# DATA-DRIVEN COMPUTATIONAL MODELING OF THE STATE AND ARCHITECTURE OF THE BREAST CANCER KINOME

Kyla A.L. Collins

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Curriculum of Bioinformatics and Computational Biology.

Chapel Hill
2017

Approved by:

Timothy C. Elston

Gary L. Johnson

Shawn M. Gomez

Lee M. Graves

Zefeng Wang

**ABSTRACT**

Kyla A.L. Collins: Data-Driven Computational Modeling of the State and Architecture of the Breast Cancer Kinome
(Under the direction of Gary L. Johnson and Shawn M. Gomez)

The complex signaling in the kinome provides a unique insight into breast cancer, which is heterogeneous with many disease states or subtypes. The kinome has been implicated in many cancers and is highly targeted by inhibitor therapies because of its importance in cell proliferation and differentiation. High-throughput data sets using proteomics help characterize the kinome and allow quantification of the baseline and perturbed states of the kinome. These high-throughput experimental methods allow for quantification of kinases that are not well-studied, or are understudied.

In this thesis, I employ machine-learning techniques to distinguish between breast cancer subtypes using a functional proteomics data set and to demonstrate that the state of the kinome looks different in proteomic and sequencing data sets. Characterized, as well as understudied, kinases are identified as important features in stratifying unperturbed breast cancer subtypes. In addition, both understudied and characterized kinases respond dynamically across breast cancer subtypes in response to kinase inhibitor therapy treatment. Further, I developed computational methodologies to characterize the architecture of the kinome network and an optimization method for choosing effective combination therapies for cancer treatment. Public protein-protein interaction databases are compiled to create the comprehensive kinome network, consisting of only kinase to kinase interactions. The comprehensive kinome network is clustered to identify functional modules, or subnetworks, and some of these subnetworks are significantly enrichment for understudied and targeted kinases. In addition, the optimization proposed here provides a computational framework for choosing effective sets of inhibitors to use concurrently, i.e. combination therapies.

To Benjamin Collins, my husband and best friend, for lifting me up when I needed it, unending support and encouragement, and your unwavering love.

## ACKNOWLEDGEMENTS

I would like to thank my advisors, Gary L. Johnson and Shawn M. Gomez, for mentoring and guiding me through my graduate work. In addition, I would also like to thank the rest of my dissertation committee–Timothy C. Elston, Lee M. Graves, and Zefeng Wang–for your insights and support as I have completed my degree.

Current and previous members of Shawn M. Gomez's lab were invaluable when preparing for oral examinations, committee meetings, and writing manuscripts. Janet Doolittle-Hall spent countless hours listening to my ideas and guiding my computational work. Claire Hall and Franklin Blum performed tissue culture work directly for my project for validation studies for my manuscripts and I am extremely grateful they were willing and able to do these experiments. I would also like to thank Elliot Krause and Nick Carlton for helping with database formatting and data fitting for my models. Additionally, I would like to thank all of the members of Gary L. Johnson's lab. In particular, Tim Stuhlmiller, Jon Zawistowski, Steven Angus, and Sara Velarde were always available to answer any questions I had regarding experimental procedures and the biological context within which we were working.

Friends, family, and friends who are like family have made this work possible. My parents and extended family supported my decision to become a "professional student" and moving half-way across the country. My husband has dealt with many late nights, added stress to him, and listening to me when discussing my projects. So many of our friends have helped me take care of Carter, sat with me in "study halls" on the weekends and at night for me to complete this work. I thank you all from the bottom of my heart.

Finally, I would like to thank Allen Holder for inspiring me to go to graduate school and for showing me the beauty in applied mathematics. My life and career would be very different if our paths had not crossed and I am eternally grateful our lives have intersected.

## PREFACE

Chapter 2 of this dissertation has been submitted as a preprint to bioRXiv:

https://doi.org/10.1101/122739

## TABLE OF CONTENTS

LIST OF TABLES ..... xi

LIST OF FIGURES ..... xii

LIST OF ABBREVIATIONS ..... xiii

1 INTRODUCTION ..... 1

1.1 Breast Cancer Subtypes: Diagnosis, Prognosis, and Treatment ..... 1

1.2 Kinases and Their Role in Cancer ..... 2

1.3 Methods for Quantifying Kinase Function ..... 3

1.4 Clinical Treatment ..... 4

1.5 Computational Methodologies Used to Find Patterns in High – Throughput Data in Cancer ..... 5

1.6 Network Clustering Methods for Characterizing the Architecture of the Kinome ..... 6

1.7 Existing Mathematical Frameworks for Choosing Combination Therapies ..... 6

1.8 Outlook on Utilizing Computational Approaches Clinically to Aide in Diagnosis and Treatment of Cancers ..... 7

1.9 Thesis Contributions ..... 8

2 PROTEOMIC ANALYSIS DEFINES KINASE TAXONOMIES SPECFIC FOR SUBTYPES OF BREAST CANCER ..... 9

2.1 Introduction ..... 9

2.2 Results ..... 10

2.2.1 Multiplexed Kinase Inhibitor Beads capture kinases from every subfamily and provide a means to assay understudied kinases ..... 10

2.2.2 Integrating understudied and well-characterized kinases by kinome proteomic profiling defines breast cancer subtypes ..... 12

2.2.3 Kinase MIB-binding activity is independent from mRNA expression level . . . 13

2.2.4 Kinome profiles accurately define tumor biopsies . . . 15

2.2.5 A functional interaction network of MIB-binding kinases . . . 17

2.2.6 Kinome MIB-binding profile and response to drug perturbation . . . 17

2.3 Discussion . . . 19

2.4 Materials and Methods . . . 20

2.4.1 MIB affinity chromatography . . . 20

2.4.2 Mass Spectrometry and Analysis . . . 21

2.4.3 RNA-seq . . . 21

2.4.4 Data analysis . . . 22

2.4.5 Comparison of MIB-binding to transcript abundance . . . 23

2.4.6 Prediction of subtypes . . . 23

2.4.7 Network Analysis . . . 23

3 CHARACTERIZATION OF THE ARCHITECTURE OF UNDERSTUDIED AND TARGETED SUBNETWORKS WITHIN THE KINOME . . . 38

3.1 Introduction . . . 38

3.2 Biological Data . . . 39

3.2.1 Kinome Network Databases . . . 39

3.2.2 Inhibitor Databases . . . 39

3.2.3 Defining Understudied and Targeted Kinases and Subnetworks in the Kinome . . . 40

3.3 Methods . . . 40

3.3.1 Subnetwork Identification . . . 40

3.3.2 Comparison of Subnetwork Identification Across Algorithms . . . 41

3.3.3 Statistical Assessment of Understudied and Targeted Subnetworks in the KIN . . . 42

3.4 Biological Assumptions . . . 42

3.5 Results & Discussion . . . 42

3.5.1 Clustering the KIN ..... 42
3.5.2 Functional Annotation of Subnetworks in the KIN ..... 43
3.5.3 Understudied Kinase Enrichment in Subnetworks in the KIN ..... 44
3.5.4 Targeted Kinase Enrichment in Subnetworks in the KIN ..... 44
3.5.5 Integration of Functional Annotation with Understudied and Targeted Enrichment ..... 45
3.6 Conclusions ..... 45
4 OPTIMIZATION FRAMEWORK FOR RATIONALLY CHOOSING COMBINATION DRUG THERAPIES ..... 53
4.1 Introduction ..... 53
4.2 Methods ..... 54
4.2.1 Kinome Network Databases ..... 54
4.2.2 Kinome Subnetwork Identification ..... 54
4.2.3 Inhibitor Databases ..... 55
4.2.4 Optimization of Targeted Drug Therapy Combination ..... 55
4.2.4.1 Costs and Objective Function for Optimization ..... 56
4.2.4.2 Constraints for Optimization: Subnetworks ..... 57
4.2.4.3 Constraints for Optimization: Targeting of Specific Nodes ..... 57
4.2.4.4 Constraints for Optimization: Inhibitor Interactions ..... 58
4.2.4.5 Solving with Linear Programming ..... 58
4.2.5 Access to Software/Code and Example Datasets ..... 59
4.3 Results ..... 59
4.3.1 Example #1: Targeting MTOR, TESK1 and ERBB2 ..... 59
4.3.2 Example #2: Subnetwork Targeting ..... 60
4.3.3 Example #3: HER2+ Breast Cancer ..... 60
4.3.4 Example #4: Flavopiridol and Leukemia ..... 61
4.4 Discussion ..... 62
5 CONCLUSIONS ..... 69

5.1 Summary of Thesis Contributions . . . . . 69

5.2 Impact on Cancer Biology . . . . . 69

5.3 Challenges . . . . . 70

5.4 Future Directions . . . . . 70

REFERENCES . . . . . 72

## LIST OF TABLES


2.1 Understudied kinases . . . . . 25

2.2 Specificity, sensitivity, and precision across subtypes in Leave-one-out Cross Validation . . . . . 25

2.3 Uniquely bound kinases by breast cancer subtype . . . . . 27

2.4 Subnetworks in Kinome . . . . . 28

3.1 Comparison of Clustering Methods . . . . . 46

3.2 Significantly Enriched GO Biological Processes in FastGreedy Clusters . . . . . 47

3.3 Significantly Enriched GO Biological Processes in SpinGlass Clusters . . . . . 48

3.4 Enrichment for Understudied and Targeted Kinases in KIN . . . . . 49

4.1 Example #1 Results . . . . . 64

4.2 Example #2 Results . . . . . 65

4.3 Example #3 Results . . . . . 66

4.4 Example #4 Results . . . . . 66

## LIST OF FIGURES


2.1 Assessment of multiplexed kinase inhibitor beads (MIBs) for kinase capture across breast cancer subtypes. . . . . 27

2.2 MIB/MS kinome profiling assigns breast cancer cell lines to functional subtypes. . . . . 30

2.3 Overall MIB-binding and mRNA expression levels are not correlated. . . . . 32

2.4 Baseline kinome of cell lines and tumors across breast cancer subtypes. . . . . 33

2.5 Subnetworks in the functional kinome. . . . . 35

2.6 Kinome drug response overall and by subnetwork. . . . . 37

2.7 Sensitivity, specificity, and precision across subtypes to determine number of features needed for classification . . . . . 37

3.1 Methods for Clustering Networks . . . . . 50

3.2 Results for Clustering Kinome into Subnetworks . . . . . 51

3.3 Understudied and Undertargeted Subnetworks in the Kinome . . . . . 52

4.1 Optimization and Network Clustering Methodology Description . . . . . 67

4.2 Results of Optimization in Examples . . . . . 68

## LIST OF ABBREVIATIONS

| | |
|---|---|
| ADME | Absorption, Distribution, Metabolism, and Excretion |
| AGC | Protein Kinase A, G, and C Family |
| AML | Acute Myeloid Leukemia |
| ARI | Adjusted Rand Index |
| BioGRID | A General Repository for Interaction Datasets |
| cDNA | Complementary Deoxyribonucleic Acid |
| CLL | Chronic Lymphocytic Leukemia |
| CMGC | CMGC Kinase Family |
| CML | Chronic Myeloid Leukemia |
| Cas9 | CRISPR-Associated Protein-9 Nuclease |
| CRISPR | Clustered Regularly Interspaced Short Palindromic Repeats |
| DIP | Database of Interacting Proteins |
| EGFR | Epidermal Growth Factor Receptor |
| EpCAM | Epithelial Cell Adhesion Molecule |
| ER | Estrogen Receptor |
| ERBB2 | Human Epidermal Growth Factor Receptor 2 |
| FDA | Food & Drug Administration |
| GO | Gene Ontology |
| HER2 | Human Epidermal Growth Factor Receptor 2 (Same as ERBB2) |
| HIPPIE | Human Integrated Protein-Protein Interaction rEference |
| HPRD | Human Protein Reference Database |
| I2D | Interologous Interaction Database |
| IC50 | Half-maximal Inhibitory Concentration |
| IHC | Immunohistochemistry |
| KEGG | Kyoto Encylopedia of Genes and Genomes |
| KIN | Kinome Interaction Network |
| LC-MS/MS | Liquid Chromatography-Mass Spectrometry/Mass Spectrometry |
| LINCS | Library of Integrated Network-based Cellular Signatures |

| | |
|---|---|
| LOOCV | Leave One Out Cross Validation |
| LP | Linear Programming |
| MIBs | Multiplexed Inhibitor Beads |
| MIB/MS | Multiplexed Inhibitor Beads coupled with Mass Spectrometry |
| MINT | Molecular INTeraction Database |
| MTOR | Mechanistic Target of Rapamycin |
| NeTOPT | Network Targeting using OPTimization |
| NMI | Normalized Mutual Information |
| ODE | Ordinary Differential Equation |
| PAM50 | Prediction Analysis of Microarray 50 |
| PANTHER | Protein ANalysis THrough Evolutionary Relationships |
| PC | Principal Component |
| PCA | Principal Components Analysis |
| PDR | Progressive Disease Rate |
| PPIN | Protein-Protein Interaction Network |
| PR | Progesterone Receptor |
| Reactome | Reactome Pathway Database |
| RNAi | Ribonucleic Acid Interference |
| RNA-seq | Ribonucleic Acid Sequencing |
| RPA | ‘Reverse-phase’ Protein lysate microarray Assay |
| RR | Response Rate |
| SJD | Split-Join Distance |
| SNX | Subnetwork X |
| SVM | Support Vector Machine |
| TK | Tyrosine Kinase Family |
| TNBC | Triple Negative Breast Cancer |
| URI | Unadjusted Rand Index |
| VI | Variation of Information |

# CHAPTER 1
# INTRODUCTION

## 1.1 Breast Cancer Subtypes: Diagnosis, Prognosis, and Treatment

Breast cancer is a highly heterogeneous disease that is divided into subtypes [1, 2, 3]. These subtypes were discovered using classical experimental methods in combination with molecular profiling. Historically, breast cancer has been divided into subtypes using histology, gene expression (microarrays, next-generation sequencing), and mutations in the genome [4]. Breast cancer subtypes are separated into two major groups: HER2-enriched/Luminal and Triple-Negative. These major groups have different receptor statuses with respect to three different receptors (estrogen receptor: ER, progesterone receptor: PR, Human Epidermal Growth Factor Receptor 2: HER2) and clinical prognosis [4, 5, 6, 7].

