The Landscape of Host Transcriptional Response Programs Commonly Perturbed by Infectious Pathogens: Towards Host-Oriented Broad-Spectrum Drug Targets

Yared Habteselassie Kidane

Dissertation 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

Genetics, Bioinformatics, and Computational Biology

T. M. Murali, Co-Chair

Christopher B. Lawrence, Co-Chair

David R. Bevan

Josep Bassaganya-Riera

Stephen B. Melville

March 27, 2012

Blacksburg, Virginia

Keywords: host-oriented broad-spectrum drug targets, host transcriptional responses, immunomodulation, infectious diseases, drug resistance

The Landscape of Host Transcriptional Response Programs Commonly Perturbed by Infectious Pathogens: Towards Host-Oriented Broad-Spectrum Drug Targets

Yared Habteselassie Kidane

ABSTRACT

The threat from infectious diseases dates as far back as prehistoric times. Pathogens continue to pose serious challenges to human health. The emergence and spread of diseases such as HIV/AIDS, Severe Acute Respiratory Syndrome (SARS), avian influenza, and the threats of bioterrorism have made infectious diseases major public health concerns. Despite many successes in the discovery of anti-infective medications, the treatment of infectious diseases faces serious challenges, which include (i) the emergence and reemergence of infectious pathogens, (ii) the ability of pathogens to adapt and develop resistance to drugs, and (ii) a shortage in the development and discovery of new anti-infective drugs.

Host-Oriented Broad-Spectrum (HOBS) treatments have the promising potential to alleviate these problems. The HOBS treatment paradigm focuses on finding drug targets in human host that are simultaneously effective against a wide variety of infectious agents and toxins. In this dissertation, we present a computational approach to predict HOBS treatments by integrative analysis of three types of data, namely, (a) gene expression data representing host responses upon infection by a pathogen, (b) annotations of genes to pre-defined biological pathways and processes, and (iii) genes that are targets of known drugs. Our methods combine gene set-level enrichment with biclustering.

We applied our approach to a compendium of gene expression data sets derived from host cells exposed to bacterial or to fungal pathogens, to functional annotation data from multiple databases, and to drug targets from DrugBank. We present putative host drug targets and drugs with extensive support in the literature for their potential to treat multiple bacterial and fungal infections. These results showcase the potential of our computational approach to predict HOBS drug targets that may be effective against two or more pathogens.

Our study takes a clean-slate approach that promises to yield unsuspected or unknown associations between pathogens and biological processes, and thus discern candidate gene/proteins to be further probed as HOBS targets. Furthermore, by focusing on host responses to pathogens as captured by transcriptional data, our proposed approach stimulates host-oriented drug target identification, which has potential to alleviate the problem of drug resistance.

This work was supported by the Initiative for Maximizing Student Development Program at Virgina Tech (VT-IMSD), the graduate school, the GBCB Interdisciplinary Ph.D. Program of Virginia Tech, and the Southern Regional Education Board.

Acknowledgments

First of all, I am indebted to my Orthodox Tewahedo religion for teaching me the beginning of wisdom, the fear of the Lord. My faith has been a continual source of perseverance throughout the years in graduate school. I am also indebted to my parents, my late father Habteselassie Kidane and my mother Muluemebet Yegletu. I am thankful to them for instilling in me the values of a good work ethic and discipline. I am grateful to my dearest fiancé, Meaza Yerdaw, for her tremendous support. Meaza not only encouraged me to re-join graduate school and fulfill my dream of obtaining a Ph.D., but also she did not complain (except a few times, of course), when we had to put off other plans for the love of this game. Also, I would like to thank my brothers and sisters for encouraging me all these years.

I would like to wholeheartedly thank my advisors Prof. T.M. Murali and Prof. Christopher Lawrence not only for their guidance and mentorship in my research but also for the financial support that I received during my final years at Virginia Tech. I am grateful to my committee members, Prof. Josep Bassaganya-Riera, Prof. David Bevan, Prof. Stephen

Melville for their support. I would like to thank them especially for writing me letters of support when I applied for a dissertation fellowship at the Southern Regional Education Board in spring of 2011.

I would like to acknowledge the help I received from my friend Mr. Albert Kwansa when I was preparing for my second GBCB seminar in the spring of 2012. I also enjoyed his company over the years. I would like to appreciate Mrs. Dennie Munson for being such a wonderful program liaison. I am grateful for the excellent support that I received from members of the Murali and Lawrence research groups. I would especially like to acknowledge Dr. Christopher Lasher, Mr. Christopher Poirel, and Mr. Ha Dang for their help.

Finally, I would like to extend my sincere gratitude to the following organizations for their financial support: (1) Virginia Tech's Initiative for Maximizing Student Development (VT-IMSD) Program, (2) the graduate school and the GBCB Interdisciplinary Ph.D. Program at Virginia Tech, and (3) the Southern Regional Education Board.

Dedication

This work is dedicated to my father Habteselassie Kidane who was my role model in many ways and my brother Melaku Habteselassie who first inspired me to pursue science.

Contents

List of Abbreviations

CORUM	Comprehensive Resource of Mammalian protein complexes
GEO	Gene Expression Omnibus
GSEA	Gene Set Enrichment Analysis
HOBS	Host-Oriented Broad-Spectrum
MSigDB	Molecular Signatures Database
NCBI	National Center for Biotechnology Information
NCI PID	National Cancer Institute Pathway Interaction Database

List of Figures

List of Tables

Chapter 1

Introduction

The threat from infectious diseases dates as far back as prehistoric times. Pathogens continue to pose serious challenges to human health. Despite many successes in the discovery of anti-infective drugs, infectious diseases continue to remain a major cause of death both in the United States and around the world. For instance, the death toll from infectious diseases accounted for more than 25% of the annual death worldwide (Figure 1.1). Infectious diseases such as respiratory infections, HIV/AIDS, tuberculosis, and malaria are among the top deadly diseases [1]. Although there is an increasing demand for novel anti-infective therapeutics, this area of drug discovery and development faces serious challenges. These challenges are associated mainly with (i) the diversity of disease causing agents, (ii) the reduction in the effectiveness of anti-infective therapeutics due to a high rate of mutation in pathogens, and (ii) the decline in the development and discovery of anti-infectives. Below we elaborate upon these challenges.

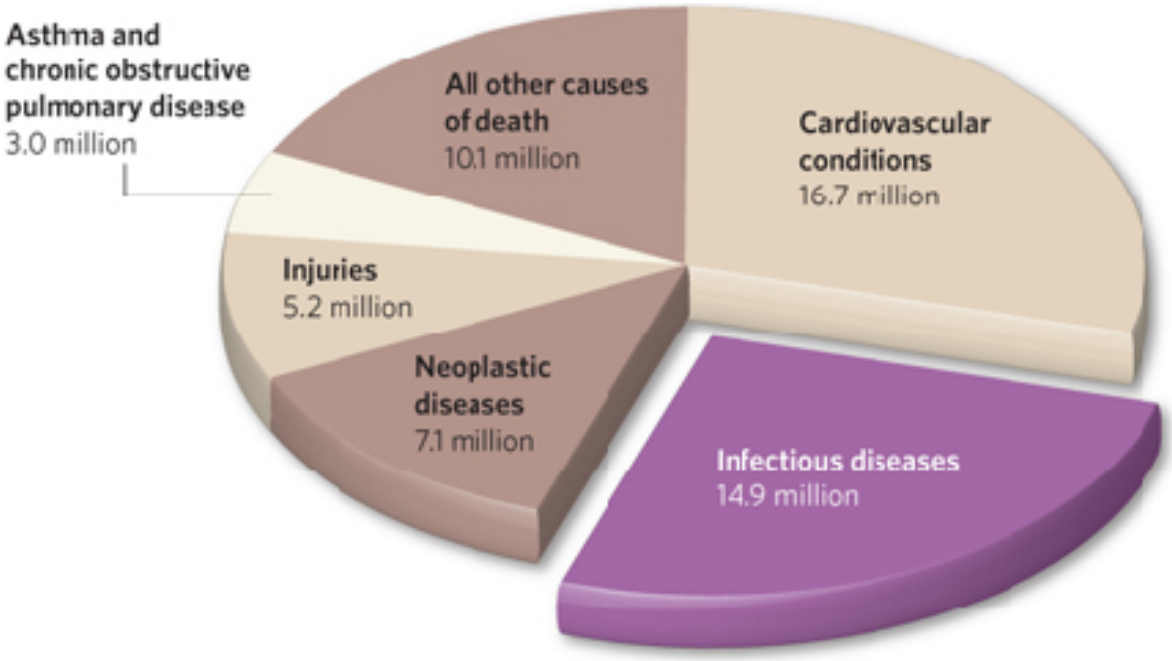

Figure 1.1: Leading causes of death worldwide. Reprinted by permission from Macmillan Publishers Ltd: Nature [1], copyright (2004)

1.1 Challenges in the treatment of infectious diseases

The one-bug-one-drug approach is not economically scalable. The "one-bug-one-drug" approach is a drug discovery and development strategy that refers to the development of a treatment regimen for each disease causing organism separately. There are as many as 1,407 recognized species of human pathogens: 538 (38%) are bacteria, 208 (15%) are viruses, 317 (23%) are fungi, and the remaining 344 (24%) are protozoa and helminths [2] (Figure 1.2). Furthermore, new pathogens are being discovered regularly e.g., the pandemic swine flu H1N1 virus recognized in 2009 [3]. Due to the diversity of infectious pathogens, tackling each pathogen individually has become practically impossible. Moreover, the discovery and development of a new drug is a lengthy process. On average, it

costs $1 billion to bring a drug to market [4]. The hight cost associated with drug discovery has made the "one-bug-one-drug" approach economically infeasible [5].

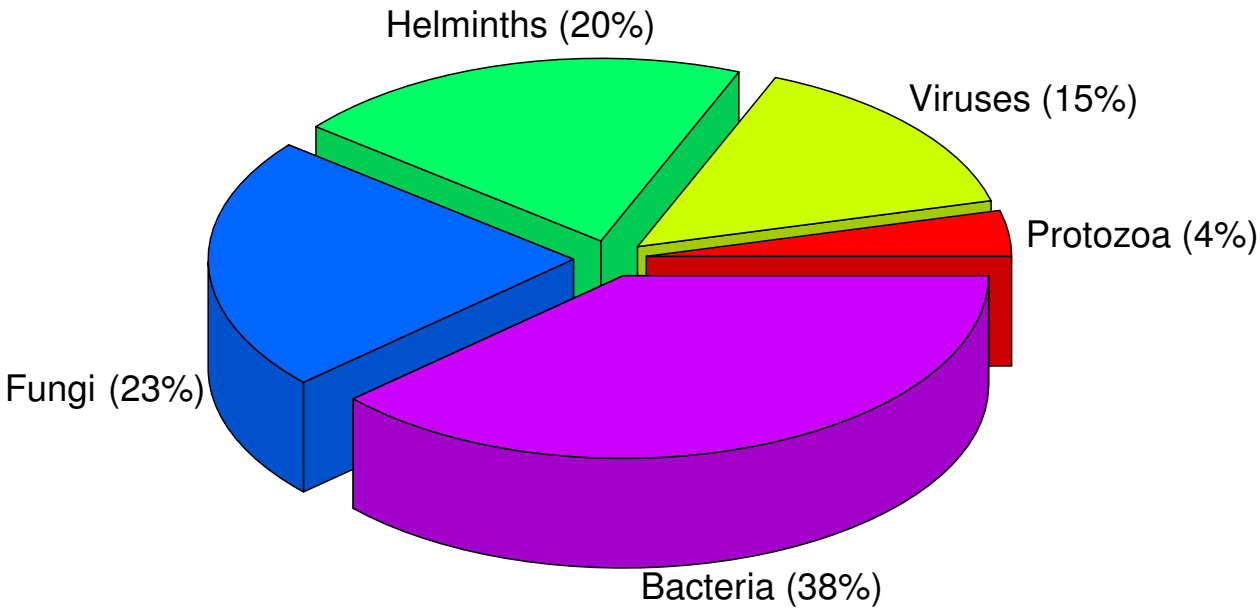

Figure 1.2: Major groups of human pathogens and their proportion. Chart created using data from Woolhouse and Gowtage-Sequeria [2].

Pathogens are becoming increasingly drug resistant. Infectious agents rapidly mutate and become resistant to drugs. For instance, multidrug-resistant strains of *Mycobacterium tuberculosis*, *Salmonella*, multidrug-resistant *Candida glabrata*, azole-resistant central nervous-system *Aspergillus fumigatus* have emerged in the past few years [6,7]. More than half of the total number of deaths due to bacterial infections are caused by pathogens that are resistant to commonly used antibiotics. Moreover, drug resistant pathogens have the ability to spread rapidly [8]. The conventional approach of targeting pathogen enzymes and proteins has accelerated the spread of resistance, resulting in the re-occurrence of in-

fectious diseases [9]. For instance, Figure 1.3 shows the different drugs that *Staphylococcus aureus* have developed resistant to in the time period from 1940 to 1970. We see that this pathogen became resistant to seven drugs in these years, giving it the name multi-drug-resistant *Staphylococcus aureus* (MRSA). In particular, *Staphylococcus aureus* developed resistance to Streptomycin in the same year this drug was introduced, indicating how some pathogens quickly adapt to drugs.

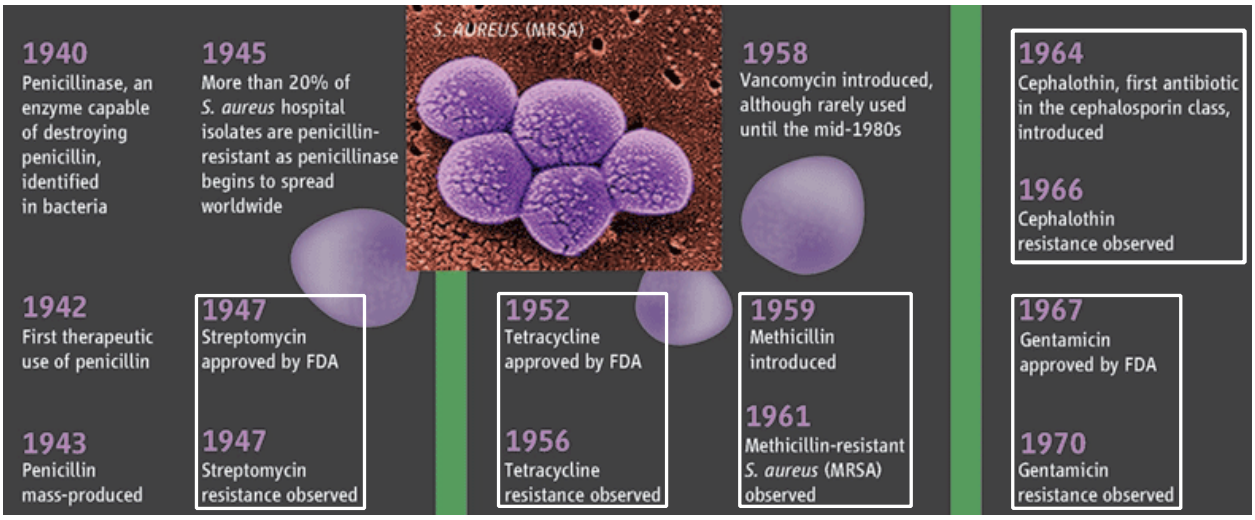

Figure 1.3: The different drugs that *Staphylococcus aureus* have developed resistant to in the time period from 1940 to 1970. From Taubes G (2008) The bacteria fight back. Science 321: 356361. Reproduced with permission from AAAS.

Anti-infective drugs are not adequately available. Drugs for infectious diseases are less profitable than other types of drugs. This trend is primarily attributed to the fact that anti-infective drugs have short-term use in comparison to other types of drugs [10]. As a result, major pharmaceutical companies are either decreasing or abandoning anti-infective drug discovery and development. This trend is reflected by the continual de-

cline in the number of antibiotics approved by the Federal Food and Drug Administration (FDA) in the years from 1983-2007 (Figure 1.4) [8, 11]. This declining trend in the development of new anti-infectives is expected to continue in the future [12].

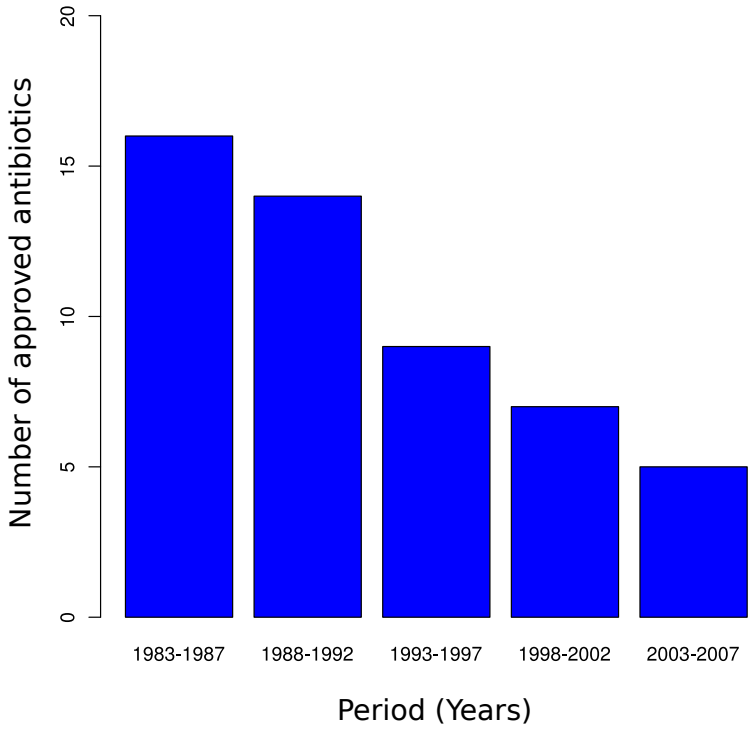

Figure 1.4: Trend in antibiotic discovery. Number of antibiotics approved by the Federal Food and Drug Administration in the time period from 1983-2007. Reproduced from the Lancet, Vol.156, Huh and Kwon, Nanoantibiotics: A new paradigm for treating infectious diseases using nanomaterials in the antibiotics resistant era, 128-145., Copyright (2011), with permission from Elsevier.

The task of meeting these challenges requires new ideas and techniques in the discovery and development of anti-infectives. Below we will discuss current paradigms that can potentially alleviate these problems.

1.2 Paradigms in the treatment of infectious diseases

Although the treatment of infectious diseases faces daunting challenges, there are many ongoing efforts that aim to decrease the health impact associated with these problems. Treatment paradigms that are specific to the problems discussed above include: the use of broad-spectrum drugs, host-oriented drug discovery approach, and repositioning known drugs (Figure 1.5). Below we describe these approaches in detail.

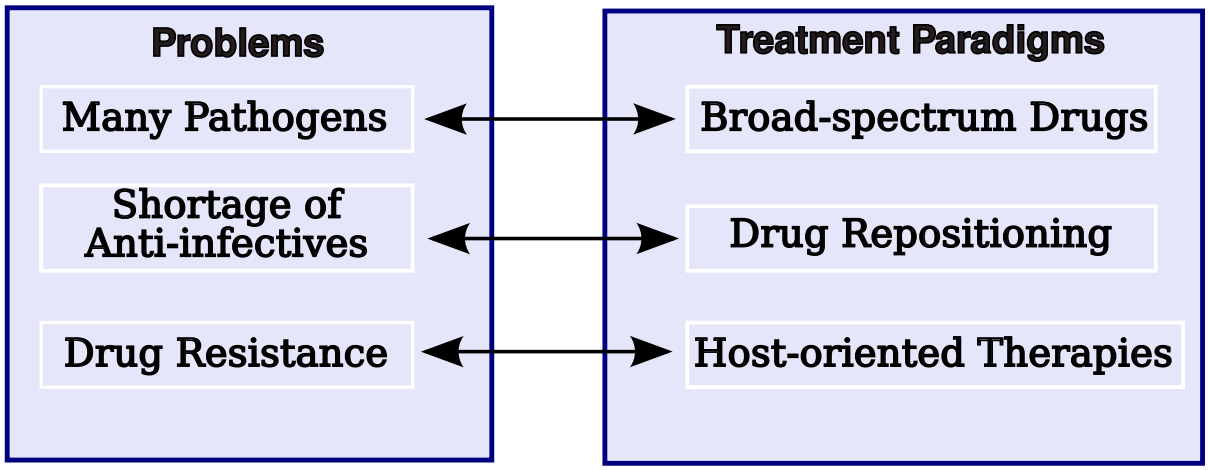

Figure 1.5: Current state of infectious diseases treatment, problems and treatment paradigms. Figure shows the three major problems surrounding the treatment of infectious diseases and corresponding paradigms.

Broad-spectrum drug discovery. The diversity of pathogens has necessitated the need for implementation of broad-spectrum interventions, i.e., develop anti-infectives that are effective against multiple infectious agents [5]. Alexander Fleming discovered the first broad-spectrum antibiotic, penicillin, in 1928. Most chemical scaffolds that are basis for

today's broad-spectrum agents were discovered between the mid-1930s and the early 1960s [13]. Sine then, much progress has not been made in discovering novel broad-spectrum agents. In fact, there were no novel classes of antibiotics discovered in the time period from 1962 to 2000. This time period is referred as the era of "innovation gap" in antibiotics discovery (Figure 1.6) [9]. In addition, the use of currently available broad-spectrum agents has been linked to the rise of drug resistance [14], which in turn calls for better broad-spectrum therapeutics, e.g., host-oriented broad-spectrum drugs.

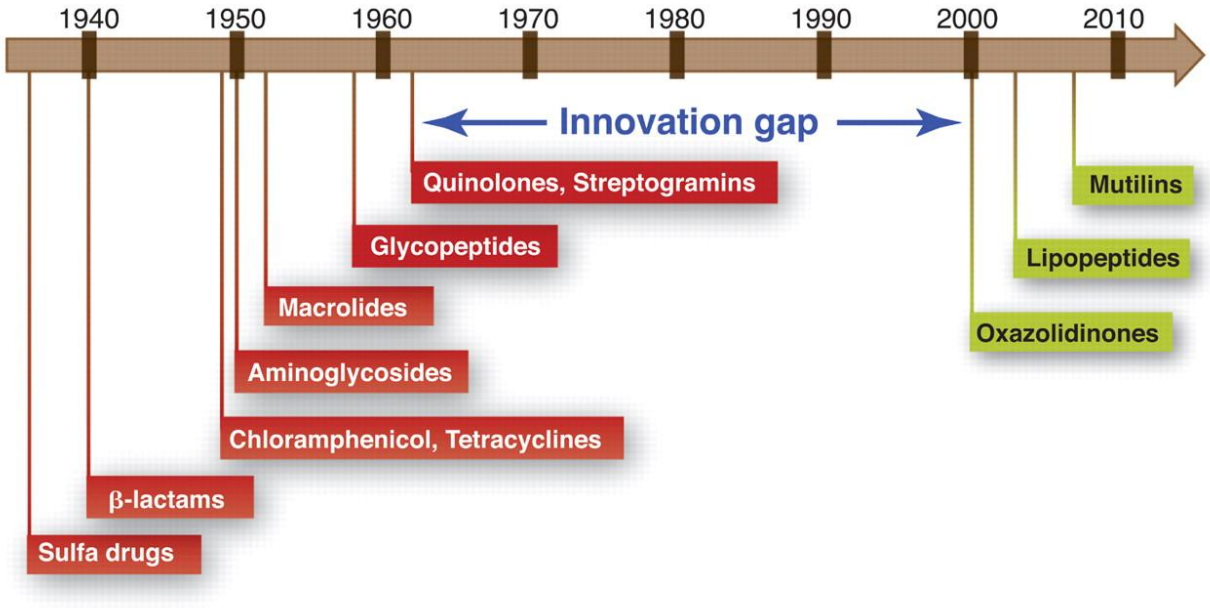

Figure 1.6: Innovation gap in antibiotics discovery. Classes of antibiotics and years they were discovered. Between the time period from 1962 to 2000, there were no major classes of antibiotics introduced [13]. From Fischbach MA, Walsh CT (2009) Antibiotics for emerging pathogens. Science (New York, NY) 325: 1089-1093. Reprinted with permission from AAAS.

