A METABOLOMIC INVESTIGATION INTO \textit{CLOSTRIDIUM DIFFICILE} INFECTION

KARISSA THAL
SPRING 2015

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree in Biology
with honors in Biology

Reviewed and approved* by the following:

Andrew Patterson
Assistant Professor of Molecular Toxicology
Thesis Supervisor

Stephen W. Schaeffer
Professor of Biology
Honors Adviser

* Signatures are on file in the Schreyer Honors College.
ABSTRACT

*Clostridium difficile* (*C. difficile*) is the leading cause of healthcare-associated diarrhea in North America and Europe. The infection has been increasing steadily in incidence, severity, and mortality for the past decade due to the development of the epidemic strain BI/NAP1/027. Currently, a biomarker for *C. difficile* infection (CDI) does not exist, but could significantly improve early diagnosis of the infection, thus reducing morbidity and mortality. This study featured a comparative analysis of the urinary metabolic profiles of control subjects to those with CDI. It was hypothesized that differences in the urinary CDI metabolite profile, urinary endogenous metabolite profile, or both, would be detectable by metabolomics, and these differences will be of great potential value in identifying a biomarker of CDI. Comparison of the ultra high pressure liquid chromatography coupled with electrospray ionization quadruple time-of-flight mass spectrometry results between the Surgical Control and CDI groups revealed two endogenous compounds, *p*-cresol sulfate and indoxylsulfate, to be significantly elevated in the CDI group. However, this result was not consistent in the comparison between the Control and CDI groups. *P*-cresol sulfate and indoxylsulfate were therefore not conclusively implicated to be biomarkers of CDI. The results of $^1$H nuclear magnetic resonance spectroscopy revealed several endogenous compounds that were either significantly elevated or depressed in the CDI group relative to the Controls. However, due to the heavy antibiotic dosage that the CDI group was exposed to and the altered gut microbiota, we could not conclusively attribute a differential metabolite to be an endogenous biomarker of CDI.
# TABLE OF CONTENTS

LIST OF FIGURES .......................................................................................................................... iv

LIST OF TABLES ............................................................................................................................. v

ACKNOWLEDGEMENTS ................................................................................................................ vi

Chapter 1 Introduction .................................................................................................................. 1

  Introduction to Clostridium Difficile ......................................................................................... 1
  Introduction to Metabolomics .................................................................................................... 8

Chapter 2 Methods and Materials ............................................................................................... 11

  Human Volunteers ..................................................................................................................... 11
  $^1$H NMR Sample Preparation and Analysis .......................................................................... 14
  Spectral Data Processing and Multivariate Data Analysis ....................................................... 14
  UPLC-ESI-QTOFMS Sample Preparation and Analysis ......................................................... 16
  Chemometric Data Processing ................................................................................................. 17
  Metaboanalyst Software ........................................................................................................... 18

Chapter 3 Results ......................................................................................................................... 20

  $^1$H NMR Analysis Results ..................................................................................................... 20
  LC-MS Analysis Results ........................................................................................................... 27
  Metaboanalyst Results ............................................................................................................. 35

Chapter 4 Discussion ................................................................................................................... 38

  NMR Results Discussion ....................................................................................................... 39
  LC-MS Results Discussion ..................................................................................................... 43

Appendix A  List of Abbreviations ............................................................................................... 50

Appendix B  Medication Codes .................................................................................................... 51

BIBLIOGRAPHY ............................................................................................................................ 55
LIST OF FIGURES

Figure 1 Overview of the methods used in this study..................................................13
Figure 2: Markerview Parameters..............................................................................18
Figure 3: S-Line Plot of 1H NMR Data......................................................................20
Figure 4: OPLS-DA Plot of 1H NMR Data.................................................................23
Figure 5: 3-Dimensional PCA plot of 1H NMR Data.................................................25
Figure 6: 2-Dimensional PCA Plot of 1H NMR Data.................................................26
Figure 7: Markerview PCA Plots of Propofol Glucuronide from LC-MS Data..............27
Figure 8: Peakview MS and MS/MS of Propofol Glucuronide from LC-MS Data........28
Figure 9: Markerview PCA Plots of Phenylalanine from LC-MS Data..........................29
Figure 10: Peakview MS and MS/MS of Phenylalanine from LC-MS Data..................30
Figure 11: Markerview PCA Plots of p-Cresol sulfate from LC-MS Data......................31
Figure 12: Peakview MS and MS/MS of p-Cresol sulfate from LC-MS Data...............32
Figure 13: Markerview PCA Plots of Indoxylsulfate from LC-MS Data........................33
Figure 14: Peakview MS and MS/MS of Indoxylsulfate from LC-MS Data..................34
Figure 15: Volcano Plot from Metaboanalyst...............................................................35
Figure 16: Random Forest Plot from Metaboanalyst....................................................36
LIST OF TABLES

Table 1: Subject Demographics .................................................................................................................. 12
Table 2: Summary of Differential Metabolites ................................................................................................. 39
ACKNOWLEDGEMENTS

I would like to first thank my thesis advisor Dr. Andrew Patterson for allowing me to be a part of his lab for the past three and a half years. The knowledge that I have gained about metabolomics and its applications to translational medicine will likely be of great value to me in my future career as a physician. I would also like to thank him for his guidance and continued patience with me in the process of writing this thesis.

Other members of the Patterson lab that I would like to thank include Dr. Philip Smith for all his help with data analysis, and more importantly, for always bringing a smile to my face. Dr. Limin Zhang proved to be an invaluable asset for this project, and the NMR analysis would not have been possible without his expertise.

Next, I would like to thank Drs. David Stewart, John Hegarty, and William Sangster for allowing me to be a part of their clinical study, and for answering any questions that I had along the way.

I would also like to thank the Penn State Office of Undergraduate Education for awarding me with a 2104 Summer Discovery Grant, which allowed me to complete this project at a time when I could invest all of my energy into it. I would like to thank the Eberly College of Science for generously increasing the value of the grant.

I would like to thank my honors advisor, Dr. Stephen Schaeffer, for his support all throughout my college career.

Finally, I would like to thank Dr. James Strauss for being like a dad to me.
Chapter 1
Introduction

Introduction to Clostridium Difficile

*Clostridium difficile* (*C. difficile*) is a pathogenic, spore-forming bacterium that is a rising public health threat.\(^1\) This gram-positive, anaerobic bacillus is the leading cause of healthcare-related diarrhea in North America and Europe.\(^2,3\) *C. difficile* is the cause of approximately 15,000-20,000 deaths per year in the US and is responsible for an estimated 1 billion dollars in US health care expenditures.\(^4\) The infection rate has been steadily climbing since 2001 and its severity has also increased due to the development of hypervirulent and antibiotic resistant strains.\(^5\) The hypertoxic strain BI/NAP1/027 (restriction nuclease pattern BI, North American pulsovar 1, PCR ribotype 027), has largely been responsible for the recent epidemic status of *C. difficile*. A ribotype is a molecular genetic profile of ribosomal genes, and the technique of ribotyping is useful for differentiating between different strains of bacteria. It was found that individuals infected with this outbreak strain were more than twice as likely to die within 30 days of the onset of infection compared to individuals who were infected with strains of other ribotypes.\(^6\)

CDI is commonly associated with the use of antibiotics because the antibiotics kill-off beneficial bacterial that normally inhabit the gut and allow *C. difficile* to cause disease. The greatest risk is associated with broad-spectrum antibiotic treatments such as clindamycin and ciprofloxacin.\(^7,8\) Clindamycin is a member of the lincosamide class of
antibiotics, which kill microbes by interfering with bacterial protein synthesis. Specifically, clindamycin binds to the large ribosomal subunit in a site that causes inhibition of the peptidyl transferase reaction. Ciprofloxacin, a member of the fluoroquinolone class of antibiotics, combats microbes by preventing cell division. The target of ciprofloxacin is DNA gyrase, an essential bacterial enzyme that catalyzes the supercoiling of DNA. By inhibiting DNA gyrase and facilitating double-stranded DNA breaks, ciprofloxacin ultimately promotes lysis of the bacterium.

In addition to killing off disease-causing bacteria that inhabit the gut, antibiotics also incidentally disrupt the normal gut ecology by killing off health-promoting microflora. Bacterial species belonging to the Bacteroidetes and Firmicutes phyla comprise the vast majority of the microbiota community in the colon. While the colon of the human host provides a protective and nutrient-rich environment for the microbiota, the Bacteroidetes perform several functions that are mutually beneficial to the host. Fermentation of carbohydrates by Bacteroidetes yields a variety of volatile fatty acids that are subsequently reabsorbed by the colon and used as an energy source by the host. Bacteroidetes role in carbohydrate metabolism likely explains why the microbiota of obese individuals is comprised of decreased proportions of this bacterial phylum. Certain strains of Bacteroidetes were found to direct the expression of genes that modulate a variety of gut functions including nutrient absorption, mucosal barrier reinforcement, and anigogenic factor production. Additionally, Bacteroidetes were found to be a necessary factor in mediating the generation of a normal, mature immune system. Specifically, the species B. fragilis is responsible for the activation of the T-cell dependent immune response, thereby minimizing the entrance and proliferation of
The species *B. thetaiotaomicron* is known to stimulate the production of antibacterial peptides by Paneth cells that are thought to prevent the colonization of pathogens. A species belonging to the genus Bacteroides is thought to play a role in the prevention of development of CDI. A study comparing gut ecology in patients with recurrent CDI to healthy controls found a significant decrease in the proportion of Bacteroidetes, with prominent increases in the proportions of bacterial phyla that are normally only minor components of the gut microflora in patients with CDI. The drop and even disappearance of Bacteroidetes allowed for the proliferation of bacteria belonging to Proteobacteria and Verrucomicrobia phyla. It was also found that overall bacterial species diversity within each phylum was markedly lower in patients with recurrent CDI. The decreased presence of the normal populations of microflora allow *C. difficile* to colonize and proliferate in the colon.

