A Systems Biology Approach to Microbiology and Cancer

Seda Arat

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Mathematics

Reinhard Laubenbacher, Chair
John A. Burns
M. Stanca Ciupe
Stefan Hoops

July 31, 2015
Blacksburg, Virginia

Keywords: data analysis, mathematical modeling, microbiome, drug repositioning, polynomial dynamical system

A Systems Biology Approach to Microbiology and Cancer

Seda Arat

ABSTRACT

Systems biology is an interdisciplinary field that focuses on elucidating complex biological processes (systems) by investigating the interactions among its components through an iterative cycle composed of data generation, data analysis and mathematical modeling. Our contributions to systems biology revolve around the following two axes:

- **Data analysis**: Two data analysis projects, which were initiated when I was a co-op at GlaxoSmithKline, are discussed in this thesis. First, next generation sequencing data generated for a phase I clinical trial is analyzed to determine the altered microbial community in human gut before and after antibiotic usage (Chapter 2). To our knowledge, there have not been similar comparative studies in humans on the impacts on the gut microbiome of an antibiotic when administered by different modes. Second, publicly available gene expression data is analyzed to investigate human immune response to tuberculosis (TB) infection (Chapter 3). The novel feature of this study is systematic drug repositioning for the prevention, control and treatment of TB using the Connectivity map.
- **Mathematical modeling**: Polynomial dynamical systems, a state- and time- discrete logical modeling framework, is used to model two biological processes. First, a denitrification pathway in *Pseudomonas aeruginosa* is modeled to shed light on the reason of greenhouse gas nitrous oxide accumulation (Chapter 4). It is the *first* mathematical model of denitrification that can predict the effect of phosphate on the denitrification performance of this bacterium. Second, an iron homeostasis pathway linked to iron utilization, oxidative stress response and oncogenic pathways is constructed to investigate how normal breast cells become cancerous (Chapter 5). To date, our intracellular model is the only expanded core iron model that can capture a breast cancer phenotype by overexpression and knockout simulations.

This work was funded by National Institute of Health grant number NCI-NIH 1R21CA156133-01A1, U.S. Army Research Office grant number W911NF-14-1-0486, GlaxoSmithKline and the Department of Education in Turkey. Results were in no way influenced by the funding agencies.

Acknowledgments

I would like to acknowledge the assistance of my advisor, my committee members, faculty and staff at the Department of Mathematics and Virginia Bioinformatics Institute at Virginia Tech (VT), members of the Computational Biology group at GlaxoSmithKline (GSK), faculty and staff at the Center for Quantitative Medicine at the University of Connecticut Health Center (UCHC), and my collaborators for providing an excellent and diverse working environment.

This achievement would not be possible without the insight, inspiration, challenges and encouragement presented by my advisor, mentor and friend, Reinhard Laubenbacher. I feel privileged to have been working with him. I would like to thank my committee, John Burns, Stanca Ciupe and Stefan Hoops for their valuable feedback. I would also like to thank my supervisor at GSK, Jim Brown for his help, letting me dive into the "data analysis" field and a fruitful collaboration.

I'm especially thankful to my mentors, collaborators and friends, who have been patient, supportive and responsive: Peter Haskell, Nicole Sutphin, Ken Hinson, Bill Reilly and Cigdem Arca from VT; Madison Brandon, Blake Treadway, Cory Brunson, Kathy Black, Suzy Torti and Chris Heinen from the UCHC; Claus Kadelka from the Uni. of Zurich; Shernita Lee from the Uni. of North Carolina-Chapel Hill; David Murrugarra from the Uni. of Kentucky; Matt Oremland from Ohio State Uni.; George Bullerjahn from Bowling Green State Uni.; Julia Chifman from Wake Forest Uni.; Zhang Wang, Michal Magid-Slav, Craig Volker and Kent Goklen from GSK; Aaron Spivak from Harvard Uni.; Lun Yang from Bayer and Maria Laubenbacher.

Finally, I thank my parents Serpil and Adnan, my grandma Fatma, my brother Emre and my husband Ali for shaping my life.

Contents

1 Introduction **1**

1.1 Systems Biology Approach 2

1.2 Outline 4

2 Microbiome Analysis in an Antibiotic Clinical Trial **6**

2.1 Introduction 7

2.2 Results and Discussion 8

2.2.1 Microbial Community Diversity 8

2.2.2 Functional Analysis 13

2.3 Materials and Methods 15

2.3.1 Clinical Study Design 15

2.3.2 Sample Collection, DNA Extraction and Sequencing 16

2.3.3 Microbiome Data Analysis 17

2.4 Conclusion 18

2.5 Acknowledgments 20

3 Meta-analysis of Human Immune Response to Tuberculosis **21**

3.1 Introduction 22

3.2 Results and Discussion 23

3.2.1 Dataset Selection and Quality Filtering . . . 23
3.2.2 Differential Expression Analysis . . . 25
3.2.3 Pathway Enrichment Analysis . . . 26
3.2.4 Upstream Regulator Analysis . . . 28
3.2.5 Drug Repositioning Analysis . . . 29
3.3 Materials and Methods . . . 30
3.3.1 Data Sources, Selection and Quality Control . . . 30
3.3.2 Differential Expression, Pathway and Upstream Regulator Analyses . . . 31
3.3.3 Drug Repositioning . . . 31
3.4 Conclusion . . . 32
3.5 Acknowledgments . . . 32

4 A Denitrification Network Model of *Pseudomonas aeruginosa* 35
4.1 Introduction . . . 36
4.2 Results and Discussion . . . 38
4.3 Materials and Methods . . . 43
4.3.1 Mathematical Model . . . 43
4.3.2 Experimental Methods . . . 47
4.4 Conclusion . . . 47
4.5 Acknowlegments . . . 48

5 Intracellular Iron Model: from Normal Breast Cells to Cancer Cells 49
5.1 Introduction . . . 50
5.2 Results and Discussion . . . 51
5.2.1 An expanded iron core network . . . 51

5.2.2 Logical Model 56
5.2.3 Model Validation 57
5.2.4 Differential regulation of pathways in cancer 58
5.3 Materials and Methods 59
5.3.1 Mathematical Model 59
5.3.2 Experimental Methods 62
5.4 Conclusion 63
5.5 Acknowledgments 63

Bibliography **66**

Appendix A A Denitrification Network Model of *Pseudomonas aeruginosa* **89**

A.1 Transition tables of Dnr, NirQ, nar and NO_2 89

Appendix B Modeling Iron-dependent Oxidative Stress in Breast Cancer **95**

B.1 Construction of the *update polynomial* for ferritin (Ft) 95
B.2 Variables and their Update Polynomials 96

List of Figures

1.1 Systems biology approach taken for this study, in which two data analysis and two mathematical modeling projects are discussed. The colon cancer project is briefly mentioned. 4

2.1 Schematic representation of the microbiome analysis pipeline synopsis for NGS data. 9

2.2 α-diversity rarefaction plots using *chao1* measure of within-group diversity. 10

2.3 Histogram of proportional changes in bacterial OTU abundance at the class level. (A) placebo and i.v.-only dosing regimen; (B) oral-i.v. dosing regimen. 11

2.4 β-diversity plots of between-group diversity based on treatment (A) and dosage (B). 14

2.5 Predictions of the functional composition of oral-i.v. metagenome. Pre-study (blue bars); End-of-study (green bars). 15

3.1 Outline of the iterative analysis process. 24

3.2 Differential gene expression profile. FC: fold change; Blue: FC $<$ -1.5; Yellow: FC $>$ 1.5; White: -1.5 $\leq$ FC $\leq$ 1.5 26

4.1 Denitrification regulatory network of *Pseudomonas aeruginosa*. 37

4.2 Steady states of the denitrification network under different environmental conditions. 40

5.1 Comprehensive pipeline for systems biology approach taken for this study. 51

5.2 Expanded iron homeostasis pathway. Arrows depict upregulation and hammer heads depict downregulation. Dashed connections are assumed to exist. 52

5.3 Effects of IRP2 overexpression on the network. GAPDH is a loading control. . . . 58

List of Tables

2.1 Significantly changed OTUs in subjects receiving oral-i.v.-administered antibiotic in pre- versus post-study comparisons based on mean proportional abundance. . . . 12

2.2 Overview of dosing regimen and clinical study design. (BID: two times a day) . . . 16

3.1 Profiles of the GEO datasets that pass all criteria filters. 25

3.2 The common enriched pathways in at least 5 PTB datasets. 27

3.3 The common upstream regulators in all PTB datasets. 33

3.4 The approved Cmap compounds in the DrugBank database. 34

4.1 Discretization of external parameters and nitrogen oxides. 39

4.2 Nitrous oxide concentration in *P. aeruginosa* cultures grown in glucose minimal medium at varying phosphate concentrations, normalized to 10^8 cells. 41

4.3 Biological interpretation of the steady states (phenotypes) of the system under different environmental conditions. 42

4.4 Summary of the model variables, update rules and supportive argument. 44

5.3 Transition tables for activation and inhibition. 60

5.4 Transition tables for IRP1 and IRP2 regulating Ft. 60

5.1 Summary of all model variables and their update rules. IH, iron homeostasis; IU, iron utilization; OSR, oxidative stress response; IR, inflammatory response; Onc, Oncogenic. 64

5.2 Attractors (steady states) under certain perturbations. 65

A.1 Transition table of Dnr 90

A.2 Transition table of NirQ 91

A.3 Transition table of nar 93

A.4 Transition table of NO_2 94

Chapter 1

Introduction

"The essence of mathematics is its freedom."
Georg Cantor (1845 - 1918)

Systems biology is an interdisciplinary field that aims to elucidate complex biological processes (systems) by investigating the interactions among its components through a *cycle* composed of data generation, data analysis and mathematical modeling [1, 2, 3]. In order to understand a living organism at the systems-level, we need to start studying a cell as a system, in which genes, transpiration factors, microRNAs, proteins and metabolites function individually and within a larger network. Depending on the system and the scientific question of interest, high-throughput data is generated and analyzed to detect which components are significantly altered under different conditions or perturbations. These altered components can be further used to generate hypotheses (e.g. pathway analysis) or to be the building blocks of a mathematical model. A mathematical model consists of a *network* that summarizes all the information about the system, its components and their interactions, and a *set of rules* that correspond to the regulation of components over time. Simulations of the model act as *in silico* experiments, reproducing a behavior of the system and generating hypotheses. Hypotheses, generated by data analysis results or model simulations, are then validated by experimentation (new data generation) or existing data that was not used form model construction. This iterative process requires collaboration across many disciplines including biology, chemistry, statistics, computer science and mathematics. One of the challenges in systems biology is to overcome jargon barrier and facilitate crosstalk among researchers in different fields.

1.1 Systems Biology Approach

A generic systems biology approach begins with a system of interest and a well-defined scientific question formulated by life and computational scientists together. Following that, existing literature (published data and known facts) related with this question is surveyed thoroughly using well-known databases such as PubMed (http://www.ncbi.nlm.nih.gov/pubmed) and Google Scholar (https://scholar.google.com/) to gather information about the system, its major components and the interactions between its components. This step is so-called literature mining.

One can also consider generating high-throughput (HTP) data (e.g. sequencing data, gene expression data, mass spectrometry data) to identify significantly changed components of the system of interest under different conditions or perturbations (e.g. healthy vs diseased, pre-treatment vs post-treatment, wildtype vs mutant). Several methods have been developed to preprocess, filter and analyze HTP data (for extensive reviews, see [4, 5, 6, 7]). After supporting some of the data analysis results via existing knowledge, these results are used for hypothesis generation.

The data analysis results, specifically the significantly changed components, can be used to construct a mathematical model. A mathematical model consists of a *network* that is a coherent picture of the system, its components and their interactions, and a *set of rules* that correspond to the evolution of components over time. Various approaches have been developed to infer the interactions between these components other than literature-mining. The most common inference tools are based on machine learning methods (e.g. Bayesian networks) and correlation-based methods.

The system components and their interactions can be visualized as a directed graph in which the components are represented by nodes (e.g. genes, transcription factors, microRNAs and proteins) and their interactions are represented by directed edges (e.g. upregulation/activation/induction and downregulation/inhibition/reduction). If the nodes are genes, the directed graph is called a gene regulatory network; if they are proteins, it is called a protein-protein interaction network. A *network* is a static representation of the (dynamic) biological system.

We now need to determine how each node is regulated by others in the network. Must all regulators be active/present to regulate the target node or is at least one active regulator sufficient to regulate this target node? Such information is not always available in pertinent biological literature and should be specified as model assumptions. There are several ways to model biological networks: discrete/qualitative (boolean models, logical models, difference equation models), continuous/quantitative (differential equation models), hybrid (combining continuous and discrete)

and their stochastic variants. Using an appropriate modeling framework, regulation of all components are then translated into *a set of rules*, which is a mathematical representation of the evolution of the (dynamical) system over time. There is no one best, one-size-fits-all modeling framework; mathematicians can decide which one is the most appropriate to be used depending on the existing knowledge of the system, the scientific question of interest and size of the network.

The mathematical model, the network and the set of rules, can be simulated for several purposes such as determining time response or dynamics of the system. The model simulations has to be tested against existing knowledge of the system (published data and known facts that were not used in model construction) before hypothesis generation. This model may not be optimal or a true depiction of reality, but it is considered useful and sufficient as long as it matches well with the existing literature and its predictions can be validated. For hypothesis generation, the model is run under different conditions or perturbations: How would the whole system respond to certain perturbations? How would the system be affected if a specific node is abundant/overproduced or a specific edge (regulation) is disrupted/removed?

Hypotheses, generated by data analysis results or model simulations, are then validated by experimentation (new data generation) or existing data that was not used. The whole systems biology process is typically iterative. Some steps can be omitted depending upon the questions of interest, known facts and generated data.

While the systems biology field has been evolving towards generating, storing and analyzing different types of data (e.g. genomics, transcriptomics, proteomics, metabolomics, lipidomics) to decipher biological processes, it comes with enormous challenges such as merging different types of -omics data and integrating them all into computational models [2]. The current work does not propose method development for analyzing different types of data and incorporating them into a mathematical model for a systems-level understanding of biological processes. However, it was highly motivated by hypothesis generation and validation using techniques of comparative and integrative data analysis and mathematical modeling, specifically in medicine and environmental science. We mean to discover new components, interactions and factors, understand health and disease states, reposition approved drugs, determine drug effects and investigate environmental factors in biological processes using mathematical and statistical tools. This document consists of four manuscripts each of which emphasizes a different aspect of these goals from a systems biology point of view.

1.2 Outline

Figure 1.1 is an illustration of the systems biology approach taken in this thesis. Briefly, data is generated for a biological system of interest and a well-defined problem, and then analyzed in order to construct a mathematical model for simulations and hypothesis generation. The dissertation follows this workflow. Each chapter is its own manuscript, meant to stand alone outside the dissertation and was prepared for journal submission. The appendices also contain sufficient detail that all studies can be replicated.

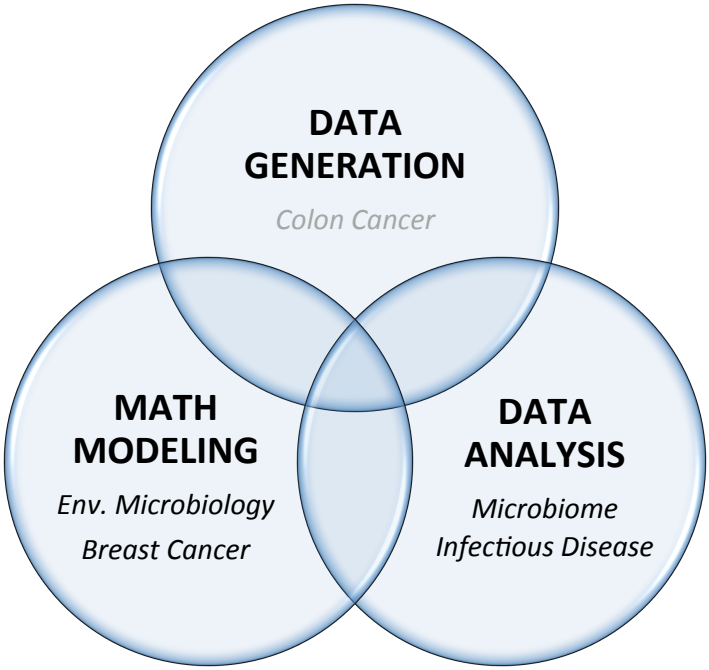

Figure 1.1: Systems biology approach taken for this study, in which two data analysis and two mathematical modeling projects are discussed. The colon cancer project is briefly mentioned.

I have been involved in the data generation aspect of systems biology, through a project on colon cancer. The goal is to identify key upstream regulators of a DNA mismatch repair (MMR) pathway and the role of microRNAs on MMR genes in colon cancer. It was not a part of this thesis, but it was discussed in Claus Kadelka's PhD thesis, "Robustness Analysis of Gene Regulatory Networks", 2015. My contribution to the study was to test model predictions through wet lab experiments and keep the network model up-to-date by active literature mining.

Chapter 2 and 3 focus on the data analysis aspect of systems biology. Chapter 2 introduces a microbiome analysis pipeline for antibiotic clinical trials. We used QIIME [8] for next generation

sequencing (NGS) data filtering and analysis, and PICRUSt [9] and STAMP [10] for hypothesis generation and visualization. To our knowledge, there have not been similar comparative studies in humans on the effects on the gut microbiome of an antibiotic when administered by different modes. This chapter is based on our published paper [11]. Chapter 3 describes gene expression data analysis and pathway analysis to find therapeutic targets for tuberculosis (TB) infection. For filtering steps (e.g. MAD score, PCA) and obtaining a gene expression profile, we used Array Studio v8.0 (OmicSoft Corporation, Cary, NC, USA). The novel feature of this study is systematic drug repositioning against human immune response for the prevention, control and treatment of TB using the Connectivity map [12]. This chapter is based on a paper in preparation for submission to PLoS Pathogens. For these studies, we acquired data (NGS data for the microbiome study and microarray data for TB study) for filtering and analysis. After data analysis results were convincingly supported by existing knowledge (published data and known facts), some pathways for the microbiome study and therapeutic drugs for the TB study were suggested for further investigation.

Chapter 4 and 5 focus on the predictive mathematical modeling aspect of systems biology. Chapter 4 describes a denitrification network model in a microbe, *Pseudomonas aeruginosa*. It is the *first* mathematical model of denitrification that predicts the effect of phosphate on its denitrification performance. This chapter is based on our published paper [13]. Chapter 5 provides an iron homeostasis pathway linked to iron utilization, oxidative stress response and oncogenic pathways to investigate how normal breast cells transition to cancer cells. To date, our normal intracellular model is the only expanded core iron model that can capture a breast cancer phenotype by overexpression and knockout simulations. This chapter is based on a paper in preparation for submission to PLoS Computational Biology. For these studies, we constructed a network (denitrification network in environmental microbiology study and expanded core iron network in breast cancer study) through literature mining. To model these networks, we used polynomial dynamical systems (PDS), a state- and time- discrete logical modeling framework described by a set of polynomials over a finite field [14]. Following model validation, computer simulations provided predictions, which were validated by either experimentation or existing knowledge that was not used for model construction.

I also involved in the study on developing a stochastic modeling framework, which was not discussed here. The paper was published as: D. Murrugarra, A. Veliz-Cuba, B. Aguilar, S. Arat, and R. Laubenbacher, "Modeling stochasticity and variability in gene regulatory networks", *EURASIP J Bioinform Syst Biol*, vol. 2012, no. 1, p. 5, 2012. My contributions to the study were to code the stochastic framework, run the simulations and edit the manuscript.

Chapter 2

Microbiome Analysis in an Antibiotic Clinical Trial

GSK1322322 (GSK‘322) is a novel antibacterial agent under development, and it has known antibacterial activities against respiratory and skin pathogens. We used next-generation sequencing (NGS) of the bacterial 16S rRNA genes from stool samples collected from 61 healthy volunteers at the pre- and end-of-study time points to determine the effects of GSK‘322 on the gastrointestinal (GI) microbiota in a phase I clinical trial. GSK‘322 was administered either intravenously (i.v.) only or in an oral-i.v. combination. We found no significant changes in the relative abundances of GI microbiota between the pre- and end-of-study samples for either the placebo- or i.v.-only-treated subjects. However, oral-i.v. treatment resulted in significant decreases in the Firmicutes and Bacteroidales, and increases in the Betaproteobacteria and Bifidobacteriaceae. To our knowledge, there have not been similar comparative studies in humans on the effects on the gut microbiome of an antibiotic when administered by different modes. Our study shows that dosing regimen is an important factor when considering the impact of antibiotic usage on GI microbiota.

This chapter is based on the published paper: S. Arat*, A. Spivak*, S. V. Horn, E. Thomas, C. Traini, G. Sathe, G. P. Livi, K. Ingraham, L. Jones, K. Aubart, D. J. Holmes, O. Naderer, and J. R. Brown, “Microbiome changes in healthy volunteers treated with GSK1322322, a novel antibiotic targeting bacterial peptide deformylase, *Antimicrob. Agents Chemother.*, vol. 59, no. 2, pp. 11821192, 2015. Aaron Spivak and I equally contributed to this paper: We (1) performed the computational analysis; (2) wrote a Python script for the microbiome analysis pipeline; (3) wrote the manuscript with Jim Brown.

2.1 Introduction

The roles of microbial organisms in human health and wellbeing are complex and multifaceted. Maintaining proper symbiosis between the human host and our endogenous gastrointestinal (GI) microbial ecosystem, or microbiome, is essential for a well-functioning immune system and may delay the onset of many chronic diseases [15, 16]. Recently, there has been increased interest in and concern about the potential damaging effects of prolonged antibiotic use on the microbiome [17, 18, 19], since oral dosing can significantly alter gut microbial communities [20, 21]. Human epidemiological studies suggest that long-term intake of antibiotics is potentially linked with later-life onset of inflammatory bowel diseases [22, 23], asthma [24], and *Clostridium difficile* infections [25]. Animal studies involving the dosing of different antibiotics show a potential disruption in intestinal homeostasis based on changes in the gut microbiota composition and bacterial metabolites [26, 27]. With the increasing global threat of drug-resistant bacteria, new antibiotics are urgently needed. However, recent studies suggest that more systematic analyses of the potential effects of these agents on the human microbiome should be carried out during drug development.

Community-acquired bacterial pneumonia (CABP) is the leading cause of infectious death in developed countries, with mortality rates of 16 to 47% in subjects hospitalized with CABP who progress to intense care unit (ICU) care [28]. GSK1322322 is a new antibiotic with a novel mechanism of action that is effective against predominant CABP pathogens, including *Streptococcus pneumoniae*, *Haemophilus influenzae*, *Moraxella catarrhalis*, *Mycoplasma pneumoniae*, *Chlamydia pneumoniae*, Legionella spp., methicillin-resistant *Staphylococcus aureus* (MRSA), and methicillin-susceptible *S. aureus* (MSSA) [29]. As a new synthetic molecule of the hydrazide class, GSK1322322 is a potent inhibitor of bacterial peptide deformylase (PDF), a highly conserved and essential bacterial metalloprotease that matures newly synthesized peptides by N-formyl group cleavage [30, 31]. GSK1322322 (GSK‘322) was shown to be well tolerated in a randomized, double-blinded, placebo-controlled, six-cohort phase I clinical trial with 62 healthy volunteers with repeat oral-intravenous (oral-i.v.) combination or i.v.-only dosing for 5 to 6 days [32, 33].

Targeted DNA sequencing of the bacterial 16S rRNA gene regions using next-generation sequencing (NGS) platforms, such as Illumina MiSeq, has been shown to be highly useful for cataloguing microbiome diversity [34, 35]. In this study, we report the results of a microbiome analysis of stool samples taken for the GSK‘322 dose-ranging phase I clinical trial in healthy volunteers [32]. Fecal samples were collected with consent from all volunteers at the pre-dosing and post-dosing end-

of-study time points. Subsequently, DNA was extracted, and Illumina MiSeq DNA sequencing analysis of the 16S rRNA variable region 4 (V4) was performed. The changes in the microbiome community before and after dosing, as well as across dosing regimens, were determined.

2.2 Results and Discussion

A total of 119 stool samples from 62 subjects were processed for DNA purification, 16S rRNA gene amplification, and amplicon DNA sequencing. A single oral-i.v. combination treated subject (with both pre-study and end-of-study samples collected) was removed because of failed PCR DNA amplification, leaving 61 subjects and 117 samples. No sample failed DNA sequencing, with the fewest reads assigned to any sample being 33,586 and the median being 64,935 reads. After preprocessing, the entire data set contained 20,744,830 sequences, of which 20,447,605 were assigned to OTUs. A total of 4,662 distinct OTUs were detected among the 117 samples. Figure 5.1 illustrates the microbiome analysis pipeline for antibiotic clinical trials.

2.2.1 Microbial Community Diversity

A comparison of α-diversity rarefaction curves (*chao1* OTU richness estimation) showed overall high levels of biodiversity for the samples categorized based on dosing regimen (placebo, oral-i.v. GSK‘322 combination, and GSK‘322 by i.v. only) and time of collection (prestudy and end of study) (Figure 2.2). The curves show a plateau trend, suggesting that the depth of coverage was sufficient to capture most of the OTU proportional abundance. However, the curves for the samples categorized as oral-i.v. combination GSK‘322 dosing at the end of the study showed markedly lower *chao1* values [36], reflecting lower OTU richness in this sample subset.

Absolute bacterial abundance cannot be determined by 16S rRNA amplicon sequencing, given the efforts to normalize the sample DNA amounts required for NGS. Assays of select bacterial species using gene specific quantitative PCR (qPCR) reactions would be required to determine absolute bacterial abundance (which was not performed as part of this study). However, it is known that the use of antibiotics with bioavailability in the intestine will result in lower bacterial cell counts, depending on the type of drug, its dosage, and level of exposure [37]. Comparisons of the α-rarefaction curves (Figure 2.2) show a much lower bacterial diversity in the end-of-study samples from oral-i.v.-dosed subjects than those dosed with i.v. only, which suggests that GSK‘322 taken

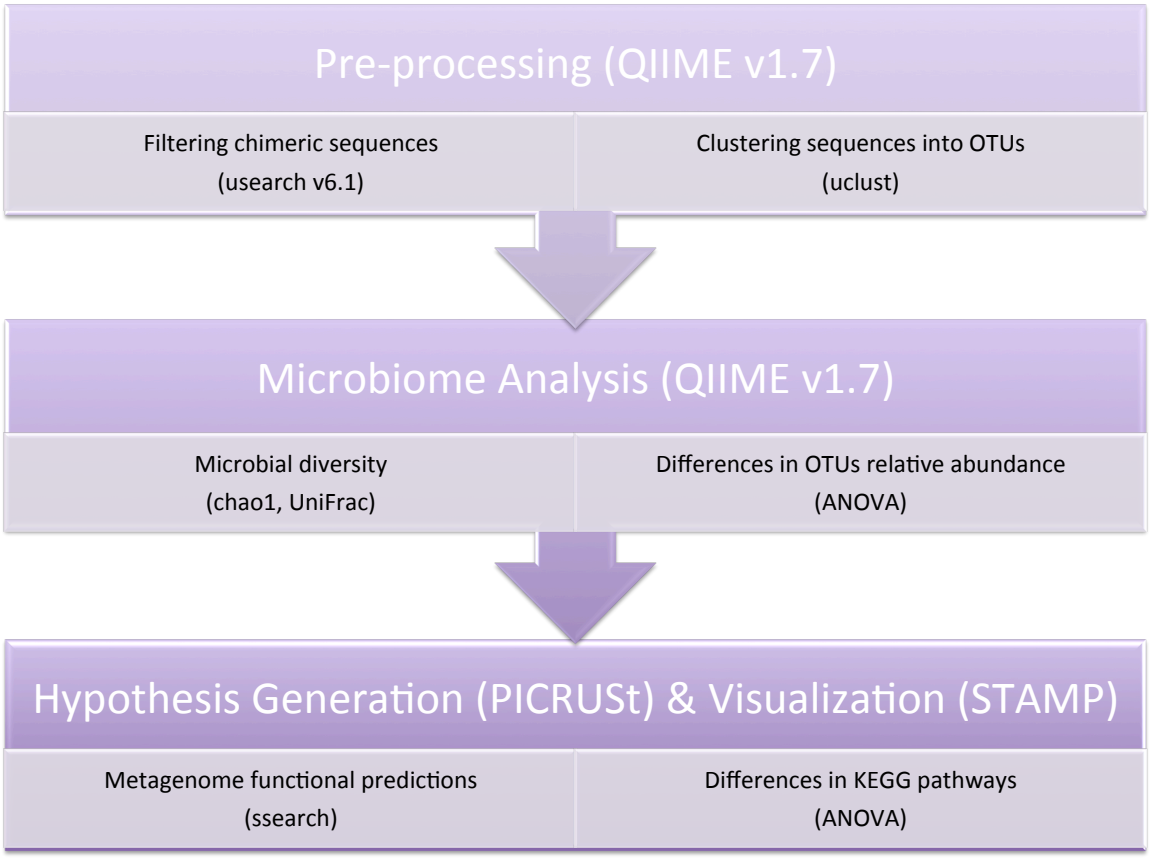

Figure 2.1: Schematic representation of the microbiome analysis pipeline synopsis for NGS data.

orally does have certain antibacterial activity in the gut. However, systemic levels of i.v.-only-administered GSK‘322 have no significant impact on the gut microbiome, similar to the results seen in placebo-treated subjects.