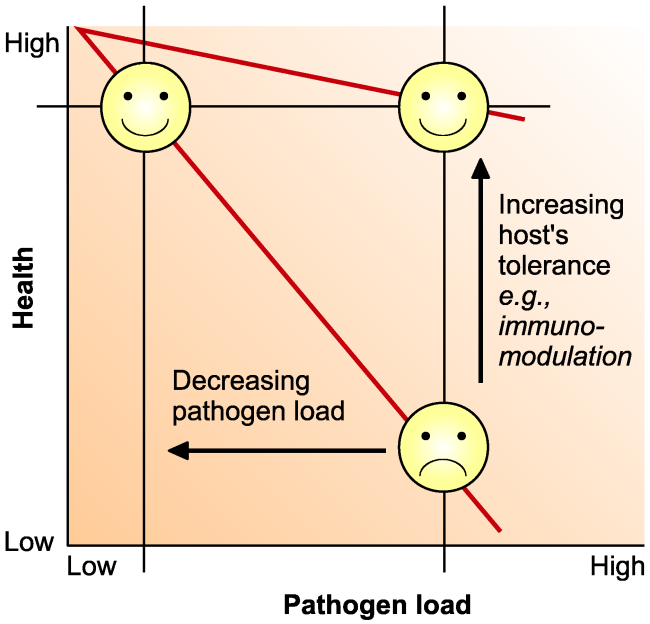

Figure 1.7: Two ways by which hosts can survive infections. An increase in the health of the host can be attained in two ways: by decreasing the pathogen load; or by increasing the hosts tolerance to pathogens without altering pathogen levels. Adapted by permission from Macmillan Publishers Ltd: Nature Reviews Immunology [15], copyright (2008)

Host-oriented treatments. There are two ways by which a host can survive infections: decreasing the pathogen load (the amount of disease causing organism) in the host and increasing the host's tolerance to damage that can be caused by pathogens [15]. The first approach kills and eradicates pathogens in the host, such methods are referred to as pathogen-oriented therapies. The second approach focuses on manipulating or subverting host genes, protein and pathways that are essential for the propagation of pathogen, but have a minimal effect on the normal functioning of the host system, i.e., host-oriented therapy (Figure 1.7). Pathogen-oriented approaches are prone to cause drug resistance due to the high rates of mutation associated with genes/proteins of pathogenic agents. In

an effort to combat the issue of drug resistance, anti-infective drug discovery is shifting to host-oriented therapeutics as host genes/proteins are relatively evolutionarily resilient. For instance, Mao *et al.* [16] identified novel host targets that are essential for the HIV viral life cycle but which can be safely perturbed without overt cytotoxicity to the host. Murali *et al.* computationally predicted HIV Dependency Factors (HDFs) by combining data from previous RNAi screens with protein-protein interaction networks [17]. HDFs are host gene/proteins that are critical for HIV replication but not lethal to the host when silenced [18]. Tseng and Perfect [19] conducted a survey of various strategies and goals of immunomodulators, which are components of the host's immune system that can be modulated in order to make the host more tolerant to a wide range of fungal infections (Figure 1.8). They grouped the strategies into four categories: (i) increase in the number of neutrophils, e.g., using granulocyte colony stimulating factor (G-CSF), (ii) activation of macrophages and dendritic cells, .e.g., using cytokines (iii) activation of the host's cellular immunity, e.g., using vaccines, and (iv) increasing host's humoral immunity by administrating antibodies.

Drug re-positioning. Due to the shortage of anti-infective drugs, there is a growing interest in discovering new uses (indications) for existing/known drugs, i.e., using a drug approved for one disease in the treatment of a new infectious disease. This process is also referred as drug repositioning or drug repurposing [20]. Repositioning an already approved/known drug offers several advantages over discovery and development of a new

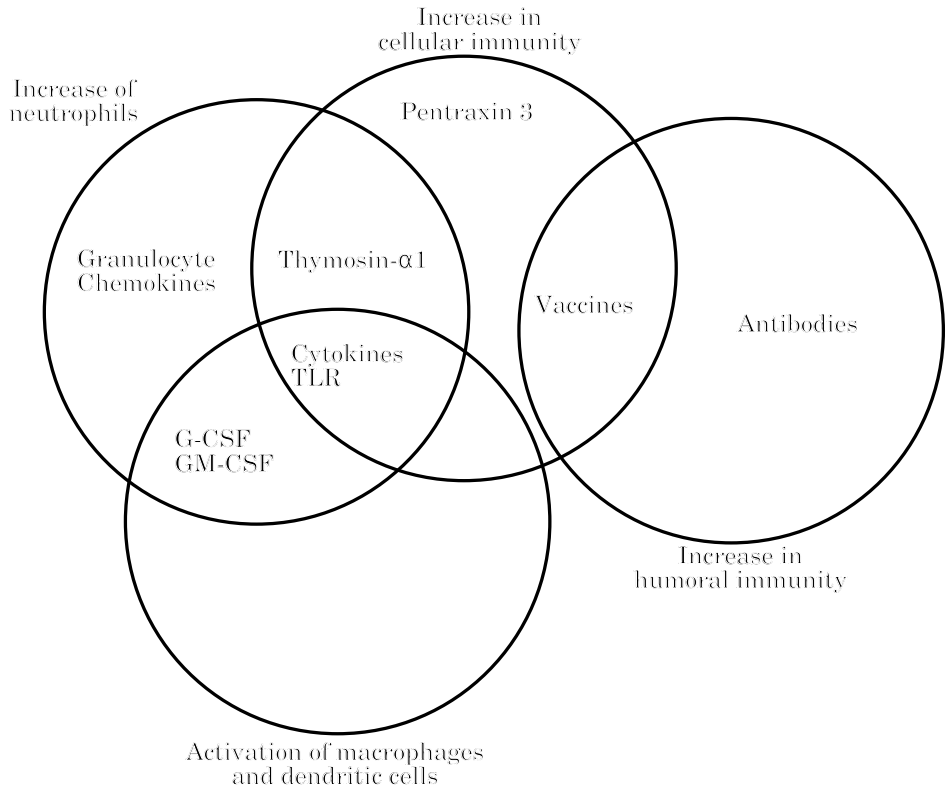

Figure 1.8: Goals and strategies underlying host-oriented therapies in fungal infections. Chart created using data from Tseng and Perfect [19]. The text outside the circles represent goals of host-oriented therapies and text inside the circles represent strategies to achieve those goals. G-CSF: Granulocyte colony stimulating factor; GM-CSF: Granulocyte-macrophae colony stimulating factor; TLR: Toll-like receptor

drug. Repositioned drugs are less costly, it takes more than 15 years and $1 billion to develop and bring a new drug to market [4]. Repositioned drugs are less likely to fail for reasons of toxicity as they already passed previous clinical tests. Dudley *et al.* [21] computationally predicted and experimental validated the use of topiramate in the treatment of inflammatory bowel disease (IBD). Topiramate has not been previously known/described to have efficacy for IBD or any related disorders of inflammation or the gastrointestinal tract. Studies have indicated the potential use of anticancer/antineoplastic agents such

as camptothecin derivatives for the treatment of fungal infections, due to the common eukaryotic heritage between humans and fungi [22,23].

1.3 Goals of this dissertation

The focus of this dissertation is to develop and use computational techniques for discovering drug targets in the host that are effective against multiple infectious pathogens. We term such drug targets as *Host-Oriented Broad-Spectrum (HOBS)* drug targets. We aim at developing a method that enables us to combine the three treatment paradigms discussed earlier in order to predict HOBS drug targets (Figure 1.5). To this end, we formulated three objectives that correspond to each of these paradigms, namely, host oriented approach (Objective 1), broad-spectrum strategy (Objective 2) , and drug repositioning (Objective 3).

Objective 1 *Identify pathways that are perturbed in the host upon infection by a pathogen; such pathways may contain target genes/proteins for host-oriented therapies.*

Objective 2 *Identify pathways that are perturbed by multiple pathogens; such pathways may contain target genes/proteins for HOBS therapies.*

Objective 3 *Identify known drugs that influence genes/proteins perturbed by pathogens; such drugs may serve as candidates in the treatment of these pathogens.*

The study of host-pathogen interaction has been useful in the identification of genes/proteins that are important for the survival of the host or critical in the infection process [16]. Ma-

nipulating these host response programs has a potential in countering pathogens either by inhibiting genes/proteins critical for the pathogen or inducing genes/proteins that would make the host more tolerant to infections (Figure 1.7). Previous findings such as this one are basis for *Objective 1*. An important component of *Objective 1* is the identification of host response programs that are involved in host-pathogen interaction. The advent of high-throughput technologies has created a favorable environment in the study of host responses to different diseases/pathogen types by providing a wide range of data types such as gene expression, metabolites, and protein-protein interactions. Gene expression data sets are among the most widely available data types. For instance, at the time of this writing, the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) houses over 700 thousand experimental DNA microarray samples. Therefore, in this study we decided to take advantage of the deluge of gene expression data that is currently available in publicly accessible repositories. We aimed at collecting and using DNA microarray gene expression data that pertains to host transcriptional responses upon infection by pathogens. Using this data and a gene set enrichment analysis (GSEA) [24] we identified host response programs at the level of biological processes.

Objective 2 is based on the premises that diseases that have a high degree of transcriptional similarity may be treated with a similar drug [25]. An important aspect of this objective is the identification of multi-way association among host response programs and pathogens. We asked if there are pathogens that commonly up- or down- regulate a subset of host response programs. In order to detect such patterns, we used a biclustering

technique (Section 2.5.5).

A common approach used in the initial phases of drug repositioning is to search for previously know drug(s) that target gene/protein of interest [26,27]. We envisioned a similar approach in the search of HOBS drugs and on the basis of which we formulated *Objective 3*. We integrate previously known drug/targets from publicly accessible databases such as DrugBank [28] and elicit drugs that target the common host response programs.

1.4 Contributions of this dissertation

This dissertation presents a computational approach to predict host-oriented broad spectrum drug targets by integrative analyses of three types of data, namely, (i) transcriptional data sets that capture host responses to various kinds of infectious agents, (ii) a compendium of gene sets that correspond to genes in a biological pathway, process, function, or members of a protein complex (collectively these were referred to as *gene sets*), and (iii) previously known drug targets. The three major contributions of this dissertation are:

(i) We combined gene set enrichment and biclustering techniques in order to develop a computational approach for detecting subsets of pathogens that commonly up- or down- regulated subsets of gene sets. Then, we integrated known drug target genes into our analysis, on the basis of which we predicted putative HOBS targets for specific group of infectious pathogens.

(ii) We applied this approach on a compendium of gene expression data representing 38 bacterial pathogens and pathogen strains. We identified 84 statistically significant associations among sub sets of pathogens and gene sets at a 0.05 p-value cutoff, after adjusting for multiple hypothesis testing. We predicted new uses of the drugs Anakinra, Etanercept, and Infliximab for gastrointestinal pathogens *Yersinia enterocolitica*, *Helicobacter pylori* kx2 strain, and enterohemorrhagic *Escherichia coli* and the drug Simvastatin for hematopoietic pathogen *Ehrlichia chaffeensis*.

(iii) We applied this approach in the prediction of common immunomodulators for the treatment of fungal infections. We analyzed transcriptional data sets from 5 fungal pathogens. We observed statistically significant associations between host responses to *Aspergillus fumigatus* and *Candida albicans*. Our analysis enabled the identification of consistently perturbed biological processes by these two pathogens containing both immune response inducing genes such as MALT1, SERPINE1, ICAM1, and IL8 and immune response repressing genes such as DUSP8, DUSP6 and SPRED2. We predicted that these genes belong to the pool of common immunomodulators for the treatment of aspergillosis and candidiasis, based on evidences found in the literature.

The computational approach that we developed in this dissertation may complement existing approaches that are developed to discover transcriptional responses common to many diseases. Unlike previous approaches, our method takes advantage of biclustering to detect gene set-specific relationships only among subsets of

pathogens. In addition, the predictions made in this dissertation might be utilized by drug target researchers to generate experimentally testable hypotheses in the discovery of broad spectrum drug targets for bacterial and fungal infectious.

1.5 Prior work

A principal component of our approach include detection of host response programs that are commonly perturbed by different pathogens. This involves discovery of similarities among gene expression profiles representing different pathogens. Below we will review some of the latest computational techniques in integrative microarray gene expression data sets analysis.

1.5.1 Detecting similarities in transcriptional profiles

Different computational approaches have been developed to discover similarities among gene expression profiles that correspond to different experimental variables such as microarray platforms, disease conditions, cell/tissue types, host species, development stages, and pathogen types. These methods have been useful in the discovery of disease signatures, disease-similarity network, drug re-purposing, prediction and classification of previously uncategorized samples.

For instance, Russ and Futschik [29] analyzed four transcriptional data sets that were

generated using three different microarray platforms in order to assess similarity among expression patterns of genes across different platforms. They started by selecting non-redundant genes that were common to all data sets. Then, they computed correlation coefficients among expression values of all genes within a data set for each of the four data sets. They later computed a second set of correlation coefficients between the correlation coefficients obtained earlier in order to determine the similarity of expression profiles between data sets representing different platforms. A large correlation of correlations indicated global similarity in the expression of genes from two platforms.

Hu and Agarwal [26] identified disease-similarity network from 7,000 publicly available transcriptional profiles. First, they obtained gene expression data sets from the NCBI Gene Expression Omnibus (GEO). Then, they computed differential gene expression profiles for each data set using a *t*-statistic. These authors defined similarity between gene expression profiles in two ways: (1) Using a correlation coefficient of *t*-statistic values of two disease profiles. A higher correlation coefficient correspond to a strong similarity between disease profiles. (2) Defining signature genes in one profile and measuring the distribution of these signature genes in the other ranked profile using a non parametric test such as the Gene Set Enrichment Analysis [24]. The distribution of signature genes at the top or bottom of the ranked profile indicates a strong similarity between disease profiles in comparison. Furthermore, Hu and Agarwal integrated drug molecular profiles obtained from Connectivity Map [30] in disease-similarity network they discovered. They created a drug-disease network based upon which they suggested new indications

for existing drugs.

Suthram *et al.* [27] performed integrated analysis of 54 disease-related gene expression profiles with protein-protein interaction network in order to detect protein functional modules that are commonly perturbed by diseases. They started with 4,620 pre-computed module of functionally interacting proteins. Then, they defined perturbation score of functional modules by disease as the mean normalized transcriptional activity of its component genes in the disease molecular profile. Then, they used hierarchical clustering of these scores to find disease-disease similarities. Finally, they integrated previously known drug targets in this functional modules perturbed by multiple pathogens based on which they predicted drug targets that can be effective against multiple diseases.

Jenner *et al.* [31] identified common host transcriptional responses from gene expression profiles. They started with gene expression profiles that correspond to 77 pathogens obtained from publicly accessible repositories. Then, they computed differential expression of genes in each data set using which they created a matrix representing the expression of genes across the different pathogens. Finally, they used hierarchical clustering technique in order to identify pathogens that have similar expression profiles.

Another study conducted by Dudley *et al.* [32] analyzed 429 disease associated transcriptional data sets, representing 238 diseases and 122 tissues in order to study if disease-related expression profiles are similar across tissue or experiment types. the authors started by computing a "disease state vector", which represent differential expression of all genes in each data set. Then, they computed the correlation between all pairs of disease

state vectors. They showed that the distribution of correlation coefficient representing the same disease in different tissues were higher than those that came from different diseases in same tissue type.

We noticed that the computational approaches described above tend to identify global similarities among expression profiles. Correlation and hierarchical clustering techniques were the two main ingredients in these approaches. Although these techniques were very useful in deriving generalized relationships among expression profiles they tend to obscure relationship that may exist over only a sub set of genes or biological conditions. Our approach is based on a biclustering technique that can detect expression profile similarities among pathogens across a subset of the genes or biological pathways.

Another common feature of the approaches discussed above is that they are based on differential expression values of individual genes, except in the case of Suthram *et al.* [27] where similarities in expression profiles were derived from perturbation scores of predefined protein functional modules. We argue that gene-based approach are less sensitive as they do not take the functional association among genes into account. In order to remedy this, in our approach, we characterized host response as perturbation of predefined gene sets which contains genes that annotate a biological process, pathway, or protein complexes.

1.6 Overview of Chapters

Chapter 2 describes a computational procedure to predict host-oriented broad-spectrum (HOBS) drug targets for bacterial pathogens. This work has been submitted to *BMC Systems Biology, 2012* and is currently under review. Chapter 3 describes computational techniques to predict immunomodulators that can act against multiple fungal pathogens, based on publicly available transcriptional data sets.

Chapter 2

The Landscape of Host Transcriptional Response Programs Commonly Perturbed by Bacterial Pathogens: Towards Host-Oriented Broad-Spectrum Drug Targets

2.1 Attribution

This chapter contains material originally submitted as Yared H Kidane, Christopher Lawrence, T. M. Murali (2012). The Landscape of Host Transcriptional Responses Commonly Perturbed by Bacterial Pathogens: Towards Host-Oriented Broad-Spectrum Drug Targets. *BMC Systems Biology*

2.2 Abstract

The emergence of drug-resistant pathogen strains and new infectious agents pose major challenges to public health. A promising approach to combat these problems is to target the host's genes or proteins, especially to discover targets that are effective against multiple pathogens, i.e., host-oriented broad-spectrum (HOBS) drug targets. An important first step in the discovery of such drug targets is the identification of host responses that are commonly perturbed by multiple pathogens.

In this paper, we present a methodology to identify common host responses elicited by multiple pathogens. First, we identified host responses perturbed by each pathogen individually using a gene set enrichment analysis of publicly available genome-wide transcriptional datasets. Then, we used a biclustering approach to identify host pathways and biological processes that were perturbed only by a subset of the analyzed pathogens. Finally, we tested the enrichment of each bicluster in human genes that are known drug targets, on the basis of which we elicited putative HOBS targets for specific groups of bacterial pathogens.

We identified 84 up-regulated and three down-regulated statistically significant biclusters. Each bicluster contained a group of pathogens that commonly dysregulated a group of biological processes. We validated our approach by checking whether these biclusters correspond to known hallmarks of bacterial infection. Indeed, these biclusters contained biological process such as inflammation, activation of dendritic cells, pro- and anti-

apoptotic responses and other innate immune responses. Next, we identified biclusters containing pathogens that infected the same tissue. Based on the drug targets contained in these biclusters, we predicted new uses of the drugs Anakinra, Etanercept, and Infliximab for gastrointestinal pathogens *Yersinia enterocolitica*, *Helicobacter pylori* kx2 strain, and enterohemorrhagic *Escherichia coli* and the drug Simvastatin for hematopoietic pathogen *Ehrlichia chaffeensis*.

The results from our study may be utilized by researchers to generate concrete hypothesis on which gene sets to probe further in their quest for HOBS drug targets for bacterial infections. All our results are available at the following supplementary website: http://bioinformatics.cs.vt.edu/∼murali/supplements/2011-kidane-bmc-systems-biology.

2.3 Introduction

Infectious diseases are the second leading cause of death worldwide, next to cardiovascular diseases [33]. Bacterial infections such as tuberculosis and food- and waterborne infections from *Salmonella enterica* and *Escherichia coli* still present many challenges to biomedical researchers. Foremost among these challenges is that infectious agents rapidly mutate and become resistant to drugs [6]. The conventional approach of targeting pathogen proteins has accelerated the spread of resistance, resulting in re-occurrence of infectious disease, such as those caused by multidrug-resistant strains of *Mycobacterium tuberculosis*, *Staphylococcus aureus*, and *Salmonella enterica* [34]. In an effort to combat the

issue of drug resistance, anti-infective drug discovery is shifting to a new approach that targets the host instead of pathogens. "Host-oriented" drug discovery focuses on manipulating or subverting biological processes in the host that pathogens utilize [35]. Another problem facing the treatment of infectious diseases is the increasing number of pathogenic agents [2]. Furthermore, new pathogens are appearing regularly, e.g., the pandemic swine flu H1N1 virus recognized in 2009. The expanding range of infectious agents coupled with the high cost associated with drug discovery have made it economically infeasible and practically impossible to tackle each pathogen individually. These factors have necessitated treatment regimens that are effective against a wide variety of infectious agents.

These factors have encouraged efforts in host-oriented broad-spectrum (HOBS) drug discovery, i.e., finding targets in the host that can simultaneously cure multiple infections [34, 36]. Examples of HOBS drugs currently available in the market include Statins and Isoprinosine. Statins are used in the treatment of *Leishmania, Staphylococcus aureus*, and HIV infections [37–39]. Statins lower the cholesterol level in human body. They are effective against pathogens that utilize cholesterol in binding and internalization to the host cell. Isoprinosine, which stimulates the proliferation of T-cells, is used in the treatment of *Herpes simplex*, *Hepatitis*, and *Epstein-Barr* virus infections [40].

A first and important step in HOBS drug discovery is the development of computational tools to discover common physiological processes and cellular pathways that different pathogens utilize to infect, proliferate, and spread in the host. We hypothesized that comprehensive molecular datasets of host responses to diverse varieties of pathogens

might form a powerful resource to discover such pathways. Transcriptional datasets that correspond to different infectious diseases, cell/tissue types, and organisms are the most abundantly available. Meta-analysis of transcriptional datasets have been performed for a wide range of diseases. For instance, Rhodes *et al.* [41] analyzed 40 cancer related microarray datasets to identify common signatures of cancer. English and Butte [42] integrated 49 obesity-related genome-wide experiments obtained from human, mouse, rat, and worm to predict new genes that may be associated with obesity. Magalhaes *et al.* [43] performed meta-analysis of 27 age-related gene expression profile datasets from human, mouse, and rat to reveal several common signatures of aging. Jenner *et al.* [31] used hierarchical clustering of gene expression profiles of 77 pathogens in order to find genes that exhibited similar expression profiles across several disease types.

Recent approaches have taken meta-analysis of DNA microarray datasets one step further by incorporating drug targets into the analysis and inferring new uses for existing drugs on the basis of disease similarities. The premise underlying these approaches is that diseases with a high degree of transcriptional similarity might be treated with similar drugs [25]. Hu *et al.* [26] discovered disease-disease links by using correlation coefficients and gene set enrichment analysis to measure the similarities between gene expression profiles of diseases. They also integrated gene expression profiles that pertain to responses of cell lines to drugs derived from the Connectivity Map [30] to create a drug-disease network where clusters of drugs and diseases suggested shared drug mechanisms and molecular disease pathology. Suthram *et al.* [27] performed integrative analysis of 54

disease-related mRNA expression datasets. They measured the perturbation of predefined protein functional modules using the mean normalized transcriptional activity of each module's component genes in the disease's transcriptional profile. Furthermore, they identified known drug targets in the modules that are perturbed by multiple disease types, which they proposed as pluripotent/broad-spectrum drug targets.

The goal of our work is similar to that of Jenner *et al.*, Hu *et al.*, and Suthram *et al.*: to discover transcriptional responses common to many diseases, specifically those caused by bacterial pathogens, and to discover existing drug targets within those transcriptional signatures. The previous authors have used global correlation measures to detect disease associations, which may obscure relationships that exist over only a subset of the diseases or genes. In contrast, we use a combination of gene set level enrichment and biclustering. This approach has helped us group sets of host genes that are dysregulated only by a subset of the pathogens, enabling us capture pathway-specific relationships among groups of pathogens.