Just as the presence of Bacteroidetes in the colon benefits the human host, Firmicutes function is to produce butyrate, a compound that yields the preferential energy source of colonocytes, and also assist in carbohydrate breakdown. Other species of Firmicutes, particularly those belonging to Clostridial clusters IV and XIVa, function to stimulate the absorption of water and sodium, decrease colonic pH, inhibit the growth of pathogens and aid in cholesterol synthesis. The species *Faecalibacterium prausnitzii*, a prominent member of Firmicutes, is known to have anti-inflammatory properties. Jump et al. reported that administration of isolates from the *Lachnospiraceae* bacterial family of cluster XIVa demonstrated to be effective in restoring colonization resistance to *C. difficile* in mice.
In addition to antibiotics, the use of proton pump inhibitors (PPIs) can increase susceptibility to CDI. PPIs work by inhibiting the H^+/K^+ ATPase pump in parietal cells of the gastrointestinal tract, which causes an increase in stomach pH. Jump et al. reported that the non-soporific, or vegetative form of *C. difficile*, is better able to survive in the presence of decreased stomach acidity in human patients.\textsuperscript{19}

*C. difficile* spreads via the fecal-oral route, which involves the ingestion of spores via the mouth. While the vegetative form of *C. difficile* requires moist surfaces to survive, *C. difficile* spores are capable of lying dormant on dry surfaces for periods of months, which increases the opportunity for transmission into a host.\textsuperscript{5,19} It is thought that the primary means of transmission in non-outbreak periods is the temporarily contaminated hands of healthcare workers.\textsuperscript{5} Eradicating *C. difficile* spores from surfaces in healthcare facilities such as bedrails and call buttons is complicated by its resistance to common disinfectants.\textsuperscript{5} However, disinfectant solutions containing 1:10 diluted sodium hypochlorite have been found to be effective in killing spores.\textsuperscript{2}

Once established in the gut, *C. difficile* secretes two principle protein toxins, Toxin A and Toxin B, which destroy the epithelial cells lining the colon reducing water uptake and resulting in diarrhea.\textsuperscript{20} Toxins A and B are monoglucosyltransferases and function to glycosylate Rho GTPases, which govern many cellular functions, such as immune cell signaling and migration, cytokine production, and epithelial barrier function.\textsuperscript{21} Glycosylation of the Rho GTPases by Toxins A and B inactivates these proteins, thereby blocking all downstream signaling pathways and proves to be cytotoxic to involved cells.\textsuperscript{21} Specifically, Toxin A induces harm to colonic cells by causing the epithelium to secrete an inflammatory fluid, while the mechanism of Toxin B involves inference with
the actin component of the cytoskeleton, which has a cytotoxic effect.\textsuperscript{22} \textit{C. difficile} strains that produce one or both of these toxins are virulent.\textsuperscript{3} Increased levels of these pro-inflammatory toxins are correlated with increased virulence of CDI.\textsuperscript{20}

The elderly are also a highly susceptible population due to an age-related decline in immune functioning.\textsuperscript{8} Previous studies have estimated the incubation period of CDI to be 2-3 days.\textsuperscript{5} The classic clinical manifestation of CDI is the occurrence of diarrhea, which is defined as the movement of loose, watery stools three or more times in a period of 24 hours or less. Other common symptoms include bloody stools, abdominal discomfort and cramping, fever and weight loss.\textsuperscript{5} Severe forms of the infection are characterized by the development of fulminant \textit{Clostridium difficile} colitis (FCDC), toxic megacolon and perforation.\textsuperscript{23} In addition to acute inflammation of the colon, individuals with FCDC can present with hypotension, fever, leukocytosis, oliguria and azotemia.\textsuperscript{24} Although FDCD presents in only 3-8% of CDI cases, it has a mortality rate of 30-90%.\textsuperscript{22,24} Toxic megacolon is characterized by the dilation of colonic diameter by at least 5 cm, in addition to systemic toxicity.\textsuperscript{25,26} FDCD and toxic megacolon each increase the likelihood of perforation, which results in the spillage of colonic contents in the abdominal cavity, a phenomenon that often causes peritonitis. If not promptly treated, the infection can enter the bloodstream. Sepsis and multiple organ dysfunction syndrome (MODS) ultimately cause death.\textsuperscript{26}

One challenge that clinicians face when treating \textit{C. difficile} is combating its rate of recurrence, which is estimated to be between 15-20%.\textsuperscript{27} Traditionally, \textit{C. difficile} is treated with antimicrobials such as vancomycin and metronidazole.\textsuperscript{4} Vancomycin is combative against gram-positive bacterial pathogens like \textit{C. difficile}, as it functions to
inhibit peptidoglycan synthesis, the major component of the microbe cell walls. Specifically, vancomycin binds to the C-terminal dipeptide D-alanyl-D-alanine of the muramyl-pentapeptide moiety of peptidoglycan precursors, thereby interfering with synthesis.\textsuperscript{28} Metronidazole combats CDI with a different mechanism of action than that of vancomycin. Inactive metronidazole passively diffuses into bacterial cells and becomes activated in the cytoplasm, or in various organelles within the bacterium. Once activated, metronidazole undergoes intracellular reduction, which temporarily converts it to a nitroso free radical.\textsuperscript{29} Now cytotoxic, the metronidazole inhibits DNA synthesis as well as causing both single and double-stranded breaks in the bacterial DNA, which leads to cell death.\textsuperscript{29}

However, resistant spores can germinate into vegetative forms after a patient completes a course of antimicrobials. The technique of tapering antimicrobial dosage has yielded slight success in reducing recurrence of infection.\textsuperscript{4} More notably, a recent study found that the use of monoclonal antibody (MOA) therapies reduced recurrent infections to a quarter of the rate experienced by patients being treated with the aforesaid antimicrobials.\textsuperscript{30} The injected MOAs worked by binding to the antigens on the surface of the principle toxins emitted by \textit{C. difficile}.\textsuperscript{30}

Even more impressive results are associated with Fecal Microbiota Transplantation (FMT), which is proving to be the most promising method of treatment.\textsuperscript{4} This technique involves the transfer of feces from a healthy donor to a patient with CDI. The fecal matter from the donor promotes the colonization of the normal microflora in the gut of the recipient, thereby restoring the colonic bacterial composition that had been disrupted by the consumption of antibiotics. In addition to reported cure rates of 96%,
FMT is considered to be a favorable treatment for recurrent CDI because donor fecal content is both readily available and inexpensive. Cure rates have been noted to be slightly higher in cases where the recipient and donor are closely related, either genetically or intimately (eg. spouse). Although no adverse effects have been directly attributed to FMT, one reported downside of this treatment is rapid, unintended weight gain in the recipient, which has been documented in both human and animal models. FMT triggered obesity was found in cases in which the fecal donor had an obese microbiota.

Impact: The absence of an existing biomarker of CDI, or a metabolite that is indicative of the biological processes associated with CDI, limits the ability of clinicians to diagnose, and therefore properly treat the infection. Early diagnosis is key to a positive prognosis, as CDI is known to progress to colitis, toxic megacolon, and possibly even death. Additionally, failure to promptly diagnose CDI can result in an increased risk of the widespread development of cases in a community setting such as a hospital or nursing home. Discovery of such a biomarker would thus be of great value in early detection and treatment of CDI, thereby reducing its overall morbidity and mortality.

Hypothesis: I hypothesize that urinary metabolite profiles will be useful in distinguishing CDI from normal, healthy controls. Further, these metabolites may be potential diagnostic biomarkers of CDI. Urine is the concentrated output of many metabolic processes and therefore represents a potentially rich source of biomarkers.
Introduction to Metabolomics

Metabolomics may be considered the terminal part of the omics cascade (genomics, transcriptomics, proteomics) representing the final output of cellular responses to endogenous and exogenous stimuli. The metabolome, the collective composition of all the small molecule metabolites present in an organism, reflects the environmental influences, disease state, and genetic changes that affect organisms.\textsuperscript{34} Essentially, metabolomics is the study of the identification and quantification of the chemical fingerprints that are the residue of biological processes. Tracking the presence of certain metabolites and their levels can reveal insight into how the organism mechanistically and biologically changes in response to exposure to a disease or toxin. Metabolomics is therefore a useful tool in predicting physiological state.

The idea of using small molecule metabolites to identify and predict disease has been previously established. One prominent example of metabolomics in a clinical setting is the detection of sarcosine as a biomarker of prostate cancer progression. Sarcosine is the N-methyl derivative of the amino acid glycine, and can be detected non-invasively through urine collection.\textsuperscript{35} Through the analysis of metabolic profiles using a combination of Liquid Chromatography-Mass Spectrometry (LC-MS) and Gas Chromatography-Mass Spectrometry (GC-MS), sarcosine was identified as a differential metabolite that was significantly increased both in patients with localized, malignant prostate cancer and those with metastatic cancer, compared to patients with a benign version of the disease.\textsuperscript{35} Inactivation of glycine-N-methyl transferase, the enzyme responsible for converting glycine to sarcosine, resulted in mitigation of prostate cancer progression.\textsuperscript{35} Biomarkers like sarcosine are of great potential value to physicians, as they
can reveal disease state, be used for the prediction of the possible development of comorbidities, and shed light on treatment options.

Similarly, in addition to confirming glucose as a biomarker for diabetes, Suhre et al. utilized metabolomics to identify metabolites with differential concentrations in affected patients and in healthy control subjects. On average, fatty acids of medium length such as caproate, as well as arachidonate, were found to be decreased in the diabetes group, while long chain fatty acids such as desoxyhexose were increased. This suggests that fatty acid length and corresponding abundance, which can be deduced by metabolomics, can be predictive of a diabetic disease-state. Additionally, metabolomics can monitor the success of treatment methods by observing if a metabolic profile shifts back to a healthy state. Metabolomics therefore proposes a significant contribution to the field of individualized medicine by allowing pharmacological treatments to be tailored to metabolic phenotype.

Multiple platforms, such as NMR and mass spectrometry coupled to separate techniques, have been used for metabolomics applications. Advantages of NMR as a platform include the simplicity of sample preparation. For urine in particular, no extraction is necessary. Another advantage of NMR is that it produces peaks that correlate directly and linearly to metabolite abundance. NMR is a non-discriminating and non-destructive technique, and it is also favorable because it has potential for high-throughput fingerprinting. A downside to the use of NMR as a platform is that only metabolites with medium to high abundances are detected, and the identification of chemical shifts in complex mixtures, such as human urine, is difficult.
The platform of mass spectrometry is advantageous in that it offers quantitative analysis with high sensitivity and specificity that allows for coverage and potential identification of a broad range of metabolites. Coupling mass spectrometry with a separation technique such as liquid (LC) or gas chromatography (GC) lessens the complexity of mass spectra due to metabolite separation in a time dimension, and it reveals additional data on the physio-chemical properties of the metabolites. Specific advantages of LC-MS include its capability to reduce ion suppression that is caused by coeluting compounds and to separate isomers.

