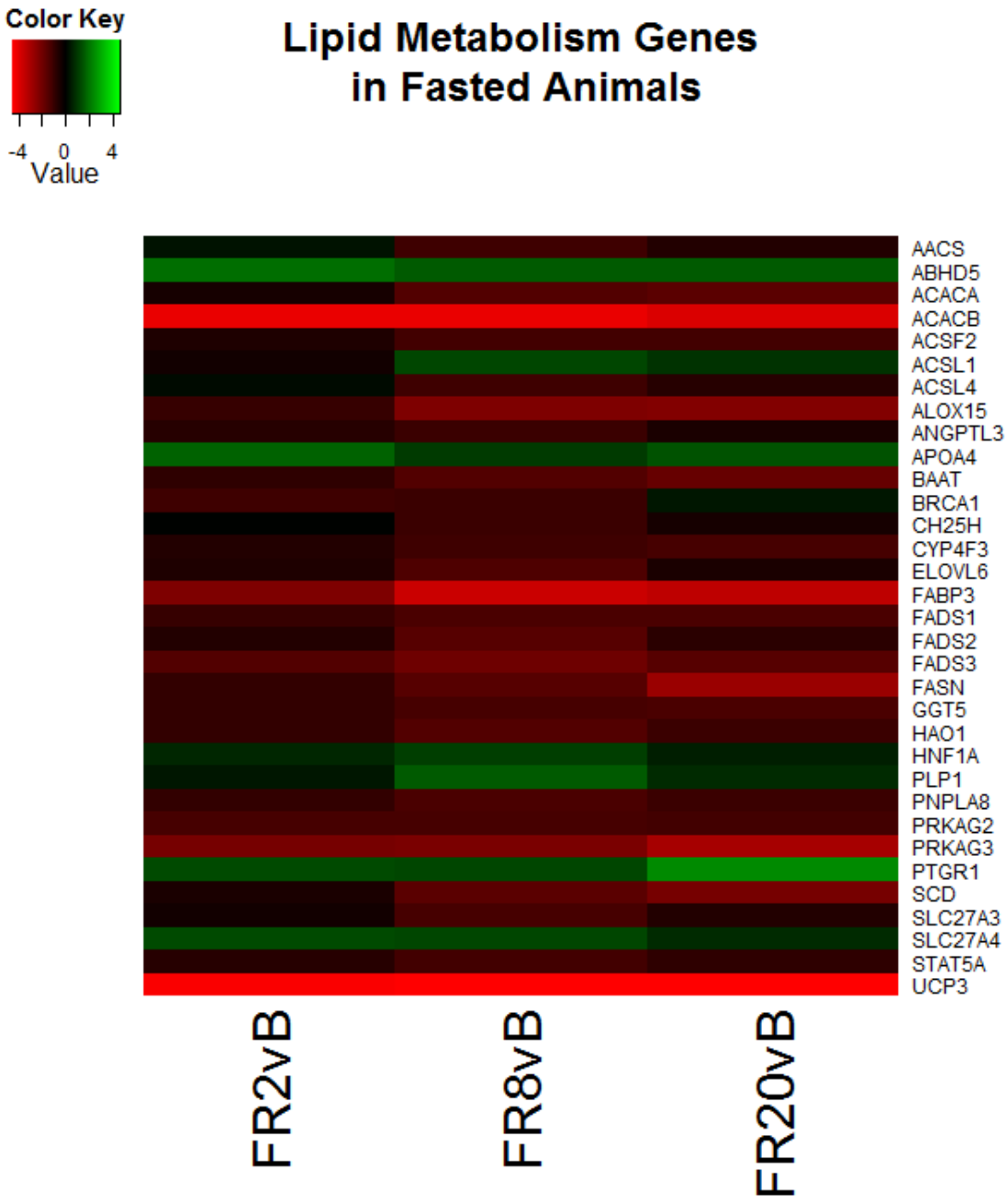

Figure 9 – Heatmap of log2 fold changes of genes associated with carbohydrate metabolism in fasted (FS) animals at each resuscitation timepoint (2, 8, and 20 hours) relative to Baseline (B). Rows are differentially expressed genes (DEGs) following RNA sequencing. Columns are denote the respective timepoints Baseline (B), 2 hours full resuscitation change from Baseline (FR2vB), 8 hours full resuscitation change from Baseline (FR8vB), and 20 hours full resuscitation change from Baseline (FR20vB). Green denotes increased mRNA expression whereas red denotes the opposite.

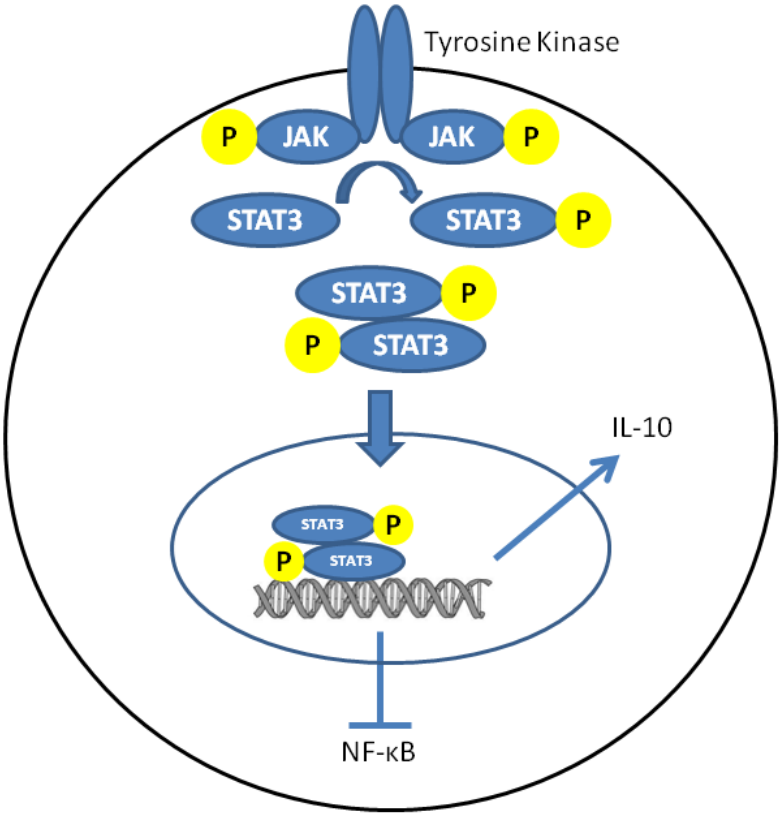

Figure 10 – Diagram of simplified Interleukin-10 (IL-10) pathway. A hyperosmotic condition induces cellular shrinkage and activates tyrosine phosphorylation of Janus kinases (JAK) to subsequently phosphorylate STAT3 and IL-10 production.

CHAPTER 9

Conclusions

Concluding Remarks

Hemorrhagic shock, following trauma, has continued to be a leading cause of death in both military and civilian settings. This public health interest has spurred research over decades leading to improvements in treatment. However, the metabolic mechanisms altered in response to shock and subsequent resuscitation require further understanding to not only treat patients post-injury but potentially also advance pre-conditioning strategies to mitigate potential injury in the military sector. Transcriptomics is an established sub-discipline of systems biology that can profile altered genome level processes and provide insight into overall physiologic responses. Metabolomics is an additional and rapidly growing sub-discipline of systems biology that facilitates the identification of altered pathways and biomarker identification. Together, these two systems biology approaches applied towards hemorrhagic shock facilitate advancements in the understanding of the numerous complex responses to injury. The focus of this dissertation has been to evaluate current approaches towards analyzing metabolomics data (Chapter 6) and analyze the impact of fed state on both the metabolomic (Chapter 7) and transcriptomic (Chapter 8) response to hemorrhagic shock and resuscitation.

Herein, established NMR and RNA-sequencing techniques were applied explore the metabolomic and transcriptomic responses, respectively, within our clinically relevant, porcine model of hemorrhagic shock and resuscitation. The response to hemorrhagic shock and resuscitation is a complex process consisting of multiple systemic changes. Metabolomic and transcriptomic techniques facilitate the analysis of these systemic changes at the small molecule and RNA transcript level.

The first study discussed was designed to evaluate commonly used multivariate analysis techniques applied to metabolomics datasets (Chapter 6). Metabolomics is a

rapidly growing field as the cost decreases and popularity increases. Previous methods have been developed for transcriptomic and proteomic techniques in addition to other multivariate methods that have been applied towards metabolomics. The diversity of methods, in addition to potentially conflicting results of alternate methods, necessitated an analysis that would facilitate the rapid comparison of multiple algorithms. This would provide not only an objective means to identify the optimal algorithm but also facilitate more robust analyses by providing comprehensive results of multiple accepted methods. We therefore developed a new tool to rapidly assess algorithms for the analysis of metabolomics data. The *OmicsMarkeR* program is an open-source package written for the extensive R Statistical Program[157] and is made available through Bioconductor. The results demonstrate how no algorithm is optimal for all problems as overall performance varies between datasets. It is therefore important to compare and utilize multiple methods when exploring high-dimensional datasets such as metabolomics.

The following study explored the impact of fed state upon the liver metabolic response to hemorrhagic shock and resuscitation (Chapter 7). The objective was to determine if a carbohydrate prefeed would alter the liver metabolic processes following shock and if so, what processes were altered. The results reported a clear distinction between fed and fasted animals via PLS-DA analysis. The metabolites that contributed most towards the discrimination were associated with alternate carbon energy sources, amino acid metabolism, oxidative stress, and membrane maintenance. This analysis revealed that carbohydrate prefeeding alters metabolic pathways during the response to shock and resuscitation.

Lastly, we conducted a transcriptomics analysis to evaluate the impact of fed state on gene expression following hemorrhagic shock and resuscitation (Chapter 8). The objective was to determine if upstream metabolic pathways were altered at the gene expression level in response to a carbohydrate fed state. Our findings reported changes in carbohydrate metabolism as expected but also changes in cytoskeleton reorganization and inflammation. Our results and previous evidence in hypertonic resuscitation techniques suggest that the fasted state facilitates a shrunken, catabolic state of the liver cells providing an anti-inflammatory state that partially mitigates hepatocellular damage.

This collective work demonstrates the application of high-throughput, systems level analysis to advance our understanding of the complex responses elicited by hemorrhagic shock and resuscitation. We have successfully determined that fed state is an important consideration in hemorrhagic shock experiments thereby advancing upon a previous study that reported increased lung injury and a trend towards increased mortality in prefed pigs following hemorrhagic shock and resuscitation[155]. These results provide a firm foundation from which future experiments in trauma should be expanded upon to include the fed state of the model organism. Furthermore, the fed state could be readily determined through the early stages of shock via metabolomics and transcriptomics. However, this difference appears to dissipate in later stages likely a consequence of the metabolic effect of shock and injury overwhelming the metabolic effect of fed state. This provides a potential opportunity to specifically treat patients in the early stages of shock that have eaten or not. Transcriptomics results also provide further support for hypertonic resuscitation techniques. These results suggest that by stimulating alternate conditions, it is theoretically possible to convert cells into a catabolic and anti-

inflammatory state similar to ischemia precondition models. Further research is needed to explore and advance fluid resuscitation techniques which should be guided by the inherent cellular processes.