2.4 Results and Discussion

We start with an overview of the method (Figure 2.1). We obtained genome-wide transcriptional data sets of host responses after infection by bacterial pathogens from the NCBI's Gene Expression Omnibus (GEO) (Figure 2.1A). After data filtering (see Methods), we retained 29 gene expression profiling studies which represent 213 host samples

and 38 bacterial pathogens or pathogen strains. We sub-divided the datasets into four major kinds of infection: gastrointestinal, oral cavity, hematopoietic, and respiratory. A complete description of these datasets along with the GEO accession numbers are provided on the supplementary website.

Since these datasets were generated by different research groups with different objectives in mind, they tended to be very diverse, e.g., in the microarray platform used, the infected host, and the tissue or cell type from which the gene expression measurements were taken. Such variations made the direct comparison of the datasets difficult. To alleviate this problem, we computed gene sets perturbed by each pathogen using Gene Set Enrichment Analysis (GSEA) (Figure 2.1B), thereby enabling comparison across pathogens at the level of perturbed gene sets. We recorded all pathogens and the gene sets they perturbed in a matrix. Next, we biclustered this matrix in order to identify all subsets of the gene sets that were co-perturbed across a subset of the pathogens (Figure 2.1C). We assessed the statistical significance of the biclusters by comparing their sizes to biclusters found in randomized matrices. This process yielded 84 up-regulated and three down-regulated significant biclusters at a 0.05 p-value cutoff, after adjusting for multiple-hypothesis testing [44]. In this paper, we focus our discussion on up-regulated biclusters as they are far greater in number than down-regulated biclusters. We used Fisher's exact test to estimate the enrichment of a bicluster in known drug targets (Figure 2.1D). Finally, we sought for support in the literature for biologically meaningful connections among the gene sets, pathogens, and drug targets in a bicluster (Figure 2.1E).

We organize the results from our study into two major sections. First, we asked if the biclusters we computed could reveal well-known immunological responses in the host to bacterial infection. To this end, we identified host gene sets that were contained in those biclusters that also included many pathogens. Our analysis revealed that biological functions pertaining to the up-regulation of inflammatory gene sets, Lipopolysaccharide (LPS)-inducible gene sets, innate immunity response, induction and inhibition of apoptosis, and maturation of dendritic cells are host responses that are triggered by most of the bacterial pathogens. Rediscovering well known host responses to infection established the validity of our approach in detecting common host signatures. Second, we analyzed the biclusters for putative HOBS targets. Out of the 84 significantly up-regulated biclusters, 47 of them were enriched in known drug targets at the 0.05 significance level. We identified seven biclusters where all the pathogens contained in each of these biclusters infected a particular tissue or organ in the human body. For instance, in bicluster 38, we found four gastrointestinal pathogens namely *Yersinia enterocolitica* wap and p60 strains, *Helicobacter pylori* kx2 strain, and Enterohemorrhagic *Escherichia coli*. From this bicluster, we suggested the potential use of chronic inflammation suppressor such as Anakinra, Etanercept and Infliximab in treating infection caused by these four pathogens.

2.4.1 Signature Gene Sets Perturbed in Response to Bacterial Infection

There are several stages and outcomes that are hallmarks of generalized infection. On one hand, pathogens try to enter, multiply, and spread in the host, causing disease. On

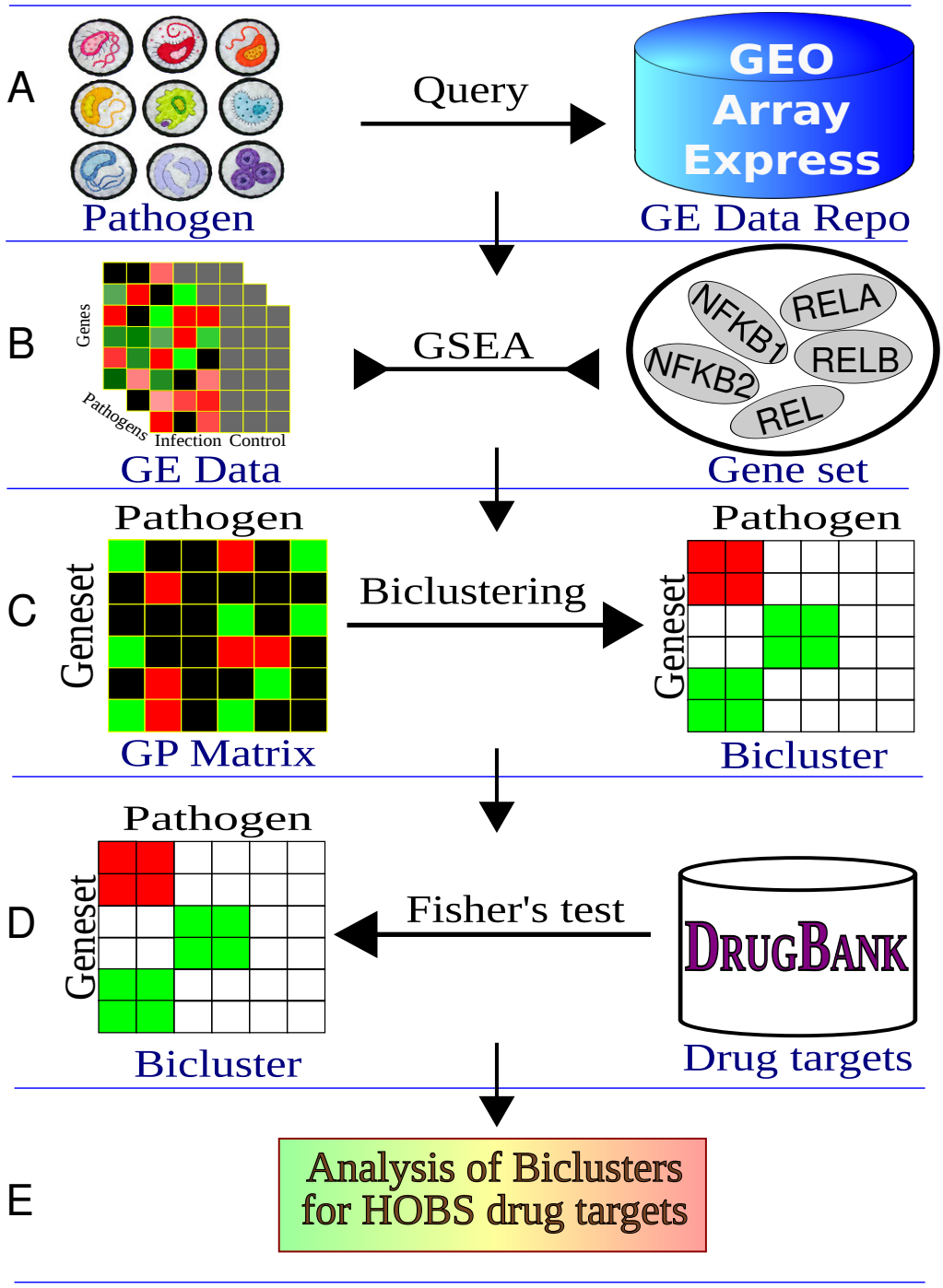

Figure 2.1: Overview of our computational system to compute host-oriented broad-spectrum drug targets. (A) Obtaining relevant collection of taxonomic names for human bacterial pathogens. Querying the GEO metadatabase in search of relevant transcriptional datasets. (B) Gene Set Enrichment Analysis of the transcriptional datasets collected in Step A. (C) Identification of pathogen-gene set biclusters and estimation of statistical significance of biclusters (D) Testing bicluster enrichment for known drug targets. (E) Literature analysis of putative HOBS drug targets contained in biclusters.

the other hand, hosts attempt to defend the attack from pathogens using processes conferring innate and adaptive immunity, leading to the elimination of pathogens. There are different strategies that are utilized by pathogens and by hosts to achieve these objectives. Among other things, pathogens induce or inhibit apoptosis, import their genetic material into the host, and replicate their genome [45, 46]. Hosts utilize various arms of the immune system such as inflammation, response to stimulus, maturation of dendritic cells and activation of various components of the innate immunity to lessen pathogenicity.

The 84 statistical significant up-regulated biclusters contained 1,364 distinct gene sets and 34 pathogens. To determine if our biclusters capture the hallmarks of infection mentioned above, we asked which gene sets belonged to the the largest number of biclusters. To compute these signature gene sets, we ranked the gene sets in decreasing order of number of biclusters they were perturbed in. We observed that the number of biclusters that a gene set was contained in had a high positive correlation (r=0.89) with the number of pathogens that perturb the gene set (Figure 2.2). Table 2.1 shows the top ten gene sets in this ranked list. Then, for each gene set, we assigned Gene Ontology (GO) biological processes for intuitive interpretation (Table 2.2) using the procedure described in Methods. We now proceed to discuss the signature gene sets we computed and correlate them to well-known hallmarks of infection.

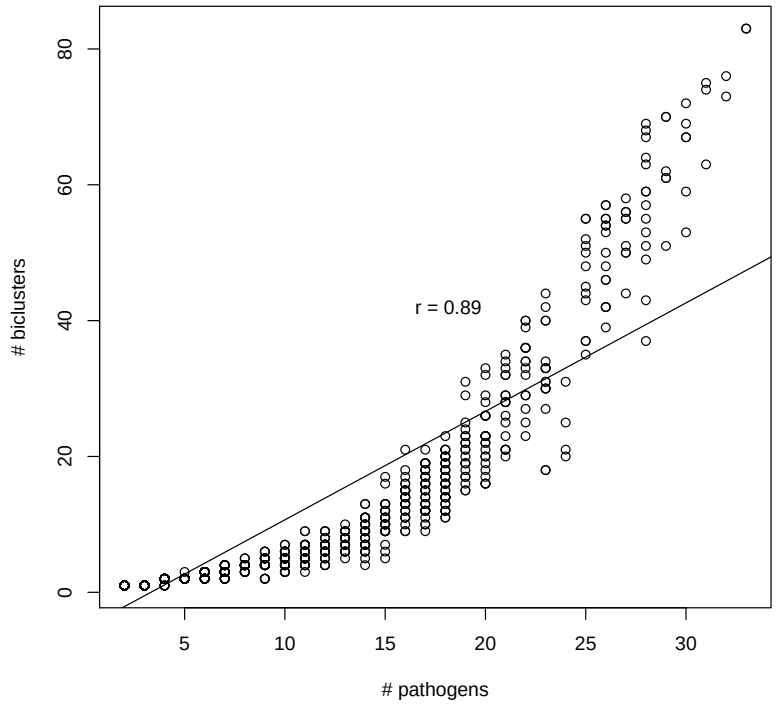

Figure 2.2: Correlation between number of pathogens that perturb gene set and the number of biclusters that contain the gene set. Plot indicates that number of pathogens perturbing a gene set are positively correlated with the number of biclusters a particular gene set appeared in.

Gene Set	# Pathogens	# Biclusters
ZHANG RESPONSE TO IKK INHIBITOR AND TNF UP	33	83
SEKI INFLAMMATORY RESPONSE LPS UP	33	83
DIRMEIER LMP1 RESPONSE EARLY	32	76
DAUER STAT3 TARGETS UP	31	75
HINATA NFKB TARGETS KERATINOCYTE UP	31	74
TIAN TNF SIGNALING VIA NFKB	32	73
LINDSTEDT DENDRITIC CELL MATURATION B	30	67
UZONYI RESPONSE TO LEUKOTRIENE AND THROMBIN	31	63
NETPATH IL 4 PATHWAY DOWN	30	59
MAHADEVAN RESPONSE TO MP470 UP	30	53

Table 2.1: Signature gene sets. For each gene set, the table shows the number of pathogens that perturb it and the number of biclusters it appears in.

Gene Set	GO Enriched Processes (Top Three)	p-value
ZHANG RESPONSE TO IKK INHIBITOR AND TNF UP	INFLAMMATORY RESPONSE	2.89×10^{-6}
	RESPONSE TO WOUNDING	1.28×10^{-4}
	DEFENSE RESPONSE	4.56×10^{-4}
SEKI INFLAMMATORY RESPONSE LPS UP	LOCOMOTORY BEHAVIOR	1.19×10^{-6}
	RESPONSE TO EXTERNAL STIMULUS	1.36×10^{-5}
	DEFENSE RESPONSE	6.49×10^{-5}
DIRMEIER LMP1 RESPONSE EARLY	APOPTOSIS GO	9.42×10^{-3}
	PROGRAMMED CELL DEATH	9.56×10^{-3}
	VIRAL GENOME REPLICATION	1.38×10^{-2}
DAUER STAT3 TARGETS UP	CYCLIC NUCLEOTIDE METABOLIC PROCESS	1.94×10^{-3}

Continued on next page...

Table 2.2 – Continued

Gene Set	GO Enriched Processes (Top Three)	p-value
	PROTEIN IMPORT INTO NUCLEUS TRANSLOCATION	1.94×10^{-3}
	DNA DAMAGE RESPONSE SIGNAL TRANSDUCTION RESULTING IN INDUCTION OF APOPTOSIS	3.66×10^{-3}
HINATA NFKB TARGETS KERATINOCYTE UP	RESPONSE TO WOUNDING	1.09×10^{-6}
	INFLAMMATORY RESPONSE	1.31×10^{-6}
	RESPONSE TO STRESS	3.63×10^{-5}
TIAN TNF SIGNALING VIA NFKB	DEFENSE RESPONSE	1.77×10^{-3}
	REGULATION OF I KAPPAB KINASE NF KAPPAB CASCADE	3.09×10^{-3}
	RESPONSE TO WOUNDING	3.4×10^{-3}
LINDSTEDT DENDRITIC CELL MATURATION B	APOPTOSIS GO	9.05×10^{-4}
	PROGRAMMED CELL DEATH	9.22×10^{-4}
	CELL DEVELOPMENT	2.47×10^{-3}
UZONYI RESPONSE TO LEUKOTRIENE AND THROMBIN	HEART DEVELOPMENT	1.72×10^{-2}
	INFLAMMATORY RESPONSE	3.19×10^{-2}
	REGULATION OF TRANSCRIPTION	3.26×10^{-2}
NETPATH IL 4 PATHWAY DOWN	ACTIVATION OF INNATE IMMUNE RESPONSE	5.27×10^{-6}
	PATTERN RECOGNITION RECEPTOR SIGNALING PATHWAY	5.27×10^{-6}
	TOLL-LIKE RECEPTOR SIGNALING PATHWAY	5.27×10^{-6}
MAHADEVAN RESPONSE TO MP470 UP	LOCOMOTORY BEHAVIOR	1.6×10^{-7}
	DEFENSE RESPONSE	3.62×10^{-7}

Continued on next page...

Table 2.2 – Continued

Gene Set	GO Enriched Processes (Top Three)	p-value
	INFLAMMATORY RESPONSE	1.01×10^{-6}

Table 2.2: Mapping of Gene Sets to GO Biological Processes. The table shows top three GO biological processes that have the highest overlap with each of the ten most frequently perturbed gene sets (in Table 2.1). The p-value indicates the statistical significance of the overlap, based on Fisher's exact test.

Inflammatory Response

Inflammation is one of the immediate reactions by the host against pathogenic infections. Of the top ten gene sets, four gene sets namely, "ZHANG RESPONSE TO IKK INHIBITOR AND TNF UP", "UZONYI RESPONSE TO LEUKOTRIENE AND THROMBIN", "HINATA NFKB TARGETS KERATINOCYTE UP", and "MAHADEVAN RESPONSE TO MP470 UP" have a high overlap with genes annotated with GO's inflammatory response process (GO:0006954). For each of these gene sets, we describe the experiment that generated it. We note that these experiments were conducted in diverse tissues and were not directly related to pathogen infection. Nevertheless, by examining the connection between each of these gene sets and inflammation, we demonstrate that inflammation is a non-specific response triggered by many of the pathogens irrespective of the type of cell being infected. The gene set "ZHANG RESPONSE TO IKK INHIBITOR AND TNF UP" is perturbed in 83 biclusters spanning 33 different bacterial pathogens. This

gene set contains 219 genes that are up-regulated in BxPC3 pancreatic cancer cells after treatment with tumor necrosis factor (TNF)-α, a pro-inflammatory cytokine [47]. This gene set consists of genes encoding for pro-inflammatory mediators such as IL1A, IL1B, TNFSF10 and a number of other chemokines including CCL20, CCL5, CXCL1, CXCL10, CXCL11, CXCL16, CXCL2, and CXCL3. The next set in the list is "HINATA NFKB TARGETS KERATINOCYTE UP", which was perturbed by 31 pathogens and appeared in 74 biclusters. This gene set contains 71 genes that were up-regulated in primary keratinocyte cells after transduction with NF-kappa B [48]. The majority of the genes in this gene set are cytokines and growth factor genes including chemokines (CCL20, CCL5, CXCL10, CXCL11, CXCL3, CXCL6); interleukins (IL15, IL1B, IL1RN, IL6, IL8); and growth factor genes (TNC, VEGFA, ESM1, MP2). The "UZONYI RESPONSE TO LEUKOTRIENE AND THROMBIN" gene set is perturbed by the same number of pathogens as "HINATA NFKB TARGETS KERATINOCYTE UP". It contains 37 genes that were up-regulated in Human Umbilical Vein Endothelial Cells (HUVEC) after stimulation with leukotriene LTD4, a leukocyte produced at sites of inflammation [49]. The fourth gene set is "MAHADEVAN RESPONSE TO MP470 UP", which is perturbed by 30 pathogens and appeared in 53 biclusters. This gene set contains 19 genes that were up-regulated in gastrointestinal stromal tumor cell-line after treatment with protein-kinase inhibitor drug (MP470) [50]. This gene set also contains chemokines and proinflammatory cytokines such as CCL5, CXCL1, CXCL10, CXCL3, CXCL5, CXCL6, IL8, and IL6.

Activation of Innate Immunity

In addition to inflammation, innate immunity also involves the activation of anatomical barriers, mechanical removal of antigens, pattern-recognition receptors, complement pathways, and phagocytosis. Previously we have shown that inflammation is a biological process most widely perturbed by the bacterial pathogens we analyzed. The "NETPATH IL 4 PATHWAY DOWN" gene set (which contains 90 genes that are supposed to be transcriptionally down-regulated by the activation of IL4 pathway) is among the top ten most perturbed gene sets. It is perturbed by 30 pathogens and is implicated in 59 biclusters. This gene set has a high overlap with three GO biological process namely "ACTIVATION OF INNATE IMMUNE RESPONSE", "PATTERN RECOGNITION RECEPTOR SIGNALING PATHWAY", and "TOLL-LIKE RECEPTOR SIGNALING PATHWAY". The perturbation of this gene set indicated that in addition to inflammation, other components of the innate immunity process are also perturbed by multiple bacterial pathogens.

Maturation of Dendritic Cells

Dendritic cells have the ability to develop from immature antigen-capturing cells to more specialized antigen-presenting cells. The maturation of dendritic cells is a very important aspect of the host response to bacterial infection. This step indicates the stimulation of various cytokines, chemokines, and other co-stimulatory molecules that are necessary for the onset of adaptive immunity [51]. A number of factors drive the maturation of den-

dritic cells including the type of antigen (e.g., lipopolysaccharide) and the presence of inflammatory cytokines (e.g., IL-1 and TNF-alpha). In our study, we found that the "LINDSTEDT DENDRITIC CELL MATURATION A" gene set was perturbed by 30 pathogens and implicated in 67 biclusters. This gene set contains 54 genes that are up-regulated in a transcriptional study involving stimulation of human monocyte-derived dendritic cells with inflammatory stimuli, consisting of tumor necrosis factor (TNF)-α and IL-1β [52].

Induction and Inhibition of Apoptosis

Induction and inhibition of apoptosis are important mechanisms of bacterial pathogenesis [45]. The "DIRMEIER LMP1 RESPONSE EARLY" gene set, which has a high overlap with GO's "APOPTOSIS GO" (GO:0006915) and "PROGRAMMED CELL DEATH" (GO:0012501) biological processes is the second most highly perturbed gene set across the significant biclusters. It is perturbed by 32 pathogens spanning 76 biclusters. This gene set contains 54 genes that are dysregulated in B lymphocyte cells after induction of LMP1, an oncogene. This gene set contains both pro- and antiapoptotic genes whose balance permitted survival of B lymphocyte cells [53]. Perturbation of the "DIRMEIER LMP1 RESPONSE EARLY" gene set by most of the pathogens we analyzed indicated that genes with opposing activities involved in cell survival were up-regulated during bacterial infection. This gene set contains tumor suppressors (KLF6, TNFAIP3), oncogenes (BIRC3, CXCR7, HERPUD1, HSP90AB1, LCP1, MYC, NFKB2), cell differentiation markers (CD69, CD83, ICAM1, SLAMF1), and growth markers (LTA, NPPB, TNFSF9).

Response to Lipopolysaccharide Stimulation

The host responds in a variety of ways against internal or external stimuli. An example of an external stimulus is a lipopolysaccharide (LPS). LPS is a molecule found on the outer membrane of Gram-negative bacteria. It triggers the expression of a number of signaling molecules, pro-inflammatory cytokines, and antibacterial genes when interacting with the Toll-like receptor of the host cell [54]. The "SEKI INFLAMMATORY RESPONSE LPS UP" gene set [55,56] contains genes that were up regulated in hepatic stellate cells of the mouse after stimulation with bacterial lipopolysaccharide. This gene set is up-regulated in as many as 83 biclusters (similar to "ZHANG RESPONSE TO IKK INHIBITOR AND TNF UP" gene set) indicating that, like inflammatory response, genes related to LPS stimulation are predominantly perturbed across a significant number of pathogens.

We expected this gene set would be perturbed only by Gram-negative bacteria, as LPS is a characteristic of these bacteria [54]. However, we observed that 30% of the pathogens that up-regulated this gene set are Gram-positive. Table 2.3 shows 20 distinct pathogens (without counting strains of the same pathogen) that up-regulated the "SEKI INFLAMMATORY RESPONSE LPS UP" gene set. Six of these pathogens are Gram-positive, namely *Streptococcus pneumoniae*, *Listeria monocytogenes*, *Bifidobacterium bifidum*, *Streptococcus pyogenes*, *Lactobacillus acidophilus*, and *Bacillus anthracis*. We noted that this gene set has a significant overlap with genes annotated with the GO biological process"RESPONSE TO EXTERNAL STIMULUS" (GO:0009605). This biological process represents the cells's re-

sponse to external stimuli. Of the 83 genes annotated to this GO term, 14 genes also belong to "SEKI INFLAMMATORY RESPONSE LPS UP" gene set (p-value 1.36×10^{-5}). This high degree of overlap suggests that many genes that respond to LPS may belong to a broader class of genes that are perturbed by any external stimulus, including a pathogenic bacterium. This possibility may explain our finding that many Gram-positive bacteria perturb the gene set "SEKI INFLAMMATORY RESPONSE LPS UP".

Pathogen Name	q-value
Gram-negative bacteria	
Aeromonas cavia	2.00×10^{-5}
Aggregatibacter actinomycetemcomitans	6.86×10^{-3}
Brucella melitensis	1.00×10^{-4}
Brucella neotomae	4.10×10^{-4}
Brucella ovis	7.60×10^{-4}
Burkholderia pseudomallei	1.26×10^{-2}
Ehrlichia chaffeensis	0
Escherichia coli	0
Helicobacter pylori	0
Mycobacterium tuberculosis	0
Porphyromonas gingivalis	1.25×10^{-3}
Pseudomonas aeruginosa	0
Shigella dysenteriae	0
*Yersinia enterocolitica*6	0
Gram-positive bacteria	

Continued on next page...

Table 2.3 – Continued

Pathogen Name	q-value
Bacillus anthracis	0
Bifidobacterium bifidum	0
Lactobacillus acidophilus	0
Listeria monocytogenes	0
Streptococcus gordonii	1.65×10^{-1}
Streptococcus pneumoniae	0
Streptococcus pyogenes	0

Table 2.3: Pathogens that perturb the SEKI_INFLAMMATORY_RESPONSE_LPS_UP gene set. The second column contains the q-values as well as a color indicating the magnitude of the q-value. Figure 2.3 contains the legend mapping q-values to colors. All pathogens up-regulate this gene set, except *Streptococcus gordonii*; which down-regulate it.