Sample preparation is more complex with mass spectrometry than with NMR, which could result in a loss of the metabolites. It is also possible that certain classes of metabolites may be discriminated based on the specific introduction system and ionization technique used. Reverse-Phase High Performance Liquid Chromatography (RP-HPLC) is a standard technique used to separate non-polar and medium polar metabolites, as very polar metabolites usually elute with the void volume. Quadruple-Time of Flight mass analyzers are advantageous due to their ability to perform MS/MS experiments, which are useful in determining the structural elucidation of biomarkers. Electrospray Ionization is the ionization technique of choice employed in LC-MS metabolomics. In order to obtain broad range coverage of the metabolome, ionization is done in both positive and negative modes, which allows for the accounting of differences in the physicochemical properties of chemicals found in a sample.
Chapter 2

Methods and Materials

Human Volunteers

Urine samples were collected from study subjects (n = 60) at the Milton S. Hershey Medical Center. All subjects were patients of David Stewart, M.D. at the Colon and Rectal Surgery outpatient clinic. Documented approval from The Penn State College of Medicine Institutional Review Board was obtained and kept on file at the Penn State Milton S. Hershey Medical Center. Urine samples were collected from all subjects in a sterile cup as either a clean catch or from a urinary catheter. Samples were stored at -80°C upon collection and were thawed and aliquoted into 1 ml Eppendorf tubes prior to being shipped to the Metabolomics Facility at University Park. Urine from the CDI subjects (n = 20, 12 males, 8 females, mean age 53.25 years) was collected within 24 hours after the subject was diagnosed with CDI via a rapid C. difficile stool antigen test. Subject demographics are reported in Table 1. Subjects comprising the surgical control group (n=20, 12 males, 8 females, mean age 52.75 years) were randomly selected from a pool of patients who had undergone colon and rectal surgery. Post-operative urine was collected from this group. Urine was collected from a second control group, referred to as controls, of randomly selected subjects (n=20, 8 males, 12 females, mean age 40.05 years) either prior to or at least 4 weeks following their surgical procedure. Subjects who had known renal failure were excluded.
<table>
<thead>
<tr>
<th>Subject</th>
<th>Age</th>
<th>Gender</th>
<th>Race</th>
<th>BMI</th>
<th>Medications</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD1</td>
<td>44</td>
<td>0</td>
<td>0</td>
<td>29</td>
<td>BJ, F, DB, D, TT, CCC, DC, CL, XX, C, E</td>
</tr>
<tr>
<td>CD2</td>
<td>30</td>
<td>1</td>
<td>0</td>
<td>37</td>
<td>G, LLL, XX, D, H, BB</td>
</tr>
<tr>
<td>CD3</td>
<td>66</td>
<td>1</td>
<td>0</td>
<td>40</td>
<td>DE, DF, DG, HH, G, DH, DI, U, BS, AZ, AS, DJ, DK, BB, C</td>
</tr>
<tr>
<td>CD4</td>
<td>60</td>
<td>0</td>
<td>0</td>
<td>38</td>
<td>N, MMM, DL, AAA, DM, CCC, XX, TTT, DN, T, E, C</td>
</tr>
<tr>
<td>CD5</td>
<td>19</td>
<td>0</td>
<td>0</td>
<td>19</td>
<td>DO, F, DO, DP</td>
</tr>
<tr>
<td>CD6</td>
<td>18</td>
<td>1</td>
<td>3</td>
<td>19</td>
<td>BB</td>
</tr>
<tr>
<td>CD7</td>
<td>53</td>
<td>1</td>
<td>0</td>
<td>23</td>
<td>GG, HH, BS, AZ, AW, BB, C</td>
</tr>
<tr>
<td>CD8</td>
<td>54</td>
<td>0</td>
<td>0</td>
<td>24</td>
<td>D, I, J, K, E, L, M, H</td>
</tr>
<tr>
<td>CD9</td>
<td>73</td>
<td>0</td>
<td>0</td>
<td>60</td>
<td>D, N, O, P, Q, R, S, T, U, V, W, X, Y, Z, AA, BB, H</td>
</tr>
<tr>
<td>CD10</td>
<td>88</td>
<td>0</td>
<td>0</td>
<td>32</td>
<td>CC, DD, EE, FF, GG, HH, II, JJ, G, KK, LL, MM, M, NN, OO, PP, H</td>
</tr>
<tr>
<td>CD11</td>
<td>30</td>
<td>1</td>
<td>0</td>
<td>43</td>
<td>QO, RR, SS, TT, UU, C</td>
</tr>
<tr>
<td>CD12</td>
<td>64</td>
<td>1</td>
<td>0</td>
<td>28</td>
<td>D, N, VV, WW, XX, YY, C</td>
</tr>
<tr>
<td>CD13</td>
<td>59</td>
<td>1</td>
<td>0</td>
<td>34</td>
<td>ZZ, AAA, BBB, C</td>
</tr>
<tr>
<td>CD15</td>
<td>85</td>
<td>1</td>
<td>0</td>
<td>20</td>
<td>B, RR, CCC, C</td>
</tr>
<tr>
<td>CD16</td>
<td>50</td>
<td>0</td>
<td>0</td>
<td>21</td>
<td>RR, DDD, EEE, FFF, GGG, CI, XX, X, H</td>
</tr>
<tr>
<td>CD17</td>
<td>24</td>
<td>0</td>
<td>0</td>
<td>38</td>
<td>HHH, III, JJ, HH, KKK, FFF, LLL, XX, C</td>
</tr>
<tr>
<td>CD18</td>
<td>65</td>
<td>1</td>
<td>0</td>
<td>20</td>
<td>MMM, D, N, NNN, OOO, CCC, PPP, OQQ, RRR, C</td>
</tr>
<tr>
<td>CD19</td>
<td>43</td>
<td>0</td>
<td>1</td>
<td>52</td>
<td>SSS, TTT, UUU, VVV, TT, QQ, WWW, H</td>
</tr>
<tr>
<td>CD20</td>
<td>71</td>
<td>1</td>
<td>0</td>
<td>32</td>
<td>OOO, YYY, RR, C</td>
</tr>
<tr>
<td>CD21</td>
<td>69</td>
<td>1</td>
<td>0</td>
<td>37</td>
<td>D, ZZZ, AB, T, CCC, X, F, C</td>
</tr>
<tr>
<td>U1</td>
<td>60</td>
<td>0</td>
<td>0</td>
<td>27</td>
<td>AC, AD, AE, A, AF, DDD, A, AG, II, AH, BB, AI, AJ, AK</td>
</tr>
<tr>
<td>U2</td>
<td>49</td>
<td>0</td>
<td>0</td>
<td>47</td>
<td>CCC, AL, AD, C, AF, DDD, A, AG, III, AM, AH, AI, B, AI, AK</td>
</tr>
<tr>
<td>U3</td>
<td>53</td>
<td>1</td>
<td>0</td>
<td>23</td>
<td>U, AL, AD, C, AM, MI, AA, AN, AD, AF, DDD, A, AG, BB, AH, AI, AI, AK</td>
</tr>
<tr>
<td>U4</td>
<td>34</td>
<td>0</td>
<td>0</td>
<td>39</td>
<td>TTT, AP, AQ, A, AF, DDD, A, AG, II, AA, AK, AH, AI, B, BB</td>
</tr>
<tr>
<td>U5</td>
<td>62</td>
<td>1</td>
<td>0</td>
<td>31</td>
<td>AR, AS, AI, AD, C, AF, DDD, A, AG, BB, BH, AH, AI</td>
</tr>
<tr>
<td>U6</td>
<td>72</td>
<td>1</td>
<td>0</td>
<td>22</td>
<td>D, AL, AD, C, DDD, A, AG, AI, AO, III, AT, BB, AJ</td>
</tr>
<tr>
<td>U7</td>
<td>74</td>
<td>0</td>
<td>0</td>
<td>31</td>
<td>D, N, UUU, OOO, G, U, CCC, AU, AE, AD, C, DDD, A, AG, III, BB, BB, AI, AJ, AI</td>
</tr>
<tr>
<td>U8</td>
<td>70</td>
<td>0</td>
<td>0</td>
<td>37</td>
<td>JJJ, AV, UUU, U, CCC, XX, AA, DU, AW, AE, AD, C, AF, DDD, A, AG, AA, AT, AI, III, BB, AX, AO</td>
</tr>
<tr>
<td>U9</td>
<td>71</td>
<td>1</td>
<td>0</td>
<td>17</td>
<td>F, AT, AZ, BA, C, AF, DDD, A, AG, III, BB, AH, AI, B</td>
</tr>
<tr>
<td>U10</td>
<td>75</td>
<td>0</td>
<td>0</td>
<td>31</td>
<td>D, F, AB, BC, U, CCC, AS, AL, AD, D, DDD, A, AG, III, JJ, AK, B, AH, AI, AI</td>
</tr>
<tr>
<td>U11</td>
<td>45</td>
<td>1</td>
<td>0</td>
<td>25</td>
<td>F, BA, BD, AL, C, AF, DDD, A, AG, BE, AO, BB, BH, AF, BJ, AI</td>
</tr>
<tr>
<td>U12</td>
<td>58</td>
<td>1</td>
<td>0</td>
<td>19</td>
<td>AE, AD, C, AF, DDD, A, AG, BB, III, AO, AT, BB, BH, AI, AJ</td>
</tr>
<tr>
<td>U13</td>
<td>52</td>
<td>1</td>
<td>0</td>
<td>22</td>
<td>YY, C, AF, DDD, AG, AM, BB, AJ, AI, BH</td>
</tr>
<tr>
<td>U14</td>
<td>62</td>
<td>1</td>
<td>0</td>
<td>33</td>
<td>RR, AE, AD, C, AF, DDD, A, AG, III, BB, AT, AI, AH, B, AJ</td>
</tr>
<tr>
<td>U15</td>
<td>51</td>
<td>0</td>
<td>0</td>
<td>28</td>
<td>C, AF, DDD, A, AG, III, B, BB, AH, AJ, AI</td>
</tr>
<tr>
<td>U16</td>
<td>24</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>BI, C, AF, DDD, A, AG, III, B, BB, AH, AI</td>
</tr>
<tr>
<td>U17</td>
<td>37</td>
<td>1</td>
<td>0</td>
<td>22</td>
<td>BJ, BK, BL, DO, D, UUU, CJ, AL, AD, C, AF, DDD, A, AG, III, BB, AH, AI</td>
</tr>
<tr>
<td>U18</td>
<td>37</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>F, FFF, BI, AE, AD, C, AF, DDD, A, AG, DR, BE, AO, BB, BH, AH, AI</td>
</tr>
<tr>
<td>U19</td>
<td>64</td>
<td>1</td>
<td>0</td>
<td>26</td>
<td>BM, BN, TT, AN, YYY, PPP, C, DDD, A, AG, BB, BH, AH, AJ, AI</td>
</tr>
<tr>
<td>U20</td>
<td>55</td>
<td>1</td>
<td>0</td>
<td>23</td>
<td>BO, OOO, XXX, BP, K, C, AF, DDD, A, AG, AA, AO, III, AK, AH, B, AI</td>
</tr>
<tr>
<td>U21</td>
<td>41</td>
<td>1</td>
<td>0</td>
<td>28</td>
<td>CW, CK, CY</td>
</tr>
<tr>
<td>U22</td>
<td>35</td>
<td>0</td>
<td>0</td>
<td>38</td>
<td>BT, R, M, U</td>
</tr>
<tr>
<td>U23</td>
<td>36</td>
<td>0</td>
<td>0</td>
<td>49</td>
<td>E, BU</td>
</tr>
<tr>
<td>U24</td>
<td>76</td>
<td>0</td>
<td>0</td>
<td>28</td>
<td>BV, BW, D, N, SS, BX, BT, T, U, AT, CY, CB, CD</td>
</tr>
<tr>
<td>U25</td>
<td>81</td>
<td>0</td>
<td>0</td>
<td>27</td>
<td>D, CE, T, U, CCC</td>
</tr>
<tr>
<td>U26</td>
<td>35</td>
<td>0</td>
<td>2</td>
<td>25</td>
<td>none</td>
</tr>
<tr>
<td>U27</td>
<td>76</td>
<td>0</td>
<td>0</td>
<td>22</td>
<td>B, CF, BO, CE, BC, FF, HH, II, U, CG</td>
</tr>
<tr>
<td>U28</td>
<td>32</td>
<td>1</td>
<td>0</td>
<td>33</td>
<td>BA, CH</td>
</tr>
<tr>
<td>U29</td>
<td>65</td>
<td>0</td>
<td>0</td>
<td>38</td>
<td>D, TTT, CI, HH, CJ, AR, AJ, CK, CX, YYY, CL</td>
</tr>
<tr>
<td>U30</td>
<td>66</td>
<td>0</td>
<td>0</td>
<td>unknown</td>
<td></td>
</tr>
<tr>
<td>U31</td>
<td>67</td>
<td>0</td>
<td>0</td>
<td>unknown</td>
<td></td>
</tr>
<tr>
<td>U32</td>
<td>32</td>
<td>1</td>
<td>0</td>
<td>27</td>
<td>BT, M</td>
</tr>
<tr>
<td>U33</td>
<td>26</td>
<td>0</td>
<td>0</td>
<td>38</td>
<td>CM, CN, M</td>
</tr>
<tr>
<td>U34</td>
<td>76</td>
<td>0</td>
<td>0</td>
<td>28</td>
<td>D, BC, BN, OOO, CO, GGG, AR, AE, CP, AS, AW, AAA, AY</td>
</tr>
<tr>
<td>U35</td>
<td>26</td>
<td>1</td>
<td>0</td>
<td>26</td>
<td>none</td>
</tr>
<tr>
<td>U36</td>
<td>53</td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>B, E, CR, CS, JJ, HH, BF, AG, CT</td>
</tr>
<tr>
<td>U37</td>
<td>58</td>
<td>1</td>
<td>0</td>
<td>35</td>
<td>B, E, MMM, DDD, CU, BN, OOO, XXX, DS, G, CCC, AZ, AS, RRR, CV</td>
</tr>
<tr>
<td>U38</td>
<td>31</td>
<td>1</td>
<td>0</td>
<td>25</td>
<td>CW, CK, CY</td>
</tr>
<tr>
<td>U39</td>
<td>25</td>
<td>1</td>
<td>0</td>
<td>23</td>
<td>none</td>
</tr>
<tr>
<td>U40</td>
<td>50</td>
<td>1</td>
<td>0</td>
<td>27</td>
<td>B, DT, CZ, BK, DA, CW</td>
</tr>
</tbody>
</table>

