Computational Prediction of Pharmacodynamic Drug Interactions Using Public Gene Expression Data.

BY

GURUNADH PARINANDI
B.S., Ohio State University, 2008

THESIS

Submitted as partial fulfillment of the requirements for the degree of Master of Science in Bioinformatics in the Graduate College of the University of Illinois at Chicago, 2013

Chicago, Illinois

Defense Committee:

Yves A. Lussier, Chair and Advisor
Yang Dai
Haiquan Li, Medicine
ACKNOWLEDGEMENTS:

I would like to thank my thesis advisor Dr. Yves A. Lussier, for his support and guidance throughout my training in bioinformatics. He has taught me a great deal about the research process, from gestation to final product. Without his help, I would not have been able to accomplish this project. I would also like to thank Dr. Haiquan Li for assisting me in conceptualizing all the different facets of this project. I would also like to thank individual members of our lab, Dr. Vincent Gardeux, Dr. Ikbel Achour, and Ms. Colleen Kenost, for assisting me in both the theory and the presentation of this project. I would also like to thank Dr. Joe G.N. Garcia for providing me with financial support in order to finish this project, as well as my master’s degree. I would also like to thank Dr. Yang Dai for giving me proper guidance throughout the graduate school process and in completing this thesis. I would also like to thank my family, Dr. Narasimham Parinandi, Mrs. Nagamani Parinandi, and Mr. Srinivas Parinandi, as well as my fiancé, Ms. Sarah Warner, for their love and guidance throughout my graduate training.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. INTRODUCTION</td>
<td></td>
</tr>
<tr>
<td>1.1 Drug Interactions</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Pharmacogenomic Drug Interactions</td>
<td>2</td>
</tr>
<tr>
<td>1.3 Pharmacokinetic Drug Interactions</td>
<td>3</td>
</tr>
<tr>
<td>1.4 Pharmacodynamic Drug Interactions</td>
<td>5</td>
</tr>
<tr>
<td>1.5 Clinical Relevance of Pharmacodynamic Drug Interactions</td>
<td>7</td>
</tr>
<tr>
<td>1.6 Computational Advances in Translational Research</td>
<td>8</td>
</tr>
<tr>
<td>1.6.1 Connectivity Map</td>
<td>10</td>
</tr>
<tr>
<td>1.6.2 Gene Set Enrichment Analysis</td>
<td>11</td>
</tr>
<tr>
<td>1.7 Drug Repurposing</td>
<td>12</td>
</tr>
<tr>
<td>1.8 Purpose of Study</td>
<td>14</td>
</tr>
<tr>
<td>2. METHODS</td>
<td></td>
</tr>
<tr>
<td>2.1 Data Preparation</td>
<td>15</td>
</tr>
<tr>
<td>2.2 Case Study</td>
<td>16</td>
</tr>
<tr>
<td>2.3 Control Study</td>
<td>18</td>
</tr>
<tr>
<td>2.4 Experimental Design</td>
<td>18</td>
</tr>
<tr>
<td>2.4.1 Computing Chemical Deregulation via Connectivity Map</td>
<td>20</td>
</tr>
<tr>
<td>2.4.2 Method for Imputing KEGG pathway deregulation</td>
<td>23</td>
</tr>
<tr>
<td>CHAPTER</td>
<td>PAGE</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>3.</td>
<td></td>
</tr>
<tr>
<td>RESULTS</td>
<td></td>
</tr>
<tr>
<td>3.1</td>
<td>24</td>
</tr>
<tr>
<td>Chemical Deregulation Imputed By CMAP</td>
<td></td>
</tr>
<tr>
<td>Overlaps Between Drugs</td>
<td></td>
</tr>
<tr>
<td>3.2</td>
<td>26</td>
</tr>
<tr>
<td>KEGG Pathway Deregulation After Application of Drugs By GSEA</td>
<td></td>
</tr>
<tr>
<td>3.3</td>
<td>28</td>
</tr>
<tr>
<td>Control Results</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td></td>
</tr>
<tr>
<td>CONCLUSION AND DISCUSSION</td>
<td></td>
</tr>
<tr>
<td>4.1</td>
<td>31</td>
</tr>
<tr>
<td>Summary</td>
<td></td>
</tr>
<tr>
<td>4.2</td>
<td>31</td>
</tr>
<tr>
<td>Limitations</td>
<td></td>
</tr>
<tr>
<td>4.3</td>
<td>32</td>
</tr>
<tr>
<td>Future Work</td>
<td></td>
</tr>
<tr>
<td>4.4</td>
<td>32</td>
</tr>
<tr>
<td>Conclusion</td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td></td>
</tr>
<tr>
<td>SUPPLEMENTARY FIGURES</td>
<td>34</td>
</tr>
<tr>
<td>6.</td>
<td></td>
</tr>
<tr>
<td>REFERENCES</td>
<td>39</td>
</tr>
<tr>
<td>7.</td>
<td></td>
</tr>
<tr>
<td>CURRICULUM VITAE</td>
<td>41</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Preliminary studies and respective associated computational analyses</td>
<td>15</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Schema of Drug Interactions</td>
<td>2</td>
</tr>
<tr>
<td>2. Schema of Connectivity Map</td>
<td>11</td>
</tr>
<tr>
<td>3. Schema of Gene Set Enrichment Analysis</td>
<td>12</td>
</tr>
<tr>
<td>4. Diagram of Experimental Study</td>
<td>19</td>
</tr>
<tr>
<td>5. Correlated and Anticorrelated drugs with respect to untreated vs. Aplidin treated SKI-DLCL cell lines</td>
<td>21</td>
</tr>
<tr>
<td>6. Correlated and Anticorrelated drugs with respect to untreated vs. ARAC treated SKI-DLCL cell lines</td>
<td>22</td>
</tr>
<tr>
<td>7. Schema for calculating pathway deregulation between two studies via GSEA</td>
<td>23</td>
</tr>
<tr>
<td>8. Imputed chemical transcriptome deregulation applied to Aplidin exposed and ARAC exposed SKI-DLCL cell line</td>
<td>24</td>
</tr>
<tr>
<td>9. Imputed transcriptome deregulation by GSEA-KEGG pathway applied to Aplidin exposed and ARAC exposed SKI-DLCL cell line</td>
<td>26</td>
</tr>
<tr>
<td>10. Connectivity Map and GSEA overlaps of Aplidin treated SKI-DLCL cell line and ATRA treated HL60 cell line.</td>
<td>28</td>
</tr>
<tr>
<td>11. Connectivity Map and GSEA overlaps of Aplidin treated SKI-DLCL cell line and ATRA treated TEX cell line.</td>
<td>29</td>
</tr>
<tr>
<td>FIGURE</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>12.</td>
<td>Control Studies overlaps and associated p-values</td>
</tr>
<tr>
<td>13.</td>
<td>Case Studies overlaps and associated p-values</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