The Luminal A and B subtypes are the most prevalent types (approximately 70%) of breast cancer in the general population [4, 7]. These tumors are derived from the luminal cells on the inner lining of the mammary ducts in the breast, hence their name [7]. Tumors in these subtypes can be ER- and/or PR-positive but are usually HER2-negative [4, 7]. Of all of the breast cancer subtypes, Luminal tumors have the best prognosis, highest survival, and lowest recurrence rates [4]. These tumors are generally driven by the PI3K/AKT/MTOR pathway [7, 8, 9].

The HER2-enriched subtype accounts for approximately 15% of all breast cancers [4]. These tumors are ER- and PR-negative but are positive for HER2. Treatment efficacy with chemotherapy greatly increased with the use of HER2 inhibitors. Current standard of care for this tumor type involves treatment with a combination of chemotherapy and trastuzumab (Herceptin), a HER2/ERBB2 inhibitor [10].

Triple-negative breast cancer is a particularly aggressive group of subtypes that tends to have poorer prognosis [4]. This group is divided into two different subtypes: Basal-like and Claudin-low [11]. Both Basal-like and Claudin-low subtypes cannot be treated with hormone therapy due to

their negative receptor status and targeted therapies have yet to be FDA approved for these specific subtypes [12]. The tumors are ER-, PR-, and (mostly) HER2-negative and are derived from the basal cells on the outer lining of the mammary ducts [4, 5, 6].

While these subtype classifications have helped tremendously in the identification of this heterogeneous disease, treatment based on these data have fallen short [3, 7, 13, 14]. The drug therapies designed to treat this set of diseases largely target the kinome, namely at the level of post-translational modifications. The direct changes occurring in the cell after treatment with targeted therapies are happening down stream of gene expression changes in the cell. These targeted therapies are mostly composed of kinase inhibitors, drugs that inhibit the activity of particular kinases in cancer cells. Studies of the level at which these therapies are acting in cells are very limited but can provide unique insights into establishing effective therapies for each breast cancer subtype and, eventually, individual patients.

### 1.2 Kinases and Their Role in Cancer

Protein kinases are key components of cellular signaling pathways, mediating a vast array of cellular processes. With over 500 members, this family forms a core component of intracellular signaling by interacting with each other as well as other proteins to form a network of interactions. While still poorly understood, dysregulation of this network of kinases, or the kinome, is implicated in many diseases, with cancer perhaps being the prototypical example. For instance, key kinases have been labeled oncogenic due to their aberrant regulation [15]. Similarly, protein kinases are drug targets of great interest, with numerous clinical trials underway and several kinase inhibitors already having been approved for use in clinical therapies [16]. While kinase inhibitors have been shown to be effective for certain cancers, some of the cancers treated circumvent the therapies through "reprogramming" [17].

Rich systems exist downstream and upstream of well-known kinases, e.g. networks of serine/threonine and tyrosine kinases, many of which are understudied and untargeted by chemical probes. Kinases deemed understudied or untargeted are crucial in these signaling networks and help to mediate a cell's response to a drug therapy or therapies [18]. Lack of information regarding function, interactions with other proteins, etc. for a particular kinase does not dictate how important a

particular kinase is in the signaling network. Understanding how network perturbation can reestablish normal behavior or cell death is essential for the development of potential new therapeutic targets and cancer treatment strategies.

### 1.3 Methods for Quantifying Kinase Function

While its importance is recognized, our understanding of the architecture and dynamics of the kinome in cancer and in response to therapy is severely limited. This is in part due to the difficulty of experimentally characterizing its behavior. Traditional experimental techniques typically allow for only a small fraction of the entire kinome to be studied at any given time ($\sim$5-10%) [17]. The most common experimental method for quantifying the levels of proteins in cell lysates is via Western blotting [19]. This method requires the use of specific antibodies that bind to particular forms of proteins if any posttranslational modifications are to be detected. One of the major drawbacks of this method is that each protein of interest must have a specific antibody, which may not exist, to be able to quantify its levels in the cell lysate [19]. Additionally, each lane on the gel for this method is used for a single protein, making this method very labor intensive (or impossible) when trying to analyze more than a handful of proteins [19]. Another method for quantifying protein levels is the 'Reverse-phase' Protein lysate microarray Assay (RPA), which is a method similar to gene expression microarrays but for proteins [20]. This method allows many proteins to be quantified at once, which is an improvement on the Western blot, but still requires the use of specific antibodies to pull down proteins of different forms, similar to the Western blot [20].

Gary Johnson's lab has developed a technology, Multiplexed Inhibitor Beads coupled with Mass Spectrometry (MIB/MS), that allows for a global capture of kinases that bind to the beads in a sample. In the MIB/MS method, kinase inhibitors, which bind preferentially to functional forms of kinases, are bound to sepharose beads. Cell lysates are run over the beads to preferentially pull down the functional kinase forms in the sample. The beads are washed off and the proteins remaining are run on LC-MS/MS to determine the species of the proteins that bound to the MIBs [17, 21]. The outcome of one of these runs allows for a global snapshot (over 85%) of the cell's functional kinome, which means we can analyze any changes in posttranslational modifications to kinases.

While the MIB/MS technology gives the opportunity to get a global view of the functional kinome, it does have some aspects that should be considered when analyzing and interpreting data from it. The MIBs take advantage of the affinity of the proteins to the kinase inhibitors that are bound to the beads to pull down the functional kinases in the cells. This does preferentially pull down functional kinases in the lysate but other proteins do compete for binding to the beads. Because of this, the final quantification of the levels of functional kinases is directly affected by the other types of proteins in the cell lysate. Another consideration is the competition for binding to the beads among just the kinases. The beads with varying kinase inhibitors, each having its own binding profile across the kinases, are pooled together. Each kinase inhibitor has a different affinity to each of the kinases in the lysate so the kinases compete with each other to bind to the beads to be in the final quantification. While these aspects do need to be considered when interpreting the quantification using this technology, it is important to note that all biological quantifications are biased in some way but that does not make them useless in providing biological insight.

### 1.4 Clinical Treatment

Current clinical treatments using single therapeutic agents are not very effective across the breast cancer subtypes due to poor drug targets or cancer tissue/cells reprogramming [22]. Polypharmacology and drug combinations are fields with high potential to address these issues. Many drug combinations are currently in clinical trials to test their efficacy over their individual components and a few have been successful already [23, 24]. Drug combinations differ from polypharmacology because they are a set of drugs administered together, not a single agent with multiple targets [25, 26]. Currently, there is no systematic way to select inhibitors to compose effective combination therapies (the use of multiple kinase inhibitors together) [7, 27]. Serial treatment of drugs has repeatedly led to failure but initial treatment with multiple drug therapies at the same time can do significantly better [22].

## 1.5 Computational Methodologies Used to Find Patterns in High – Throughput Data in Cancer

Computational approaches for characterizing and analyzing intracellular mechanisms in cancer have a strong history, particularly with regard to gene expression. The identification of specific gene expression profiles from microarray data has been used to discriminate breast cancer subtypes using unsupervised clustering with correlations to potential response to treatment [28, 29]. These approaches are particularly strong with breast cancer subtypes that have genetic mutations specifically tied to changes in gene expression [28, 29]. However, several breast cancer subtypes are not well characterized by these gene expression signatures because the difference between cancerous cells and normal tissue is in the activation levels of multiple kinases rather than changes in gene expression [29]. Current inhibitor therapies largely rely on alteration of protein kinase activity rather than kinase expression. Thus, understanding the activity state of the kinome is critical for both cancer subtype characterization and understanding systematic effects of inhibitors. There is a great need to develop approaches for defining the kinome through gene expression profiles and kinase activation data. Many computational methods (supervised and unsupervised) exist for analyzing large data sets like these.

Principal Component Analysis (PCA) is a dimension-reduction method commonly used in the analysis of high-dimensional data sets [30]. PCA projects data onto a lower dimension space for classification and interpretation. PCA ranks the principal components based on the amount of variance accounted for in the data by a particular component. The first principal component (PC) will account for the most variation in the data set, the second component will account for the next highest amount of variation, and so on. In addition to PCA, regression methods can be used to determine which features (i.e. proteins/kinases, genes) are important for describing the variation in the data set [31, 32].

To better understand what kinase activities discriminate between subtypes, a variety of feature ranking approaches are available [33]. Generally, the metrics generated by such approaches measure how much information a particular feature adds to a classification that discriminates between two classes. The features that provide the most information to distinguish the two classes in question will be ranked at the top.

To classify as well as investigate how many features are needed to distinguish subtypes, Support Vector Machine (SVM), a kernel-based machine-learning methodology that determines the separation of subtypes based on features in a data set, can be used [34]. This thesis will primarily focus on kinases as features. SVM can be utilized in combination with a "leave-one-out cross-validation" (LOOCV) strategy to assess precision, sensitivity, and specificity metrics for the classifier [35, 36]. This SVM can be trained on data and tested on any future samples collected.

### 1.6 Network Clustering Methods for Characterizing the Architecture of the Kinome

Many types of network clustering algorithms exist and the types of algorithms encompass both deterministic and nondeterministic algorithms [37]. The work presented here will focus on two algorithms, FastGreedy (deterministic) and SpinGlass (nondeterministic). Because FastGreedy is a deterministic algorithm, the results of clustering a network will always be the same. This algorithm takes a bottom-up approach by starting with individual nodes (subnetworks) and merging them based on finding the subnetwork configuration that gives the maximal modularity score for all subnetworks [38]. SpinGlass is a nondeterministic algorithm developed originally in statistical physics for clustering nodes on a lattice. This algorithm is also a bottom-up approach but, unlike FastGreedy, the result of running this algorithm on the same network can vary from run to run because seed nodes are randomly chosen at the beginning of a run [39, 40, 41].

### 1.7 Existing Mathematical Frameworks for Choosing Combination Therapies

The concept of intelligently choosing individual agents to combine to make an effective combination therapy is not new. A variety of studies to this effect have been done within the last few decades. The experimental data informing these decisions has recently grown to the point of computational usefulness on a large scale [42]. Choosing drug combinations is experimentally intractable. The FDA has approved over 1500 compounds; therefore, to test all of these drugs individually in all biological indications would be experimentally and logistically challenging even before the added complexity of drug combinations [43, 44, 45]. Some experimental design work has helped to reduce time and money spent on experiments needed to search the possible drug combinations [43]. Unsu-

pervised clustering of high-dimensional data has helped identify important genes/proteins in cancer, in particular [46].

In addition to unsupervised clustering of data, some methods exist to predict effective combination therapies for specific contexts [47]. Most methods focus on a small number of proteins (usually a pathway). In these methods, a single therapy is selected and pathway information is used to decide on a therapy to add to the original treatment [3, 22]. Many computational methodologies have been developed to help address this problem that use a variety of quantitative techniques. These techniques include quantifying individual drugs and drug combinations using synergy measures (i.e. combination indices) [48, 49, 50], pathway modeling using Ordinary Differential Equations (ODEs) or logic modeling [51, 52], network analyses [53, 54], and optimization [55, 56].

One shortcoming of previous methods is the exclusion of drug information regarding absorption, distribution, metabolism, and excretion (ADME) processes when combining drugs [42]. This does rely on experimental data to inform the computational methods, which can slow down computational assistance, but this information is critical when combining drugs for patient use. Another factor that has not been incorporated into existing computational methodologies is off-target effects of drug treatments [42]. [56] does incorporate some of these shortcomings of other methods by using optimization to minimize off-targets and, to an extent, adverse drug effects but fails to incorporate other clinically relevant aspects of using certain drugs (i.e. monetary cost, side effect information, etc.).

## 1.8 Outlook on Utilizing Computational Approaches Clinically to Aide in Diagnosis and Treatment of Cancers

Prediction Analysis of Microarray 50 (PAM50) is an example of computational work that is being applied in the clinic for diagnosis of the breast cancer subtypes [57]. While PAM50 is a step in the correct direction, this diagnostic tool does little to provide information regarding which treatment would be most effective. PAM50 performs well to be able to distinguish the Luminal A and B subtypes in patients but has trouble detecting accurately TNBC tumors [58]. This is largely due to the fact that Basal-like and Claudin-low tumors are truly different diseases from the Luminal/HER2-enriched subtypes [58]. Additionally, heterogeneity within the same class of tumors between patients,

not to mention inter-tumor heterogeneity within one patient or intra-tumor heterogeneity, makes it difficult to determine a generalized effective therapy for a particular class of tumors. Personalized medicine is the next "big" therapeutic approach for treating a variety of cancer types.

Due to the amount of data that needs to be collected and analyzed, computational approaches will allow for the identification of effective therapies on a patient-by-patient basis. Covering the entire realm of possible therapeutics (and all combinations thereof) by simply performing more experiments with many different therapies (and combinations of therapies) is not feasible. Using computational and modeling methods to explore the possible therapeutics for a given tumor class is much more feasible and can guide experiments to the most likely effective options for therapy.

## 1.9 Thesis Contributions

In this work, I use computational methods to assess the state of the kinome before and after inhibitor treatment, to characterize the architecture of the kinome network, and to choose effective combination therapies. Chapter 2 uses machine-learning techniques to identify functional kinase profiles that are important in distinguishing breast cancer subtypes. Chapter 3 presents a method for clustering the kinome network into functional modules (subnetworks) and results of enrichment analyses in the identified subnetworks. In Chapter 4, an optimization method for choosing combination therapies based on relevant biological information is proposed. Overall, this work provides quantitative characterization of the kinome and its importance in breast cancer diagnosis and treatment.

# CHAPTER 2

# PROTEOMIC ANALYSIS DEFINES KINASE TAXONOMIES SPECFIC FOR SUBTYPES OF BREAST CANCER

## 2.1 Introduction

Breast cancer has two primary subtypes that includes luminal A and B as well as the majority of HER2+ breast cancers and triple negative breast cancer (TNBC) that can be divided into basal-like and claudin-low [1]. Interestingly, basal-like breast cancer using molecular taxonomies is as different from luminal and HER2+ breast cancers as lung cancer, leading to the proposal that basal-like breast cancers are in fact a unique disease [59, 60]. Estrogen and progesterone receptor dependence and HER2 addiction define vulnerabilities in luminal/HER2+ breast cancers. However, in basal-like and claudin-low triple negative breast cancer, there are no oncogenic drivers that define a vulnerability that can be therapeutically targeted.