Using analysis of variance (ANOVA) [38] with multiple test corrections, we evaluated the potential occurrences of relative abundances with the available subject metadata variables, which included subject identifier, gender, age, body mass index (BMI), dosing levels (in mg), treatment (placebo, i.v. only, or oral-i.v. combination), and minor diarrhea adverse events (GSK‘322 was well tolerated, and no subjects withdrew from the study due to severe adverse events [32, 39]). No significant differences in the bacterial relative abundances at any OTU level (L2 through L7) were found for the subject variables of age, gender, BMI, adverse events, or subject identifier. Comparisons of pre-study versus end-of-study distributions of bacterial relative abundances at all OTU levels showed no significant differences among the subjects receiving the placebo or i.v.-only dosing regimens (Figure 2.3A). However, in similar comparisons between pre-study versus the

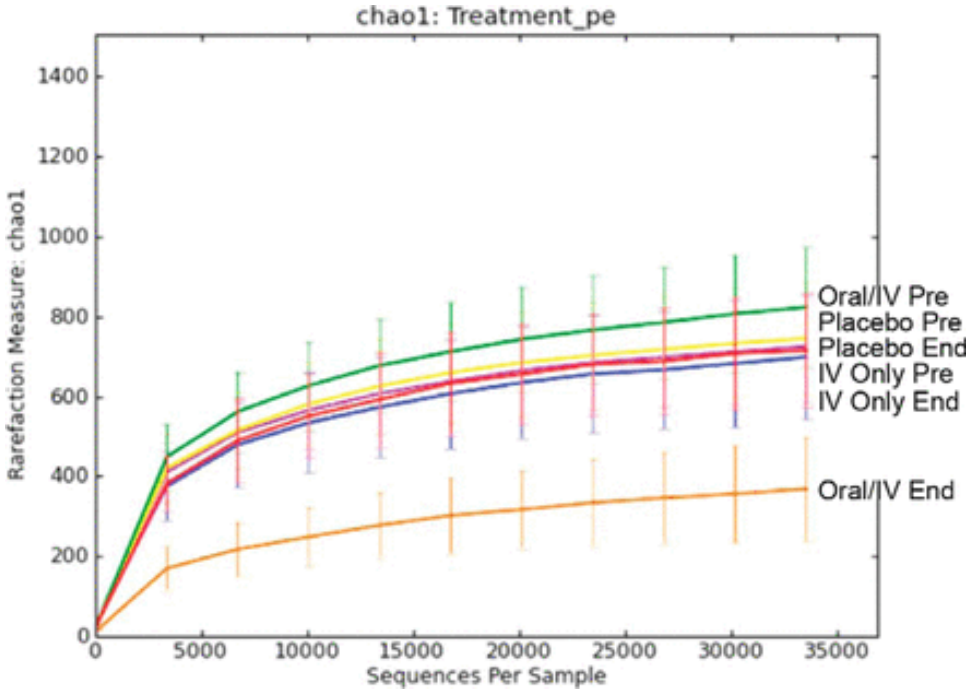

Figure 2.2: α-diversity rarefaction plots using *chao1* measure of within-group diversity.

end of the study, sample plots of relative bacterial taxon abundance show that several OTUs were markedly changed for subjects receiving oral-i.v. combination regimens (Figure 2.3B). ANOVAs with Bonferroni's correction [40, 41] for multiple tests revealed significant (P value ≤ 0.05) increases and decreases in specific OTUs (Table 2.1). The bacterial families showing significant decreases in relative abundances include various members of the phyla Firmicutes, such as Ruminococcaceae, Lachnospiraceae, and other Clostridiales families, which correlates with the antibacterial activity of GSK'322 against Gram-positive (Firmicutes) bacterial pathogens, such as *S. aureus* and *S. pneumoniae*. In addition, various Bacteroidales families were reduced. Some specific species identified as having decreases in relative abundances were *Faecalibacterium prausnitzii*, *Parabacteroides distasonis*, *Bacteroides uniformis*, and *Blautia obeum*. Other OTUs increased under oral-i.v. GSK'322 treatment, which included members of the Betaproteobacteria (Sutterella spp.), Gammaproteobacteria (Enterobacter spp.), and Bifidobacteriaceae (the species *B. pseudolongum*, *B. adolescentis*, and others).

Whether or not GSK'322 has a negative, neutral, or positive effect on gut microbiota communities is difficult to assess, mainly because our knowledge of the interplay between specific GI tract bacterial species and human health is still evolving. For example, *F. prausnitzii* has been perceived to have a protective role in the gut and in minimizing the effects of Crohn's disease [42]; however,

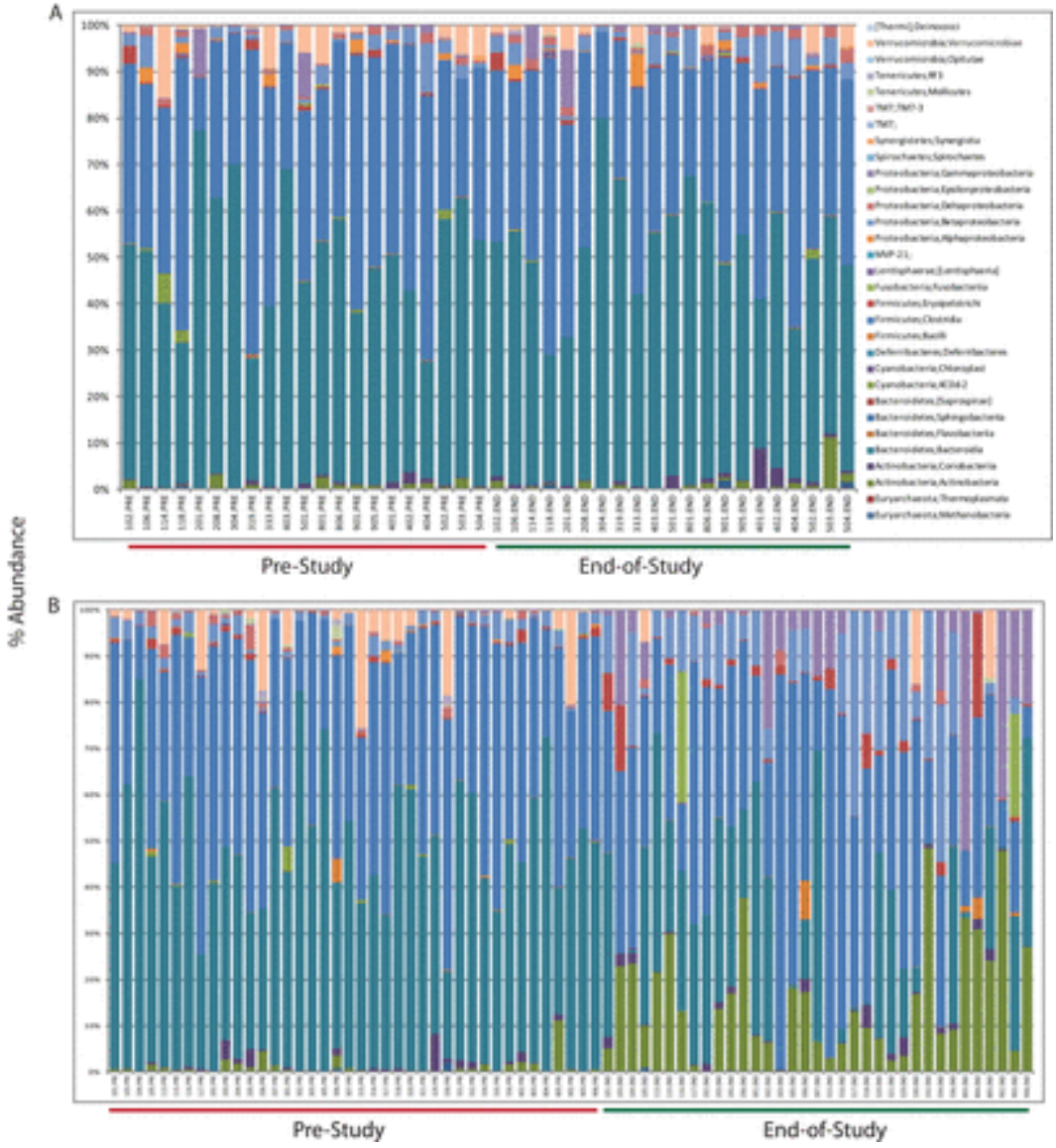

Figure 2.3: Histogram of proportional changes in bacterial OTU abundance at the class level. (A) placebo and i.v.-only dosing regimen; (B) oral-i.v. dosing regimen.

recent clinical studies have questioned the causative association of this species with disease improvement [43]. Bacteroides spp., such as *B. fragilis*, have been shown to produce polysaccharides that have beneficial immunomodulatory effects [44], while others, such as *P. distasonis*, can be carriers of multi-drug resistance loci [45]. Proteobacterial species, including Enterobacter spp., have been associated with instigating pro-inflammatory cascades leading to various disease pathologies [46]. Conversely, Bifidobacterium spp. have potential immunomodulatory properties, and several strains are actively being developed as probiotics [47]. Regardless, GSK‘322 was well tolerated in

OTU	% Change	Corr. P-value
Firmicutes; Clostridia; Clostridiales; Ruminococcaceae; Faecalibacterium prausnitzii	-4.82	1.2582E-09
Firmicutes; Clostridia; Clostridiales; Ruminococcaceae	-7.85	7.5614E-07
Bacteroidetes; Bacteroidia; Bacteroidales[Odoribacteraceae]; Odoribacter	-0.40	8.4503E-07
Firmicutes; Clostridia; Clostridiales; Ruminococcaceae; Faecalibacterium	-0.01	2.2920E-05
Bacteroidetes; Bacteroidia; Bacteroidales; Porphyromonadaceae; Parabacteroides distasonis	-0.76	3.5451E-05
Firmicutes; Clostridia; Clostridiales; Lachnospiraceae; Blautia	-0.29	6.6706E-05
Proteobacteria; Betaproteobacteria; Burkholderiales; Alcaligenaceae; Sutterella	+9.43	2.4570E-04
Actinobacteria; Actinobacteria; Bifidobacteriales; Bifidobacteriaceae; Bifidobacterium pseudolongum	+0.15	3.9237E-04
Actinobacteria; Actinobacteria; Bifidobacteriales; Bifidobacteriaceae; Bifidobacterium adolescentis	+7.58	1.1377E-03
Bacteroidetes; Bacteroidia; Bacteroidales[Barnesiellaceae]	-1.34	1.3239E-03
Firmicutes; Clostridia; Clostridiales[Mogibacteriaceae]	-0.05	1.6670E-03
Firmicutes; Clostridia; Clostridiales; Lachnospiraceae	+7.65	3.0110E-03
Actinobacteria; Actinobacteria; Bifidobacteriales; Bifidobacteriaceae; Bifidobacterium	+6.44	6.4763E-03
Bacteroidetes; Bacteroidia; Bacteroidales[Odoribacteraceae]; Butyricimonas	-0.27	2.0858E-02
Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae	+7.40	2.5116E-02
Bacteroidetes; Bacteroidia; Bacteroidales; Bacteroidaceae; Bacteroides uniformis	-4.13	2.5985E-02
Firmicutes; Clostridia; Clostridiales; Lachnospiraceae; Anaerostipes	-0.06	4.1837E-02
Firmicutes; Clostridia; Clostridiales; Lachnospiraceae; Blautia obeum	-0.01	4.6599E-02

Table 2.1: Significantly changed OTUs in subjects receiving oral-i.v.-administered antibiotic in pre- versus post-study comparisons based on mean proportional abundance.

this phase I clinical trial subject group, and gastrointestinal AEs, as monitored, were uncommon and did not prevent volunteers from completing the dosing regimen [32, 39].

The concurrent increase in some bacterial groups with the decrease of others for the end-of-study oral-i.v. dosing regimen of GSK‘322 potentially reflects dynamic changes in the gastrointestinal tract microbial ecosystem caused by antibiotic exposure. Dethlefsen and Relman [48] characterized the microbiome from three volunteers taking two courses of the antibiotic ciprofloxacin for 10 months and found profound shifts in the overall gut microbial ecosystem, reflecting potential niche replacement of one bacterial species or group by another. Similar to terrestrial and aquatic ecosystems, certain bacterial species might have more prominent roles in shaping the overall composition of the microbiome and be potential keystone species; alterations of their abundances might lead

to new niches being available for exploitation by lower-abundance species that might be from the same or different taxonomic groups [49, 50]. We can only speculate that some of the dynamic changes in bacterial species abundances induced by oral-i.v. GSK‘322 dosing regimens might reflect a disruption in interspecies interactions. While beyond the scope of the present study, it would also be interesting to monitor how the observed changes in bacterial species are reversible over time, especially the resilience of the microbiome to return to its pre-dosing state.

We assessed the overall differences between the bacterial communities of the samples, as partitioned by all available metadata variables (i.e. age, gender, BMI, subject identifier, adverse events, dose level, and treatment) using β-diversity indices calculated for unweighted UniFrac measures of OTU phylogenetic distance [51]. Principal coordinate analysis (PCoA) plots of β-diversity indices revealed that the clearest separation of bacterial communities occurred for either dosing regimen (mg) or treatment (i.e., oral-i.v., i.v. only, and placebo). Here, the end-of-study samples from the subjects receiving the oral-i.v. dosing regimen formed a well-separated cluster from all the pre-study samples, regardless of treatment, as well as end-of-study samples from either the placebo or i.v.-only-treated subjects (Figure 2.4).

2.2.2 Functional Analysis

We looked for potential functional changes in oral-i.v.-dosed bacterial communities by comparing pre-study to end-of-study matched samples using the software PICRUSt [9], which uses 16S rRNA sequence profiles to estimate metagenomic content based on reference bacterial genomes and the KEGG pathway database [52]. The most significant increased pathway representations in the end-of-study samples were membrane transport, which includes multi-drug transporters, xenobiotic metabolism and degradation, and signal transduction (Figure 2.5). The pathways with decreased presence in the samples were metabolism of terpenoids and polyketides, protein folding, sorting and degradation, and metabolism of cofactors and vitamins.

The functional analyses suggest that the greatest changes occurred in pathways potentially involving resistance mechanisms, such as efflux pumps, xenobiotic (antibiotic) metabolism, and adaptive resistance via lowering of the growth rate (i.e., slowing of metabolic pathways and DNA repair) [53, 54, 55]. However, our functional analyses are highly provisional, insofar as the computational method employed by PICRUSt infers metagenomic content indirectly based on computed linkages between 16S rRNA gene signatures and reference bacterial genomes [9]. Direct metagenomic DNA sequencing will be required to substantiate the inferences of the genomic potential of the

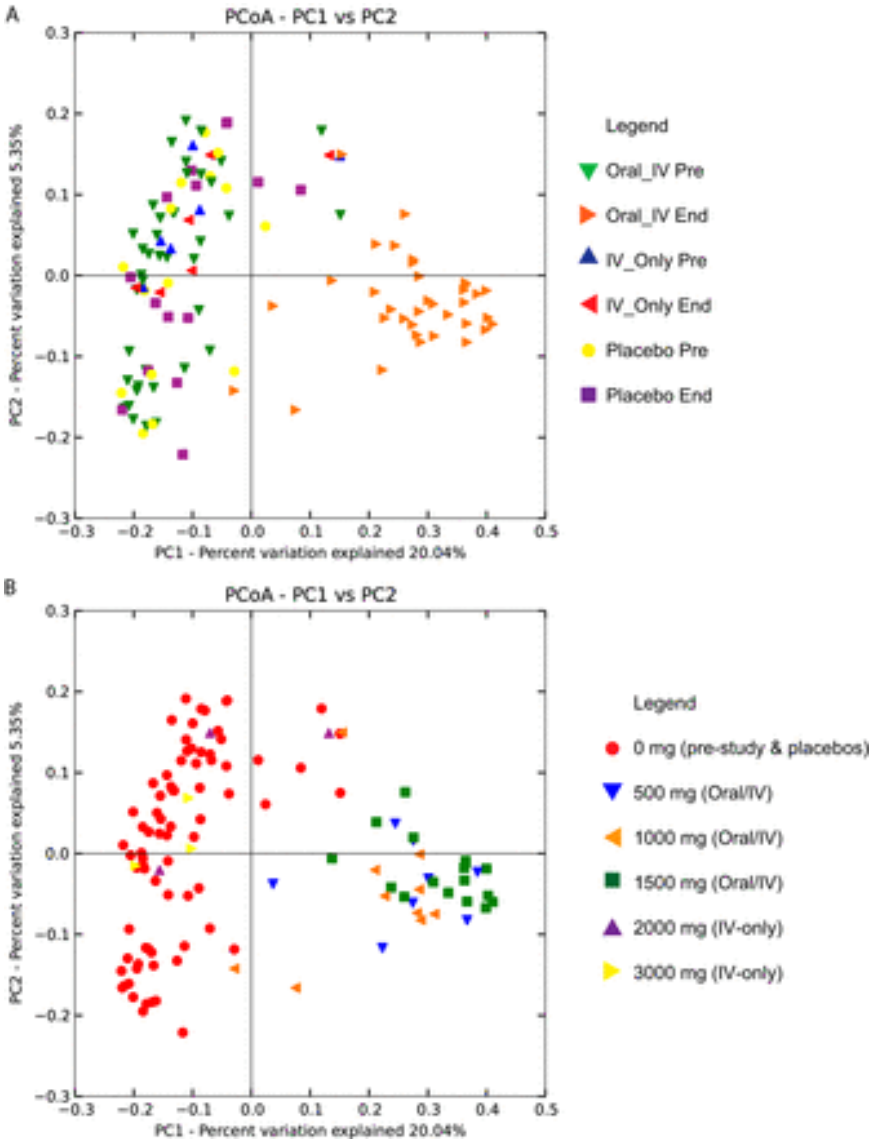

Figure 2.4: β-diversity plots of between-group diversity based on treatment (A) and dosage (B).

microbiome made here in this study.

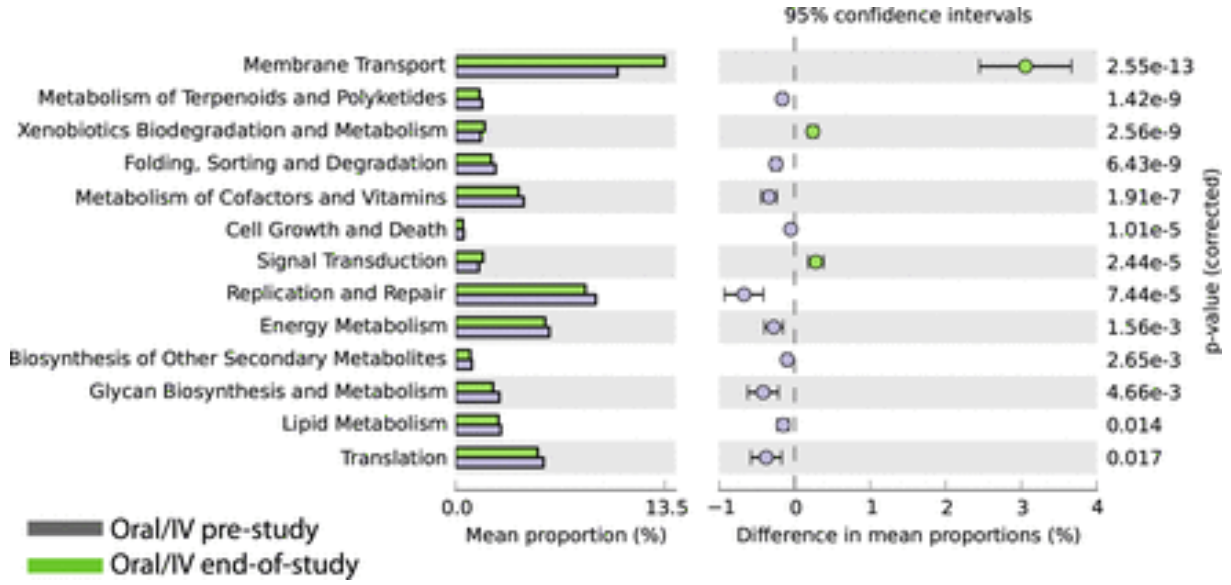

Figure 2.5: Predictions of the functional composition of oral-i.v. metagenome. Pre-study (blue bars); End-of-study (green bars).

2.3 Materials and Methods

2.3.1 Clinical Study Design

The study design and subject population were described previously by Naderer et al. [32, 39]. Briefly, adults age 18 to 65 years with a body mass index (BMI) of 18.5 to 29.9 kg/m2 and in generally good health were eligible for enrollment. Female volunteers of non-childbearing potential were also eligible. Single-dose oral tablet (500 mg each) doses of GSK‘322 (for a total of 1,000 or 1,500 mg) were administered in two cohorts (B and C) to determine absolute bioavailability, mean absorption time, and systemic exposure of oral tablet administration (Table 2.2). Single-dose escalation of i.v. GSK‘322 from 500 to 3,000 mg was administered in six cohorts (A1, A2, B, C, D, and E) to determine tolerability, dose proportionality, urinary excretion, and systemic exposure. The highest dosages, 2,000 mg and 3,000 mg, were administered in i.v.-only formulations. Repeat-dose escalation of i.v. GSK‘322 from 500 to 1,500 mg was administered in six cohorts (A1, A2, B, C, F1, and F2) to evaluate tolerability, the accumulation ratio, time invariance, and systemic exposure. In addition, two formulations for i.v. administration of GSK‘322 were evaluated for safety and tolerability: a free-base formulation (in cohorts A1, A2, B, and C) and a more stable mesylate salt preparation (in cohorts D, E, F1, and F2).

Cohort	Dosed/Placebo	Dose (mg) Protocol
A1, A2	8/4	Single dose, 500 i.v. GSK‘322/placebo → BID for 4 days
B	6/2	Single dose 1,000 oral → 1,000 i.v. GSK‘322/placebo→ BID for 4 days
C	18/3	Single dose 1,500 oral → 1,500 i.v. GSK‘322/placebo → BID for 4 days
D	3/1	Single dose 2,000 i.v. GSK‘322/placebo
E	3/1	Single dose 3,000 i.v. GSK‘322/placebo
F1, F2	8/4	1,000 i.v. GSK‘322/placebo BID for 4 days

Table 2.2: Overview of dosing regimen and clinical study design. (BID: two times a day)

Volunteers were admitted to the study unit the day before drug administration and discharged after the study procedures were completed on day 3, 5, or 7, depending on the cohort. All oral doses were administered following an overnight fast of at least 10 h. Standardized meals were served daily while volunteers were housed within the unit. Each volunteer provided written informed consent. The study was approved by an institutional review board (Western International Review Board, Olympia, WA) and was conducted in accordance with good clinical practice and the Declaration of Helsinki. Four volunteers withdrew from the study due to mild adverse events (AEs) (two patients with irritation, one patient with urticaria at the i.v. infusion site, and one patient with moderate pruritic rash) [39].

2.3.2 Sample Collection, DNA Extraction and Sequencing

With the consent of the subjects, stool samples were collected during this study for measuring the microbiome. The samples were collected pre-dose and at the end of the study treatment. The stool samples were collected with a sterile spoon, transferred into a pre-labeled stool collection tube containing 8 ml of stool DNA stabilizer, mixed by shaking, and then immediately stored frozen at $20\,^{\circ}\mathrm{C}$ prior to shipment. Care was taken to minimize any sample contamination and reduce prolonged exposure to air. For DNA extraction, the frozen stool samples were completely thawed, and DNA was isolated from approximately 1.4 ml of each homogenized sample using the PSP Spin Stool DNA Plus kit (Invitek, Berlin, Germany). DNA was isolated, as per the manufacturers instructions. Each DNA sample was quantified by spectrophotometry (NanoDrop, ND-1000; Thermo Scientific, DE, USA) prior to PCR amplification.

Multiplex bar-coded primers were used for paired-end sequencing of the 16S rRNA variable 4 (V4) region on the Illumina MiSeq platform (Illumina, San Diego, CA). We used the 16S rRNA V4 region primers 515F (5=-GTGCCAGCMGCCGCGGTAA-3=) and 806R (5=-GGACTACHVGGGT WTCTAAT-3=), as recommended by Caporaso et al. [56], for maximal coverage of bacterial phylogeny. The 16S rRNA gene primers were combined with the appropriate barcode and linker oligonucleotides. Illumina amplicon library generation was performed as described previously [56], except for the additional steps of purification of the PCR products after amplification by AMPure (Beckman Coulter, Brea, CA) and quantification by spectrophotometry (NanoDrop, ND-1000; Thermo Scientific, DE), followed by normalization using SequalPrep (Life Technologies, Carlsbad, CA). The amplified bar-coded DNAs from 119 samples were then pooled. The samples were controlled for quality and quantity using an Agilent Bioanalyzer (Santa Clara, CA) chip for the absence of primer-dimers and accurate sizing of product, as well as quantified using quantitative PCR (qPCR) (Kapa Biosystems, Wilmington, MA). The samples were diluted to a final dilution of 7 pM, combined at a 95:05 ratio with 7pM of PhiX control, and run on a MiSeq 2150 cycle run. The Illumina sequencer instrument, reagents (MiSeq reagent kit version 2 300 cycle), and pooled samples were prepared according to Illumina MiSeq protocols. Data collection and base calling were performed on the MiSeq instrument using CASAVA 1.8 (Illumina). After the removal of sequences that failed the Illumina quality filtering, the reads were converted to the FASTQ format. Sequence quality using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was determined before and after demultiplexing of the samples.

2.3.3 Microbiome Data Analysis

Additional quality filtering and analyses of multiplexed DNA sequencing reads from the forward primer were performed using the software QIIME 1.7 [8]. The reads were truncated at the base preceding the first low-quality stretch, and only reads of 75 bases long were retained. The reads were discarded if the sequence contained one or more ambiguous base calls or if the barcode sequence contained any mismatch errors. PCR chimera filtering was accomplished using usearch version 6.1 [57]. The closed-reference QIIME protocol was used with the UCLUST method [57] to select operational taxonomic units (OTUs). The sequences with 97% identity were clustered together. A representative sequence from each cluster was used to identify the bacterial taxa from the May 2013 edition of the Greengenes 16S rRNA database [58, 59]. OTUs containing fewer than two sequences were discarded, and sequences with 60% similarity to those in the Greengenes

database were also discarded to remove potential contaminants from the data set.

The OTU table was rarefied to a depth of 33,582 sequences, and the resultant table was used for diversity analyses, as per the recommended guidelines. β-diversity was estimated using the UniFrac metric to calculate the distances between the samples and visualized by principal coordinate analysis (PCoA) [60]. Analyses of variance (ANOVAs) [38] of the treatment effects and each OTU category (L2 [phylum] to L7 [OTU]) were calculated using QIIME 1.7 [8] and custom R scripts.

Potential changes in the microbiome at the functional level were determined using the software PICRUSt [9], with default settings, and the Kyoto Encyclopedia of Genes and Genomes (KEGG) database release 70.0 [61, 52], and they were visualized using STAMP [10]. The human-specific pathways were removed from the results to focus on true bacterial pathways. Bonferroni-corrected P values of 0.05 were used to determine the statistical significances for all analyses.

2.4 Conclusion

In this study, we evaluated the effects of a novel antibacterial drug, GSK1322322 (GSK‘322), on the human gut microbiota in a randomized dose-escalating phase I clinical trial with matched placebo controls in healthy volunteers. We found the most significant effect on microbiota was the method of drug administration. No significant differences in relative bacterial abundances were observed for either the placebo or i.v.-only-treated subjects. However, significant changes in relative bacterial abundances and distinct β-diversity clustering were found for the oral-i.v. combination treatment. Our study adds to the growing body of literature on the potential short- and long-term effects of antibiotics on the gut microbiome in several important ways. First, our longitudinal interventional study design with matching pre-study and end-of-study samples from the same subjects on oral-i.v., i.v.-only, and placebo dosing regimens allowed us to evaluate the potential effects of GSK‘322 on the microbiome in terms of changes within individuals as well as by the drug delivery method. Previous studies on the effects of antibiotics on human microbiota have compared population cohorts delineated by treatment or nontreatment, which, although informative, can be confounded by interindividual variation in personal microbiome compositions [62]. Other reports of the longitudinal monitoring of microbiome changes over time under antibiotic treatment have been limited to much smaller numbers of human subjects than those used in our study [48].

Second, GSK‘322 is a new generation of antibiotics with a very specific mechanism of action and a known bacterial target, peptide deformylase (PDF), which is essential for bacterial survival and

proliferation [63]. Additional insights are gained from the fact that PDF is an unexploited bacterial target, and there has been little clinical exposure to PDF inhibitors [64]. Therefore, we were able to evaluate the effects of GSK‘322 on the microbiota in terms of the potential variation in the target sequence and pathways across bacterial taxa without the confounding effects of selection for resistance due to the historical clinical usage of this class of drugs.

In the growing discourse over the potentially negative effects of long-term antibiotic exposure on beneficial microbes in our GI tract, it is important to consider the route of drug administration into the human body [17]. To our knowledge, there have not been similar comparative studies in humans on the effects on the gut microbiome of an antibiotic when administered by different modes. Zhang et al. [65] inoculated mice with antibiotic-resistant tet(M)-carrying Enterobacter spp. or blaCMY-2-carrying *E. coli* and then administered different doses of either tetracycline hydrochloride or ampicillin by oral and i.v. routes. They found that mice receiving orally administered antibiotics had significantly higher levels of antibiotic-resistant bacterial strains, while i.v.-treated animals had negligible changes. These differences might be due to renal versus gastrointestinal tract secretion of i.v. and orally administered antibiotics, respectively. The greatest current medical need for new antibiotics is to treat emerging drug-resistant bacterial strains in hospitalized critical care patients. In this patient population, i.v. administration is the most likely delivery route for these future drugs, which will potentially minimize their impact on the GI microbiome and mitigate any selection pressure favoring the propagation of antibiotic resistance genes.

Our study shows specific changes in the microbiome associated with the novel antibacterial compound GSK‘322. Importantly, the effects are mitigated by the drug delivery method, and no significant changes in the relative abundances of any bacterial taxa were observed for subjects treated with the i.v.-only dosing regimen. Dosing with an oral component did result in significant changes in some bacterial taxa and shifted overall bacterial community diversity measures away from those of samples taken from subjects either pre-study for placebo, i.v.-only, and oral-i.v. dosing or at the end of the study for placebo and i.v.-only dosing. However, while the changes in the microbiota were subtle, the differences were significant and specific to particular bacterial groups. Future microbiome analyses involving different antibiotics in clinical trials will be valuable for understanding potential interactions between antibiotics and human gastrointestinal microbial communities.

2.5 Acknowledgments

The original clinical trial and this microbiome study were entirely funded by GlaxoSmithKline. We all worked for GlaxoSmithKline at the time this work was performed.

Chapter 3

Meta-analysis of Human Immune Response to Tuberculosis

As the etiologic agent of tuberculosis (TB), *Mycobacterium tuberculosis* (Mtb) infects approximately 2 billion people worldwide and is the leading cause of morbidity and mortality due to infectious disease. Current standard of TB therapy involves a regimen of four antibiotics: isoniazid, rifampicin, pyrazinamide, and ethambutol. Although effective at the onset of administration, the confounded intervention and prolonged period of treatment quickly present as challenges for both patient compliance and the notorious multidrug-resistance (MDR) of Mtb. Therefore, there is an imperative need to develop novel therapeutic approaches for the prevention, control and treatment of TB by targeting the host factors essential for pathogen invasion. Here, we perform an integrative analysis of publicly available human gene expression data to investigate immune response to TB infection. We identify host pathways that are significantly enriched across different comparisons in pulmonary and latent TB infections and proposed several opportunities of repositioning 13 existing drugs toward the TB therapeutics. Although 6 of these potential repositioned drugs are already involved in TB-related studies, the remaining 7 are approved drugs for the treatment of cardiovascular disease and neurological disorders.

This chapter is based on a paper draft: S. Arat*, Z. Wang*, J.R. Brown, and M. Magid-Slav, "Meta-analysis of human host gene expression response to *Mycobacterium tuberculosis* infection", *PLoS Pathogens* (in preparation). Zhang Wang and I contributed equally to this paper. We (1) played a major role on the computational analysis of the study; (2) wrote R scripts for filtering the data; (3) analyzed and interpreted the results; (4) wrote the manuscript.