The integration of multiple levels of systems biology continues to be a difficult task. To our knowledge, there is no standard or recommended means to integrate these levels beyond correlations, mapping to pathways, and literature based knowledge of the biological processes. Within our experiments, the limited number of metabolites resolved by NMR analysis renders any pathway analysis software unusable and therefore interpretations must be made with caution. However, there were some consistencies between these two studies. Altered carbohydrate metabolism was readily identified thereby confirming both methods were successful. Additionally, indications of membrane maintenance that follow cytoskeletal remodeling and cellular shrinkage are potentially reflected by 3-Hydroxyisovalerate and sn-Glycero-3-phosphocholine (Chapter 7). The higher resolution of mass spectrometry analysis may provide further insights into the metabolome.

This work has provided insight into the metabolomic and transcriptomic responses to hemorrhagic shock and resuscitation in addition to improving analytical methods. However, each of these components demonstrates the need for further research both within each specialty as well as integration methods. Although we provide metabolomic and transcriptomic data, proteomics also provides invaluable information on the compounds conducting the processes involved. Future work should also continue to develop improved algorithms with standardized means of evaluation. An exploration of the impact of alternate diets is strongly encouraged following the clear discrimination

between carbohydrate prefed and fasted pigs at both the metabolomic and transcriptomic levels. Furthermore, this work suggests a renewed interest in the 'metabolic state' of cells whereby different conditions poise a cell to respond to a stimulus in a different and potentially beneficial manner. Finally, methodology in systems biology integration needs to be more readily accessible. This is rapidly becoming possible as costs continue to drop through advancements in next-generation and high-throughput technologies. The movement beyond the classical biological dogma, wherein each component is considered separately, towards an integrative and holistic perspective will advance our understanding of not only the complex processes following hemorrhagic shock and resuscitation but other systemic pathologies as well.

CHAPTER 10

Bibliography

1. Kauvar, D. S. & Wade, C. E. The epidemiology and modern management of traumatic hemorrhage: US and international perspectives. *Crit. Care* **9,** 1–9 (2005).
2. Kauvar, D. S., Lefering, R. & Wade, C. E. Impact of Hemorrhage on Trauma Outcome: An Overview of Epidemiology, Clinical Presentations, and Therapeutic Considerations. *J. Trauma Inj. Infect. Crit. Care* **60,** S3–S11 (2006).
3. Kelly, J. F. *et al.* Injury Severity and Causes of Death From Operation Iraqi Freedom and Operation Enduring Freedom: 2003–2004 Versus 2006. *J. Trauma Inj. Infect. Crit. Care* **64,** S21–S27 (2008).
4. Pfeifer, R., Tarkin, I. S., Rocos, B. & Pape, H.-C. Patterns of mortality and causes of death in polytrauma patients—Has anything changed? *Injury* **40,** 907–911 (2009).
5. Dutton, R. P. *et al.* Trauma Mortality in Mature Trauma Systems: Are We Doing Better? An Analysis of Trauma Mortality Patterns, 1997–2008. *J. Trauma Inj. Infect. Crit. Care* **69,** 620–626 (2010).
6. Kvarstein, G., Mirtaheri, P. & Tønnessen, T. I. Detection of organ ischemia during hemorrhagic shock. *Acta Anaesthesiol. Scand.* **47,** 675–686 (2003).
7. Semenza, G. L., Roth, P. H., Fang, H. M. & Wang, G. L. Transcriptional regulation of genes encoding glycolytic enzymes by hypoxia-inducible factor 1. *J. Biol. Chem.* **269,** 23757–23763 (1994).
8. Abramov, A. Y., Scorziello, A. & Duchen, M. R. Three Distinct Mechanisms Generate Oxygen Free Radicals in Neurons and Contribute to Cell Death during Anoxia and Reoxygenation. *J. Neurosci.* **27,** 1129–1138 (2007).
9. Kinugasa, Y. *et al.* Allopurinol Improves Cardiac Dysfunction After Ischemia-Reperfusion via Reduction of Oxidative Stress in Isolated Perfused Rat Hearts. *Circ. J.* **67,** 781–787 (2003).
10. Xia, Y. & Zweier, J. L. Substrate Control of Free Radical Generation from Xanthine Oxidase in the Postischemic Heart. *J. Biol. Chem.* **270,** 18797–18803 (1995).
11. Walder, C. E. *et al.* Ischemic Stroke Injury Is Reduced in Mice Lacking a Functional NADPH Oxidase. *Stroke* **28,** 2252–2258 (1997).
12. Nicholson, J. K., Lindon, J. C. & Holmes, E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* **29,** 1181–1189 (1999).
13. Lusczek, E. R. *et al.* Urinary 1H-NMR Metabolomics Can Distinguish Pancreatitis Patients from Healthy Controls. *JOP J. Pancreas* **14,** 161–170 (2013).
14. Putri, S. P. *et al.* Current metabolomics: Practical applications. *J. Biosci. Bioeng.* **115,** 579–589 (2013).
15. Altavilla, D. *et al.* Oxidative stress causes nuclear factor-κB activation in acute hypovolemic hemorrhagic shock. *Free Radic. Biol. Med.* **30,** 1055–1066 (2001).
16. Lemasters, J. J. V. Necrapoptosis and the mitochondrial permeability transition: shared pathways to necrosis and apoptosis. *Am. J. Physiol. - Gastrointest. Liver Physiol.* **276,** G1–G6 (1999).
17. Threlfall, C. J. & Stoner, H. B. Carbohydrate Metabolism in Ischæmic Shock. *Exp. Physiol.* **39,** 1–9 (1954).
18. Seok, J. *et al.* Genomic responses in mouse models poorly mimic human inflammatory diseases. *Proc. Natl. Acad. Sci.* **110,** 3507–3512 (2013).

19. Hannon, J. P., Bossone, C. A. & Rodkey, W. G. Splenic red cell sequestration and blood volume measurements in conscious pigs. *Am. J. Physiol. - Regul. Integr. Comp. Physiol.* **248,** R293–R301 (1985).
20. Holcomb, J. B. *et al.* Effect of dry fibrin sealant dressings versus gauze packing on blood loss in grade V liver injuries in resuscitated swine. *J. Trauma* **46,** 49–57 (1999).
21. Bollen, M., Keppens, S. & Stalmans, W. Specific features of glycogen metabolism in the liver. *Biochem. J.* **336,** 19–31 (1998).
22. Grant, D. M. Detoxification pathways in the liver. *J. Inherit. Metab. Dis.* **14,** 421–430 (1991).
23. Caetano, A. C. *et al.* The antioxidant response of the liver of male Swiss mice raised on a AIN 93 or commercial diet. *BMC Physiol.* **13,** 3 (2013).
24. Jancova, P., Anzenbacher, P. & Anzenbacherova, E. Phase ii drug metabolizing enzymes. *Biomed. Pap.* **154,** 103–116 (2010).
25. Klaassen, C. D. & Lu, H. Xenobiotic Transporters: Ascribing Function from Gene Knockout and Mutation Studies. *Toxicol. Sci.* **101,** 186–196 (2008).
26. Ma, Y., Wang, P., Kuebler, J. F., Chaudry, I. H. & Messina, J. L. Hemorrhage induces the rapid development of hepatic insulin resistance. *Am. J. Physiol. - Gastrointest. Liver Physiol.* **284,** G107–G115 (2003).
27. Maitra, S. R., Geller, E. R., Pan, W., Kennedy, P. R. & Higgins, L. D. Altered cellular calcium regulation and hepatic glucose production during hemorrhagic shock. *Circ. Shock* **38,** 14–21 (1992).
28. Cairns, C. B. Rude unhinging of the machinery of life: metabolic approaches to hemorrhagic shock. *Curr. Opin. Crit. Care* **7,** 437–443 (2001).
29. Chaudry, I. H., Ohkawa, M., Clemens, M. G. & Baue, A. E. Alterations in electron transport and cellular metabolism with shock and trauma. *Prog. Clin. Biol. Res.* **111,** 67–88 (1983).
30. Mela, L. M., Miller, L. D., Bacalzo, L. V., Jr, Olofsson, K. & White, R. R., 4th. Alterations of mitochondrial structure and energy-linked functions in hemorrhagic shock and endotoxemia. *Adv. Exp. Med. Biol.* **33,** 231–242 (1972).
31. Ikai, I. *et al.* Significance of hepatic mitochondrial redox potential on the concentrations of plasma amino acids following hemorrhagic shock in rats. *Circ. Shock* **27,** 63–72 (1989).
32. Sharma, P., Walsh, K. T., Kerr-Knott, K. A., Karaian, J. & Morgan, P. Pyruvate Modulates Hepatic Mitochondrial Functions and Reduc... : Anesthesiology. *Anesthesiology* **103,** 65–73 (2005).
33. Mongan, P. D. *et al.* Pyruvate improves redox status and decreases indicators of hepatic apoptosis during hemorrhagic shock in swine. *Am. J. Physiol. - Heart Circ. Physiol.* **283,** H1634–H1644 (2002).
34. Paxian, M., Keller, S. A., Huynh, T. T. & Clemens, M. G. Perflubron Emulsion Improves Hepatic Microvascular Integrity... : Shock. *Shock* **20,** 449–457 (2003).
35. Chang, C. G., Van Way III, C. W., Dhar, A., Helling Jr., T. & Hahn, Y. The Use of Insulin and Glucose during Resuscitation from Hemorrhagic Shock Increases Hepatic ATP. *J. Surg. Res.* **92,** 171–176 (2000).
36. Jaeschke, H. Mechanisms of reperfusion injury after warm ischemia of the liver. *J. Hepatobiliary. Pancreat. Surg.* **5,** 402–408 (1998).
37. Jaeschke, H. Reactive oxygen and ischemia/reperfusion injury of the liver. *Chem. Biol. Interact.* **79,** 115–136 (1991).
38. Adamson, G. M. & Billings, R. E. Tumor necrosis factor induced oxidative stress in isolated mouse hepatocytes. *Arch. Biochem. Biophys.* **294,** 223–229 (1992).