Even with the growing databases of genomic information for breast cancers, it is often still unclear how molecular taxonomies translate to phenotype. Additional methods characterizing proteomic taxonomies are needed to understand signaling networks, particularly of protein kinases due to their druggability. Important for this analysis of the breast cancer kinome is a characterization of understudied kinases, representing nearly half of the kinome and lacking essential functional characterization as well as molecular tools for their manipulation and study [18]. These understudied kinases need to be functionally integrated into kinase networks for a global understanding of kinome dynamics to be achieved both at baseline and in response to perturbation.

To this end, we have developed methods using multiplexed inhibitor beads (MIBs) coupled with mass spectrometry (MIB/MS) that have the ability to bind and identify a large percentage of kinases in the human kinome [17, 21]. By RNA-seq, most cell lines express 350 or so kinases and our MIB-binding profiling captures a significant percentage of the expressed kinome [61]. In the current study we have compiled the baseline kinase MIB-binding profile using MIB/MS for

15 cell lines across breast cancer subtypes in addition to patient tumors. Using feature selection methodologies, it was possible to define kinase taxonomies for breast cancers based on the MIB/MS profile of 50 kinases among the kinases captured by MIB/MS that includes understudied protein kinases, lipid and metabolic kinases. Using the baseline MIB-binding state in a machine-learning framework further allows accurate classification of breast cancer in both cell lines and primary tumors. Kinases identified within these distinguishing profiles are distributed throughout the kinome, representing multiple subnetworks with a significant representation of understudied kinases. The findings demonstrate that determining the functional kinome based on MIB-binding has prognostic value in defining the integration of signaling networks that is not currently possible using genomic strategies.

## 2.2 Results

### 2.2.1 Multiplexed Kinase Inhibitor Beads capture kinases from every subfamily and provide a means to assay understudied kinases

Multiplexed kinase inhibitor beads (MIBs) are a set of Sepharose beads each with a specific covalently-attached kinase inhibitor [17, 62]. Coupling MIB gravity-flow affinity chromatography with mass spectrometry (MIB/MS) provides the ability to capture and identify kinases from whole cell lysates on a kinome scale. Binding of kinases is dependent on the functional expression of the kinase and affinity for the different immobilized inhibitors. To determine the inhibitor bead selective distribution of bound kinases, we assayed kinase capture by six different inhibitors individually covalently coupled to Sepharose beads [21, 62]: CTx-0294885, VI-16832, PP58, Purvalanol B, and two custom synthesized molecules, UNC-8088A and UNC-2147A. Four cell lines representative of breast cancer subtypes: HCC1806 (basal-like), SUM159 (claudin-low), MCF7 (luminal), and SKBR-3 (HER2-enriched) were used for analysis (Figure 2.1A). Of these, CTx-0294885 (CTx) and VI-16832 (VI) captured the most total kinases (265 and 254, respectively) and the most unique kinases (32 and 29, respectively). The other four beads bound a lesser number of kinases (PP58, 194 kinases; Purvalanol B, 164; UNC-8088A, 162; UNC-2147A, 130, Figure 2.1B). Although UNC-8088A binds the fewest unique kinases (only five), these include the atypical bromodomain and extraterminal (BET) domain-containing family of chromatin readers BRD2, -3, and -4 [63].

Hierarchical clustering of identified kinase peptides shows each bead binds a unique set of kinases and UNC-2147A displays the most distinct binding profile selectively enriching the AGC kinases (Figure 2.1C).

Understudied kinases make up a large proportion of kinases captured across the inhibitor beads, ranging from 23-34% of all kinases captured for any individual bead (Figure 2.1D). Understudied kinases [18] represent approximately 40% (Table S2) of the expressed kinome. Characteristics of understudied kinases includes: i) integration of the protein kinase in signaling networks is poorly defined, ii) function and/or regulation is poorly defined, iii) activation loop phospho-antibodies and/or IHC grade antibodies may not exist, iv) lack of selective chemical tools for use in characterization of function (e.g., small molecule inhibitors), v) RNAi and CRISPR/Cas9 for knockout/altered expression and cDNAs for overexpression may be primary tools, vi) kinase knockout or altered expression may not provide readily assayable phenotypes (e.g., growth, migration, apoptosis or in vivo function in mouse organ physiology).

Across all MIB/MS runs, 381 kinases in total were identified. Of these, 35 are metabolic and lipid kinases, 346 are protein kinases of which 142 can be considered understudied (41% of protein kinases identified) (Figure 2.1E). The overall distribution of kinases bound indicates CTx and VI are clearly pan-kinase inhibitors (Figure 2.1F, circle size proportional to number of unique peptides identified per kinase). Purvalanol B also binds kinases across families but to a lesser extent. PP58 has some preference for tyrosine kinases (TKs), and UNC-8088A has preference for TKs, CMGC, and atypical kinases over other families. UNC-2147A, designed for interaction with the binding pocket of AKT, has a strong affinity for AGC kinases lacking in most of the other five kinase inhibitors. CTx, VI, and PP58 have a strong affinity for the PRKDC (DNA-PK) not seen with the other three inhibitors. All inhibitor beads display high affinity for many understudied kinases (green circles and text). The most-highly captured understudied kinases across the four cell lines were GAK, SLK, MRCKB, AAK1, TBK1, and NEK9.

Kinases known to be oncogenic drivers in general and/or nodal signaling kinases display anticipated MIB-binding profiles across the different breast cancer subtype cell lines (Figure 2.1G). For example, SKBR3 (luminal HER2+) and MCF7 (luminal) cells have abundant AKT1/2 MIB-binding. Other well-characterized kinases are highly represented in a specific cell line, such as EGFR and FAK1 in HCC1806, EPHA2 and UFO (AXL) in SUM159, IGF1R and KS6B1 (p70 S6K) in

MCF7, and HER2/ERBB2 and TGFBR1 in SKBR-3 cells. Several understudied kinases also show high selectivity in functional MIB-binding including CDK13, DMPK, SIK3 and TESK1 in MCF7 and CLK4, CDK14 and NLK in SUM159 cells (Figure 2.1H). Figure 2.1I shows kinases whose MIB-binding is greatest in each of the four cell lines, proportional to the number of unique peptides identified. Unsupervised hierarchical clustering illustrates the differences in MIB-binding throughout the kinome for each cell line (Figure 2.1J). These findings indicate the four cell lines display a unique MIB/MS binding profile for both well-characterized and understudied kinases.

### 2.2.2 Integrating understudied and well-characterized kinases by kinome proteomic profiling defines breast cancer subtypes

We characterized the baseline kinome of 15 breast cancer cell lines representing the four major breast cancer subtypes defined by gene expression profiles [1]. Cell lysates were passed over an affinity column composed of the six kinase inhibitor beads and processed for LC-MS/MS (Figure 2.2A). Using label-free peptide quantitation measurements a total of 360 kinases were identified having at least 3 unique peptides. Claudin-low and basal-like cells (TNBC) are readily distinguished from HER2-enriched/luminal cells by MIB profiling of the cellular kinomes shown by consensus clustering in Figure 2.2B. The basal-like HER2-amplified cell line HCC1954 clusters with basal-like lines by kinome profile and is thus separated from the luminal HER2+ lines. Interestingly, the SKBR-3 HER2-enriched cell line shows an intermediate clustering between HCC1954 and other HER2+/luminal cell lines, and a previous report demonstrated SKBR-3 patterns as basal-like in functional RNAi screens [64]. Hierarchical clustering of kinases further separated cell lines with the claudin-low phenotype, SUM159, MDA-MB-231 and MDA-MB-468 (basal-like), showing the greatest difference from other cell lines (Figure 2.2C). SUM229 cells have two subpopulations, a basal-like EpCAM positive/E-cadherin positive (SUM229pos) and a claudin-low EpCAM negative/E-cadherin negative population (SUM229neg). The two populations are genomically similar by exome sequencing, but differ epigenetically [65] and cluster together based on their kinome MIB-binding profile (Figure 2.2C).

Principal components analysis (PCA) of baseline MIB-binding kinase profiles revealed significant differences between subtypes within the first principal component, clearly separating triple-negative from HER2-enriched and luminal cell lines (Figure 2.2D). Further separation of

the triple-negative group into claudin-low and basal subtypes is also readily achieved. Appreciable separation of HER2-enriched cell lines from luminal cell lines is observed in the second principal component, as is that of the basal-like/HER2-amplified cell lines from the basal-like and claudin-low lines. A loadings plot, which defines relationships between MIB-binding for each kinase, highlights those kinases with significant variation within subtypes, with numerous understudied kinases being apparent (Figure 2.2E). Examples of understudied kinases with differences in MIB-binding among cell lines include ADCK1, PKN3, STK17A and TESK1. Similarly, well-characterized kinases known for their relevance in breast cancer are observed, such as ERBB2, EPHA1, MET and TGFBR2.

Supervised differential expression analysis of MIB-captured kinases from claudin-low/basal-like (TNBC) versus HER2/luminal cell lines defined several statistically significant differences (Figure 2.2F). Multiple Ephrin receptors (EPHA2/A7/B2) and members of the TGF-beta superfamily (TGFBR2, ACVR1) are among the kinases most associated with TNBC while ERBB3 and RET are over-represented in HER2+ and luminal cell lines. Many understudied kinases display higher MIB-binding in HER2+/luminal cells, including DMPK, ADCK1, and TESK1. Individual plots for selected kinases are shown in Figure 2.2G, showing distinctive patterns of MIB-binding across subtypes.

### 2.2.3 Kinase MIB-binding activity is independent from mRNA expression level

Our results clearly demonstrate that kinase MIB-binding displays strong variation across breast cancer subtypes. Global gene expression measurements have similarly shown subtype-specific dynamics, with expression of selected gene sets being utilized in subtype determination and diagnosis [59, 66, 67]. We compared baseline RNA-seq measurements with corresponding MIB-capture of protein kinases to assess the relationship between transcript abundance and functional kinome behavior. Comparison of kinase transcript expression with MIB-binding using label-free quantitation of kinase peptide abundance for each subtype showed relatively low correspondence, with correlation coefficients of 0.25 or below, implying that less than 7% of the observed variation between MIB-binding and RNA abundance in breast cancer subtypes is explained through this relationship (Figure 2.3A–D). These results are consistent with earlier work that found the average correlation between gene expression and protein abundance in TCGA colorectal cancer samples to be approximately 0.47, with a lesser correlation of 0.23 when comparing gene expression and protein variation [68]. The low

correlation between RNA-seq and MIB-binding suggests that the use of MIB/MS provides a picture of kinome behavior complementary to that provided through expression measures. In particular, these results support the potentially significant role of post-translational and transcriptional regulation in kinome dynamics [69, 70]. The relationship between these two measures is shown as a scatter plot for cell lines representing the basal subtype in Figure 2.3A (left panel), with the Z-scores for both read count and MIB-binding plotted on the x- and y-axes, respectively, and further emphasize the lack of strong correlation between these two measures of kinase behavior. Kinases along the gray line (y=x) represent those where transcript expression directly corresponds to MIB-binding (i.e. a unit increase in expression corresponds to a unit increase in MIB-binding). A point above the gray line in the scatter plot indicates a kinase that has more MIB-binding than expected given the expression level of that kinase in that subtype; these kinases are referred to as "functionally dominant" for their increased MIB-binding relative to transcript expression. Examples of functionally dominant kinases in the basal-like subtype are KCC2A, HCK, and EPHA5; in the claudin-low subtype KCC2A, MK10, and FGFR2; in the HER2-enriched subtype KCC2A, HCK, and EPHB1; and in the luminal subtype KCC2A, HCK, and FYN. Similarly, points below the gray line indicate kinases that have less MIB-binding than expected given the expression level of that kinase and are labeled as "expression dominant" kinases. Examples of expression dominant kinases include KPYM, TIF1B and K6PL. The significant number of both functionally and expression dominant kinases common across several subtypes is readily detected by MIB/MS and suggests a post-transcriptional regulation of different kinases likely related to differential covalent regulatory modifications and protein stability. Such proteomic behavioral properties can not be detected by RNA-seq. In addition, recombinant kinases often used to profile on-target/off-target inhibitor profiles are not representative of endogenous kinases in cell lysates that have associated regulatory subunits and post-translational modifications [71, 72]. These data describing endogenous kinases captured in cell lysates by MIBs provides an integrated perspective, with both understudied and well-characterized kinases having similar functional versus expression dominant properties as well as demonstrating the potential for common mechanisms of regulation. Furthermore, the properties of a subset of characterized and understudied kinases suggest the dynamic regulation of a subset of the kinome not previously detected using other methods.

In contrast to the lack of correlation between expression and MIB-binding of specific kinases, there are a significant subset of kinases that show a high MIB-mRNA correlation across cell lines, including some having a correlation value greater than 0.8. These include ERBB2, expressed at high levels in HER2+ breast cancer, as well as PDPK1 and several other kinases such as PKN1 and others (Figure 2.3E), consistent with these kinases having transcriptional regulation as an important component of their overall functional output. Similar to the low correlation observed in Figure 2.3A-D, the distribution of the Pearson correlation coefficients of all kinases across all 15 cell lines shows a low correspondence between MIB-binding and RNA-seq (Figure 2.3F), with the mean correlation being 0.2. As also observed in Figure 2.3G there are a number of kinases that are only observed with one of the applied methods, MIB/MS or RNA-seq. This discrepancy is partly due to the 50+ RSEM read threshold used here as a positive identification in RNA-seq, potentially missing very lowly expressed kinases. Similarly, kinases not observed with MIB/MS but identified in RNA-seq may be missed due to being in an inactive, nonfunctional state and/or failure of chosen inhibitors to bind these kinases with adequate affinity. Pseudokinases that do not bind ATP will generally not be captured by MIBs.

### 2.2.4 Kinome profiles accurately define tumor biopsies

Cumulatively, our data show measurement of kinases by MIB capture allows integration of a significant fraction of the expressed kinome, including both understudied and well-characterized kinases. MIB-binding profiles define a taxonomy of breast cancer determined by the functional behavior of protein kinases. This functional taxonomy is used below to define subnetworks within the kinome that integrates understudied and well-characterized kinases that is not possible using transcriptome data.