2.4.2 Putative HOBS Drug Targets

We now turn our attention to discovering potential HOBS drug targets in our biclusters. To this end, we further filtered the 84 significant biclusters based on the type of infection caused by the pathogens they contained. Table 2.4 shows biclusters that contained pathogens that cause a single kind of infection, e.g., respiratory. We identified seven such biclusters: five gastrointestinal, one respiratory, and one hematopoietic. For discussion in this paper we selected the most statistically significant bicluster from each category.

Gastrointestinal Pathogens

Bicluster 38 consisting of the Gram-negative pathogens *Yersinia enterocolitica* wap and p60 strains, *Helicobacter pylori* kx2 strain, and enterohemorrhagic *Escherichia coli* is the bicluster most enriched with gastrointestinal pathogens (p-value 1.5×10^{-3}). *Yersinia enterocolitica* causes a broad range of gastrointestinal syndromes ranging from acute diarrhea, terminal ileitis, mesenteric lymphadenitis, and pseudoappendicitis [57]. *Helicobacter pylori* Kx2 strain is responsible for causing gastric adenocarcinoma [58]. Enterohemorrhagic *Escherichia coli* causes diarrhea or hemorrhagic colitis in humans [59]. The four pathogens jointly up-regulate 227 gene sets (Figure 2.3A shows gene sets in this bicluster that contain drug targets). There are 18 known drug targets in this bicluster (p-value 3.7×10^{-7}). Below we will discuss the drug targets IL1R1 and TNF, which are both primary pro-inflammatory cytokines.

Interleukin-1 type 1 receptor (IL-1R1) is a target molecule for the drug Anakinra (DrugBank ID DB00026). Anakinra is designed to treat rheumatoid arthritis by competitively binding to IL-1R1 thereby inhibiting the action of elevated levels of the pro-inflammatory cytokine IL-1. Previous studies have shown that *Yersinia enterocolitica*, *Helicobacter pylori* kx2 strain, and Enterohemorrhagic *Escherichia coli* induce chronic inflammation [58,60,61]. These observations suggest the potential use of drugs that suppress elevated levels of IL-1, such as Anakinra, in the treatment of gastrointestinal infections caused by these four pathogens. Another pro-inflammatory molecule produced by cells infected with bacteria

is TNF-α, which can cause TNF-α-induced apoptosis. TNF-α has been implicated as a target molecule for a number of FDA-approved drugs. Etanercept (DrugBank ID: DB00005) and Infliximab (DrugBank ID: DB00065) are TNF-α blockers. Anti-TNF therapies have shown to be effective in the treatment of Crohn's disease and ulcerative colitis, which are both disease of the gastrointestinal tract that are characterized by inflammation [62, 63]. Although we did not find supporting evidence on the use of these drugs in the treatment of infections caused by *Yersinia enterocolitica*, *Helicobacter pylori* kx2 strain, and Enterohemorrhagic *Escherichia coli*, the potential use of TNF-α blockers such as Etanercept and Infliximab in the treatment of infection caused by these four pathogens may be worth investigating.

Respiratory Pathogens

Bicluster 72 is enriched with respiratory pathogens (p-value 3.0×10^{-2}). It contains the pathogens *Pseudomonas aeruginosa* and *Mycobacterium tuberculosis*. *Pseudomonas aeruginosa* causes major infections in immunocompromised patients. It is also a leading cause of hospital-acquired infections such as pneumonia [64]. *Mycobacterium tuberculosis* is a causative agent of tuberculosis. The two pathogens jointly perturb 245 gene sets including the IL-12 and IL-23 pathways (Figure 2.3B shows gene sets in this bicluster that contain drug targets). The role of IL-12 induction in the treatment of *M. tuberculosis* has been reported in previous studies. For instance, Lowrie *et al.* have shown that up-regulation of IL-12 suppressed proliferation of *M.tuberculosis* in mice [65]. They further suggested

the inclusion of this cytokine in tuberculosis vaccines. IL-12 plays a significant role in the host response against *P.aeruginosa*. It is an important molecule in the generation of IFN-γ and TNF-α, which are essential to promote bacterial clearance. Up-regulation of IL-12 by the host cell is a common strategy used by the host to fight infections caused by these two pathogens. Boosting the level of this molecule when needed, e.g., in immunocompromised patients, might be a viable strategy to treat infection caused by *Pseudomonas aeruginosa* and *Mycobacterium tuberculosis*. Studies suggest that *Pseudomonas aeruginosa* up-regulate IL-23 there by creating airway inflammation in the host. Dubin *et al.* [66] suggested suppression of IL-23 as a potential avenue for immunotherapy to infection with this pathogen. Another study indicated that IL-23 is not required by the host to control *Mycobacterium tuberculosis* infection [67]. This indicates the down-regulation of IL-23 may not disrupt the host defense mechanism during *M.tuberculosis* infection. Therefore, we suggest that down-regulating IL-23 might be a common strategy to treat infection caused by *Pseudomonas aeruginosa* and *Mycobacterium tuberculosis*.

Hematopoietic Pathogens

Bicluster 0 contains two *E.chaffeensis* species, Arkansa and Wakulla. Infection with *Ehrlichia chaffeensis* causes ehrlichiosis, which is characterized by an influenza-like illness, elevation of transaminase levels and sepsis [68]. These two strains commonly up-regulated as many as 979 gene sets, which is not surprising considering the fact that they are different strains of the same bacterial pathogen. However, what is interesting is that the

E.chaffeensis Liberty strain, which is a part of our study, is not part of this bicluster. This result indicates *E.chaffeensis* Arkansa and *E.chaffeensis* wakulla elicit similar host responses that are different from those perturbed by the Liberty strain. Considering the similarity in the host transcriptional responses it is tempting to speculate that a common treatment regimen may exist for infection caused by the strains Arkansa and Wakulla.

Among the commonly up-regulated gene sets, "HSIAO LIVER SPECIFIC GENES" contains the highest number of known drug targets (Figure 2.3C shows gene sets in this bicluster that contain drug targets). There are 49 known drug-target proteins in this gene set alone. The "HSIAO LIVER SPECIFIC GENES" gene set determined by Hsiao *et al.* [69] contains 255 genes that are selectively expressed in the human liver in a gene expression profiling study that involved 59 human samples of 19 different tissue types. The genes in "HSIAO LIVER SPECIFIC GENES" genes are annotated with liver-specific function including blood coagulation (GO:0007596) and homeostasis (GO:0007599). The up-regulation of the "HSIAO LIVER SPECIFIC GENES" gene sets by by Wakulla and Arkansas (but not by Liberty) might indicate that *E.chaffeensis* Liberty is inactive in the liver.

The liver is an important organ in cholesterol synthesis, regulation, and export to the other cells. The "HSIAO LIVER SPECIFIC GENES" gene set contains the protein F2, coagulation factor II (thrombin), which is linked to the cholesterol lowering drug Simvastatin (DrugBank ID: DB00641). Simvastatin reduces total and LDL-cholesterol as well as plasma triglycerides and apolipoprotein B. Previous studies have indicated that *E.chaffeensis*

requires cholesterol for survival and growth. However, *E.chaffeensis* does not have the genes for synthesizing cholesterol. Instead, it depends on the host cell to acquire this molecule [70]. In another study, treatment of *E.chaffeensis* with cholesterol extraction reagent methyl-β-cyclodextrin hampred the ability of this pathogen to infect leukocytes [71]. With this observation in mind, we reasoned that cholesterol lowering drugs such as Simvastatin can be used in the treatment of *E.chaffeensis* infection.

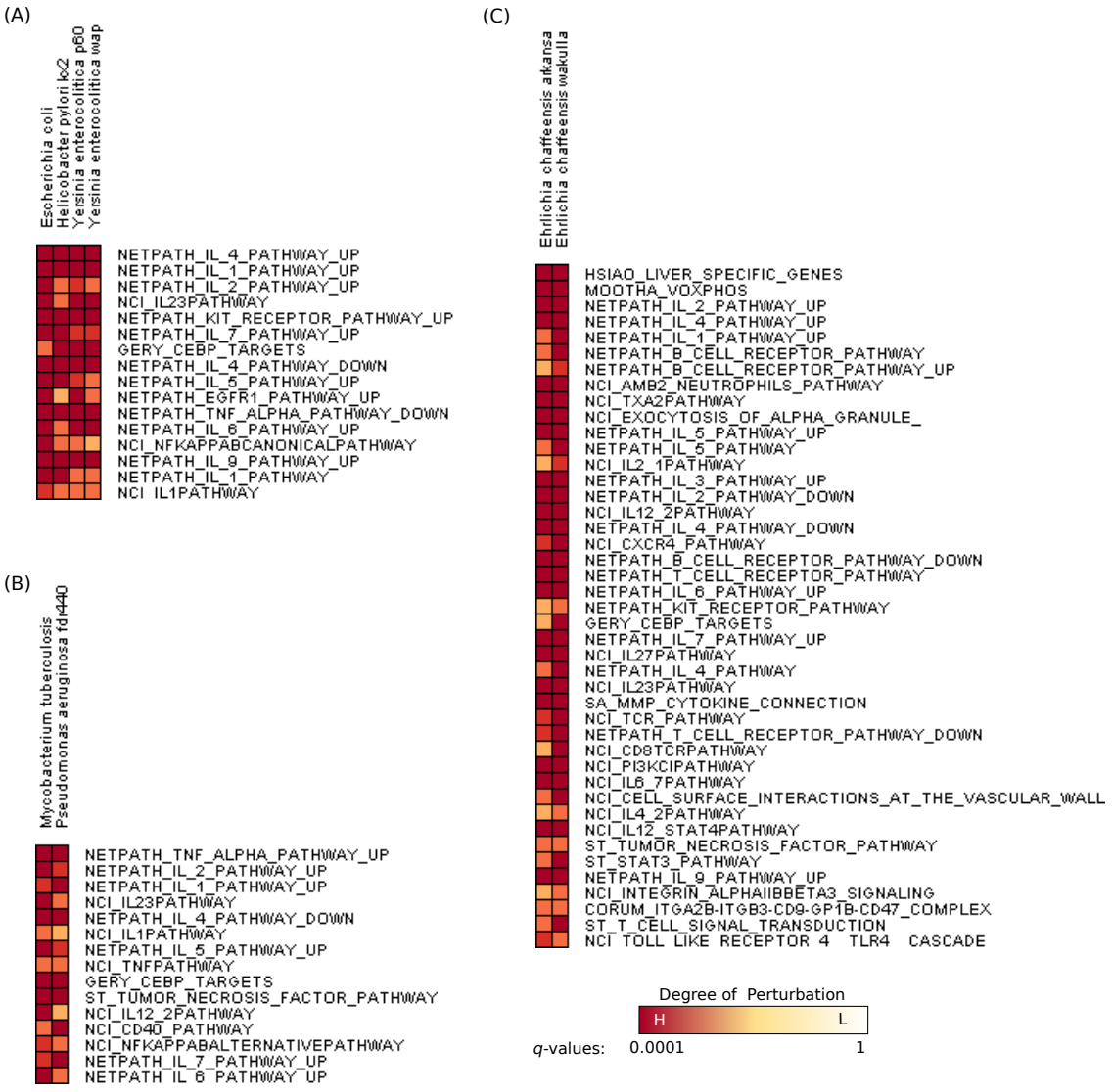

Figure 2.3: Dendrogram of hierarchical clustering of gene sets for three tissue-specific biclusters. (A) *Yersinia enterocolitica* wap and p60 strains, *Helicobacter pylori* kx2 strain, and enterohemorrhagic *Escherichia coli*. (B) *Pseudomonas aeruginosa* and *Mycobacterium tuberculosis*. (C) *E.chaffeensis* Arkansa and Wakulla strains. The figure only shows gene sets that contain one or more known human drug targets.

Pathogens	Bicluster p-value	# Gene Sets	# Targets	Target Enrich. (p-value)
Gastrointestinal				
Yersinia Enterocolitica wap and p60 strains, *Helicobacter Pylori*, and *Escherichia Coli*	1.5×10^{-3}	227	18	3.7×10^{-7}
Yersinia Enterocolitica, *Lactobacillus Acidophilus*, *Listeria Monocytogenes*, and *Helicobacter Pylori*	1.3×10^{-2}	173	11	9.2×10^{-4}
Yersinia Enterocolitica and *Helicobacter Pylori*	1.7×10^{-2}	272	21	7.9×10^{-6}
Yersinia Enterocolitica, *Listeria Monocytogenes*, and *Bifidobacterium Bifidum*	1.8×10^{-2}	269	17	2.1×10^{-4}
Yersinia Enterocolitica, *Bifidobacterium Bifidum*, *Streptococcus Pyogenes*, and *Helicobacter Pylori*	3.6×10^{-2}	101	6	9.7×10^{-3}
Respiratory				
Pseudomonas Aeruginosa, and *Mycobacterium Tuberculosis*	3.0×10^{-2}	245	16	4.7×10^{-4}
Hematopoietic				
Ehrlichia Chaffeensis; Strains: arkansa and wakulla	$< 5.4 \times 10^{-9}$	979	186	4.1×10^{-55}

Table 2.4: Biclusters divided by kind of infection. The columns from left to right are: (i) list of pathogens contained in a bicluster, (ii) a p-value indicating the statistical significance of the bicluster, (iii) the number of gene sets in the bicluster, (iv) the number of known human target genes/proteins in the bicluster, and (v) p-value indicating the enrichment of the bicluster in know human drug-target genes/proteins.

2.5 Materials and Methods

2.5.1 Gene Expression Datasets

We retrieved 808 distinct taxonomic names of bacterial pathogens from the American Biological Safety Association database of human pathogens. We downloaded the GEO meta database [72] that contains metadata associated with the NCBI's Gene Expression Omnibus (GEO) [73] samples, platforms, and datasets. Next, we queried the meta database using the taxonomic names as keywords. We obtained gene expression datasets for 105 of the 808 bacterial pathogens. Next, we pruned the datasets using the following criteria: (i) We removed time-course data to avoid complications that could arise due to temporal variation of cellular responses to the various pathogens. (ii) We excluded datasets that have less than six samples (infected and healthy samples combined) so that our datasets conform to the recommended sample size for conducting *t*-tests. (iii) We considered DNA microarray data collected from three hosts, namely, *Homo sapiens*, *Mus musclus*, and *Rattus norvegicus*. (iv) We considered experiments that involved the comparison of normal and infected samples. After this process, we retained 29 GEO datasets for subsequent analysis. Details on these datasets are given in the supplementary website.

S/N	GEO Acc #	Pathogen Name	Gram Stain	Platform	Organism	Target Cell/Tissue
1	GSE6765	Aeromonas caviae (epa72)	Negative	GPL1261	*M. musculus*	Neonatal mouse small intestinal tissue
2	GSE9723	Aggregatibacter actinomycetemcomitans	Negative	GPL96	*H. sapiens*	Human immortalized gingival keratinocyte
3	GSE2600	Anaplasma phagocytophilum	Negative	GPL570	*H. sapiens*	Promyelocytic cells (NB4)
4	GSE14390	Bacillus anthracis (7702)	Positive	GPL570	*H. sapiens*	Human alveolar macrophage
5	GSE14686	Bacteroides thetaiotaomicron	Negative	GPL1261	*M. musculus*	Proximal colon of mouse
6	GSE20302	Bifidobacterium bifidum	Positive	GPL1261	*M. musculus*	Bone marrow-derived dendritic cells (BMDD)
7	GSE8385	Brucella melitensis	Negative	GPL81	*M. musculus*	Macrophage cell lines (RAW 264.7)
8	GSE8385	Brucella neotomae	Negative	GPL81	*M. musculus*	Macrophage cell lines (RAW 264.7)
9	GSE8385	Brucella ovis	Negative	GPL81	*M. musculus*	Macrophage cell lines (RAW 264.7)
10	GSE7577	Burkholderia pseudomallei	Negative	GPL96	*H. sapiens*	Human monocytic macrophage ell lines (THP-1)
11	GSE6688	Chlamydia pneumoniae	Negative	GPL339	*M. musculus*	Lung
12	GSE8966	Ehrlichia chaffeensis (arkansas, wakulla, liberty)	Negative	GPL1261	*M. musculus*	Liver
13	GSE19315	Escherichia coli	Negative	GPL570	*H. sapiens*	Human monocytic macrophage ell lines (THP-1) cells
14	GSE14686	Eubacterium rectale	Positive	GPL1261	*M. musculus*	Proximal colon of mouse
15	GSE6927	Fusobacterium nucleatum	Negative	GPL96	*H. sapiens*	Human immortalized gingival keratinocytes cells (HIGC)

Continued on next page...

Table 2.5 – Continued

S/N	GEO Acc #	Pathogen Name	Gram Stain	Platform	Organism	Target Cell/Tissue
16	GSE10262	Helicobacter pylori (kx1, kx2)	Negative	GPL1261	*M. musculus*	Gastric epithelial progenitor and non-progenitor cells
17	GSE581	Helicobacter pylori	Negative	GPL193	*H. sapiens*	Gastric biopsies
18	GSE20302	Lactobacillus acidophilus (ncfm)	Positive	GPL1261	*M. musculus*	Bone marrow-derived dendritic cells (BMDD)
19	GSE9946	Listeria monocytogenes	Positive	GPL96	*H. sapiens*	Monocyte-derived dendritic cells
20	GSE17477	Mycobacterium tuberculosis	none	GPL571	*H. sapiens*	Human monocytic macrophage ell lines (THP-1)
21	GSE9723	Porphyromonas gingivalis	Negative	GPL96	*H. sapiens*	Human immortalized gingival keratinocytes cells(HIGC)
22	GSE1469	Pseudomonas aeruginosa (pak)	Negative	GPL91	*H. sapiens*	Lung pneumocytes cell line (A549)
23	GSE923	Pseudomonas aeruginosa (fdr1234, fdr875,fdr1, fdr440)	Negative	GPL96	*H. sapiens*	Calu-3 human lung epithelial cells
24	GSE19315	Shigella dysenteriae	Negative	GPL570	*H. sapiens*	Human monocytic macrophage ell lines (THP-1) cells
25	GSE6802	Staphylococcus aureus	Positive	GPL571	*H. sapiens*	Bronchial epithelial cells beas-2b
26	GSE6927	Streptococcus gordonii	Positive	GPL96	*H. sapiens*	Human immortalized gingival keratinocytes cells (HIGC)
27	GSE8527	Streptococcus pneumoniae (d39, g54, tigr4)	Positive	GPL570	*H. sapiens*	Pharyngeal epithelial cell lines (Detroit 562)
28	GSE11494	Streptococcus pyogenes (90-226)	Positive	GPL1261	*M. musculus*	Nasal-associated lymphoid tissue (NALT)

Continued on next page...

Table 2.5 – Continued

S/N	GEO Acc #	Pathogen Name	Gram Stain	Platform	Organism	Target Cell/Tissue
29	GSE2973	Yersinia enterocolitica (wap, p60)	Negative	GPL339	*M. musculus*	Bone marrow-derived macrophages (BMDM) BALB/C and C57BL/6

Table 2.5: Details of DNA Microarray Datasets Used in the Study.

2.5.2 Gene Set Compendium

We built comprehensive functional annotation data sets encompassing biological pathways and functionally associated genes. We integrated data from four sources:

1. National Cancer Institute-Nature Pathway Interaction Database (NCI-PID): The NCI-PID contains a collection of curated and peer-reviewed pathways of molecular signaling, regulatory events, and cellular processes [74].

2. NetPath: The NetPath database contains cancer and immune signaling pathways, such as the T- and B- cell receptor signaling pathways [75].

3. CORUM: The CORUM database houses protein complexes mainly from human, rat, and mouse. A protein complex contains multiple gene products annotated by the same function or localization e.g., respiratory chain "protein complex-mitochondrial" [76].

4. The Molecular Signature Database (MsigDB): MsigDB contains genes that are biologically related. This relatedness can be defined by participation in the same biological pathway, chromosomal location, or response to some treatment as evidenced by high-throughput experiments such as gene expression profiling. MsigDB houses four categories of gene sets namely, positional gene sets, curated gene sets, motif gene sets, and computational gene sets. In our analyses we used only curated gene sets.

We collected 449 curated pathways from NCI-PID, 20 curated pathways from the NetPath database, 1,765 protein protein complexes from the CORUM database, and 3,272 curated gene sets from MsigDB.

2.5.3 Drugs and Drug Targets Data

We collected 1652 human drug target proteins from DrugBank [28]. These drug targets were linked to 6796 therapeutically-validated and experimental drugs.

2.5.4 Computation of Gene Sets Perturbed in the Host by a Pathogen

We downloaded the raw gene expression profiles (CEL files) from the NCBI's Gene Expression Omnibus (GEO) [73] for the 29 GEO accessions identified above. We normalized the datasets with the Microarry Analysis Suite (MAS5) [77] using the ExpressionFileCreator Module of the Gene Pattern genomic analysis platform [78]. We ran Gene Set Enrichment Analysis (GSEA) [24] on each gene expression dataset using the compendium of gene sets collected above. We collected the resulting q-values (FDR values) into a matrix that indicates the significance of perturbation of each gene set by each pathogen.

2.5.5 Biclustering the *q*-value Matrix

A q-value is the expected probability that GESA's assessment that a pathogen perturbs a gene set represents a false positive finding. For instance, a q-value of 0.2 indicates that a gene-pathogen set association is valid an expected 4 out of 5 times. Using a cutoff of 0.2 for the q-value, we created two binary matrices representing up-regulated and down-regulated biclusters, respectively. In each matrix, each row corresponded to a gene set and each column to a pathogen. An entry in one of these matrices had a value of 1 if and only if the GSEA q-value for that gene set-pathogen pair was at least 0.2. We applied the BiMax algorithm [79] implemented in the BicAT biclustering analysis toolbox [80] on these matrices to obtain two sets of biclusters, one for up-regulated gene sets and another for down-regulated gene sets.

2.5.6 Computing the Statistical Significance of Biclusters

We generated 10,000 randomized binary matrices using the swap randomization algorithm [81]. Given a binary matrix M with values 0 and 1, the swap randomization algorithm creates a random matrix M' such that each row (respectively, column) of M' has the same number of 1s as the corresponding row (respectively, column) of M. The algorithm achieves this goal through a series of steps that swap row-column pairs. We used our own Perl implementation of this algorithm. We computed biclusters in each of these matrices. We built two sets of distributions reflecting the number of pathogens and the number of

genes sets in random biclusters. First, for every integer $k \geq 1$, we recorded the number of biclusters that contained k pathogens and at least l gene sets, for different values of l. Next, we repeated this process for each integer k, considering the number of gene sets in a bicluster. Now, given a bicluster in the original data containing k pathogens and l gene sets, we computed two p-values. One p-value was the fraction of random biclusters that contained k pathogens and at least l gene sets. The second p-values was the fraction of random biclusters that contained l gene sets and at least k pathogens. These p-values indicate the probability of observing a bicluster that contains at least a certain number of pathogens or gene sets in the original dataset by chance. We adjusted the p-values for multiple hypothesis testing using the method of Benjamin and Hochberg [44]. Finally, we chose the greater of the two p-values as a p-value for each bicluster. We further considered only biclusters with p-value of at most 0.05.

2.5.7 Computation of Bicluster Enrichment

We computed the enrichment of each bicluster in various attributes such as the number of known drug targets, host type (human, mouse, and rat), infected cell type (epithelial, dendritic, and macrophage), Gram stain of the pathogen (positive and negative), and infection kind (gastrointestinal, respiratory, oral cavity, and hematopoietic). We used Fisher's exact test for testing the significance of enrichment of a bicluster in each of these attributes.