Table 1: Subject Demographics

Subject Demographics. CD1-21 = CDI group. U1-20 = Surgical Control Group. U21-40 = Controls. A medication key is found in Appendix A.
Figure 1 Overview of the methods used in this study
1H NMR Sample Preparation and Analysis

1H NMR spectroscopy was used to analyze the urine samples. 550µl of each urine sample was combined with 55 µl of a PBS solution in a 1.5mL microcentrifuge tube. The PBS solution was comprised of 1.5 M K₂HPO₄: Na₂HPO₄ (4:1) which have high solubility and are stable at low temperatures, and contained D₂O and TSP. The mixture was vortexed for 30 seconds with the Denville Scientific Inc. Vortexer 59A, and then centrifuged at 4°C at 10,000xg with the Eppendorf Centrifuge 5430R for 10 minutes. 550µl of the resulting supernatant was transferred to a 5-mm outer-diameter Norell NMR tube.

1H NMR spectra of the urine samples was collected at a temperature of 298°K by a Bruker Avance III 600 MHz spectrometer, which operated at 600.08 MHz and contained a Bruker inverse detection cryogenic probe (Bruker Biospin, Germany). A standard, one-dimensional NMR spectrum was recorded for each of the samples, using the first increment of NOESY pulse sequence (NOESYPR1D). The water signal was suppressed through the application of a continuous wave of irradiation to the water peak during the recycle delay (2 s) and the mixing time (100 ms). For each sample the 90° pulse length was adjusted to approximately 10 µs and 64 transients were collected into 32 K data points for each spectrum with spectral width of 20 ppm.

Spectral Data Processing and Multivariate Data Analysis

Following data collection, spectral data processing and multivariate data analyses were carried out. All of the spectra were referenced to the TSP signal (δ 0.0) and were manually corrected for baseline and phase distortions. AMIX software package (V3.8, Bruker-Biospin,
Germany) was employed to integrate the spectral region δ 0.5-9.5 into regions of equal widths of 0.004 ppm (2.4 Hz). In order to avoid the effects of imperfect water saturation, region δ 4.6-5.20 was excised. Each of the integrated regions was then normalized to the total sum of the spectral integrals to compensate for the overall concentration differences preceding statistical data analysis.

SIMCAP+ software (version 13.0, Umetrics, Sweden) was used to conduct the multivariate data analyses. Principal Component Analysis (PCA) was performed for each of the groups in order to produce an overview of the data dispersal pattern. Appendix A can be referenced for all abbreviations. PCA facilitates the identification of groups of variables that are interrelated via phenomena that cannot be directly observed (ie. The presence of latent variables). It accentuates the important variables from the noise. Identification of outliers and clustering patterns allowed for the assessment of data quality. The PCA plot can be seen in Figure 2.1. Following PCA, Orthogonal Projection to Latent Structures Discriminant Analysis (OPLS-DA) was carried out on the $^1$H NMR data. A 7-fold cross validation technique was used to validate the OPLS-DA plot, and the quality of the plots was assessed by the parameters $R^2_X$ and $Q^2$. $R^2_X$ denotes the total of explained variations and $Q^2$ denotes the plot predictability in relation to its statistical validity. Before the loadings plots were generated, back-scale transformation loadings were performed to facilitate interpretation. The plot was further interpreted by the incorporation of color-coded correlation coefficients of the metabolites that were responsible for the separation. The color-coded plot was generated using an in-house developed script for the MATLAB (version 7.1, The Mathworks Inc.; Natwick, MA). The color of the correlation coefficient denotes the significance of the metabolite contributing to the
separation of the two groups. “Hot” colors (eg. red) denote a higher significance than do “cold” colors (eg. blue). The OPLS-DA plot can be seen in Figure 4.

**UPLC-ESI-QTOFMS Sample Preparation and Analysis**

Ultra High Pressure Liquid Chromatography Coupled with Electrospray Ionization Tandem Mass Spectrometry-Quadruple Time of Flight Mass Spectrometry (UPLC-ESI-QTOFMS) analysis was performed on the urine samples. This is an advantageous analysis method due to its high sensitivity and capability to detect a broad range of metabolites in biofluids such as urine, serum, or those metabolites obtained from tissue extracts. In this study, 100µl of urine was combined with 100µl of a solution comprised of H₂O:MeOH (1:1) with 1 µM chloropropamide standard. The mixture was vortexed for 30 seconds with the Denville Scientific Inc. Vortexer 59A, and the supernatant was collected via centrifugation at 10,000xg and 4°C with the Eppendorf Centrifuge 5430R for 20 minutes. Supernatant was transferred to an 800µl plastic Thermo Scientific crimp vial. 11mm crimp Teflon seals were secured to the vials with an 11mm Agilent Technologies Electronic Crimper. Samples were randomized and stored at -20°C prior to use.

Samples (5µl) were separated by reverse phase HPLC using a Prominence 20 UFLCXR system (Shimadzu, Columbia MD) with a Waters BEH C18 column (100 x 2.1mm 1.7 µm) maintained at 55C and a 20 minute aqueous/acetonitrile gradient, at a flow rate of 250 µl/min. Solvent A was water with 0.1% formic acid, and solvent B was acetonitrile with 0.1% formic acid. The initial conditions were 3% B increasing to 45%B
at 10 min and 75%B at 12 min where it was held until 17.5 min before returning to initial conditions at 18 min. The eluate was delivered into a 5600 (QTOF) TripleTOF using a Duospray™ ion source (all AB SCIEX, Framingham, MA). The capillary voltage was set at 5.5 kV in positive ion mode and 4.0 kV in negative ion mode with a declustering potential of 80V used in both modes. The mass spectrometer was operated in IDA (Information Dependent Acquisition) mode with a 250 ms survey scan from 50 to 1250 m/z, and up to 10 MS/MS product ion scans (100 ms) per duty cycle using a collision energy of 50V with a 20V spread.

**Chemometric Data Processing**

The data was processed using MarkerView software (AB SCIEX), which allowed for the generation of a multivariate data set (a table consisting of m/z values, retention time, and peak areas). The parameters for this software are reported in Figure 3. This data set was subjected to Pareto scaling, a method that augments the significance of ions that are present in low-abundance without notably amplifying the noise. The analysis revealed information pertaining to the retention time and m/z value of ions present in the samples. The product ions were then searched by accurate mass in the online published metabolite database METLIN (https://metlin.scripps.edu/index.php). Metabolites were searched within a tolerance of 30 ppm and according to the ion mode the Markerview data was processed in (positive or negative). In order to further confirm the identification of the metabolites, the collected MS/MS ion mass spectra were compared to those within 20 ppm of the m/z value published in METLIN.
Specific parameters that were applied to the Markerview software program during data analysis. Retention time is the time interval between when the component was injected and when it was detected. Retention time tolerance and Mass Tolerance are the windows for error in assignment of retention time and mass, respectively. Together these parameters mandate strict assignment of components. The Subtraction Offset, Subtraction Mult. Factor, Noise Threshold, Min. Spectral Peak Width and Min. RT Peak Width are quality control filters for the LC-MS.

Metaboanalyst Software

Metaboanalyst 3.0 ([http://www.metaboanalyst.ca/MetaboAnalyst/](http://www.metaboanalyst.ca/MetaboAnalyst/)) was used as a tool to further analyze and interpret the data that was aligned in MarkerView (AB SCIEX). The data was exposed to an Interquartile Range (IQR) filter to remove ineffectual variables. Data was not normalized because the intent was to detect smaller differences between the groups that were anticipated to be caused by endogenous compounds. The data was subjected to Pareto scaling. Univariate analysis was conducted first by using the volcano plot, which is a combination of t-tests and fold change. The volcano plot can be viewed in
Figure 15. The fold change threshold was set to 50 and the P-value was set to 0.001 with unequal group variance, which is useful to detect differences in sample means when the population variances are unknown and unequal. A bar graph, which displayed the original values in each group, and a box and whisker plot, which displayed the normalized values in each group, were provided for each feature above threshold parameters. Features were then searched in METLIN.