3.1 Introduction

As the etiologic agent of tuberculosis (TB), *Mycobacterium tuberculosis* (Mtb) infects approximately 2 billion people worldwide and is the leading cause of morbidity and mortality due to infectious disease [66]. Tuberculosis principally occurs in the lung, but could also infect other body sites such as kidney, spine and brain and become fatal if not properly treated. Both a latent and an active phase exist for the TB infection. The majority of infected individuals are thought to have a latent infection, in which they are asymptomatic with no clinical evidence of disease for years or decades. During the latent infection, Mtb is phagocytosed by macrophages, which trigger host immune response and the recruitment of additional macrophages and monocytes that ultimately form an organized structure called granuloma. Mtb was kept dormant and non-replicative within granuloma that suppresses the immediate threat of active infection and successfully evades further immune response. Approximately 5-10% of latent patients will go on to develop active TB, in which Mtb would provoke an active host immune response should the host be immunosuppressed. And when the granuloma breaks down, the quarantined Mtb will be released to the airways and transmitted to the next host, leading to the spreading of the infection. Current standard of TB therapy involves a regimen of four antibiotics: isoniazid, rifampicin, pyrazinamide, and ethambutol, taken with an initiation of 2 months and a continuation phase of 4-7 months. Although effective at the onset of administration, the confounded intervention and prolonged period of treatment quickly present as challenges for both patient compliance and the notorious multidrug-resistance (MDR) of Mtb. Therefore, there is an imperative need to develop novel therapeutic approaches for the prevention, control and treatment of TB.

The rapid emergence of antibiotic resistance that outpaced the development of novel antibacterials has brought the attention of a new therapeutic paradigm by targeting the host factors essential for pathogen invasion [67, 68]. Bacterial pathogens exploit host cells for its survival and proliferation through modulating multiple host pathways. For example, gram negative bacteria such as Escherichia, Salmonella, Shigella and Yersinia invade host cells through manipulating the host heterotrimeric GTP-binding protein (G protein) signaling, ubiquitination and MAPK-signaling pathways leading to cytoskeletal remodeling. Several intracellular bacterial pathogens like *Legionella pneumophila*, *Brucella abortus*, and *Salmonella enteric* secrete virulence factors to manipulate the host phosphoinositide-3-kinase (PI3K) lipid signaling pathway. Also, specific actin-rich pedestal structures made from host microtubules and other cytoskeleton components are essential for the infection of enterohemorrhagic and enteropathogenic *Escherichia coli* and some

intracellular pathogens.

Targeting the host-pathogen interactions has its advantages over the traditional antibacterials on therapeutic intervention of infectious diseases in several ways. Firstly, unlike the fast evolvement of antibiotic resistance of pathogens, host pathways are potentially less susceptible to the strong selective pressure that drives the bacterial genome evolution to an extreme. Consequently, compounds targeting the host factors could be more recalcitrant to the rapid development of drug resistance as compared to antibiotics. Secondly, compared with antibacterial drugs, there are more existing and developable compounds targeting host related factors in the current pharmaceutical collections, thereby offering greater opportunities for repositioning known drugs to new indications. At the mean time, a potential drawback of such host-directed compounds could be the toxic side effects to the host by mis-targeting undesirable host factors or compromising the host immune system. Therefore, a precise and comprehensive understanding of host-pathogen interactions is required to carefully select host factors as therapeutic targets.

Several studies have generated extensive human microarray gene expression datasets for subjects infected with M. tuberculosis with the primary objective of developing diagnostic biomarkers for detection of TB infections [69, 70, 71, 72, 73]. Few studies have systematically searched for potential therapeutic opportunities aimed at targeting human cellular pathways during *M. tuberculosis* infection. We have previously demonstrated the possibility of identifying host targets in response to both bacterial and viral infections using public microarray datasets [74, 75]. Here we adopted a similar computational biology framework and performed a comprehensive meta-analysis of host gene expression profiles in response to both latent and active TB infections. We identified host pathways that were significantly enriched across different comparisons and proposed several opportunities of repurposing existing drugs toward TB treatment/therapeutics.

3.2 Results and Discussion

3.2.1 Dataset Selection and Quality Filtering

We employed an iterative process of database querying, data filtering and common pathway analysis across all datasets (Figure 5.1). At the time of this study, there were 39 Gene Expression Omnibus (GEO) datasets for human gene expression during either latent or active TB infection. After filtering based on inclusion criteria (see Material and Methods), a total of 11 unique datasets

(GSE19435, GSE19439, GSE19444, GSE29536, GSE34608, GSE41055, GSE42826, GSE54992, GSE56153, GSE65517, GSE62525) were retained. GSE41055, GSE42826, GSE56153 and GSE62525 were excluded because of the extreme non-normal distribution of their raw data as suggested by kernel density. A final set of 7 datasets with 7 human gene expression profiles for active/pulmonary tuberculosis (PTB) comparisons and 3 profiles for latent tuberculosis (LTB) comparisons were selected for downstream analysis (Table 3.1).

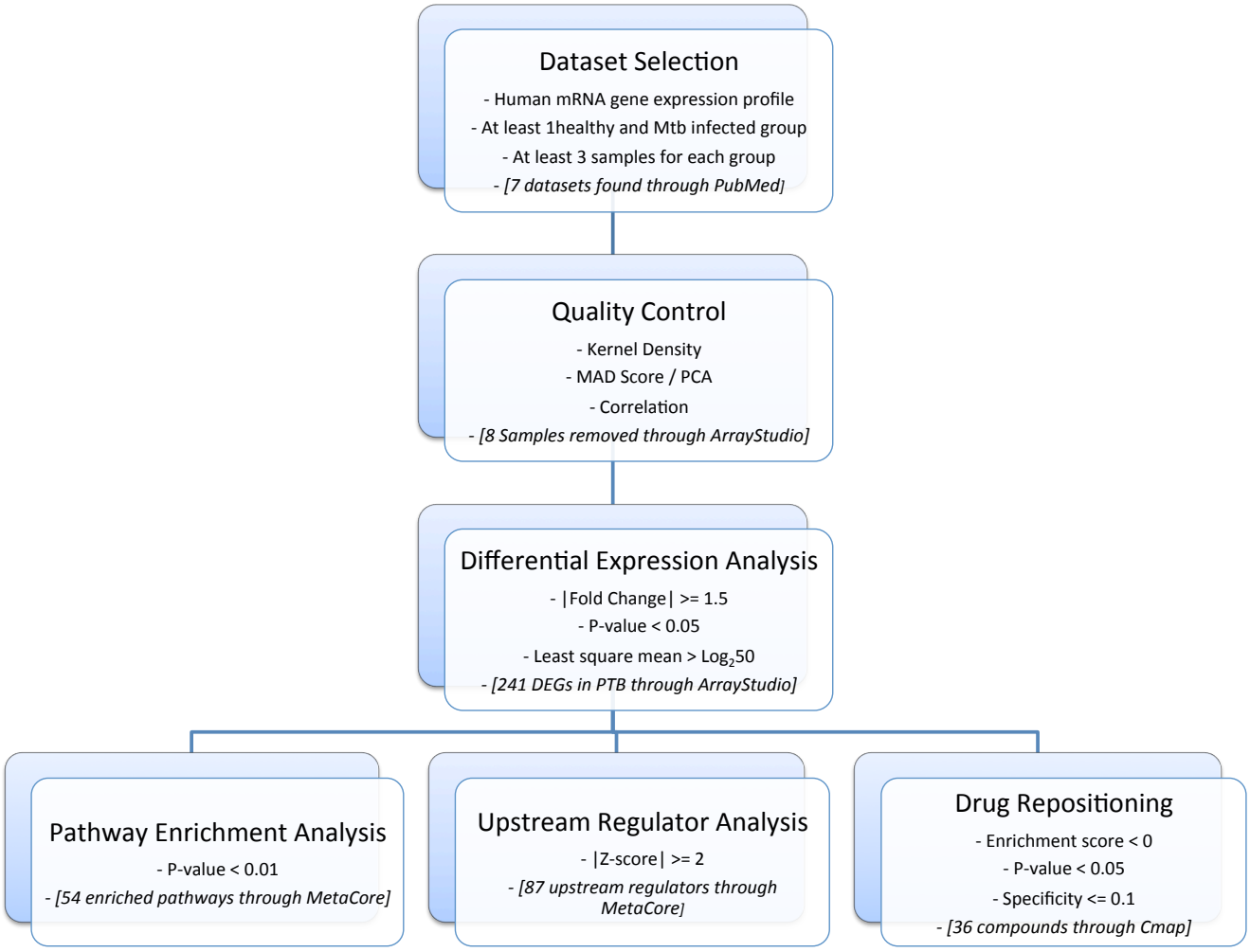

Figure 3.1: Outline of the iterative analysis process.

Within each dataset, we further filtered out potential outliers using our quality control (QC) criteria in Array Studio v8.0 (OmicSoft Corporation, Cary, NC, USA). For example, sample GSM484369 from GSE1943 failed to meet the criteria for both pairwise correlation and kernel density and were thus excluded from downstream analysis. Similarly, samples GSM484458 and GSM484465 in

Datasets	Reference	Cell line/Patient sample	Platform	PTB	LTB	CTRL	Outliers
GSE19435	[69]	PBMCs	Illumina	12	0	7	1
GSE19439	[69]	PBMCs	Illumina	17	13	6	2
GSE19444	[69]	PBMCs	Illumina	20	20	12	0
GSE29536	[70]	Blood samples	Illumina	9	0	6	0
GSE34608	[71]	Blood samples	Agilent	8	0	18	2
GSE54992	[72]	PBMCs	Affymetrix	9	6	6	3
GSE65517	[73]	PBMCs	Illumina	3	0	3	0

Table 3.1: Profiles of the GEO datasets that pass all criteria filters.

GSE19439, and samples GSM851876 and GSM851889 in GSE34608 were filtered out as outliers in both MAD scores and pairwise correlation. In total, 58 control, 39 latent and 78 active samples from all the 7 datasets passed the QC criteria for differential expression analysis.

3.2.2 Differential Expression Analysis

Differential expression analysis was performed for each of the 7 datasets using Array Studio v8.0 (OmicSoft Corporation, Cary, NC, USA). The gene expression levels in each comparison group were $\log_2$ transformed and fit into a general linear model. A moderated t-test was performed to identify differentially expressed genes (DEGs) with a False Discovery Rate (FDR) corrected p-value < 0.05, a fold change of 1.5 or above in either direction and a least square mean (LSM) greater than $\log_2 50$. For LTB comparisons, we used the raw p-value as criteria since no gene was retained under the adjusted p-value cutoff for all datasets. Genes were further ranked by the number of comparison groups in which they were differentially expressed at the same direction. For the PTB comparisons, there were 241 DEGs that were significantly up- or down-regulated in at least 4 out of 7 datasets, with LAP3, JAK2, COP1, C1QB and PARP9 identified as consistently unregulated in all datasets. Figure 3.2 shows a heatmap of differentially expressed genes for PTB comparisons. Blue represents significantly down-regulated genes and yellow represents significantly upregulated genes. The genes that were not significantly changed are colored in white. On the other hand, only one gene, FCGR1B, was identified as significantly up-regulated in 2 out of the

3 LTB comparison groups, indicating almost unobservable host responses were provoked during the latent TB infections.

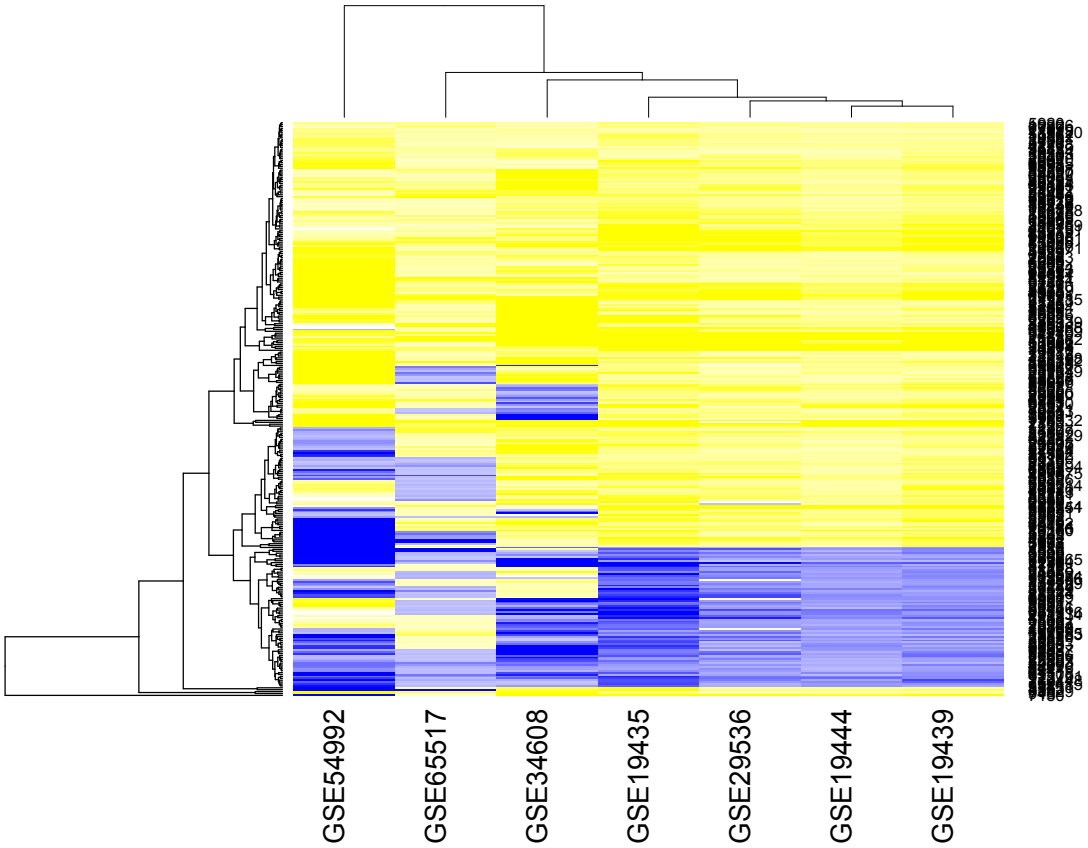

Figure 3.2: Differential gene expression profile. FC: fold change; Blue: FC $<$ -1.5; Yellow: FC $>$ 1.5; White: -1.5 $\leq$ FC $\leq$ 1.5

3.2.3 Pathway Enrichment Analysis

Differentially expressed genes (DEGs) from each comparison group were then analyzed for enriched functional pathways using MetaCore v5.0 from Thomson Reuters. In total, there were 22 pathways significantly enriched in at least 5 out of the 7 PTB comparison groups (p-value$<$ 0.01) (Table 3.2), most of which were involved in host immune and inflammatory responses. Of these, two pathways, Role of PKR in stress-induced antiviral cell response and Bacterial infections in CF airways, were significantly enriched in all the 7 PTB comparisons. 19 out of 22 genes that were connected with PKR in the Role of PKR in stress-induced antiviral cell response pathway were

identified as DEGs in at least one dataset, indicating a central role of PKR as both regulator and effector in host response to active TB infections. Notably, several pathways not usually associated with immune response were also identified as pervasively enriched, such as Cytokine-mediated regulation of megakaryopoiesis, Thrombopoetin signaling via JAK-STAT pathway and Integrin inside-out signaling in T cells.

Pathway	**TB-related Study**
Immune response: Role of PKR in stress-induced antiviral cell response	x
Bacterial infections in CF airways	x
Apoptosis and survival: Apoptotic TNF-family pathways	x
Immune response: IL-1 signaling pathway	x
Immune response: HMGB1/TLR signaling pathway	x
Development: Cytokine-mediated regulation of megakaryopoiesis	-
IL-6 signaling in multiple myeloma	-
Immune response: Bacterial infections in normal airways	x
Development: Thrombopoetin signaling via JAK-STAT pathway	-
Immune response : IFN gamma signaling pathway	x
Signal transduction: PTMs in IL-12 signaling pathway	x
Immune response: IL-15 signaling via JAK-STAT cascade	x
Immune response: IFN alpha/beta signaling pathway	x
Immune response: IL-10 signaling pathway	x
Immune response: Generation of memory CD4+ T cells	x
Immune response: IL-18 signaling	x
Stimulation of TGF-beta signaling in lung cancer	x
Cell adhesion: Integrin inside-out signaling in T cells	x
Development: Growth hormone signaling via STATs and PLC/IP3	-
Development: Prolactin receptor signaling	-
LRRK2 and immune function in Parkinson's disease	$\sim$
Immune response: Inflammasome in inflammatory response	x

Table 3.2: The common enriched pathways in at least 5 PTB datasets.

Almost all the pathways are already studied in TB-related research except cytokine-mediated regulation of megakaryopoiesis, IL-6 signaling in multiple myeloma, thrombopoetin signaling via JAK-STAT pathway, growth hormone signaling via STATs and PLC/IP3, and prolactin receptor

signaling. LRRK2 gene (and immune function in Parkinson's disease) and its association with susceptibility to TB infection has been being investigated by Gutierrez Group supported by Michael J. Fox Foundation since 2013.

TB is also considered in the context of excessive cytokine production that might trigger immunoendocrine response. del Rey and her colleagues highlighted that several hormone levels were altered in active TB patients compared to healthy controls [76]. Although growth hormone and prolactine hormone plasma levels are elevated in this study, they are not in our DEGs list; yet growth hormone (GH) signaling and prolactin (PRL) receptor signaling pathways were significantly enriched in 5 PTB comparisons. Some studies suggest that GH and PRL stimulate cell-mediated immunity by activating macrophages human phagocytic cells and might be involved in the progression of TB.

In comparison, analysis of differentially expressed genes for the 3 LTB comparison groups reveals a total of 19 pathways that were significantly enriched in either one of the three groups. However, none of these pathways was shared by any two groups.

3.2.4 Upstream Regulator Analysis

For each comparison group, we also looked for upstream regulators that are highly connected with and predicted to regulate DEGs in each dataset using MetaCore. The connectivity of each upstream regulator was assessed by z-score, which essentially measures the deviation of the actual number of connections in the pathways from the expected number of connections from hypergeometric distribution. A number of 87 genes were identified as highly connected upstream regulators in at least 4 out of the 7 PTB comparisons (average z-score ≥ 2), 28 of which were highly connected in all the 7 comparisons (Table 3.3. Among all the 7 PTB comparisons, the upstream regulator with the highest average z-score was IRF1 (14.6), which was followed by IRF8, MYC, CREB1, STAT1 and IRF4. As members of the interferon regulatory factor (IRF) family, both IRF1 and IRF8 are implicated in host resistance to intracellular infection and known to play a critical role in the pathogenesis of TB. Yamada et. al. investigated the role of IRF-1 in the mycobacterial infection in vivo and found that IRF-1 knockout mice were highly susceptible to Mtb infection [77]. Similarly, Marquis et al. suggested that mice bearing a defective IRF-8 allele were extremely susceptible to *M. tuberculosis* with a rapid and uncontrolled Mtb replication in various tissues [78]. IRF8 acts as a co-activator of IRF1 to stimulate transcription activation of IFN-gamma during host response to TB. Consistently, our results suggest that both IRF1 and IRF8 are important regulators of host immune response to TB by connecting to the largest number of DEGs in the pathways.

In addition, BCL6, MYC, NFKB, STAT1 and STAT2 were among the 241 DEGs in at least 4 out of the 7 comparisons, providing independent evidence for the role of these genes in controlling the downstream effectors through their own expression levels. Not surprisingly, a dramatically reduced set of 13 upstream regulators were identified in either one of the three LTB comparisons, with SP1 being the only gene shared by at least two datasets (GSE54992 and GSE19439).

3.2.5 Drug Repositioning Analysis

The Connectivity map (Cmap) is a powerful tool for drug repurposing based on the similarity of gene expression signatures between diseases and compounds [12]. We performed a Cmap analysis to associate the gene expression profile of each of the PTB comparison groups with the gene expression signature of all public compounds, to look for potential opportunities of repurposing existing drugs to TB therapeutics. A maximum of 250 up-regulated and 250 down-regulated genes in each comparison group was used to generate the gene expression signature. And the compounds that have a gene signature inverse to the disease signature were identified. A total of 36 compounds were identified in at least 2 out of the 7 PTB comparisons. 11 of which is approved compounds in the DrugBank database (Table 3.4). Among them, the compound sulfadimidine had the highest prevalence by being present in 5 comparisons. This was followed by oleandomycin, harmaline and sulfacetamide which were present in at least 3 PTB datasets.

These 11 approved drugs can be classified in 4 groups based on their treatment area: infectious disease, cardiovascular, cancer and neurological disorder. The drugs used to treat infectious disease are already used in TB-related research. Since we are mainly interested in human host-targeting drugs, the drugs to treat cardiovascular diseases (Ajmaline, debrisoquin, methoxamine), neurological disorders (phensuximide, trazodone) and cancer (aminoglutethimide) can be further investigated as a potential therapeutics for TB infection. On the other hand, no compound was found to be shared by at least 2 out of the 3 LTB comparisons, suggesting a lack of consistency in gene expression signature during latent TB infections.

3.3 Materials and Methods

3.3.1 Data Sources, Selection and Quality Control

The National Center for Biotechnology Informations (NCBI) Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/ accessed May 2015) was queried for human mRNA datasets involving *Mycobacterium tuberculosis* (Mtb). Filtering steps reduced the number of datasets to 7 datasets (Table 3.1). GEO datasets were selected based on the following inclusion criteria:

- Gene expression profile is derived from human cells and probed using a human-based genome array platform and not other species
- The array platform is supported by either Illumina, Agilent or Affymetrix
- The bacterium infected is wild-type Mtb
- There is at least one control group and treatment group
- The control group consists of only healthy volunteers
- The treatment group consists of patients only infected by Mtb
- Each treatment group and control group have at least three samples

Here, group is a collection of blood samples, each of which originates from their own array chip. Control group is a collection of blood samples from healthy volunteers whereas treatment group is a collection of blood samples from Mtb patients.

The raw expression data was processed using ArrayStudio v8.0 (OmicSoft Corporation, Cary, NC, USA). The raw data, the study design table and the annotation table were imported from the GEO database for each dataset. The raw data was assumed to be normalized and then the intensity values were log2 transformed. Samples irrelevant to the main study design were marked as excluded from our downstream analysis such as Sarcoidosis (SARC) patients samples in the dataset GSE34608.

Quality control (QC) analysis were performed and the samples that were outliers in kernel density [84, 85], principle component analysis (PCA) [86, 87], median absolute deviation (MAD) score [88] and/or pair-wise Pearson correlation plots were excluded from the further statistical analysis.

3.3.2 Differential Expression, Pathway and Upstream Regulator Analyses

A comparison group means the treatment group compared to a control group. Each dataset has 7 PTB comparison groups and 3 LTB comparison groups in this study. The probes in the comparison groups were mapped to their corresponding genes based on the following criteria:

- The absolute value of the $\log_2$ fold change of the probe is greater than 1.5
- P-value of the probe is less than 0.05
- The least square mean (LSM) [89] of the probe is greater than $\log_2 50$

If more than one probe mapped to the same gene, then the probe with the highest magnitude fold change value was used for that gene. The mapped genes satisfying the three criteria above, so-called differentially expressed genes, from each comparison group were analyzed for generating enriched pathways and upstream regulators through the use of MetaCore. Pathway significance/enrichment was defined as a pathway with its p-value is less than 0.01 (i.e. pathway $\log_{10}$(p-value) > 2). The enriched pathways were ranked by their frequency, which is defined by the number of comparison groups in which it appeared. Similar to the pathway analysis, upstream regulator significance was defined as an upstream regulator with the absolute value of its activation z-score is at least 2. The upstream regulators were also ranked by their frequency, which is defined by the number of comparison groups in which it is consistently activated or inhibited. An upstream regulator that is activated in one comparison group and inhibited in another comparison group was considered as inconsistent and excluded from further analysis.

3.3.3 Drug Repositioning

Repositioned drug candidates for the differentially expressed genes in each comparison group were analyzed using the Connectivity Map (Cmap) [12]. The Cmap results were filtered based on the following criteria:

- There is at least 3 samples
- The enrichment score of the compound is less than 0
- The p-value is less than 0.05

- The specificity of the compounds is no more than 0.1

The targets, mechanism of actions and indications of compounds were determined from the National Center for Biotechnology Informations PubChem (http://pubchem.ncbi.nlm.nih.gov/ accessed July 2015) and DrugBank (http://www.drugbank.ca/ accessed July 2015) [90]. The compounds used as a veterinary medicine or for diagnostic purposes or pain management were ignored. GoPubMed (http://www.gopubmed.org/web/gopubmed/ accessed July 2015) was utilized to detect which Cmap compounds were already tested in TB-related research.

3.4 Conclusion

In our study, we used computational biology techniques to discover novel human-host responses and drug targets to active and latent TB infections through an integrative data analysis pipeline for public human GEO datasets. Based on this analysis as well as criteria for clinical importance, we identified several differentially expressed genes, enriched pathways, upstream regulators and potential drugs for the treatment of active TB. To our knowledge, it is the first integrative study on drug repositioning for TB therapeutics. In comparison, very few host processes were triggered in response to the latent TB infections. This is consistent with the understanding that a paucity of host immune response was provoked by the Mtb during the latent stage, and only when progressing to the active state would the host immune responses be activated.

Pigment epithelium derived factor (PEDF a.k.a serpin F1) is a potent neurotropic, anti-angiogenic, anti-tumorigenic and immunomodulatory protein that involves activation of ERK and NFkB signaling pathways. PEDF signaling pathway was significantly enriched in 4 PTB comparisons (not shown in Table 3.2). PEDF was discovered as a neurotropic factor in retinal cells in 1989 and following studies show that PEDF is expressed in multiple body regions including the plasma. PEDF has been suggested to a therapeutic target for several diseases including cardiovascular disease, diabetes and cancer. In light of disease areas of Cmap approved drugs, PEDF signaling pathway can be a potential target for TB therapeutics.

3.5 Acknowledgments

We acknowledge GlaxoSmithKline R&D for their support of this study.

Gene symbol	Average z-score	DEG list
IRF1	14.603	-
IRF8	14.065	-
MYC	14.061	x
CREB1	12.605	-
STAT1	12.042	x
IRF4	11.475	-
SPI1	10.267	-
RELA	8.514	-
TP53	8.276	-
FOXP3	7.646	-
NR3C1	7.6433	-
REL	7.558	-
SP1	7.179	-
ESR1	7.064	-
RELB	6.996	-
STAT3	6.858	-
NFKB1	6.834	-
EGR1	6.706	-
CUX1	6.681	-
FOXO1	6.515	-
E2F1	6.321	-
HIF1A	5.893	-
HSF1	5.595	-
GATA3	5.479	-
JUN	5.0852	-
GABPA	4.928	-
TP63	4.926	-
SRF	4.564	-

Table 3.3: The common upstream regulators in all PTB datasets.

Compound	Description	Disease Area	TB-related Study
sulfadimidine (sulfamethazine)	Anti-infective agent	Infectious disease	[79]
oleandomycin (troleandomycin)	Antibiotic	Infectious disease	[80]
harmaline (ajmaline)	Antiarrhythmic agent	Cardiovascular disease	-
sulfacetamide	Anti-infective agent	Infectious disease	[81]
aminoglutethimide	Antineoplastic agent	Cancer	-
debrisoquine	Antihypertensive agent	Cardiovascular disease	-
lomefloxacin	Antibiotic	Infectious disease	[82]
methoxamine	Antihypotensive agent	Cardiovascular disease	-
phensuximide	Anticonvulsant	Neurological disorder	-
sulfafurazole (sulfisoxazole)	Antibacterial agent	Infectious disease	[83]
trazodone	Antidepressive agent	Neurological disorder	-

Table 3.4: The approved Cmap compounds in the DrugBank database.

Chapter 4

A Denitrification Network Model of *Pseudomonas aeruginosa*

Pseudomonas aeruginosa is a metabolically flexible member of the Gammaproteobacteria. Under anaerobic conditions and the presence of nitrate, *P. aeruginosa* can perform (complete) denitrification, a respiratory process of dissimilatory nitrate reduction to nitrogen gas via nitrite (NO_2), nitric oxide (NO) and nitrous oxide (N_2O). This study focuses on understanding the influence of environmental conditions on bacterial denitrification performance, using mathematical modeling tools. To our knowledge, this is the *first* mathematical model of denitrification for *P. aeruginosa*. Analysis of the long-term behavior of the network under changing concentration levels of oxygen (O_2), nitrate (NO_3), and phosphate (PO_4) suggests that PO_4 concentration strongly affects bacterial denitrification performance. The model provides predictions on denitrification activity of *P. aeruginosa* under various environmental conditions, and these predictions are either experimentally validated or supported by pertinent biological literature. One motivation for this study is to capture the effect of PO_4 on a denitrification network of *P. aeruginosa* in order to shed light on mechanisms for greenhouse gas N_2O accumulation during seasonal oxygen depletion in aquatic environments such as Lake Erie (Laurentian Great Lakes, USA).

This chapter is based on the published paper: S. Arat, G. Bullerjahn, and R. Laubenbacher, "A network biology approach to denitrification in *Pseudomonas aeruginosa*", *PLoS One*, vol.10, no. 2, p. e0118235, 2015. My contribution to this paper were (1) playing a major role on the design of the study; (2) constructing and modeling the network; (3) programming for steady state analysis and interpreting the results; (4) writing the manuscript.

4.1 Introduction

Denitrification is a facultative anaerobic process in which nitrate (NO_3) is utilized as an alternative terminal electron receptor and dissimilatory nitrate is reduced to nitrogen gas via nitrogen oxides [91, 92, 93].

$$NO_3 \rightarrow NO_2 \rightarrow NO \rightarrow N_2O \rightarrow N_2$$

Since denitrification is one of the few pathways for producing atmospheric N_2, it is a major component of the nitrogen cycle [94]. Denitrification occurs in several habitats such as soils, lakes, rivers and oceans [95]. Nitrogen fluxes from marine systems to the atmosphere are between 25×10^9 and 179×10^9 kilograms per year via microbial denitrification [96]. *Pseudomonas aeruginosa*, a facultative ubiquitous, and metabolically flexible member of the Gammaproteobacteria, can perform (complete) denitrification under anaerobic conditions and the presence of nitrate. Complete denitrification consists of four sequential steps to reduce nitrate (NO_3) to dinitrogen (N_2) via nitrite (NO_2), nitric oxide (NO), and nitrous oxide (N_2O), and each step of the pathway is catalyzed by (denitrification) enzymes such as nitrate reductase (nar), nitrite reductase (nir), nitric oxide reductase (nor), and nitrous oxide reductase (nos). The identification and transcriptional control of denitrification genes encoding *nar*, *nir*, *nor* and *nos* has been largely established. Transcription is dependent on a hierarchy of the FNR-like Crp family transcription factors Anr and Dnr, the two-component system NarXL, and the CbbQ family protein NirQ [97, 98], summarized in [94], allowing for experimental validation of N_2O yield as environmental parameters change.