2.5.8 Translating Gene Identifiers

Different data sources use different naming schemes for identifying genes . For instance, the molecular signature database uses HUGO symbols while DrugBank uses UniProt namespaces. We used HUGO gene symbols as the common gene identifier in our study. We used the Synergizer service for translating gene/protein's identifiers from other namespaces to HUGO [82].

2.5.9 Assigning Gene Ontology Biological Processes to a Gene Set

Some of the gene set names in the MsigBD are not self-explanatory, affecting intuitive interpretation of results. In order alleviate this problem, we considered the Gene Ontology biological processes that have the highest overlaps with each respective gene set. To this end, we used the pre-computed overlap/hypergeometric p-values between a gene set and GO processes that are provided on the MsigDB website. For the "NETPATH IL 4 PATHWAY DOWN" gene set, we obtained the corresponding GO biological processes using GOrilla [83].

2.6 Conclusions

In this paper, we have presented a computational approach to identify potential host-oriented broad-spectrum drug targets. Gene set enrichment and biclustering were key

ingredients of our method. We combined these two techniques to compute subsets of pathogens that commonly up- or down- regulated sets of biological pathways, gene sets, or protein complexes. We applied this approach on a compendium of gene expression data that represented 38 bacterial pathogens and pathogen strains, from which we identified 84 up-regulated and three down-regulated statistically significant biclusters. Using this approach we were successful in detecting common host responses that are hallmarks of bacterial infections.

Motivated by the premise that diseases that have high degree of transcriptional similarity may be treated with similar drugs [25], we integrated drug target information into our analysis to predict HOBS targets for bacterial infections. Focusing on biclusters that contained pathogens that infected same tissue, we predicted new uses of the drugs Anakinra, Etanercept, and Infliximab for gastrointestinal pathogens *Yersinia enterocolitica*, *Helicobacter pylori* kx2 strain, and enterohemorrhagic *Escherichia coli* and the drug Simvastatin for hematopoietic pathogen *Ehrlichia chaffeensis*.

Broadly, the approach we presented in this paper falls in the realm of integrative DNA microarray data analysis. It can be viewed as an alternative approach to the existing methods developed to discover transcriptional responses common to many diseases [26,27,31]. Unlike previous approaches, our method leverages on biclustering to detect pathway-specific relationships only among subsets of pathogens.

A difficultly that may arise with our approach is that the number of pathways in a bicluster can sometimes be overwhelming for subsequent analysis. A rational extension

to our work is to design methods to prioritize non-redundant biclusters and biological processes. Recent techniques for functional enrichment [84] may be appropriate for this task.

In this study, we analyzed host response data from bacterial infections. In the future, we plan to apply the approach developed here to fungal and viral data sets as well. The results from our studies and related approaches [27] may serve as powerful resources for researches engaged in host-oriented broad-spectrum drug target discovery.

Chapter 3

Computational Discovery of Common Immunomodulators in Fungal Infections: Towards Broad-Spectrum Immunotherapeutic Interventions

3.1 Abstract

Fungi are the second most abundant type of human pathogen, next to bacteria. Invasive fungal pathogens are leading causes of life-threatening infections in clinical settings. Toxicity to the host and drug-resistance are two major deleterious issues associated with existing antifungal agents. Increasing a host's tolerance to fungal pathogens has the potential to alleviate these problems. A host's tolerance may be improved by modulating the immune system so that it responds more rapidly and robustly in all facets, ranging from the recognition of pathogens to their clearance from the host. An understanding of

biological processes and genes that are perturbed during fungal exposure, colonization, and/or invasion will help guide the identification of endogenous immunomodulators and/or small molecules that activate host immune responses such as specialized adjuvants.

In this study, we present computational techniques using publicly available transcriptional data sets, to predict immunomodulators that may act against multiple fungal pathogens. Our study analyzed data sets derived from host cells exposed to five fungal pathogens, namely, *Alternaria alternata*, *Aspergillus fumigatus*, *Candida albicans*, *Pneumocystis jirovecii*, and *Stachybotrys chartarum*. We observed statistically significant associations between host responses to *A. fumigatus* and *C. albicans*. Our analysis identified biological processes that were consistently perturbed by these two pathogens. These processes contained both immune response inducing genes such as MALT1, SERPINE1, ICAM1, and IL8, and immune response repressing genes such as DUSP8, DUSP6 and SPRED2. We hypothesize that these genes belong to a pool of common immunomodulators that can potentially be activated or suppressed (agonized or antagonized) in order to make the host more tolerant to infections caused by *A. fumigatus* and *C. albicans*. We hope that these immunomodulators may be used to generate experimentally testable hypotheses that could help in the discovery of broad-spectrum immunotherapeutic interventions.

3.2 Introduction

Fungi are the second most abundant pathogens next to bacteria, accounting for 307 of the 1,407 species of recognized human infectious pathogens [2]. The health-related impacts of fungal pathogens have increased over the past several years. For instance, in the years from 2000 to 2005, the fraction of people who were admitted to hospitals primarily due to the presence of *Candida* species in their blood increased by 52% [85]. Furthermore, the death toll from invasive fungal infections has increased by 320% from 1980 through 1997 in the United States alone [86]. The increase in immunocompromised individuals, the prevalence of cancer, chemotherapy treatments, organ transplantation, and autoimmune diseases are major factors that have contributed to the rise in opportunistic fungal infections [87].

Despite the medical importance of fungal pathogens, the discovery and development of new antifungal agents has been very slow. There are only four major classes of antifungals for the treatment of systemic infections, namely fluoropyrimidine analogs, polyenes, azoles, and echinocandins [88]. Fungi and humans have similar cellular structure and molecular machinery. Consequently, the number of fungal-specific targets are few, partly contributing to the shortage of antifungal drugs [89]. Even for approved drugs, e.g., polyenes, toxicity to the host has been a major problem [90]. Over the past two decades, drug resistant fungal pathogens have emerged, resistance of *Candida* species to azoles being the most common type [91]. Recent examples of drug-resistant fungal pathogens

include multidrug resistant *C. glabrata* and azole-resistant central nervous-system *A. fumigatus* [7]

There are two mechanisms in the treatment of fungal pathogens: (1) decreasing pathogen load by clearing pathogen from the host and (2) increasing the ability of a host to limit the health impact of a pathogen by making the host more tolerant to infections [15]. The host's defense capacity primarily constitutes the sum of these two mechanisms. Medical treatments that merely focus on eradicating pathogens from host cells are prone to the development of drug resistance. On the contrary, treatment strategies that target host tolerance were shown to be viable approaches to circumvent the problem of drug resistance [92].

A host's tolerance to pathogens can be improved by modulating the immune system so that it responds rapidly and robustly in all facets, ranging from the recognition of pathogens to their clearance from the host. For example, this strategy may involve the use of granulocyte colony stimulating factor (G-CSF) to increase the activation of immune cells such as neutrophils, macrophages, and dendritic cells or the activation of toll-like receptors to promote recognition of pathogens. Strategies may also involve the use of vaccines to increase the host's humoral immunity [19], which is an antibody mediated immunity but this approach may have limitations in the context of immunosuppression. For instance, mice that are deficient in pentraxin 3, a gene involved in pathogen recognition, have been shown to be more susceptible to *A. fumigatus* infection [93]. In another study, administration of pentraxin 3 improved survival rate of immunocompromised rats in-

fected with *A. fumigatus* and decreased the overall fungal burden [94]. Another example is the use of thymosin $\alpha 1$ (Tα1), a naturally occurring thymic peptide that was shown to facilitate induction of interleukin 12 (IL-12) and functional maturation of dendritic cells [95]. Following this study, chemically synthesized molecules have become available e.g., Thymalfasin which mimics the human Thymosin $\alpha 1$. The benefit of immunomodulatory agents may be maximized by designing them so that they can be effective against multiple pathogens. For instance, activation of antifungal $CD4^+$ T_h1 immunity using epitope *p41* of *A. fumigatus* extracellular cell wall glucanase *Crf1* succeeded in clearing of both invasive aspergillosis and candidiasis [96].

An understanding of biological processes and genes that are perturbed during fungal exposure, colonization, and/or invasion will help guide the identification and development of therapies that are targeted to enhance the hosts' tolerance against fungal pathogens. In this study, we present computational techniques to predict immunomodulators that can act against multiple fungal pathogens, based on publicly available transcriptional data sets.

3.3 Results and Discussion

We obtained genome-wide transcriptional data sets of host responses upon infection by fungal pathogens from the NCBI's Gene Expression Omnibus (GEO) [72] and ArrayExpress [97]. We filtered data using the criterion described in Section 2.5.1, after which we

retained nine data sets. The data sets involved five fungal pathogens, namely *Alternaria alternata*, *Aspergillus fumigatus*, *Candida albicans*, *Pneumocystis jirovecii*, and *Stachybotrys chartarum* [98–103]. They covered seven target cell/tissue types including macrophages, epithelial cells, dendritic cells, monocytes, neutrophils, endothelial cells and lung cells, totaling 107 samples (Table 3.1).

We computed biclusters using the procedure described in Section 2.5.5. Briefly, first, we computed gene sets up- and down- regulated by each fungal pathogen using Gene Set Enrichment Analysis (GSEA) [24]. Then, we created two binary matrices for up- and down-regulated gene sets, respectively. These matrices captured if a gene set was perturbed or unperturbed by a pathogen to a statistically significant extent. We biclustered these matrices using the Bimax biclustering software [79] in order to identify subsets of fungal pathogens that commonly up- or down- regulated subsets of gene sets. We obtained 27 up-regulated and 13 down-regulated biclusters (see supplementary website for details on these biclusters). We assessed the statistical significance of biclusters by comparing their sizes to biclusters found in randomized data sets. After multiple hypothesis correction using the method of Benjamin and Hochberg [44], we retained three significant biclusters (two up-regulated and one down-regulated) at a 0.05 p-value cutoff (Table 3.2). All the significant biclusters contained two pathogens, namely *A. fumigatus* and *C. albicans*. Among the two significantly up-regulated biclusters, we noticed that there were more up-regulated gene sets in bicluster B1 than in bicluster B2. We reasoned that this difference arose from the variation in the target cell type used in *A. fumigatus* infection, since

the *C. albicans* data sets in the two biclusters were identical. The *A. fumigatus* in bicluster B1 came from lung epithelial cells whereas the one in bicluster B2 came from dendritic cells (Table 3.1). Epithelial cells of the lung are the primary entry points for *A. fumigatus* infection. *A. fumigatus* has been shown to adhere to and enter epithelial cells of the lung in order to escape the hosts' phagocytic cells and thereby invade respiratory tissue [104]. We reasoned this fact may explain why *A. fumigatus* perturbed more genes in epithelial cells as compared to in dendritic cells. Hence, in this paper we decided to focus our discussion on up-regulated bicluster B1 and down-regulated bicluster B3.

ID	Author(s)	Acc #	Pathogen	Platform	Target Cell/Tissue
	H. sapiens				
D1	Babiceanu *et al.* [*]	GSE32893	*A. alternata*	HG-U133_Plus_2	Epithelial
D2	Mezger *et al.* [98]	GSE6965	*A. fumigatus*	HG-U133_Plus_2	Dendritic
D3	Sharon *et al.* [100]	GSE24983	*A. fumigatus*	HG-U133A_2	Epithelial
D4	Mattingsdal *et al.* [*]	E-MEXP-1103	*A. fumigatus*	HG-U133_Plus_2	Monocytes
D5	Rubin-Bejerano *et al.* [*]	E-MEXP-914	*C. albicans*	HG-U133A_2	Neutrophils
D6	Rizzetto *et al.* [101]	E-MTAB-135	*C. albicans*	Illumina HumanHT-12 v3.0	Dendritic
D7	Mller *et al.* [102]	GSE7355	*C. albicans*	HG-U133A	Endothelial
	R. norvegicus				
D8	Cheng *et al.* [103]	GSE20149	*P. carinii*	RG_U34A	Macrophages
	M. musculus				
D9	Shimodaira *et al.* [*]	GSE23178	*S. chartarum*	Mouse430_2	Lung

Table 3.1: Description of gene expression data sets of host responses to fungal pathogens. The columns from left to right are (i) dataset id, (ii) authors and publication associated with dataset, (iii) GEO or ArrayExpress accession number, (iv) pathogen name, (v) microarray platform, and (vi) cell/tissue type from which gene expression measurements were taken. Data sets are categorized by the host, namely, *H. sapiens*, *R. norvegicus* and *M. musculus*.. [*] indicates data set that do not have associated publication.

ID	Pathogens	Bicluster p-value	Number of gene sets
	Up-regulated		
B1	*A. fumigatus* (D3) and *C. albicans* (D6)	$< 1.35 \times 10^{-2}$	205
B2	*A. fumigatus* (D2) and *C. albicans* (D6)	$< 1.35 \times 10^{-2}$	174
	Down-regulated		
B3	*A. fumigatus* (D3) and *C. albicans* (D6)	$< 10^{-5}$	133

Table 3.2: Statistically significant biclusters. The columns from left to right are: (i) bicluster identification code, (ii) list of pathogens contained in a bicluster. The texts in parenthesis indicated the gene expression data identification code (see Table 3.1), (iii) a p-value indicating the statistical significance of the bicluster, (iv) the number of gene sets in a bicluster. The table shows up- and down- regulated biclusters

Pathogens may commonly perturb a gene set without perturbing a single gene in common. Therefore, we started our analysis by detecting if gene sets perturbed by *A. fumigatus* and *C. albicans* share common genes. To this end, we considered the leading edge of each gens set perturbed by each pathogen. As computed by GSEA, the leading edge genes for a gene set-pathogen pair constitute those genes that contribute the most to the perturbation of a gene set by a pathogen [24]. We computed the intersection of leading edge genes for the two pathogens for each gene set in biclusters B1 and B3. We retained 115 gene sets in up-regulated bicluster B1 and 49 gene sets in down-regulated bicluster B3 whose leading edge in the two pathogens have a non-empty intersection.

Next, we ranked the remaining gene sets in increasing order of McNemar's Chi-squared statistic. We used this statistic to measure whether or not the two pathogens perturbed different numbers of genes in each gene set. We focused on gene sets where both pathogens

perturbed the same number of genes; i.e with small value of this statistic (see Methods). We called such gene sets *consistently and commonly perturbed gene sets* (see Methods).

Gene set	$\lvert A \cap C \rvert$	$\lvert A \setminus C \rvert$	$\lvert C \setminus A \rvert$	$\lvert A' \cap C' \rvert$	$\lvert n \rvert$
Up-regulated					
ADAPTIVE IMMUNE RESPONSE (MsigDB)	1	4	4	15	24
DISSOLUTION OF FIBRIN CLOT (NCI)	2	1	2	3	8
GRANULOCYTES PATHWAY (BIOCARTA)	3	0	1	10	14
INACTIVATION OF MAPK ACTIVITY (MsigDB)	3	1	1	9	14
VIRAL GENOME REPLICATION (MsigDB)	3	2	2	14	21
HEDGEHOG PATHWAY UP (NETPATH)	3	2	1	16	22
POSITIVE REGULATION OF IMMUNE RESPONSE (MsigDB)	3	3	2	21	29
CD28 DEPENDENT PI3K AKT SIGNALING (REACTOME)	4	2	3	10	19
PEPTIDYL TYROSINE PHOSPHORYLATION (MsigDB)	4	5	6	12	27
CDMAC PATHWAY (BIOCARTA)	5	2	3	6	16
CARDIACEGF PATHWAY (BIOCARTA)	5	3	3	7	18
CHUK NFKB2 REL IKBKG SPAG9 NFKB1 NFKBIE COPB2 TNIP1 NFKBIA RELA TNIP2 COMPLEX (CORUM)	6	1	2	3	12
CD40 PATHWAY (BIOCARTA)	7	1	1	6	15
EGFR1 PATHWAY DOWN (NETPATH)	8	13	13	64	98
NFKAPPABCANONICALPATHWAY (NCI)	9	2	3	10	24
TRAF6 MEDIATED INDUCTION OF THE ANTIVIRAL CYTOKINE IFN ALPHA BETA CASCADE (REACTOME)	10	7	6	30	53
NEGATIVE REGULATION OF APOPTOSIS (MsigDB)	19	18	19	91	147
MAPK SIGNALING PATHWAY (KEGG)	24	34	34	175	267
SIGNALING IN IMMUNE SYSTEM (REACTOME)	28	47	47	244	366

Continued on next page...

Table 3.3 – Continued

Gene set	$\lvert A \cap C \rvert$	$\lvert A \setminus C \rvert$	$\lvert C \setminus A \rvert$	$\lvert A' \cap C' \rvert$	$\lvert n \rvert$
Down-regulated					
LSM1-7 COMPLEX (CORUM)	4	2	1	0	7
RESPIRATORY CHAIN COMPLEX I (INCOMPLETE INTERMEDIATE) MITOCHONDRIAL (CORUM)	4	3	4	0	11
MITOCHONDRIAL RESPIRATORY CHAIN (MsigDB)	8	7	7	2	24

Table 3.3: Consistently perturbed gene sets. The columns from left to right are: (i) name of gene set, (ii) $\lvert A \cap C \rvert$: the number of genes in the gene set perturbed by both *A. fumigatus* and *C. albicans*, (iii) $\lvert A \setminus C \rvert$: the number of genes perturbed by *A. fumigatus* but not by *C. albicans*, (iv) $\lvert C \setminus A \rvert$: the number of genes perturbed by *C. albicans* but not by *A. fumigatus*, (v) $\lvert A' \cap C' \rvert$: the number of genes unperturbed by both pathogens, and (vi) $\lvert n \rvert$: the total number of genes in a gene set. For each gene set, we indicate the source database in parentheses (see Section 3.5.2). This table shows gene set that have a McNemar's test statistic value of zero.

We hypothesize that such consistently and commonly perturbed gene sets represent coherent host responses against the pathogens *A. fumigatus* and *C. albicans*, and that they may contain genes with the potential to serve as common immunomodulators. We selected 19 gene sets from bicluster B1 and 3 gene sets from bicluster B3 that have a McNemar's test statistic value of zero (Table 3.3). Then, we re-ranked these gene sets in increasing order of the number of genes commonly perturbed by *A. fumigatus* and *C. albicans*. Among these gene sets, we selected the top ten gene set from up-regulated bicluster B1 and all three gene sets from down-regulated bicluster B3 (a total of 13 gene sets) for discussion in this paper (Table 3.4).

Gene set	**Genes**
Up-regulated	
ADAPTIVE IMMUNE RESPONSE (MsigDB)	MALT1
DISSOLUTION OF FIBRIN CLOT (NCI)	SERPINE1, PLAUR
GRANULOCYTES PATHWAY (BIOCARTA)	ICAM1, IL8, IL1A
INACTIVATION OF MAPK ACTIVITY (MsigDB)	DUSP8, DUSP6, SPRED2
VIRAL GENOME REPLICATION (MsigDB)	TNIP1, CCL2, IL8
HEDGEHOG PATHWAY UP (NETPATH)	PMP22, THBD, MYC
POSITIVE REGULATION OF IMMUNE RESPONSE (MsigDB)	FYN, EREG, MALT1
CD28 DEPENDENT PI3K AKT SIGNALING (REACTOME)	MAP3K14, FYN, TRIB3, MAP3K8
PEPTIDYL TYROSINE PHOSPHORYLATION (MsigDB)	CLCF1, STAT1, IL12A, LYN
CDMAC PATHWAY (BIOCARTA)	NFKB1, NFKBIA, FOS, MYC, RELA
Down-regulated	
LSM1-7 COMPLEX (CORUM)	LSM4, LSM6, LSM2, LSM7
RESPIRATORY CHAIN COMPLEX I INCOMPLETE INTERMEDIATE MITOCHONDRIAL (CORUM)	NDUFS6, NDUFV2, NDUFS4, NDUFS7
MITOCHONDRIAL RESPIRATORY CHAIN (MsigDB)	UQCRC1, NDUFAB1, BCS1L, NDUFS4, NDUFS7, NDUFS3, NDUFS8, SURF1

Table 3.4: Top ten consistently perturbed gene sets, ranked by the number of genes they contain.

We organize our results into two sections:

1. We evaluated the putative immunomodulatory activity of the 13 selected gene sets and the genes they contain, based on evidence found in the literature.

2. We grouped gene sets depending on whether they induced or repressed immune response, based on the observation made in Step 1. Immune response-inducing genes represent up- or down- regulated genes that would make the host more tolerant to infections by *A. fumigatus* and *C. albicans* , while immune response repressing gene genes are those that would have an opposite effect when up- or down- regulated. We generated a network of these gene sets/genes (see Section 3.3.2) and analyzed their combined immunomodulatory roles.

3.3.1 Predicted immunomodulatory activity

Most immunocompetent humans are immune to infections caused by *A. fumigatus* and *C. albicans*. The innate and adaptive immune systems of the host are versatile enough to prevent infection by these microorganisms. The host becomes prone to infection only when one or more of the immune system components such as physical barriers, cellular or humoral components are compromised [105, 106]. Consistently and commonly perturbed gene sets represent biological processes that are important for pathogen survival or for defending the host from being invaded by the pathogen. In other words, these gene sets are important for maintaining the balance between pathogen clearance, colonization and invasion. Identifying these host responses is the first step in the search for immunotherapies that are effective against both *A. fumigatus* and *C. albicans*. Augmenting these host responses may assist an immunocompromised individual in the fight against these pathogens. Guided by these hypotheses we examined the 13 highly consistently

and commonly perturbed gene sets identified earlier (Table 3.4). These gene sets contain a total of 41 unique genes. Below we discuss the potential immunomodulatory activity of these genes in relation to the gene sets to which they belong, based on evidence found in the literature.

ADAPTIVE IMMUNE RESPONSE (GO:0002460)

In the "ADAPTIVE IMMUNE RESPONSE" gene set, *A. fumigatus* and *C. albicans* commonly up-regulated the mucosa associated lymphoid tissue lymphoma translocation gene 1 (MALT1) (Table 3.4). MALT1 is involved in the activation of Th-17 based adaptive antifungal immunity. This process involves the following steps: Zymosan, a component of the fungal cell wall, activates dectin-1, a pattern recognition receptor in the host. This interaction induces a protein scaffold consisting of caspase recruitment domain 9 (Card9), B cell lymphoma 10 (Bcl10) and Malt1 [107]. Specifically, the activation of Malt1 is responsible for the activation of the c-Rel component of NF-κB, which then induces Th-17 polarizing cytokines such as IL-1β and IL-23 [108].

Inhibition of MALT1 has been shown to prevent c-Rel activation and Th-17 immunity in human primary dendritic cells infected with *Candida* species [109]. Th-17 cells secrete IL-17, which is important for mobilizing neutrophils against fungal infections [108]. Gringhuis *et al.* [108] have experimentally verified the importance of Malt1 in the induction of adaptive immunity against various species of *Candida*. In addition, they pointed out that a similar mechanism may function similarly in *A. fumigatus* infection, due to the

presence of glucans in both the cell wall of *C. albicans* and *A. fumigatus*. Our analysis supports their premises that up-regulation of MALT1 might be a common host response mechanism against *A. fumigatus* and *C. albicans*. Hence, we hypothesize that increasing the expression of MALT1 might help to prompt pathogen recognition by dectin-1 and may serve as a viable strategy for immune response modulation (Figure 3.1).

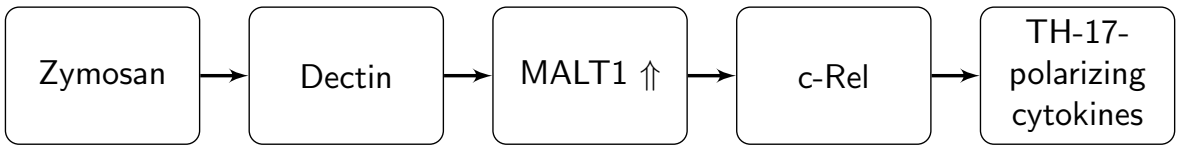

Figure 3.1: Immunomodulation of Th-17 adaptive immunity using MALT1. The figure shows the cascade of events in Malt1 dependent activation of Th-17 type adaptive immunity.