39. Jaeschke, H. & Woolbright, B. L. Current strategies to minimize hepatic ischemia–reperfusion injury by targeting reactive oxygen species. *Transplant. Rev.* **26,** 103–114 (2012).
40. Jaeschke, H. Molecular mechanisms of hepatic ischemia-reperfusion injury and preconditioning. *Am. J. Physiol. - Gastrointest. Liver Physiol.* **284,** G15–G26 (2003).
41. Carrola, J. *et al.* Metabolic Signatures of Lung Cancer in Biofluids: NMR-Based Metabonomics of Urine. *J. Proteome Res.* **10,** 221–230 (2011).
42. Lusczek, E. R., Lexcen, D. R., Witowski, N. E., Mulier, K. E. & Beilman, G. Urinary metabolic network analysis in trauma, hemorrhagic shock, and resuscitation. *Metabolomics* **9,** 223–235 (2013).
43. Bodi, V. *et al.* Metabolomic Profile of Human Myocardial Ischemia by Nuclear Magnetic Resonance Spectroscopy of Peripheral Blood SerumA Translational Study Based on Transient Coronary Occlusion Models. *J. Am. Coll. Cardiol.* **59,** 1629–1641 (2012).
44. Brindle, J. T., Nicholson, J. K., Schofield, P. M., Grainger, D. J. & Holmes, E. Application of chemometrics to 1H NMR spectroscopic data to investigate a relationship between human serum metabolic profiles and hypertension. *The Analyst* **128,** 32–36 (2003).
45. Lin, C. Y., Wu, H., Tjeerdema, R. S. & Viant, M. R. Evaluation of metabolite extraction strategies from tissue samples using NMR metabolomics. *Metabolomics* **3,** 55–67 (2007).
46. Scribner, D. M. *et al.* Liver Metabolomic Changes Identify Biochemical Pathways in Hemorrhagic Shock. *J. Surg. Res.* **164,** e131–e139 (2010).
47. Sitter, B. *et al.* Quantification of metabolites in breast cancer patients with different clinical prognosis using HR MAS MR spectroscopy. *NMR Biomed.* **23,** 424–431 (2010).
48. Chen, J.-H. *et al.* Resolution of creatine and phosphocreatine 1H signals in isolated human skeletal muscle using HR-MAS 1H NMR. *Magn. Reson. Med.* **59,** 1221–1224 (2008).
49. Griffin, J. l., Williams, H. j., Sang, E. & Nicholson, J. k. Abnormal lipid profile of dystrophic cardiac tissue as demonstrated by one- and two-dimensional magic-angle spinning 1H NMR spectroscopy. *Magn. Reson. Med.* **46,** 249–255 (2001).
50. Swanson, M. G. *et al.* Quantitative analysis of prostate metabolites using 1H HR-MAS spectroscopy. *Magn. Reson. Med.* **55,** 1257–1264 (2006).
51. Swanson, M. G. *et al.* Quantification of choline- and ethanolamine-containing metabolites in human prostate tissues using 1H HR-MAS total correlation spectroscopy. *Magn. Reson. Med.* **60,** 33–40 (2008).
52. Barding, G. A., Béni, S., Fukao, T., Bailey-Serres, J. & Larive, C. K. Comparison of GC-MS and NMR for Metabolite Profiling of Rice Subjected to Submergence Stress. *J. Proteome Res.* **12,** 898–909 (2013).
53. Kettunen, J. *et al.* Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nat. Genet.* **44,** 269–276 (2012).
54. Madsen, R., Lundstedt, T. & Trygg, J. Chemometrics in metabolomics—A review in human disease diagnosis. *Anal. Chim. Acta* **659,** 23–33 (2010).
55. Zhang, A., Sun, H., Wang, P., Han, Y. & Wang, X. Modern analytical techniques in metabolomics analysis. *Analyst* **137,** 293–300 (2012).
56. Weljie, A. M., Newton, J., Mercier, P., Carlson, E. & Slupsky, C. M. Targeted Profiling: Quantitative Analysis of 1H NMR Metabolomics Data. *Anal. Chem.* **78,** 4430–4442 (2006).
57. Van den Berg, R. A., Hoefsloot, H. C., Westerhuis, J. A., Smilde, A. K. & Werf, M. J. van der. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* **7,** 142 (2006).
58. Lee, J. W., Lee, J. B., Park, M. & Song, S. H. An extensive comparison of recent classification tools applied to microarray data. *Comput. Stat. Data Anal.* **48,** 869–885 (2005).

59. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10,** 57–63 (2009).
60. Rapaport, F. *et al.* Comprehensive evaluation of differential expression analysis methods for RNA-seq data. (2013). at <http://arxiv.org/abs/1301.5277>
61. Twine, N. A., Janitz, K., Wilkins, M. R. & Janitz, M. Whole Transcriptome Sequencing Reveals Gene Expression and Splicing Differences in Brain Regions Affected by Alzheimer's Disease. *PLoS ONE* **6,** e16266 (2011).
62. Kvam, V. M., Liu, P. & Si, Y. A comparison of statistical methods for detecting differentially expressed genes from RNA-seq data. *Am. J. Bot.* **99,** 248–256 (2012).
63. Soneson, C. & Delorenzi, M. A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics* **14,** 91 (2013).
64. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2010).
65. Metzker, M. L. Sequencing technologies — the next generation. *Nat. Rev. Genet.* **11,** 31–46 (2010).
66. Chu, Y. & Corey, D. R. RNA Sequencing: Platform Selection, Experimental Design, and Data Interpretation. *Nucleic Acid Ther.* **22,** 271–274 (2012).
67. Marguerat, S. & Bähler, J. RNA-seq: from technology to biology. *Cell. Mol. Life Sci.* **67,** 569–579 (2010).
68. Kent, W. J. BLAT—The BLAST-Like Alignment Tool. *Genome Res.* **12,** 656–664 (2002).
69. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25,** 1105–1111 (2009).
70. Au, K. F., Jiang, H., Lin, L., Xing, Y. & Wong, W. H. Detection of splice junctions from paired-end RNA-seq data by SpliceMap. *Nucleic Acids Res.* **38,** 4570–4578 (2010).
71. Wang, K. *et al.* MapSplice: Accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res.* **38,** e178–e178 (2010).
72. Wu, T. D. & Nacu, S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26,** 873–881 (2010).
73. Bona, F. D., Ossowski, S., Schneeberger, K. & Rätsch, G. Optimal spliced alignments of short sequence reads. *BMC Bioinformatics* **9,** O7 (2008).
74. Garber, M., Grabherr, M. G., Guttman, M. & Trapnell, C. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat. Methods* **8,** 469–477 (2011).
75. Zerbino, D. R. & Birney, E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18,** 821–829 (2008).
76. Robertson, G. *et al.* De novo assembly and analysis of RNA-seq data. *Nat. Methods* **7,** 909–912 (2010).
77. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28,** 511–515 (2010).
78. Guttman, M. *et al.* Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat. Biotechnol.* **28,** 503–510 (2010).
79. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5,** 621–628 (2008).

80. Bullard, J. H., Purdom, E., Hansen, K. D. & Dudoit, S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* **11,** 94 (2010).
81. Jiang, H. & Wong, W. H. Statistical inferences for isoform expression in RNA-Seq. *Bioinformatics* **25,** 1026–1032 (2009).
82. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11,** 1–12 (2010).
83. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25,** 25–29 (2000).
84. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28,** 27–30 (2000).
85. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4,** 44–57 (2008).
86. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 15545–15550 (2005).
87. Khatri, P., Draghici, S., Ostermeier, G. C. & Krawetz, S. A. Profiling gene expression using onto-express. *Genomics* **79,** 266–270 (2002).
88. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37,** 1–13 (2009).
89. Khatri, P., Sirota, M. & Butte, A. J. Ten Years of Pathway Analysis: Current Approaches and Outstanding Challenges. *PLoS Comput Biol* **8,** e1002375 (2012).
90. Han, X. *et al.* Metabolomics in Early Alzheimer's Disease: Identification of Altered Plasma Sphingolipidome Using Shotgun Lipidomics. *PLoS ONE* **6,** e21643 (2011).
91. Davis, V. W., Schiller, D. E., Eurich, D. & Sawyer, M. B. Urinary metabolomic signature of esophageal cancer and Barrett's esophagus. *World J. Surg. Oncol.* **10,** 271 (2012).
92. Guan, W. *et al.* Ovarian cancer detection from metabolomic liquid chromatography/mass spectrometry data by support vector machines. *BMC Bioinformatics* **10,** 259 (2009).
93. Nishiumi, S. *et al.* Serum metabolomics as a novel diagnostic approach for pancreatic cancer. *Metabolomics* **6,** 518–528 (2010).
94. Bain, J. R. *et al.* Metabolomics Applied to Diabetes Research Moving From Information to Knowledge. *Diabetes* **58,** 2429–2443 (2009).
95. Determan Jr., C. *et al.* Carbohydrate fed state alters the metabolomic response to hemorrhagic shock and resuscitation in liver. *Metabolomics* 1–8 doi:10.1007/s11306-014-0621-6
96. Ein-Dor, L., Zuk, O. & Domany, E. Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. *Proc. Natl. Acad. Sci.* **103,** 5923–5928 (2006).
97. Michiels, S., Koscielny, S. & Hill, C. Prediction of cancer outcome with microarrays. *The Lancet* **365,** 1684–1685 (2005).
98. Zucknick, M., Richardson, S. & Stronach, E. A. Comparing the Characteristics of Gene Expression Profiles Derived by Univariate and Multivariate Classification Methods. *Stat. Appl. Genet. Mol. Biol.* **7,** Article7 (2008).
99. Wishart, D. S. in *Bioinforma. Methods Clin. Res.* (Matthiesen, R.) 283–313 (Humana Press, 2010). at <http://link.springer.com/protocol/10.1007/978-1-60327-194-3_14>
100. Wongravee, K. *et al.* Monte-Carlo methods for determining optimal number of significant variables. Application to mouse urinary profiles. *Metabolomics* **5,** 387–406 (2009).

101. Eisner, R. *et al.* Learning to predict cancer-associated skeletal muscle wasting from 1H-NMR profiles of urinary metabolites. *Metabolomics* **7,** 25–34 (2011).
102. Ametaj, B. N. *et al.* Metabolomics reveals unhealthy alterations in rumen metabolism with increased proportion of cereal grain in the diet of dairy cows. *Metabolomics* **6,** 583–594 (2010).
103. Xiao, J. F. *et al.* LC–MS Based Serum Metabolomics for Identification of Hepatocellular Carcinoma Biomarkers in Egyptian Cohort. *J. Proteome Res.* **11,** 5914–5923 (2012).
104. Haug, K. *et al.* MetaboLights--an open-access general-purpose repository for metabolomics studies and associated meta-data. *Nucleic Acids Res.* **41,** D781–D786 (2013).
105. Wold, S. in *Quant. Sociol. Int. Perspect. Math. Stat. Model.* (Blalock, H. M., Aganbegian, A., Borodkin, F. M., Boudon, R. & Capecchi, V.) 307–357 (Academic Press, 1975).
106. Cassel, C., Hackl, P. & Westlund, A. H. Robustness of partial least-squares method for estimating latent variable quality structures. *J. Appl. Stat.* **26,** 435–446 (1999).
107. Bogdanov, M. *et al.* Metabolomic profiling to develop blood biomarkers for Parkinson's disease. *Brain* **131,** 389–396 (2008).
108. Carraro, S. *et al.* Metabolomics Applied to Exhaled Breath Condensate in Childhood Asthma. *Am. J. Respir. Crit. Care Med.* **175,** 986–990 (2007).
109. Chen, T. *et al.* Serum and Urine Metabolite Profiling Reveals Potential Biomarkers of Human Hepatocellular Carcinoma. *Mol. Cell. Proteomics* **10,** (2011).
110. Duarte, I. F. *et al.* Can nuclear magnetic resonance (NMR) spectroscopy reveal different metabolic signatures for lung tumours? *Virchows Arch.* **457,** 715–725 (2010).
111. Qiu, Y. *et al.* Urinary Metabonomic Study on Colorectal Cancer. *J. Proteome Res.* **9,** 1627–1634 (2010).
112. Lexcen, D. R., Lusczek, E. R., Witowski, N. E., Mulier, K. E. & Beilman, G. J. Metabolomics classifies phase of care and identifies risk for mortality in a porcine model of multiple injuries and hemorrhagic shock. *J. Trauma Acute Care Surg.* **73,** S147–S155 (2012).
113. Sanchez, G. DiscriMiner: Tools of the Trade for Discriminant Analysis. *R Package Version 01-29* (2012). at <http://CRAN.R-project.org/package=DiscriMiner>
114. Zou, H. & Hastie, T. Regularization and variable selection via the Elastic Net. *J. R. Stat. Soc. Ser. B* **67,** 301–320 (2005).
115. Rohart, F. *et al.* Phenotypic prediction based on metabolomic data for growing pigs from three main European breeds. *J. Anim. Sci.* **90,** 4729–4740 (2012).
116. Wahl, S. *et al.* Metabolomics reveals determinants of weight loss during lifestyle intervention in obese children. *Metabolomics* 1–11 (2013). doi:10.1007/s11306-013-0550-9
117. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **33,** 1–22 (2010).
118. Breiman, L. Random Forests. *Mach. Learn.* **45,** 5–32 (2001).
119. Folleco, A., Khoshgoftaar, T. M., Van Hulse, J. & Bullard, L. Software quality modeling: The impact of class noise on the random forest classifier. in *IEEE Congr. Evol. Comput. 2008 CEC 2008 IEEE World Congr. Comput. Intell.* 3853 –3859 (2008). doi:10.1109/CEC.2008.4631321
120. Svetnik, V. *et al.* Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. *J. Chem. Inf. Comput. Sci.* **43,** 1947–1958 (2003).
121. Verikas, A., Gelzinis, A. & Bacauskiene, M. Mining data with random forests: A survey and results of new tests. *Pattern Recognit.* **44,** 330–349 (2011).
122. Hische, M. *et al.* A distinct metabolic signature predicts development of fasting plasma glucose. *J. Clin. Bioinforma.* **2,** 1–10 (2012).
123. Houtkooper, R. H. *et al.* The metabolic footprint of aging in mice. *Sci. Rep.* **1,** (2011).