We have built a combined gene regulatory and metabolic network for the denitrification pathway in *Pseudomonas aeruginosa* PAO1, a well-studied denitrifier strain (Figure 5.2). With this study, we hope to shed light on the environmental factors contributing to greenhouse gas N_2O accumulation, of particular interest in Lake Erie (Laurentian Great Lakes, USA). Environments such as Lake Erie experience seasonal periods of hypoxic conditions favorable for denitrification, and the endemic microbial community regulates expression of alternative respiratory pathways to adapt to low oxygen (O_2) tension.

We are interested in using the model to investigate the effect of PO_4 on the denitrification performance of *P. aeruginosa* under anaerobic conditions with high NO_3. Although there are several studies on regulation of denitrification by kinetic mathematical modeling approaches (e.g. [99, 100, 101, 102]), these attempts are not enough to cover the phenomenon at different levels [92].

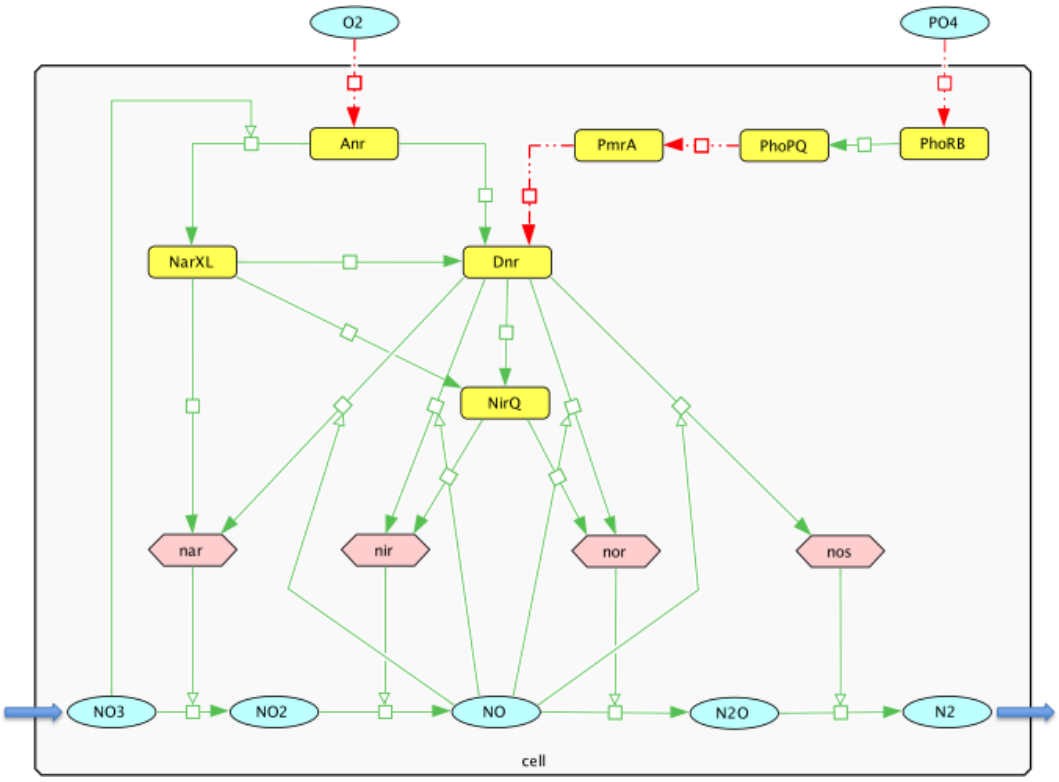

Figure 4.1: Denitrification regulatory network of *Pseudomonas aeruginosa*.

One of the challenges in building kinetic mathematical models of networks, such as systems of differential equations, is that many of the needed parameters are either not known or unmeasurable. Furthermore, for large networks, kinetic models are difficult to analyze mathematically. Therefore, we take a qualitative approach to model denitrification distinct from the quantitative denitrification models attempted previously. We use a discrete model framework that provides coarse-grained information about the temporal biochemical output of the network in response to environmental conditions. This framework captures attractors (and their biological correspondence, phenotypes) yet it does not render any measurements of time or concentration. In particular, we prefer a time-discrete and multi-state deterministic framework, polynomial dynamical systems (PDS) [14], to model our denitrification network in *Pseudomonas aeruginosa*.

4.2 Results and Discussion

The denitrification network consists of molecules, proteins and genes all of which can play an important role in the denitrification process in *Pseudomonas aeruginosa*. Figure 5.2 illustrates a static representation of the variables and their regulations. CellDesigner was used for visualization [103]. The blue circular nodes are molecules (O_2, PO_4, NO_3, NO_2, NO, N_2O, N_2), the yellow rectangular nodes are proteins (PhoRB, PhoPQ, PmrA, Anr, Dnr, NarXL, NirQ) and the pink hexagonal nodes are genes (*nar*, *nir*, *nor*, *nos*) in the network. The large gray rectangle represents the bacterial cell. The regulatory edges between the nodes are either upregulation/activation (green solid arrows) or downregulation/inhibition (red dashed arrows). The pathway begins with the phosphate-sensing two component regulatory system PhoRB [104]. PhoRB, the main PO_4 sensor activating the pho regulon, has been recently shown to be a regulator of PhoPQ transcription in the gammaproteobacterium *Escherichia coli* [105]. In light of the fact that *Pseudomonas aeruginosa* possesses a similar regulatory system to PhoRB in E. coli [106], it is appropriate to label the PO_4-sensing regulatory protein as PhoRB in the denitrification network. In this case, the red dashed arrow from PO_4 to PhoRB means that the availability of phosphate, PO_4, reduces PhoRB function, and the green arrow from PhoRB to PhoPQ means that PhoPQ is activated by PhoRB. Thus, the availability of PO_4 downregulates PhoPQ via PhoRB. The green solid arrow from Anr (anaerobic regulation of arginine deiminase and nitrate reduction) to NarXL and the green solid arrow from NO_3 to the arrow between Anr and NarXL indicate that Anr activates NarXL in the presence of NO_3. In the same setting, PhoPQ inhibits the expression of PmrA [107]. Low oxygen (O_2) tension, which is the major initial signal to turn on the denitrification pathway, can be sensed by Anr [91]. Under anaerobic conditions, Anr primarily promotes Dnr (dissimilatory nitrate respiration regulator) transcription [94]. The effect of Anr on Dnr can be amplified by NarXL [98]. The mechanism of inhibitory effect of PmrA on Dnr [107] is not known, so we assumed that the effect of Anr on Dnr can be reduced by PmrA. The regulatory protein NirQ, which can be activated by NarXL or Dnr, regulates *nir* and *nor* coordinately to keep the level of NO low because of toxicity of NO [94]. A NO_3-responding regulatory protein, NarXL, directly activates *nar*, and indirectly activates *nir* and *nor* via NirQ [94]. The main regulator of the system, Dnr, controls the expression of all denitrification genes (*nar*, *nir*, *nor*, *nos*) in the presence of NO [108]. Of particular note is the influence of the two-component system PhoPQ on PmrA expression and, subsequently, Dnr expression [107], suggesting that phosphorus (P) availability influences denitrification gene expression (see Figure 5.2). This is particularly relevant, since linkages between anaerobic Fe(III) reduction and P release adsorbed to $FeOOH$ in sediments have been recognized for many years

[109, 110], and recently documented in Lake Erie by stable isotope methods [111].

The actual mechanisms of the relationships in the denitrification network (Figure 5.2) may be quite complex and involve several intermediates. Thus, the network does not represent a biochemical reaction network, for instance, but rather captures the regulatory logic driving the network in a similar way that a circuit diagram explains the function of a circuit board. In the network (Figure 5.2), O_2, PO_4 and NO_3 are *external parameters* and the remaining nodes are *variables*. In the discrete setting that is used to model the denitrification network, each node (e.g. an external parameter O_2 or a variable *nos*) can take up to three states (low, medium, high), and time is implicit and progresses in discrete steps. Our interest lies in perturbation of the external parameters and their effect on the long-term behavior of the variables in the system. Table 4.1 indicates the discretization values (low/high) for external parameters and nitrogen oxides. Information in the table was obtained from [112, 113, 114, 115, 116].
Such values incorporate appropriate ranges of long-term nutrient and seasonal oxygen concentrations for Lake Erie [117, 118].

Molecules	Low	High
O_2	$< 0.125mM$	$\geq 0.4mM$
PO_4	$< 0.01mM$	$> 0.9mM$
NO_3	$< 0.1mM$	$> 14.0mM$
NO_2	$< 0.2mM$	$> 2.0mM$
NO	$< 0.1mM$	$> 1.0mM$
N_2O	$< 0.5mM$	$> 5.0mM$

Table 4.1: Discretization of external parameters and nitrogen oxides.

The denitrification network is an open system; it exchanges molecules with the outside environment and responds to external stimuli [119]. The molecule NO_3 enters the bacterium and N_2 exits the system once the system is triggered by low O_2. The model predicts the long-term behavior of the denitrification pathway under various environmental conditions and these predictions are either supported by the literature or validated by experimental results. Figure 4.2 indicates the (predicted) attractors of the system under some possible configuration of the external parameters. The first condition (low O_2, low PO_4 and high NO_3) corresponds to the perfect condition for

denitrification and the second condition (low O_2, high PO_4 and high NO_3) corresponds to the denitrification condition disrupted by PO_4 availability. The remaining conditions can be labeled as aerobic conditions. We did not focus on two conditions that are not included in the analysis. The low NO_3 and low PO_4 condition and the low NO_3 and high PO_4 condition, while possible, are less likely in freshwaters based on a worldwide survey of lakes revealing that N:P stoichiometric ratios average above the ideal Redfield ratio of 16 [120]. Besides, these conditions would be less relevant to current conditions in Lake Erie, for example, as current measurements of nitrate concentrations (averaging 14μM) typically exceed the Km (Michaelis-Menten constant) for nitrate-dependent denitrification in Pseudomonas spp. (for more information, see [121, 122]). However, a high P, high NO_3 condition can arise in lakes affected by agricultural nutrient inputs and deposition of P in sediments.

EXTERNAL PARAMETERS			VARIABLES														
O2	PO4	NO3	PhoRB	PhoPQ	PmrA	Anr	NarXL	Dnr	NirQ	nar	nir	nor	nos	NO2	NO	N2O	N2
low	low	high	high	high	low	high	high	high	high	high	high	high	high	high	high	high	high
low	high	high	low	low	high	high	high	medium	high	high	high	high	medium	high	high	high	medium
high	low	low	high	high	low	low	low	low	low	low	low	low	low	low	low	low	low
high	low	high	high	high	low	low	low	low	low	low	low	low	low	low	low	low	low
high	high	low	low	low	high	low	low	low	low	low	low	low	low	low	low	low	low
high	high	high	low	low	high	low	low	low	low	low	low	low	low	low	low	low	low

Figure 4.2: Steady states of the denitrification network under different environmental conditions.

- **Prediction 1:** If the concentration levels of O_2 and PO_4 are low, and NO_3 is high, then it is a perfect condition for complete denitrification to N_2. The model suggests that all variables in the network except PmrA are expected to be high and the bacterium reduces NO_3 to N_2 via nitrogen oxides. This prediction is *supported* by the following studies [94, 98, 91]. In this condition, Anr senses low O_2 and activates NarXL in the presence of NO_3 [94]. Since the effect of Anr on Dnr is amplified by NarXL but is not reduced by PmrA under low PO_4 conditions, Dnr is highly expressed [98]. The main regulator of the system, Dnr, promotes activation of all denitrification genes (*nar*, *nir*, *nor*, *nos*), so NO_3 is reduced to N_2 via NO_2, NO and N_2O [91].

- **Prediction 2:** If the concentration level of O_2 is low, and PO_4 and NO_3 are high, then the model suggests that all variables except PhoRB-PhoPQ are medium or high. Thus, lower complete denitrification activity to N_2 is expected because the *nar*, *nir* and *nor* levels are high whereas the *nos* level is intermediate. This can cause lower rates of reduction of N_2O to N_2 i.e. higher rates of accumulation of N_2O. These predictions *coincide* with the following studies [98, 107] and experimentation. In this condition, Dnr level is intermediate and

induces the expression of denitrification genes (*nar*, *nir*, *nor*, *nos*) due to the fact that the effect of Anr on Dnr is amplified by NarXL and is reduced by PmrA [98, 107]. Moreover, our experimental results in Table 4.2 show a modest increase in N_2O production with a high PO_4 level. There is about a 2-fold increase in N_2O concentration in comparison of the anaerobic *P. aeruginosa* culture with $1.0mM$ PO_4 to the anaerobic *P. aeruginosa* culture with $7.5mM$ PO_4. Under these conditions, the culture at $1.0mM$ PO_4 is grown under the ideal total N:P ratio of 16 reflecting the 16:1 N:P elemental stoichiometry of aquatic plankton [123]. The cultures grown at elevated PO_4 ($3.0mM$ and $7.5mM$) thus reflect a condition in which PO_4 is available at surplus levels that repress the PhoRB-dependent gene activation. This is an example of how PO_4 can influence the expression of denitrification gene, *nos*, distant from PO_4 acquisition and subsequently greenhouse gas N_2O accumulation.

Culture (mM PO_4)	[N_2O] ppm, 24 h	[N_2O] ppm, 72 h
$1.0mM$	760.3 ∓ 109.34	813.8 ∓ 52.1
$3.0mM$	856.0 ∓ 121.5	872.3 ∓ 63.3
$7.5mM$	1484.0 ∓ 146.2	1786 ∓ 98.0

Table 4.2: Nitrous oxide concentration in *P. aeruginosa* cultures grown in glucose minimal medium at varying phosphate concentrations, normalized to 10^8 cells.

- **Prediction 3:** If the concentration level of O_2 is high, then, the model suggests that there is no denitrification activity regardless of the values of the other external parameters (PO_4 or NO_3). This prediction is *supported* by Zumft's extensive review paper, which states that under aerobic conditions, *Pseudomonas aeruginosa* cannot perform denitrification because Anr cannot activate the main regulator of the system, Dnr, in the presence of oxygen [91].

Figure 4.2 indicates the attractors of the system under different environmental condition. These attractors indeed are steady states, each of which corresponds to one environmental condition. This agrees with biology; Palsson highlighted that open systems eventually reach a (homeostatic) steady state and are in balance with their environment until the environmental conditions are perturbed [119]. Phenotypes, biological interpretations of the long-term behavior (steady states), of the system under various environmental conditions can be found in Table 4.3. Based on the steady state analysis above, the Pseudomonas network model predicts that elevated PO_4, hypothesized to increase under hypoxia, acts to modulate the transcriptional network to limit *nos* gene expression.

Thus, the physiological output under this condition will be an increased yield of N_2O relative to N_2. Given the prediction that increased PO_4 will influence the N_2O yield, our experimental results thus far indicate that PO_4 availability modestly, but significantly increases N_2O yield in this model species (ANOVA $p = 0.012$; Table 4.2). While other studies have suggested linkages between N_2O accumulation and factors such as *nosZ* vs. *nirS/K* abundance [124, 125], *nirS* (heme dependent nitrite reductase) genetic diversity [126], or soil pH [127], the data presented here are the first to suggest a role for PO_4 in regulating the denitrification pathway. Given the elevated PO_4 release from $FeOOH$ complexes following sedimentary anaerobic Fe(III) reduction [109, 110], hypoxia may yield a high P, high NO_3 condition that enhances N_2O production.

O_2	PO_4	NO_3	Phenotype
low	low	high	high denitrification performance
low	high	high	low denitrification performance
high	low	low	no denitrification
high	low	high	no denitrification
high	high	low	no denitrification
high	high	high	no denitrification

Table 4.3: Biological interpretation of the steady states (phenotypes) of the system under different environmental conditions.

Current efforts can be expanded to determine how PO_4 affects greenhouse gas N_2O accumulation during denitrification in *P. aeruginosa*. According to the model, the activation of Dnr by Anr or the activation of *nos* in the presence of NO by Dnr can be prevented by high PO_4. These hypotheses will be tested utilizing quantitative reverse transcriptase PCR (qRT-PCR) to determine Dnr, *norB* (nitric oxide reductase large subunit gene) and *nosZ* (encoding nitrous oxide reductase) transcript levels in denitrifying cultures grown in increasing P. Synergistic interactions between individual members of population of *Pseudomonas aeruginosa* may need to be taken into account and incorporated to the model. For instance, Toyofuku and his colleagues stated that denitrification performance of *P. aeruginosa* does not only depend upon activation of denitrification genes (*nar*, *nir*, *nor*, *nos*) but also cell-cell communications under denitrifying conditions [128].

4.3 Materials and Methods

4.3.1 Mathematical Model

Our network consists of two different sub-networks (metabolic and gene regulatory) and consequently different time scales. From a discrete modeling perspective, this issue can be tackled or ignored only if the long-term behavior of the system is of interest. One could address this issue either (1) using a stochastic framework such as Stochastic Discrete Dynamical System (SDDS) [129] if how fast/slow the reactions are in the network are known/inferred out of a time-course experimental data or (2) introducing time delays by an asynchronous update schedule. Due to inadequate information on the reaction rates, we do not focus on a stochastic framework. Even with a fully asynchronous update schedule, the attractors are preserved for each configuration of external parameters; however, this asynchronous update schedule requires more time steps to reach a steady state than a synchronous update schedule does. Since an asynchronous update schedule provides us more on transient behavior of the system and we are interested in long-term behavior of the system, we prefer to use a deterministic framework with a synchronous update schedule, polynomial dynamical systems (PDS), which allows us to model regulatory networks over a finite field [14].

Definition 1. Let $x_1, x_2, \ldots, x_n$ be variables which can take values in finite fields $X_1, X_2, \ldots, X_n$ respectively. Let $\mathbf{X} = X_1 \times \cdots \times X_n$ be the Cartesian product. For each $i = 1, 2, \ldots, n$, we define $f_i : \mathbf{X} \to X_i$ which is an update function that describes the regulation of x_i through interaction with other variables in the system. A *polynomial dynamical systems* is a collection of n update functions

$$f = (f_1, f_2, \ldots, f_n) : \mathbf{X} \to \mathbf{X}$$

In the model, all external parameters (O_2, PO_4, NO_3) and some variables (PhoPQ, PmrA, Anr, NarXL) are Boolean (low or high), and other variables are ternary (low, medium or high). There are 15 variables, each of which is labeled for the mathematical formulation. Table 4.4 indicates the variables, their discretization, update rules and the literature evidence that support these update rules. The update rules with an asterix (*) means this update rule is very close to the biological correspondence but not quite. Inflow substances (i.e. external parameters: O_2, PO_4, NO_3) in this model give inputs to variables and are involved in the update rules. They do not have update rules because not only they do not have regulators but also we are interested in analyzing the long-term

behavior of the model under different configurations of them. The model has only one outflow substance, N_2, whose regulation depends upon the greenhouse gas N_2O and its reductase, *nos*.

Index	Variable	Discretization	Update Rule	Evidence
1	PhoRB	Boolean	$NOT(PO_4)$	[104]
2	PhoPQ	Boolean	$PhoRB$	[130, 131, 105]
3	PmrA	Boolean	$NOT(PhoPQ)$	[107]
4	Anr	Boolean	$NOT(O_2)$	[91]
5	NarXL	Boolean	$MIN(Anr, NO_3)$	[94]
6	Dnr	Ternary	$^*MIN(Anr, MAX(NarXL, NOT(PmrA)))$	[107, 98]
7	NirQ	Ternary	$^*MAX(NarXL, Dnr)$	[94]
8	*nar*	Ternary	$^*MIN(NarXL, MIN(Dnr, NO))$	[94, 108]
9	*nir*	Ternary	$MAX(NirQ, MIN(Dnr, NO))$	[94, 108]
10	*nor*	Ternary	$MAX(NirQ, MIN(Dnr, NO))$	[94, 108]
11	*nos*	Ternary	$MIN(Dnr, NO)$	[108]
12	NO_2	Ternary	$^*MIN(NO_3, nar)$	[91]
13	NO	Ternary	$MIN(NO_2, nir)$	[91]
14	N_2O	Ternary	$MIN(NO, nor)$	[91]
15	N_2	Ternary	$MIN(N_2O, nos)$	[91]

Table 4.4: Summary of the model variables, update rules and supportive argument.

Based on the literature, we formulate the regulation of the variables with MIN, MAX and NOT, which correspond to AND, OR and NOT in a Boolean setting. The following are examples for how the update rules are decided:

- An update rule of NarXL can be defined as "MIN (Anr, NO_3)" because NarXL is activated by Anr only in the presence of NO_3, i.e. both Anr and NO_3 need to be high for NarXL regulation.
- An update rule of *nir* can be labeled as "MAX (NirQ, MIN(Dnr, NO))" due to the fact that *nir* is activated by NirQ or Dnr in the presence of NO.

- An update rule of PhoRB can be "NOT (PO_4)" since PO_4 downregulates PhoRB, i.e. one is low when another is high.

From the update rules in Table 4.4, for each network variable, we constructed a corresponding transition table, which describes how a specific variable responds to different configurations of their regulators. Although the regulations for most variables can be formulated by MIN, MAX and/or NOT, the regulations of a few variables are very close to some formulation but not quite. For the sake of consistency with biology, we decided to slightly modify the transition table of Dnr, NirQ, *nar* and NO_2, whose update rules are marked with an asterix (*) in Table 4.4. The transition tables of these variables can be found in Appendix A.

Besides, if the variable takes three states (low, medium, high), the current state of the variable is included its own transition table. This does not mean autoregulation/self-regulation; but it is to prevent the variable from jumps between the low (0) state and the high (2) state at the next time step. In other words, including the current state of a ternary variable in its transition table provides a smooth transition among its own states. On the other hand, such jumps cannot occur in a Boolean variable.

After constructing a transition table for each variable x_i, an update function can be obtained by interpolating its transition table using the polynomial form:

$$f_i(x) = \sum_{(c_{i_1},\ldots,c_{i_r})\in F_p^r} f_i(c_{i_1},\ldots,c_{i_r}) \prod_{j\in\{i_1,\ldots,i_r\}} (1-(x_j-c_j)^{p-1}) \pmod p \tag{4.1}$$

where $\mathbf{x} = (x_{i_1},\ldots,x_{i_r})$ is a vector; $c_{i_1},\ldots,c_{i_r}$ are the values of the variables $x_{i_1},\ldots,x_{i_r}$, which affect the update of x_i in the transition table of x_i; $f_i(c_{i_1},\ldots,c_{i_r})$ is the value in the last column of the transition table of x_i; p is the maximum (prime) number of the different discrete values that all variables can take on [132]. In our model, all computations were done in modulo 3.

Here are the update polynomials of all variables in the network:

$\mathbf{f_1} = -PO_4^2 + 1$
$\mathbf{f_2} = -x_1^2 - x_1$
$\mathbf{f_3} = -x_2^2 + 1$
$\mathbf{f_4} = -O_2^2 + 1$
$\mathbf{f_5} = NO_3^2 * x_4^2 + NO_3^2 * x_4 + NO_3 * x_4^2 + NO_3 * x_4$
$\mathbf{f_6} = -x_3^2 * x_4^2 * x_5^2 * x_6 + x_3^2 * x_4^2 * x_5 * x_6^2 - x_3^2 * x_4 * x_5^2 * x_6^2 - x_3 * x_4^2 * x_5^2 * x_6^2 - x_3^2 * x_4^2 * x_5^2 + x_3^2 *$

$x_4^2 * x_5 * x_6 + x_3 * x_4^2 * x_5^2 * x_6 + x_3^2 * x_4^2 * x_6^2 - x_3^2 * x_4 * x_5 * x_6^2 + x_3 * x_4^2 * x_5 * x_6^2 - x_3^2 * x_5^2 * x_6^2 - x_4^2 * x_5^2 * x_6^2 + x_3^2 * x_4^2 * x_5 - x_3^2 * x_4 * x_5^2 + x_3 * x_4^2 * x_5^2 - x_3^2 * x_4^2 * x_6 - x_3 * x_4^2 * x_5 * x_6 + x_3^2 * x_5^2 * x_6 - x_4^2 * x_5^2 * x_6 - x_3 * x_4^2 * x_6^2 + x_3^2 * x_5 * x_6^2 + x_3 * x_5^2 * x_6^2 + x_4 * x_5^2 * x_6^2 - x_3^2 * x_4^2 + x_3^2 * x_4 * x_5 - x_3 * x_4^2 * x_5 + x_3 * x_4 * x_5^2 - x_4^2 * x_5^2 + x_3 * x_4^2 * x_6 - x_3^2 * x_5 * x_6 + x_4^2 * x_5 * x_6 - x_3 * x_5^2 * x_6 - x_3^2 * x_6^2 + x_4^2 * x_6^2 - x_3 * x_5 * x_6^2 + x_4 * x_5 * x_6^2 - x_5^2 * x_6^2 - x_3^2 * x_4 + x_3 * x_4^2 - x_3 * x_4 * x_5 + x_4^2 * x_5 - x_4 * x_5^2 + x_3^2 * x_6 - x_4^2 * x_6 + x_3 * x_5 * x_6 + x_5^2 * x_6 + x_3 * x_6^2 + x_5 * x_6^2 + x_3 * x_4 - x_4^2 + x_4 * x_5 - x_3 * x_6 - x_5 * x_6 - x_6^2 - x_4 + x_6$

$\mathbf{f_7} = -x_5^2 * x_6^2 * x_7^2 + x_5^2 * x_6^2 * x_7 - x_5^2 * x_6 * x_7^2 - x_5 * x_6^2 * x_7^2 - x_5^2 * x_6^2 + x_5^2 * x_7^2 - x_5^2 * x_7 - x_6^2 * x_7 + x_6 * x_7^2 - x_5^2 + x_6^2 - x_7^2 - x_5 + x_7$

$\mathbf{f_8} = x_5^2 * x_6^2 * x_{13}^2 * x_8^2 + x_5^2 * x_6^2 * x_{13}^2 * x_8 + x_5^2 * x_6^2 * x_{13} * x_8^2 + x_5^2 * x_6 * x_{13}^2 * x_8^2 - x_5 * x_6^2 * x_1 3^2 * x_8^2 - x_5^2 * x_6^2 * x_{13}^2 - x_5^2 * x_6^2 * x_{13} * x_8 - x_5^2 * x_6 * x_{13}^2 * x_8 - x_5^2 * x_6 * x_{13} * x_8^2 + x_6^2 * x_{13}^2 * x_8^2 + x_5^2 * x_6^2 * x_{13} + x_5^2 * x_6 * x_{13}^2 + x_5^2 * x_6 * x_{13} * x_8 - x_6^2 * x_{13}^2 * x_8 - x_6^2 * x_{13} * x_8^2 - x_6 * x_{13}^2 * x_8^2 - x_5^2 * x_6 * x_{13} + x_6^2 * x_{13}^2 + x_6^2 * x_{13} * x_8 + x_6 * x_{13}^2 * x_8 + x_5^2 * x_8^2 + x_6 * x_{13} * x_8^2 - x_6^2 * x_{13} - x_6 * x_{13}^2 - x_5^2 * x_8 - x_6 * x_{13} * x_8 - x_5^2 + x_6 * x_{13} - x_8^2 - x_5 + x_8$

$\mathbf{f_9} = x_6^2 * x_{13}^2 * x_7^2 * x_9 - x_6^2 * x_{13}^2 * x_7 * x_9^2 - x_6^2 * x_{13} * x_7^2 * x_9^2 - x_6 * x_{13}^2 * x_7^2 * x_9^2 - x_6^2 * x_{13}^2 * x_7^2 - x_6^2 * x_{13}^2 * x_9^2 + x_6^2 * x_{13} * x_7 * x_9^2 + x_6 * x_{13}^2 * x_7 * x_9^2 + x_6 * x_{13} * x_7^2 * x_9^2 - x_6^2 * x_{13}^2 * x_9 - x_6^2 * x_{13} * x_9^2 - x_6 * x_{13}^2 * x_9^2 - x_6 * x_{13} * x_7 * x_9^2 + x_6^2 * x_{13}^2 + x_6 * x_{13} * x_9^2 - x_7^2 * x_9 + x_7 * x_9^2 + x_7^2 - x_9^2 + x_9$

$\mathbf{f_{10}} = x_6^2 * x_{13}^2 * x_7^2 * x_{10} - x_6^2 * x_{13}^2 * x_7 * x_{10}^2 - x_6^2 * x_{13} * x_7^2 * x_{10}^2 - x_6 * x_{13}^2 * x_7^2 * x_{10}^2 - x_6^2 * x_{13}^2 * x_7^2 - x_6^2 * x_{13}^2 * x_{10}^2 + x_6^2 * x_{13} * x_7 * x_{10}^2 + x_6 * x_{13}^2 * x_7 * x_{10}^2 + x_6 * x_{13} * x_7^2 * x_{10}^2 - x_6^2 * x_{13}^2 * x_{10} - x_6^2 * x_{13} * x_{10}^2 - x_6 * x_{13}^2 * x_{10}^2 - x_6 * x_{13} * x_7 * x_{10}^2 + x_6^2 * x_{13}^2 + x_6 * x_{13} * x_{10}^2 - x_7^2 * x_{10} + x_7 * x_{10}^2 + x_7^2 - x_{10}^2 + x_{10}$

$\mathbf{f_{11}} = -x_6^2 * x_{13}^2 * x_{11}^2 - x_6^2 * x_{13}^2 * x_{11} - x_6^2 * x_{13} * x_{11}^2 - x_6 * x_{13}^2 * x_{11}^2 + x_6^2 * x_{13}^2 + x_6 * x_{13} * x_{11}^2 - x_{11}^2 + x_{11}$

$\mathbf{f_{12}} = NO_3^2 * x_8^2 * x_{12} - NO_3^2 * x_8 * x_{12}^2 - NO_3^2 * x_8^2 + NO_3 * x_8^2 * x_{12} - NO_3^2 * x_{12}^2 - NO_3 * x_8 * x_{12}^2 - NO_3 * x_8^2 + NO_3^2 * x_{12} + NO_3 * x_{12}^2 - NO_3 * x_{12} - x_{12}^2 + x_{12}$

$\mathbf{f_{13}} = -x_9^2 * x_{12}^2 * x_{13}^2 - x_9^2 * x_{12}^2 * x_{13} - x_9^2 * x_{12} * x_{13}^2 - x_9 * x_{12}^2 * x_{13}^2 + x_9^2 * x_{12}^2 + x_9 * x_{12} * x_{13}^2 - x_{13}^2 + x_{13}$

$\mathbf{f_{14}} = -x_{10}^2 * x_{13}^2 * x_{14}^2 - x_{10}^2 * x_{13}^2 * x_{14} - x_{10}^2 * x_{13} * x_{14}^2 - x_{10} * x_{13}^2 * x_{14}^2 + x_{10}^2 * x_{13}^2 + x_{10} * x_{13} * x_{14}^2 - x_{14}^2 + x_{14}$

$\mathbf{f_{15}} = -x_{11}^2 * x_{14}^2 * x_{15}^2 - x_{11}^2 * x_{14}^2 * x_{15} - x_{11}^2 * x_{14} * x_{15}^2 - x_{11} * x_{14}^2 * x_{15}^2 + x_{11}^2 * x_{14}^2 + x_{11} * x_{14} * x_{15}^2 - x_{15}^2 + x_{15}$

After having all update functions, we computed the basin of attraction of the whole system under the environmental conditions of interest (see Figure 4.2). For model construction and steady state analysis, we used customized Ruby and Perl scripts, which are a part of the source code of Analysis of Dynamic Algebraic Models (ADAM), a free of charge web-tool to analyze the dynamics of discrete biological systems [133].