DISSOLUTION OF FIBRIN CLOT (NCI-PID)

Dissolution of fibrin clot (also known as fibrinolysis) refers to the degradation of fibrin. The main enzyme in this process is plasmin. Plasmin is obtained when the precursor plasminogen is converted to plasmin by plasminogen activators. However, the conversion of plasminogen to plasmin can be inhibited by plasminogen activators inhibitors [110].

Previous studies have linked the activation of plasmin and hence fibrinolysis to increased pathogenicity of both *A. fumigatus* and *C. albicans*. The binding of *Candida* to the host's plasmin via its surface cell receptors resulted in an increased ability to cross an *in-vitro* blood-brain barrier [111]. Urokinase plasminogen activator (uPA) and urokinase plas-

minogen activator receptor (uPAR), agents that covert plasminogen to plasmin, were up-regulated when human monocytes were co-cultured with *A. fumigatus* [112].

In our study, *A. fumigatus* and *C. albicans* up-regulated the genes serpin peptidase inhibitor clade E member 1 (SERPINE1) and plasminogen activator urokinase receptor (PLAUR) in the "DISSOLUTION OF FIBRIN CLOT" gene set (Table 3.4). PLAUR encodes for the receptor of urokinase plasminogen activator (uPA) and SERPINE1 is an inhibitor of tissue plasminogen activator (tPA) and urokinase plasminogen activator (uPA). The genes SERPINE1 and PLAUR counteract eachother, in that PLAUR tends to increase the level of plasmin while SERPINE1 tends to inhibit plasmin formation. The up-regulation of these two genes in our analysis might indicate competition between the host and the pathogen where *A. fumigatus* and *C. albicans* attempted to create a favorable environment for their pathogenicity, and the host reactes to overcome this process. This observation led us to speculate that up-regulating SERPINE1 and/or down-regulating PLAUR may help an immunocompromised individual to become more tolerant to infections caused by either of these pathogens (Figure 3.2).

GRANULOCYTES PATHWAY (BIOCARTA)

A. fumigatus and *C. albicans* up-regulated three genes, namely, intercellular adhesion molecule 1 (ICAM1), interleukin 8 (IL-8) and interleukin 1 alpha (IL-1α) in the "GRANULOCYTES PATHWAY" gene set (Table 3.4). Adhesion molecules play an important role in the host defense mechanism against pathogens. Oral epithelial cells have been shown to induce

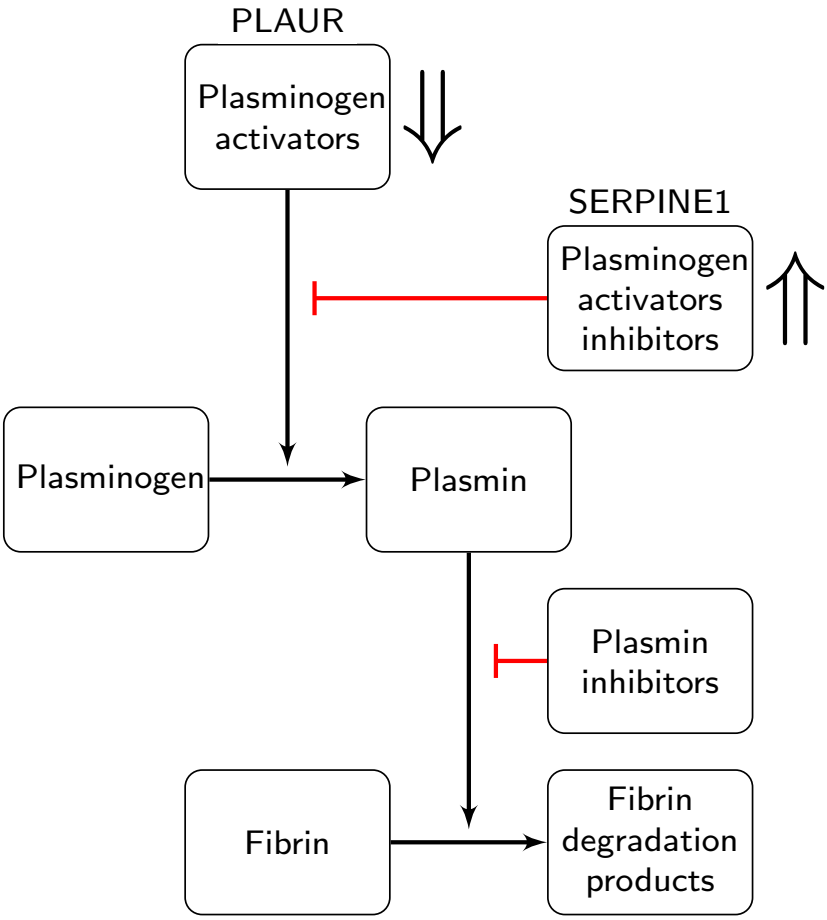

Figure 3.2: Immunomodulation of the dissolution of fibrin clot by SERPINE1 and PLAUR genes. The figure shows a simplified model of the dissolution of fibrin clot and the involvement of the genes SERPINE1 and PLAUR in this process.

ICAM1 in order to recruit and maintain neutrophils during infection by *Actinobacillus actinomycetemcomitans* and *Porphyromonas gingivalis* [113]. Inhibition of ICAM1 has been shown to hinder adherence of *C. albicans* to human gingival epithelial cells (HGECs) and resulted in decreased secretion of IL-8, an important pro-inflammatory molecule secreted during *A. fumigatus* and *C. albicans* infections [114–116]. Another study linked the up-regulation of IL-1α with the secretion of IL-8 in oral epithelial cells infected with *C. albicans* [117]. All these studies indicated the common role of ICAM1, IL-8 and IL-1α in the host defense against pathogens. Our analysis suggests that the up-regulation of the granulocytes pathway, in particular of these three genes, may be an important aspect of the host's defense against infections caused by *A. fumigatus* and *C. albicans*. We suggest that ICAM1, IL-8 and IL-1α as candidates for a host-oriented therapy that exploits the importance of granulocytes pathway.

INACTIVATION OF MAPK ACTIVITY (GO:0000188)

Mitogen-activated protein kinases (MAPKs) encompass a group of protein kinases that regulate a number of cellular processes ranging from cellular differentiation and proliferation to apoptosis. Extracellular signal-regulated kinases (ERKs) such as ERK1, ERK2, and ERK3 were among the first recognized MAPKs in mammals [118]. Inactivation of MAPKs has a negative effect on the normal functioning of host systems including immune responses. For instance, Dubourdeau *et al.* indicated that inactivation of MAPK/ERK correlates with a decrease in the activation of innate immunity against *A. fumigatus* in a mouse

model [119].

In our analysis, *A. fumigatus* and *C. albicans* commonly up-regulated three members of the "INACTIVATION OF MAPK ACTIVITY" gene set, namely, dual specificity phosphatase 6 (DUSP 6), dual specificity phosphatase 8 (DUSP8), and sprouty-related EVH1 domain-containing 2 (SPRED2) genes (Table 3.4). DUSP6 and DUSP8 are known negative regulators of MAPK activity [120]. Another study indicated that SPRED2 negatively regulates growth factor-mediated ERK signaling and hematopoiesis [121]. Given these observations, we hypothesize that the up-regulation of the "INACTIVATION OF MAPK ACTIVITY" gene set might make the host more vulnerable to infections caused by *A. fumigatus* and *C. albicans*, and that the genes DUSP6, DUSP8 and SPRED2 might play a vital role in this aspect. Hence, modulating the expression or activities of these gene products may work to maintain cellular signaling via MAPKs.

VIRAL GENOME REPLICATION (GO:0019079)

The "VIRAL GENOME REPLICATION" gene set annotates 21 genes. Among these genes, *A. fumigatus* and *C. albicans* commonly up-regulated the genes monocyte chemotactic protein-1 (MCP-1), TNFAIP3 interacting protein 1 (TNIP1), and interleukin 8 (IL-8) (Table 3.4). We have discussed the positive role of up-regulation of IL-8 in the host defense against *A. fumigatus* and *C. albicans* above under the "GRANULOCYTES PATHWAY" gene set. Likewise, activation of MCP-1 has been shown to have a positive role in the host defense against aspergillosis and candidiasis. Neutralizing the MCP-1 gene resulted

in increased mortality and pathogen burden in the lungs of mice with invasive aspergillosis [122]. MCP-1 was also produced by oral and vaginal epithelial cells when challenged with *C. albicans* [123]. TNIP1, through its interaction with tumor necrosis factor α-induced protein 3 (TNFAIP), is able to inhibit NF-κB [124,125]. The inhibition of NF-κB is reported to decrease the number of neutrophils as well as the phagocytic and microbicide capacity against *C. albicans* [126]. Hence, in this gene set, while the up-regulation of IL-8 and MCP-1 might help the host to better tolerate the damage caused by *A. fumigatus* and *C. albicans*, down-regulation of TNIP1 might be advantageous to the host.

HEDGEHOG PATHWAY UP (NETPATH)

The hedgehog (Hh) signaling pathway regulates the expression of genes that are important for various cellular processes including growth, cell cycle regulation and embryogenesis [127]. In this pathway, *A. fumigatus* and *C. albicans* up-regulated the genes peripheral myelin protein 22 (PMP22), thrombomodulin (THBD), and MYC (Table 3.4). We did not find evidence in the literature that linked the Hh signaling pathway to fungal infection. However, a recent study showed that up-regulation of this biological pathway increased cellular permissiveness in hepatitis C infection [128]. Paya *et al.* [129] indicated that fungal infections such as those caused by *Candida* species and hepatitis infections reoccur in liver transplant patients, which may suggest an indirect association between the up-regulation of the Hh signaling pathway and the pathogenesis of *C. albicans*. Guided by this observation, we propose that the up-regulation of this pathway might help in increas-

ing the pathogenesis of opportunistic infections such as those caused by *A. fumigatus* and *C. albicans*.

POSITIVE REGULATION OF IMMUNE RESPONSE (GO:0050778)

In the "POSITIVE REGULATION OF IMMUNE RESPONSE" gene set, *A. fumigatus* and *C. albicans* commonly up-regulated the genes tyrosine-protein kinase FYN, epiregulin (EREG) and mucosa associated lymphoid tissue lymphoma translocation gene 1 (MALT1) (Table 3.4). We have discussed the role of MALT1 in the activation of Th-17 host immunity (see the section on the "ADAPTIVE IMMUNE RESPONSE" gene set). Experimental evidence has suggested that the up-regulation of FYN and EREG is associated with an increase in the host's defense against *A. fumigatus* and *C. albicans*. FYN is involved in the control of Th1-Th2 type cellular differentiation. Kudlacz *et al.* [130] showed that FYN-knockout mice exhibited an increase in Th2-type immune response and a decrease in allergic airway inflammation. A shift in the host's immune response towards Th2-type will generally aggravate invasive aspergillosis as well as invasive candidiasis [96]. This observation suggest that the up-regulation of FYN might be linked to the maintenance of host's Th1 type immunity. Th1 type immune response is an integral component of host response in the protection against both *A. fumigatus* and *C. albicans* [131,132]. In addition EREG, which encodes for epiregulin (a protein involved in the formation of epidermal growth factor receptor agonist), plays a vital role in the proliferation of immune cells such as macrophages [133]. EREG is up-regulated when alveolar macrophages are chal-

lenged with *A. fumigatus* [134]. The up-regulation of FYN and EREG by *A. fumigatus* and *C. albicans* in our analysis, might indicate the hosts attempt to shift towards a Th1 type immune response and an increase in the proliferation of phagocytic cells.

CD28 DEPENDENT PI3K AKT SIGNALING (REACTOME)

Phosphatidylinositol 3-kinase (PI3K) signaling is among the first signal transduction events that occurs when an antigen interacts with host cell surface receptors. Activated PI3K will recruits pleckstrin homology domain-containing proteins such as Akt to initiate the PI3K/Akt signaling cascade. PI3K/Akt signaling plays an important role in various cellular activities ranging from cellular differentiation to motility [135].

The "CD28 DEPENDENT PI3K AKT SIGNALING" gene set annotates 19 genes. Among these, the genes mitogen-activated protein 3 kinase 8 (MAP3K8), mitogen-activated protein 3 kinase 14 (MAP3K14), tyrosine-protein kinases FYN and tribbles homolog 3 (Drosophila) (TRIB3) were commonly up-regulated by *A. fumigatus* and *C. albicans* (Table 3.4). We have discussed the role of FYN in the regulation of Th1-Th2 type immune response earlier (see section on "POSITIVE REGULATION OF IMMUNE RESPONSE" gene set).

MAP3K8 is known to activate extracellular signal-regulated kinases (ERKs). MAP3K8-deficient mice exhibited low levels of TNF-α and ERK production [136]. Similarly, MAP3K14 also participates in ERK signaling [137]. The activation of MAPK/ERK is important in the recognition of *A. fumigatus* by innate immunity [119]. The fourth gene perturbed

in this gene set is TRIB3. Schwarzer *et al.* [138] have shown that TRIB3 is up-regulated when there is a shortage of nutrient supplies in cells. With this observation in mind, we reasoned that the the up-regulation of the "CD28 DEPENDENT PI3K AKT SIGNALING" gene set is part of the host defense mechanism against the two pathogens. An immunomodulation strategy might focus on the four genes discussed above.

PEPTIDYL TYROSINE PHOSPHORYLATION (GO:0018108)

The "PEPTIDYL TYROSINE PHOSPHORYLATION" gene set annotates 27 genes. Among these genes, *A. fumigatus* and *C. albicans* commonly up-regulated the genes signal transducer and activator of transcription 1 (STAT1), cardiotrophin-like cytokine factor 1 (CLCF1), interleukin 12A (IL12A), and v-yes-1 Yamaguchi sarcoma viral related oncogene homolog (LYN). STAT1 plays a vital role in the induction of Th17 and Th1 type host immune responses. It encodes for an adapter molecule that is important for the activation of the IL-23 (IL23R) and IL-12 (IL12R) receptor pathways. IL23R and IL12R are vital in inducing Th-17 and Th-1 type immune responses. Previous studies indicated that mutation in STAT1 resulted in defective Th-17-type and Th-1-type responses [139, 140]. Hence, this gene set may have a positive role in the host defense mechanism.

CDMAC PATHWAY (BIOCARTA)

The cadmium-induced DNA synthesis and proliferation in macrophages (CDMAC) pathway annotates 16 genes that were dysregulated when macrophages were exposed to

cadmium ions ($Cd2^{+}$). Among these genes, *A. fumigatus* and *C. albicans* commonly up-regulated five genes namely NFKB1, RELA, NFKBIA, FOS, and MYC. Cadmium induces both cellular proliferation promoting and immune response inhibiting genes. The cellular proliferation genes include NFKB1, REL1, FOS, and MYC. NFKB1 and REL1 are involved in the formation of NF-κB complexes and NFKBIA is an inhibitor of NF-κ-B/REL complexes [141]. Taken together, the up-regulation of this gene set might indicate the proliferation of macrophages in response to infection.

LSM1-7 COMPLEX (CORUM)

The "LSM1-7 COMPLEX" gene set annotates seven genes that are involved in mRNA degradation [142]. Among these genes *A. fumigatus* and *C. albicans* commonly down-regulated four genes, namely LSM4, LSM6, LSM2, and LSM7 (Table 3.4). Unstable mRNA degradation has been linked to a number of diseases including the inflammatory disease, arthritis [143]. Previous studies also indicated the potential regulation of mRNA stability in the treatment of a wide range of diseases including those caused by infectious pathogens [144, 145]. The down-regulation of genes related to mRNA degradation in our analysis might be linked to the pathogenesis of *A. fumigatus* and *C. albicans*. The genes LSM4, LSM6, LSM2, and LSM7 are potential candidates for an immunological strategy that targets the LSM1-7 complex in countering these two pathogens.

MITOCHONDRIAL RESPIRATORY CHAIN (GO:0005746) and RESPIRATORY CHAIN COMPLEX I MITOCHONDRIAL (CORUM)

About 90% of cellular energy, in the form of adenosine-tri-phosphate (ATP), is produced inside mitochondria. Mitochondria also play a vital role in cellular processes such as the regulation of reactive oxygen, calcium homeostasis, programmed cell death and metabolic processes. The classical mitochondrial respiratory chain is composed of four complexes (Complex I-IV) [146]. *A. fumigatus* and *C. albicans* down-regulate 10 unique genes in the "MITOCHONDRIAL RESPIRATORY CHAIN" and "RESPIRATORY CHAIN COMPLEX I MITOCHONDRIAL" gene sets. Studies have shown that mitochondrial respiratory chain has an important role in antiviral processes. A decrease in Coxsackievirus B3 (CVB3) viral load was observed with an increase in mitochondrial complexes I in a mouse model [147]. Although we did not find literature evidence describing the role of the mitochondrial respiratory chain in bacterial and fungal infections, the perturbation of this gene set, in our analysis, may shed some light on its importance. The down-regulation of the mitochondrial respiratory chain gene set might favor the pathogenesis of *A. fumigatus* and *C. albicans*.

3.3.2 Immune response- inducing and repressing gene sets and genes

In the preceding section, we discussed the immunological relevance of 13 consistently and commonly perturbed gene sets. Among these gene sets, we observed that the up-

regulation of seven gene sets namely, "ADAPTIVE IMMUNE RESPONSE", "GRANULOCYTES PATHWAY", "VIRAL GENOME REPLICATION", "POSITIVE REGULATION OF IMMUNE RESPONSE", "CD28 DEPENDENT PI3K AKT SIGNALING", "PEPTIDYL TYROSINE PHOSPHORYLATION", and "CDMAC PATHWAY" favors the host immune response. On the other hand, the up-regulation of gene sets such as "DISSOLUTION OF FIBRIN CLOT", "INACTIVATION OF MAPK ACTIVITY" and "HEDGEHOG PATHWAY UP" and the down-regulation of "LSM1-7 COMPLEX", "RESPIRATORY CHAIN COMPLEX I MITOCHONDRIAL", and "MITOCHONDRIAL RESPIRATORY CHAIN" disfavors the host immune response. In the same way, the perturbed genes also fall into two groups. For instance, genes such as MALT1, SERPINE1, ICAM1, and IL8 have a positive impact on the immune response when up-regulated, while the up-regulation of genes such as DUSP8, DUSP6 and SPRED2 may negatively impact the host immune response. We refer to the first set of gene sets/genes as immune response-inducing and to the second set as immune response-repressing gene sets/genes.

We generated a network of these gene sets using the Markov chain Monte Carlo Biological process Networks (MCMC-BPN) method developed by Lasher [148] (Figure 3.3). Given a gene expression data, protein-protein interaction, functional annotations of genes, and a set of biological process. MCMC-BPN first creates a Bayesian network among all pairs of process and pairs of interacting genes that are cross annotated by the processes. Then, it seeks to find the smallest number of process links that explain as many as perturbed interactions. Also, we created a functional interaction network among the genes using

Cytoscape [149] and data obtained from the STRING database [150] (Figure 3.4). The goal is to identify highly interacting gene sets and genes, upon which we can prioritize immunomodulatory strategies. We noticed that "CD28 DEPENDENT PI3K AKT SIGNALING" and "CDMAC PATHWAY" are the top two highly connected gene sets (Figure 3.3). In addition, genes such as such as NF-κB1 and LSM-7 have many interactors. In general an immunomodulatory strategy against *A. fumigatus* and *C. albicans* might focus on increasing the expression level of genes in the immune response-inducing category and/or suppressing immune response-repressing genes.

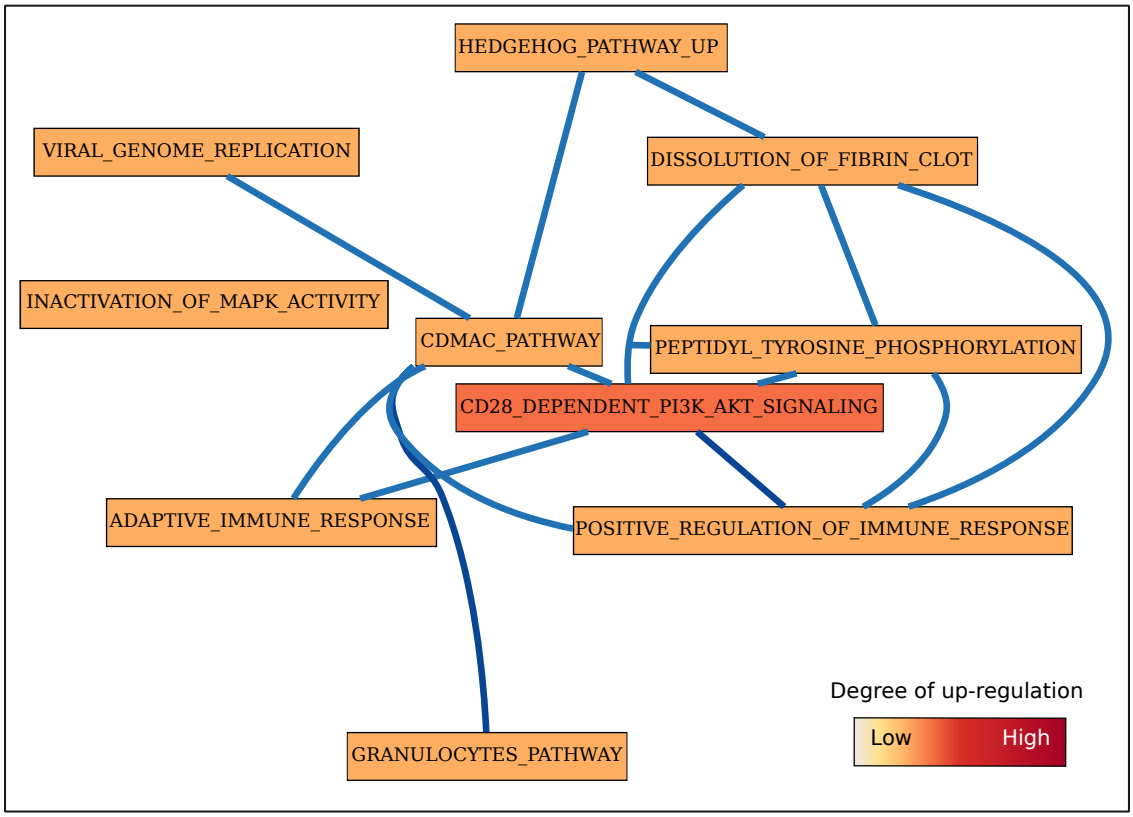

Figure 3.3: A network of immune response-inducing and repressing gene sets. Gene sets are represented with a rectangular shape. Gene sets are connected by an edge if they are interacting.