1.1 Drug Interactions

A drug interaction is defined as a situation where a substance, usually a drug, alters the activity of another drug, when administered together. Normally, interactions between two or more drugs come to mind. However, interactions between drugs and food, as well as drugs and medicinal plants could happen. There are many reasons for drug interactions to occur. One example is age, where physiological changes may affect drug metabolism. This is most noticeable by drugs that are metabolized by either the kidneys or liver, such as amphetamines. Another example is an individual’s genetic factors. Genes will synthesize enzymes that will metabolize drugs, and certain races have genes that could increase or decrease the activity of these enzymes. This can be evidenced in the genotypic variations in the isozymes of the enzyme cytochrome P450. Drug interactions are described as being either synergistic (when a drug’s effect is increased) or antagonistic (when a drug’s effect is decreased). Synergism usually implies that the total effect of the drug combination is greater than the sum of the effects taken independently, which is usually described as an additive effect. An example of a synergistic effect of two drugs includes combined consumption of alcohol and prescription painkillers, whereas an example of an antagonistic effect is the reduction of the anticoagulant effect of warfarin when an agent that accelerates hepatic metabolism like phenobarbital is given. Pertaining to the field of
pharmacology, drug interactions have three classifications: pharmacokinetic, pharmacogenomic, or pharmacodynamic. Figure 1 below represents the drug interaction scheme in the form of a tree.

**Figure 1 Schema of Drug Interactions.**

**1.2 Pharmacokinetic Drug Interactions**

Pharmacokinetics is the field of pharmacology that studies the process of how a drug is absorbed, distributed, metabolized, and excreted by the body. Substances of interest not only include pharmaceutical agents, but also hormones, nutrients, and toxins. There are four major types of pharmacokinetic drug interactions.

- Transport and Distribution: This happens when the drug that arrives first binds with plasma protein, while the plasma dissolves
the second drug. This reaction therefore reduces the concentration of the second drug.

- **Absorption Interactions:** The passage of a drug moves into the bloodstream. Absorption can be increased or decreased by either a combination of different foods or drugs. For example, foods high in fat content such as avocado reduce the absorption of a drug.

- **Metabolism Interactions:** Drug combinations can alter the process a drug is metabolized by the body. For instance cytochrome P450 (CYP450) is a large family of hemoproteins that are characterized both by their enzyme activity and their role in metabolizing drugs. This enzyme metabolizes nearly seventy-five percent of all drugs. CYP450 is known to metabolize caffeine. On the other hand nicotine is also known to inhibit the enzyme activity of CYP450. Therefore if nicotine and caffeine are both consumed, then caffeine will be present in the blood stream for longer periods of time.

- **Excretion Interactions:** This is the process where a drug is removed by the kidneys. Drugs that are low in pH are more likely to be removed by the kidneys. An example of this includes amphetamines.

### 1.3 Pharmacogenomic Drug Interactions

Pharmacogenomics is the study of how an individual’s genetic makeup will affect an individual’s response to drugs i.e. the influence of genetic
variation on drug response in patients. This is a more recent approach to the study of drug interactions, brought to the forefront by advances in computation and the completion of the human genome project. Certain questions addressed by pharmacogenomics include what genes are involved in a drug’s mechanism of action, how a drug’s effects are passed through pathways, and how pharmacogenomic information could be used in decisions regarding prescriptions (Altman, 2012). Key questions about adverse effects due to single drug consumption have also been seriously studied in the field of pharmacogenomics. One such example is that of mercaptopurine, which is used to treat child leukemia (Altman, 2012). The drug in fact works by suppressing both white and blood cell production. In the 1950’s it was noticed that when certain children take mercaptopurine, they begin experiencing bone marrow toxicity, immunosuppression, and life-threatening infections. Therefore, even though the drug is known to be therapeutic, the same drug can have dangerous effects on certain individuals. There is a debate right now as to whether or not pharmacokinetics and pharmacodynamics are subfields of pharmacogenomics, the reason being that transcription regulates both pharmacokinetic and pharmacodynamic reactions. With advances in high-throughput sequences, pharmacogenomics is an intense area of study at this present time and will continue to grow exponentially.
1.4 Pharmacodynamic Drug Interactions

Pharmacodynamics is a major branch of pharmacology that studies the biochemical and physiological effects that drugs have on living organisms. These drug interactions can occur at the same biochemical or molecular site, on the same target organ, or on different target organs sharing a common physiological effect. Some of the processes examined by pharmacodynamics include time course of a drug, intensity of therapeutic, and the adverse effects of the drug on the living organism. There are three main types of pharmacodynamics interactions.

- Pharmacological Receptors: Most pharmacodynamic interactions occur at receptor. Specific questions may include the effect a drug’s concentration has on its receptors, and the downstream effect of this interaction. If two drugs act on the same receptor, then the interaction is called homodynamic. If two drugs act on different receptors, then the interaction is called heterodynamic.

- Signal Transduction Mechanism: These are molecular processes that occur after the drug interacts with its receptor.