Random Forest classification was then used to select prominent features in the data matrix, as can be seen in Figure 16. Random Forest plots show a combination of tree predictors, such that each tree depends on the values of a random vector that was sampled independently and with the same distribution for all trees in a forest. Random Forest, which is a machine learning approach, was used because it generates a more sensitive classification model. 500 trees were grown and 7 predictors were tried for each node. Selected features were ranked by their contribution to classification accuracy, and the top five potential outliers were identified. Features were then search in METLIN.
Chapter 3

Results

$^1$H NMR Analysis Results

Figure 3: S-Line Plot of $^1$H NMR Data

The S-Line Plot for the models from the NMR spectra plotted as ppm vs. colored-coded loadings with correlation coefficients. Metabolites were identified by their chemical shifts (in ppm) along the horizontal axis. As can be seen on the far right of the plot, the color-coded meter ranges from 0.0-1.0. The cut-off value for this data set was 0.4. Any value and peak of its corresponding color above 0.4 is considered to be significant.

Figure 3 depicts the results of the $^1$H NMR Spectroscopy Analysis of the samples. This data was used to determine whether there was a significant difference in the concentration of the named metabolites between the CDI and the Control groups. The S-Line Plot displays the coefficient-coded loading plots for the models from NMR spectra of the urine. As can be seen on the far right of the plot, the color-coded meter ranges from 0.0-1.0. The cut-off value for this data set was 0.4, which is based on the sample size used in a two-sided t-test. Any value and peak of its corresponding color above 0.4 is considered to be significantly different from all other...
metabolites in the urine samples. Thus, all of the labeled compounds are considered to be either significantly elevated or depressed in the CDI group relative to the controls. Hippurate, phenylacetylglucose (PAG), trimethylamine N-oxide (TMAO), Urea, and Creatinine are significantly reduced in the CDI group compared to the Control group. Allantoate, Creatine, Glycine, 2-oxoglutarate, Acetate, Butyrate, and Methylmalonate are present in significantly higher levels in the CDI group than in the Control group.

Hippurate, (PAG), and TMAO are all bacterial metabolites. Acetate and Butyrate are bacterial fermentation products. Glycine is an amino acid. Creatine is associated with muscle contraction. 2-oxoglutarate is a Krebs cycle intermediate. Creatinine is associated with kidney function.

Hippurate is known as a mammalian-microbe co-metabolite because it is the product of metabolism that is shared by the host and the bacterial species. Specifically, it is the product of the conjugation of glycine with benzoate. Since both the gut microbial metabolism and the shared host-microbial metabolism generate unique products, changes in the presence or amount of urinary co-metabolites is therefore indicative of varied metabolism of the gut microbiota. Studies have found that the concentration of hippurate in urine is altered according to composition of the gut microbiome in animal models.

PAG is the end product of the metabolism of the amino acid phenylalanine. Previous studies have reported that increases in PAG levels in urine may be related to alternations in the gut microflora. Doessegger et al. reported that elevated levels of urinary PAG correlated with phospholipidosis, a condition involving the accumulation of excess phospholipids in intracellular spaces.
TMAO is a metabolite of the choline moiety of phosphatidylcholine, which results from gut-microbiota metabolism. Previous studies have shown that increased levels of TMAO in urine are associated with an increased risk in the development of cardiovascular disease. Acetate is a known fermentation product of the lactic acid bacterium *Lactococcus lactis*. Another known fermentation product of the gut microbiome, butyrate is noted to be essential to the proper functioning of the colon. Butyrate is the preferential energy source of colonocytes that comprise the pouch epithelium, and it is hypothesized that deficiencies in butyrate can cause inflammation of the ileoanal pouch. Butyrate, which is produced via the *but* and *buk* pathways, regulates gene expression and differentiation, as well as apoptosis in host cells. Elevated levels of methylmalonate in blood and urine serve as a biomarker for a vitamin B12 deficiency. In the absence of adequate levels of vitamin B12, methylmalonate is not converted into the energy-rich metabolite of succinate acid.

The bacterial species *B. fastidiosus* from the phylum Firmicutes is known to degrade allantoate. The enzyme responsible for the degradation of allantoate to ureidoglycolate and ammonia, allantoate-amidinohydrolase, is thus present in this phylum in the intestinal flora.
Figure 4: OPLS-DA Plot of 1H NMR Data

R²X=0.28, Q²=0.52

OPLS-DA plot depicting the data distribution plotted as t1P vs. t2O. t1P and t2O represent the principal components. The R²X value represents the model’s fit and the Q² value represents the model’s predicative value. Red Squares: Control Group. Blue Diamond: CDI Group.

Figure 4 details the OPLS-DA plot, which was constructed to measure the model fit and reliability. The OPLS-DA scores quantify the deviation of the principal components in the dataset and measure the reliability of the NMR. The R²X score measured the model fit and the
Q² value measured the model’s predictive value. The R²X score was 0.28 and the Q² was 0.52. The nearer each of the respective values is to 1, the better the fit and the more reliable the data is considered to be. With human data, as opposed to simpler animal models, the standard for the values is lower, especially with low sample size. Taking these conditions into consideration, the Q² score is indicative of a favorable predictive value.
Figure 5: 3-Dimensional PCA plot of 1H NMR Data


Figure 5 details the three-dimensional PCA plot. The majority of the data points are plotted on the PC2 vs. PC3 axis, however clear distinction between the three groups was not exhibited.
Figure 6: 2-Dimensional PCA Plot of 1H NMR Data

The two-dimensional PCA plot that displays the distribution of the data plotted component 1 vs. component 2. Red Squares: Control Group. Blue Diamonds: CDI Group.
LC-MS Analysis Results

Figure 7: Markerview PCA Plots of Propofol Glucuronide from LC-MS Data

PCA of LC-MS data from Markerview. Propofol glucuronide is elevated in Surgical Controls. Panel A is the Score Plot. Panel B is the Loading Plot. Panel C displays the percent abundance of the selected feature in each subject. Because the metabolite is present in all but two of the Surgical Controls and is absent in all but two of the CDI patients, it is intuitive that this metabolite is a synthetic compound that was administered to the Surgical Control group.

Figure 7 displays the differential metabolite propofol glucuronide. A metabolite that is present in high concentration in nearly all members of one group and absent in nearly all members of the other group is most likely indicative of xenobiotic or drug. Propofol is an anesthetic that was administered to the Surgical Control group prior to their surgeries. The mass and retention time are listed in the top left corner of Panel C.
Figure 8 displays the MS and MS/MS spectra of propofol glucuronide. The spectra was chosen from the Surgical Control subject denoted U11, as this subject exhibited high levels of propofol in the Percent Abundance of Feature Plot in Panel C of Figure 7. The m/z and retention time of the compound, denoted in Panel C of Figure 7 were used to locate and identify the MS spectra of propofol in Panel A in Figure 8. The red asterisk within the broader red circle in Panel A indicates the data point that was selected. The MS/MS chromatogram displays the fragmentation pattern of propofol. In addition to accurate mass, the m/z values of the fragments were compared with published values, which confirmed that the compound was propofol glucuronide.
Figure 9: Markerview PCA Plots of Phenylalanine from LC-MS Data

PCA of LC-MS data from Markerview. Phenylalanine, an endogenous compound, present in both the CDI and Control groups. Panel A is the Score Plot. Panel B is the Loading Plot. Panel C displays the percent abundance of the selected feature in each subject. Because this compound is present in nearly equal levels in both the Control and the CDI groups, it is intuitive that it is an endogenous compound.

Figure 9 displays the metabolite phenylalanine. A metabolite that is present in similar concentrations in both groups is evidence that it is an endogenous compound. Phenylalanine is an essential amino acid that functions in protein formation, which explains why it would be found in both study groups regardless of their health status. The mass and retention time are listed in the top left corner of Panel C.
Panel A in the panel displays metabolites within the specified m/z range that triggered an MS in the Peakview software program. The metabolite of interest is denoted by a red asterisk within a broader red circle. Panel B displays the MS/MS chromatogram of the metabolite that was selected in the previous figure. The labeled peaks are representative of the fragment ions. MS/MS refers to tandem mass spectrometry, which measures the m/z of the features, followed by a fragmentation step, then measures the m/z of the fragments. The m/z of each fragment is denoted at the top of each peak. Panel C displays the published MS/MS chromatogram of the endogenous compound L-Phenylalanine on the Metlin database. The presence of fragment peaks with identical m/z values as those displayed in the lower left figure confirm that the metabolite is L-Phenylalanine.

Figure 10 displays the MS and MS/MS spectra of phenylalanine. The spectra was chosen from the Control Subject denoted U30, as this subject exhibited high levels of phenylalanine in the
Percent Abundance of Feature Plot in Panel C of Figure 9. The m/z and retention time of the compound, denoted in Panel C of Figure 9, were used to locate and identify the MS spectra of phenylalanine in Panel A in Figure 10. The red asterisk within the broader red circle in Panel A indicates the data point that was selected. The MS/MS chromatogram displays the fragmentation pattern of phenylalanine. In addition to accurate mass, the m/z values of the fragments were compared with the published MS/MS on the Metlin database. The appearance of identical peaks on the Metlin MS/MS confirmed that the metabolite was phenylalanine.

**Figure 11: Markerview PCA Plots of p-Cresol sulfate from LC-MS Data**

PCA of LC-MS data from Markerview. p-Cresol sulfate is elevated in the CDI group. Panel A is the Score Plot. Panel B is the Loading Plot. Panel C displays the percent abundance of the selected feature in each subject. Because the metabolite is present in 9 of the CDI patients at high levels, and only in 2 of the controls at much lower levels, p-Cresol sulfate can be considered a differential metabolite.

Figure 11 displays the metabolite p-Cresol sulfate. The compound is present in 9 of the 20 CDI patients at high levels, but only in 2 of the 20 controls at much lower levels. This distribution is most likely indicative of a differential metabolite. The mass and retention time are listed in the top left corner of Panel C.
Panel A displays the MS/MS chromatogram of p-Cresol sulfate. The labeled peaks are representative of the fragment ions. MS/MS is a multistep tandem mass spectrometry, which measures the mass of the features, followed by a fragmentation step, then measures the mass of the fragments. The mass of each fragment is denoted at the top of each peak. Panel B displays the published MS/MS chromatogram of p-Cresol sulfate from Cao et al. The presence of fragment peaks with identical m/z values as those displayed in the lower left figure confirm that the metabolite is p-Cresol sulfate.