Given both the variation in kinase MIB-binding profiles observed across subtypes, as well as their differing information content when compared to RNA expression measurements, we sought to better understand which kinases were key nodes in the subtype-selective baseline breast cancer kinome. To address this question we investigated the MIB-binding behavior of kinases across all four subtypes. We predicted three major classes of kinases: 1) those that show variation in MIB-binding across all subtypes, 2) those that exhibit more limited subtype-specific behaviors, and 3) kinases that have nominal distinguishing behavior. Standard application of PCA identifies those kinases displaying

the greatest variation across all samples ("pan-subtype kinases"; Figure 2.2E) and thus we used a feature selection approach based on the Bhattacharyya distance [73] to determine subtype-specific kinases that are highly distinguishing/informative for a single cancer subtype (see Methods). The combined set of the 50 most informative kinases is shown in Figure 2.4A, with column ordering based on similarity of the kinome profile and recapitulating similarity between claudin-low and basal subtypes as well as HER2-enriched and luminal. The HER2+ cell line that profiles as basal-like (HCC1954, in purple) is displayed separately. Recognized cancer-related kinases are again observed in this set, including ERBB2, FGFR2, PTK6, RAF1 and RON (MST1R) as well as 22 understudied kinases. Using the 50 most informative pan- and subtype-specific kinases in Figure 2.4A, we assessed their effectiveness in subtype identification using a support vector machine (SVM) classifier along with leave-one-out cross-validation. We found that the ability to classify a cell line's subtype from measurement of these kinases was highly accurate, with perfect precision and specificity for claudin-low and basal subtypes (Table 2.2).

The kinases shown in Figure 2.4A have the greatest variation within and across subtypes and are representative of each of the major subfamilies of kinases in addition to three metabolic kinases captured by MIBs (Figure 2.4B). Under the assumption that TNBC (represented by basal/claudin-low cell lines) and HER2/luminal breast cancer are separate diseases, we again used unsupervised feature selection of MIB/MS data to identify kinases that distinguish TNBC (basal/claudin-low) from HER2/luminal breast cancer [73]. As shown in Figure 2.4C, obvious differences in the kinome profiles of TNBC and HER2/luminal are observed, demonstrating the unique functional phenotypic features of the kinome in the two different breast cancers. Sixteen understudied kinases showed strong variance between TNBC and HER2/luminal breast cancer, with higher-ranked understudied kinases being DAPK3, ADCK1, MRCKA (CDC42BPA), STK17A, DMPK and VRK2.

Using the kinases chosen from feature selection in Figure 2.4C, we evaluated the ability to use MIB-binding profiles to define subtypes of human HER2+ needle biopsies and TNBC breast tumors (Figure 2.4D & E). Diagnostic needle biopsies of 2 patient tumors (2 HER2+) having ~1 mg of total protein were processed using MIB enrichment. With just 1 mg of tumor lysate protein, the total number of kinases purified from each biopsy ranged from ~200 to 275. Utilizing the kinases in Figure 2.4C in a SVM classifier, it was possible to clearly identify HER2+ and TNBC primary patient tumors. As with cell lines, application of PCA to MIB-binding profiles showed a clear separation

between TNBC tumors and HER2+ tumors (Figure 2.4F). Kinases driving the variation within the data included ERBB2 as expected and several understudied kinases including TESK1 and DMPK (Figure 2.4G). Thus, the MIB-binding activity of as few as 50 kinases is sufficient to discriminate the functional phenotypic nature of the kinome in breast cancer.

### 2.2.5 A functional interaction network of MIB-binding kinases

To determine the basic architecture of the kinome using MIB-bound kinases, we compiled protein interaction and phosphorylation data from multiple data sources (HIPPIE, HPRD, I2D, PhosphoSitePlus, Reactome; see Methods) and established a functional interaction network among 246 of the 254 kinases commonly identified in the panel of 15 breast cancer cell lines. Spectral clustering of this network further enabled the identification of 16 subnetworks (Figure 2.5A). Understudied kinases (green glyphs) are widely distributed across all the major subnetworks, demonstrating these poorly characterized kinases are integrated into subnetworks with well-characterized kinases. The 50 distinguishing kinases identified for cell lines in Figure 2.4A (triangle glyphs) were also distributed throughout the network and associated subnetworks. The scope of subnetwork coverage by these kinases suggests that their predictive value in our subtype classification comes from their distribution across many subnetworks, providing an overall estimate of the state of many functional processes simultaneously. Example subnetworks with enriched GO terms for the innate immune response (subnetwork 1) and cell division and mitosis (subnetwork 2) are shown in Figures 2.5B and 2.5C, respectively.

### 2.2.6 Kinome MIB-binding profile and response to drug perturbation

To assess how the baseline kinome and associated understudied kinases change in their functional MIB-binding profile in response to targeted drug perturbation, we exposed four cell lines to three subtype-relevant kinase inhibitors: SUM159 and HCC1806 with trametinib (a MEK1/2 inhibitor); SKBR3 with lapatinib (a HER2/EGFR inhibitor); and MCF7 with buparlisib (a PI3K inhibitor). The four inhibitors each strongly inhibited growth of the treated cell line (Figure 2.6A). We have previously shown that the kinome is dynamic and rapidly adapts to targeted perturbation by kinase inhibitors [17, 21]. This adaptive response is readily observed by changes in the MIB-binding profiles for each drug treatment (Figure 2.6B), with SUM159 cells showing the strongest dynamic response

to drug perturbation relative to the other cell lines, but each line clearly shows an adaptive response of the kinome measured by MIB-binding profiles. Figure 2.6C shows the kinases that are unique to each subtype defined in Figure 2.2C (Table 2.3).

Scatter plots of the SUM159 and HCC1806 dataset defines specific kinases and kinome subnetworks that drive the adaptive response to MEK1/2 inhibition that are represented by both understudied and well-characterized kinases (Figure 2.6D). Understudied (i.e. NEK2 and PASK) and well-characterized (i.e. DDR1 and EPHA4) kinases respond differently in the two subtypes (basal-like and claudin-low) when they are treated with the same kinase inhibitor (trametinib). The kinases in the subnetworks defined in Figure 2.5 also respond uniquely in the basal-like and claudin-low subtypes when treated with trametinib (Figure 2.6E). The adaptive kinome response measured by dynamic changes in MIB-binding profiles is more clearly seen when specific subnetworks are analyzed (Figure 2.6F). The global response of the seven largest subnetworks to these drug perturbations is shown with their functional annotation as estimated from Gene Ontology terms and KEGG pathway enrichment shown on the x-axis and MIB-binding response reported on the y-axis as a mean across all cell lines and drugs. Subnetworks have heterogeneous responses, with some subnetworks being fairly coordinated in response and others having kinase members acting in a more strongly divergent manner. For instance, subnetwork 3 (SN3) is enriched with many kinases relevant to cytoskeleton, adhesion and motility and has many of its members strongly up-regulated in response to drug perturbation. In comparison, SN2, involved in cell cycle and cell division, contains both strongly up-regulated and down-regulated kinases, with the largest responses being loss of MIB-binding, consistent with the inhibition of cell growth. Understudied kinases (green labels in Figure 2.6F) often display large responses to drug treatment within a given subnetwork, demonstrating these kinases actively contribute to adaptive kinome reprogramming in response to targeted kinase inhibition. Similarly, a more detailed look at targeted inhibition of specific subnetworks for each of the cell lines shows the dynamic response of the kinome to be highly dependent on the drug, subtype and subnetwork context (Figure 2.6G, Table 2.4). More broadly, the response of kinases in subnetworks is consistent with a unique functional regulation of the kinome in cancer subtypes and in response to different perturbations.

## 2.3 Discussion

While the creation of molecular taxonomies has established the existence of subtypes in many tissue-specific cancers, how these taxonomies can be leveraged to characterize phenotype or to guide the development of targeted therapeutics remains unclear. A complication for improving therapeutic intervention with targeted kinase inhibitors in cancer is the extensive number of understudied kinases, whose poor characterization presents significant challenges to understanding their role in emergent processes such as adaptive bypass reprogramming and resistance to kinase inhibitors. Despite such challenges, understudied kinases do have the potential as novel drug targets once their functional integration into signaling networks is more clearly determined. Methods generally have been lacking to capture kinases, both well-characterized and understudied, to define the functional kinome en masse. Characterization of kinase MIB-binding in tumor cell lysates has proven to be a powerful technique for characterizing functional architectures of the kinome that provides the capability to identify prognostic signatures and differential response to perturbations such as targeted kinase inhibition, as well as better establishing the integration and function of understudied kinases. This is clearly seen in the 50 kinase profile distinguishing TNBC from HER2+/luminal breast cancer with many of the 50 kinases representing understudied kinases.

The highest weighted understudied kinases distinguishing TNBC from HER2+/luminal breast cancer include ADCK1 (AarF Domain Containing Kinase whose function is unclear), DAPK3 (Death-associated protein kinase thought to be involved in apoptosis), DMPK (Dystrophia myotonica protein kinase whose function is not well-defined), MRCKA (Myotonic dystrophy kinase-related CDC42 binding protein kinase alpha that may signal CDC42 control of the actin cytoskeleton and is related to DMPK), STK17A (Serine/threonine kinase 17A has apoptosis-inducing activity and is a member of the DAP kinase-related family), TLK2 (Tousled-like kinase 2 is involved in chromatin assembly and possibly DNA repair) and VRK2 (Vaccinia-related kinase 2 that is believed to regulate apoptosis and cell growth). Screening of the cBioPortal for Cancer Genomics (http://www.cbioportal.org/public-portal/) indicates amplification of MRCKA (CDC42BPA) in 13-25% of invasive breast cancer while TLK2 is amplified in 10% of invasive breast cancer and 25% of adenoid cystic breast cancer. ADCK1, DAPK3, DMPK, STK17A and VRK2 were found to be similarly amplified in other cancers including prostate adenocarcinoma, uterine carcinosarcoma and

pancreatic adenocarcinoma. Prominent MIB-binding signatures combined with potential increased expression in tumors suggests these understudied kinases have important functions for the tumor cell phenotype that have not been characterized to date.

The dynamic nature of the kinome is clearly captured in the kinase MIB-binding profiles characterizing baseline versus post-drug treated cells. This adaptive reprogramming of the kinome is involved in the epigenetic development of resistance to kinase inhibitors [74]. We have proposed that blocking this adaptive reprogramming is important clinically for making single kinase inhibitors more durable [21]. Pre- and post-drug treatment MIB/MS analysis allows for the quantitative measure of kinome adaptive responses and the rapid screening of combinations of kinase or epigenetic inhibitors that would block the adaptive behavior of the kinome [17, 21, 74]. This analysis can be done in preclinical models as well as patient trials where biopsy accessible tumor specimens are available. We have been able to capture more than 200 kinases with as little as 300 $\mu$g of tumor biopsy protein. Thus, MIBs provide a proteomic approach to characterize the functional state and dynamics of the kinome and thus define therapeutic response and targetable adaptive resistance networks. Importantly, MIBs capture both well-characterized and understudied kinases for a comprehensive measure of the functional kinome.

## 2.4 Materials and Methods

### 2.4.1 MIB affinity chromatography

Broad spectrum Type I kinase inhibitors (CTx-0294885, VI-16832, PP58, Purvalanol B, UNC-2147A, and UNC-8088A) were custom-synthesized with hydrocarbon linkers and terminal amine groups and covalently attached to ECH-activated Sepharose beads as previously described [21]. Cells were rinsed in PBS and processed in lysis buffer (50 mM HEPES, 150 mM NaCl, 0.5% Triton X-100, 1 mM EDTA, 1 mM EGTA, at pH 7.5 containing 10 mM NaF, 2.5 mM NaVO4, cOmplete protease Inhibitor Cocktail (Roche), and 1% Phosphatase Inhibitor Cocktails 2 and 3 (Sigma)). Tumor biopsies obtained from UNC Tissue Procurement were manually homogenized with a chilled mortar and pestle in lysis buffer. For individual bead profiling (Figure 2.1), 2mg of total protein was gravity-flowed over 100uL of each bead. For Figure 2.2 (cell lines), 5mg of total protein lysate and for Figure 2.4 (human tumor biopsies), 1mg of total protein was gravity-flowed over a mixture

of the six kinase inhibitor-linked beads (175$\mu$L total beads). Beads were washed with at least 30 volumes of high salt (1M NaCl) and low salt (150mM NaCl) lysis buffer, then 500$\mu$L of low salt lysis buffer containing 0.1% SDS. Bound proteins were eluted by boiling with 0.5% SDS and 1% $\beta$-mercaptoethanol in 100mM Tris-HCl, pH 6.8, 2X 15 minutes, treated with DTT (5mM, 25min at 60°C) and Iodoacetamide (20mM, 30min in the dark at RT), and spin-concentrated to 100$\mu$L (Millipore Amicon Ultra-4, 10K cutoff) before Methanol/Chloroform precipitation. Proteins were trypsinized overnight at 37°C and then dried down in a speed-vac. Peptides were cleaned with C-18 spin columns (Pierce).

### 2.4.2 Mass Spectrometry and Analysis

Peptides were resuspended in 2% ACN and 0.1% Formic Acid. For Figure 2.1 (bead profiling) 20% of each sample was injected onto a Thermo Easy-Spray 75$\mu$m x 15cm C-18 column using an Easy nLC-1000 in technical triplicate and separated on a 150 min gradient (5-40% ACN). For Figures 2.2 and 2.4 (cell lines and tumor biopsies), 40% of the final peptide suspension was injected onto an Easy-Spray 75$\mu$m x 25cm C-18 column and separated on a 300min gradient (cell lines) or a 180min gradient (tumor biopsies). For all runs, ESI parameters: 3e6 AGC MS1, 80ms MS1 max inject time, 1e5 AGC MS2, 100ms MS2 max inject time, 20 loop count, 1.8 m/z isolation window, 45s dynamic exclusion. Spectra were searched against the Uniprot/Swiss-Prot database with Sequest HT on Proteome Discoverer software (Figures 2.1 and 2.2) or MaxQuant (Figure 2.4). Only peptides with medium or greater confidence (5% FDR) were considered for quantitation, and only kinases having 3 or more unique peptides were considered for further analysis. Heat maps were generated using GENE-E software (BROAD institute). Kinome trees were generated using Kinome Render (http://bcb.med.usherbrooke.ca/kinomerender.php) [75].