- Antagonic Physiological Systems: When two drugs act on different organs, increasing concentrations of the same substance, thereby causing adverse effects. An example of this is the combination of digoxin and furosemide, which increases digoxin binding and causes toxicity to the heart.
For the majority of organisms, the intensity of a drug’s effect is determined by the concentration of the drug at its receptors. However there are other factors that may affect a drug’s response. These include the density of the receptors on the cell surface, mechanisms by which a signal is transmitted into the cell by secondary messengers, and how a drug’s effect is controlled by regulatory factors that control gene expression and protein production. On an individual basis, factors such as these will result in a variation of the effectiveness of a drug.

Traditionally, pharmacodynamics is studied in a laboratory setting, where pharmacologists tediously study the effect of a drug on a specified target. One of the many traditional paradigms of pharmacodynamics is that of tolerance, the process by which a drug’s effectiveness can be reduced with continued use. With respect to pharmacodynamics, tolerance can occur when the same concentration of a particular drug has repeated exposure at a receptor site, thereby reducing the effect of the drug. A particular example of pharmacodynamic drug tolerance is the use of opiates in the therapy of chronic pain. It is common that over a period of time, patients will require increased dosage of opiates in order to treat their chronic pain. One of the key classical questions in pharmacodynamics is effectively predicting a patient’s drug regimen. One would like to maximize a drug’s therapeutic properties while minimizing their dosage increases when predicting a patient’s drug regimen.
In the 21st century, pharmacodynamics has merged with systems biology to create a new discipline called systems pharmacology (Van Der Graaf et al., 2011). Single drugs affecting multiple targets may have many downstream effects on regulation expression. In addition, few of the so-called targeted therapies really have a single target (Yang et al., 2010). One can assess quantitatively the dynamic interactions between one or many biological systems and drug(s). It, therefore, can be surmised that gene expression data could indeed play a role in predicting pharmacodynamic interactions.

1.5 Clinical Relevance of Pharmacodynamic Drug Interactions

Pharmacodynamic drug interactions are a key concern in a clinical setting. This is due to the growing practice of polytherapy, the prescribing of more than one drug for treating a patient’s ailments. An individual patient can consume up to ten different types of drugs per day. The key idea is that the more drugs a patient is prescribed, the more likely they will interact. Whereas pharmacokinetic drug interactions are easily verified because of available mathematical tools and equation, pharmacodynamic drug interactions are more difficult to predict because different drugs use different therapeutic targets and different mechanism of actions (Agins, 2012). While combination of treatment is thought to increase therapy, adverse drug effects are a common problem. For example, a combination of terfenadine and cefaclor can potentiate arrhythmia (irregular heart beat). And combined use of central nervous depressants, such as benzodiazepines, opioid analgesics, or alcohol
can lead to respiratory depression, loss of consciousness, coma, or even death. These are just a few examples of the pejorative effects of drug interactions that clinicians will often try to avoid. In the year 2000, the Federal Food and Drug Administration reported that there were on average two million adverse drug reactions per year. Of these adverse drug reactions, one hundred thousand of them resulted in death. Adverse drug reactions are also the fourth leading cause of death in the United States, ahead of pulmonary disease, AIDS, pneumonia, and automobile accidents. Therefore a clinician must be very judicious when prescribing drugs to patients when treating diseases. Effective prediction of adverse drug reactions is of major relevance in the clinical sciences.

### 1.6 Computational Advances in Translational Research

With a more direct focus on translational research, computational methods have been applied to study drug-target associations by utilizing network-based approaches. These modern techniques take advantage of high-throughput data such as drug chemical structures and screens, side effect profiles, transcriptional responses after drug treatment, and genome wide association studies. The ultimate goal of these techniques is to make medicine more personalized to the individual by catering drugs unique to an individual’s DNA. These ideas are also currently being used in predicting pharmacodynamic drug interactions on an individualized basis. Currently, there are two major approaches to computationally inferring
pharmacodynamic drug interactions (Huang et al., 2013). The first method called “similarity-based-drug-drug-interactions” measures the similarity of drug information from either phenotypic or genotypic data sets, such as protein structures (Keiser et al., 2009), indications (Berger et al., 2009), or side effect (Campillos et al., 2008). Originally intended to predict new drug targets, this method uses a multitude of information about any given drug, such as drug structures, targets, indications, side effects, and gene expression profiles (Lamb et al., 2006), to predict drug interactions. The second method utilizes a “knowledge based approach” to predict drug interactions from a collection of sources, which includes the compendia scientific literature, electronic medical record databases, and the Adverse Event Reporting System maintained by the FDA. Currently, these two methods, albeit robust in their own rights, do suffer from limitations. One such limitation is the inability to handle newly discovered drugs, due to insufficient knowledge, and the inability of databases to keep up with the rapid rate of drug discovery (Duke et al., 2012). Currently, newer methods are being developed to study drug interactions. One such method uses a network based approach that consists of inferring pharmacodynamic drug interactions based off of their protein-protein interaction network (Huang et al., 2013).

We will now briefly discuss two popular computational tools used in translational bioinformatics. These are the Gene Set Enrichment Analysis and The Connectivity Map.
1.6.1 Connectivity Map

Connectivity Map (CMAP) is a tool kit created by The Broad Institute at MIT (Lamb et al., 2006). It is intended to use gene expression data in order to find connections between drugs and human diseases, as well as being a hypothesis-generating tool. The connectivity map contains information on 1309 small molecules (drugs) on four cultured human cell lines. Originally when the paper was published, there were only 164 drugs. It can test biological instances like disease vs. normal case. The database contains mRNA expression data for drugs, so it deals with the transcriptome level. The tool kit finds drugs that are correlated (similar to up regulated effect) and anti-correlated (similar to down regulated effect) with an associated p-value based on the Kolmogorov-Smirnov nonparametric statistic. For any given study of interest, one will first create a query signature, i.e., the collection of up regulated and down regulated genes for a specific biological instance in question. Next, the reference gene expression profiles in connectivity map are represented in a nonparametric fashion. Genes are rank ordered based on differences with respect to control. The output of connectivity map is drugs ranked by the connectivity score, which is a metric that ranges between -1 and +1. The query signature is compared to each rank ordered list to determine whether up regulated genes are at top of the list and down regulated genes are at the bottom of the list. If this is the case, the connectivity score is positive. If this is not the case, then the connectivity score is negative. Those drugs at the top of the list have a positive connectivity score and are functionally
correlated to the biological effect studied. Those drugs at the bottom of the list are negative and are anti correlated with the query state, and have the opposite biological effect. Figure 2 gives a schema of how connectivity map is used in our experiment.