Figure 12 displays the MS/MS spectrum of p-Cresol sulfate. The spectrum was chosen from the CDI subject denoted CD13, as this subject exhibited high levels of p-Cresol sulfate in the Percent Abundance of Feature Plot in Panel C of Figure 11. The m/z and retention time of the compound, denoted in Panel C of Figure 11, were used to locate and identify the MS/MS spectrum of p-Cresol sulfate in Panel A in Figure 12. The MS/MS chromatogram displays the fragmentation pattern of p-Cresol sulfate. In addition to accurate mass, the m/z values of the fragments were compared with the published MS/MS in the report by Cao et al. The appearance of identical peaks in Panel B confirmed that the metabolite was p-Cresol sulfate.
Figure 13: Markerview PCA Plots of Indoxylsulfate from LC-MS Data

PCA of LC-MS data from Markerview. Indoxylsulfate is elevated in the CDI group. Panel A is the Score Plot. Panel B is the Loading Plot. Panel C displays the percent abundance of the selected feature in each subject. Because the metabolite is present in 14 members of the CDI group at high levels, and only in 7 of the controls at much lower levels, indoxylsulfate can be considered a differential metabolite.

Figure 13 displays the metabolite indoxylsulfate. The compound is present in 14 of the 20 CDI patients at high levels, but only in 7 of the 20 controls at much lower levels. This distribution is most likely indicative of a differential metabolite. The mass and retention time are listed in the top left corner of Panel C.
Panel A displays the MS/MS chromatogram of indoxylsulfate. The labeled peaks are representative of the fragment ions. MS/MS is a multistep tandem mass spectrometry, which measures the mass of the features, followed by a fragmentation step, then measures the mass of the fragments. The mass of each fragment is denoted at the top of each peak. Panel B displays the published MS/MS chromatogram of indoxylsulfate from the Human Metabolome Database (http://www.hmdb.ca/). The presence of fragment peaks with identical m/z values as those displayed in the lower left figure confirm that the metabolite is indoxylsulfate.

Figure 14 displays the MS/MS spectrum of indoxylsulfate. The spectrum was chosen from the CDI subject denoted CD13, as this subject exhibited high levels of indoxylsulfate in the Percent Abundance of Feature Plot in Panel C of Figure 13. The m/z and retention time of the compound, denoted in Panel C of Figure 13, were used to locate and identify the MS/MS spectrum of indoxylsulfate in Panel A in Figure 14. The MS/MS chromatogram displays the fragmentation pattern of indoxylsulfate. In addition to accurate mass, the m/z values of the fragments were compared with the published MS/MS in the Human Metabolome Database.
The appearance of identical peaks in Panel B confirmed that the metabolite was indoxylsulfate.

**Metaboanalyst Results**

**Figure 15: Volcano Plot from Metaboanalyst**

Important features selected by volcano plot with fold change threshold (x) 80 and t-tests threshold (y) 0.01. The purple circles represent features above the threshold. Note both fold changes and p values are log transformed. The further its position away from the (0,0), the more significant the feature.

Figure 15 displays the volcano plot generated from the Metaboanalyst software tool. The significant features are denoted by the purple circles, which are above the fold change
and the t-test threshold parameters, as indicated by their placement above both dotted lines. When selected, the m/z and retention time of the significant features were listed. The m/z values were searched in the Metlin database.

Figure 16: Random Forest Plot from Metaboanalyst

Significant features identified by Random Forest. The features are ranked by the mean decrease in classification accuracy when they are permuted.

Figure 16 displays the Random Forest plot generated from the Metaboanalyst software tool. The m/z values of the significant features are listed to the leftmost side of the graph and the retention times are listed immediately to the right. The m/z values were searched in the Metlin database.
The color-coded key on the far right of the figure shows which features are elevated in which groups. Red squares signify an elevated abundance of the particular feature while green squares signify a lower abundance.
Chapter 4
Discussion

This study featured a comparative analysis of the urinary metabolic profiles of patients with CDI to those of control subjects (both normal, healthy controls and surgical controls). Both $^1$H NMR and LC-MS metabolomics were carried out on the samples, and results were analyzed with Markerview, Peakview, Metaboanalyst and multivariate statistics such as PCA and OPLS-DA. When the CDI group was compared to the Surgical Controls, two endogenous compounds, $p$-cresol sulfate and indoxylsulfate, were found to be significantly elevated in the CDI group, leading support to the status of each of these endogenous compounds as a potential biomarker of CDI. Further analysis revealed that the anesthetic propofol glucuronide, which was significantly elevated in the Surgical Control group, was the compound that was responsible for driving the distinction between the two groups. A new control group was drafted to eliminate the effect of propofol on the data. Unexpectedly, when compared to the Control group, the CDI group did not express significantly elevated levels of $p$-cresol sulfate and indoxylsulfate. $^1$H NMR revealed several significantly different compounds between the CDI and Control group. However, none of the compounds could conclusively be implicated as biomarkers of CDI and we are currently expanding enrollment in the study to increase the power of our study. The LC-MS data similarly did not reveal a conclusive biomarker, as there was no endogenous compound that was distinctly elevated or depressed in either the CDI or Control groups.
Table 2: Summary of Differential Metabolites

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Platform</th>
<th>Abundance in CDI Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-Cresol sulfate</td>
<td>LC-MS</td>
<td>Increased</td>
</tr>
<tr>
<td>Indoxylsulfate</td>
<td>LC-MS</td>
<td>Increased</td>
</tr>
<tr>
<td>Hippurate</td>
<td>NMR</td>
<td>Decreased</td>
</tr>
<tr>
<td>PAG</td>
<td>NMR</td>
<td>Decreased</td>
</tr>
<tr>
<td>TMAO</td>
<td>NMR</td>
<td>Decreased</td>
</tr>
<tr>
<td>Acetate</td>
<td>NMR</td>
<td>Increased</td>
</tr>
<tr>
<td>Butyrate</td>
<td>NMR</td>
<td>Increased</td>
</tr>
<tr>
<td>Creatine</td>
<td>NMR</td>
<td>Increased</td>
</tr>
<tr>
<td>Allantoate</td>
<td>NMR</td>
<td>Increased</td>
</tr>
<tr>
<td>Glycine</td>
<td>NMR</td>
<td>Increased</td>
</tr>
<tr>
<td>2-oxoglutarate</td>
<td>NMR</td>
<td>Increased</td>
</tr>
<tr>
<td>Methylmalonate</td>
<td>NMR</td>
<td>Increased</td>
</tr>
<tr>
<td>Creatinine</td>
<td>NMR</td>
<td>Decreased</td>
</tr>
<tr>
<td>Urea</td>
<td>NMR</td>
<td>Decreased</td>
</tr>
</tbody>
</table>

Differential Metabolites. p-Cresol sulfate and indoxylsulfate were found to be elevated in the CDI group when it was compared to the surgical controls. All other metabolites were found to be significant in the comparison of the CDI group to the controls.

NMR Results Discussion

A study by Romick-Rosendale et al. that utilized NMR-based metabolomics to measure the effects of broad-spectrum antibiotic treatment on urinary metabolites found decreased acetate...
and TMAO in the urine of treated mice. These results partially support the findings of this study, as TMAO was decreased in the CDI group relative to the Controls. The treatment group in the previous study is similar to the CDI group in this study, as the CDI group is comprised of patients who had likely been treated with broad-spectrum antibiotics prior to being diagnosed with CDI, as these medications often facilitate the development of the infection. The decrease in TMAO can possibly be attributed to the loss of bacterial catabolism of dietary choline. Yap et al. reported increased levels of choline after vancomycin treatment. Accumulation of choline, and thus decreased breakdown of choline, supports the result of decreased TMAO in this study and that by Romick-Rosendale et al. Acetate, however, was found to be significantly elevated in the CDI group, which is exhibited in Figure 3.

Romick-Rosendale et al. also reported that the treatment group had elevated levels of creatine and creatinine. Consistent with Romick-Rosendale et al, increased creatine and creatinine in fecal content following treatment with the antibiotic clindamycin was reported by Jump et al. The CDI group expressed significantly increased levels of creatine relative to the controls, likely due to the loss of bacterial enzyme degradation, yet depressed levels of creatinine, as can be seen in Figure 3.

Consistency between this study and that of Romick-Rosendale et al. was exhibited by depressed levels of urea in both the CDI group and the treatment group, which is likely due to a loss of bacterial purine catabolism. In contrast to these results, Jump et al. identified urea as a potential biomarker of colonization resistance, or the prevention of colonization of pathogenic bacteria in the host by the indigenous microbiota, as urea demonstrated significant elevation during clindamycin treatment followed by a return to baseline levels within 8 days post-treatment. Urea, an intermediate in carbohydrate metabolism, was likely elevated due to the loss
of metabolic digestion by microbiota. It is intuitive that after the antimicrobial treatment was stopped, recolonization by normal, anaerobic bacteria correlated with the reduction of urea to normal levels. This finding conflicts with that of Romick-Rosendale et al, as well as this study in that urea was depressed in the CDI group.

Yap et al. reported depressed levels of urinary excretion of the bacterial co-metabolites hippurate and PAG, which is a result that is consistent with this study, as is demonstrated in Figure 3.

Theriot et al reported increased levels of the amino acid glycine in the urine of mice following antibiotic treatment. This finding is consistent with the increased glycine levels in the CDI group, as is demonstrated in Figure 3. Previous studies have reported that glycine enhances the growth of C. difficile.

It is important to consider that Romick-Rosendale, Yap and Theriot all used mice as models in their respective studies. While mice reflect close similarities to humans in anatomy, physiology and genetics, there are crucial differences in the gut microbiomes between the two mammals that must be taken into account when translating results from mice to human models. Casteleyn et al. reported that the surface ratio of small intestine to colon is 18 in mice, yet 400 in humans. It is known that humans have transverse folds along the length of their colons, while in mice these folds are restricted to the proximal colon and cecum. Mice are also known to have greater fermentation capacity in the cecum than humans. These differences in structure and compartmentalization can influence the creation of diverse ecological micro-niches that are capable hosting different microbial communities. Differences also exist between the two species in the distribution of Paneth and goblet cells in the GI tract, which suggests possible differences in local immune responses, which may influence the composition of the gut
microbiota. While Bacteroidetes and Firmicutes are the most prominent phyla in both humans and mice, Ley et al. reported that 85% of the bacterial genera present in mice are not found in humans. It is possible that these anatomical differences between humans and mice are shaped by variance in diets, feeding patterns, body size, and metabolic requires.

In addition to the intrinsic differences between mice and humans, divergence between the mammalian systems may be due to environmental factors. Mice involved with studies are typically housed in specific pathogen free (SPF) conditions, where the presence of microbes and diet are tightly controlled. Chow diet, which consists mostly of plant-based material, vastly differs from the daily human diet in composition and variation. The diets between humans and mice may differ in components that have a prominent influence on gut microbiota composition. It is also possible that stress resulting from human handling and living in a confined space affects the gut microbiota in mice. Additionally, mice used in studies are typically inbred in order to maintain genetic homogeneity, which enables researchers to elucidate mechanism and causality. However, these inbred strains cannot model the natural genetic variation in the human population.