4.3.2 Experimental Methods

Pseudomonas aeruginosa PAO1 cultures were grown in stoppered $20mL$ serum vials containing glucose minimal medium [134] supplemented with $110mM$ glucose and $16mM$ nitrate (NO_3). Phosphate (PO_4) concentration varied from $1.0mM$ to $7.5mM$, and triplicate culture vials were sampled for headspace gases at $24h$ and $72h$ post-inoculation. Gases were dispensed into evacuated exetainers and assayed for nitrous oxide by gas chromatography. Gas production was normalized to cell counts obtained by flow cytometry of culture fluids.

4.4 Conclusion

In an aquatic system, oxygen dissolves in water to be available to living aerobic organisms. Hypoxia is the phenomenon of dissolved oxygen below $4mgO_2$ per liter. Common reasons for hypoxia include aerobic respiration of decaying algal biomass from bloom events. Such blooms are fueled by increased availability of N and P due to anthropogenic inputs such as agricultural runoff and industrial pollutants [135]. The linkage between high nutrient (N,P) loads and N losses (N_2 and N_2O) through dissimilatory anaerobic processes was described recently [136]. Hypoxic (low-oxygen) areas, so-called dead zones, often occur in several large bodies of water affected by human activity, including Lake Erie, which is of particular interest. Establishing a better understanding of the nutrient cycling of Lake Erie has quite wide ranging socioeconomic impacts on its recreational area and economy, primarily fisheries. Through denitrification, dead zones lead to microbial production of the greenhouse gas nitrous oxide (N_2O), which plays a crucial role in ozone layer depletion and climate change. Simulating the microbial production of greenhouse gases in anaerobic aquatic systems such as Lake Erie allows a deeper understanding of the contributing environmental effects that will inform studies on, and remediation strategies for, other hypoxic sites worldwide. During hypoxia, the denitrification rate in Lake Erie is about $150\mu\text{mol}N_2m^{-2}h^{-1}$ [137]. In addition to oxygen, the intersections of the nitrogen cycle with other geochemical cycles may be important factors influencing denitrification and nitrogen (N) sinks in aquatic systems. In particular, the increased availability of phosphorus (P) has been shown to dictate the rate of nitrogen removal in aquatic systems [136]. Indeed, the transcriptional regulatory network developed for *P. aeruginosa* indicates that bioavailable phosphate (PO_4) is an environmental factor that should be considered.

The bacterium *Pseudomonas aeruginosa* is an example of an abundant microbe in aquatic systems

[138], and analysis of Lake Erie metagenomic data sets reveals abundant pseudomonads capable of denitrification (Unpublished data, DOE-JGI). This study describes a computational model of a denitrification network of this bacterium to capture the effect of PO_4 on its denitrification performance in order to shed light on greenhouse gas N_2O accumulation during oxygen depletion. To our knowledge, this is the *first* mathematical model of denitrification for this bacterium. Transcription is dependent on a hierarchy of the FNR-like Crp family transcription factors Anr and Dnr, the two-component system NarXL, and the CbbQ family protein NirQ [97, 139, 98], allowing for experimental measurement of N_2O as external (environmental) parameters change. The model was constructed based on the pertinent biological literature. Model predictions either agree with current published results or are validated by experimentation. The new biology that our model discovers is that PO_4 availability strongly affects the denitrification activity of *P. aeruginosa* under anaerobic conditions and the presence of nitrate; high PO_4 can cause less N_2O reduction to N_2 during denitrification. The data presented here are the first to suggest a role for PO_4 in regulating the denitrification pathway in *Pseudomonas aeruginosa*.

The model described here works well for cultured *Pseudomonas*, and the next step is to test natural complex microbial communities from different denitrification sites. The effects of PO_4 on N_2O production will be tested in mesocosms of hypoxic Lake Erie water samples to see if the model described here predicts the community as a whole. By testing the model on environmental samples in mesocosms from Lake Erie and elsewhere, the study can likely be applied broadly to other marine dead zones such as those that routinely occur in the Gulf of Mexico.

4.5 Acknowlegments

We thank Mr. Michael Schlais (Bowling Green State University, Bowling Green, OH) for assistance with the Pseudomonas culture experiment and Dr. Richard A. Bourbonniere (Environment Canada, Burlington, ON) for the gas chromatography measurements.

Chapter 5

Intracellular Iron Model: from Normal Breast Cells to Cancer Cells

Iron is required for many processes such as energy production and proliferation. Iron can be toxic due to its ability to cause high oxidative stress levels and consequently DNA damage. In order to prevent damage, all organisms that require iron have developed mechanisms to tightly control iron levels. Dysregulation of iron metabolism is detrimental and can lead to a wide range of diseases including cancer. In this study, we build a predictive mathematical model of an iron metabolism linked to other pathways such as iron utilization, oxidative stress response and oncogenic pathways in a normal breast cell. We used a time- and state-discrete mathematical framework, polynomial dynamical systems (PDS), to model regulatory networks over a finite field. Simulations of our model provide two predictions, which are either validated by experimentation (new data) or published data which was not used for model construction. To date, our intracellular model is the only expanded iron metabolism model that can capture a breast cancer phenotype by overexpression and knockout simulations. One motivation for this study is to understand how normal cells become malignant cells and identify key players in the system for therapeutic targets.

This chapter is based on a paper draft: J. Chifman*, S. Arat*, Z. Deng, C. Lopez, S. Torti, and R. Laubenbacher, "Intracellular Iron Model: from Normal Breast Cells to Cancer Cells", *PLoS Computational Biology* (in preparation). Julia Chifman and I contributed equally to this paper. We (1) designed the study; (2) constructed and modeled the network; (3) wrote the manuscript. I also wrote a Perl script to compute the basin of attraction of the system.

5.1 Introduction

Every aerobic organism requires iron for many processes such as energy production, DNA synthesis, oxygen transport and cellular respiration. At the same time, this essential element can be toxic and damaging due to its ability to exist in various oxidation states. In order to prevent damage, all organisms that require iron have developed a complicated machinery to tightly control iron at both the systemic and the cellular levels. Inappropriately low levels of iron or excess iron are detrimental and lead to a wide range of diseases. For example, chronic iron overload increases the risk of neurodegenerative diseases (Alzheimer's, Parkinson's, Huntington's), heart failure, liver cirrhosis and cancer [140]. Dysregulation of iron metabolism in cancer is well documented and it has been suggested that there is an interdependence between excess iron and increased cancer incidence [141]. Without a doubt, cancer is more than an iron disorder, nevertheless, many findings point out to a direct relationship between iron and cancer, and thus, our goal is to clarify the precise nature of this relationship.

Breast cancer is the leading cause of cancer death among women worldwide and second leading cause of cancer death among women in the US [142]. Recently we have made first attempts at understanding the dynamics of the intracellular iron metabolism by constructing a continuous model of the iron core control system in the form found in normal breast epithelial cells [143]. This choice of cell type was motivated by our interest in the role of intracellular iron homeostasis in the pathogenesis of breast cancer. Based on the hypothesis that major pathways activated in cancer do disrupt iron pathway, we have set out on the quest of connecting the iron core network to several known molecules whose expression levels are altered in cancer. In this study, we present and analyze a mathematical model of an expanded intracellular iron homeostasis pathway linked to iron utilization, oxidative stress response, inflammatory and oncogenic pathways. We used a time- and state-discrete mathematical framework, polynomial dynamical systems (PDS), which allows us to model regulatory networks over a finite field. Model simulations allow us to identify key players in the system that lead to different phenotypes without having to perform lengthy laboratory experiments. Figure 5.1 summarizes the steps taken from network reconstruction to hypothesis validation through mathematical model and simulations. To our knowledge, our intracellular model is the only expanded iron metabolism model that can capture a breast cancer phenotype by overexpression and knockout simulations.

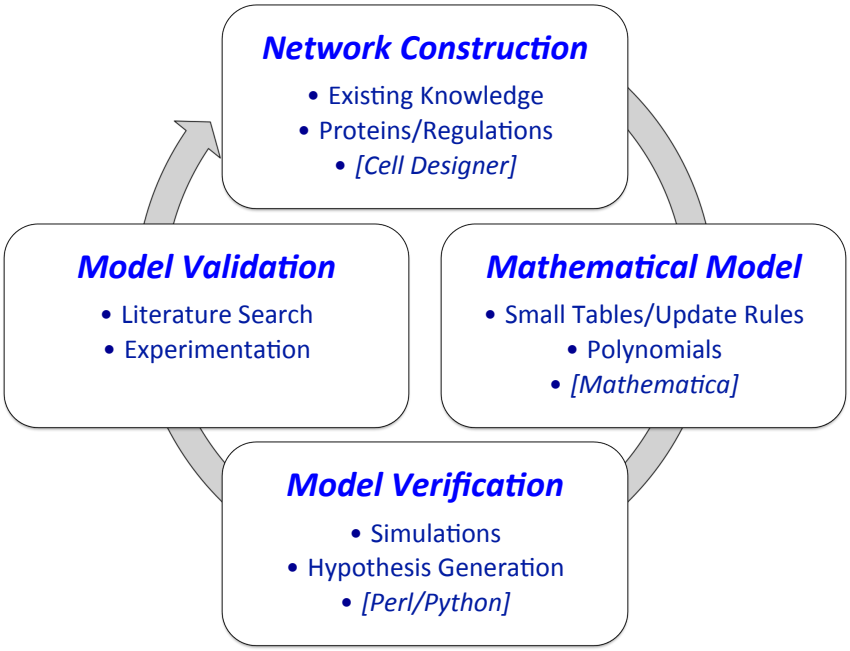

Figure 5.1: Comprehensive pipeline for systems biology approach taken for this study.

5.2 Results and Discussion

5.2.1 An expanded iron core network

Figure 5.2 is a static representation of the expanded iron homeostasis pathway (LIP, TfR1, Fpn, Ft, IRP1, IRP2, Hep) with iron utilization (Mfrn, LIPmt, Ftmt, ALAS1, heme, HO-1), oxidative stress response (ROS, Keap1, Nfr2, Antioxidant enzymes) and oncogenic (EGFR, SOS, GAPs, Ras, ERK, c-Myc) pathways. IL-6 is the only inflammatory response protein in the network. CellDesigner was used for visualization [103]. Each pathway, its function, its components and their interactions were discussed below.

Intracellular Iron Metabolism. Free ferrous iron contributes to the formation of the hydroxyl radical through the Fenton reaction, thus to reduce toxicity intracellular iron is meticulously maintained. Iron levels are controlled by iron-regulatory proteins (IRPs) that coordinate intracellular iron uptake, utilization, storage and excretion. What follows is a brief description of the iron core control system. For an overview of the intracellular and systemic iron homeostasis the reader is encouraged to consult [140]).

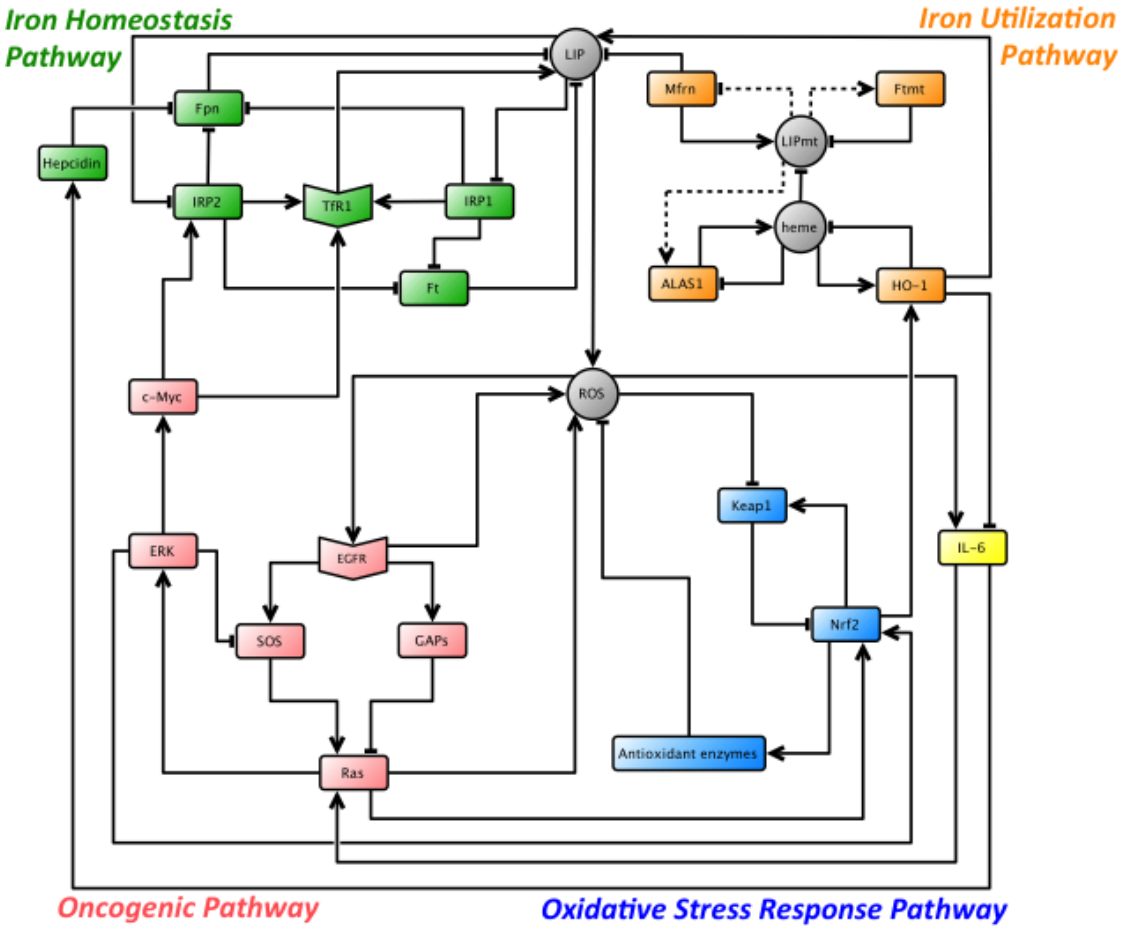

Figure 5.2: Expanded iron homeostasis pathway. Arrows depict upregulation and hammer heads depict downregulation. Dashed connections are assumed to exist.

Ferric iron, Fe^{3+}, circulates in plasma bound to transferrin (Tf), a glycoprotein with two binding sites for ferric iron. Tf retains iron in a soluble form, which limits the formation of toxic radicals, and delivers iron to cells. Cells acquire iron predominantly through transferrin receptor 1 (TfR1), a major iron importer. From the endosomes, iron enters the *labile iron pool* (LIP), a cytosolic pool of weakly bound iron. Ferroportin (Fpn), located on the plasma membrane, is believed to be the only ferrous iron exporter. Excess ferrous iron that was not exported or utilized is oxidized by a cytosolic protein ferritin (Ft) and is sequestered into its ferrihydrite mineral core.

Iron regulatory proteins, IRP1 and IRP2, regulate iron homeostasis post-transcriptionally by binding to iron responsive elements (IREs). In iron-deplete cells, IRPs are active and have high affinity for IREs. This stabilizes TfR1 and inhibits ferroportin and ferritin. In iron-replete cells, IRPs have no affinity for IREs. This leads TfR1 degradation and ferroporting and ferritin expression.

The peptide hormone hepcidin (Hep) regulates systemic iron homeostasis by inhibiting iron release from duodenal enterocytes, macrophages, and hepatocytes. Hepcidin binds to the iron exporter ferroportin and triggers its internalization and degradation in lysosomes. Hepcidin is transcriptionally induced by the inflammatory cytokine interleukin-6 (IL-6) [144]. The induction of hepcidin by IL-6 is thought to be a major contributor to the hypoferremia that frequently accompanies chronic infections, acute inflammation and cancer [30]. Recently, it has been established that breast epithelial cells also express hepcidin and that it plays an important role in peripheral tissue by regulating ferroportin [145].

Iron utilization. The mitochondrion is the major site of iron utilization. Cytosolic iron (LIP) is imported into the mitochondrion by SLC transporter mitoferrin (Mfrn) to be incorporated into protoporphyrin IX (PPIX) to make heme. There are two homologs mitoferrin-1 (SLC25A37), which is expressed at high levels in erythroblasts and at low levels in other tissue, and mitoferrin-2 (SLC25A28), which is expressed ubiquitously [146]. Once iron is transported into the mitochondrion, mitochondrial labile iron pool (LIPmt) is then used in heme synthesis, iron sulfur cluster (ISC) synthesis or enters mitochondrial ferritin (Ftmt). Just like cytosolic ferritin, Ftmt is an iron storage protein. Primary function of Ftmt is not fully understood but evidence indicates that its role is to protect mitochondrion from iron-dependent oxidative damage [147].

It is well established that intracellular heme regulates its own production and degradation through delta aminolevulinate synthase 1 (ALAS1) and heme oxygenase 1 (HO-1), respectively. HO-1 is responsible for maintaining heme homeostasis by initiating the oxidative cleavage of heme to ferrous iron (Fe^{2+}), carbon monoxide (CO), and biliverdin. Moreover, HO-1 inhibits the expression of IL-6 thus also taking on an anti-inflammatory function [148]. Heme synthesis involves several steps that occur in two compartments: (i) mitochondrion with the initial and final steps, and (ii) cytosol with intermediate steps. The ALA synthase reaction is the committed step of heme synthesis. Heme represses the transcription of the gene for delta aminolevulinate synthase and translocates ALAS1. The mitochondrion exports ALAS1 to the cytoplasm, where the next four reactions occur. The final step occurs in mitochondrion where Fe^{2+} is incorporated into PPIX via ferrochelatase, which completes heme synthesis.

In LIPmt-deplete cells, heme synthesis is not completed and so, we assumed a feedback regulation from LIPmt to ALAS1 to cover this. Although our understanding about precise regulation of Ftmt and Mfrn is partial, an experimental evidence suggests that a feedback mechanism must exist, which responds to the levels of LIPmt [149, 147], and some studies imply that there might be

even a cross-talk between cytosolic and mitochondrial iron metabolism [150]. Thus, our model assumed three regulations (dashed arrows in Figure 5.2) through some unknown species/mechanism to permit for mitochondrial iron homeostasis.

Oxidative Stress. Oxidative stress results from an imbalance between reactive oxygen species (ROS) and antioxidants, resulting from either excessive amounts of ROS or a deficiency in antioxidants. ROS, a family of oxygen species with one or more unpaired electrons, are generated during many cell processes and overcome by antioxidants to prevent DNA damage and to support genomic stability [151].

Iron can contribute to the formation of ROS. In aerobic organisms, oxygen (O_2) is mostly bound to hydrogen (H_2) as water. However a small portion of O_2 can be converted to a variety of reactive oxygen species (ROS), including superoxide radical $(O_2\cdot)^-$, hydrogen peroxide H_2O_2 and hydroxyl radical $\cdot$OH [152, 153]. Ferrous iron Fe(II) can interact with O_2 to form $(O_2\cdot)^-$ and H_2O_2, which then leads to formation of highly active, unstable and the most damaging oxidant $\cdot$OH via iron-catalyzed Fenton reaction [151, 154, 155].

$$\text{Fe(II)} + O_2 \rightarrow \text{Fe(III)} + (O_2\cdot)^-$$

$$\text{Fe(II)} + (O_2\cdot)^- + 2H^+ \rightarrow \text{Fe(III)} + H_2O_2$$

$$\text{Fe(II)} + H_2O_2 \rightarrow \cdot OH + OH^- + \text{Fe(III)}$$

Nuclear factor (erythroid-derived 2)-like2 (Nrf2) is an important contributor to reduction of oxidative stress. The main function of Nrf2 is to transcriptionally activate genes containing antioxidant response elements (ARE). Kelch-like ECH-associated protein 1 (Keap1), is the main regulator of Nrf2, and plays a central role in sensing and protecting cells against ROS. Under normal conditions, Nfr2 binds to Keap1, which promotes degradation of Nrf2 [156]. Upon exposure to ROS, Keap1 is inactivated, Nrf2 disassociates from Keap1 and becomes stabilized and heterodimerizes with small masculoaponeurotic fibrosarcoma (Maf) proteins [157]. ARE-containing genes counterbalance the harmful effects of ROS through a variety of mechanisms [158, 159]. For this study, we focused on 4 antioxidant enzymes that contribute to the antioxidant response: superoxide dismutase (SOD), catalase (CAT), glutathione peroxidase (GTPx) and heme oxygenase-1 (HO-1). Of these, GTPx and HO-1 are directly inducible by Nrf2. Recall that HO-1 is a part of the iron utilization pathway and thus this enzyme is modeled in our network as a separate node. On the

other hand SOD, CAT and GTPx, which can eliminate specific reactive oxygen species, are represented as a single node, labeled Antioxidant enzymes. SOD in cytosol and mitochondria catalyzes the dismutation of superoxide radical $(O_2\cdot)^-$ to form hydrogen peroxide H_2O_2 and O_2 [151, 154]. GTPx in cytosol and mitochondria and CAT in tissue peroxisomes reduce H_2O_2 to water and O_2 to control production of DNA-damaging molecule hydroxyl radical ·OH [151, 153]. ·OH, which has a very short half-life $\sim 10^{-9}$ seconds, cannot be scavenged by an enzymatic reaction. Thus, endogenous and dietary ·OH scavengers (e.g. melatonin, vitamin E) are out of scope of this study.

Besides having a destructive role, ROS can also act as signaling molecules to promote cell proliferation, survival, apoptosis, differentiation and migration [155, 160]. It has been suggested that ROS induces inflammatory response protein IL-6 production in this fashion [161, 162]. ROS also involves regulation of the epidermal growth factor receptor in the oncogenic pathway as a signaling molecule, which is discussed further in the following part.

Oncogenic Pathway. The epidermal growth factor receptor (EGFR) regulates cell growth, differentiation, and motility through interaction with its ligand, epidermal growth factor (EGF). EGFR activation stimulates transient activation of Ras-GTP and this eventually leads to activation of extracellular-signal regulated kinases (ERKs) [163], which in turn results in phosphorylation and stabilization of c-Myc [164]. It has been established that c-Myc stimulates the expression of IRP2 [165] and activates TfR1 [166], which provides a link between oncogenic and iron homeostasis pathways.

Ras is a small guanosine triphosphatase (GTPase) and its activity is controlled by a regulated GDP/GTP cycle. The duration of Ras activity (time being in the GTP-bound form) and the level of activation (GTP-bound form / total Ras) are controlled by (a) the guanine nucleotide exchange factors (GEFs) that promote exchange of GDP for GTP, and (b) GTPase-activating proteins (GAPs) that stimulate the intrinsic GTPase activity of Ras to promote formation of the inactive, GDP-bound form of Ras. The activator of Ras is a GEF protein son of sevenless (SOS), which facilitates the switch from Ras-GDP to Ras-GTP. Both SOS and Ras-GAP are recruited to the phosphorylated EGFR [163, 167]. ERK phosphorylates SOS resulting in its dissociation from growth factor receptor-bound protein 2 (Grb2) providing a negative feedback and thus limiting activation of Ras [168, 167]. Ras also is activated by IL-6 [169, 170].

The oncogenic pathway and reactive oxygen species (ROS) have a close and intricate relationship. Our model is not refined enough to capture all the complexities of this interaction, but we do

include many known established connections. In particular, it has been shown that activated Ras induces the production of ROS, which is required for oncogene-mediated cellular transformation and Ras dependent proliferation [171, 172, 173, 174]. Moreover, there is a direct induction of EGFR by endogenous H_2O_2 and a localized generation of H_2O_2 by EGFR through a NADPH oxidase (Nox)-mediated process [160, 175]. Extracellular-signal regulated kinases (ERKs) and Ras are also involved in oxidative pathway by activating Nrf2 [176, 177].

5.2.2 Logical Model

Logical models have been used for modeling biological systems in the simple and intuitive manner by viewing the network in terms of a collection of logical rules with discrete time steps [178]. Each species in our network is represented by ternary logic, which is an extension of Boolean logic. For most of the components in our network, 1 denotes *normal*, 0 denotes *lower than normal* and 2 denotes *higher than normal* concentration levels. In some components (e.g. IRP1, IRP2 and Ras), 0 or 2 mean *inactive* or *active* form respectively. Regardless of the labeling, the interactions in the network can be translated into update rules analogous to Boolean functions. Update rules are then used for simulations and hypothesis generation. Table 5.1 is a summary of all species in our network, their update rules and relevant citations. We used a time- and state-discrete mathematical framework, polynomial dynamical systems (PDS), to model our network over a finite field, $\mathbb{F}_3$. A detailed description of the construction of the model and definitions of the logic gates (Max, Min and Not) can be found in 5.3.1. The entire set of update rules in their polynomial form used for simulations is provided in Appendix B.2.

The (normal cell) model is constructed in a way that it has one steady state, which indicates all species to be in their normal levels no matter what the initial state is. We detected different attractors depending on the knockout and overexpression simulations. For knockout simulation, the update polynomial of the knockout component is set to 0. Regardless of its regulators, it stays in lower than its normal concentration levels or inactive form. For overexpression simulations, the update polynomial of the knockout component is set to 2. Regardless of its regulators, it stays in higher than its normal concentration levels or active form. The main purpose is to determine the effect of perturbations (knockout and overexpression) on the long-term behavior of the system. We simulated a normal cell, a mitoferrin knockout (Mfrn k/o) cell, an IRP2 overexpressed (IRP2 o/e) cell and spontaneous k/o and o/e simulations to capture a breast cancer phenotype.

5.2.3 Model Validation

After obtaining all 24 polynomials, we ran a customized Perl and Python script to simulate the whole state space and the basin of attraction of the system. Table 5.2 shows the variables and their long-term behavior under different simulation types. Recall that the (normal cell) model has one steady state, which indicates all species to be 1 (in their normal levels) no matter what the initial configuration is. Mitoferrin knockout (Mfrn k/o) cell simulation is to test the model whether it reproduces the known fact on the iron utilization pathway. IRP2 overexpressed (o/e) cell simulation for hypothesis generation, which are validated by experimentation. Breast cancer cell simulation results are discussed in 5.2.4.

Mitoferrin knockout simulation matches with the current literature. Mitoferrin (Mfrn) transports cytosolic iron (LIP) into mitochondria and mitochondria LIP (LIPmt) is utilized for heme production [146, 147]. When Mfrn is knockout, LIP is not transported into mitochondria and heme synthesis is not initiated. The Mtfrn k/o cell model can generate this state. Based on Mfrn k/o simulation, the system has only one steady state, in which all the components are in their normal levels except the iron utilization proteins (see the third column in Table 5.2). Low levels of LIPmt, Ftmt, ALAS1 and heme are a consequence of Mfrn knockout. While the effect of Mfrn k/o on HO-1 is not HO-1 is not much affected because it can possibly be regulated by Nrf2, an oxidative stress response protein.

IRP2 overexpression only alters the core iron system. The IRP2 overexpressed (o/e) cell model predicts that IRP2 o/e only affects the iron homeostasis pathway but not others (see the forth column in Table 5.2). As IRP2 is constantly active, TfR1 levels are elevated whereas ferritin and ferroportin levels are declined. No significant effect is expected on other components in the network.

To test this prediction experimentally, at least one protein in each pathway was decided to be measured in IRP2 o/e cells. Figure 5.3 indicates the experimental data, in which iron-related proteins TfR1, ferritin and HO-1, the oxidative stress sensor protein Keap1, the inflammatory response protein IL-6, and oncogenic proteins EGFR and c-Myc were measured in normal and IRP2 o/e cells. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) is a loading control, which indicates equal loading in each lane. IRP2 o/e increases TfR1 recruitment as it moderately decreases ferritin production. However, there is not significant change between the levels of other proteins in

normal cells and in IRP2 o/e cells. This validates our prediction: IRP2 o/e only affects the iron homeostasis pathway in our network.

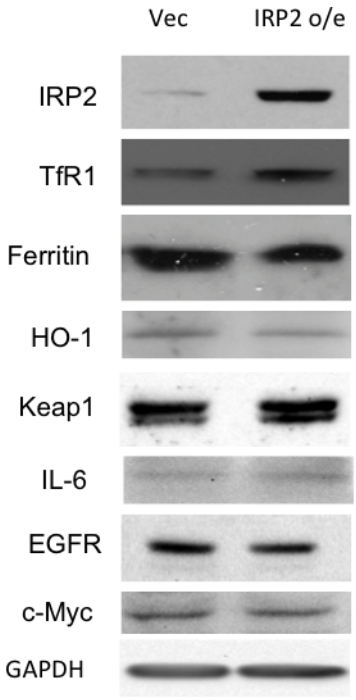

Figure 5.3: Effects of IRP2 overexpression on the network. GAPDH is a loading control.

5.2.4 Differential regulation of pathways in cancer

One main goal of this study is to capture a breast cancer-like behavior from our normal cell model to investigate how normal cells become malignant. First of all, we determined a breast cancer phenotype based on pertinent literature. We are aware of the fact that there are many types of breast cancers and breast cancer cell lines. Therefore, we have a unified breast cancer phenotype in which the components are known as higher or lower in general when compared to the components in normal breast cells. Iron homeostasis, oxidative stress response and oncogenic pathways and their differential regulation in breast cancer is well-known and supported by several studies. However, differential regulation of iron utilization pathway in (breast) cancer can not determined due to lack of information and studies in the current literature.

In many breast cancers, free iron (LIP), TfR1, IRP2 and hepcidin levels are high whereas ferroportin, ferritin and IRP1 levels are low [145, 181]. It has been known that breast cancer cells are

frequently under persistent oxidative stress [182]. *In vitro*, human tumor cell lines have higher levels of ROS ($(O_2\cdot)^-$, H_2O_2, $\cdot OH$) than their non-tumorigenic versions [151]. Several mechanisms have been reported for the increased activity of Nrf2 in breast cancer, one of which is loss-of-mutations in Keap1 and gain-of-function mutations in Nrf2 [158]. The levels of antioxidant enzymes are detected in several breast cancer studies. Increased expression of SOD and HO-1 are clinically observed but CAT and GTPx levels are declined in breast cancer tissues [183, 184, 179]. In addition, Sen et al. highlights that catalese protein levels are high, yet catalase bioactivity is too low to overcome high ROS levels in breast cancer cells [185]. This suggests that the inhibitory effect of antioxidants on ROS is disrupted in cancer. High serum levels of inflammatory cytokine IL-6 are detected in breast cancer patients [186]. Determining the levels of oncogenic pathway proteins is a bit tricky. For instance, EGFR levels are elevated in 50% of the breast cancers and Ras is mutated (hyper-active) only in 5% of the breast cancers [168, 187, 188]. c-Myc levels are higher in 50%-100% of breast cancer cases [189, 190].