124. Patterson, A. D. *et al.* Aberrant Lipid Metabolism in Hepatocellular Carcinoma Revealed by Plasma Metabolomics and Lipid Profiling. *Cancer Res.* **71,** 6590–6600 (2011).
125. Romero, R. *et al.* Metabolomics in premature labor: a novel approach to identify patients at risk for preterm delivery. *J. Matern. Fetal Neonatal Med.* **23,** 1344–1359 (2010).
126. Liaw, A. & Wiener, M. Classification and Regression by randomForest. *R News* **2,** 18–22 (2002).
127. Friedman, J. H. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Stat.* **29,** 1189–1232 (2001).
128. Hastie, T., Tibshirani, R. & Friedman, J. in *Elem. Stat. Learn.* 337–387 (Springer New York, 2009). at <http://link.springer.com/chapter/10.1007/978-0-387-84858-7_10>
129. Ridgeway, G. gbm: Generalized Boosted Regression Models. (2013). at <http://CRAN.R-project.org/package=gbm>
130. Vapnik, V. in *Nonlinear Model.* (Suykens, J. A. K. & Vandewalle, J.) 55–85 (Springer US, 1998). at <http://link.springer.com/chapter/10.1007/978-1-4615-5703-6_3>
131. Mahadevan, S., Shah, S. L., Marrie, T. J. & Slupsky, C. M. Analysis of Metabolomic Data Using Support Vector Machines. *Anal. Chem.* **80,** 7562–7570 (2008).
132. Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A. & Leisch, F. e1071: Misc Functions of the Department of Statistics (e1071). (2012). at <http://CRAN.R-project.org/package=e1071>
133. Guyon, I., Weston, J., Barnhill, S. & Vapnik, V. Gene Selection for Cancer Classification using Support Vector Machines. *Mach. Learn.* **46,** 389–422 (2002).
134. Tibshirani, R., Hastie, T., Narasimhan, B. & Chu, G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc. Natl. Acad. Sci.* **99,** 6567–6572 (2002).
135. Ray, S. *et al.* Classification and prediction of clinical Alzheimer's diagnosis based on plasma signaling proteins. *Nat. Med.* **13,** 1359–1362 (2007).
136. Sadanandam, A. *et al.* A colorectal cancer classification system that associates cellular phenotype and responses to therapy. *Nat. Med.* **19,** 619–625 (2013).
137. Saeys, Y., Inza, I. & Larrañaga, P. A review of feature selection techniques in bioinformatics. *Bioinformatics* **23,** 2507–2517 (2007).
138. Real, R. & Vargas, J. M. The Probabilistic Basis of Jaccard's Index of Similarity. *Syst. Biol.* **45,** 380–385 (1996).
139. Dice, L. R. Measures of the Amount of Ecologic Association Between Species. *Ecology* **26,** 297–302 (1945).
140. Sørensen, T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Biol Skr* **5,** 1–34 (1948).
141. Ochiai, A. Zoogeographical studies on the soleoid fishes found in Japan and its neigbouring regions. *Bull. Jpn. Soc. Sci. Fish.* **22,** 526–530 (1957).
142. Shi, L. *et al.* Cross-platform comparability of microarray technology: Intra-platform consistency and appropriate data analysis procedures are essential. *BMC Bioinformatics* **6,** S12 (2005).
143. Kuncheva, L. I. A Stability Index for Feature Selection. in (ACTA Press, 2007). at <http://www.actapress.com/Abstract.aspx?paperId=29484>
144. Jurman, G. *et al.* Algebraic stability indicators for ranked lists in molecular profiling. *Bioinformatics* **24,** 258–264 (2008).
145. Saeys, Y., Abeel, T. & Peer, Y. V. de. in *Mach. Learn. Knowl. Discov. Databases* (Daelemans, W., Goethals, B. & Morik, K.) 313–325 (Springer Berlin Heidelberg, 2008). at <http://link.springer.com/chapter/10.1007/978-3-540-87481-2_21>

146. Abeel, T., Helleputte, T., Peer, Y. V. de, Dupont, P. & Saeys, Y. Robust biomarker identification for cancer diagnosis with ensemble feature selection methods. *Bioinformatics* **26,** 392–398 (2010).
147. Davis, C. A. *et al.* Reliable gene signatures for microarray classification: assessment of stability and performance. *Bioinformatics* **22,** 2356–2363 (2006).
148. Breiman, L. Bagging predictors. *Mach. Learn.* **24,** 123–140 (1996).
149. Wolpert, D. H. & Macready, W. G. No free lunch theorems for optimization. *IEEE Trans. Evol. Comput.* **1,** 67 –82 (1997).
150. Holcomb, J. B., Stansbury, L. G., Champion, H. R., Wade, C. & Bellamy, R. F. Understanding combat casualty care statistics. *J. Trauma* **60,** 397–401 (2006).
151. Spinella, P. C. & Holcomb, J. B. Resuscitation and transfusion principles for traumatic hemorrhagic shock. *Blood Rev.* **23,** 231–240 (2009).
152. Holcomb, J. *et al.* Causes of death in US Special Operations Forces in the global war on terrorism: 2001-2004. *US Army Med. Dep. J.* 24–37 (2007).
153. Dubick, M. A. Current concepts in fluid resuscitation for prehospital care of combat casualties. *US Army Med. Dep. J.* 18–24 (2011).
154. Eastridge, B. J. *et al.* Death on the battlefield (2001-2011): implications for the future of combat casualty care. *J. Trauma Acute Care Surg.* **73,** S431–437 (2012).
155. Iyegha, U. P. *et al.* Does the fed state impact outcome in polytrauma and hemorrhagic shock? *J. Am. Coll. Surg.* **215,** S52 (2012).
156. Tyagi, R. K., Azrad, A., Degani, H. & Salomon, Y. Simultaneous extraction of cellular lipids and water-soluble metabolites: Evaluation by NMR spectroscopy. *Magn. Reson. Med.* **35,** 194–200 (1996).
157. R Core Team. R: A Language and Environment for Statistical Computing. (2013). at <http://www.R-project.org/>
158. Bijlsma, S. *et al.* Large-Scale Human Metabolomics Studies: A Strategy for Data (Pre-) Processing and Validation. *Anal. Chem.* **78,** 567–574 (2006).
159. Trygg, J., Holmes, E. & Lundstedt, T. Chemometrics in Metabonomics. *J. Proteome Res.* **6,** 469–479 (2007).
160. Liland, K. H. Multivariate methods in metabolomics – from pre-processing to dimension reduction and statistical analysis. *TrAC Trends Anal. Chem.* **30,** 827–841 (2011).
161. Simpson, G. L. permute: Functions for generating restricted permutations of data. R package version 0.7-0. (2012). at <http://CRAN.R-project.org/package=permute>
162. Szymanska, E., Saccenti, E., Smilde, A. K. & Westerhuis, J. A. Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics* **8,** 3–16 (2012).
163. Westerhuis, J. A. *et al.* Assessment of PLSDA cross validation. *Metabolomics* **4,** 81–89 (2008).
164. McNelis, J. *et al.* Prolonged lactate clearance is associated with increased mortality in the surgical intensive care unit. *Am. J. Surg.* **182,** 481–485 (2001).
165. Orman, M. A., Androulakis, I. P., Berthiaume, F. & Ierapetritou, M. G. Metabolic network analysis of perfused livers under fed and fasted states: Incorporating thermodynamic and futile-cycle-associated regulatory constraints. *J. Theor. Biol.* **293,** 101–110 (2012).
166. Orman, M. A., Ierapetritou, M. G., Androulakis, I. P. & Berthiaume, F. Effect of Fasting on the Metabolic Response of Liver to Experimental Burn Injury. *PLoS ONE* **8,** e54825 (2013).
167. Voelckel, W. G. *et al.* Arginine vasopressin, but not epinephrine, improves survival in uncontrolled hemorrhagic shock after liver trauma in pigs*. *Crit. Care Med.* **31,** 1160–1165 (2003).