1.6.2 Gene Set Enrichment Analysis

Gene Set Enrichment Analysis (GSEA) is a tool kit created by the Broad Institute at MIT (Subramanian et al., 2005). This tool is used to find deregulated pathways from curated sets such as GOBP, GOMF, and the Kyoto Encyclopedia of Genes and Genomes (KEGG). As of now, there are 450 known KEGG pathways. The data is inputted in two separate classes: normal vs. disease. The program will rank an input set of genes based on the difference between expression levels in both classes. It will also indicate significant pathways with an associated p-value based on the Kolmogorov-
Smirnov nonparametric statistic. GSEA also reports an enrichment score based on each pathway. The score uses a random walk to reflect the degree to which a curated set of genes associated to each pathway, is overrepresented in the ranked set of genes, associated to the biological study of interest. The enrichment score is indeed the maximum derivation from zero, encountered in a random walk. Figure 3 gives a schema of how GSEA is used in our experiment.

**Figure 3 Schema of Gene Set Enrichment Analysis**

### 1.7 Drug Repurposing

Drug repurposing is a computational technique intended to use existing known FDA approved drugs in order to predict new disease indications. Armed with increasing computing power and a compendium of public gene
expression data, e.g. NIH GEO, computational techniques have allowed drug repurposing as a whole (Lu et al., 2013). The most widespread example of drug repurposing is sildenafil (VIAGRA), which is used to treat erectile dysfunction. The approach to drug repurposing is using established computational tools and techniques for screening libraries of lead compounds against biological targets of interest, in early stage drug discovery. Drug discovery normally involves billions of dollars of investment, not to mention that ninety percent of drugs fail during the phase I clinical trial period. By using already approved drugs, repurposing reduces the costs involved in early stage drug development. Increasing computational tools have made drug repurposing very effective. High throughput screening has allowed researchers to use libraries of approved drugs against a multitude of biological targets, in order to find robust drug signatures. By using drug expression signatures from the connectivity map, Atul Butte’s team at Stanford University, was able to screen expression profiles from human cancer cell lines treated with 164 drugs against 100 diseases from NIH GEO (Sirota et al., 2011). They found 16,000 drug disease combinations, of which 2,664 were statistically significant (Lu et al., 2013). Of the 100 diseases profiles, 53 have a significant drug disease relationship. The next step after drug repurposing would be moving from in silico to in vitro, where one would test to see if the repurposed drugs indeed had a biological effect. Currently, this is the problem being addressed. Increasing drug libraries and computational techniques imply that drug repurposing has a bright future.
1.8 Purpose of Study

In this study we hypothesize that a compendium of public gene expression data can be utilized to predict pharmacodynamic drug interactions. Armed with increasing computation power, and a compendium of public gene expression data, e.g. NIH-GEO, computational techniques have allowed drug repurposing as a whole (Sirota et al., 2011). However, to our knowledge, this approach has never been utilized for predicting drug interactions. For this purpose, we conduct a proof of concept study using two different analytic methods, and one biologically proven drug interaction for which expression data was available. In order to do this, we utilize Gene Set Enrichment Analysis and Connectivity Map.
2. METHODS

<table>
<thead>
<tr>
<th>Description of Studies</th>
<th>Case Study</th>
<th>Control Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEO Reference ID:</td>
<td>GSE5681</td>
<td>GSE14672</td>
</tr>
<tr>
<td>Authors:</td>
<td>Humeniuk R et al.</td>
<td>Schenk T et al.</td>
</tr>
<tr>
<td>Date of Publication:</td>
<td>12/22/12</td>
<td>3/11/12</td>
</tr>
<tr>
<td>Date Downloaded:</td>
<td>9/13/13</td>
<td>9/23/13</td>
</tr>
<tr>
<td>Platform:</td>
<td>Affymetrix Human Genome U133A 2.0 Array</td>
<td>Illumina HumanHT-12 V4.0 expression beadchip</td>
</tr>
<tr>
<td>Cell line:</td>
<td>SKI-BEC1</td>
<td>H160</td>
</tr>
<tr>
<td>Drug Combination</td>
<td>5nM of Aplidin</td>
<td>ATRA</td>
</tr>
<tr>
<td>applied to cell line:</td>
<td>20nM of araC</td>
<td>TCP and ATRA</td>
</tr>
</tbody>
</table>

| Number of deregulated genes from SAM analyses: | Upregulated=11,954 | Upregulated=43,325 | Upregulated=882 | Upregulated=1,400 | Upregulated=3 | Upregulated=409 |
| FDR cutoff:           | 0.05 | 0.05 | 0.25 | 0.25 | 0.25 | 0.25 |
| Fold Change Cutoff:   | 1.44 | 1.88 | 1 | 1.4 | 1 | 1 |
| Number of deregulated genes following FDR and fold change cutoff: | Upregulated=280 | Upregulated=539 | Upregulated=371 | Upregulated=373 | Upregulated=14 | Upregulated=179 |

Table I: Preliminary studies and respective associated computational analyses.