The breast cancer phenotype for each pathway can be found at the last column in Table 5.2. Our breast cancer simulation, in which Ras was overexpressed and the inhibitory effect of Antioxidant enzymes on ROS was removed, matches with the existing literature.

5.3 Materials and Methods

5.3.1 Mathematical Model

Our network has 24 nodes, all of which can take either 1 (normal), 0 (lower than normal/inactive) or 2 (higher than normal/active). Update rules were determined based on existing literature (Table 5.1). Here, we provide the steps taken to build a (discrete) mathematical model using the update rules.

If species X is inducing species Y ($X \rightarrow Y$) or species X is inhibiting species Y ($X \dashv Y$) then we represent these relations via a transition table as depicted in Table 5.3.1.

$X \to Y$	
0	0
1	1
2	2

$X \dashv Y$	
0	2
1	1
2	0

Table 5.3: Transition tables for activation and inhibition.

Note that inhibition in Table 5.3.1 is just a logic NOT gate (denoted by $\overline{X_i}$ where $X_i \in \{0, 1, 2\}$). The other two fundamental gates, OR and AND, for two species X and Y regulating species Z ($X \to Z \leftarrow Y$), can be defined as $\max\{X_i, Y_j\}$ and $\min\{X_i, Y_j\}$ respectively, for $X_i, Y_j \in \{0, 1, 2\}$. To differentiate from the Boolean OR and AND gates, we will denote these gates by Max and Min, respectively.

We can express the above relations as polynomials over a finite field on three elements, $\mathbb{F}_3$. Notice though, that different polynomials can give rise to the same function, e.g. $x^3y + 1$ and $xy + 1$ have the same output for any $x, y \in \mathbb{F}_3$. Thus, polynomials below are not unique.

$$\begin{aligned} \overline{x} &= 2 + 2x \\ \text{Max}(x, y) &= x^2y^2 + x^2y + xy^2 + 2xy + x + y \\ \text{Min}(x, y) &= 2x^2y^2 + 2x^2y + 2xy^2 + xy \end{aligned} \tag{5.1}$$

Various adjustments to the strength of a particular regulation can be made by altering entries in the Table 5.3.1. For example, it has been suggested that IRP1, when active, contributes less to the regulation of ferritin (Ft) than IRP2 (see Table 5.3.1). These tables mean that when IRP2 = 2 (active) it will inhibit Ft, whereas when IRP1 = 2 (active) it will have a lesser affect on Ft.

IRP1 $\dashv$ Ft	
0	2
1	1
2	1

IRP2 $\dashv$ Ft	
0	2
1	1
2	0

Table 5.4: Transition tables for IRP1 and IRP2 regulating Ft.

Thus, we can represent regulation of Ft by IRP2 in Table 5.3.1 using Equation 5.1:

$$\overline{IRP2} = 2 + 2 \cdot \text{IRP2}$$

Now for IRP1 regulating Ft according to this new adjustment one can also find a polynomial representing Table 5.3.1 (left table). For convenience, whenever we use an adjusted regulation we will place an asterisk (*) in front of the variable inside the logic gate.

$$*\overline{IRP1} = 2 + 2 \cdot (\text{IRP1})^2$$

We can update the state of each of the species in our network either *synchronously* or *asynchronously*. Synchronous update simply means that all species in the network are updated simultaneously, while asynchronous, as the name suggests, means that not all species are updated at the same time. We use *synchronous* update and hence each state in our network will belong to the basin of attraction of only one attractor. The attractor can be a *point* attractor (steady state) or a *cycle* attractor (limit-cycle). These attractors can be considered as phenotypes in the biological context.

To make sure that we preserve continuity (i.e. each species changes at most one unit up or down), we are going to employ methodology as described in [178]. The logic behind this is to take into account the previous state (e.g. concentration) of the regulated species. The future value of the regulated species under continuity is computed as follows. Let f_i be the update function for x_i. To ensure that each variable changes at most 1 unit, define a function $h(x_i, f_i)$ for the updated value of the variable x_i at the next time step:

$$h(x_i, f_i) = \begin{cases} x_i + 1 & \text{if } f_i > x_i \\ x_i & \text{if } f_i = x_i \\ x_i - 1 & \text{if } f_i < x_i \end{cases} \tag{5.2}$$

We would like to note that this methodology imposes self-regulation, which is not a problem for many species in our network as many of them self-degrade. LIP, Heme and H_2O_2 do not undergo self-regulation and hence we do not apply continuity to these species. In order to compute final

polynomials, we are going to make a use of the following property of finite fields:

Remark 5.3.1. If $h : \mathbb{F}_p^n \to \mathbb{F}_p$ is any function then there is a polynomial $f : \mathbb{F}_p^n \to \mathbb{F}_p$ so that $h(x) = f(x)$ for all $x \in \mathbb{F}_p^n$.

One can find f by using the following formula,

$$f(x) = \sum_{c \in \mathbb{F}_p^n} h(c) \prod_{j}^{n} (1 - (x_j - c_j)^{p-1}), \tag{5.3}$$

where $h(c)$ is the update function as defined by (5.2), c is a vector of input variables, and the right-hand side is computed *modulo p*.

All of these logic gates, transition tables describing different strength of regulation and continuity, are then appropriately translated into final polynomial functions over a finite field with three elements. These polynomial functions then form what we call a *polynomial dynamical systems* (PDS) over a finite field [14]. We fully described a construction of the *update polynomial* for ferritin (Ft) in iron homeostasis pathway in Appendix B.1. In a similar fashion, the update polynomial of all variables were constructed, which can be found in Appendix B.2.

5.3.2 Experimental Methods

MCF10A, non-tumorigenic immortalized human mammary epithelial cells were obtained from the Wake Forest University Comprehensive Cancer Center Tissue Culture Core facility. The cells were maintained in a suggested condition by ATCC.

To overexpress IRP2 in MCF10A cells, the lentiviral vector pSL2-IRP2 [191] were applied. Briefly, MCF10A cells were infected with the concentrated viral particles from pSL2-IRP2 and pLS2 empty vector (as a control). The infection efficiencies for both infections were over 90% based on GFP fluorescence in cells. The cell lysates were harvested for subsequent analysis seven days after infection.

Western blotting was performed as previously described [191]. Antibodies: GAPDH (Fitzgerald), TfR1 and c-Myc (Invitrogen), IRP2 and EGFR (Santa Cruz Biotechnology), Keap1(Cell Signaling Technology), HO-1 and IL-6 (Abcam) , ferritin H ([192]).

5.4 Conclusion

In this study, we took a systems biology approach to investigate how normal cells become cancerous. According to our literature-based predictive mathematical model of an expanded intracellular iron metabolism, the model predictions were validated through experimentation or current literature. The new cancer biology that we discovered is that IRP2 overexpression only alters the iron homeostasis pathway. To our knowledge, there has not been a similar model that can capture a breast cancer phenotype by overexpression and knockout simulations. The breast cancer simulation suggests that the iron utilization and antioxidant mechanism are disrupted in breast cancer. Dysfunctional iron utilization in breast cancer can be further examined as a follow-up study.

5.5 Acknowledgments

We acknowledge support from National Institute of Health NCI-NIH 1R21CA156133-01A1.

Variable	Classification	Update Rule	Evidence
LIP	IH	Min(Max(TfR1, HO-1), Min($\overline{\text{Fpn}}$, $\overline{\text{Ft}}$, $\overline{\text{Mfrn}}$))	[140, 179]
TfR1	IH	Max(IRP1, IRP2, c-Myc)	[140, 166]
Fpn	IH	Min($\overline{\text{IRP1}}$, $\overline{\text{IRP2}}$, $\overline{\text{Hep}}$)	[140, 145]
Ft	IH	Min($\overline{*\text{IRP1}}$, $\overline{\text{IRP2}}$)	[140]
IRP1	IH	$\overline{\text{LIP}}$	[140]
IRP2	IH	Max($\overline{\text{LIP}}$, c-Myc)	[140, 165]
Hep	IH	IL-6	[144]
Mfrn	IU	$\overline{\text{LIPmt}}$	[146]
LIPmt	IU	Min(Mfrn, $\overline{\text{Ftmt}}$, $\overline{\text{heme}}$)	[146, 147]
Ftmt	IU	LIPmt	[147]
ALAS1	IU	Min($\overline{\text{heme}}$, LIPmt)	[149, 147]
heme	IU	Min(ALAS1, $\overline{\text{HO-1}}$)	[149, 147]
HO-1	IU	Max(heme, Nrf2)	[149, 159]
ROS	OSR	Min(Max(LIP, EGFR), $\overline{\text{Antioxidant enzymes}}$)	[152, 154, 175, 158]
Keap1	OSR	Min($\overline{\text{ROS}}$, Nfr2)	[157, 156]
Nrf2	OSR	Max($\overline{\text{Keap1}}$, Ras, ERK)	[157, 176, 156, 180]
Antioxidant enzymes	OSR	Nrf2	[157, 158]
IL-6	IR	Max($\overline{\text{HO-1}}$, ROS)	[148, 161, 162]
EGFR	Onc	ROS	[175]
SOS	Onc	Max(EGFR, $\overline{\text{ERK}}$)	[163, 167, 168]
GAPs	Onc	EGFR	[163, 167]
Ras	Onc	Min(Max(IL-6, SOS), $\overline{\text{GAPs}}$)	[163, 169, 170]
ERK	Onc	Ras	[163]
c-Myc	Onc	ERK	[164]

Table 5.1: Summary of all model variables and their update rules. IH, iron homeostasis; IU, iron utilization; OSR, oxidative stress response; IR, inflammatory response; Onc, Oncogenic.

Variable	Normal	Mfrn k/o	IRP2 o/e	Breast Cancer
LIP	1	1	1	2
TfR1	1	1	2	2
Fpn	1	1	0	0
Ft	1	1	0	0
IRP1	1	1	1	0
IRP2	1	1	2	2
Hep	1	1	1	2
Mfrn	1	0	1	0
LIPmt	1	0	1	0
Ftmt	1	0	1	0
ALAS1	1	0	1	2
heme	1	0	1	0
HO-1	1	1	1	2
ROS	1	1	1	2
Keap1	1	1	1	0
Nrf2	1	1	1	2
AEs	1	1	1	2
IL-6	1	1	1	2
EGFR	1	1	1	2
SOS	1	1	1	2
GAPs	1	1	1	2
Ras	1	1	1	2
ERK	1	1	1	2
c-Myc	1	1	1	2

Table 5.2: Attractors (steady states) under certain perturbations.

Bibliography

[1] A. S. Rodin, G. Gogoshin, and E. Boerwinkle, "Systems biology data analysis methodology in pharmacogenomics," *Pharmacogenomics*, vol. 12, no. 9, pp. 1349–60, 2011.

[2] M. Schneider, *In Silico Systems Biology*, ser. Methods in Molecular Biology. Springer Science+Business Media, LLC, 2013, vol. 1021.

[3] M. Altaf-Ul-Amin, F. M. Afendi, S. K. Kiboi, and S. Kanaya, "Systems biology in the context of big data and networks," *Biomed Res Int*, vol. 2014, p. 428570, 2014.

[4] G. Parmigiani, *The analysis of gene expression data: methods and software*. New York: Springer-Verlag, 2003.

[5] Y. Saeys, I. Inza, and P. Larranaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, pp. 2507–2517, 2007.

[6] R. Bao, L. Huang, J. Andrade, W. Tan, W. Kibbe, H. Jiang, and G. Feng, "Review of current methods, applications, and data management for the bioinformatics analysis of whole exome sequencing," *Cancer Informatics*, p. 67, 2014.

[7] G. V. P. Raju, S. R. Peri, and C. S. Vasamsetty, "Gene expression analysis methods on microarray dataa review," *Global Journal of Computer Science and Technology*, vol. 14, 2014.

[8] J. A. Navas-Molina, J. M. Peralta-Sánchez, A. González, P. J. McMurdie, Y. Vázquez-Baeza, Z. Xu, L. K. Ursell, C. Lauber, H. Zhou, S. J. Song, J. Huntley, G. L. Ackermann, D. Berg-Lyons, S. Holmes, J. G. Caporaso, and R. Knight, "Advancing our understanding of the human microbiome using qiime," in *Methods in Enzymology*. Elsevier BV, 2013, pp. 371–444. [Online]. Available: http://dx.doi.org/10.1016/b978-0-12-407863-5.00019-8

[9] M. G. I. Langille, J. Zaneveld, J. G. Caporaso, D. McDonald, D. Knights, J. A. Reyes, J. C. Clemente, D. E. Burkepile, R. L. V. Thurber, R. Knight, R. G. Beiko, and C. Huttenhower, "Predictive functional profiling of microbial communities using 16s rrna marker gene sequences," *Nat Biotechnol*, vol. 31, no. 9, pp. 814–821, aug 2013. [Online]. Available: http://dx.doi.org/10.1038/nbt.2676

[10] D. H. Parks and R. G. Beiko, "Identifying biologically relevant differences between metagenomic communities," *Bioinformatics*, vol. 26, no. 6, pp. 715–721, feb 2010. [Online]. Available: http://dx.doi.org/10.1093/bioinformatics/btq041

[11] S. Arat, A. Spivak, S. V. Horn, E. Thomas, C. Traini, G. Sathe, G. P. Livi, K. Ingraham, L. Jones, K. Aubart, D. J. Holmes, O. Naderer, and J. R. Brown, "Microbiome changes in healthy volunteers treated with gsk1322322, a novel antibiotic targeting bacterial peptide deformylase," *Antimicrob. Agents Chemother.*, vol. 59, no. 2, pp. 1182–1192, feb 2015. [Online]. Available: http://dx.doi.org/10.1128/aac.04506-14

[12] J. Lamb, "The connectivity map: a new tool for biomedical research," *Nat Rev Cancer*, vol. 7, no. 1, pp. 54–60, jan 2007. [Online]. Available: http://dx.doi.org/10.1038/nrc2044

[13] S. Arat, G. S. Bullerjahn, and R. Laubenbacher, "A network biology approach to denitrification in pseudomonas aeruginosa," *PLoS One*, vol. 10, no. 2, p. e0118235, 2015.

[14] R. Laubenbacher and B. Stigler, "A computational algebra approach to the reverse engineering of gene regulatory networks," *J Theor Biol*, vol. 229, no. 4, pp. 523–37, 2004.

[15] M. Blaser, P. Bork, C. Fraser, R. Knight, and J. Wang, "The microbiome explored: recent insights and future challenges," *Nat Rev Micro*, vol. 11, no. 3, pp. 213–217, feb 2013. [Online]. Available: http://dx.doi.org/10.1038/nrmicro2973

[16] E. Holmes, J. V. Li, J. R. Marchesi, and J. K. Nicholson, "Gut microbiota composition and activity in relation to host metabolic phenotype and disease risk," *Cell Metabolism*, vol. 16, no. 5, pp. 559–564, nov 2012. [Online]. Available: http://dx.doi.org/10.1016/j.cmet.2012.10.007

[17] S. Zeissig and R. S. Blumberg, "Life at the beginning: perturbation of the microbiota by antibiotics in early life and its role in health and disease," *Nat Immunol*, vol. 15, no. 4, pp. 307–310, mar 2014. [Online]. Available: http://dx.doi.org/10.1038/ni.2847

[18] M. J. Blaser and S. Falkow, "What are the consequences of the disappearing human microbiota?" *Nat Rev Micro*, vol. 7, no. 12, pp. 887–894, nov 2009. [Online]. Available: http://dx.doi.org/10.1038/nrmicro2245

[19] M. Blaser, "Antibiotic overuse: Stop the killing of beneficial bacteria," *Nature*, vol. 476, no. 7361, pp. 393–394, aug 2011. [Online]. Available: http://dx.doi.org/10.1038/476393a

[20] A. E. Pérez-Cobas, A. Artacho, H. Knecht, M. L. Ferrús, A. Friedrichs, S. J. Ott, A. Moya, A. Latorre, and M. J. Gosalbes, "Differential effects of antibiotic therapy on the structure and function of human gut microbiota," *PLoS ONE*, vol. 8, no. 11, p. e80201, nov 2013. [Online]. Available: http://dx.doi.org/10.1371/journal.pone.0080201

[21] A. E. Perez-Cobas, M. J. Gosalbes, A. Friedrichs, H. Knecht, A. Artacho, K. Eismann, W. Otto, D. Rojo, R. Bargiela, M. von Bergen, S. C. Neulinger, C. Daumer, F.-A. Heinsen, A. Latorre, C. Barbas, J. Seifert, V. M. dos Santos, S. J. Ott, M. Ferrer, and A. Moya, "Gut microbiota disturbance during antibiotic therapy: a multi-omic approach," *Gut*, vol. 62, no. 11, pp. 1591–1601, dec 2012. [Online]. Available: http://dx.doi.org/10.1136/gutjnl-2012-303184

[22] S. Y. Shaw, J. F. Blanchard, and C. N. Bernstein, "Association between the use of antibiotics and new diagnoses of crohns disease and ulcerative colitis," *The American Journal of Gastroenterology*, vol. 106, no. 12, pp. 2133–2142, sep 2011. [Online]. Available: http://dx.doi.org/10.1038/ajg.2011.304

[23] L. Jostins, S. Ripke, R. K. Weersma, R. H. Duerr, D. P. McGovern, K. Y. Hui, J. C. Lee, L. P. Schumm, Y. Sharma, C. A. Anderson, J. Essers, M. Mitrovic, K. Ning, I. Cleynen, E. Theatre, S. L. Spain, S. Raychaudhuri, P. Goyette, Z. Wei, C. Abraham, J.-P. Achkar, T. Ahmad, L. Amininejad, A. N. Ananthakrishnan, V. Andersen, J. M. Andrews, L. Baidoo, T. Balschun, P. A. Bampton, A. Bitton, G. Boucher, S. Brand, C. Bning, A. Cohain, S. Cichon, M. D'Amato, D. D. Jong, K. L. Devaney, M. Dubinsky, C. Edwards, D. Ellinghaus, L. R. Ferguson, D. Franchimont, K. Fransen, R. Gearry, M. Georges, C. Gieger, J. Glas, T. Haritunians, A. Hart, C. Hawkey, M. Hedl, X. Hu, T. H. Karlsen, L. Kupcinskas, S. Kugathasan, A. Latiano, D. Laukens, I. C. Lawrance, C. W. Lees, E. Louis, G. Mahy, J. Mansfield, A. R. Morgan, C. Mowat, W. Newman, O. Palmieri, C. Y. Ponsioen, U. Potocnik, N. J. Prescott, M. Regueiro, J. I. Rotter, R. K. Russell, J. D. Sanderson, M. Sans, J. Satsangi, S. Schreiber, L. A. Simms, J. Sventoraityte, S. R.

Targan, K. D. Taylor, M. Tremelling, H. W. Verspaget, M. D. Vos, C. Wijmenga, D. C. Wilson, J. Winkelmann, R. J. Xavier, S. Zeissig, B. Zhang, C. K. Zhang, H. Zhao, M. S. Silverberg, V. Annese, H. Hakonarson, S. R. Brant, G. Radford-Smith, C. G. Mathew, J. D. Rioux, E. E. Schadt, M. J. Daly, A. Franke, M. Parkes, S. Vermeire, J. C. Barrett, and J. H. Cho, "Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease," *Nature*, vol. 491, no. 7422, pp. 119–124, oct 2012. [Online]. Available: http://dx.doi.org/10.1038/nature11582

[24] A. Semic-Jusufagic, D. Belgrave, A. Pickles, A. G. Telcian, E. Bakhsoliani, A. Sykes, A. Simpson, S. L. Johnston, and A. Custovic, "Assessing the association of early life antibiotic prescription with asthma exacerbations, impaired antiviral immunity, and genetic variants in 17q21: a population-based birth cohort study," *The Lancet Respiratory Medicine*, vol. 2, no. 8, pp. 621–630, aug 2014. [Online]. Available: http://dx.doi.org/10.1016/s2213-2600(14)70096-7

[25] G. L. Peterfreund, L. E. Vandivier, R. Sinha, A. J. Marozsan, W. C. Olson, J. Zhu, and F. D. Bushman, "Succession in the gut microbiome following antibiotic and antibody therapies for clostridium difficile," *PLoS ONE*, vol. 7, no. 10, p. e46966, oct 2012. [Online]. Available: http://dx.doi.org/10.1371/journal.pone.0046966

[26] I. Cho, S. Yamanishi, L. Cox, B. A. Methé, J. Zavadil, K. Li, Z. Gao, D. Mahana, K. Raju, I. Teitler, H. Li, A. V. Alekseyenko, and M. J. Blaser, "Antibiotics in early life alter the murine colonic microbiome and adiposity," *Nature*, vol. 488, no. 7413, pp. 621–626, aug 2012. [Online]. Available: http://dx.doi.org/10.1038/nature11400

[27] M. Membrez, F. Blancher, M. Jaquet, R. Bibiloni, P. D. Cani, R. G. Burcelin, I. Corthesy, K. Mace, and C. J. Chou, "Gut microbiota modulation with norfloxacin and ampicillin enhances glucose tolerance in mice," *The FASEB Journal*, vol. 22, no. 7, pp. 2416–2426, jul 2008. [Online]. Available: http://dx.doi.org/10.1096/fj.07-102723

[28] M. R. Jacobs, "The alexander project 1998-2000: susceptibility of pathogens isolated from community-acquired respiratory tract infection to commonly used antimicrobial agents," *Journal of Antimicrobial Chemotherapy*, vol. 52, no. 2, pp. 229–246, jul 2003. [Online]. Available: http://dx.doi.org/10.1093/jac/dkg321

[29] K. ODwyer, M. Hackel, S. Hightower, D. Hoban, S. Bouchillon, D. Qin, K. Aubart, M. Zalacain, and D. Butler, "Comparative analysis of the antibacterial activity

of a novel peptide deformylase inhibitor, gsk1322322," *Antimicrobial Agents and Chemotherapy*, vol. 57, no. 5, pp. 2333–2342, mar 2013. [Online]. Available: http://dx.doi.org/10.1128/aac.02566-12

[30] J.-P. Guilloteau, M. Mathieu, C. Giglione, V. Blanc, A. Dupuy, M. Chevrier, P. Gil, A. Famechon, T. Meinnel, and V. Mikol, "The crystal structures of four peptide deformylases bound to the antibiotic actinonin reveal two distinct types: A platform for the structure-based design of antibacterial agents," *Journal of Molecular Biology*, vol. 320, no. 5, pp. 951–962, jul 2002. [Online]. Available: http://dx.doi.org/10.1016/s0022-2836(02)00549-1

[31] J. Huang, G. S. V. Aller, A. N. Taylor, J. J. Kerrigan, W.-S. Liu, J. M. Trulli, Z. Lai, D. Holmes, K. M. Aubart, J. R. Brown, and M. Zalacain, "Phylogenomic and biochemical characterization of three legionella pneumophila polypeptide deformylases," *Journal of Bacteriology*, vol. 188, no. 14, pp. 5249–5257, jun 2006. [Online]. Available: http://dx.doi.org/10.1128/jb.00866-05

[32] O. J. Naderer, L. S. Jones, J. Zhu, M. Kurtinecz, and E. Dumont, "Safety, tolerability, and pharmacokinetics of oral and intravenous administration of gsk1322322, a peptide deformylase inhibitor," *The Journal of Clinical Pharmacology*, pp. n/a–n/a, aug 2013. [Online]. Available: http://dx.doi.org/10.1002/jcph.150

[33] O. J. Naderer, E. Dumont, J. Zhu, M. Kurtinecz, and L. S. Jones, "Single-dose safety, tolerability, and pharmacokinetics of the antibiotic gsk1322322, a novel peptide deformylase inhibitor," *Antimicrobial Agents and Chemotherapy*, vol. 57, no. 5, pp. 2005–2009, feb 2013. [Online]. Available: http://dx.doi.org/10.1128/aac.01779-12

[34] J. Peterson, S. Garges, M. Giovanni, P. McInnes, L. Wang, J. A. Schloss, V. Bonazzi, J. E. McEwen, K. A. Wetterstrand, C. Deal, C. C. Baker, V. D. Francesco, T. K. Howcroft, R. W. Karp, R. D. Lunsford, C. R. Wellington, T. Belachew, M. Wright, C. Giblin, H. David, M. Mills, R. Salomon, C. Mullins, B. Akolkar, L. Begg, C. Davis, L. Grandison, M. Humble, J. Khalsa, A. R. Little, H. Peavy, C. Pontzer, M. Portnoy, M. H. Sayre, P. Starke-Reed, S. Zakhari, J. Read, B. Watson, and M. Guyer, "The nih human microbiome project," *Genome Research*, vol. 19, no. 12, pp. 2317–2323, oct 2009. [Online]. Available: http://dx.doi.org/10.1101/gr.096651.109

[35] L. M. Proctor, "The human microbiome project in 2011 and beyond," *Cell Host & Microbe*, vol. 10, no. 4, pp. 287–291, oct 2011. [Online]. Available: http://dx.doi.org/10.1016/j.chom.2011.10.001

[36] A. Chao, "Nonparametric estimation of the number of classes in a population," *Scandinavian Journal of Statistics*, vol. 11, no. 4, pp. 265–270, 1984.

[37] S. Panda, I. E. khader, F. Casellas, J. L. Vivancos, M. G. Cors, A. Santiago, S. Cuenca, F. Guarner, and C. Manichanh, "Short-term effect of antibiotics on human gut microbiota," *PLoS ONE*, vol. 9, no. 4, p. e95476, apr 2014. [Online]. Available: http://dx.doi.org/10.1371/journal.pone.0095476

[38] R. A. Fisher, "The comparison of samples with possibly unequal variances," *Annals of Eugenics*, vol. 9, no. 2, pp. 174–180, 1939. [Online]. Available: http://dx.doi.org/10.1111/j.1469-1809.1939.tb02205.x

[39] O. J. Naderer, E. Dumont, J. Zhu, M. Kurtinecz, and L. S. Jones, "Safety, tolerability and pharmacokinetics of repeat dosing of the antibiotic GSK1322322, a peptide deformylase inhibitor: a randomized placebo-controlled study," *Journal of Antimicrobial Chemotherapy*, vol. 68, no. 8, pp. 1901–1909, apr 2013. [Online]. Available: http://dx.doi.org/10.1093/jac/dkt097

[40] J. Neyman and E. S. Pearson, "On the use and interpretation of certain test criteria for purposes of statistical inference: Part i," *Biometrika*, vol. 20A, p. 175, 1928.

[41] O. J. Dunn, "Multiple comparisons among means," *Journal of the American Statistical Association*, vol. 56, p. 52, 1961.