One possible conclusion that can be drawn from the comparison of results from the studies by Romick-Rosendale, Yap and Theriot to those of this study is that the changes in the differential metabolites that were consistent among the studies may be due to the inherent differences between human and murine gut microbiota, or to the common treatment of broad spectrum antibiotics. Thus, the urinary metabolites that were found to be significant in this study cannot conclusively be attributed to the presence of CDI.
LC-MS Results Discussion

One major obstacle of analyzing of the LC-MS data was that of identifying synthetic compounds and removing them from the data set. Initially, the study consisted of only the Surgical Control group and the CDI group. The rationale for selecting patients who had undergone a surgical procedure to comprise the control group in this study was to account for physiological stress. Because all of the CDI subjects had at some point in their course of treatment undergone procedures ranging from body scans to GI surgery, we thought it was necessary to negate the potential effect of physiological stress to the body as the agent of differential metabolism between the two groups. When the LC-MS data from the Surgical Control and the CDI groups was analyzed in Markerview, two endogenous compounds, \( p \)-cresol sulfate and indoxylsulfate, were identified and found to be significantly elevated in the CDI group. It is known that \( C. difficile \) has the ability to ferment tyrosine into \( p \)-cresol sulfate.\(^{61}\) \( p \)-cresol sulfate is a phenolic, bacteriostatic compound that \( C. difficile \) is able to tolerate, and it is thought that \( p \)-cresol sulfate may provide \( C. difficile \) with an advantage that enables it to outcompete other gut microbiota, thereby allowing it to proliferate and cause CDI.\(^{57}\) Thus, it is reasonable to conclude that elevated levels of \( p \)-cresol sulfate is indicative of CDI.

Indoxylsulfate, a derivative of indole, is chemically similar to \( p \)-cresol sulfate. Both compounds are involved with cytokine and chemokine activation, and have demonstrated the ability to decrease endothelial proliferation and interfere with wound repair.\(^{62}\) Indoxylsulfate is also known to accumulate to significantly high levels in patients with chronic kidney disease.\(^{63}\) Nicholson et al. reported indoxylsulfate to be a metabolite that is contributed by a member of the Clostridium class, \( C. sporogenes \), and that it is associated with GI pathologies.\(^{16}\) It is therefore plausible that indoxylsulfate is a metabolic product common of multiple members of the
Clostridium class, including that of *C. difficile*, which could explain why it was shown to be elevated in the CDI group when compared to the Surgical Control group.

A separate study by Nicholson et al. reported *p*-cresol sulfate, PAG, and hippurate to be metabolic products of *C. difficile*.16 Nicholson et al. also suggested that urinary hippurate may be a biomarker of obesity and hypertension in humans, and that elevated levels of *p*-cresol sulfate in urine is associated with colorectal cancer.16 While we found *p*-cresol sulfate to be elevated in the CDI group when it was compared to the Surgical Controls, PAG and hippurate were significantly lower in the CDI group relative to the Controls, as can be seen in Figure 3. Because PAG and hippurate are metabolic products of *C. difficile*, it would be expected that they would be elevated in the CDI group. However, because all of the CDI patients were being treated with antibiotics in heavy doses, it is possible that these metabolites were depressed in the group because the antibiotics were killing off the *C. difficile* population. Theriot et al. reported that antibiotic use and the disruption of the gut microbiota that it causes alters the global gut microbiome profile to one that favors the germination and proliferation of *C. difficile*.55

Nicholson et al. also reported that TMAO is a metabolite associated with the bacteria *Faecalibacterium prausnitzii*, which is a member of the Clostridia class. The population of *F. prausnitzii*, which is known to have anti-inflammatory properties, was shown to be reduced in abundance in patients with Crohn’s Disease (CD).64 CD is similar to CDI in that it affects the beginning of the colon and causes inflammation.60 Due to the comparable effects of CD and CDI on the GI tract and the reported finding of a reduction in *F. prausnitzii* populations in CD patients, it would be reasonable to expect reduced *F. prausnitzii* abundance, and thus reduced levels of TMAO in CDI patients. As can be seen in Figure 3, the result of decreased TMAO in the CDI group supports this expectation. However, Romick-Rosendale et al. found decreased
TMAO in the urine of mice that had been treated with broad-spectrum antibiotics. The application of broad-spectrum antibiotics is often what triggers CDI, and CD patients are prescribed antibiotics for bacterial infections in the GI tract. Thus, it is possible that antibiotic use, as opposed to the presence of *C. difficile*, is the driving force of the reduced abundance of TMAO.

Acetate and Butyrate are reported to be associated with bacterial species belonging to clusters IV and XIVa of the class Clostridia and the phylum Firmicutes. Figure 3 shows each of these compounds to be elevated in the CDI group. While both acetate and butyrate are metabolites expected to be present in healthy subjects, it is possible that the diminished population of Bacteroidetes in the colon of individuals with CDI allows for an increase in proportion in members of the phylum Firmicutes, which could explain the significant increase in acetate and butyrate in the CDI group.

Further analysis of the LC-MS results revealed that the significant distinction between the Surgical Control and CDI groups was driven by the anesthetic propofol glucuronide, which was present in high concentration in the Surgical Controls. Because the aim of this study was to examine the effect of CDI on the gut microbial community, only differences in endogenous metabolites are of value in implicating a biomarker. Thus, twenty new control patients were selected to comprise the Control group. These patients had either never undergone a surgical procedure, or the urine samples were collected from them either prior to or at least four weeks following such a procedure, thereby excluding the effect of anesthetics like propofol on the data. The clustering of the LC-MS data between the Controls and the CDI group, however, did not exhibit clear differentiation. Another unexpected result from the analysis of the Controls and the CDI group was that *p*-cresol sulfate and indoxylsulfate did not present to be significantly higher
in the CDI group, as it had whenever the CDI was compared to the Surgical Controls. Although the control group had been changed, the same urine samples from subjects comprising the CDI group were used in both rounds of LC-MS analysis. It is therefore unclear why p-cresol sulfate and indoxylsulfate did not present to be significantly higher in the CDI group when compared to the Controls. We cannot conclude that that p-cresol sulfate and indoxylsulfate are biomarkers of CDI.

Even though propofol was no longer present in the data set, nearly all of the patients in both the CDI and Control group were on multiple prescription medications. It is possible that potentially relevant endogenous compounds, although present in the urine, never triggered an MS due to their low concentration relative to the high concentrations of the synthetic compounds. This effect was undesirable because the sought after endogenous biomarker of CDI may have been deemed insignificant due to the normalization of the data set. To correct for this effect the data was not normalized and the fold change was measured for each metabolite by dividing the CDI group value by that of the Control group. Metabolites with fold changes of zero, or those that could not be calculated because the Control group value was equal to zero, were eliminated from the data set before it was uploaded to Metaboanalyst. Any metabolite absent in all members of one group, yet present in all members of the other group was determined to be a synthetic component, and was thus not a potential biomarker of CDI. However, we were not able to identify the endogenous compounds that were the top hits from the random forest and volcano plots.

Study Limitations: One limitation of this study is the uneven distribution of gender among the groups. The age of the males in the CDI group was significantly higher than that of the males in the control group (p< 0.05), which is another limiting factor. Health history and
medication information is also unknown for two subjects in the control group. The time of the day that the samples were collected was not consistent in the study.

The enormous variation in chemical composition and in enzymatic activities among individuals could account for differences in metabolism. This biochemical variability among individuals is likely due in part to the vast variability in the microbiota that comprises the ecological community of the gut in each person. The factors responsible for the diversity among microbial communities are the genetics of the individual hosts and their nutritional status, including dietary intake, exposure to environmental toxicity and drug toxins.\textsuperscript{65} The composition and activity of the gut microbiota being to develop along with the host at birth. It is also thought that maternal microbiota may influence the health of the fetus both prior to and after parturition.\textsuperscript{16} In an effort to measure heterogeneity among human populations, Wijeyesekera et al. found that the metabolic activity of the microbial communities resulted in different concentrations of microbial end products between individuals of different gender, age, race and ethnicity.\textsuperscript{66} For example, Japanese subjects were found to excrete significantly lower levels of hippuric acid in urine than other populations.\textsuperscript{62} Another metabolomics study found that half of the metabolites present in normal urine were statistically unique between males and females.\textsuperscript{67} Saude et al. reported that nine of the twenty-four metabolites present in normal urine statistically differed between the morning and afternoon sampling times, thereby displaying variance within individuals.\textsuperscript{63} Thus, in addition to the widespread variability among individuals, studies have shown that within individuals, the microbial community changes throughout a person’s lifetime.\textsuperscript{65}

\textbf{Future Studies:} It would be beneficial to perform future studies with a greater sample size. Trends in the data would likely be more evident with a larger sample size and
reproducibility would be more feasible. Ideally, the future pool of subjects would not be consuming any medications other than the necessary CDI treatment for the CDI group, although the feasibility of this scenario is low due to the likelihood of pre-existing conditions in the affected patients. It may also be beneficial to re-run the study with a targeted metabolomics approach, which involves carrying out quantitative and qualitative analysis of specific metabolites of interest, such as \( p \)-cresol sulfate, indoxylsulfate, hippurate or TMAO, as opposed to global profiling of differential metabolites between the groups.

Additionally, while urine was the only biofluid examined in this study, analyzing fecal content and serum samples may provide a more complete understanding of how the functioning of the gut-microbiota is affected by CDI. Urine is a favorable biofluid to use because it can be collected in large volumes non-invasively, and it requires less sample preparation due to its low protein quantities.\(^6\) However, examining fecal content would be advantageous because it allows for bacteriological culture analysis. Determining the microbial populations and grouping them by shared characteristics, as well as species identification, is made possible by the technique of 16S ribosomal RNA (rRNA) gene sequencing. This technique works because the 16S rRNA gene is highly conserved among bacteria, and the presence of nine hypervariable regions of the sequences allows for the classification of microbes.\(^6\) If fecal content was obtained from study subjects, the proportions of certain bacterial populations and any alterations in microbial composition could shed light on the effect of CDI on the gut microbial community.