168. Zhang, S. *et al.* Effect of Endogenous Adenosine Augmentation on Ischemia and Reperfusion Injury to the Liver. *Transplant. Proc.* **29,** 1336–1337 (1997).
169. Guinzberg, R. *et al.* Inosine released after hypoxia activates hepatic glucose liberation through A3 adenosine receptors. *Am. J. Physiol. - Endocrinol. Metab.* **290,** E940–E951 (2006).
170. Snodgrass, P. J., Lin, R. C., Müller, W. A. & Aoki, T. T. Induction of urea cycle enzymes of rat liver by glucagon. *J. Biol. Chem.* **253,** 2748–2753 (1978).
171. Kooij, A., Schiller, H. J., Schijns, M., Van Noorden, C. J. F. & Frederiks, W. M. Conversion of xanthine dehydrogenase into xanthine oxidase in rat liver and plasma at the onset of reperfusion after ischemia. *Hepatology* **19,** 1488–1495 (1994).
172. Elias-Miró, M., Bibiana Jiménez-Castro, M., Rodés, J. & Peralta, C. Current knowledge on oxidative stress in hepatic ischemia/reperfusion. *Free Radic. Res.* 1–32 (2013). doi:10.3109/10715762.2013.811721
173. Hoberman, H. D. & Prosky, L. Evidence of reduction of fumarate to succinate in perfused rat liver under conditions of reduced O2 tension. *Biochim. Biophys. Acta BBA - Gen. Subj.* **148,** 392–399 (1967).
174. Eisenberg, D., Gill, H. S., Pfluegl, G. M. U. & Rotstein, S. H. Structure–function relationships of glutamine synthetases. *Biochim. Biophys. Acta BBA - Protein Struct. Mol. Enzymol.* **1477,** 122–145 (2000).
175. Nissen, S. *et al.* β-Hydroxy-β-Methylbutyrate (HMB) Supplementation in Humans Is Safe and May Decrease Cardiovascular Risk Factors. *J. Nutr.* **130,** 1937–1945 (2000).
176. Weitzman, S. A. & Stossel, T. P. Effects of oxygen radical scavengers and antioxidants on phagocyte-induced mutagenesis. *J. Immunol.* **128,** 2770–2772 (1982).
177. Paulin, A., Droillard, M. J. & Bureau, J. M. Effect of a free radical scavenger, 3,4,5-trichlorophenol, on ethylene production and on changes in lipids and membrane integrity during senescence of petals of cut carnations (Dianthus caryophyllus). *Physiol. Plant.* **67,** 465–471 (1986).
178. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
179. Durinck, S. *et al.* BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21,** 3439–3440 (2005).
180. Smedley, D. *et al.* BioMart – biological queries made easy. *BMC Genomics* **10,** 22 (2009).
181. Oshlack, A., Robinson, M. D. & Young, M. D. From RNA-seq reads to differential expression results. *Genome Biol.* **11,** 220 (2010).
182. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11,** R25 (2010).
183. Theodoropoulos, P. A. *et al.* Hepatocyte swelling leads to rapid decrease of the G-/total actin ratio and increases actin mRNA levels. *FEBS Lett.* **311,** 241–245 (1992).
184. Pedersen, S. ., Hoffmann, E. . & Mills, J. . The cytoskeleton and cell volume regulation. *Comp. Biochem. Physiol. A. Mol. Integr. Physiol.* **130,** 385–399 (2001).
185. Hoffmann, E. K., Lambert, I. H. & Pedersen, S. F. Physiology of Cell Volume Regulation in Vertebrates. *Physiol. Rev.* **89,** 193–277 (2009).
186. Cinel, I. & Opal, S. M. Molecular biology of inflammation and sepsis: A primer*. *Crit. Care Med. January 2009* **37,** 291–304 (2009).
187. Tsan, M.-F. & Gao, B. Cytokine function of heat shock proteins. *Am. J. Physiol. - Cell Physiol.* **286,** C739–C744 (2004).

188. Kol, A., Lichtman, A. H., Finberg, R. W., Libby, P. & Kurt-Jones, E. A. Cutting Edge: Heat Shock Protein (HSP) 60 Activates the Innate Immune Response: CD14 Is an Essential Receptor for HSP60 Activation of Mononuclear Cells. *J. Immunol.* **164,** 13–17 (2000).
189. Angele, M. K., Schneider, C. P. & Chaudry, I. H. Bench-to-bedside review: Latest results in hemorrhagic shock. *Crit. Care* **12,** 218 (2008).
190. Peitzman, A. B. *et al.* Hemorrhagic shock. *Curr. Probl. Surg.* **32,** 925–1002 (1995).
191. Benkoël, L. *et al.* Effect of Ischemia–Reperfusion on Bile Canalicular F-Actin Microfilaments in Hepatocytes of Human Liver Allograft. *Dig. Dis. Sci.* **46,** 1663–1667 (2001).
192. Gaussin, V., Baquet, A. & Hue, L. Cell shrinkage follows, rather than mediates, the short-term effects of glucagon on carbohydrate metabolism. *Biochem. J.* **287,** 17–20 (1992).
193. Haussinger, D. The role of cellular hydration in the regulation of cell function. *Biochem. J.* **313,** 697–710 (1996).
194. Häussinger, D., Gerok, W., Roth, E. & Lang, F. Cellular hydration state: an important determinant of protein catabolism in health and disease. *The Lancet* **341,** 1330–1332 (1993).
195. Haussinger, D., Lang, F. & Gerok, W. Regulation of cell function by the cellular hydration state. *Am. J. Physiol. - Endocrinol. Metab.* **267,** E343–E355 (1994).
196. Molitoris, B. A. Putting the actin cytoskeleton into perspective: pathophysiology of ischemic alterations. *Am. J. Physiol. - Ren. Physiol.* **272,** F430–F433 (1997).
197. Flores, J., DiBona, D. R., Frega, N. & Leaf, D. A. Cell volume regulation and ischemic tissue damage. *J. Membr. Biol.* **10,** 331–343 (1972).
198. Leaf, A. Cell Swelling A Factor in Ischemic Tissue Injury. *Circulation* **48,** 455–458 (1973).
199. Tranum-Jensen, J. *et al.* Tissue osmolality, cell swelling, and reperfusion in acute regional myocardial ischemia in the isolated porcine heart. *Circ. Res.* **49,** 364–381 (1981).
200. Manciet, L. H., Poole, D. C., McDonagh, P. F., Copeland, J. G. & Mathieu-Costello, O. Microvascular compression during myocardial ischemia: mechanistic basis for no-reflow phenomenon. *Am. J. Physiol. - Heart Circ. Physiol.* **266,** H1541–H1550 (1994).
201. Gavin, J. B., Thomson, R. W., Humphrey, S. M. & Herdson, P. B. Changes in vascular morphology associated with the no-reflow phenomenon in ischaemic myocardium. *Virchows Arch. A* **399,** 325–332 (1983).
202. Rezkalla, S. H. & Kloner, R. A. No-Reflow Phenomenon. *Circulation* **105,** 656–662 (2002).
203. Oreopoulos, G. D. *et al.* Hypertonic preconditioning prevents hepatocellular injury following ischemia/reperfusion in mice: A role for interleukin 10. *Hepatology* **40,** 211–220 (2004).
204. Coimbra, R. *et al.* Hypertonic Saline Resuscitation Decreases Susceptibility to... : Journal of Trauma and Acute Care Surgery. *J. Trauma-Inj. Infect. Surg. Care* **42,** 602–607 (1997).
205. Bahrami, S. *et al.* Small-Volume Fluid Resuscitation with Hypertonic Saline Prevents Inflammation but not Mortality in a Rat Model of Hemorrhagic Shock. *Shock March 2006* **25,** 283–289 (2006).
206. Murao, Y. *et al.* Hypertonic Saline Resuscitation Reduces Apoptosis and Tissue Damage of the Small Intestine in a Mouse Model of Hemorrhagic Shock: *Shock* **20,** 23–28 (2003).
207. Ke, Q.-H., Zheng, S.-S., Liang, T.-B., Xie, H.-Y. & Xia, W.-L. Pretreatment of Hypertonic Saline Can Increase Endogenous Interleukin 10 Release to Attenuate Hepatic Ischemia Reperfusion Injury. *Dig. Dis. Sci.* **51,** 2257–2263 (2006).
208. Benkhart, E. M., Siedlar, M., Wedel, A., Werner, T. & Ziegler-Heitbrock, H. W. L. Role of Stat3 in Lipopolysaccharide-Induced IL-10 Gene Expression. *J. Immunol.* **165,** 1612–1617 (2000).
209. Gatsios, P. *et al.* Activation of the Janus Kinase/Signal Transducer and Activator of Transcription Pathway by Osmotic Shock. *J. Biol. Chem.* **273,** 22962–22968 (1998).

210. Yoshidome, H., Kato, A., Edwards, M. J. & Lentsch, A. B. Interleukin-10 suppresses hepatic ischemia/reperfusion injury in mice: Implications of a central role for nuclear factor κB. *Hepatology* **30,** 203–208 (1999).
211. Fiorentino, D. F., Zlotnik, A., Mosmann, T. R., Howard, M. & O'Garra, A. IL-10 inhibits cytokine production by activated macrophages. *J. Immunol.* **147,** 3815–3822 (1991).

APPENDIX A

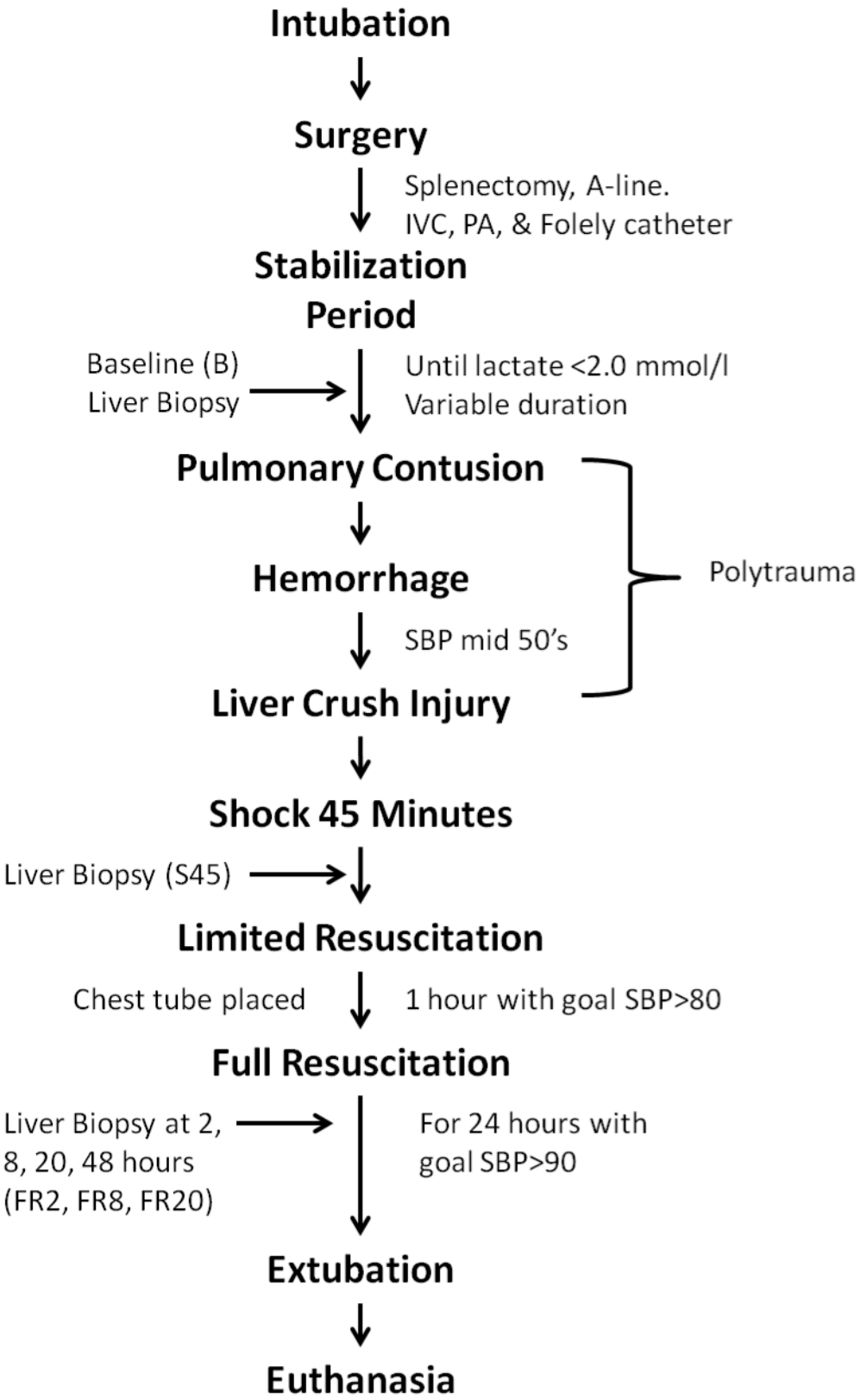

Figure 1: Graphical representation of the experimental timeline described in Methods.