### 2.4.3 RNA-seq

Total RNA was spin column purified using RNeasy Plus Mini kit (Qiagen). Library construction was performed at the UNC Lineberger Comprehensive Cancer Center Genomics Core and the sequencing at the UNC High-Throughput Sequencing Facility. mRNA-Seq libraries were constructed with 1 $\mu$g total RNA using the Illumina TruSeq$^{TM}$RNA Sample Prep Kit according to the manufacturer's protocol. 50-cycled single-end sequencing runs with multiplexing were produced using an

Illumina HiSeq2000. CASAVA 1.8.2 generated bases and assessed sequence quality. The QC-passed reads were aligned to the hg19 human reference genome using MapSplice and the alignment profile was determined by Picard Tools v1.64 [76]. Aligned reads were sorted and indexed using SAMtools, and then translated to transcriptome coordinates and filtered for indels, large inserts, and zero mapping quality using UBU v1.0. Transcript abundance estimates for each sample were determined using an Expectation-Maximization algorithm [77]. Publicly available data from [78, 79] were also processed using this computational method.

### 2.4.4 Data analysis

Hierarchical clustering, Principal Components Analysis (PCA), and feature selection were performed in MATLAB. PCA is a commonly-used data analysis and dimension-reduction technique that transforms variables into a set of linearly uncorrelated principal components [80]. Application of PCA also provides the ability to assign a weight to each feature (kinase) in the data set that can be used as a relative measure of its ability to distinguish subtypes. To identify kinases that dominate individual PCs, kinases having weights in the 90th-percentile (i.e. those weighted in the top 10% of weights) per PC were selected from the first three PCs and used in downstream classification tasks. Feature selection using the Bhattacharyya distance was also used as a secondary mechanism for ranking kinases in terms of their ability to distinguish subtypes [73]. Pairwise classification between subtypes (e.g. basal-like subtype from all others, claudin-low from all others, etc) was iteratively performed to identify the most informative features.

Kinases identified through feature ranking and PCA are combined to create a list of the most distinguishing kinases in MIB-binding across the breast cancer subtypes. Subtype-specific signature kinases are defined as the top 5% of the highest ranking kinases found using the Bhattacharyya feature ranking coefficient for each subtype are compiled for the overall list. Pan-subtype signature kinases are defined as the most heavily weighted kinases (top 10%) from the first three PCs are used. Subtype-specific signature kinases are compiled from each of the breast cancer subtypes then the global signature kinases are added (in order from most heavily weighted to less heavily weighted) starting with PC1 kinases then moving to PC2 then to PC3 until a maximum of 50 kinases is reached to make up the list of distinguishing kinases. A total of 50 kinases was chosen as classification accuracy of the breast cancer subtypes plateaued at this level (Figure 2.7).

### 2.4.5 Comparison of MIB-binding to transcript abundance

The Z-score is calculated by sample based on the average log2 value per kinase and using the standard deviation of all kinases for a given sample; the samples within a particular subtype are then averaged (Figure 2.3 A–D). The Pearson correlation calculated by subtype (Figure 2.3 A–D) is the correlation between the MIB-binding profile to the RNA transcript levels within a single subtype. The Pearson correlation of individual kinases (Figure 2.3F) is calculated for each kinase across the 15 cell lines (not distinguished by subtype) between MIB-binding and RNA transcript levels. The correlation calculated by subtype (Figure 2.3 A–D) is the correlation between MIB-binding and RNA transcript levels to see how the profiles of each data type correlate between MIB-binding and RNA. The correlation calculated for each kinase is a correlation across cell lines to determine how a kinase correlates across all cell lines analyzed here.

### 2.4.6 Prediction of subtypes

Classification of subtype based on a previously unobserved kinome signature was performed using a Support Vector Machine (SVM). The SVM is a commonly used machine learning technique used in supervised classification, and thus requires a training set on which to learn parameters that can then be applied towards prediction of previously unobserved data [34]. The SVM used here was trained on the 50 distinguishing kinases previously identified. Performance of the SVM was analyzed using leave-one-out cross validation, where training is performed on all samples except for one and a classification prediction being carried out for the left-out sample. Predictions are made in this way for every sample and final sensitivity, specificity, and precision are calculated on classification performance across all samples. Human tumors are classified into one of the major groups (TNBC or HER2+/Luminal) or as “other” using the SVM trained on the 50 distinguishing kinases identified from breast cancer cell line samples.

### 2.4.7 Network Analysis

Protein-protein interaction information was compiled from multiple public data sources for the 254 kinases analyzed in this data set and included, the Human Integrated Protein-Protein Interaction rEference (HIPPIE) (updated 9/1/2015; [81]), Human Protein Reference Database (HPRD Release 9;

[82]), Interlogous Interaction Database (I2D version 2.9; [83, 84]), PhosphoSitePlus (phosphosite.org - downloaded 10/15/2015; [85]) and Reactome protein-protein interactions (downloaded 12/15/2015; [86]). The union of all interactions between the 254 kinases was used to form a single network that was then clustered into communities/subnetworks with the spectral method in Mathematica (ver 10.3).

Table 2.1: Understudied kinases.

| | | | | | |
|---|---|---|---|---|---|
| AAK1 | CDKL4 | MAP3K15 | NEK8 | PRPF4B | STK36 |
| AATK | CIT | MAP3K2 | NEK9 | PXK | STK38 |
| ADCK1 | CLK1 | MAP3K3 | NIM1 | RIOK1 | STK38L |
| ADCK2 | CLK2 | MAP3K4 | NLK | RIOK2 | STK40 |
| ADCK3 | CLK3 | MAP3K5 | NPR1 | RIPK4 | STRADA |
| ADCK4 | CLK4 | MAP3K8 | NPR2 | RNASEL | STRADB |
| ADCK5 | DAPK1 | MAP4K1 | NRBP1 | SBK1 | TAOK1 |
| ALPK1 | DAPK2 | MAP4K2 | NRBP2 | SBK2 | TAOK2 |
| ALPK2 | DAPK3 | MAP4K3 | NRK | SCYL1 | TAOK3 |
| ALPK3 | DMPK | MAP4K5 | NUAK1 | SCYL2 | TBCK |
| AMHR2 | DSTYK | MAST1 | NUAK2 | SCYL3 | TESK1 |
| ANKK1 | DYRK1A | MAST2 | OBSCN | SGK071 | TESK2 |
| BCKDK | DYRK1B | MAST3 | OXSR1 | SGK110 | TEX14 |
| BMP2K | DYRK2 | MAST4 | PASK | SGK196 | TLK1 |
| BRD2 | DYRK3 | MASTL | PBK | SGK223 | TLK2 |
| BRD3 | DYRK4 | MATK | PDIK1L | SGK494 | TNIK |
| BRD4 | EEF2K | MELK | PDK1 | SIK1 | TNNI3K |
| BRSK1 | ERN1 | MINK1 | PDK2 | SIK2 | TP53RK |
| BRSK2 | ERN2 | MKNK1 | PDK3 | SIK3 | TRIB1 |
| CASK | GAK | MKNK2 | PDK4 | SLK | TRIB2 |
| CDC42BPA | GSG2 | MLKL | PEAK1 | SNRK | TRIB3 |
| CDC42BPB | GUCY2D | MOK | PHKG1 | SPEG | TRIM24 |
| CDC7 | GUCY2F | MST4 | PHKG2 | SRMS | TRIM28 |
| CDK10 | HIPK1 | MYLK | PIK3R4 | SRPK1 | TRIO |
| CDK11A | HIPK2 | MYLK2 | PIM1 | SRPK2 | TRPM6 |
| CDK12 | HIPK3 | MYLK3 | PIM2 | SRPK3 | TRPM7 |
| CDK13 | HIPK4 | MYLK4 | PIM3 | STK10 | TSSK1B |
| CDK14 | HUNK | MYO3A | PINK1 | STK16 | TSSK2 |
| CDK15 | ICK | MYO3B | PKDCC | STK17A | TSSK3 |
| CDK16 | INSRR | NEK1 | PKMYT1 | STK17B | TSSK4 |
| CDK17 | KALRN | NEK10 | PKN1 | STK24 | TSSK6 |
| CDK18 | LMTK2 | NEK11 | PKN2 | STK25 | TTK |
| CDK19 | LMTK3 | NEK2 | PKN3 | STK31 | TTN |
| CDK20 | LRRK1 | NEK3 | PRKCI | STK32A | TXK |
| CDK7 | LRRK2 | NEK4 | PRKG1 | STK32B | UHMK1 |
| CDKL1 | MAK | NEK5 | PRKG2 | STK32C | VRK1 |
| CDKL2 | MAP3K1 | NEK6 | PRKX | STK33 | VRK2 |
| CDKL3 | MAP3K14 | NEK7 | PRKY | STK35 | WNK4 |

Table 2.2: Specificity, sensitivity, and precision across subtypes in Leave-one-out Cross Validation.

| | **Specificity** | **Sensitivity** | **Precision** |
|---|---|---|---|
| **Basal-like** | 0.9615 | 1 | 0.8571 |
| **Claudin-low** | 1 | 0.8333 | 1 |
| **HER2-enriched** | 0.9545 | 0.9 | 0.9 |
| **Luminal** | 1 | 1 | 1 |

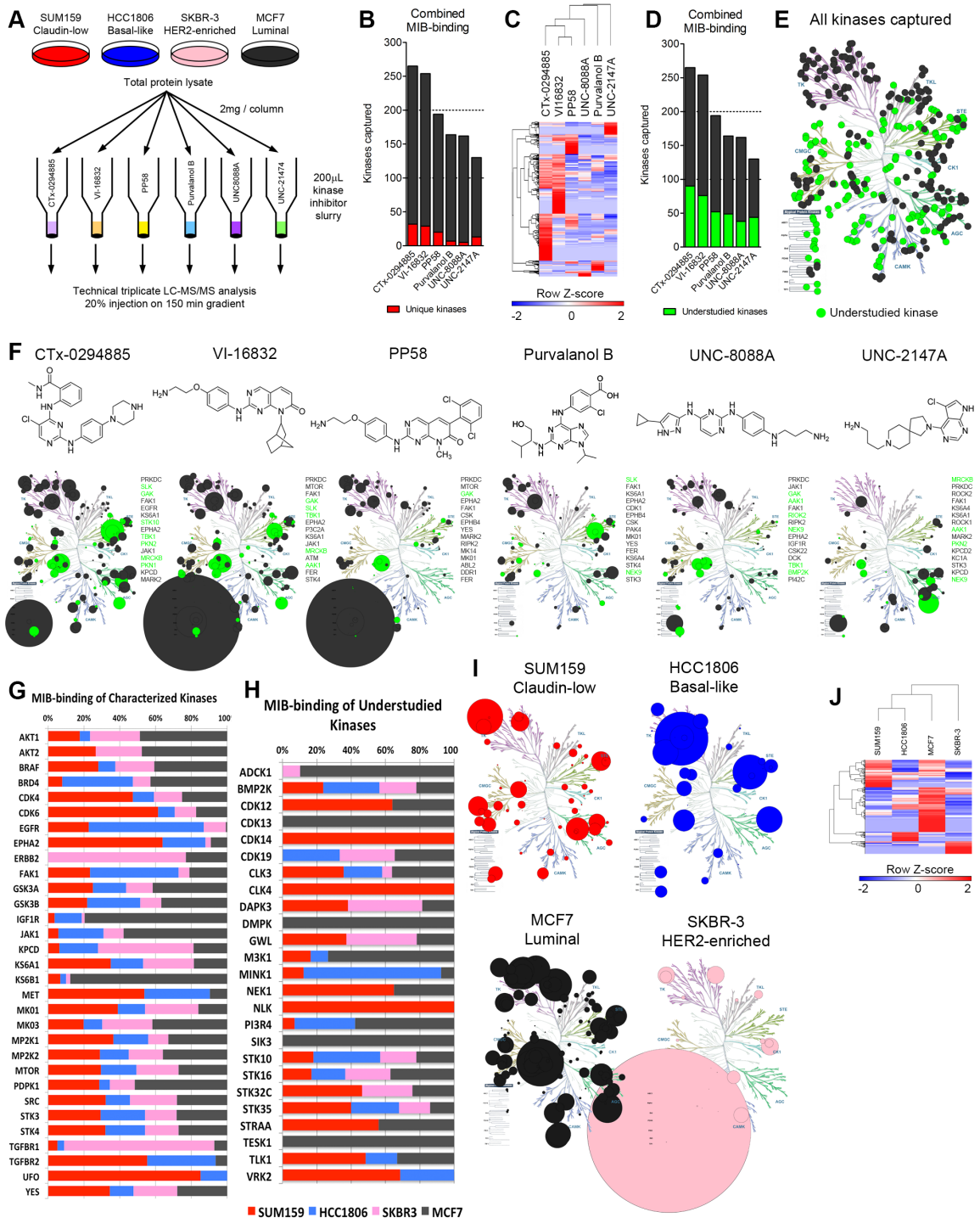

A
SUM159 Claudin-low
HCC1806 Basal-like
SKBR-3 HER2-enriched
MCF7 Luminal
Total protein lysate
2mg / column
CTx-0294885
VI-16832
PP58
Purvalanol B
UNC8088A
UNC-21474
200µL kinase inhibitor slurry
Technical triplicate LC-MS/MS analysis
20% injection on 150 min gradient
B
Combined MIB-binding
Kinases captured
Unique kinases
C
Row Z-score
D
Combined MIB-binding
Understudied kinases
E
All kinases captured
Understudied kinase
F
CTx-0294885
VI-16832
PP58
Purvalanol B
UNC-8088A
UNC-2147A
G
MIB-binding of Characterized Kinases
H
MIB-binding of Understudied Kinases
SUM159
HCC1806
SKBR3
MCF7
I
SUM159 Claudin-low
HCC1806 Basal-like
MCF7 Luminal
SKBR-3 HER2-enriched
J
Row Z-score

Table 2.3: Uniquely bound kinases to MIBs by breast cancer subtype.