2.1 Data Preparation

In order to test our hypothesis, we obtained two data sets catalogued from the NIH-GEO database (http://www.ncbi.nlm.nih.gov/geo/). The publication date, and the date when each data set was download for our study is given in Table I. Our case study is entitled “Aplidin synergizes with cytosine arabinoside: functional relevance of mitochondria in Aplidin-induced toxicity.” (Humeniuk et al., 2007). Our control study is entitled “Inhibition of the LD51 (KDM1A) demethylase reactivates the all-trans-retinoic acid differentiation pathway in acute myeloid leukemia” (Schenk et al., 2012). Both of these studies biologically verify their respective drug interactions. For the case study, CEL files were downloaded from the GEO database and then converted to log2 normalized expression data using the oligo package from bioconductor.
With respect to the control study, the GEO dataset was already log2 normalized.

2.2 Case Study

Our case study investigates the synergism between aplidin and cytosine arabinoside (ARAC). Aplidin is a marine derived anti-tumor agent currently undergoing phase II clinical trials. It is widely used to treat leukemia because of the lack of bone marrow toxicity. Multiple signaling pathways have been shown to be involved in aplidin induced cell cycle arrest, during the G1 and G2 phases. The investigators also demonstrate that mitochondria associated (or localized) processes are potential targets of aplidin. ARAC generates oxidative stress and shows synergism with aplidin in leukemia models, both in vivo and in vitro. This is evidenced by the reduction of tumor size when both drugs are applied in combination. The effective combination of aplidin and ARAC are currently being evaluated in phase I and phase II clinical trials.

The investigators show that aplidin potentiates the anti tumor effect of ARAC in vivo. The gene expression data was taken 24 hours after application of both drugs, individually and together, to the SKI-DLCL human lymphoma cell line. The investigators observe that after 24 hours, aplidin invokes a similar change in gene expression as that of ARAC and aplidin in combination. They also find that aplidin exhibits a pleotropic effect on the cell line, since multiple pathways are affected by the
application of this drug. Some of these include fatty acid metabolism, sterol biosynthesis, and energy metabolism.

While neither drug alone significantly increased cell death, both drugs together enhanced it. A potential reason is given. Aplidin alone increased the number of cells in G0/G1 phase and decreased the number of cells in S phase of the cell cycle. ARAC alone inhibits cells in S phase and significantly decreased cells in G1 phase. Since both drugs are acting at different parts of the cell cycle, this could explain synergistic effects on cell death.

Reactive oxygen species (ROS) are important chemically active molecules that contain oxygen. During times of environmental stress, such as UV exposure, ROS levels increase dramatically. This process is known as oxidative stress. In aerobic organisms (those that need oxygen to survive), energy needed to fuel biological functions are produced in the mitochondria. This includes ROS. The paper also shows that both aplidin and ARAC alone, or in combination, resulted in significant increases of ROS. This implies that the application of aplidin is indeed targeting mitochondria associated processes. The investigators show that the source of aplidin induced ROS is the electron transport chain, a sequence of reactions that takes place within the mitochondria. An increase in mitochondrial membrane potential following aplidin treatment could be attributed to ROS production as well, evidenced by mitochondrial swelling and disruption of the outer membrane. On the other hand, ARAC
induced ROS could be from a different source, evidenced by antioxidants used to prevent ARAC toxicity.

2.3 Control Study

Our control study analyzes the effect of two drugs, all trans-retinoic acid (ATRA) and tranylcypromine (TCP), on two different cell lines. These cell lines are TEX and HL60. The TEX cell line is derived from primitive human cord blood cells. The HL60 cell line is a popular in vitro model, used in laboratory research, to study how certain kinds of blood cells are formed. This study analyzes the effect of the two drugs on the two different cell lines both independently, and in combination. Treatment of severe, non-obese, immunodeficient diabetic mice with ATRA and TCP in combination, showed lower engraftment of primary human AML cells in vivo. This suggests that the drug combination could target cells that initiate leukemia. They also note the drug combination could have an anti-leukemia effect as well.

2.4 Experimental Design

Here we present the experimental design for our study. Figure 4 represents a schema of the experimental design.
Utilizing both the control and case studies, we proceeded to conduct a series of experiments. From the case study dataset we conducted two experiments. These include untreated vs. aplidin treated SKI-DLCL cell line and untreated vs. ARAC treated SKI-DLCL cell line. From the control study we conducted four experiments. These include untreated vs. the TEX and HL60 cell lines treated with ATRA, as well as untreated vs. the TEX and HL60 cell lines treated with the drug combination of TCP and ATRA.
2.4.1 Computing Chemical Deregulation via Connectivity Map

We first proceeded to calculate chemical deregulation by using the connectivity map tool. Utilizing the Significant Analyses of Microarrays (SAM) technique (Tusher et al., 2001), we calculated the list of deregulated genes for each dataset. In order to properly use the connectivity map database, one has to have an aggregate sum of 900 differentially expressed genes. In order to obtain such a rigid number of genes, we found an FDR and fold change cutoff to reduce the number of deregulated genes obtained from the SAM analysis. In each case, the final number of genes used is represented in Table I. The FDR and fold change cutoffs were obtained through a process of trial and error. We also represent FDR and fold change cutoffs for each experiment in Table I.

The Connectivity Map database also required that the entire probe set ID’s have representation from the Affymetrix HG-U133A platform. While the case study dataset had equal representation as the HG-U133A platform, we had to convert the probe set ID’s of the control study, from the Illumina HumanHT-12 version 4.0 expression beadchip platform to the HG-U133A platform. In order to convert the probe set ID’s for control study, we utilized the BioMart Portal. The list of up regulated and down regulated probe set ID’s corresponding the HG-U133A platform were separately uploaded on the connectivity web portal as “grp” files. We then executed the algorithm.
A ranked list of drugs corresponding to each dataset will be the output of connectivity map. With each drug there is an associated p-value. We then computed the false discovery rate (FDR) associated to each p-value to account for false positives. We then specified a FDR cutoff of 0.25 for both the case and control studies. Following the results from this procedure, we had our final collection of correlated and anti-correlated drugs pertaining to each specified experiment.

Figure 5 Correlated and Anticorrelated drugs with respect to untreated vs. Aplidin treated SKI-DLCL cell line.