[42] W. Jia, R. N. Whitehead, L. Griffiths, C. Dawson, R. H. Waring, D. B. Ramsden, J. O. Hunter, and J. A. Cole, "Is the abundance of faecalibacterium prausnitzii relevant to crohns disease?" *FEMS Microbiology Letters*, vol. 310, no. 2, pp. 138–144, aug 2010. [Online]. Available: http://dx.doi.org/10.1111/j.1574-6968.2010.02057.x

[43] K. Gerasimidis, M. Bertz, L. Hanske, J. Junick, O. Biskou, M. Aguilera, V. Garrick, R. K. Russell, M. Blaut, P. McGrogan, and C. A. Edwards, "Decline in presumptively protective gut bacterial species and metabolites are paradoxically associated with disease improvement in pediatric crohn's disease during enteral nutrition," *Inflammatory*

Bowel Diseases, vol. 20, no. 5, pp. 861–871, 2014. [Online]. Available: http://dx.doi.org/10.1097/mib.0000000000000023

[44] S. K. Mazmanian, J. L. Round, and D. L. Kasper, "A microbial symbiosis factor prevents intestinal inflammatory disease," *Nature*, vol. 453, no. 7195, pp. 620–625, may 2008. [Online]. Available: http://dx.doi.org/10.1038/nature07008

[45] J. Molina, G. Barrantes, C. Quesada-Gómez, C. Rodríguez, and E. Rodríguez-Cavallini, "Phenotypic and genotypic characterization of multidrug-resistant bacteroides , parabacteroides spp., and pseudoflavonifractor from a costa rican hospital," *Microbial Drug Resistance*, vol. 20, no. 5, pp. 478–484, oct 2014. [Online]. Available: http://dx.doi.org/10.1089/mdr.2013.0180

[46] S. E. Winter, C. A. Lopez, and A. J. Bumler, "The dynamics of gut-associated microbial communities during inflammation," *EMBO Rep*, vol. 14, no. 4, pp. 319–327, mar 2013. [Online]. Available: http://dx.doi.org/10.1038/embor.2013.27

[47] D. Groeger, L. O'Mahony, E. F. Murphy, J. F. Bourke, T. G. Dinan, B. Kiely, F. Shanahan, and E. M. Quigley, "Bifidobacterium infantis 35624 modulates host inflammatory processes beyond the gut," *Gut Microbes*, vol. 4, no. 4, pp. 325–339, jul 2013. [Online]. Available: http://dx.doi.org/10.4161/gmic.25487

[48] L. Dethlefsen and D. A. Relman, "Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation," *Proceedings of the National Academy of Sciences*, vol. 108, no. Supplement_1, pp. 4554–4561, sep 2010. [Online]. Available: http://dx.doi.org/10.1073/pnas.1000087107

[49] G. Hajishengallis, R. P. Darveau, and M. A. Curtis, "The keystone-pathogen hypothesis," *Nat Rev Micro*, vol. 10, no. 10, pp. 717–725, sep 2012. [Online]. Available: http://dx.doi.org/10.1038/nrmicro2873

[50] X. Ze, F. L. Mougen, S. H. Duncan, P. Louis, and H. J. Flint, "Some are more equal than others: the role of "keystone" species in the degradation of recalcitrant substrates," *Gut Microbes*, vol. 4, no. 3, pp. 236–240, may 2013. [Online]. Available: http://dx.doi.org/10.4161/gmic.23998

[51] C. A. Lozupone, M. Hamady, S. T. Kelley, and R. Knight, "Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial

communities," *Applied and Environmental Microbiology*, vol. 73, no. 5, pp. 1576–1585, jan 2007. [Online]. Available: http://dx.doi.org/10.1128/aem.01996-06

[52] M. Kanehisa, "Kegg: Kyoto encyclopedia of genes and genomes," *Nucleic Acids Research*, vol. 28, no. 1, pp. 27–30, jan 2000. [Online]. Available: http://dx.doi.org/10.1093/nar/28.1.27

[53] D. I. Andersson and D. Hughes, "Microbiological effects of sublethal levels of antibiotics," *Nat Rev Micro*, vol. 12, no. 7, pp. 465–478, may 2014. [Online]. Available: http://dx.doi.org/10.1038/nrmicro3270

[54] J. A. Delmar, C.-C. Su, and E. W. Yu, "Bacterial multidrug efflux transporters," *Annual Review of Biophysics*, vol. 43, no. 1, pp. 93–117, may 2014. [Online]. Available: http://dx.doi.org/10.1146/annurev-biophys-051013-022855

[55] G. Cox and G. D. Wright, "Intrinsic antibiotic resistance: Mechanisms, origins, challenges and solutions," *International Journal of Medical Microbiology*, vol. 303, no. 6-7, pp. 287–292, aug 2013. [Online]. Available: http://dx.doi.org/10.1016/j.ijmm.2013.02.009

[56] J. G. Caporaso, C. L. Lauber, W. A. Walters, D. Berg-Lyons, J. Huntley, N. Fierer, S. M. Owens, J. Betley, L. Fraser, M. Bauer, N. Gormley, J. A. Gilbert, G. Smith, and R. Knight, "Ultra-high-throughput microbial community analysis on the illumina hiseq and miseq platforms," *The ISME Journal*, vol. 6, no. 8, pp. 1621–1624, mar 2012. [Online]. Available: http://dx.doi.org/10.1038/ismej.2012.8

[57] R. C. Edgar, "Search and clustering orders of magnitude faster than blast," *Bioinformatics*, vol. 26, no. 19, pp. 2460–2461, aug 2010. [Online]. Available: http://dx.doi.org/10.1093/bioinformatics/btq461

[58] T. Z. DeSantis, P. Hugenholtz, N. Larsen, M. Rojas, E. L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu, and G. L. Andersen, "Greengenes, a chimera-checked 16s rrna gene database and workbench compatible with arb," *Applied and Environmental Microbiology*, vol. 72, no. 7, pp. 5069–5072, jul 2006. [Online]. Available: http://dx.doi.org/10.1128/aem.03006-05

[59] D. McDonald, M. N. Price, J. Goodrich, E. P. Nawrocki, T. Z. DeSantis, A. Probst, G. L. Andersen, R. Knight, and P. Hugenholtz, "An improved greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea,"

The ISME Journal, vol. 6, no. 3, pp. 610–618, dec 2011. [Online]. Available: http://dx.doi.org/10.1038/ismej.2011.139

[60] C. Lozupone, M. E. Lladser, D. Knights, J. Stombaugh, and R. Knight, "Unifrac: an effective distance metric for microbial community comparison," *The ISME Journal*, vol. 5, no. 2, pp. 169–172, sep 2010. [Online]. Available: http://dx.doi.org/10.1038/ismej.2010.133

[61] M. Kanehisa, S. Goto, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe, "Data, information, knowledge and principle: back to metabolism in kegg," *Nucleic Acids Research*, vol. 42, no. D1, pp. D199–D205, nov 2013. [Online]. Available: http://dx.doi.org/10.1093/nar/gkt1076

[62] F. Fouhy, C. M. Guinane, S. Hussey, R. Wall, C. A. Ryan, E. M. Dempsey, B. Murphy, R. P. Ross, G. F. Fitzgerald, C. Stanton, and P. D. Cotter, "High-throughput sequencing reveals the incomplete, short-term recovery of infant gut microbiota following parenteral antibiotic treatment with ampicillin and gentamicin," *Antimicrobial Agents and Chemotherapy*, vol. 56, no. 11, pp. 5811–5820, sep 2012. [Online]. Available: http://dx.doi.org/10.1128/aac.00789-12

[63] D. Butler, D. Chen, K. ODwyer, T. Lewandowski, K. Aubart, and M. Zalacain, "Potent sub-mic effect of gsk1322322 and other peptide deformylase inhibitors on in vitro growth of staphylococcus aureus," *Antimicrobial Agents and Chemotherapy*, vol. 58, no. 1, pp. 290–296, oct 2013. [Online]. Available: http://dx.doi.org/10.1128/aac.01292-13

[64] J. A. Sutcliffe, "Antibiotics in development targeting protein synthesis," *Annals of the New York Academy of Sciences*, vol. 1241, no. 1, pp. 122–152, dec 2011. [Online]. Available: http://dx.doi.org/10.1111/j.1749-6632.2011.06323.x

[65] L. Zhang, Y. Huang, Y. Zhou, T. Buckley, and H. H. Wang, "Antibiotic administration routes significantly influence the levels of antibiotic resistance in gut microbiota," *Antimicrobial Agents and Chemotherapy*, vol. 57, no. 8, pp. 3659–3666, may 2013. [Online]. Available: http://dx.doi.org/10.1128/aac.00670-13

[66] W. H. Organization, "Global tuberculosis report," Tech. Rep., 2013.

[67] A. Karlas, N. Machuy, Y. Shin, K.-P. Pleissner, A. Artarini, D. Heuer, D. Becker, H. Khalil, L. A. Ogilvie, S. Hess, A. P. Murer, E. Mller, T. Wolff, T. Rudel, and T. F. Meyer, "Genome-wide RNAi screen identifies human host factors crucial for influenza

virus replication," *Nature*, vol. 463, no. 7282, pp. 818–822, jan 2010. [Online]. Available: http://dx.doi.org/10.1038/nature08760

[68] D. Kumar, L. Nath, M. A. Kamal, A. Varshney, A. Jain, S. Singh, and K. V. Rao, "Genome-wide analysis of the host intracellular network that regulates survival of mycobacterium tuberculosis," *Cell*, vol. 140, no. 5, pp. 731–743, mar 2010. [Online]. Available: http://dx.doi.org/10.1016/j.cell.2010.02.012

[69] M. P. R. Berry, C. M. Graham, F. W. McNab, Z. Xu, S. A. A. Bloch, T. Oni, K. A. Wilkinson, R. Banchereau, J. Skinner, R. J. Wilkinson, C. Quinn, D. Blankenship, R. Dhawan, J. J. Cush, A. Mejias, O. Ramilo, O. M. Kon, V. Pascual, J. Banchereau, D. Chaussabel, and A. O/'Garra, "An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis," *Nature*, vol. 466, no. 7309, pp. 973–977, Aug. 2010. [Online]. Available: http://dx.doi.org/10.1038/nature09247

[70] J. Park, I. Munagala, H. Xu, D. Blankenship, P. Maffucci, D. Chaussabel, J. Banchereau, V. Pascual, and C. Cunningham-Rundles, "Interferon Signature in the Blood in Inflammatory Common Variable Immune Deficiency," *PLoS ONE*, vol. 8, no. 9, p. e74893, Sep. 2013. [Online]. Available: http://dx.doi.org/10.1371%2Fjournal.pone.0074893

[71] J. Maertzdorf, J. Weiner, H.-J. Mollenkopf, T. Network, T. Bauer, A. Prasse, J. Mller-Quernheim, and S. H. E. Kaufmann, "Common patterns and disease-related signatures in tuberculosis and sarcoidosis," *Proceedings of the National Academy of Sciences*, vol. 109, no. 20, pp. 7853–7858, 2012. [Online]. Available: http://www.pnas.org/content/109/20/7853.abstract

[72] Y. Cai, Q. Yang, Y. Tang, M. Zhang, H. Liu, G. Zhang, Q. Deng, J. Huang, Z. Gao, B. Zhou, C. G. Feng, and X. Chen, "Increased Complement C1q Level Marks Active Disease in Human Tuberculosis," *PLoS ONE*, vol. 9, no. 3, p. e92340, 2014. [Online]. Available: http://dx.doi.org/10.1371%2Fjournal.pone.0092340

[73] C. Bergenfelz, A.-M. Larsson, K. von Stedingk, S. Gruvberger-Saal, K. Aaltonen, S. Jansson, H. Jernstrm, H. Janols, M. Wullt, A. Bredberg, L. Rydn, and K. Leandersson, "Systemic monocytic-mdscs are generated from monocytes and correlate with disease progression in breast cancer patients," *PLoS ONE*, vol. 10, no. 5, p. e0127028, 05 2015. [Online]. Available: http://dx.doi.org/10.1371%2Fjournal.pone.0127028

[74] S. B. Smith, W. Dampier, A. Tozeren, J. R. Brown, and M. Magid-Slav, "Identification of common biological pathways and drug targets across multiple respiratory viruses based on human host gene expression analysis," *PLoS ONE*, vol. 7, no. 3, p. e33174, mar 2012. [Online]. Available: http://dx.doi.org/10.1371/journal.pone.0033174

[75] S. B. Smith, M. Magid-Slav, and J. R. Brown, "Host response to respiratory bacterial pathogens as identified by integrated analysis of human gene expression data," *PLoS ONE*, vol. 8, no. 9, p. e75607, sep 2013. [Online]. Available: http://dx.doi.org/10.1371/journal.pone.0075607

[76] A. del Rey, C. V. Mahuad, V. V. Bozza, C. Bogue, M. A. Farroni, M. L. Bay, O. A. Bottasso, and H. O. Besedovsky, "Endocrine and cytokine responses in humans with pulmonary tuberculosis," *Brain, Behavior, and Immunity*, vol. 21, no. 2, pp. 171 – 179, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0889159106002558

[77] H. Yamada, S. Mizuno, and I. Sugawara, "Interferon regulatory factor 1 in mycobacterial infection," *Microbiology and immunology*, vol. 46, no. 11, pp. 751–760, 2002. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1111/j.1348-0421.2002.tb02760.x/full

[78] J.-F. Marquis, R. LaCourse, L. Ryan, R. J. North, and P. Gros, "Disseminated and Rapidly Fatal Tuberculosis in Mice Bearing a Defective Allele at IFN Regulatory Factor 8," *The Journal of Immunology*, vol. 182, no. 5, pp. 3008–3015, Mar. 2009. [Online]. Available: http://www.jimmunol.org/cgi/doi/10.4049/jimmunol.0800680

[79] P. K. Sen, R. Chatterjee, and J. R. Saha, "Use of sulphadimidine acetylation test in determining isoniazid inactivation rate in tuberculous patients," *Indian J Med Res*, vol. 60, no. 1, pp. 28–39, 1972.

[80] G. D. Bianchera, V. Gramazio, B. Occhialini, and E. Volpe, "[clinical casuistics of 43 tuberculotics with pleural or extrapleural empyema treated with the oleandomycin-tetracycline combination. clinical and laboratory observations, and therapeutic results]," *Minerva Med*, vol. 51, pp. 425–32, 1960.

[81] M. Ungureanu, D. Diculencu, and C. A. Stirbu, "Researches concerning the synthesis and antituberculosis action of some new sulphacetamide derivatives," *Rev Med Chir Soc Med Nat Iasi*, vol. 110, no. 2, pp. 456–9, 2006.

[82] A. D. Pranger, J. W. Alffenaar, and R. E. Aarnoutse, "Fluoroquinolones, the cornerstone of treatment of drug-resistant tuberculosis: a pharmacokinetic and pharmacodynamic approach," *Curr Pharm Des*, vol. 17, no. 27, pp. 2900–30, 2011.

[83] W. Weiss, G. M. Eisenberg, and H. F. Flippin, "Kanamycin and sulfisoxazole in the treatment of pulmonary tubeculosis," *Antibiotic Med Clin Ther (New York)*, vol. 6, pp. 731–3, 1959.

[84] M. Rosenblatt, "Remarks on some nonparametric estimates of a density function," *Ann. Math. Statist.*, vol. 27, no. 3, pp. 832–837, 09 1956. [Online]. Available: http://dx.doi.org/10.1214/aoms/1177728190

[85] E. Parzen, "On estimation of a probability density function and mode," *Ann. Math. Statist.*, vol. 33, no. 3, pp. 1065–1076, 09 1962. [Online]. Available: http://dx.doi.org/10.1214/aoms/1177704472

[86] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philosophical Magazine*, vol. 2, no. 11, p. 559572, 1901.

[87] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *J Educ Psychol*, vol. 24, p. 417520, 1933.

[88] C. F. Gauss, "Bestimmung der genauigkeit der beobachtungen," *Zeitschrift fr Astronomie und verwandte Wissenschaften*, vol. 1, p. 187197, 1816.

[89] A. M. Legendre, *Nouvelles mthodes pour la dtermination des orbites des comtes*, Paris, 1805.

[90] V. Law, C. Knox, Y. Djoumbou, T. Jewison, A. C. Guo, Y. Liu, A. Maciejewski, D. Arndt, M. Wilson, V. Neveu, A. Tang, G. Gabriel, C. Ly, S. Adamjee, Z. T. Dame, B. Han, Y. Zhou, and D. S. Wishart, "DrugBank 4.0: shedding new light on drug metabolism," *Nucleic Acids Research*, vol. 42, no. D1, pp. D1091–D1097, nov 2013. [Online]. Available: http://dx.doi.org/10.1093/nar/gkt1068

[91] W. G. Zumft, "Cell biology and molecular basis of denitrification," *Microbiol Mol Biol Rev*, vol. 61, no. 4, pp. 533–616, 1997.

[92] P. Groffman, E. Davidson, and S. S., "New approaches to modeling denitrification," *Biogeochemistry*, vol. 93, no. 1/2, pp. 1–5, 2009.

[93] K. Fennel, D. Brady, D. DiToro, R. Fulweiler, W. Gardner, A. Giblin, M. McCarthy, A. Rao, S. Seitzinger, M. Thouvenot-Korppoo, and C. Tobias, "Modeling denitrification in aquatic sediments," *Biogeochemistry*, vol. 93, no. 1/2, pp. 159–178, 2009.

[94] H. Arai, "Regulation and function of versatile aerobic and anaerobic respiratory metabolism in pseudomonas aeruginosa," *Front Microbiol*, vol. 2, p. 103, 2011.

[95] S. Seitzinger, J. A. Harrison, J. K. Bohlke, A. F. Bouwman, R. Lowrance, B. Peterson, C. Tobias, and G. Van Drecht, "Denitrification across landscapes and waterscapes: a synthesis," *Ecological Applications*, vol. 16, no. 6, pp. 2064–2090, 2006.

[96] C. E. Kupchella and M. C. Hyland, "Environmental science: Living within the system of nature," *Ecology*, 1989.

[97] H. Arai, M. Mizutani, and Y. Igarashi, "Transcriptional regulation of the nos genes for nitrous oxide reductase in pseudomonas aeruginosa," *Microbiology*, vol. 10, no. 2, pp. 29–36, 2015.

[98] K. Schreiber, R. Krieger, B. Benkert, M. Eschbach, H. Arai, M. Schobert, and D. Jahn, "The anaerobic regulatory network required for pseudomonas aeruginosa nitrate respiration," *J Bacteriol*, vol. 189, no. 11, pp. 4310–4, 2007.

[99] M. J. Kampschreur, R. Kleerebezem, C. Picioreanu, L. Bakken, L. Bergaust, S. de Vries, M. S. Jetten, and M. C. van Loosdrecht, "Metabolic modeling of denitrification in agrobacterium tumefaciens: a tool to study inhibiting and activating compounds for the denitrification pathway," *Front Microbiol*, vol. 3, p. 370, 2012.

[100] A. Jayakumar, G. D. O'Mullan, S. W. Naqvi, and B. B. Ward, "Denitrifying bacterial community composition changes associated with stages of denitrification in oxygen minimum zones," *Microb Ecol*, vol. 58, no. 2, pp. 350–62, 2009.

[101] K. E. Mampaey, B. Beuckels, M. J. Kampschreur, R. Kleerebezem, M. C. van Loosdrecht, and E. I. Volcke, "Modelling nitrous and nitric oxide emissions by autotrophic ammonia-oxidizing bacteria," *Environ Technol*, vol. 34, no. 9-12, pp. 1555–66, 2013.

[102] Y. Pan, B.-J. Ni, and Z. Yuan, "Modeling electron competition among nitrogen oxides reduction and n2o accumulation in denitrification," *Environ. Sci. Technol.*, vol. 47, pp. 11 083–11 091, 2013.

[103] A. Funahashi, Y. Matsuoka, A. Jouraku, M. Morohashi, N. Kikuchi, and H. Kitano, "Celldesigner 3.5: A versatile modeling tool for biochemical networks," *Proceedings of the IEEE*, vol. 96, pp. 1254–1265, 2008.

[104] B. L. Wanner, "Gene regulation by phosphate in enteric bacteria," *J Cell Biochem*, vol. 51, no. 1, pp. 47–54, 1993.

[105] L. W. Marzan, C. M. Hasan, and K. Shimizu, "Effect of acidic condition on the metabolic regulation of escherichia coli and its phob mutant," *Arch Microbiol*, vol. 195, no. 3, pp. 161–71, 2013.

[106] A. Filloux, M. Bally, C. Soscia, M. Murgier, and A. Lazdunski, "Phosphate regulation in pseudomonas aeruginosa: cloning of the alkaline phosphatase gene and identification of phob- and phor-like genes," *Mol Gen Genet*, vol. 212, no. 3, pp. 510–3, 1988.

[107] E. Galan-Vasquez, B. Luna, and A. Martinez-Antonio, "The regulatory network of pseudomonas aeruginosa," *Microb Inform Exp*, vol. 1, no. 1, p. 3, 2011.

[108] G. Giardina, S. Rinaldo, K. A. Johnson, A. Di Matteo, M. Brunori, and F. Cutruzzola, "No sensing in pseudomonas aeruginosa: structure of the transcriptional regulator dnr," *J Mol Biol*, vol. 378, no. 5, pp. 1002–15, 2008.

[109] C. H. Mortimer, "The exchange of dissolved substances between mud and water in lakes," *J. Ecol.*, no. 29, pp. 280–329, 1941.

[110] ——, "The exchange of dissolved substances between mud and water in lakes," *J. Ecol.*, no. 30, pp. 147–201, 1942.

[111] K. E. Elsbury, A. Paytan, N. Ostrom, C. Kendall, M. B. Young, K. McLaughlin, and S. Watson, "Using oxygen isotopes of phosphate to trace phosphorus sources and cycling in lake erie," *Environ. Sci. Technol.*, no. 43, pp. 3108–3114, 2009.

[112] T. Fujiwara and Y. Fukumori, "Cytochrome cb-type nitric oxide reductase with cytochrome c oxidase activity from paracoccus denitrificans atcc 35512," *J Bacteriol*, vol. 178, no. 7, pp. 1866–71, 1996.

[113] C. L. Hulse, J. M. Tiedje, and B. A. Averill, "A spectrophotometric assay for dissimilatory nitrite reductases," *Anal Biochem*, vol. 172, no. 2, pp. 420–6, 1988.

[114] J. K. Kristjansson and T. C. Hollocher, "First practical assay for soluble nitrous oxide reductase of denitrifying bacteria and a partial kinetic characterization," *J Biol Chem*, vol. 255, no. 2, pp. 704–7, 1980.

[115] "US Environmental Protection Agency," http://www.epa.gov/glindicators/water/phosphorusb.html, accessed 03/03/2014.

[116] "US Environmental Protection Agency," http://www.epa.gov/grtlakes/monitoring/limnology/index.html, accessed 03/03/2014.

[117] J. Makarewicz, P. Bertram, and L. T.W., "Chemistry of the offshore surface waters of lake erie: Pre- and post-dreissena introduction (1983-1993)," *J. Great Lakes Res.*, vol. 26, pp. 82–93, 2000.

[118] S. Wilhelm, G. Bullerjahn, M. Eldridge, J. Rinta-Kanto, L. Poorvin, and R. Bourbonniere, "Seasonal hypoxia and the genetic diversity of prokaryote populations in the central basin hypolimnion of lake erie: evidence for abundant cyanobacteria and photosynthesis," *J. Great Lakes Res.*, vol. 32, no. 4, pp. 657–671, 2006.

[119] B. Palsson, *Systems Biology: Simulation of Dynamic Network States*. Cambridge University Press, 2011.

[120] R. E. Hecky, P. Campbell, and L. L. Hendzel, "The stoichiometry of carbon, nitrogen and phosphorus in particulate matter of lakes and oceans," *Limnol Oceanogr*, vol. 38, no. 4, pp. 709–724, 1993.

[121] J. S. Almeida, M. A. M. Reis, and M. J. T. Carrondo, "Competition between nitrate and nitrite reduction in denitrification by pseudomonas fluorescens," *Biotechnol Bioeng*, vol. 46, p. 476484, 1995.

[122] H. Bootsma, "Naturally occurring chemicals: Nutrients," *State of the Lakes Ecosystem Conference, Milwaukee, WI (1-3 November, 2006)*, 2006.

[123] I. Loladze and J. J. Elser, "The origins of the redfield nitrogen-to-phosphorus ratio are in a homoeostatic protein-to-rrna ratio," *Ecol Lett*, vol. 14, no. 3, pp. 244–50, 2011.

[124] L. Philippot, J. uhel, N. P. A. Saby, D. Chneby, A. Chrokov, D. Bru, D. Arrouays, F. Martin-Laurent, and M. imek, "Mapping field-scale spatial patterns of size and activity of the denitrifier community," *Environ. Microbiol.*, vol. 11, no. 6, pp. 1518–1526, 2009.

[125] L. Philippot, J. Andert, C. Jones, B. D., and H. S., "Importance of denitrifiers lacking the genes for nitrous oxide reductase for n2o emissions from soil," *Glob. Change Biol.*, vol. 17, no. 3, pp. 1497–1504, 2011.

[126] J. F. Salles, X. Le Roux, and F. Poly, "Relating phylogenetic and functional diversity among denitrifiers and quantifying their capacity to predict community functioning," *Frontiers in Microbiology*, vol. 3, 2012.

[127] L. R. Bakken, L. Bergaust, B. Liu, and A. Frostegrd, "Regulation of denitrification at the cellular level: a clue to the understanding of n2o emissions from soils," *Phil. Trans. R. Soc. B*, vol. 367, pp. 1226–1234, 2011.

[128] M. Toyofuku, H. Uchiyama, and N. Nomura, "Social behaviours under anaerobic conditions in pseudomonas aeruginosa," *Int J Microbiol*, vol. 2012, p. 405191, 2012.

[129] D. Murrugarra, A. Veliz-Cuba, B. Aguilar, S. Arat, and R. Laubenbacher, "Modeling stochasticity and variability in gene regulatory networks," *EURASIP J Bioinform Syst Biol*, vol. 2012, no. 1, p. 5, 2012.

[130] L. D. Kier, R. M. Weppelman, and B. N. Ames, "Regulation of nonspecific acid phosphatase in salmonella: phon and phop genes," *J Bacteriol*, vol. 138, no. 1, pp. 155–61, 1979.

[131] E. A. Groisman, F. Heffron, and F. Solomon, "Molecular genetic analysis of the escherichia coli phop locus," *J Bacteriol*, vol. 174, no. 2, pp. 486–91, 1992.

[132] A. Veliz-Cuba, A. S. Jarrah, and R. Laubenbacher, "Polynomial algebra of discrete models in systems biology," *Bioinformatics*, vol. 26, no. 13, pp. 1637–43, 2010.

[133] F. Hinkelmann, M. Brandon, B. Guang, R. McNeill, G. Blekherman, A. Veliz-Cuba, and R. Laubenbacher, "Adam: Analysis of discrete models of biological systems using computer algebra," *BMC Bioinformatics*, vol. 12, p. 295, 2011.

[134] R. Y. Stanier, N. J. Palleroni, and M. Doudoroff, "The aerobic pseudomonads: a taxonomic study," *J . Gen. Microbial*, vol. 43, pp. 159–271, 1966.

[135] C. L. Dybas, "Dead zones spreading in world oceans," *Bioscience*, vol. 55, no. 7, pp. 552–557, 2005.

[136] J. Finlay, G. Small, and R. Sterner, "Human influences on nitrogen removal in lakes," *Science*, vol. 342, no. 6155, pp. 247–250, 2013.

[137] M. McCarthy, W. Gardner, P. Lavrentyev, K. Moats, F. Jochem, and D. Klarer, "Effects of hydrological flow regime on sediment-water interface and water column nitrogen dynamics in a great lakes coastal wetland (old woman creek, lake erie)," *Journal of Great Lakes Research*, vol. 33, no. 1, pp. 219–231, 2007.

[138] U. Romling, Y. Wingender, H. Muller, and B. Tummler, "A major pseudomonas aeruginosa clone common to patients and aquatic habitats," *Applied and Environmental Microbiology*, vol. 60, no. 6, pp. 1734–1738, 1994.

[139] D. A. Rodionov, I. L. Dubchak, A. P. Arkin, E. J. Alm, and M. S. Gelfand, "Dissimilatory metabolism of nitrogen oxides in bacteria: comparative reconstruction of transcriptional networks," *PLoS Comput Biol*, vol. 1, no. 5, p. e55, 2005.

[140] J. Chifman, R. Laubenbacher, and S. V. Torti, "A systems biology approach to iron metabolism," in *A Systems Biology Approach to Blood*, ser. Advances in Experimental Medicine and Biology, S. J. Corey, M. Kimmel, and J. N. Leonard, Eds. Springer New York, 2014, vol. 844, pp. 201–225. [Online]. Available: http://dx.doi.org/10.1007/978-1-4939-2095-2_10

[141] S. V. Torti and F. M. Torti, "Ironing out cancer," *Cancer Research*, vol. 71, no. 5, pp. 1511–1514, 2011.

[142] A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman, "Global cancer statistics," *CA: A Cancer Journal for Clinicians*, vol. 61, pp. 69–90, 2011.

[143] J. Chifman, A. Kniss, P. Neupane, I. Williams, B. Leung, Z. Deng, P. Mendes, V. Hower, F. Torti, S. Akman, S. Torti, and R. Laubenbacher, "The core control system of intracellular iron homeostasis: A mathematical model," *Journal of Theoretical Biology*, vol. 300, no. 0, pp. 91 – 99, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0022519312000343

[144] E. Nemeth, S. Rivera, V. Gabayan, C. Keller, S. Taudorf, B. K. Pedersen, and T. Ganz, "Il-6 mediates hypoferremia of inflammation by inducing the synthesis of the iron regulatory hormone hepcidin," *Journal of Clinical Investigation*, vol. 113, no. 9, pp. 1271–1276, 05 2004.

[145] Z. K. Pinnix, L. D. Miller, W. Wang, R. D'Agostino, T. Kute, M. C. Willingham, H. Hatcher, L. Tesfay, G. Sui, X. Di, S. V. Torti, and F. M. Torti, "Ferroportin and iron regulation in breast cancer progression and prognosis," *Science translational medicine*, vol. 2, no. 43, pp. 43ra56–43ra56, 08 2010. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3734848/

[146] P. N. Paradkar, K. B. Zumbrennen, B. H. Paw, D. M. Ward, and J. Kaplan, "Regulation of mitochondrial iron import through differential turnover of mitoferrin 1 and mitoferrin 2," *Molecular and Cellular Biology*, vol. 29, no. 4, pp. 1007–1016, 2009.

[147] A. Campanella, E. Rovelli, P. Santambrogio, A. Cozzi, F. Taroni, and S. Levi, "Mitochondrial ferritin limits oxidative damage regulating mitochondrial iron availability: hypothesis for a protective role in friedreich ataxia," *Human Molecular Genetics*, vol. 18, no. 1, pp. 1–11, 2009.

[148] C. Zou, H. Zhang, Q. Li, H. Xiao, L. Yu, S. Ke, L. Zhou, W. Liu, W. Wang, H. Huang, N. Ma, Q. Liu, X. Wang, W. Zhao, H. Zhou, and X. Gao, "Heme oxygenase-1: a molecular brake on hepatocellular carcinoma cell migration," *Carcinogenesis*, vol. 32, no. 12, pp. 1840–1848, 2011. [Online]. Available: http://carcin.oxfordjournals.org/content/32/12/1840.abstract

[149] G. Nie, A. D. Sheftel, S. F. Kim, and P. Ponka, "Overexpression of mitochondrial ferritin causes cytosolic iron depletion and changes cellular iron homeostasis," *Blood*, vol. 105, no. 5, pp. 2161–2167, 2004.

[150] D. R. Richardson, D. J. R. Lane, E. M. Becker, M. L.-H. Huang, M. Whitnall, Y. S. Rahmanto, A. D. Sheftel, and P. Ponka, "Mitochondrial iron trafficking and the integration of iron metabolism between the mitochondrion and cytosol," *Proceedings of the National Academy of Sciences*, vol. 107, no. 24, pp. 10 775–10 782, 2010.

[151] D. H. Kang, "Oxidative stress, dna damage, and breast cancer," *AACN Advanced Critical Care*, vol. 13, p. 540549, 2002.

[152] C. A. Cohn, S. R. Simon, and M. A. Schoonen, "Comparison of fluorescence-based techniques for the quantification of particle-induced hydroxyl radicals," *Part Fibre Toxicol*, vol. 5, p. 19, 2008.

[153] D. R. Gough and T. G. Cotter, "Hydrogen peroxide: a jekyll and hyde signalling molecule," *Cell Death and Disease*, vol. 2, p. e213, 2011.

[154] E. Birben, U. M. Sahiner, C. Sackesen, S. Erzurum, and O. Kalayci, "Oxidative stress and antioxidant defense," *World Allergy Organization Journal*, vol. 5, p. 9, 2012.

[155] J. A. McCubrey, M. M. LaHair, and R. A. Franklin, "Reactive oxygen species-induced activation of the map kinase signaling pathways," *Antioxidants & redox signaling*, vol. 8, p. 17751789, 2006.

[156] K. Taguchi, H. Motohashi, and M. Yamamoto, "Molecular mechanisms of the keap1-nrf2 pathway in stress response and cancer evolution: Molecular mechanisms of the keap1-nrf2 pathway," *Genes to Cells*, vol. 16, pp. 123–140, 2011.

[157] Y. Mitsuishi, H. Motohashi, and M. Yamamoto, "The keap1nrf2 system in cancers: stress response and anabolic metabolism," *Frontiers in Oncology*, vol. 2, 2012.

[158] M. C. Jaramillo and D. D. Zhang, "The emerging role of the nrf2-keap1 signaling pathway in cancer," *Genes Dev*, vol. 27, no. 20, pp. 2179–91, 2013.

[159] J. Alam, D. Stewart, C. Touchard, S. Boinapally, A. M. Choi, and J. L. Cook, "Nrf2, a capncollar transcription factor, regulates induction of the heme oxygenase-1 gene," *Journal of Biological Chemistry*, vol. 274, p. 2607126078, 1999.

[160] C. E. Paulsen, T. H. Truong, F. J. Garcia, A. Homann, V. Gupta, S. E. Leonard, and K. S. Carroll, "Peroxide-dependent sulfenylation of the egfr catalytic site enhances kinase activity," *Nature Chemical Biology*, vol. 8, pp. 57–64, 2011.

[161] Y. Yoshida, M. Maruyama, T. Fujita, N. Arai, R. Hayashi, J. Araya, S. Matsui, N. Yamashita, E. Sugiyama, and M. Kobayashi, "Reactive oxygen intermediates stimulate interleukin-6 production in human bronchial epithelial cells," *American Journal of Physiology-Lung Cellular and Molecular Physiology*, vol. 276, p. L900L908, 1999.

[162] C. Madeddu, G. Gramignano, C. Floris, G. Murenu, G. Sollai, and A. Macci, "Role of inflammation and oxidative stress in post-menopausal oestrogen-dependent breast cancer," *Journal of Cellular and Molecular Medicine*, vol. 18, pp. 2519–2529, 2014.