Another future exploration would be to track urinary metabolites over an extended period of time in a group of healthy patients who had once been diagnosed with CDI. By following these subjects, it may be possible to detect changes in the concentration of certain metabolites that are associated with recurrent infections. Detecting the onset of recurrent CDI, potentially
prior to the patient becoming symptomatic, has significant implications in reducing morbidity and mortality. The concentrations of differential metabolites may also correlate with specific strains of CDI. Subsequently, clinicians would be able to prescribe treatment according to the specific CDI strain, which would likely improve effectivity.
Appendix A

List of Abbreviations

CDI - *Clostridium difficile* Infection
NMR - Nuclear Magnetic Resonance
PBS – Phosphate-Buffered Saline
TSP – Trimethylsilyl propionate
PCA – Principal Component Analysis
OPLS-DA – Orthogonal Projection to Latent Structures Discriminant Analysis
UPLC-ESI-QTOFMS – Ultra High Pressure Liquid Chromatography Coupled with Electrospray Ionization Quadrupole Time-of-Flight Mass Spectrometry
IDA - Information Dependent Acquisition
IQR - Interquartile Range
Appendix B

Medication Codes

A – Propofol
B – Acetaminophen
C – Heparin
D - Aspirin
E – Oxycodone
F - Prednisone
G- Insulin
H – Lovenox
I – Fluoxetine
J – Vancomycin
K – Zosyn
L – Fioricet
M – Ibuprofen
N – Atorvastatin
O – Cefdinir
P – Plavix
Q – Vaseretic
R – Escitalopram
S – Febuxostat
T – Furosimide
U – Levothyroxine
V – Nitrofurantoin
W – Olmesartan
X - Potassium Chloride
Y – Pregabalin
Z – Ropinirole
AA – Sucralfate
BB – Dilaudid
CC- Lipitor
DD – Wellbutrin
EE - Catapres
FF - Ferrous Sulfate
GG – Advair
HH – Gabapentin
II – Glimepiride
JJ – Hydralazine
KK – Imudur
LL – Synthroid
<table>
<thead>
<tr>
<th>Code</th>
<th>Drug 1</th>
<th>Code</th>
<th>Drug 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>MM</td>
<td>Prilosec</td>
<td>JJJ</td>
<td>Duloxetine</td>
</tr>
<tr>
<td>NN</td>
<td>Sotalol</td>
<td>KKK</td>
<td>Lactobacillus</td>
</tr>
<tr>
<td>OO</td>
<td>Coumadin</td>
<td>LLL</td>
<td>Methadone</td>
</tr>
<tr>
<td>PP</td>
<td>Bumex</td>
<td>MMM</td>
<td>Alubuterol</td>
</tr>
<tr>
<td>QQ</td>
<td>Cefepime</td>
<td>NNN</td>
<td>Finasteride</td>
</tr>
<tr>
<td>RR</td>
<td>Amlodipine</td>
<td>OOO</td>
<td>Hydrochlorothiazide</td>
</tr>
<tr>
<td>SS</td>
<td>Calcitriol</td>
<td>PPP</td>
<td>Tamsulosin</td>
</tr>
<tr>
<td>TT</td>
<td>Lisinopril</td>
<td>QQQ</td>
<td>Ipratropium</td>
</tr>
<tr>
<td>UU</td>
<td>Niacin</td>
<td>RRR</td>
<td>Zolpidem</td>
</tr>
<tr>
<td>VV</td>
<td>Venlafaxine</td>
<td>SSS</td>
<td>Doxycycline</td>
</tr>
<tr>
<td>WW</td>
<td>Warfarin</td>
<td>TTT</td>
<td>Citalopram</td>
</tr>
<tr>
<td>XX</td>
<td>Pantoprazole</td>
<td>UUU</td>
<td>Folic Acid</td>
</tr>
<tr>
<td>YY</td>
<td>Imodium</td>
<td>VVV</td>
<td>Norvasc</td>
</tr>
<tr>
<td>ZZ</td>
<td>Acyclovir</td>
<td>WWW</td>
<td>Toradol</td>
</tr>
<tr>
<td>AAA</td>
<td>Ciprofloxacin</td>
<td>XXX</td>
<td>Hydrochlorothiazide-</td>
</tr>
<tr>
<td>BBB</td>
<td>Fluconazole</td>
<td>Losartan</td>
<td></td>
</tr>
<tr>
<td>CCC</td>
<td>Metoprolol</td>
<td>YYY</td>
<td>Pravastatin</td>
</tr>
<tr>
<td>DDD</td>
<td>Fentanyl</td>
<td>ZZZ</td>
<td>Modafinil</td>
</tr>
<tr>
<td>EEE</td>
<td>Filgrastim</td>
<td>AB</td>
<td>Ezetimibe</td>
</tr>
<tr>
<td>FFF</td>
<td>Levetiracetam</td>
<td>AC</td>
<td>Lecithin</td>
</tr>
<tr>
<td>GGG</td>
<td>Lorazepam</td>
<td>AD</td>
<td>Flagyl</td>
</tr>
<tr>
<td>HHH</td>
<td>Doxorubicin</td>
<td>AE</td>
<td>Kefzol</td>
</tr>
<tr>
<td>III</td>
<td>Dexamethasone</td>
<td>AF</td>
<td>Midazolam</td>
</tr>
<tr>
<td>Code</td>
<td>Drug Name</td>
<td>Code</td>
<td>Drug Name</td>
</tr>
<tr>
<td>------</td>
<td>----------------------------</td>
<td>------</td>
<td>----------------------------</td>
</tr>
<tr>
<td>AG</td>
<td>Rocuronium Lidocaine</td>
<td>BE</td>
<td>Hydrocortisone</td>
</tr>
<tr>
<td>AH</td>
<td>Zofran</td>
<td>BF</td>
<td>Ketorolac</td>
</tr>
<tr>
<td>AI</td>
<td>Glycopyrrolate</td>
<td>BG</td>
<td>Succinylcholine</td>
</tr>
<tr>
<td>AJ</td>
<td>Neostigmine</td>
<td>BH</td>
<td>Bupivacaine</td>
</tr>
<tr>
<td>AK</td>
<td>Indigo Carmine</td>
<td>BI</td>
<td>Clonazepam</td>
</tr>
<tr>
<td>AL</td>
<td>Cefazolin</td>
<td>BJ</td>
<td>Cellcept</td>
</tr>
<tr>
<td>AM</td>
<td>Hydromorphone</td>
<td>BK</td>
<td>Plaquenil</td>
</tr>
<tr>
<td>AN</td>
<td>Labetalol</td>
<td>BL</td>
<td>Prograf</td>
</tr>
<tr>
<td>AO</td>
<td>Phenylephrine</td>
<td>BM</td>
<td>Doxazosin</td>
</tr>
<tr>
<td>AP</td>
<td>Gentamycin</td>
<td>BN</td>
<td>Glipizide</td>
</tr>
<tr>
<td>AQ</td>
<td>Clindamycin</td>
<td>BO</td>
<td>Augmentin</td>
</tr>
<tr>
<td>AR</td>
<td>Metformin</td>
<td>BP</td>
<td>Rosuvastatin</td>
</tr>
<tr>
<td>AS</td>
<td>Simvastatin</td>
<td>BQ</td>
<td>Zrytec</td>
</tr>
<tr>
<td>AT</td>
<td>Ephedrine</td>
<td>BR</td>
<td>Flonase</td>
</tr>
<tr>
<td>AU</td>
<td>Sulfasalazine</td>
<td>BS</td>
<td>Morphine</td>
</tr>
<tr>
<td>AV</td>
<td>Flecaainide</td>
<td>BT</td>
<td>Alprazolam</td>
</tr>
<tr>
<td>AW</td>
<td>Trazodone</td>
<td>BU</td>
<td>Ambien</td>
</tr>
<tr>
<td>AX</td>
<td>Ketamine</td>
<td>BV</td>
<td>Alendronate</td>
</tr>
<tr>
<td>AY</td>
<td>Metronidazole</td>
<td>BW</td>
<td>Allopurinil</td>
</tr>
<tr>
<td>AZ</td>
<td>Omeprazole</td>
<td>BX</td>
<td>Cholecalciferol</td>
</tr>
<tr>
<td>BA</td>
<td>Adalimumumab</td>
<td>BY</td>
<td>Dicyclomine</td>
</tr>
<tr>
<td>BC</td>
<td>Fenofibrate</td>
<td>CA</td>
<td>Mirabegron</td>
</tr>
<tr>
<td>BD</td>
<td>Mesalamine</td>
<td>CB</td>
<td>Propranolol</td>
</tr>
<tr>
<td>CD – Valsartan</td>
<td>CZ – Chlorodiapoxide</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CE – Estradiol</td>
<td>DA – Enoxaparin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CF – Lomotil</td>
<td>DB – Valganciclovir</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG – Simethicone</td>
<td>DC – Spironolactone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CH – Percocet</td>
<td>DE – Amiodarone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CI – Lasix</td>
<td>DF – Bumetanide</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CJ – Magnesium Oxide</td>
<td>DG – Carvedilol</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CK – Pioglitazone</td>
<td>DH – Symbicort</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CL – Bactrim</td>
<td>DI – Keppra</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CM – Eletriptan</td>
<td>DJ – Stiagliptin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CN – Famotidine</td>
<td>DK – Spiriva</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CO – Isosorbide Mononitrate</td>
<td>DL – Ativan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP – Repaglinide</td>
<td>DM – Digoxin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CQ – Clobetasol</td>
<td>DN – Dronedarone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR – Amitriptyline</td>
<td>DO – Asacol</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CS – Cephalexin</td>
<td>DP – Retin A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CT – Metaxalol</td>
<td>DQ - Tri-Previfem</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CU – Cetirizine</td>
<td>DR – Dexamedetomidine</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CV – Ondansetron</td>
<td>DS- Hydroxychloroquine</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CW – Protonix</td>
<td>DT- Hydrocodone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CX – Diflucan</td>
<td>DU - Tramadol</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CY – Nasonex</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
BIBLIOGRAPHY


ACADEMIC VITA

Karissa Allan Thal
618 Jacksonville Road, Bellefonte, PA 16823 / kat5314@psu.edu

Education

The Pennsylvania State University
Schreyer Honors College
Major: Biology, Bachelor of Science

Anticipated Date of Graduation: May 2015

Work Experience

The Pennsylvania State University, University Park, PA 16802

Teaching Assistant August 2013-October 2013

- Instructed the laboratory portion of an introductory biology course for first-year students
- Designed lesson plans and quizzes used for course evaluation
- Fostered the development of proper laboratory skills and scientific writing among students

New Student Orientation Scheduling Advisor June 2013-August 2013

- Worked one-on-one with rising first-year students to create a first semester course schedule
- Counseled students on major choices and future course selection
- Presented educational abroad opportunities to incoming students

Laboratory Research Assistant June 2012-October 2013
The Patterson Research Group

- Conducting a portion of a research project that is funded by the NIH
- The goal of the project is to determine the cause of substantial weight gain in a group of children consuming Valproate as a treatment for epilepsy
- Duties include preparing samples, recording data, and analyzing results

Honors and Activities

- Schreyer Honors College Academic Scholarship Recipient 2011-2015
  - Recognized for academic excellence
- Dean’s List 2011-2014
  - Recognized for semester GPA greater than 3.5
- Science LionPride Ambassador 2012-2013
  - Serve as an ambassador for Penn State’s Eberly College of Science
  - Give tours of the College of Science to prospective students
  - Attend distinguished alumni events
  - Participate in service activities including THON and Relay for Life