Figure 5 above details the results from Connectivity Map for the case of aplidin treated vs. untreated SKI-DLCL cell line. Connectivity map results include a total of 18 drugs, of which 9 are correlated and 9 are anti-correlated. We proceeded to do this for ARAC treated vs. untreated SKI-DLCL cell line. The results include a total of 52 drugs, of which 32
were correlated and 20 were anti-correlated. Figure 6 includes a partial list of drugs from Connectivity Map.

![ARAC](image)

Number of Up and Down CMAP drugs via ARAC

<table>
<thead>
<tr>
<th>DRUG NAME</th>
<th>DIRECTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  vorinostat</td>
<td>+</td>
</tr>
<tr>
<td>2  trichostatin A</td>
<td>+</td>
</tr>
<tr>
<td>3  sirolimus</td>
<td>+</td>
</tr>
<tr>
<td>4  valproic acid</td>
<td>+</td>
</tr>
<tr>
<td>5  Y-294002</td>
<td>+</td>
</tr>
<tr>
<td>6  amicacin</td>
<td>-</td>
</tr>
<tr>
<td>7  trifluoridine</td>
<td>+</td>
</tr>
<tr>
<td>8  daunorubicin</td>
<td>+</td>
</tr>
<tr>
<td>9  thioridazine</td>
<td>+</td>
</tr>
<tr>
<td>10</td>
<td>pirodine</td>
</tr>
<tr>
<td>11</td>
<td>trifluoperazine</td>
</tr>
<tr>
<td>12</td>
<td>fluphenazine</td>
</tr>
<tr>
<td>13</td>
<td>rituximab</td>
</tr>
<tr>
<td>14</td>
<td>H-2</td>
</tr>
<tr>
<td>15</td>
<td>camptothecin</td>
</tr>
<tr>
<td>16</td>
<td>fulvestrant</td>
</tr>
<tr>
<td>17</td>
<td>methotrexate</td>
</tr>
<tr>
<td>18</td>
<td>prochlorperazine</td>
</tr>
</tbody>
</table>

Drugs from CMAP and their associated directions (total of 52 drugs)

Figure 6 Correlated and Anticorrelated drugs with respect to untreated vs. ARAC treated SKI-DLCL cell line.

We then proceeded to find the drugs that overlapped between aplidin treated vs. untreated SKI-DLCL cell line, and ARAC treated vs. untreated SKI-DLCL cell line. We also found the direction of the drugs with respect to the connectivity score. The final result is represented in Figure 8. This process was also executed to find connected drugs, and their respective directions, for the control study as well. These results are included in the supplement section.
2.4.2 Method for imputing KEGG pathway deregulation

We followed the same method described in 2.4.1 to impute the pathway deregulation between the two studies. In order to calculate pathway deregulation, the GSEA tool kit was used. For our study, we utilized the Kyoto Encyclopedia of Genes and Genomes (KEGG) curated collection of pathways. We first found respective deregulated KEGG pathways between all cases described earlier. We proceeded to consider only those pathways that were below a FDR threshold of 0.50. Next we sought to see which pathways were included in the intersection of each of the separate studies mentioned above. Utilizing the sign enrichment score from GSEA, we were able to find the direction of the deregulated pathway. For instance if an enrichment score for a respective pathway were positive, then we delegated this pathway as being up regulated, and down regulated if the enrichment score were negative. We give a schematic in Figure 7 to compute the pathway deregulation for the case study, via GSEA.

Figure 7 Schema for calculating pathway deregulation between two studies via GSEA
3. RESULTS

3.1 Chemical Deregulation Imputed By CMAP Overlaps between Drugs

![Figure 8 Imputed chemical transcriptome deregulation applied to Aplidin exposed and ARAC exposed SKI-DLCL cell line](image)

In this section we discuss the results obtained from connectivity map. We elucidate these results with respect to our case study. These results are presented in Figure 8. Individual connectivity map results for the control study are given in the supplementary section. After computing the intersection of the individual results from connectivity map, we obtained four drugs as proxy to both aplidin and ARAC. We note that the results from connectivity map indicate that the proxy drugs are in fact anti-
correlated with respect to aplidin and ARAC treatment, in the SKI-DLCL cell line. We next proceeded to compute the statistical significance of the overlap with respect to the anti-correlated drugs, by considering those drugs that are anti-correlated in the intersection of both studies are of interest. If drugs are correlated in the same direction within the intersection of both studies, we appropriate those drugs to the disjoint members of each study. By Fisher’s Exact Test we compute the statistical significance of the overlap to be 0.0045, with the background consisting of the 1309 drugs in the connectivity map database. Since this is below the threshold of 0.05, we identify this overlap to be statistically significant. There was no mention of the biological significance of ARAC applied to the SKI-DLCL cell line in the case study paper mentioned in section 2.2. Therefore we have no way of biologically validating that ARAC has the opposite effect on the SKI-DLCL cell line as aplidin. The paper only mentions the synergistic effects of the combined treatment of aplidin and ARAC, as well as the biological mechanisms affected by the treatment of aplidin to the SKI-DLCL cell line.

Searching various Internet databases, we next proceed to understand the drug results from connectivity map. Anisomycin is an antibiotic that inhibits protein synthesis. It can also cause partial inhibition of DNA synthesis. It activates stress related protein kinases, MAP kinase, and other signal transduction pathways. Anisomycin has a positive connectivity with respect to Aplidin. One argument could be made that
since aplidin reduces the number of cell in the S phase of the cell cycle, the point where DNA replication occurs, the drugs do exhibit similar effects. On the other hand trichostatin A altars gene expression by removing the acetyl group from histones, thereby inhibiting the ability of transcription factors to access DNA molecules from chromatin. ARAC is a drug that damages DNA, when the cell cycle is in the S phase. Therefore, even though these two drugs are doing different things, the end result could be apoptotic, which is identified in section 2.