| Subtype | Uniquely Bound Kinases |
|---|---|
| **Basal-like** | DAPK2, DUSTY, MAST1, MAST2, NEK11, NRK |
| **Claudin-low** | ADCK4, BRD2, CD11A, CTRO, HIPK3, KPCG, M3K7, MK07, MYLK, NEK4, NEK6, NEK7, PAK7, PASK, PGFRA, SGK1, STK33, STK39, UHMK1 |
| **HER2-enriched** | CDKL3, CSKP, DDR2, ERBB2, HXK1, JAK3, K6PF, KPCB M3K13, MAK, PCKGM, RIPK3, TESK2, TRIO |
| **Luminal** | BLK, ERBB4, INSRR, KAPCG, LRRK2, M3K10, MK15, VGFR3 |

Figure 2.1: Assessment of multiplexed kinase inhibitor beads (MIBs) for kinase capture across breast cancer subtypes. (A) Experimental design to assess performance of six kinase inhibitor beads. (B) Combined data from all four cell lines assayed shows CTx-0294885 binds the most number of kinases. Number of kinases captured uniquely by each bead is shown in red. (C) Euclidean hierarchical clustering kinase peptides bound by the six beads shows each bead enriches for a distinct set of kinases. UNC-2147A displays the most unique binding profile. (D) A large proportion of kinases captured by MIBs (23-24%) are understudied or poorly characterized (green). (E) 381 kinases were identified across all four cell lines, including 346 protein kinases and 35 metabolic kinases. Of these protein kinases, 142 are understudied (green). (F) Chemical structures and kinase-binding of each inhibitor bead across the kinome. Circle size is proportional to the number of unique peptides identified per kinase. PRKDC (DNA-PK) is over-represented in VI-16832 and PP58 (large circle under Atypical protein kinases). Most beads capture kinases across families but UNC-2147 preferentially enriches for AGC family kinases. Shown to the right of each kinome tree are the 15 most-highly captured kinases for each bead. Green circles and text signify understudied kinases. (G) Comparison of relative binding of characterized kinases across breast cancer cell lines/subtypes. (H) Comparison of relative binding of understudied kinases across breast cancer cell lines/subtypes. (I) Each cell line representing the different breast cancer subtypes displays a unique kinome profile. Only kinases with the greatest number of peptides identified in particular cell line are shown. Circle size is proportional to the number of peptides identified. (J) Hierarchical clustering of peptides identified for each kinase (rows) across the cell lines (columns) cluster triple-negative cell lines (SUM159, HCC1806) together and indicates HER2-positive SKBR3 cells have the most distinct kinome profile.

Table 2.4: Subnetworks in kinome.

| Subnetwork | Kinases |
|---|---|
| **SN1** | M3K5, M3K6, MK14, VRK2, IRAK1, M3K11, MP2K4, MP2K1, M3K1, MK10, KS6A5, MK09, DYR1B, M3K3, MK08, MP2K3, M4K4, M3K4, MP2K5, RIPK2, IRAK4, KC1G1, KC1G2, MP2K6, TAOK2, MK11, MLTK, M4K3, M3K2, PKN1, MK13 |
| **SN2** | CDK1, STK3, CDK2, WEE1, AURKB, PLK1, CDK8, AAK1, CDK3, CDK6, CDK7, GWL, LATS1, MK06, NEK2, NEK9, PLK4, CDK13, CDK12, NEK3, CDK16, CDK18, CDK19, E2AK1, STK38, STK10 |
| **SN3** | EPHA1, EPHA3, EPHA4, EPHB2, EPHB3, EPHB6, MET, YES, ACK1, LYN, RON, ROCK1, EPHB4, CDK17, EPHA2, CDK4, DDR1, PI42C, DAPK3, ROCK2, STK16, LIMK2 |
| **SN4** | ABL2, EGFR, FYN, JAK1, BRAF, PTK6, TBK1, FGRF1, FGFR2, FGFR3, KSYK, FRK, TGFR2, M3KL4, IGF1R, TYK2, FER, KT3K, TEC, FAK2, PKN3 |
| **SN5** | AAPK1, AAPK2, MARK4, PK3C3, SIK1, SIK2, SIK3, STRAA, TIF1B, KAPCB, CHK2, KPCD2, DYR1A, KAPCA, K6PP, KC1A, KC1D, KC1E, KKCC2 |
| **SN6** | BUB1, GSK3B, PDPK1, TTK, KS6A2, MK01, KS6A1, KS6A3, KS6A6, KS6B1, PMYT1, MK03, TLK2, INSR, KS6Z4, E2AK2, KCC2B, TNIK |
| **SN7** | PI3R4, SRPK2, AKT1, AKT2, CHK1, CLK2, CSK21, FYV1, IKKA, P4K2B, PKN2, PRPK, CLK3, MINK1, TLK1, TNK1 |
| **SN8** | MARK3, RAF1, KPYM, KPCD, GSK3A, AKT3, ARAF, KPCA, MP2K2, DMPK, KPCD3, NEK1, NLK |
| **SN9** | MTOR, PRKDC, ACVR1, CDK14, IKKB, OXSR1, TFGR1, CSK22, CDK9, RIOK2, E2AK4, SG196, STK35 |
| **SN10** | CSK, CDK5, EPHA5, EPHA7, HCK, SRC, AGK, GAK, CHKA, EPHB1, M4K5 |
| **SN11** | MARK1, MARK2, STK11, TAOK1, STK24, MST4, ERN1, TAOK3, MELK, STK25, ST38L |
| **SN12** | ABL1, BMP2K, CDKL5, KPCD1, CLK1, CLK4, PI42A, PAK4 |
| **SN13** | ERBB2, FAK1, KCC2A, FGFR4, SRPK1, KCC2D, KCC2G, PHKG2 |
| **SN14** | BMR1A, ACV1B, AVR2A, BMPR2, LIMK1, MRCKA |
| **SN15** | ILK, K6PL, IKKE, PDXK, SLK, PI4KA |
| **SN16** | PI51C, PI42B, PI51A |
| **SN17** | TESK1 |
| **SN18** | CHKB |
| **SN19** | MRCKB |
| **SN20** | MRCKG |

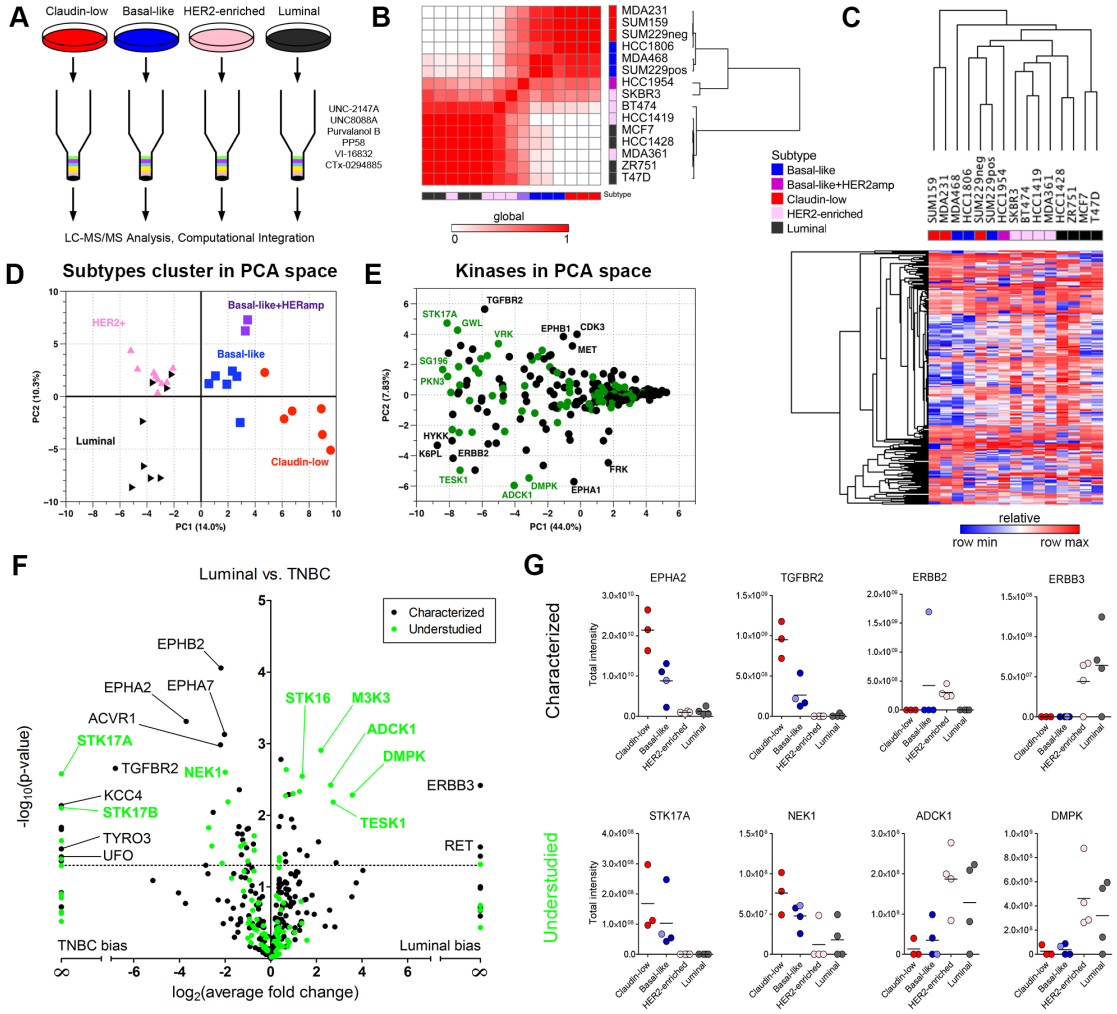
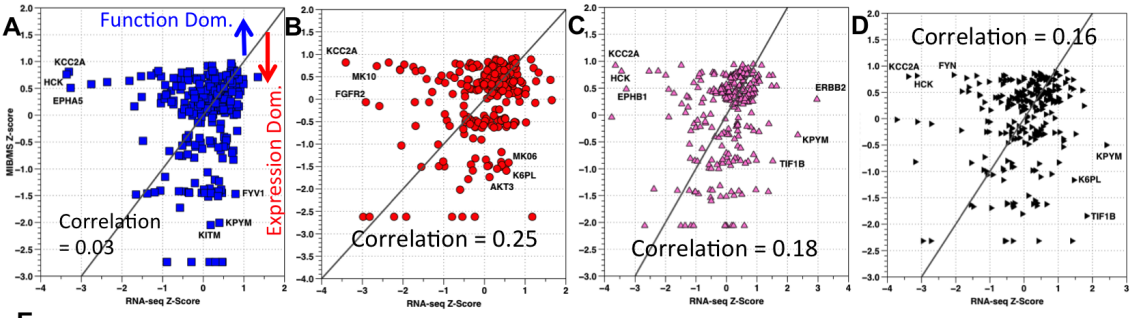
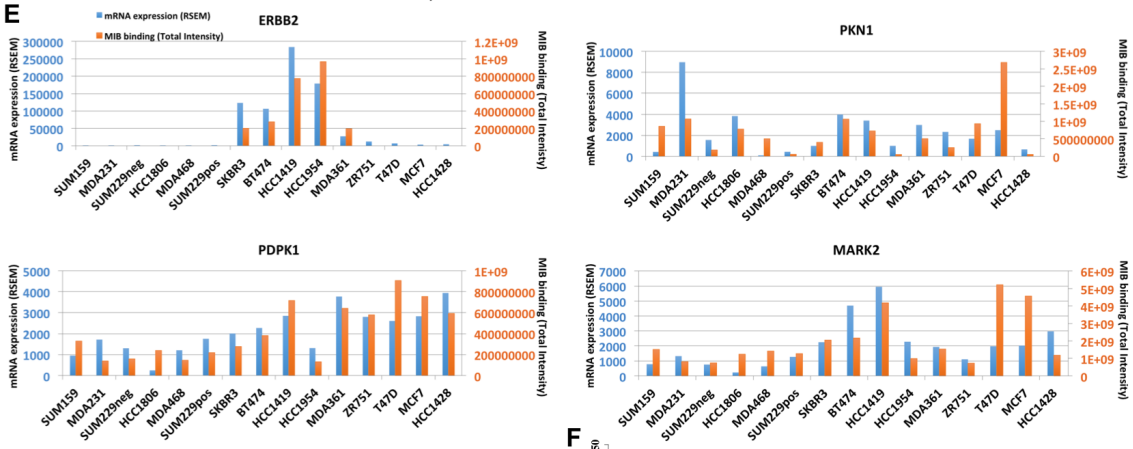
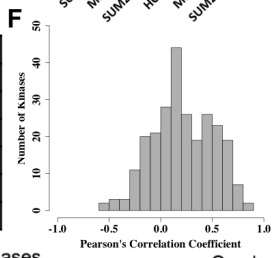

A
Claudin-low
Basal-like
HER2-enriched
Luminal
UNC-2147A
UNC8088A
Purvalanol B
PP58
VI-16832
CTx-0294885
LC-MS/MS Analysis, Computational Integration
B
MDA231
SUM159
SUM229neg
HCC1806
MDA468
SUM229pos
HCC1954
SKBR3
BT474
HCC1419
MCF7
HCC1428
MDA361
ZR751
T47D
Subtype
global
0
1
C
Subtype
Basal-like
Basal-like+HER2amp
Claudin-low
HER2-enriched
Luminal
SUM159
SUM229neg
MDA468
HCC1806
SUM229pos
HCC1954
SKBR3
BT474
HCC1419
MDA361
HCC1428
ZR751
MCF7
T47D
relative
row min
row max
D
Subtypes cluster in PCA space
HER2+
Basal-like+HERamp
Basal-like
Luminal
Claudin-low
PC1 (14.0%)
PC2 (10.3%)
E
Kinases in PCA space
TGFBR2
STK17A
GWL
VRK
EPHB1
CDK3
MET
SG196
PKN3
HYKK
K6PL
ERBB2
TESK1
DMPK
ADCK1
EPHA1
FRK
PC1 (44.0%)
PC2 (7.83%)
F
Luminal vs. TNBC
Characterized
Understudied
EPHB2
EPHA2
EPHA7
ACVR1
STK17A
TGFBR2
NEK1
KCC4
STK17B
TYRO3
UFO
STK16
M3K3
ADCK1
DMPK
ERBB3
TESK1
RET
TNBC bias
Luminal bias
-log10(p-value)
log2(average fold change)
G
Characterized
Understudied
EPHA2
TGFBR2
ERBB2
ERBB3
STK17A
NEK1
ADCK1
DMPK
Total intensity
Claudin-low
Basal-like
HER2-enriched
Luminal