APPENDIX B

NMR – BINARY - HIGH SAMPLE												
Non-Ensemble												
	Random						Correlated					
Method	PLSDA	GBM	SVM	RF	GLMNET	PAM	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	0.466	0.497	0.423	0.196	0.597	0.621	0.460	0.366	0.422	0.213	0.552	0.528
Accuracy	0.436	0.590	0.780	0.880	0.740	0.750	0.460	0.530	0.660	0.940	0.670	0.660
AUC-ROC	0.597	0.781	0.892	0.956	0.876	0.806	0.779	0.787	0.824	0.984	0.856	0.751
Stability	0.50	0.43	0.29	0.11	0.50	0.53	0.46	0.28	0.31	0.12	0.47	0.44
Ensemble												
	Random						Correlated					
Method	PLSDA	GBM	SVM	RF	GLMNET	PAM	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	0.462	0.493	0.424	0.405	0.478	0.617	0.441	0.374	0.477	0.315	0.557	0.510
Accuracy	0.445	0.530	0.720	0.920	0.710	0.780	0.432	0.470	0.670	0.920	0.520	0.650
AUC-ROC	0.648	0.765	0.884	0.960	0.884	0.783	0.593	0.870	0.816	0.980	0.708	0.769
Stability	0.48	0.46	0.30	0.26	0.36	0.51	0.45	0.31	0.37	0.19	0.60	0.42

Table 1 – Results from High Sample (50/group) NMR-scale Binary Null Classification Simulations. RPT – Robustness-Performance Trade-off, AUC-ROC – Area under the Receiver Operator Curve

NMR – BINARY - LOW SAMPLE												
Non-Ensemble												
	Random						Correlated					
Method	PLSDA	GBM	SVM	RF	GLMNET	PAM	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	0.411	0.432	0.395	0.212	0.569	0.562	0.455	0.422	0.525	0.227	0.596	0.578
Accuracy	0.365	0.425	0.825	0.900	0.775	0.700	0.410	0.425	0.850	0.900	0.650	0.750
AUC-ROC	0.565	0.663	0.925	0.975	0.900	0.806	0.603	0.706	0.900	0.975	0.900	0.831
Stability	0.47	0.44	0.26	0.12	0.45	0.47	0.51	0.42	0.38	0.13	0.55	0.47
Ensemble												
	Random						Correlated					
Method	PLSDA	GBM	SVM	RF	GLMNET	PAM	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	0.386	0.457	0.558	0.357	0.530	0.525	0.442	0.502	0.525	0.355	0.570	0.434
Accuracy	0.365	0.475	0.875	0.950	0.875	0.700	0.410	0.675	0.675	0.925	0.750	0.600
AUC-ROC	0.565	0.713	0.950	1.000	0.950	0.894	0.603	0.681	0.900	1.000	0.925	0.744
Stability	0.41	0.44	0.41	0.22	0.38	0.42	0.48	0.40	0.43	0.22	0.46	0.34

Table 2 – Results from Low Sample (25/group) NMR-scale Binary Null Classification Simulations. RPT – Robustness-Performance Trade-off, AUC-ROC – Area under the Receiver Operator Curve

NMR – MULTI-CLASS - HIGH SAMPLE												
Non-Ensemble												
	Random						Correlated					
Method	PLSDA	GBM	SVM	RF	GLMNET	PAM	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	0.357	0.340	0.418	0.211	0.462	0.462	0.299	0.342	0.460	0.211	0.432	0.466
Accuracy	0.262	0.315	0.390	0.875	0.410	0.430	0.238	0.280	0.470	0.870	0.415	0.445
AUC-ROC	0.657	0.683	0.763	0.964	0.757	0.736	0.645	0.714	0.751	0.967	0.713	0.688
Stability	0.56	0.37	0.45	0.12	0.53	0.50	0.40	0.44	0.45	0.12	0.45	0.49
Ensemble												
	Random						Correlated					
Method	PLSDA	GBM	SVM	RF	GLMNET	PAM	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	0.347	0.392	0.418	0.269	0.397	0.443	0.317	0.280	0.428	0.256	0.444	0.436
Accuracy	0.258	0.360	0.390	0.850	0.385	0.405	0.241	0.205	0.460	0.880	0.335	0.400
AUC-ROC	0.589	0.672	0.764	0.921	0.735	0.696	0.653	0.665	0.740	0.937	0.741	0.714
Stability	0.53	0.43	0.45	0.16	0.41	0.49	0.46	0.44	0.40	0.15	0.66	0.48

Table 3 – Results from High Sample (50/group) NMR-scale Multi-Class Null Classification Simulations. RPT – Robustness-Performance Trade-off, AUC-ROC – Area under the Receiver Operator Curve

NMR – MULTI-CLASS - LOW SAMPLE												
Non-Ensemble												
	Random						Correlated					
Method	PLSDA	GBM	SVM	RF	GLMNET	PAM	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	0.336	0.350	0.399	0.212	0.518	0.524	0.351	0.280	0.439	0.228	0.472	0.571
Accuracy	0.250	0.300	0.638	0.925	0.550	0.563	0.306	0.225	0.563	0.913	0.538	0.538
AUC-ROC	0.647	0.723	0.913	0.963	0.871	0.817	0.598	0.722	0.838	0.992	0.796	0.772
Stability	0.51	0.42	0.29	0.12	0.49	0.49	0.41	0.37	0.36	0.13	0.42	0.61
Ensemble												
	Random						Correlated					
Method	PLSDA	GBM	SVM	RF	GLMNET	PAM	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	0.271	0.329	0.480	0.227	0.403	0.460	0.347	0.359	0.548	0.227	0.551	0.585
Accuracy	0.263	0.275	0.600	0.900	0.475	0.450	0.306	0.325	0.700	0.888	0.525	0.600
AUC-ROC	0.733	0.782	0.858	0.954	0.804	0.741	0.598	0.749	0.850	0.967	0.800	0.794
Stability	0.28	0.41	0.40	0.13	0.35	0.47	0.40	0.40	0.45	0.13	0.58	0.57

Table 4 – Results from Low Sample (25/group) NMR-scale Multi-Class Null Classification Simulations. RPT – Robustness-Performance Trade-off, AUC-ROC – Area under the Receiver Operator Curve

MS – BINARY - HIGH SAMPLE												
Non-Ensemble												
	Random						Correlated					
Method	PLSDA	GBM	SVM	RF	GLMNET	PAM	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	0.367	0.370	0.181	0.077	0.428	0.448	0.372	0.297	0.241	0.058	0.450	0.462
Accuracy	0.500	0.590	0.940	0.940	0.820	0.810	0.520	0.420	0.860	0.880	0.820	0.770
AUC-ROC	0.580	0.805	0.964	0.952	0.892	0.891	0.580	0.809	0.952	0.924	0.984	0.808
Stability	0.29	0.27	0.10	0.04	0.29	0.31	0.29	0.23	0.14	0.03	0.31	0.33
Ensemble												
	Random						Correlated					
Method	PLSDA	GBM	SVM	RF	GLMNET	PAM	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	0.324	0.391	0.400	0.257	0.413	0.355	0.338	0.326	0.487	0.297	0.385	0.406
Accuracy	0.500	0.560	0.870	0.900	0.880	0.78	0.52	0.39	0.86	0.85	0.84	0.74
AUC-ROC	0.580	0.755	0.980	0.968	0.968	0.872	0.580	0.793	0.984	0.956	0.924	0.865
Stability	0.24	0.30	0.26	0.15	0.27	0.23	0.25	0.28	0.34	0.18	0.25	0.28

Table 5 – Results from High Sample (50/group) MS-scale Binary Null Classification Simulations. RPT – Robustness-Performance Trade-off, AUC-ROC – Area under the Receiver Operator Curve

MS – BINARY - LOW SAMPLE												
Non-Ensemble												
	Random						Correlated					
Method	PLSDA	GBM	SVM	RF	GLMNET	PAM	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	0.339	0.308	0.368	0.058	0.568	0.554	0.393	0.286	0.357	0.077	0.543	0.514
Accuracy	0.295	0.400	0.925	0.850	0.925	0.900	0.465	0.450	0.950	0.875	0.950	0.900
AUC-ROC	0.549	0.731	1.000	0.975	1.000	0.831	0.566	0.731	1.000	0.975	1.000	0.925
Stability	0.40	0.25	0.23	0.03	0.41	0.40	0.34	0.21	0.22	0.04	0.38	0.36
Ensemble												
	Random						Correlated					
Method	PLSDA	GBM	SVM	RF	GLMNET	PAM	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	0.312	0.460	0.563	0.366	0.553	0.494	0.386	0.352	0.479	0.318	0.453	0.450
Accuracy	0.295	0.525	0.900	0.900	0.950	0.900	0.465	0.475	0.875	0.975	0.925	0.900
AUC-ROC	0.549	0.825	0.950	1.000	1.000	0.919	0.566	0.719	0.975	1.000	1.000	0.900
Stability	0.33	0.41	0.41	0.23	0.39	0.34	0.33	0.28	0.33	0.19	0.30	0.30

Table 6 – Results from Low Sample (25/group) MS-scale Binary Null Classification Simulations. RPT – Robustness-Performance Trade-off, AUC-ROC – Area under the Receiver Operator Curve

MS – MULTI-CLASS - HIGH SAMPLE												
Non-Ensemble												
	Random						Correlated					
Method	PLSDA	GBM	SVM	RF	GLMNET	PAM	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	0.249	0.238	0.206	0.039	0.421	0.415	0.313	0.240	0.191	0.020	0.368	0.421
Accuracy	0.193	0.260	0.730	0.910	0.580	0.590	0.328	0.300	0.720	0.860	0.580	0.580
AUC-ROC	0.669	0.686	0.885	0.983	0.804	0.711	0.777	0.695	0.877	0.945	0.830	0.736
Stability	0.35	0.22	0.12	0.02	0.33	0.32	0.30	0.20	0.11	0.01	0.27	0.33
Ensemble												
	Random						Correlated					
Method	PLSDA	GBM	SVM	RF	GLMNET	PAM	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	0.191	0.282	0.371	0.225	0.376	0.324	0.338	0.326	0.487	0.297	0.385	0.406
Accuracy	0.238	0.285	0.645	0.840	0.535	0.500	0.520	0.390	0.860	0.850	0.840	0.740
AUC-ROC	0.659	0.683	0.834	0.899	0.775	0.715	0.580	0.793	0.984	0.956	0.924	0.865
Stability	0.16	0.28	0.26	0.13	0.29	0.24	0.25	0.28	0.34	0.18	0.25	0.28