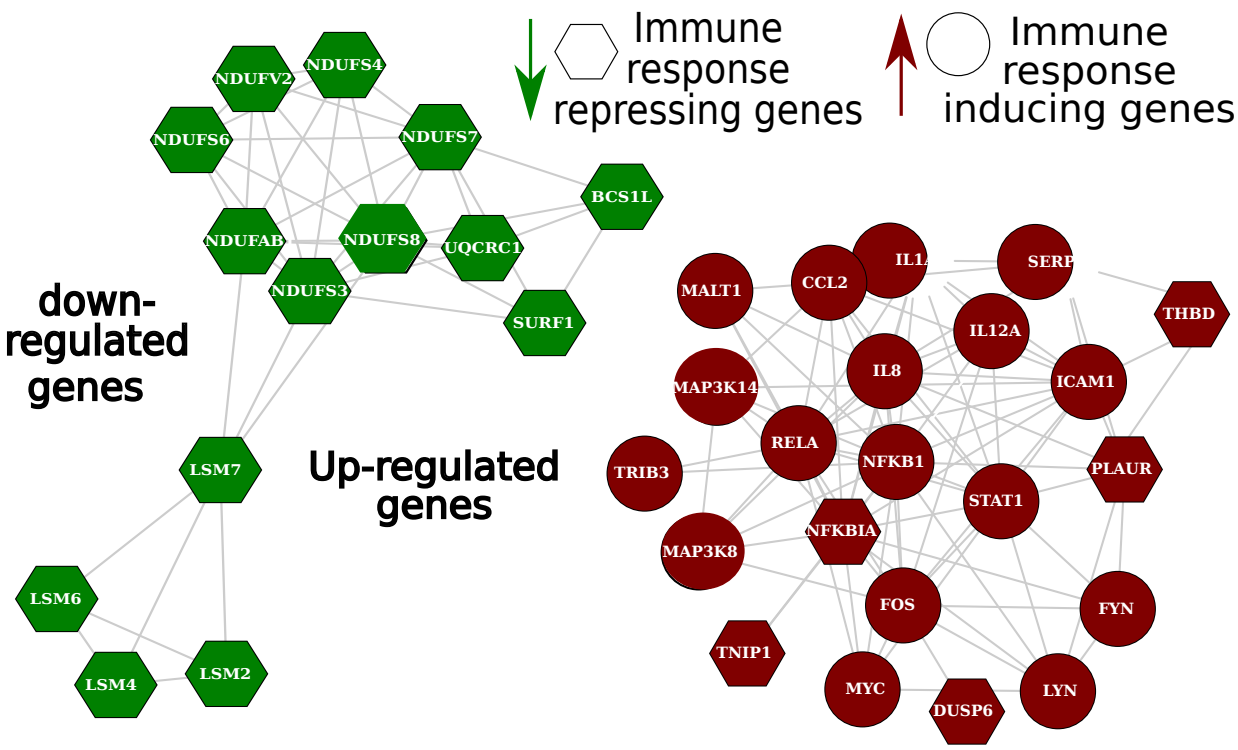

Figure 3.4: A network of immune response- inducing and repressing genes. In this diagram circle represent immune response inducing genes and hexagonal shapes represent immune response repressing genes. Genes are connected by an edge if they are functionally interacting (interaction information was obtained from the STRING database [150]

3.4 Conclusions

In Chapter 2 we used a combination of gene set enrichment analysis and biclustering in the prediction of broad-spectrum drug targets for bacterial infections. In this Chapter, we extended this approach to predict common immunomodulators that can potentially be activated or repressed in order to make the host more tolerant to fungal infections.

Our approach is based on detecting consistently perturbed gene sets among gene sets commonly perturbed by pathogens guided by McNemar's chi-square test of dependence.

The hypothesis behind our approach is that such gene sets might capture biological processes that are commonly and consistently involved in the host-pathogen interaction, and that such gene sets may contain potential putative broad-spectrum immunomodulators. Using this approach, we produced a ranking of gene sets containing immunomodulators, although we focused our analysis on ten highly consistently perturbed gene sets.

In this study we identified genes that are positively and negatively correlated with the host immune response, we called them immune response-inducing and -repressing genes. Immune response -inducing genes were genes whose up-regulation increased the host's immune response and immune response-repressing genes had an opposite effect. We believe that the perturbation of both groups of gene sets and genes are important in maintaining the balance between clearance, colonization and invasion in immunocompetent individuals.

Our approach has enabled the identification of previously known immunomodulators such as MALT1, suggesting the validity of the computational techniques that we implemented in this study. Currently, our analysis is limited by the number of publicly available gene expression data sets pertaining to host-response to fungal infections. In the future we hope to expand the approach to include more fungal pathogens as new data sets become available. Also, our analyses suffers from heterogeneity in the transcriptional data sets utilized in this study, e.g., microarray platforms. We believe that standardized experimental designs may improve such analyses.

Gene	Immunomodulatory Activity	Pathogen	Ref
Immune response inducing genes			
MALT1	Inhibition of MALT1 stopped c-Rel activation and Th-17 immunity to *Candida* species	*C. albicans*	[109]
SERPINE1	SERPINE1 Inhibits plasmin formation. Plasmin bound *C. albicans* showed increased invasiveness.	*C. albicans*	[111]
ICAM1	Inhibition of ICAM-1 stopped adherence of *C. albicans* to human gingival epithelial cells (HGECs) and resulted in a decreased activation of pro-inflammatory cytokine IL-8	*C. albicans*	[114]
IL8	see ICAM1		[114]
IL1A	IL-1α from *C. albicans* infected oral epithelial cells up-regulated secretion IL-8 and granulocyte colony-stimulating factor (GM-CSF)	*C. albicans*	[117]
FYN	FYN knockout mice were shown to have increased Th2 type immune response		[130]
STAT1	Defects in STAT1 result in defective Th17-type and Th1-type responses		[139, 140]
CCL2	Neutralizing CCL2 resulted in increased mortality and pathogen burden in the lungs of mice with invasive aspergillosis	*A. fumigatus*	[122]
MAP3K8	MAP3K8 deficient mice exhibited low-level of TNF-α and ERK production		[136]
MAP3K14	MAP3K14 Activated MAPKs		[137]
NFKB1	Component of NF-κB		

Continued on next page...

Table 3.5 – Continued

Gene	Immunomodulatory Activity	Pathogen	Ref
RELA	Component of NF-κB		
EREG	EREG is important in the proliferation and differentiation of macrophages. It was up-regulated when alveolar macrophages were challenged with conidia of *A. fumigatus*	*A. fumigatus*	[134]
IL12A	Central for the induction of TH1-type cytokines		[139, 140]
TRIB3	TRIB3 senses the presence of cellular nutrient in PI3K/AKT signaling		[138]
CLCF1	Activates the Jak-STAT signaling cascade?		
LYN	Inhibiting the ICAM-1-binding activity?		
FOS	??		
MYC	??		
Immune response repressing genes			
PLAUR	PLAUR promotes plasmin formation. Plasmin bound *C. albicans* showed increased invasiveness.	*C. albicans*	[111]
DUSP8	DUSP8 inhibits MAPK/ERK. Inactivation of MAPK/ERK correlated with a decrease in the activation of innate immunity against *A. fumigatus* in a mice model	*A. fumigatus*	[119]
DUSP6	DUSP6 inhibits MAPK/ERK (see DUSP8)		
SPRED2	SPRED2 inhibits MAPK/ERK (see DUSP8)		

Continued on next page...

Table 3.5 – Continued

Gene	**Immunomodulatory Activity**	**Pathogen**	**Ref**
NFKBIA	Inhibits NF-κB. Inhibition of NF-κ decreased neutrophil phagocytosis and microbicide capacity	*C. albicans*	[126]
TNIP1	Inhibits NF-κB		[125]
PMP22	Up-regulation of Hedgehog pathway increased cellular permissiveness for hepatitis C virus replication		[128]
THBD	Up-regulation of Hedgehog pathway increased cellular permissiveness for hepatitis C virus replication		[128]

Table 3.5: Genes and their predicted immunomodulatory activities for up-regulated bicluster B1.

3.5 Methods

3.5.1 Gene Expression Datasets

We obtained 307 distinct taxonomic names of fungal pathogens from Woolhouse and Gowtage-Sequeria [2]. We queried the GEO meta database [72] and ArrayExpress [97]) using these taxonomic names as keywords. We filtered the data using the criterion described in Section 2.5.1. In short, we removed time-course data, we excluded datasets that have less than six samples, we retained data only from three host species namely *Homo sapiens*, *Mus musculus* and *Rattus norvegicus*, and we removed data that did not involve healthy and infected samples. We obtained nine data sets. The data sets spanned across across five fungal pathogens, namely *Alternaria alternata*, *A. fumigatus*, *C. albicans*, *Pneumocystis jirovecii*, and *Stachybotrys chartarum*, and seven target cell/tissue types including macrophages, epithelial, dendritic cells, monocytes, neutrophils, endothelial cells, and lung (Table 3.1). We downloaded gene expression data and normalized using the GC Robust Multi-array Average (GCRMA) procedure.

3.5.2 Functional annotations

Functional annotation data sets were collected from four sources namely National Cancer Institute-Nature Pathway Interaction Database (NCI-PID), NetPath database (NetPath), CORUM Database of Mammalian Protein Complexes (CORUM), and Molecular Signa-

ture Database (MsigDB). Collectively these were referred to as gene sets in our analysis.

3.5.3 Computation of bicluster genes

In GSEA, the leading edge genes for a gene set-pathogen pair (GS_i, P_j) is defined as the set of genes that contribute the most to the perturbation of a gene set by a pathogen. The leading edge genes for a gene set constitute those genes that appear in the ranked list of genes at or before the point where the running sum reaches its maximum deviation from zero [24]. On the basis of this we intended to define *bicluster genes* which can be interpreted as the core of gene sets that accounts for the perturbation of the gene sets by the pathogens in a bicluster.

Consider a bicluster b that consists of m gene sets and n pathogens, and let $L_{i,j}$ be the set of leading edge genes for the i^{th} gene set and j^{th} pathogen in b. We computed the set of *bicluster genes* in the following way: First, found the intersection of the sets $L_{i,j}$ across all the pathogens for each gene set in a bicluster, and we denoted the resulting set by L_i. Then, we found the union of the sets $L_i s$ across all member gene sets which gave us *bicluster genes* denoted by BG_b. Table 3.6 demonstrates this process for a bicluster that contains two gene sets and two pathogens.

	P_1		P_2	
GS_1	L_{11}	$\cap$	L_{12}	L_1
				$\cup$
GS_2	L_{21}	$\cap$	L_{22}	L_2
				BG_b

Table 3.6: Computation of *bicluster genes* for a 2x2 bicluster. $L_{i,j}$ represent the leading-edge genes for a pathogen-gene set pair. L_i represent leading edge genes for the i^{th} gene set over all the pathogens in the bicluster. BG_b represents the *bicluster genes* for the bicluster

3.5.4 Computation of consistently and common perturbed gene sets

Pathogens may not perturb gene sets in a bicluster in the same manner, i.e., the number of perturbed genes may be skewed towards one of the pathogens, or the perturbation may be similar across the two pathogens, assuming two pathogens in a bicluster. We intended to discover gene sets where the number of genes perturbed by two pathogens remained similar. We converted this into a contingency table as shown in Figure 3.5. We used the McNemar's chi-square test statistic as an assessment criterion [151]. A smaller test statistic result indicated that the number of perturbed genes was similar for the two pathogens whereas a larger test statistic indicated that the number of genes perturbed by the two pathogens varied between the two pathogens.

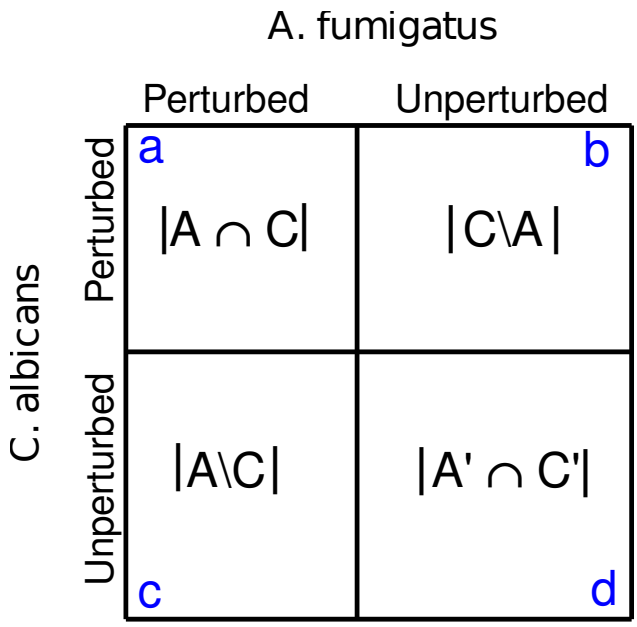

Figure 3.5: Contingency table of perturbed/unperturbed genes by *A. fumigatus* and *C. albicans*, (a) $A \cap C$: number of genes perturbed by both *A. fumigatus* and *C. albicans*, (b) $C \setminus A$: number of genes perturbed by *C. albicans* but not by *A. fumigatus*, (c) $A \setminus C$: number of genes perturbed by *A. fumigatus* but not by *C. albicans*, and (d) $A' \cap C'$: number of genes unperturbed by both pathogens.

Chapter 4

Conclusion

4.1 Summary of this dissertation

In this dissertation we aimed at developing a computational approach for detecting biological pathways that are commonly perturbed by infectious pathogens. Furthermore, we intended to discover HOBS drug targets in such pathways by integrating known drug target genes/proteins.

In Chapter 2 we presented a computational technique to identify common host response programs (combination of predefined biological pathways, gene sets, and protein complexes) that are up- or down- regulated by a group of infectious pathogens, using gene expression and functional annotation data sets. The approach that we developed combined two existing computational techniques, namely biclustering and gene set enrich-

ment analysis. We used gene set enrichment analysis in order to find host response programs that are perturbed when the host cell/tissue is infected with a pathogen. We used the biclustering technique in order to find multi-way associations among host response programs and pathogens. Furthermore, we integrated known drug targets from the DrugBank into genes/proteins that constitute the common host response programs, upon which we predicted host-oriented broad-spectrum drug targets.

We applied this approach on a compendium of gene expression data that pertains to host response to bacterial pathogens. The gene expression data was collected from NCBI's Gene Expression Omnibus (GEO). Based on these steps, (i) we identified hallmarks of bacterial infections such as inflammation, activation of dendritic cells, pro- and anti-apoptotic responses, and (ii) we predicted new uses for the drugs Anakinra, Etanercept, and Infliximab for gastrointestinal pathogens *Yersinia enterocolitica*, *Helicobacter pylori* kx2 strain, and enterohemorrhagic *Escherichia coli* and the drug Simvastatin for hematopoietic pathogen *Ehrlichia chaffeensis*.

In Chapter 3 we extended this approach to discover consistently perturbed biological processes among gene sets commonly perturbed by pathogens. We used the McNemar's chi-squared test for assessing the consistent perturbation of gene sets. Our analysis was driven by the hypothesis that such gene sets represent coherent host responses against pathogens, and that they may contain genes with the potential to serve as common immunomodulators. We analyzed data sets derived from host cells exposed to five fungal pathogens, namely, *Alternaria alternata*, *Aspergillus fumigatus*, *Candida albicans*, *Pneumo-*

cystis jirovecii, and *Stachybotrys chartarum*. We found statistically significant associations among host responses to *Aspergillus fumigatus* and *Candida albicans*. We selected 22 gene sets that were highly and consistently perturbed by *Aspergillus fumigatus* and *Candida albicans* among which we discussed the involvement of 13 gene sets in immunomodulation, based on evidence that we found in the literature.

4.2 Future work

In this study we used transcriptional data sets because they are abundantly available in publicly accessible repositories. Transcriptional data sets only indicate mRNA expression levels. However, cellular activity is governed by the amount of mRNA that is translated into proteins. A common way to remedy this is to integrate Protein-Protein Interaction Networks (PPIN). A possible extension of this work may focus on redesigning the approach in such a way that PPINs are integrated in the analyses.

Because of the heterogeneity of the data sets that were used in this study e.g., microarray platforms and infected cell/tissue types, we acknowledge that certain information might have been lost during integrative analysis e.g., while converting gene identifiers. We believe that standardized experimental designs might yield improved results.

We believe that computational predictions need to be supported by experimental validation. A future extension of this work may focus on experimentally testing some of the predictions made in this study involving host-oriented broad-spectrum drug targets.

Our analysis on host response to fungal pathogens was limited by the availability of transcriptional data sets in GEO and ArrayExpress. Out of the 307 known human fungal pathogens published by Woolhouse and Gowtage-Sequeria [2] we found transcription data sets corresponding to only five fungal pathogens that met our selection criterion. One extension of this work is to automate the analysis procedure so that new data sets are included as they become available.

We acknowledge that our analysis in this dissertation was limited to bacterial and fungal pathogens. Extending this approach to include other pathogens types, e.g., viruses may be the next logical step in this study. While we are writing this section, Smith *et al.* [152] published their work analyzing microarray datasets involving host responses to infections by seven respiratory viruses, namely, influenza A virus, respiratory syncytial virus, rhinovirus, SARS-coronavirus, metapneumonia virus, coxsackievirus and cytomegalovirus. These authors computed pathways significantly perturbed by each of these viral pathogens using a pathway enrichment approach. Then, they ranked pathways by the number of pathogens they were perturbed by; using this they identified commonly perturbed pathways. Furthermore, they identified potential repurposed drugs on the basis of known drugs that target commonly perturbed pathways/genes. A possible extension of this work may focus on increasing the number and diversity of viral pathogens in the study, e.g., include pathogens that infect various tissue/organ types.

In this dissertation, we analyzed bacterial and fungal pathogens separately and we predicted host-oriented broad-spectrum drug targets that are exclusive for a group of bacte-

rial or fungal pathogens. However we believe that some host-oriented therapies can can be effective against pathogens that belong to different groups. For instance, Munter *et al.* showed that a group of viruses, bacteria, and parasites namely *Vaccinia, Enteropathogenic Escherichia coli (EPEC), Cryptosporidium, Group B coxsackie viruses (CVBs), Plasmodium,* and *Theileria* exploit the host's Src family kinases in various ways during their infection process. *Vaccinia, EPEC,* and *Cryptosporidium* use Src signalling to induce local actin polymerization; *CVBs* use it to induce endocytosis, *Plasmodium* uses the Src signalling pathway for cell adhesion, and in the case of *Theileria* this same pathway is used for proliferation [153]. A future extension of this work may focus on a comprehensive analysis of pathogens irrespective of their type. This type of analysis may shed light on host-therapies that are effective against pathogens spanning bacteria, fungi, viruses, and protozoa.

Bibliography

[1] Morens DM, Folkers GK, Fauci AS (2004) The challenge of emerging and re-emerging infectious diseases. Nature 430: 242–249.

[2] Woolhouse ME, Gowtage-Sequeria S (2005) Host range and emerging and reemerging pathogens. Emerging Infectious Diseases 11: 1842–1847.

[3] Michaelis M, Doerr HWW, Cinatl J (2009) Novel swine-origin influenza a virus in humans: another pandemic knocking at the door. Medical microbiology and immunology 198: 175–183.

[4] Rawlins MD (2004) Cutting the cost of drug development? Nature reviews Drug discovery 3: 360–364.

[5] Gilfillan L, Smith BT, Inglesby TV, Kodukula K, Schuler A, et al. (2004) Taking the measure of countermeasures: Leaders' views on the nation's capacity to develop biodefense countermeasures. Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science 2: 320–327.

[6] Walsh C (2000) Molecular mechanisms that confer antibacterial drug resistance. Nature 406: 775–781.

[7] Pfaller MA (2012) Antifungal drug resistance: mechanisms, epidemiology, and consequences for treatment. The American journal of medicine 125.

[8] Taubes G (2008) The bacteria fight back. Science 321: 356–361.

[9] Walsh C (2003) Where will new antibiotics come from? Nature reviews Microbiology 1: 65–70.

[10] Bradley JS, Guidos R, Baragona S, Bartlett JG, Rubinstein E, et al. (2007) Anti-infective research and development–problems, challenges, and solutions. The Lancet infectious diseases 7: 68–78.

[11] Spellberg B, Powers JH, Brass EP, Miller LG, Edwards JE (2004) Trends in antimicrobial drug development: implications for the future. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America 38: 1279–1286.

[12] Nwaka S, Ridley RG (2003) Virtual drug discovery and development for neglected diseases through public-private partnerships. Nature reviews Drug discovery 2: 919–928.

[13] Fischbach MA, Walsh CT (2009) Antibiotics for emerging pathogens. Science (New York, NY) 325: 1089–1093.

[14] Huh AJJ, Kwon YJJ (2011) "nanoantibiotics": a new paradigm for treating infectious diseases using nanomaterials in the antibiotics resistant era. Journal of controlled release 156: 128–145.

[15] Schneider DS, Ayres JS (2008) Two ways to survive infection: what resistance and tolerance can teach us about treating infectious diseases. Nat Rev Immunol 8: 889–895.

[16] Mao H, Chen H, Fesseha Z, Chang S, Medoff HU, et al. (2009) Identification of novel host-oriented targets for human immunodeficiency virus type 1 using random homozygous gene perturbation. Virology Journal 6: 154+.

[17] Murali TM, Dyer MD, Badger D, Tyler BM, Katze MG (2011) Network-Based prediction and analysis of HIV dependency factors. PLoS Comput Biol 7: e1002164+.

[18] Brass AL, Dykxhoorn DM, Benita Y, Yan N, Engelman A, et al. (2008) Identification of Host Proteins Required for HIV Infection Through a Functional Genomic Screen. Science 319: 921–926.

[19] Tseng HKK, Perfect JR (2011) Strategies to manage antifungal drug resistance. Expert opinion on pharmacotherapy 12: 241–256.

[20] Ashburn TT, Thor KB (2004) Drug repositioning: identifying and developing new uses for existing drugs. Nature reviews Drug discovery 3: 673–683.

[21] Dudley JT, Sirota M, Shenoy M, Pai RK, Roedder S, et al. (2011) Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. Science translational medicine 3: 96ra76.

[22] Routh MM, Raut JS, Karuppayil SM (2011) Dual properties of anticancer agents: An exploratory study on the in vitro Anti-Candida properties of thirty drugs. Chemotherapy 57: 372–380.

[23] Cardenas ME, Cruz MC, Del Poeta M, Chung N, Perfect JR, et al. (1999) Antifungal activities of antineoplastic Agents:Saccharomyces cerevisiae as a model system to study drug action. Clinical Microbiology Reviews 12: 583–611.

[24] Subramanian A, Tamayo P, Mootha V, Mukherjee S, Ebert B, et al. (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A .

[25] Dudley JT, Deshpande T, Butte AJ (2011) Exploiting drug-disease relationships for computational drug repositioning. Briefings in Bioinformatics 12: 303–311.

[26] Hu G, Agarwal P (2009) Human disease-drug network based on genomic expression profiles. PLoS ONE 4: e6536+.

[27] Suthram S, Dudley JT, Chiang AP, Chen R, Hastie TJ, et al. (2010) Network-based elucidation of human disease similarities reveals common functional modules enriched for pluripotent drug targets. PLoS Comput Biol 6: e1000662+.

[28] Wishart DS, Knox C, Guo ACC, Cheng D, Shrivastava S, et al. (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Research 36: D901–906.

[29] Russ J, Futschik ME (2010) Comparison and consolidation of microarray data sets of human tissue expression. BMC genomics 11: 305+.

[30] Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, et al. (2006) The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. Science 313: 1929–1935.

[31] Jenner RG, Young RA (2005) Insights into host responses against pathogens from transcriptional profiling. Nature Reviews Microbiology 3: 281–294.

[32] Dudley JT, Tibshirani R, Deshpande T, Butte AJ (2009) Disease signatures are robust across tissues and experiments. Molecular Systems Biology 5.

[33] Fauci AS, Touchette NA, Folkers GK (2005) Emerging infectious diseases: A 10-year perspective from the national institute of allergy and infectious diseases. The International Journal of Risk and Safety in Medicine 17: 157–167.

[34] Schwegmann A, Brombacher F (2008) Host-directed drug targeting of factors hijacked by pathogens. Sci Signal 1: re8+.

[35] Tan SL, Ganji G, Paeper B, Proll S, Katze MG (2007) Systems biology and the host response to viral infection. Nature Biotechnology 25: 1383–1389.

[36] Finlay BB, Hancock RE (2004) Can innate immunity be enhanced to treat microbial infections? Nature reviews Microbiology 2: 497–504.

[37] Del Real G, Jiménez-Baranda S, Mira E, Lacalle RAA, Lucas P, et al. (2004) Statins inhibit HIV-1 infection by down-regulating rho activity. The Journal of Experimental Medicine 200: 541–547.

[38] Liu CII, Liu GY, Song Y, Yin F, Hensler ME, et al. (2008) A cholesterol biosynthesis inhibitor blocks *Staphylococcus aureus* virulence. Science 319: 1391–1394.