Figure 2.2: MIB/MS kinome profiling assigns breast cancer cell lines to functional subtypes. (A) Individual samples run across 6-bead composition with LC-MS/MS analysis. (B) Heat map of correlation between MIB/MS samples for cell lines analyzed. Color bars indicated the subtype of each cell line (blue: basal-like, red: claudin-low, pink: HER2-enriched, black: luminal, purple: basal-like/HER2amp). White in the heat map indicates a low correlation between samples, while red shows higher correlation. Rows and columns are hierarchically clustered. (C) Heat map of MIB/MS average for each of the 15 cell lines analyzed. Rows are kinases; columns are MIB/MS cell line averages. Color bar for columns indicates the subtype associated to each cell line. Each column is an average of 2 or 3 MIB/MS samples, depending on the cell line. Colors in the heat map are relative by row minimum (blue) and maximum (red). A total of 254 kinases passed filtering (see Methods). Rows and columns are hierarchically clustered using Euclidean distance and average linkage. (D) Principal Component Analysis (PCA) on the entire MIB/MS data set. PC1 and PC2 account for 14.0% and 10.3% of the variance in the data set, respectively. A total of 32 samples across the fours subtypes are represented by their subtype (red circle: claudin-low, blue square: basal-like, pink upward triangle: HER2-enriched, black right-pointing triangle: luminal, purple square: basal-like/HER2amp). (E) PCA on the MIB/MS data set to show highly variable kinases across the samples. Characterized and understudied kinases are shown in black and green, respectively. PC1 and PC2 account for 44.0% and 7.83% of the variance in the data set, respectively. (F) Volcano plot showing characterized (black) and understudied (green) kinases that are significantly ($p<0.05$) different between the Luminal/HER2-enriched and TNBC (basal-like/claudin-low) cell line samples in the MIB/MS data set. (G) Profiles of selected characterized (top row) and understudied (bottom row) kinases across breast cancer subtypes.




| | | | | |
|---|---|---|---|---|
| ABL2 | MARK2 | ACK1 | MARK4 | CDK16 |
| MK09 | CHK1 | MP2K4 | DDR1 | MRCKB |
| EPHA1 | MRCKG | ERBB2 | PDPK1 | FGFR2 |
| PI42C | JAK1 | PI4KA | K6PP | PI51A |
| KPCD | PKN1 | KPCD1 | TAOK1 | KPCD2 |
| TLK2 | KSYK | TNIK | M3K11 | TNK1 |
| M3KL4 | VRK2 | | | |



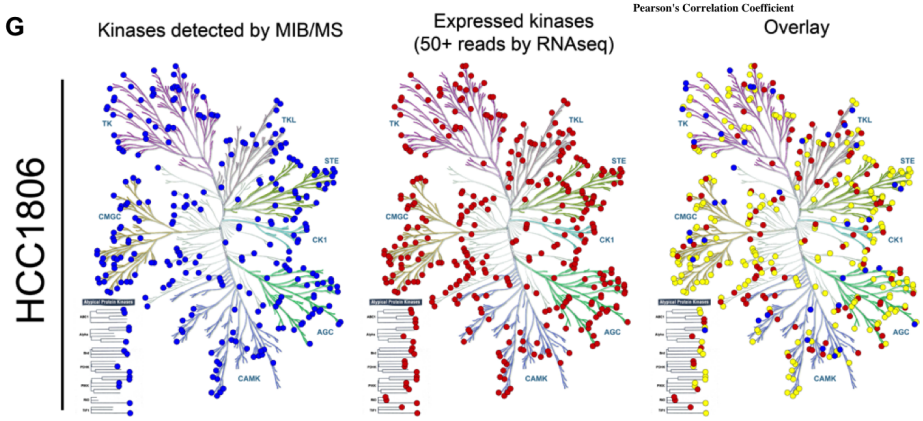



Figure 2.3: Overall MIB-binding and mRNA expression levels are not correlated. (A–D) Comparison of baseline normalized MIB/MS (y-axis) intensities and baseline normalized mRNA (x-axis) counts in cell lines for representing each of the four subtypes (A: basal-like, B: claudin-low, C: HER2-enriched, D: luminal). The grey line indicates a 1-to-1 relationship between normalized MIB/MS and RNA levels. Above the line indicates Function Dominance, while below the line indicates Expression Dominance. Correlation between the MIB/MS and RNA levels were basal-like: 0.03, claudin-low: 0.25, HER2-enriched: 0.18, and luminal: 0.16. (E) Representative raw profiles of ERBB2, PKN1, PDPK1, and MARK2 in MIB/MS intensity (orange bars) and mRNA RSEM counts (bue bars), which are both highly correlated between the two data sets (correlation >0.8). A table of 32 kinases that have a correlation greater than 0.7 across the four subtypes is also shown (bottom). (F) Frequency distribution of Pearson's correlation coefficient across all cell lines in MIB/MS and RNA-seq for each of the 254 kinases. (G) KinomeTrees for MIB-binding (left), RNA expression (middle), and the overlap between the two data sets (right) for the HCC1806 (basal-like) cell line. In the overlap, yellow indicates where the two data sets overlap.

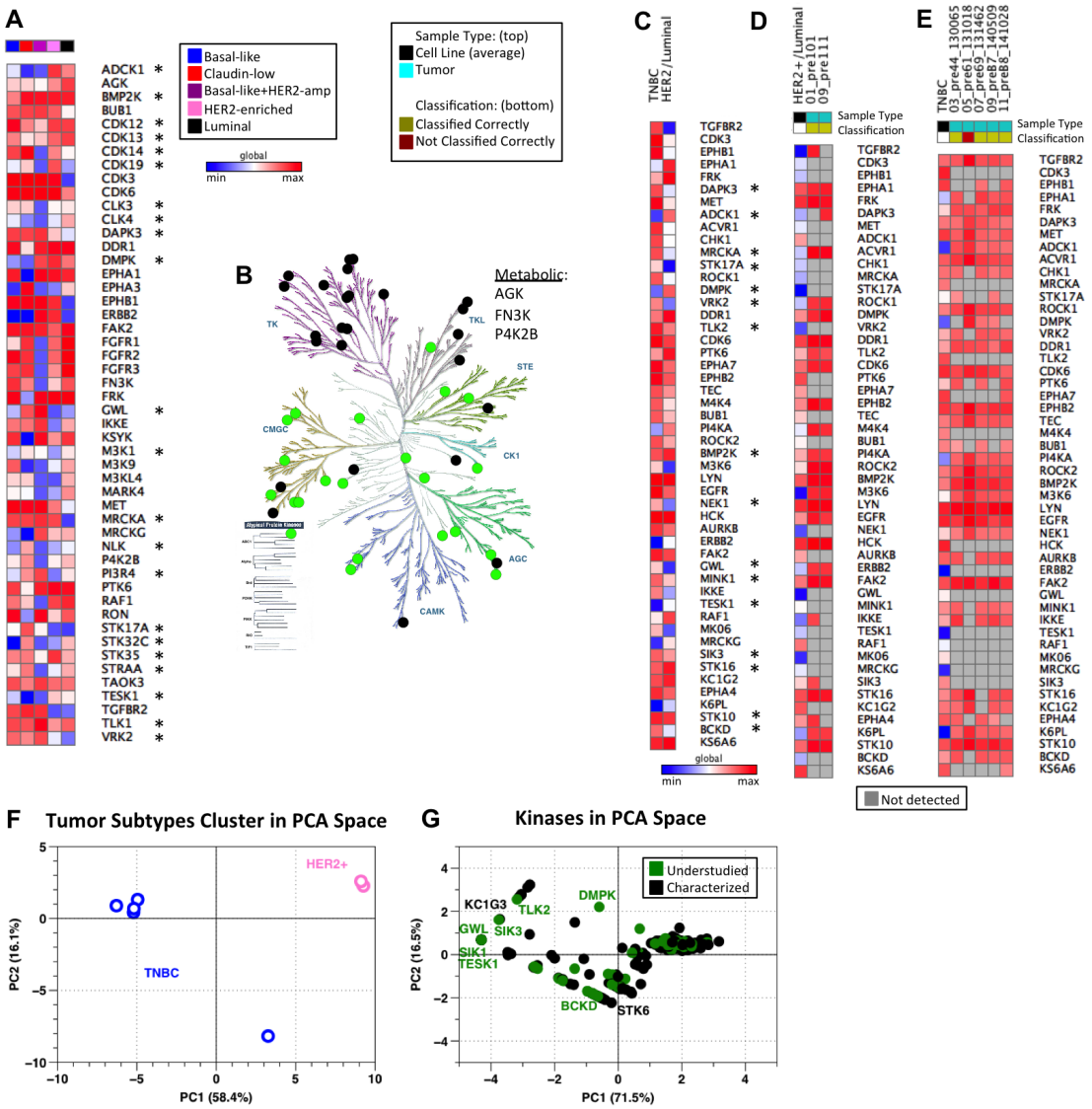


Figure 2.4: Baseline kinome of cell lines and tumors across breast cancer subtypes. (A) Compilation of subtype specific and pan-subtype kinases chosen from feature selection and PCA, respectively. All data is log2 normalized and autoscaled by sample, with heat map colors indicating low (blue) to high (red) MIB-binding. Column color bar indicates subtype (red: claudin-low, blue: basal-like, pink: HER2-enriched, black: luminal, purple: basal-like/HER2amp; Understudied kinases are denoted by *). Global maximum and minimum color assignment. (B) KinomeTree with the 50 distinguishing features from (A) are denoted. Black circles denote characterized kinases, while green circles represent understudied kinases. (C) Kinases chosen from feature selection when comparing Luminal/HER2-enriched cell line samples against basal-like/claudin-low (TNBC) cell line samples. Kinases are ordered from top to bottom in the same ordering as from the feature selection (most heavily weighted kinases are at the top of the heat map). All data is log2 normalized and autoscaled by sample, with heat map colors consistent with those in (A) (Understudied kinases are denoted by *). Global maximum and minimum color assignment. (D) Heat map of Luminal/HER2-enriched cell line average (HER2+/Luminal column; black in "Sample Type" column color bar) across the kinases shown in (C) with two tumor samples (teal in "Sample Type" column color bar). Data is log2 normalized and autoscaled by samples, as previously noted. Yellow in the "Classification" column bar shows which samples are classified correctly as Luminal/HER2-enriched by the SVM using the kinases from (C). Blue in the heat map indicates a low MIB-binding, red indicates high MIB-binding, and grey (in the tumor samples only) indicates that a kinase was not detected by MIBs in the tumor sample. Global maximum and minimum color assignment. (E) Heat map of TNBC cell line average (TNBC column; black in "Sample Type" column color bar) across the kinases shown in (C) with five tumor samples (teal in "Sample Type" column color bar). Data is log2 normalized and autoscaled by samples, as previously noted. Yellow in the "Classification" column color bar shows which samples are classified correctly as TNBC by the SVM using the kinases from (C). Dark red in "Classification" indicates that the tumor sample was incorrectly classified (not classified as TNBC) by the SVM using the kinases identified in (C). Color scheme in the heat map is consistent with that described in (D). Global maximum and minimum color assignment. (F) PCA scores plot of tumor samples with PC1 and PC2 accounting for 58.4% and 16.1% of variance, respectively. TNBC tumors are blue and HER2-enriched tumors are pink. (G) PCA loadings plot of tumor samples with PC1 and PC2 accounting for 71.5% and 16.5% of variance, respectively. Black points are characterized kinases and green points denote understudied kinases.

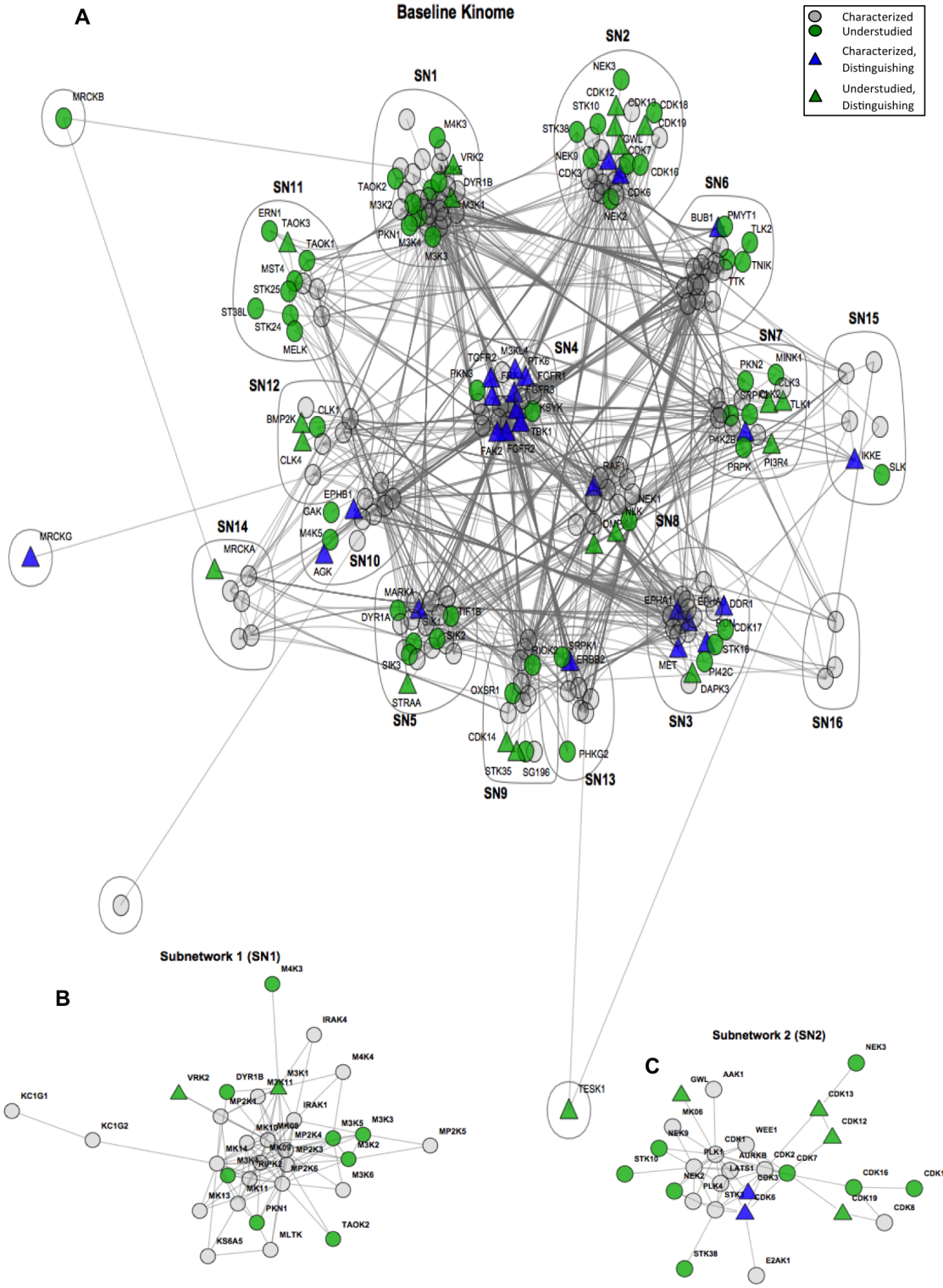

A
Baseline Kinome
Characterized
Understudied
Characterized, Distinguishing
Understudied, Distinguishing
SN1
SN2
SN3
SN4
SN5
SN6
SN7
SN8
SN9
SN10
SN11
SN12
SN13
SN14
SN15
SN16
Subnetwork 1 (SN1)
B
Subnetwork 2 (SN2)
C