Table 7 – Results from High Sample (50/group) MS-scale Multi-Class Null Classification Simulations. RPT – Robustness-Performance Trade-off, AUC-ROC – Area under the Receiver Operator Curve

MS – MULTI-CLASS - LOW SAMPLE												
Non-Ensemble												
	Random						Correlated					
Method	PLSDA	GBM	SVM	RF	GLMNET	PAM	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	0.270	0.234	0.299	0.039	0.545	0.513	0.285	0.245	0.286	0.039	0.494	0.540
Accuracy	0.213	0.288	0.875	0.888	0.812	0.750	0.231	0.263	0.900	0.938	0.788	0.700
AUC-ROC	0.726	0.804	0.954	0.967	0.913	0.827	0.630	0.689	0.958	0.996	0.913	0.769
Stability	0.37	0.28	0.18	0.02	0.41	0.39	0.37	0.23	0.17	0.02	0.36	0.44
Ensemble												
	Random						Correlated					
Method	PLSDA	GBM	SVM	RF	GLMNET	PAM	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	0.160	0.308	0.465	0.302	0.501	0.434	0.174	0.290	0.588	0.259	0.463	0.435
Accuracy	0.188	0.275	0.850	0.938	0.775	0.725	0.206	0.300	0.888	0.950	0.725	0.575
AUC-ROC	0.737	0.774	0.963	0.996	0.913	0.709	0.770	0.715	0.988	0.967	0.846	0.790
Stability	0.14	0.35	0.32	0.18	0.37	0.31	0.15	0.28	0.44	0.15	0.34	0.35

Table 8 – Results from Low Sample (25/group) MS-scale Multi-Class Null Classification Simulations. RPT – Robustness-Performance Trade-off, AUC-ROC – Area under the Receiver Operator Curve

MS - BINARY – DISCRIMINATORY - HIGH SAMPLE						
Subset						
Method	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	0.681	0.484	0.630	0.302	0.718	0.746
Accuracy	0.770	0.570	1.000	0.940	1.000	0.960
AUC-ROC	0.602	0.784	1.000	0.880	1.000	0.939
Stability	0.61	0.42	0.46	0.18	0.56	0.61
Max # Features	20	20	20	20	20	20
TP %	65.0%	54.0%	52.0%	38.5%	64.5%	62.0%
Model Derived						
Method	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	0.655	0.514	0.598	0.182	0.649	0.857
Accuracy	0.810	0.600	0.980	0.990	1.000	1.000
AUC-ROC	0.849	0.770	1.000	1.000	1.000	1.000
Stability	0.55	0.45	0.43	0.10	0.48	0.75
Max # Features	285	23	100	126	83	800
TP %	6.7%	47.4%	15.7%	9.0%	19.6%	2.5%
Ensemble						
Method	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	0.681	0.525	0.630	0.305	0.718	0.746
Accuracy	0.770	0.650	1.000	1.000	1.000	0.960
AUC-ROC	0.602	0.846	1.000	1.000	1.000	0.939
Stability	0.61	0.44	0.46	0.18	0.56	0.61
Max # Features	20	20	20	20	20	20
TP %	66.5%	54.5%	52.0%	41.5%	64.5%	62.0%

Table 9 – Results from High Sample (50/group) MS-scale Binary Classification Simulations. RPT – Robustness-Performance Trade-off, AUC-ROC – Area under the Receiver Operator Curve

MS - BINARY – DISCRIMINATORY - LOW SAMPLE						
Subset						
Method	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	0.429	0.336	0.435	0.113	0.573	0.655
Accuracy	0.410	0.475	0.975	0.975	0.950	0.950
AUC-ROC	0.552	0.794	1.000	1.000	0.950	0.925
Stability	0.45	0.26	0.28	0.06	0.41	0.50
Max # Features	20	20	20	20	20	20
TP %	33.0%	22.0%	20.0%	10.5%	32.5%	25.5%
Model Derived						
Method	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	0.492	0.292	0.614	0.112	0.670	0.874
Accuracy	0.413	0.400	0.925	0.825	0.975	0.950
AUC-ROC	0.700	0.713	0.975	0.950	1.000	0.925
Stability	0.61	0.23	0.46	0.06	0.51	0.81
Max # Features	325	85	100	68	67	881
TP %	4.1%	8.9%	9.6%	4.9%	13.6%	2.1%
Ensemble						
Method	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	0.394	0.480	0.601	0.433	0.563	0.582
Accuracy	0.410	0.625	1.000	0.950	0.950	0.950
AUC-ROC	0.552	0.663	1.000	1.000	0.975	0.925
Stability	0.38	0.39	0.43	0.28	0.40	0.42
Max # Features	20	20	20	20	20	20
TP %	31.0%	28.0%	25.0%	23.5%	33.0%	28.0%

Table 10 – Results from Low Sample (25/group) MS-scale Binary Classification Simulations. RPT – Robustness-Performance Trade-off, AUC-ROC – Area under the Receiver Operator Curve

MS - MULTI-CLASS - DISCRIMINATORY - HIGH SAMPLE						
Subset						
Method	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	0.462	0.246	0.495	0.147	0.666	0.581
Accuracy	0.380	0.180	0.845	0.930	0.800	0.735
AUC-ROC	0.723	0.698	0.957	0.985	0.938	0.753
Stability	0.59	0.39	0.35	0.08	0.57	0.48
Max # Features	20	20	20	20	20	20
TP %	57.0%	42.5%	47.5%	24.0%	53.0%	53.0%
Model Derived						
Method	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	0.475	0.208	0.547	0.244	0.655	0.931
Accuracy	0.402	0.165	0.915	0.950	0.950	0.895
AUC-ROC	0.698	0.651	0.959	0.991	0.994	0.770
Stability	0.58	0.28	0.39	0.14	0.50	0.97
Max # Features	357	207	100	252	226	991
TP %	4.7%	6.7%	14.7%	3.7%	6.4%	2.0%
Ensemble						
Method	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	0.393	0.351	0.626	0.587	0.636	0.603
Accuracy	0.370	0.280	0.865	0.925	0.795	0.760
AUC-ROC	0.618	0.639	0.979	0.977	0.920	0.768
Stability	0.42	0.47	0.49	0.43	0.53	0.50
Max # Features	20	20	20	20	20	20
TP %	55.5%	47.5%	55.5%	53.5%	55.5%	53.5%

Table 11 – Results from High Sample (50/group) MS-scale Multi-Class Classification Simulations. RPT – Robustness-Performance Trade-off, AUC-ROC – Area under the Receiver Operator Curve

MS - MULTI-CLASS - DISCRIMINATORY - LOW SAMPLE						
Subset						
Method	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	0.378	0.131	0.352	0.113	0.568	0.618
Accuracy	0.294	0.088	0.875	0.900	0.800	0.763
AUC-ROC	0.721	0.728	0.975	0.979	0.963	0.704
Stability	0.53	0.26	0.22	0.06	0.44	0.52
Max # Features	20	20	20	20	20	20
TP %	45.5%	18.0%	27.0%	14.0%	34.0%	38.5%
Model Derived						
Method	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	0.424	0.253	0.631	0.164	0.661	0.918
Accuracy	0.317	0.238	0.888	0.913	0.938	0.863
AUC-ROC	0.701	0.767	0.954	0.975	1.000	0.844
Stability	0.64	0.27	0.49	0.09	0.51	0.98
Max # Features	382	191	100	159	108	989
TP %	3.4%	4.6%	10.1%	4.1%	7.9%	2.0%
Ensemble						
Method	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	0.306	0.222	0.629	0.414	0.600	0.585
Accuracy	0.313	0.163	0.913	0.888	0.800	0.725
AUC-ROC	0.668	0.756	0.992	1.000	0.971	0.736
Stability	0.30	0.35	0.48	0.27	0.48	0.49
Max # Features	20	20	20	20	20	20
TP %	34.0%	27.5%	33.0%	30.5%	29.5%	38.0%

Table 12 – Results from Low Sample (25/group) MS-scale Multi-Class Classification Simulations. RPT – Robustness-Performance Trade-off, AUC-ROC – Area under the Receiver Operator Curve

Eisner Results

Non-ensemble Results

	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	.657	.430	.457	.331	.705	.711
Accuracy	.757	.557	.871	.957	.786	.771
AUC-ROC	.685	.773	.933	.975	.958	.744
Stability	.58	.35	.31	.20	.64	.66

Table 13 - Non-ensemble model diagnostics of Eisner dataset

Output from feature.table

PAM Feature Table

features	consistency	frequency
Adipate	10	1
Creatine	10	1
Glucose	10	1
Succinate	10	1
X3.Hydroxyisovalerate	10	1
myo.Inositol	9	0.9
Betaine	7	0.7
Glutamine	7	0.7
Quinolinate	6	0.6
cis.Aconitate	6	0.6
Acetate	5	0.5
N.N.Dimethylglycine	5	0.5
Lysine	3	0.3
Leucine	2	0.2

Table 14 – Eisner PAM feature table

GLMNET Feature Table

features	consistency	frequency
Adipate	10	1
Glucose	10	1
Leucine	10	1
Quinolinate	10	1
Valine	10	1
X3.Hydroxyisovalerate	10	1
myo.Inositol	9	0.9
Succinate	6	0.6
Betaine	4	0.4
Glutamine	4	0.4
Lysine	4	0.4
Creatine	3	0.3
N.N.Dimethylglycine	3	0.3
Acetate	3	0.3
Alanine	2	0.2
Formate	1	0.1
Xylose	1	0.1

Table 15 – Eisner GLMNET feature table

Ensemble Results

	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	.696	.486	.612	.433	.572	.721
Accuracy	.736	.529	.743	.957	.686	.743
AUC-ROC	.802	.717	.892	1.0	.967	.790
Stability	.66	.45	.52	.28	.49	.70

Table 16 - Ensemble model diagnostics of Eisner dataset

Output from feature.table

PAM Feature Table

features	consistency	frequency
Adipate	10	1
Creatine	10	1
Glucose	10	1
Succinate	10	1
X3.Hydroxyisovalerate	10	1
myo.Inositol	9	0.9
N.N.Dimethylglycine	9	0.9
Quinolinate	9	0.9
Acetate	7	0.7
Glutamine	5	0.5
Betaine	3	0.3
Leucine	2	0.2
Sucrose	2	0.2
cis.Aconitate	2	0.2
X3.Hydroxybutyrate	1	0.1
Alanine	1	0.1

Table 17 – Eisner PAM ensemble feature table

PLSDA Feature Table

features	consistency	frequency
Glucose	10	1
Leucine	10	1
Quinolinate	10	1
Valine	10	1
X3.Hydroxyisovalerate	10	1
Adipate	9	0.9
Succinate	9	0.9
myo.Inositol	7	0.7
N.N.Dimethylglycine	7	0.7
Creatine	5	0.5
Glutamine	4	0.4
X3.Hydroxybutyrate	2	0.2
Betaine	2	0.2
Methylamine	1	0.1
Alanine	1	0.1
Pyroglutamate	1	0.1
Acetate	1	0.1
Formate	1	0.1