[163] P. van der Geer, T. Hunter, and R. A. Lindberg, "Receptor protein-tyrosine kinases and their signal transduction pathways," *Annual Review of Cell Biology*, vol. 10, no. 1, pp. 251–337, 1994, pMID: 7888178. [Online]. Available: http://dx.doi.org/10.1146/annurev.cb.10.110194.001343

[164] W.-B. Tsai, I. Aiba, Y. Long, H.-K. Lin, L. Feun, N. Savaraj, and M. T. Kuo, "Activation of ras/PI3K/ERK pathway induces c-myc stabilization to upregulate argininosuccinate synthetase, leading to arginine deiminase resistance in melanoma cells," *Cancer research*, vol. 72, no. 10, pp. 2622–2633, 05 2012. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3433038/

[165] K.-J. Wu, A. Polack, and R. Dalla-Favera, "Coordinated regulation of iron-controlling genes, h-ferritin and irp2, by c- myc," *Science*, vol. 283, no. 5402, pp. 676–679, 1999. [Online]. Available: http://www.sciencemag.org/content/283/5402/676.abstract

[166] K. A. O'Donnell, D. Yu, K. I. Zeller, J.-w. Kim, F. Racke, A. Thomas-Tikhonenko, and C. V. Dang, "Activation of transferrin receptor 1 by c-myc enhances cellular proliferation and tumorigenesis," *Molecular and Cellular Biology*, vol. 26, no. 6, pp. 2373–2386, 03 2006. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1430295/

[167] P. J. S. Stork, "Directing ngf's actions: it's a rap," *Nat Cell Biol*, vol. 7, no. 4, pp. 338–339, 04 2005. [Online]. Available: http://dx.doi.org/10.1038/ncb0405-338

[168] R. J. Orton, M. E. Adriaens, A. Gormand, O. E. Sturm, W. Kolch, and D. R. Gilbert, "Computational modelling of cancerous mutations in the egfr/erk signalling pathway," *BMC Systems Biology*, vol. 3, pp. 100–100, 2009. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2764635/

[169] M. Nakafuku, T. Satoh, and Y. Kaziro, "Differentiation factors, including nerve growth factor, fibroblast growth factor, and interleukin-6, induce an accumulation of an active ras. gtp complex in rat pheochromocytoma pc12 cells," *Journal of Biological Chemistry*, vol. 267, p. 1944819454, 1992.

[170] M. Rowley and B. Van Ness, "Activation of N-ras and K-ras induced by interleukin-6 in a myeloma cell line: implications for disease progression and therapeutic response," *ONCOGENE*, vol. 21, no. 57, pp. 8769–8775, DEC 12 2002.

[171] A. C. Lee, B. E. Fenster, H. Ito, K. Takeda, N. S. Bae, T. Hirai, Z.-X. Yu, V. J. Ferrans, B. H. Howard, and T. Finkel, "Ras proteins induce senescence by altering the intracellular levels of reactive oxygen species," *Journal of Biological Chemistry*, vol. 274, no. 12, pp. 7936–7940, 1999. [Online]. Available: http://www.jbc.org/content/274/12/7936.abstract

[172] P. S. Hole, L. Pearn, A. J. Tonks, P. E. James, A. K. Burnett, R. L. Darley, and A. Tonks, "Ras-induced reactive oxygen species promote growth factor–independent proliferation in human cd34+ hematopoietic progenitor cells," *Blood*, vol. 115, no. 6, pp. 1238–1246, 2009.

[173] W. Xu, J. Trepel, and L. Neckers, "Ras, ros and proteotoxic stress: A delicate balance," *Cancer Cell*, vol. 20, no. 3, pp. 281–282, 2011. [Online]. Available: http://dx.doi.org/10.1016/j.ccr.2011.08.020

[174] M.-T. Park, M.-J. Kim, Y. Suh, R.-K. Kim, H. Kim, E.-J. Lim, K.-C. Yoo, G.-H. Lee, Y.-H. Kim, S.-G. Hwang, J.-M. Yi, and S.-J. Lee, "Novel signaling axis for ros generation during k-ras-induced cellular transformation," *Cell Death Differ*, vol. 21, no. 8, pp. 1185–1197, 08 2014. [Online]. Available: http://dx.doi.org/10.1038/cdd.2014.34

[175] T. H. Truong and K. S. Carroll, "Redox regulation of epidermal growth factor receptor signaling through cysteine oxidation," *Biochemistry*, vol. 51, p. 99549965, 2012.

[176] K. L. Cheung, J. H. Lee, L. Shu, J.-H. Kim, D. B. Sacks, and A.-N. T. Kong, "The ras gtpase-activating-like protein iqgap1 mediates nrf2 protein activation via the mitogen-activated protein kinase/extracellular signal-regulated kinase (erk) kinase (mek)-erk pathway," *Journal of Biological Chemistry*, vol. 288, no. 31, pp. 22 378–22 386, 2013. [Online]. Available: http://www.jbc.org/content/288/31/22378.abstract

[177] H.-K. Na, E.-H. Kim, J.-H. Jung, H.-H. Lee, J.-W. Hyun, and Y.-J. Surh, "(?)-epigallocatechin gallate induces nrf2-mediated antioxidant enzyme expression via activation of {PI3K} and {ERK} in human mammary epithelial cells," *Archives of Biochemistry and Biophysics*, vol. 476, no. 2, pp. 171 – 177, 2008, highlight Issue: Polyphenol and Health. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0003986108001793

[178] A. Veliz-Cuba, A. S. Jarrah, and R. Laubenbacher, "Polynomial algebra of discrete models in systems biology," *Bioinformatics*, vol. 26, no. 13, pp. 1637–1643, 2010. [Online]. Available: http://bioinformatics.oxfordjournals.org/content/26/13/1637.abstract

[179] H. Was, J. Dulak, and A. Jozkowicz, "Heme oxygenase-1 in tumor biology and therapy," *Current drug targets*, vol. 11, p. 15511570, 2010.

[180] G. M. DeNicola, F. A. Karreth, T. J. Humpton, A. Gopinathan, C. Wei, K. Frese, D. Mangal, K. H. Yu, C. J. Yeo, E. S. Calhoun, F. Scrimieri, J. M. Winter, R. H. Hruban, C. Iacobuzio-

Donahue, S. E. Kern, I. A. Blair, and D. A. Tuveson, "Oncogene-induced nrf2 transcription promotes ros detoxification and tumorigenesis," *Nature*, vol. 475, pp. 106–109, 2011.

[181] S. V. Torti and F. M. Torti, "Iron and cancer: more ore to be mined," *Nature Reviews Cancer*, vol. 13, pp. 342–355, 2013.

[182] W. A. Wlassoff, C. D. Albright, M. S. Sivashinski, A. Ivanova, J. G. Appelbaum, and R. I. Salganik, "Hydrogen peroxide overproduced in breast cancer cells can serve as an anticancer prodrug generating apoptosis-stimulating hydroxyl radicals under the effect of tamoxifen-ferrocene conjugate," *Journal of Pharmacy and Pharmacology*, vol. 59, pp. 1549–1553, 2007.

[183] F. Tas, H. Hansel, A. Belce, S. Ilvan, A. Argon, H. Camlica, and E. Topuz, "Oxidative stress in breast cancer," *Medical Oncology*, vol. 22, p. 1115, 2005.

[184] S. J. Noh, J. S. Bae, U. Jamiyandorj, H. S. Park, K. S. Kwon, S. H. Jung, H. J. Youn, H. Lee, B.-H. Park, M. J. Chung, W. S. Moon, M. J. Kang, and K. Y. Jang, "Expression of nerve growth factor and heme oxygenase-1 predict poor survival of breast carcinoma patients," *BMC Cancer*, vol. 13, pp. 516–516, 2013.

[185] S. Sen, B. Kawahara, and G. Chaudhuri, "Maintenance of higher h 2 o 2 levels, and its mechanism of action to induce growth in breast cancer cells: important roles of bioactive catalase and pp2a," *Free Radical Biology and Medicine*, vol. 53, p. 15411551, 2012.

[186] P. Sansone, G. Storci, S. Tavolari, T. Guarnieri, C. Giovannini, M. Taffurelli, C. Ceccarelli, D. Santini, P. Paterini, K. B. Marcu, P. Chieco, and M. Bonaf, "Il-6 triggers malignant features in mammospheres from human ductal breast carcinoma and normal mammary gland," *Journal of Clinical Investigation*, vol. 117, pp. 3988–4002, 2007.

[187] P. Rodriguez-Viciana, O. Tetsu, K. Oda, J. Okada, K. Rauen, and F. McCormick, "Cancer targets in the ras pathway," in *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 70. Cold Spring Harbor Laboratory Press, 2005, p. 461467.

[188] F. C. von Lintig, A. D. Dreilinger, N. M. Varki, A. M. Wallace, D. E. Casteel, and G. R. Boss, "Ras activation in human breast cancer," *Breast cancer research and treatment*, vol. 62, p. 5162, 2000.

[189] D. J. Liao and R. B. Dickson, "c-myc in breast cancer," *Endocrine-related cancer*, vol. 7, p. 143164, 2000.

[190] S. Adhikary and M. Eilers, "Transcriptional regulation and transformation by myc proteins," *Nature Reviews Molecular Cell Biology*, vol. 6, pp. 635–645, 2005.

[191] W. Wang, Z. Deng, H. Hatcher, L. D. Miller, X. Di, L. Tesfay, G. Sui, R. B. D'Agostino, F. M. Torti, and S. V. Torti, "Irp2 regulates breast tumor growth," *Cancer Research*, vol. 74, no. 2, pp. 497–507, 2014.

[192] E. C. Pietsch, J. Y. Chan, F. M. Torti, and S. V. Torti, "Nrf2 mediates the induction of ferritin h in response to xenobiotics and cancer chemopreventive dithiolethiones," *Journal of Biological Chemistry*, vol. 278, no. 4, pp. 2361–2369, 2003.

Appendix A

A Denitrification Network Model of *Pseudomonas aeruginosa*

A.1 Transition tables of Dnr, NirQ, nar and NO_2

PmrA(t)	Anr(t)	NarXL(t)	Dnr(t)	Dnr(t+1)
x_3(t)	x_4(t)	x_5(t)	x_6(t)	x_6(t+1)
0	0	0	0	0
0	0	0	1	0
0	0	0	2	1
0	0	1	0	0
0	0	1	1	0
0	0	1	2	1
0	1	0	0	1
0	1	0	1	1
0	1	0	2	1
0	1	1	0	1
0	1	1	1	2
0	1	1	2	2
1	0	0	0	0
1	0	0	1	0
1	0	0	2	1
1	0	1	0	0
1	0	1	1	0
1	0	1	2	1
1	1	0	0	1
1	1	0	1	1
1	1	0	2	1
1	1	1	0	1
1	1	1	1	1
1	1	1	2	1

Table A.1: Transition table of Dnr

NarXL(t)	Dnr(t)	NirQ(t)	NirQ(t+1)
x_5(t)	x_6(t)	x_7(t)	x_7(t+1)
0	0	0	0
0	0	1	0
0	0	2	1
0	1	0	1
0	1	1	1
0	1	2	1
0	2	0	1
0	2	1	2
0	2	2	2
1	0	0	1
1	0	1	1
1	0	2	1
1	1	0	1
1	1	1	2
1	1	2	2
1	2	0	1
1	2	1	2
1	2	2	2

Table A.2: Transition table of NirQ

NarXL(t)	Dnr(t)	NO(t)	nar(t)	nar(t+1)
x_5(t)	x_6(t)	x_{13}(t)	x_8(t)	x_8(t+1)
0	0	0	0	0
0	0	0	1	0
0	0	0	2	1
0	0	1	0	0
0	0	1	1	0
0	0	1	2	1
0	0	2	0	0
0	0	2	1	0
0	0	2	2	1
0	1	0	0	0
0	1	0	1	0
0	1	0	2	1
0	1	1	0	0
0	1	1	1	0
0	1	1	2	1
0	1	2	0	0
0	1	2	1	0
0	1	2	2	1
0	2	0	0	0
0	2	0	1	0
0	2	0	2	1
0	2	1	0	0
0	2	1	1	0
0	2	1	2	1
0	2	2	0	1
0	2	2	1	1
0	2	2	2	1

NarXL(t)	Dnr(t)	NO(t)	nar(t)	nar(t+1)
x_5(t)	x_6(t)	x_{13}(t)	x_8(t)	x_8(t+1)
1	0	0	0	1
1	0	0	1	1
1	0	0	2	1
1	0	1	0	1
1	0	1	1	1
1	0	1	2	1
1	0	2	0	1
1	0	2	1	1
1	0	2	2	1
1	1	0	0	1
1	1	0	1	1
1	1	0	2	1
1	1	1	0	1
1	1	1	1	2
1	1	1	2	2
1	1	2	0	1
1	1	2	1	2
1	1	2	2	2
1	2	0	0	1
1	2	0	1	1
1	2	0	2	1
1	2	1	0	1
1	2	1	1	2
1	2	1	2	2
1	2	2	0	1
1	2	2	1	2
1	2	2	2	2

Table A.3: Transition table of nar

NO_3(t)	nar(t)	NO_2(t)	NO_2(t+1)
NO_3(t)	x_8(t)	x_{12}(t)	x_{12}(t+1)
0	0	0	0
0	0	1	0
0	0	2	1
0	1	0	0
0	1	1	0
0	1	2	1
0	2	0	0
0	2	1	0
0	2	2	1
1	0	0	0
1	0	1	0
1	0	2	1
1	1	0	1
1	1	1	1
1	1	2	1
1	2	0	1
1	2	1	2
1	2	2	2

Table A.4: Transition table of NO_2

Appendix B

Modeling Iron-dependent Oxidative Stress in Breast Cancer

B.1 Construction of the *update polynomial* for ferritin (Ft)

Ferritin (Ft) has two inhibitors: IRP1 and IRP2. States $\{0, 1, 2\}$ for Ft will denote protein concentrations *low, normal* and *high*, respectively. It has been suggested that active IRP2 has a greater affect on Ft, thus we will adjust the strength of each IRP as described by Table 5.3.1. The logic gate between two negated IRP's is a Min gate:

$$f_{Ft} = \text{Min}(*\overline{IRP1}, \overline{IRP2}).$$

This ensures that when, for example, IRP1 = 0 (inactive) and IRP2 = 2 (active), we get that Ft is inhibited by IRP2, i.e. Ft = 0 in that case, otherwise it would be 2 with a Max gate. Now we translate the above expression into a polynomial equation. First, let $x_4 :=$ Ft, $x_5 :=$ IRP1, and $x_6 :=$ IRP2 (this is the same assignment as we have in the supplemental file **??**). The polynomial functions over a field on three elements for each transition table are:

$$*\overline{x_5} = 2x_5^2 + 2 \quad \text{and} \quad \overline{x_6} = 2x_6 + 2 \tag{B.1}$$

Using appropriate polynomial for the Min gate as described by Equation (5.1), we compute the following update function for Ft by keeping in mind that all the calculations are over $\mathbb{F}_3$.

$$\begin{aligned} f_{x_4}(x_5, x_6) &= \text{Min}(*\overline{x_5}, \overline{x_6}) \\ &= \text{Min}(2x_5^2 + 2, 2x_6 + 2) \\ &= x_5^2 x_6^2 + 2x_5^2 + 2x_6 + 2. \end{aligned}$$

Now we apply continuity process as described by Equation (5.2) and carefully substitute that into Equation (5.3) to compute final polynomial $f4$ representing an update polynomial for x_4 (computations are modulo 3).

$$\begin{aligned} f_4 = {} & h(0, f_{x_4}(0,0)) \cdot (1-(x_4-0)^2)(1-(x_5-0)^2)(1-(x_6-0)^2) \\ & + h(1, f_{x_4}(0,0)) \cdot (1-(x_4-1)^2)(1-(x_5-0)^2)(1-(x_6-0)^2) \\ & \vdots \\ & + h(1, f_{x_4}(2,2)) \cdot (1-(x_4-1)^2)(1-(x_5-2)^2)(1-(x_6-2)^2) \\ & + h(2, f_{x_4}(2,2)) \cdot (1-(x_4-2)^2)(1-(x_5-2)^2)(1-(x_6-2)^2) \\ \\ = {} & 1 + x_4^2 + 2x_4^2 x_5^2 + 2x_6 + x_4 x_6 + 2x_4^2 x_6 + x_6^2 + 2x_4 x_6^2 + x_4^2 x_5^2 x_6^2 \end{aligned}$$

B.2 Variables and their Update Polynomials

$x_1 =$ LIP
$x_2 =$ TfR1
$x_3 =$ Fpn
$x_4 =$ Ft
$x_5 =$ IRP1
$x_6 =$ IRP2
$x_7 =$ Hep
$x_8 =$ HO-1
$x_9 =$ ALAS1
$x_{10} =$ heme
$x_{11} =$ ROS
$x_{12} =$ Antioxidant enzymes
$x_{13} =$ Nrf2
$x_{14} =$ Keap1
$x_{15} =$ IL-6
$x_{16} =$ Ras
$x_{17} =$ SOS
$x_{18} =$ ERK
$x_{19} =$ c-Myc
$x_{20} =$ GAP
$x_{21} =$ EGFR
$x_{22} =$ LIPmt
$x_{23} =$ Mfrn
$x_{24} =$ Ftmt

$$\mathbf{f_1} = x2+2*x2^2*x23+2*x2*x23^2+2*x2^2*x23^2+2*x2^2*x3+x2^2*x23*x3+2*x2^2*x23^2*x3+2*x2*x3^2+2*x2^2*x3^2+2*x2^2*x23*x3^2+x2*x23^2*x3^2+2*x2^2*x4+x2^2*x23*x4+2*x2^2*x23^2*x4+x2^2*x3*x4+2*x2^2*x23*x3*x4+x2^2*x23^2*x3*x4+2*x2^2*x3^2*x4+x2^2*x23*x3^2*x4+2*x2^2*x23^2*x3^2*x4+2*x2*x4^2+2*x2^2*x4^2+2*x2^2*x23*x4^2+x2*x23^2*x4^2+2*x2^2*x3*x4^2+x2^2*x23*x3*x4^2+2*x2^2*x23^2*x3*x4^2+x2*x3^2*x4^2+2*x2^2*x23*x3^2*x4^2+2*x2*x23^2*x3^2*x4^2+2*x2^2*x23^2*x3^2*x4^2+x8+2*x2*x8+x2^2*x8+2*x23^2*x8+x2*x23^2*x8+2*x2^2*x23^2*x8+2*x3^2*x8+x2*x3^2*x8+2*x2^2*x3^2*x8+x23^2*x3^2*x8+2*x2*x23^2*x3^2*x8+x2^2*x23^2*x3^2*x8+2*x4^2*x8+x2*x4^2*x8+2*x2^2*x4^2*x8+x23^2*x4^2*x8+2*x2*x23^2*x4^2*x8+x2^2*x23^2*x4^2*x8+x3^2*x4^2*x8+2*x2*x3^2*x4^2*x8+x2^2*x3^2*x4^2*x8+2*x23^2*x3^2*x4^2*x8+x2*x23^2*x3^2*x4^2*x8+2*x2^2*x23^2*x3^2*x4^2*x8+x2*x8^2+x2^2*x8^2+2*x23*x8^2+x2^2*x23*x8^2+2*x23^2*x8^2+2*x2*x23^2*x8^2+2*x3*x8^2+x2^2*x3*x8^2+x23*x3*x8^2+2*x2^2*x23*x3*x8^2+2*x23^2*x3*x8^2+x2^2*x23^2*x3*x8^2+2*x3^2*x8^2+2*x2*x3^2*x8^2+2*x23*x3^2*x8^2+x2^2*x23*x3^2*x8^2+x2*x23^2*x3^2*x8^2+x2^2*x23^2*x3^2*x8^2+2*x4*x8^2+x2^2*x4*x8^2+x23*x4*x8^2+2*x2^2*x23*x4*x8^2+2*x23^2*x4*x8^2+x2^2*x23^2*x4*x8^2+x3*x4*x8^2+2*x2^2*x3*x4*x8^2+2*x23*x3*x4*x8^2+x2^2*x23*x3*x4*x8^2+x23^2*x3*x4*x8^2+2*x2^2*x23^2*x3*x4*x8^2+2*x3^2*x4*x8^2+x2^2*x3^2*x4*x8^2+x23*x3^2*x4*x8^2+2*x2^2*x23*x3^2*x4*x8^2+2*x23^2*x3^2*x4*x8^2+x2^2*x23^2*x3^2*x4*x8^2+2*x4^2*x8^2+2*x2*x4^2*x8^2+2*x23*x4^2*x8^2+x2^2*x23*x4^2*x8^2+x2*x23^2*x4^2*x8^2+x2^2*x23^2*x4^2*x8^2+x3*x4^2*x8^2+2*x2*x3*x4^2*x8^2+x2^2*x3*x4^2*x8^2+2*x23*x3*x4^2*x8^2+x2^2*x23*x3*x4^2*x8^2+2*x2*x23^2*x3*x4^2*x8^2+x2^2*x23^2*x3*x4^2*x8^2+x3^2*x4^2*x8^2+2*x2*x3^2*x4^2*x8^2+x2^2*x3^2*x4^2*x8^2+2*x23*x3^2*x4^2*x8^2+x2^2*x23*x3^2*x4^2*x8^2+2*x2*x23^2*x3^2*x4^2*x8^2$$

$$\mathbf{f_2} = x19^2+x2+2*x19^2*x2+2*x2^2+x19*x2^2+x5^2+2*x19^2*x5^2+2*x2*x5^2+x19^2*x2*x5^2+x2^2*x5^2+2*x19^2*x2^2*x5^2+x2^2*x6+2*x19*x2^2*x6+x19^2*x2^2*x6+x6^2+2*x19^2*x6^2+2*x2*x6^2+x19^2*x2*x6^2+x19*x2^2*x6^2+x19^2*x2^2*x6^2+2*x5^2*x6^2+x19^2*x5^2*x6^2+x2*x5^2*x6^2+2*x19^2*x2*x5^2*x6^2+2*x2^2*x5^2*x6^2+x19^2*x2^2*x5^2*x6^2$$

$$\mathbf{f_3} = 1+x3^2+2*x3^2*x5^2+2*x6+x3*x6+2*x3^2*x6+x6^2+2*x3*x6^2+x3^2*x5^2*x6^2+2*x7+x3*x7+2*x3^2*x7+x6*x7+2*x3*x6*x7+x3^2*x6*x7+2*x6^2*x7+x3*x6^2*x7+2*x3^2*x6^2*x7+x7^2+2*x3*x7^2+x3^2*x5^2*x7^2+2*x6*x7^2+x3*x6*x7^2+2*x3^2*x6*x7^2+x6^2*x7^2+2*x3*x6^2*x7^2+2*x3^2*x6^2*x7^2+2*x3^2*x5^2*x6^2*x7^2$$

$$\mathbf{f_4} = 1+x4^2+2*x4^2*x5^2+2*x6+x4*x6+2*x4^2*x6+x6^2+2*x4*x6^2+x4^2*x5^2*x6^2$$

$$\mathbf{f_5} = 1+2*x1+x1^2+x1*x5+2*x1^2*x5+x5^2+2*x1*x5^2$$

$$\mathbf{f_6} = 1+2*x1+x1^2+x1*x19^2+2*x1^2*x19^2+x1*x6+2*x1^2*x6+2*x1*x19^2*x6+x1^2*x19^2*x6+x6^2+2*x1*x6^2+x1^2*x19*x6^2+x1*x19^2*x6^2+x1^2*x19^2*x6^2$$

$$\mathbf{f_7} = x15^2+x7+2*x15^2*x7+2*x7^2+x15*x7^2$$

$\mathbf{f_8} = x10^2+x13^2+2*x10^2*x13^2+x8+2*x10^2*x8+2*x13^2*x8+x10^2*x13^2*x8+2*x8^2+x10*x8^2+x13*x8^2+2*x10*x13*x8^2+x10^2*x13*x8^2+x10*x13^2*x8^2+x10^2*x13^2*x8^2$

$\mathbf{f_9} = x22^2+2*x10*x22^2+x10^2*x22^2+x9+2*x22^2*x9+x10*x22^2*x9+2*x10^2*x22^2*x9+2*x9^2+x22*x9^2+2*x10^2*x22*x9^2+2*x10*x22^2*x9^2+2*x10^2*x22^2*x9^2$

$\mathbf{f_{10}} = x9+2*x8^2*x9+2*x8*x9^2+2*x8^2*x9^2$

$\mathbf{f_{11}} = x1+2*x1^2*x12+2*x1*x12^2+2*x1^2*x12^2+x16+2*x1*x16+x1^2*x16+2*x12^2*x16+x1*x12^2*x16+2*x1^2*x12^2*x16+x1*x16^2+x1^2*x16^2+2*x12*x16^2+x1^2*x12*x16^2+2*x12^2*x16^2+2*x1*x12^2*x16^2+x21+2*x1*x21+x1^2*x21+2*x12^2*x21+x1*x12^2*x21+2*x1^2*x12^2*x21+2*x16*x21+x1*x16*x21+2*x1^2*x16*x21+x12^2*x16*x21+2*x1*x12^2*x16*x21+x1^2*x12^2*x16*x21+x16^2*x21+2*x1*x16^2*x21+x1^2*x16^2*x21+2*x12^2*x16^2*x21+x1*x12^2*x16^2*x21+2*x1^2*x12^2*x16^2*x21+x1*x21^2+x1^2*x21^2+2*x12*x21^2+x1^2*x12*x21^2+2*x12^2*x21^2+2*x1*x12^2*x21^2+x16*x21^2+2*x1*x16*x21^2+x1^2*x16*x21^2+2*x12^2*x16*x21^2+x1*x12^2*x16*x21^2+2*x1^2*x12^2*x16*x21^2+x16^2*x21^2+x1*x16^2*x21^2+x12*x16^2*x21^2+2*x1^2*x12*x16^2*x21^2+2*x1*x12^2*x16^2*x21^2+2*x1^2*x12^2*x16^2*x21^2$

$\mathbf{f_{12}} = x12+2*x12^2+x12^2*x13+x13^2+2*x12*x13^2$

$\mathbf{f_{13}} = 1+x13^2+2*x14+x13*x14+2*x13^2*x14+x14^2+2*x13*x14^2+x13^2*x14^2*x16+x14*x16^2+2*x13*x14*x16^2+x13^2*x14*x16^2+2*x14^2*x16^2+x13*x14^2*x16^2+x13^2*x14^2*x16^2+x13^2*x14^2*x18+2*x13^2*x14^2*x16*x18+x13^2*x14^2*x16^2*x18+x14*x18^2+2*x13*x14*x18^2+x13^2*x14*x18^2+2*x14^2*x18^2+x13*x14^2*x18^2+x13^2*x14^2*x18^2+x13^2*x14^2*x16*x18^2+2*x14*x16^2*x18^2+x13*x14*x16^2*x18^2+2*x13^2*x14*x16^2*x18^2+x14^2*x16^2*x18^2+2*x13*x14^2*x16^2*x18^2$

$\mathbf{f_{14}} = 1+2*x11+x11^2+x11*x14+2*x11^2*x14+2*x11*x14^2+x11^2*x14^2+x13*x14^2+2*x11^2*x13*x14^2+2*x13^2*x14^2+x11^2*x13^2*x14^2$

$\mathbf{f_{15}} = 1+x15^2+2*x8+x11^2*x8+x15*x8+2*x11^2*x15*x8+2*x15^2*x8+x11^2*x15^2*x8+x8^2+2*x11^2*x8^2+2*x15*x8^2+x11^2*x15*x8^2+x11*x15^2*x8^2+x11^2*x15^2*x8^2$

$\mathbf{f_{16}} = x15^2+x16+2*x15^2*x16+2*x16^2+x15*x16^2+x16^2*x17+2*x15*x16^2*x17+x15^2*x16^2*x17+x17^2+2*x15^2*x17^2+2*x16*x17^2+x15^2*x16*x17^2+x15*x16^2*x17^2+x15^2*x16^2*x17^2+2*x15^2*x20+x15^2*x16*x20+2*x15^2*x16^2*x20+2*x17^2*x20+x15^2*x17^2*x20+x16*x17^2*x20+2*x15^2*x16*x17^2*x20+2*x16^2*x17^2*x20+x15^2*x16^2*x17^2*x20+x15^2*x20^2+2*x15^2*x16*x20^2+2*x15*x16^2*x20^2+2*x15^2*x16^2*x20^2+2*x16^2*x17*x20^2+x15*x16^2*x17*x20^2+2*x15^2*x16^2*x17*x20^2+x17^2*x20^2+2*x15^2*x17^2*x20^2+2*x16*x17^2*x20^2+x15^2*x16*x17^2*x20^2+2*x16^2*x17^2*x20^2+2*x15*x16^2*x17^2*x20^2$

$\mathbf{f_{17}} = 1+x17^2+2*x18+x17*x18+2*x17^2*x18+x18^2+2*x17*x18^2+x17^2*x18^2*x21+x18*x21^2+2*x17*x18*x21^2+x17^2*x18*x21^2+2*x18^2*x21^2+x17*x18^2*x21^2+x17^2*x18^2*x21^2$

$\mathbf{f_{18}} = x16^2 + x18 + 2 * x16^2 * x18 + 2 * x18^2 + x16 * x18^2$
$\mathbf{f_{19}} = x18^2 + x19 + 2 * x18^2 * x19 + 2 * x19^2 + x18 * x19^2$
$\mathbf{f_{20}} = x20 + 2 * x20^2 + x20^2 * x21 + x21^2 + 2 * x20 * x21^2$
$\mathbf{f_{21}} = x11^2 + x21 + 2 * x11^2 * x21 + 2 * x21^2 + x11 * x21^2$
$\mathbf{f_{22}} = x23 + 2 * x10^2 * x23 + 2 * x10 * x23^2 + 2 * x10^2 * x23^2 + 2 * x23^2 * x24 + x10 * x23^2 * x24 + 2 * x10^2 * x23^2 * x24 + 2 * x23 * x24^2 + x10^2 * x23 * x24^2 + 2 * x23^2 * x24^2 + 2 * x10 * x23^2 * x24^2$
$\mathbf{f_{23}} = 1 + 2 * x22 + x22^2 + x22 * x23 + 2 * x22^2 * x23 + x23^2 + 2 * x22 * x23^2$
$\mathbf{f_{24}} = x22^2 + x24 + 2 * x22^2 * x24 + 2 * x24^2 + x22 * x24^2$

For the Mfrn k/o simulation, $\mathbf{f_{23}} = 0$. For the IRP2 o/e simulation, $\mathbf{f_6} = 2$. For the breast cancer simulation (Ras o/e and no Antioxidant enxymes (x_{12}) in ROS regulation), $\mathbf{f_{16}} = 2$ and $\mathbf{f_{11}} = x_1 + x_{21} + 2 * x_1 * x_{21} + x_1^2 * x_{21} + x_1 * x_{21}^2 + x_1^2 * x_{21}^2$.