3.2 KEGG Pathway Deregulation After Application of Drugs by GSEA

![Diagram showing deregulation of pathways by Aplidin and ARAC](image)

<table>
<thead>
<tr>
<th>NAME OF DEREGULATED PATHWAY</th>
<th>DIRECTION OF APLIDIN</th>
<th>DIRECTION OF ARAC</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSA00100_BIOSYNTHESIS_OF_STEROIDS</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>HSA00071_FATTY_ACID_METABOLISM</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>HSA05110_CHOLERA_INFECTION</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>HSA00280_VALINE_LEUCINE_AND_ISOUCINE_DEGRADATION</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 9 Imputed transcriptome deregulation by GSEA-KEGG Pathway applied to Aplidin exposed and ARAC exposed SKI-DLCL cell line.
In this section, we discuss the results obtained from GSEA, and elucidate these results with respect to our case study. Results are presented in Figure 9. Individual pathway deregulation results for the control study are given in the supplementary section. After computing the intersection of the individual results from GSEA, we obtain four deregulated pathways being common to each study. The pathways mentioned in bold, in the table of Figure 9, were biologically validated as being deregulated with respect to aplidin treatment (Humeniuk et al., 2007). Further laboratory evaluation would be needed in order to ascertain whether non-bolded pathways are indeed deregulated. There was no mention of these pathways being deregulated in the case study paper. But the bolded pathways further validate GSEA as a useful computational tool for predicting relevant biological pathways. We next proceed to compute the statistical significance of the overlap with respect to deregulated pathways. We utilize the same method as described for the connectivity map in section 3.1, by only considering the pathways that are anti-correlated within the intersection of both studies. Pathways that are in the intersection, and correlated in the same direction, are delegated to disjoint members of each study. By Fisher's Exact Test we compute the statistical significance of the overlap to be 0.05, with the background consisting of 450-curated pathways in KEGG. Since this meets the threshold of 0.05, we note that the overlap is barely significant.
Nonetheless, the pathway analysis elucidates drug interaction of ARAC and aplidin on the SKI-DLCL cell line, in silico.

3.3 Control Results

We proceeded next to examine our control results. What this means, is that we expect there to be no interaction between the drugs, because of the combination of different drugs on different cell lines. We performed a cross validation between the six studies mentioned in section 2.4. Here we explain two cross validations. The rest of the results are delegated to the supplementary section. We labeled our results as either being positive controls, if there was no statistical significance between the overlap, and negative controls if there was indeed a statistical significance between the overlap. The two examples mentioned here are indeed positive controls.

Figure 10 Connectivity Map and GSEA overlaps of Aplidin treated SKI-DLCL cell line and ATRA treated HL60 cell line.
Figure 10 shows the application of our method to the aplidin treated SKI-DLCL cell line and the ATRA treated HL60 cell line. Figure 11 represents our method applied to aplidin treated SKI-DLCL cell line and ATRA treated TEX cell line. Using the methods described above with regards to the ATRA treated HL60 cell line, we found the statistical significance of the connectivity map overlap to be 1, and the statistical significance of the GSEA overlap to be 0.14. Since both of these results are greater than the threshold of 0.05, we can conclude that the overlaps show no statistical significance with respect to the HL60 cell line. With regards to the TEX cell line, we found the statistical significance of the connectivity map overlap, and the GSEA overlap, to both equal 1. Since both results are again greater than the threshold of 0.05, we can conclude that the overlaps with respect to the TEX cell line showed no statistical significance. In both cases, not only is there no significance with respect to proxy drugs from connectivity map, but also there is no significance with respect to pathways analysis. Therefore we
could surmise that both drugs affect unrelated biological pathways. Researching the literature, we found no papers exposing the interaction of aplidin and ATRA. In order to conclude this would require in vivo validation. Since both results have no statistical significance, and since no papers validate the drug interaction, we would label this study as a positive control.
4. CONCLUSION AND DISCUSSION

4.1 Summary

Using principles from drug repurposing, we developed a proof of concept using a compendium of public gene expression data, in order to predict pharmacodynamic drug interactions. We utilized the connectivity map and GSEA, two ubiquitous computational techniques in translational bioinformatics, in order to develop our concept. We found one biologically validated dataset inferring drug interaction between aplidin and ARAC on the SKI-DLCL human lymphoma cell line. We were able to validate this drug interaction in-silico. We next tested our method on another data set that contains drug combinations applied to an individual cell lines, though this study does not comment on the interaction between the two drugs. We next proceeded to determine whether or not an arbitrary drug interaction could occur. Our preliminary methods show that this method could be applied to a larger number of datasets of biological relevance, in order to determine drug interactions.

4.2 Limitations

While our method has validity in the preliminary study, there are some limitations that should be discussed. The first limitation is that connectivity map has only 1309 drugs treated on four different cell lines. This is by no means comprehensive since there are hundreds of different cell lines, and potential drugs, at this moment in time. We did not find aplidin or ARAC in the connectivity map database. If it were the case that aplidin and ARAC were in the
database, this would have streamlined our proof-of-concept, thereby recapitulating the interaction of independent samples. Another limitation is that there are very few publications with biologically validated drug interactions and associated GEO datasets. In fact, our case study is the only data set that examines the biological interaction of aplidin on a respective cell line. Computational biologists, bioinformaticians, and laboratory-trained biologists would have to work in unison to generate the correct experimental design of pharmacodynamic interactions for which the mechanism remains unclear.

4.3 Future Work

We tested our method on only one control data set. In order to statistically verify our method, we would need to select 20 data sets at random, that are negative controls for drug interactions. Our hope is to reach a statistical significance of 5%, thereby showing no interaction with either ARAC, or aplidin, in GSEA and CMAP.