Table 18 – Eisner GLMNET ensemble feature table

Ametaj Results

Non-ensemble Results

	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	.597	.327	.711	.383	.806	.850
Accuracy	.488	.300	.975	.950	.900	.850
AUC-ROC	.775	.796	1.0	1.0	1.0	.875
Stability	.77	.39	.56	.24	.73	.85

Table 19 - Non-ensemble model diagnostics of Ametaj dataset

Output from feature.table()

PAM Feature Table

features	consistency	frequency
Alanine	10	1
Dimethylamine	10	1
Endotoxin	10	1
Ferulate	10	1
Fumarate	10	1
Glucose	10	1
Methylamine	10	1
Uracil	10	1
Xanthine	8	0.8
Lactate	7	0.7
Lysine	3	0.3
Propionate	1	0.1
Valerate	1	0.1

Table 20 – Ametaj PAM Feature Table

GLMNET Feature Table		
features	consistency	frequency
Acetate	10	1
Endotoxin	10	1
Glucose	10	1
Lactate	10	1
Methylamine	10	1
Propionate	10	1
Uracil	10	1
Lysine	7	0.7
Phenylacetate	7	0.7
Butyrate	5	0.5
Dimethylamine	4	0.4
Alanine	3	0.3
Xanthine	2	0.2
Isoleucine	1	0.1
X3.PP	1	0.1

Table 21 – Ametaj GLMNET Table

PLSDA Feature Table

features	consistency	frequency
Acetate	10	1
Glucose	10	1
Isoleucine	10	1
Lactate	10	1
Methylamine	10	1
Benzoate	9	0.9
Endotoxin	9	0.9
Fumarate	9	0.9
Aspartate	8	0.8
Ferulate	7	0.7
Alanine	3	0.3
Dimethylamine	2	0.2
X3.PP	1	0.1
Butyrate	1	0.1
Propionate	1	0.1

Table 22 – Ametaj PLSDA Feature Table

Ensemble Results

	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	.575	.364	.813	.638	.813	.825
Accuracy	.438	.250	.950	.950	.950	.875
AUC-ROC	.779	.750	1.0	1.0	1.0	.892
Stability	.84	.67	.71	.48	.71	.78

Table 23 - Ensemble model diagnostics of Ametaj dataset

Output from feature.table

PAM Feature Table

features	consistency	frequency
Alanine	10	1
Dimethylamine	10	1
Endotoxin	10	1
Ferulate	10	1
Fumarate	10	1
Glucose	10	1
Methylamine	10	1
Uracil	10	1
NDMA	6	0.6
Lactate	5	0.5
Propionate	3	0.3
Hypoxanthine	3	0.3
Xanthine	2	0.2
Glycerol	1	0.1

Table 24 – Ametaj PAM Ensemble Table

SVM Feature Table

features	consistency	frequency
Endotoxin	10	1
Formate	10	1
Glucose	10	1
Methylamine	10	1
Uracil	10	1
Ferulate	9	0.9
Fumarate	8	0.8
Histidine	8	0.8
X3.HP	7	0.7
Isovalerate	6	0.6
Maltose	6	0.6
X1.3.D	3	0.3
Dimethylamine	2	0.2
X3.PP	1	0.1

Table 25 – Ametaj SVM Ensemble Table

PLSDA Feature Table

features	consistency	frequency
Alanine	10	1
Dimethylamine	10	1
Endotoxin	10	1
Glucose	10	1
Methylamine	10	1
Uracil	10	1
Valine	10	1
Leucine	9	0.9
NDMA	9	0.9
Glycerol	7	0.7
Propionate	2	0.2
Cadaverine	2	0.2
Xanthine	1	0.1

Table 26 – Ametaj PLSDA Ensemble Table

GLMNET Feature Table

features	consistency	frequency
Acetate	10	1
Alanine	10	1
Butyrate	10	1
Isovalerate	10	1
X3.HP	10	1
X3.PP	10	1
Propionate	9	0.9
Benzoate	7	0.7
Phenylacetate	7	0.7
Lysine	7	0.7
Isoleucine	3	0.3
X3.HB	2	0.2
Glucose	1	0.1
X1.3.D	1	0.1
Endotoxin	1	0.1
Isobutyrate	1	0.1
Lactate	1	0.1

Table 27 – Ametaj GLMNET Ensemble Table

Xiao Results

Negative Mode

1fv2f

	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	.579	.298	.481	.163	.606	.679
Accuracy	.589	.511	.889	.889	.822	.678
AUC-ROC	.597	.608	.855	.865	.835	.610
Stability	.57	.21	.33	.09	.48	.68

Table 28 – Xiao Negative 1fv2f Model Diagnostics

1fv2r

	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	.536	.224	.390	.180	.626	.662
Accuracy	.607	.444	.889	.900	.867	.644
AUC-ROC	.578	.645	.805	.870	.885	.603
Stability	.48	.15	.25	.10	.49	.68

Table 29 – Xiao Negative 1fv2r Model Diagnostics

1rv2f

	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	.635	.258	.514	.197	.680	.679
Accuracy	.663	.456	.967	.933	.867	.678
AUC-ROC	.622	.659	.950	.925	.865	.724
Stability	.61	.18	.35	.11	.56	.68

Table 30 – Xiao Negative 1rv2f Model Diagnostics

1rv2r

	PLSDA	GBM	SVM	RF	GLMNET	PAM
RPT	.665	.294	.455	.229	.665	.746
Accuracy	.650	.444	.944	.956	.867	.733
AUC-ROC	.590	.625	.920	.915	.855	.681
Stability	.68	.22	.30	.13	.54	.76

Table 31 – Xiao Negative 1rv2r Model Diagnostics

Positive Mode

1fv2f

	PLSDA	**GBM**	**SVM**	**RF**	**GLMNET**	**PAM**
RPT	.453	.267	.574	.147	.646	.643
Accuracy	.522	.300	.956	.900	.911	.667
AUC-ROC	.625	.651	.895	.855	.865	.605
Stability	.40	.24	.41	.08	.50	.62

Table 32 – Xiao Positive 1fv2f Model Diagnostics

1fv2r

	PLSDA	**GBM**	**SVM**	**RF**	**GLMNET**	**PAM**
RPT	.495	.266	.548	.164	.612	.610
Accuracy	.583	.444	.922	.922	.844	.656
AUC-ROC	.659	.645	.805	.845	.830	.660
Stability	.43	.19	.39	.09	.48	.57

Table 33 – Xiao Positive 1fv2r Model Diagnostics

1rv2f

	PLSDA	**GBM**	**SVM**	**RF**	**GLMNET**	**PAM**
RPT	.546	.298	.600	.164	.686	.658
Accuracy	.633	.511	.944	.922	.911	.656
AUC-ROC	.660	.600	.925	.910	.930	.6625
Stability	.48	.21	.44	.09	.55	.66

Table 34 – Xiao Positive 1rv2f Model Diagnostics

1rv2r

	PLSDA	**GBM**	**SVM**	**RF**	**GLMNET**	**PAM**
RPT	.573	.403	.560	.181	.694	.737
Accuracy	.637	.544	.933	.933	.911	.733
AUC-ROC	.562	.614	.930	.905	.905	.688
Stability	.52	.32	.40	.10	.56	.74

Table 35 – Xiao Positive 1rv2r Model Diagnostics

Number Features Identified

PAM

Mode	Comparison	# Ions Selected
Positive	1F v. 2F	157
Positive	1F v. 2R	154
Positive	1R v. 2F	154
Positive	1R v. 2R	136
Negative	1F v. 2F	100
Negative	1F v. 2R	104
Negative	1R v. 2F	86
Negative	1R v. 2R	74

Table 36 – Xiao PAM Number of Features Identified by Comparison

GLMNET

Mode	Comparison	# Ions Selected
Positive	1F v. 2F	191
Positive	1F v. 2R	177
Positive	1R v. 2F	189
Positive	1R v. 2R	185
Negative	1F v. 2F	153
Negative	1F v. 2R	136
Negative	1R v. 2F	113
Negative	1R v. 2R	114

Table 37 – Xiao GLMNET Number of Features Identified by Comparison

APPENDIX C

Response to Shock (S45-B) VIP Scores			
Metabolite	**VIP Score**	**Metabolite**	**VIP Score**
Glucose	3.301	Hypoxanthine	0.338
Lactate	2.237	Nacinamide	0.334
Aspartate	2.042	Betaine	0.332
Maltose	1.993	Pyruvate	0.319
Sucrose	1.967	Histidine	0.317
Glutathione	1.4122	Glycine	0.290
Xanthine	1.391	Lysine	0.264
Asparagine	1.375	Leucine	0.214
Adenosine	1.296	NADP	0.196
3-Hydroxyisovalerate	1.257	3-Aminoisobutyrate	0.161
O-Phosphocholine	1.173	Tyrosine	0.091
Sn-Glycero-3-phosphocholine	1.162	Glutamine	0.078
Fumarate	1.087	Choline	0.076
Succinate	1.052	Creatine	0.072
Valine	0.997	ATP	0.063
Dimethylamine	0.927	Isovalerate	0.047
AMP	0.876	Acetate	0.027
Isoleucine	0.811	Proline	0.026
Glutamate	0.785	UDP-Glucose	0.012
3-Hydroxybutyrate	0.764		
Serine	0.702		
Citrate	0.631		
Phenylalanine	0.626		
Alanine	0.621		
Arginine	0.5852		
Taurine	0.555		
Methionine	0.5476		
ADP	0.537		
Formate	0.4947		
Creatinine	0.489		
Benzoate	0.414		
NAD	0.352		

Table 1 - Complete Metabolite VIP Scores during the response to shock (S45-B) comparing Fasted and Carbohydrate Prefed pigs.

Response to Resuscitation (FR2-S45) VIP Scores			
Metabolite	**VIP Score**	**Metabolite**	**VIP Score**
Glucose	3.311	Hypoxanthine	0.388
3-Hydroxyisovalerate	2.083	3-Aminoisobutyrate	0.325
Valine	1.659	Xanthine	0.318
Fumarate	1.658	Tyrosine	0.294
Benzoate	1.639	Sucrose	0.270
Leucine	1.590	Glycine	0.268
Choline	1.367	Succinate	0.258
Isoleucine	1.364	Taurine	0.2534
Lysine	1.356	Maltose	0.242
Sn-Glycero-3-Phosphocholine	1.355	ATP	0.238
Arginine	1.309	Asparagine	0.207
Niacinamide	1.306	Histidine	0.190
Adenosine	1.242	3-Hydroxybutyrate	0.168
Aspartate	1.111	Glutamate	0.157
Formate	1.0754	Betaine	0.152
Alanine	0.998	Phenylalanine	0.150
Proline	0.938	Dimethylamine	0.121
Serine	0.896	Isovalerate	0.036
Creatine	0.835	AMP	0.020
O-Phosphocholine	0.808		
Glutathione	0.802		
Glutamine	0.779		
NADP	0.763		
NAD	0.722		
Creatinine	0.688		
Citrate	0.687		
Lactate	0.637		
Acetate	0.606		
UDP-Glucose	0.600		
ADP	0.5906		
Pyruvate	0.467		
Methionine	0.418		

Table 2 - Complete Metabolite VIP Scores during the response to resuscitation (FR2-S45) comparing Fasted and Carbohydrate Prefed pig