[39] Pucadyil TJ, Chattopadhyay A (2007) Cholesterol: a potential therapeutic target in *Leishmania* infection? Trends in Parasitology 23: 49–53.

[40] Hamill P, Brown K, Jenssen H, Hancock R (2008) Novel anti-infectives: is host defence the answer? Current Opinion in Biotechnology 19: 628–636.

[41] Rhodes D, Yu J, Shanker K, Deshpande N, Varambally R, et al. (2004) Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. Proc Natl Acad Sci U S A 101: 9309-14.

[42] Perusse L, Rankinen T, Zuberi A, Chagnon YC, Weisnagel SJ, et al. (2005) The human obesity gene map: The 2004 update. Obesity 13: 381–490.

[43] De Magalhães JaP, Curado Ja, Church GM (2009) Meta-analysis of age-related gene expression profiles identifies common signatures of aging. Bioinformatics 25: 875–881.

[44] Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society 57: 289–300.

[45] Boya P, Roques B, Kroemer G (2001) Viral and bacterial proteins regulating apoptosis at the mitochondrial level. The EMBO Journal 20: 4325–4331.

[46] Hayden MS, West AP, Ghosh S (2006) NF-κB and the immune response. Oncogene 25: 6758–6780.

[47] Zhang Y, Gavriil M, Lucas J, Mandiyan S, Follettie M, et al. (2008) IκBα kinase inhibitor IKI-1 conferred tumor necrosis factor α sensitivity to pancreatic cancer cells and a xenograft tumor model. Cancer Research 68: 9519–9524.

[48] Hinata K, Gervin AM, Jennifer Zhang Y, Khavari PA (2003) Divergent gene regulation and growth effects by NF-κB in epithelial and mesenchymal cells of human skin. Oncogene 22: 1955–1964.

[49] Uzonyi B, Lötzer K, Jahn S, Kramer C, Hildner M, et al. (2006) Cysteinyl leukotriene 2 receptor and protease-activated receptor 1 activate strongly correlated early genes in human endothelial cells. Proceedings of the National Academy of Sciences of the United States of America 103: 6326–6331.

[50] Mahadevan D, Cooke L, Riley C, Swart R, Simons B, et al. (2007) A novel tyrosine kinase switch is a mechanism of imatinib resistance in gastrointestinal stromal tumors. Oncogene 26: 3909–3919.

[51] Théry C (2001) The cell biology of antigen presentation in dendritic cells. Current Opinion in Immunology 13: 45–51.

[52] Lindstedt M, Johansson-Lindbom B, Borrebaeck CAK (2002) Global reprogramming of dendritic cells in response to a concerted action of inflammatory mediators. International Immunology 14: 1203–1213.

[53] Dirmeier U, Hoffmann R, Kilger E, Schultheiss U, Briseño C, et al. (2005) Latent membrane protein 1 of Epstein-Barr virus coordinately regulates proliferation with control of apoptosis. Oncogene 24: 1711–1717.

[54] Takeda K, Kaisho T, Akira S (2003) Toll-like receptors. Annual Review of Immunology 21: 335–376.

[55] Foster SL, Hargreaves DC, Medzhitov R (2007) Gene-specific control of inflammation by TLR-induced chromatin modifications. Nature 447: 972–978.

[56] Seki E, De Minicis S, Osterreicher CH, Kluwe J, Osawa Y, et al. (2007) TLR4 enhances TGF-β signaling and hepatic fibrosis. Nature Medicine 13: 1324–1332.

[57] Bercovier H, Brenner D, Ursing J, Steigerwalt A, Fanning G, et al. (1980) Characterization of *Yersinia enterocolitica* sensu stricto. Current Microbiology 4: 201–206.

[58] Giannakis M, Chen SL, Karam SM, Engstrand L, Gordon JI (2008) *Helicobacter pylori* evolution during progression from chronic atrophic gastritis to gastric cancer and its impact on gastric stem cells. Proceedings of the National Academy of Sciences of the United States of America 105: 4358–4363.

[59] Campos LC, Whittam TS, Gomes TA, Andrade JR, Trabulsi LR (1994) *Escherichia coli serogroup O111* includes several clones of diarrheagenic strains with different virulence properties. Infection and Immunity 62: 3282–3288.

[60] Saebø A, Lassen J (1994) *Yersinia enterocolitica*: an inducer of chronic inflammation. International Journal of Tissue Reactions 16: 51–57.

[61] Ritchie JM, Thorpe CM, Rogers AB, Waldor MK (2003) Critical roles for *stx2*, *eae*, and *tir* in *enterohemorrhagic Escherichia coli*-induced diarrhea and intestinal inflammation in infant rabbits. Infection and Immunity 71: 7129–7139.

[62] Teshima CW, Thompson A, Dhanoa L, Dieleman LA, Fedorak RN (2009) Long-term response rates to infliximab therapy for crohn's disease in an outpatient cohort. Canadian Journal of Gastroenterology 23: 348–352.

[63] Rutgeerts P, Sandborn WJ, Feagan BG, Reinisch W, Olson A, et al. (2005) Infliximab for induction and maintenance therapy for ulcerative colitis. The New England Journal of Medicine 353: 2462–2476.

[64] Bodey GP, Bolivar R, Fainstein V, Jadeja L (1983) Infections caused by *Pseudomonas aeruginosa*. Reviews of Infectious Diseases 5.

[65] Lowrie DB, Tascon RE, Bonato VLD, Lima VMF, Faccioli LH, et al. (1999) Therapy of tuberculosis in mice by DNA vaccination. Nature 400: 269–271.

[66] Dubin PJ, Kolls JK (2007) IL-23 mediates inflammatory responses to mucoid *Pseudomonas aeruginosa* lung infection in mice. American Journal of Physiology - Lung Cellular and Molecular Physiology 292: L519–L528.

[67] Khader SA, Pearl JE, Sakamoto K, Gilmartin L, Bell GK, et al. (2005) IL-23 compensates for the absence of IL-12p70 and is essential for the IL-17 response during tuberculosis but is dispensable for protection and antigen-specific IFN-gamma responses if IL-12p70 is available. Journal of Immunology 175: 788–795.

[68] Paddock CD, Childs JE (2003) *Ehrlichia chaffeensis*: a prototypical emerging pathogen. Clinical Microbiology Reviews 16: 37–64.

[69] Hsiao LL, Dangond F, Yoshida T, Hong R, Jensen RV, et al. (2001) A compendium of gene expression in normal human tissues. Physiological Genomics 7: 97–104.

[70] Rikihisa Y (2010) *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis*: subversive manipulators of host cells. Nature Reviews Microbiology 8: 328–339.

[71] Lin M, Rikihisa Y (2003) *Ehrlichia chaffeensis* and *Anaplasma phagocytophilum* lack genes for lipid biosynthesis and incorporate cholesterol for their survival. Infect Immun 71: 5324–5331.

[72] Zhu Y, Davis S, Stephens R, Meltzer PS, Chen Y (2008) GEOmetadb: powerful alternative search engine for the gene expression omnibus. Bioinformatics 24: 2798–2800.

[73] Barrett T, Suzek T, Troup D, Wilhite S, Ngau W, et al. (2005) NCBI GEO: mining millions of expression profiles–database and tools. Nucleic Acids Res 33 Database Issue: D562-6.

[74] Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, et al. (2009) PID: the Pathway Interaction Database. Nucleic Acids Res 37: D674-9.

[75] Kandasamy K, Mohan SS, Raju R, Keerthikumar S, Kumar GS, et al. (2010) NetPath: a public resource of curated signal transduction pathways. Genome Biol 11: R3.

[76] Ruepp A, Brauner B, Dunger-Kaltenbach I, Frishman G, Montrone C, et al. (2008) CORUM: the comprehensive resource of mammalian protein complexes. Nucleic acids research 36: D646–650.

[77] MAS5. http://media.affymetrix.com/support/technical/whitepapers/saddwhitepaper.pdf.

[78] Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, et al. (2006) GenePattern 2.0. Nature Genetics 38: 500–501.

[79] Prelić A, Bleuler S, Zimmermann P, Wille A, Bühlmann P, et al. (2006) A systematic comparison and evaluation of biclustering methods for gene expression data. Bioinformatics 22: 1122–1129.

[80] Barkow S, Bleuler S, PreliÄ A, Zimmermann P, Zitzler E (2006) BicAT: a biclustering analysis toolbox. Bioinformatics 22: 1282–1283.

[81] Gionis A, Mannila H, Mielikäinen T, Tsaparas P (2007) Assessing data mining results via swap randomization. ACM Trans Knowl Discov Data 1: 14+.

[82] Berriz GF, Roth FP (2008) The synergizer service for translating gene, protein and other biological identifiers. Bioinformatics 24: 2272–2273.

[83] Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z (2009) GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics 10: 48+.

[84] Bauer S, Gagneur J, Robinson PN (2010) GOing bayesian: model-based gene set analysis of genome-scale data. Nucleic acids research 38: 3523–3532.

[85] Zilberberg MD, Shorr AF, Kollef MH (2008) Secular trends in candidemia-related hospitalization in the united states, 2000-2005. Infection control and hospital epidemiology 29: 978–980.

[86] McNeil MM, Nash SL, Hajjeh RA, Phelan MA, Conn LA, et al. (2001) Trends in mortality due to invasive mycotic diseases in the united states, 19801997. Clinical Infectious Diseases 33: 641–647.

[87] Maschmeyer G (2006) The changing epidemiology of invasive fungal infections: new threats. International journal of antimicrobial agents 27 Suppl 1: 3–6.

[88] Vandeputte P, Ferrari S, Coste AT (2012) Antifungal resistance and new strategies to control fungal infections. International journal of microbiology 2012.

[89] Ostrosky-Zeichner L, Casadevall A, Galgiani JN, Odds FC, Rex JH (2010) An insight into the antifungal pipeline: selected new molecules and beyond. Nat Rev Drug Discov 9: 719–727.

[90] Dismukes WE (2000) Introduction to antifungal drugs. Clinical Infectious Diseases 30: 653–657.

[91] Kontoyiannis DP, Lewis RE (2002) Antifungal drug resistance of pathogenic fungi. Lancet 359: 1135–1144.

[92] Romani L (2011) Immunity to fungal infections. Nat Rev Immunol 11: 275–288.

[93] Garlanda C, Hirsch E, Bozza S, Salustri A, De Acetis M, et al. (2002) Non-redundant role of the long pentraxin PTX3 in anti-fungal innate immune response. Nature 420: 182–186.

[94] Lo Giudice P, Campo S, Verdoliva A, Rivieccio V, Borsini F, et al. (2010) Efficacy of PTX3 in a rat model of invasive aspergillosis. Antimicrobial agents and chemotherapy 54: 4513–4515.

[95] Romani L, Bistoni F, Gaziano R, Bozza S, Montagnoli C, et al. (2004) Thymosin alpha 1 activates dendritic cells for antifungal th1 resistance through toll-like receptor signaling. Blood 103: 4232–4239.

[96] Stuehler C, Khanna N, Bozza S, Zelante T, Moretti S, et al. (2011) Cross-protective TH1 immunity against aspergillus fumigatus and candida albicans. Blood 117: 5881–5891.

[97] Parkinson H, Sarkans U, Kolesnikov N, Abeygunawardena N, Burdett T, et al. (2011) ArrayExpress update–an archive of microarray and high-throughput sequencing-based functional genomics experiments. Nucleic acids research 39: D1002–D1004.

[98] Mezger M, Wozniok I, Blockhaus C, Kurzai O, Hebart H, et al. (2008) Impact of mycophenolic acid on the functionality of human polymorphonuclear neutrophils and dendritic cells during interaction with aspergillus fumigatus. Antimicrobial agents and chemotherapy 52: 2644–2646.

[99] Gomez P, Hackett TL, Moore MM, Knight DA, Tebbutt SJ (2010) Functional genomics of human bronchial epithelial cells directly interacting with conidia of aspergillus fumigatus. BMC genomics 11.

[100] Sharon H, Amar D, Levdansky E, Mircus G, Shadkchan Y, et al. (2011) PrtT-regulated proteins secreted by aspergillus fumigatus activate MAPK signaling in exposed a549 lung cells leading to necrotic cell death. PloS one 6.

[101] Rizzetto L, Kuka M, De Filippo C, Cambi A, Netea MG, et al. (2010) Differential IL-17 production and mannan recognition contribute to fungal pathogenicity and commensalism. The Journal of Immunology 184: 4258–4268.

[102] Müller V, Viemann D, Schmidt M, Endres N, Ludwig S, et al. (2007) Candida albicans triggers activation of distinct signaling pathways to establish a proinflammatory gene expression program in primary human endothelial cells. Journal of immunology (Baltimore, Md : 1950) 179: 8435–8445.

[103] Cheng BHH, Liu Y, Xuei X, Liao CPP, Lu D, et al. (2010) Microarray studies on effects of pneumocystis carinii infection on global gene expression in alveolar macrophages. BMC microbiology 10: 103+.

[104] Paris S, Boisvieux-Ulrich E, Crestani B, Houcine O, Taramelli D, et al. (1997) Internalization of aspergillus fumigatus conidia by epithelial and endothelial cells. Infection and immunity 65: 1510–1514.

[105] Hube B (2004) From commensal to pathogen: stage- and tissue-specific gene expression of candida albicans. Current opinion in microbiology 7: 336–341.

[106] Brakhage AA (2005) Systemic fungal infections caused by aspergillus species: epidemiology, infection process and virulence determinants. Current drug targets 6: 875–886.

[107] Gross O, Gewies A, Finger K, Schäfer M, Sparwasser T, et al. (2006) Card9 controls a non-TLR signalling pathway for innate anti-fungal immunity. Nature 442: 651–656.

[108] Ouyang W, Kolls JK, Zheng Y (2008) The biological functions of t helper 17 cell effector cytokines in inflammation. Immunity 28: 454–467.

[109] Gringhuis SI, Wevers BA, Kaptein TM, van Capel TMM, Theelen B, et al. (2011) Selective C-Rel activation via malt1 controls Anti-Fungal TH-17 immunity by dectin-1 and dectin-2. PLoS Pathog 7: e1001259+.

[110] Cesarman-Maus G, Hajjar KA (2005) Molecular mechanisms of fibrinolysis. British Journal of Haematology 129: 307–321.

[111] Jong AY, Chen SHM, Stins MF, Kim KS, Tuan TL, et al. (2003) Binding of candida albicans enolase to plasmin(ogen) results in enhanced invasion of human brain microvascular endothelial cells. Journal of Medical Microbiology 52: 615–622.

[112] Loeffler J, Haddad Z, Bonin M, Romeike N, Mezger M, et al. (2009) Interaction analyses of human monocytes co-cultured with different forms of aspergillus fumigatus. Journal of Medical Microbiology 58: 49–58.

[113] Huang GT, Haake SK, Kim JW, Park NH (1998) Differential expression of interleukin-8 and intercellular adhesion molecule-1 by human gingival epithelial cells in response to actinobacillus actinomycetemcomitans or porphyromonas gingivalis infection. Oral microbiology and immunology 13: 301–309.

[114] Egusa H, Nikawa H, Makihira S, Jewett A, Yatani H, et al. (2005) Intercellular adhesion molecule 1-dependent activation of interleukin 8 expression in candida albicans-infected human gingival epithelial cells. Infection and immunity 73: 622–626.

[115] Mostefaoui Y, Bart C, Frenette M, Rouabhia M (2004) Candida albicans and streptococcus salivarius modulate IL-6, IL-8, and TNF- expression and secretion by engineered human oral mucosa cells. Cellular Microbiology 6: 1085–1096.

[116] Borger P, Koëter GH, Timmerman JA, Vellenga E, Tomee JF, et al. (1999) Proteases from aspergillus fumigatus induce interleukin (IL)-6 and IL-8 production in airway epithelial cell lines by transcriptional mechanisms. The Journal of infectious diseases 180: 1267–1274.

[117] Dongari-Bagtzoglou A, Kashleva H, Villar CC (2004) Bioactive interleukin-1alpha is cytolytically released from candida albicans-infected oral epithelial cells. Medical mycology 42: 531–541.

[118] Pearson G, Robinson F, Beers Gibson T, Xu BE, Karandikar M, et al. (2001) Mitogen-activated protein (MAP) kinase pathways: regulation and physiological functions.

Endocrine reviews 22: 153–183.

[119] Dubourdeau M, Athman R, Balloy V, Huerre M, Chignard M, et al. (2006) Aspergillus fumigatus induces innate immune responses in alveolar macrophages through the MAPK pathway independently of TLR2 and TLR4. Journal of immunology (Baltimore, Md : 1950) 177: 3994–4001.

[120] Keyse SM (2008) Dual-specificity MAP kinase phosphatases (MKPs) and cancer. Cancer metastasis reviews 27: 253–261.

[121] Nonami A, Kato R, Taniguchi K, Yoshiga D, Taketomi T, et al. (2004) Spred-1 negatively regulates interleukin-3-mediated ERK/mitogen-activated protein (MAP) kinase activation in hematopoietic cells. The Journal of biological chemistry 279: 52543–52551.

[122] Morrison BE, Park SJ, Mooney JM, Mehrad B (2003) Chemokine-mediated recruitment of NK cells is a critical host defense mechanism in invasive aspergillosis. The Journal of clinical investigation 112: 1862–1870.

[123] Steele C, Fidel PL (2002) Cytokine and chemokine production by human oral and vaginal epithelial cells in response to candida albicans. Infection and immunity 70: 577–583.

[124] Oshima S, Turer EE, Callahan JA, Chai S, Advincula R, et al. (2009) ABIN-1 is a ubiquitin sensor that restricts cell death and sustains embryonic development. Nature 457: 906–909.

[125] Verstrepen L, Carpentier I, Verhelst K, Beyaert R (2009) ABINs: A20 binding inhibitors of NF-B and apoptosis signaling. Biochemical Pharmacology 78: 105–114.

[126] Giraldo E, Martin-Cordero L, Hinchado MD, Garcia JJ, Ortega E (2010) Role of phosphatidylinositol-3-kinase (PI3K), extracellular signal-regulated kinase (ERK) and nuclear transcription factor kappa (NF-k) on neutrophil phagocytic process of <i>Candida</i> <i>albicans</i>. Molecular and Cellular Biochemistry 333: 115–120.

[127] Varjosalo M, Taipale J (2008) Hedgehog: functions and mechanisms. Genes & development 22: 2454–2472.

[128] Choi SS, Bradrick S, Qiang G, Mostafavi A, Chaturvedi G, et al. (2011) Up-regulation of hedgehog pathway is associated with cellular permissiveness for hepatitis c virus replication. Hepatology (Baltimore, Md) 54: 1580–1590.

[129] Paya CV (2001) Prevention of fungal and hepatitis virus infections in liver transplantation. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America 33 Suppl 1.

[130] Kudlacz EM, Andresen CJ, Salafia M, Whitney CA, Naclerio B, et al. (2001) Genetic ablation of the src kinase p59fynT exacerbates pulmonary inflammation in an allergic mouse model. American journal of respiratory cell and molecular biology 24: 469–474.

[131] Hebart H, Bollinger C, Fisch P, Sarfati J, Meisner C, et al. (2002) Analysis of t-cell responses to aspergillus fumigatus antigens in healthy individuals and patients with hematologic malignancies. Blood 100: 4521–4528.

[132] Vultaggio A, Lombardelli L, Giudizi MGG, Biagiotti R, Mazzinghi B, et al. (2008) T cells specific for candida albicans antigens and producing type 2 cytokines in lesional mucosa of untreated HIV-infected patients with pseudomembranous oropharyngeal candidiasis. Microbes and infection / Institut Pasteur 10: 166–174.

[133] Shirakata Y, Komurasaki T, Toyoda H, Hanakawa Y, Yamasaki K, et al. (2000) Epiregulin, a novel member of the epidermal growth factor family, is an autocrine growth factor in normal human keratinocytes. The Journal of biological chemistry 275: 5748–5753.

[134] Cornish EJ, Hurtgen BJ, McInnerney K, Burritt NL, Taylor RM, et al. (2008) Reduced nicotinamide adenine dinucleotide phosphate Oxidase-Independent resistance to aspergillus fumigatus in alveolar macrophages. The Journal of Immunology 180: 6854–6867.

[135] Pietrella D, Rachini A, Pandey N, Schild L, Netea M, et al. (2010) The inflammatory response induced by aspartic proteases of candida albicans is independent of proteolytic activity. Infection and Immunity 78: 4754–4762.

[136] Dumitru CD, Ceci JD, Tsatsanis C, Kontoyiannis D, Stamatakis K, et al. (2000) TNF-alpha induction by LPS is regulated posttranscriptionally via a Tpl2/ERK-

dependent pathway. Cell 103: 1071–1083.

[137] Dhawan P, Richmond A (2002) A novel NF-kappa b-inducing kinase-MAPK signaling pathway up-regulates NF-kappa b activity in melanoma cells. The Journal of biological chemistry 277: 7920–7928.

[138] Schwarzer R, Dames S, Tondera D, Klippel A, Kaufmann J (2006) TRB3 is a PI 3-kinase dependent indicator for nutrient starvation. Cellular signalling 18: 899–909.

[139] Liu L, Okada S, Kong XFF, Kreins AY, Cypowyj S, et al. (2011) Gain-of-function human STAT1 mutations impair IL-17 immunity and underlie chronic mucocutaneous candidiasis. The Journal of experimental medicine 208: 1635–1648.

[140] van de Veerdonk FL, Plantinga TS, Hoischen A, Smeekens SP, Joosten LA, et al. (2011) STAT1 mutations in autosomal dominant chronic mucocutaneous candidiasis. The New England journal of medicine 365: 54–61.

[141] Li H, Lin X (2008) Positive and negative signaling components involved in TNFalpha-induced NF-kappaB activation. Cytokine 41: 1–8.

[142] Bouveret E, Rigaut G, Shevchenko A, Wilm M, Séraphin B (2000) A sm-like protein complex that participates in mRNA degradation. The EMBO journal 19: 1661–1671.

[143] Hollams EM, Giles KM, Thomson AM, Leedman PJ (2002) MRNA stability and the control of gene expression: implications for human disease. Neurochemical research 27: 957–980.

[144] Nakagawa J (2008) Transient responses via regulation of mRNA stability as an immuno-logical strategy for countering infectious diseases. Infectious disorders drug targets 8: 232–240.

[145] Eberhardt W, Doller A, Akool ES, Pfeilschifter J (2007) Modulation of mRNA stability as a novel therapeutic approach. Pharmacology & Therapeutics 114: 56–73.

[146] Martins VdPdeP, Dinamarco TMM, Curti C, Uyemura SAA (2011) Classical and alternative components of the mitochondrial respiratory chain in pathogenic fungi as potential therapeutic targets. Journal of bioenergetics and biomembranes 43: 81–88.

[147] Ebermann L, Wika S, Klumpe I, Hammer E, Klingel K, et al. (2012) The mitochondrial respiratory chain has a critical role in the antiviral process in coxsackievirus b3-induced myocarditis. Laboratory investigation; a journal of technical methods and pathology 92: 125–134.

[148] Lasher CD (2011) Discovering contextual connections between biological processes using high-throughput data. Ph.D. thesis, Genetics, Bioinformatics, and Computational Biology Program, Virginia Polytechnic Institute and State University, Blacksburg, VA.

[149] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13: 2498–504.

[150] Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, et al. (2011) The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic acids research 39: D561–D568.

[151] Agresti A (2009) Categorical data analysis, New York: Wiley. pp. 350-354.

[152] Smith SB, Dampier W, Tozeren A, Brown JR, Magid-Slav M (2012) Identification of common biological pathways and drug targets across multiple respiratory viruses based on human host gene expression analysis. PLoS ONE 7: e33174+.

[153] Munter S, Way M, Frischknecht F (2006) Signaling during pathogen infection. Sci STKE 2006: re5+.