4.4 Conclusion

Polypharmacology remains to be one of the major challenges in prescribing medicine to patients. Drug molecules will often interact with multiple targets, resulting in unintended drug-target interactions that could result in side effects. The Food and Drug Administration (FDA) individually approve drugs. There are not subsequent tests to analyze how these drugs will interact. Development of in silico methods to test drug interactions is a key concept for
future work. Ongoing collaborations between the FDA and Dr. Lussier’s research group at University of Illinois at Chicago are using in silico methods in order to predict adverse effects of kinase inhibitors, and drug interactions. This proposed proof-of-concept is the first use of public gene expression data to systematically evaluate pharmacodynamics drug interactions in silico. It opens new avenues to design experiments for biologically validating adverse drug interactions, avert serious injury to patients, and protect useful drugs from limitations. It can also serve as a hypothesis-generating tool to predict biological mechanisms, such as pathways, implicated in specific pharmacodynamic drug interactions. Despite being its nascent stages, the methodology to predict drug interactions, in silico, will continue to grow in time.
5. SUPPLEMENTARY FIGURES

CONTROL STUDIES

**HL60 Cell Line**

- ATRA
- TCP+ATRA

Connectivity Map

- p-value = 1
- p-value of overlap without accounting for direction: 0.002
- p-value = 0.6824
- p-value of overlap without accounting for direction: 0.0015

**TEX Cell Line**

- ATRA
- TCP+ATRA

Connectivity Map

- p-value = 1
- p-value of overlap without accounting for direction: 3.13 e07
- p-value = 0.6959
- p-value of overlap without accounting for direction: 0.0011

*p*-values without accounting for direction are included in this figure. The reason being that the drugs and pathways are not anti-correlated. In fact addition of both drugs increases the effect in the same direction.
CASE STUDIES

**Connectivity Map**

- **ARAC**
  - SKI-DLCL cell line
  - Up: 30, Down: 20
- **ATRA**
  - HL60 cell line
  - Up: 5, Down: 2

**GSEA (KEGG)**

- **ARAC**
  - SKI-DLCL cell line
  - Up: 4, Down: 0
- **ATRA**
  - HL60 cell line
  - Up: 4, Down: 2

*p-value* = 0.046

- **Connectivity Map**
  - **ARAC**
    - SKI-DLCL cell line
    - Up: 26, Down: 19
  - **TCP+ATRA**
    - HL60 cell line
    - Up: 5, Down: 1

**GSEA (KEGG)**

- **ARAC**
  - SKI-DLCL cell line
  - Up: 1, Down: 0
- **TCP+ATRA**
  - HL60 cell line
  - Up: 26, Down: 39

*p-value* = 0.50

*p-value* = 0.189
CASE STUDIES (continued)

p-value=1

p-value=0.002
CASE STUDIES (continued)

Connectivity Map

APLIDIN
SKI-DLCL cell line

APLIDIN
SKI-DLCL cell line

ATRA
HL60 cell line

ATRA
HL60 cell line

p-value=1

p-value=0.15

Connectivity Map

APLIDIN
SKI-DLCL cell line

APLIDIN
SKI-DLCL cell line

TCP+ATRA
HL60 cell line

TCP+ATRA
HL60 cell line

p-value=0.003

p-value=0.7244
CASE STUDIES (continued)

p-value=1

p-value=0.12

p-value=1

p-value=0.45
6. REFERENCES


GURUNADH PARINANDI

Citizenship: United States

EDUCATION:

August 2012-Present
THE UNIVERSITY OF ILLINOIS AT CHICAGO – Chicago, IL. 60607
Masters of Science in Bioinformatics

September 2008
THE OHIO STATE UNIVERSITY – Columbus, OH. 43210
Bachelor of Science in Pure Mathematics

September 2000-May 2002
BERKLEY COLLEGE OF MUSIC – Boston, MA. 02215
Completed two years of study towards a Bachelor of Music degree in performance.

PUBLICATIONS:


WORK EXPERIENCE:

Graduate Research Assistant August 2012-Present
University of Illinois at Chicago
Section of General Internal Medicine
1819 West Polk Street, 342 CMW
Chicago, Illinois 60612-7323
Supervisor: Dr. Yves A. Lussier, (312)-355-0478, lussier.y@gmail.com. Contact may be made.
- Graduate training in bioinformatics. Took coursework in both theoretical and applied bioinformatics.
- Contributed to research papers by assisting team members with paper writing.

Visiting Research Specialist in Health Science May 2011-August 2012
University of Illinois at Chicago
Pulmonary, Critical Care, Sleep, and Allergy (MC 719)
Department of Medicine
COMRB 3099, 909 S. Wolcott Avenue
Chicago, Illinois 60612-7227
Supervisor: Dr. Joe G.N. Garcia, (312)-996-9450, jggarcia@uic.edu,
Dr. Yves A. Lussier, (312)-355-0478, ylussier@uic.edu. Contact may be made.

- Research training in Bioinformatics and Genomics
- Assisted PI's with statistical analysis of gene expression data and mathematical formulations of biomedical phenomena.

The Ohio State University, Columbus, OH. 43210
Supervisor: Dr. Sergei Chmutov, (419)-755-4287, chmutov@math.ohio-state.edu, contact may be made.

- A participating member of the “VIGRE” summer undergraduate working group focusing on knot and graph theory.
- Worked on the “Reconstruction Conjecture” for the case of Ribbon Graphs.

Summer Research Intern May 2002-September 2002
Johns Hopkins University Bayview Medical Center
Baltimore, MD. 21224
Supervisor: Dr. Joe G.N. Garcia, (312) 996-9450
jggarcia@uic.edu, contact may be made.

- Assisted in the creation of a genetic database for the “Human Genome Project” by cataloging various “Affymetrix” gene chips for certain biological systems.
- Assisted in the design of the Johns Hopkins Medical School’s “Human Genome Project” webpage.

AWARDS/ACADEMIC HONORS:
- Undergraduate Student Award in accomplishment- The Ohio State University, Department of Mathematics, May 2007
- Eagle Scout Award- Boy Scouts of America, 2000
- Jimi Hendrix National Guitar Competition, National Semi-finalist, 1999

SKILLS:

Computer: Microsoft Excel, Word, Powerpoint,
- Basic programming skills in Java, Perl, and R

Language: Studied four quarters of French at The Ohio State University. Speaking fluency in Telugu.