Figure 2.5: Subnetworks in the functional kinome. (A) Compiled and clustered protein-protein interaction network from public data sources of the 254 kinases analyzed in the MIB/MS cell line data set. Green nodes represent understudied kinases, while grey and blue nodes represent characterized kinases. Triangles are kinases that are also in the distinguishing features found in Figure 2.4A. (B) The largest subnetwork in (A) that contains understudied (10 kinases; 34.5%) and distinguishing (2 kinases; 6.9%) kinases. This subnetwork is enriched for the GO term "Innate Immune Response". (C) The second largest subnetwork within (A) that contains 12 (41.4%) understudied and 6 (20.7%) distinguishing kinases and is enriched for "Cell Cycle", "Cell Division", and "Mitosis" GO terms.

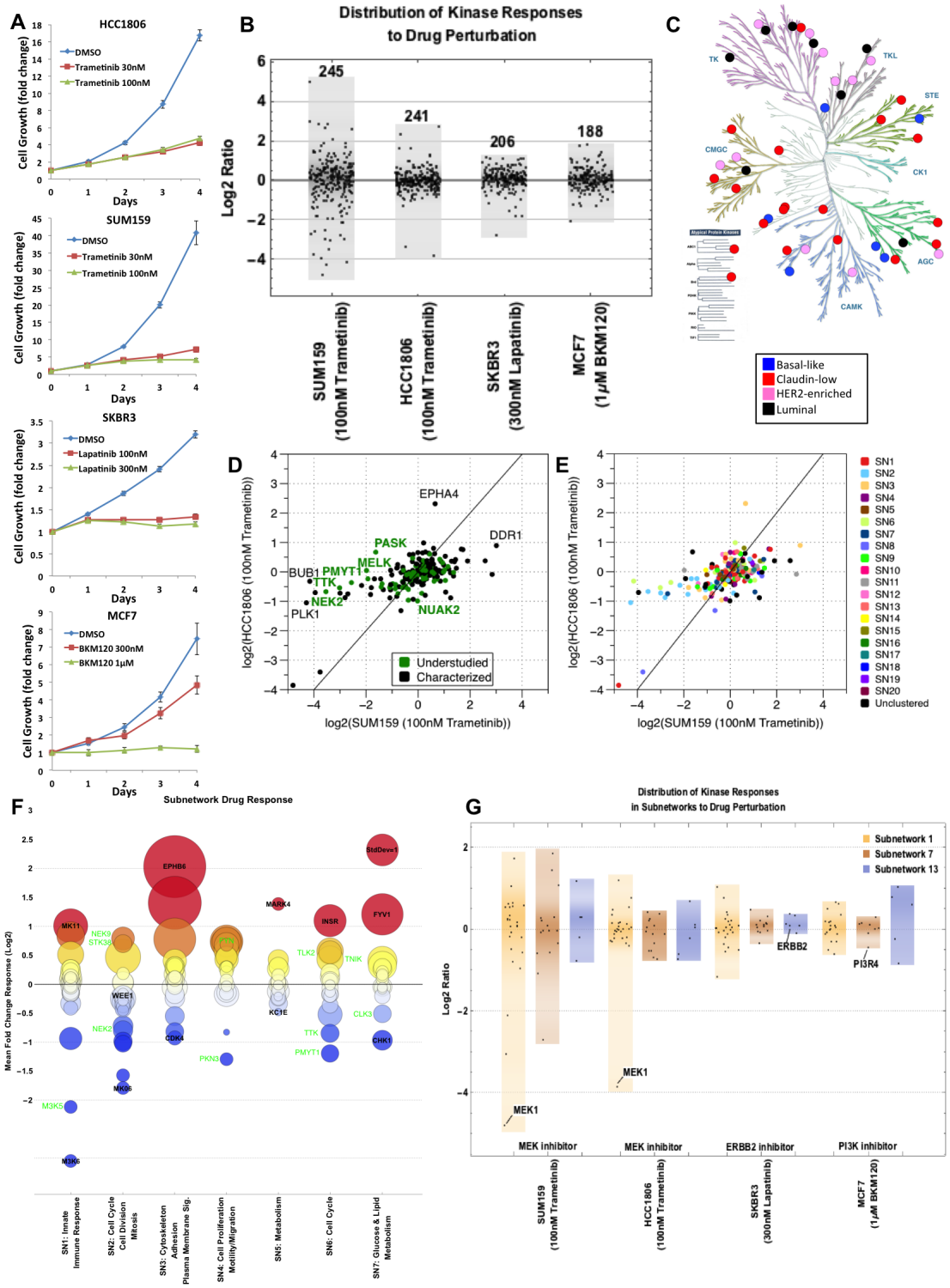

A
HCC1806
DMSO
Trametinib 30nM
Trametinib 100nM
Cell Growth (fold change)
Days
SUM159
DMSO
Trametinib 30nM
Trametinib 100nM
Cell Growth (fold change)
Days
SKBR3
DMSO
Lapatinib 100nM
Lapatinib 300nM
Cell Growth (fold change)
Days
MCF7
DMSO
BKM120 300nM
BKM120 1μM
Cell Growth (fold change)
Days
B
Distribution of Kinase Responses
to Drug Perturbation
Log2 Ratio
245
241
206
188
SUM159 (100nM Trametinib)
HCC1806 (100nM Trametinib)
SKBR3 (300nM Lapatinib)
MCF7 (1μM BKM120)
C
TK
TKL
STE
CMGC
CK1
AGC
CAMK
Basal-like
Claudin-low
HER2-enriched
Luminal
D
EPHA4
DDR1
PASK
MELK
BUB1
PMYT1
TTK
NEK2
PLK1
NUAK2
Understudied
Characterized
log2(HCC1806 (100nM Trametinib))
log2(SUM159 (100nM Trametinib))
E
SN1
SN2
SN3
SN4
SN5
SN6
SN7
SN8
SN9
SN10
SN11
SN12
SN13
SN14
SN15
SN16
SN17
SN18
SN19
SN20
Unclustered
log2(HCC1806 (100nM Trametinib))
log2(SUM159 (100nM Trametinib))
F
Subnetwork Drug Response
Mean Fold Change Response (Log2)
EPHB6
StdDev=1
MK11
NEK9
STK38
MARK4
INSR
FYV1
TLK2
TNIK
WEE1
KC1E
CLK3
NEK2
CDK4
TTK
CHK1
PMYT1
PKN3
MK06
M3K5
M3K6
SN1: Innate Immune Response
SN2: Cell Cycle Cell Division Mitosis
SN3: Cytoskeleton Adhesion Plasma Membrane Sig.
SN4: Cell Proliferation Motility/Migration
SN5: Metabolism
SN6: Cell Cycle
SN7: Glucose & Lipid Metabolism
G
Distribution of Kinase Responses
in Subnetworks to Drug Perturbation
Subnetwork 1
Subnetwork 7
Subnetwork 13
Log2 Ratio
MEK1
MEK1
ERBB2
PI3R4
MEK inhibitor
MEK inhibitor
ERBB2 inhibitor
PI3K inhibitor
SUM159 (100nM Trametinib)
HCC1806 (100nM Trametinib)
SKBR3 (300nM Lapatinib)
MCF7 (1μM BKM120)

Figure 2.6: Kinome drug response overall and by subnetwork. (A) Growth curves for HCC1806 + Trametinib, SUM159 + Trametinib, SKBR3 + Lapatinib, and MCF7 + BKM120. All curves were done at two different doses. (B) Repsonse of kinome in representative cell lines across four subtypes of breast cancer (claudin-low: SUM159, basal-like: HCC1806, HER2-enriched: SKBR3, luminal: MCF7) when treated with the indicated kinase inhibitor. Distribution of the kinome response on the log2-scale is shown for each cell line/subtype; each point represents a kinase. (C) KinomeTree showing the kinases that are uniquely captured in each of the subtypes in the baseline data set. Blue circles denote kinases bound to the MIBs only in basal-like samples. Similarly, red circles represent claudin-low, pink circles represent HER2-enriched, and black circles represent luminal uniquely bound kinases. (D) Scatter plot of the response of the basal-like vs. claudin-low cell lines to treatment with 100nM Trametinib. All values are fold change to untreated cells and log2-transformed. Kinase points are colored black for characterized and green for understudied. (E) Scatter plot of the response of the basal-like vs. claudin-low cell lines to treatment with 100nM Trametinib (same as in (D)). Kinases are colored by subnetwork assigned to each kinase from Figure 2.5. (F) Subnetwork response to drug perturbation showing mean fold change across the four representative cell lines (SUM159, HCC1806, SKBR3, and MCF7) for the top 7 subnetworks identified from Figure 2.5. Characterized and understudied kinases in each subnetwork are labeled in black and green, respectively. The color of each circle indicates the mean fold change (red=high/above 0, blue=low/below 0), while the area of the circle denotes the standard deviation of the fold changes across the representative cell lines. (G) Distribution of the kinome response in the three subnetworks (SN1, SN7, and SN13) on the log2-scale is show for each cell line/subtype.

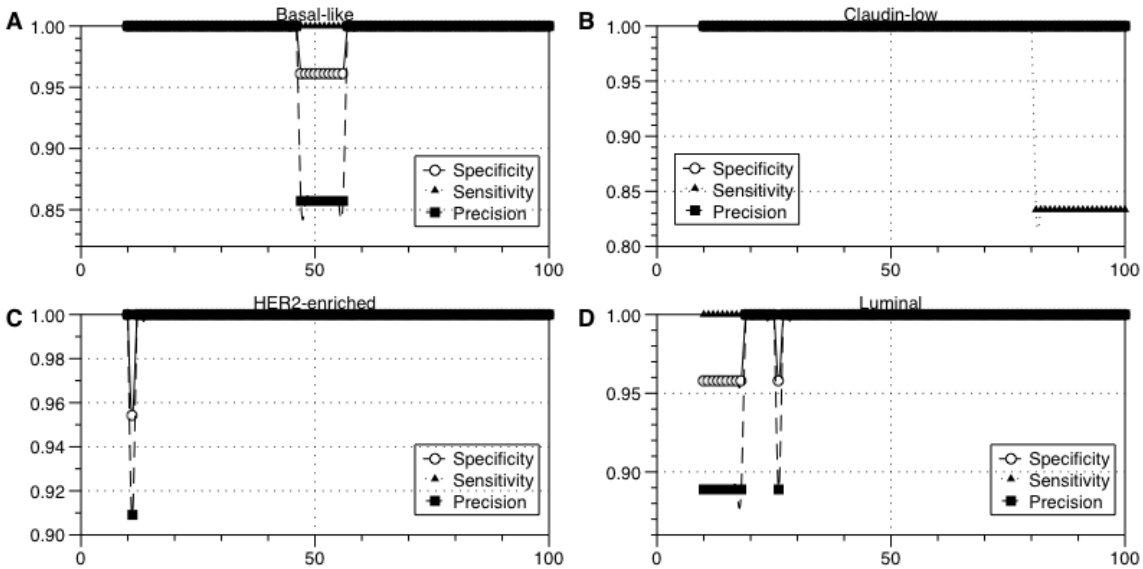



Figure 2.7: Sensitivity, specificity, and precision across subtypes to determine number of features needed for classification. (A) Basal-like, (B) Claudin-low, (C) HER2-enriched, and (D) Luminal. Open circles represent specificity, filled triangles sensitivity, and filled squares precision for each panel.

# CHAPTER 3

# CHARACTERIZATION OF THE ARCHITECTURE OF UNDERSTUDIED AND TARGETED SUBNETWORKS WITHIN THE KINOME

## 3.1 Introduction

The network of kinases and their interactions with each other, or the kinome, has importance in cancer diagnosis, treatment, and resistance due to its role in mediating cellular signals and cell-fate decisions [87]. While their importance is recognized, functions of many kinases are not known (i.e. "understudied") [18]. Aberrant regulation of kinases is a known factor in the progression of cells toward a cancer phenotype [15]. Being able to target these proteins, and their interactions with each other, is important in the study of cancer and its treatment.

The ability to target the kinome with small molecules (or kinase inhibitors) has been studied but not all kinases have small molecules that can bind to them (i.e. "undertargeted") [18]. While it would be ideal to be able to target each kinase directly, this is not always possible due to physical properties of the proteins [88]. Indirectly targeting kinases in the kinome requires knowledge of the interactions a kinase has to perform its functions. Due to the size of the kinome, computational approaches or methodologies need to be used to identify functional modules within the network to determine which groups of kinases closely interact with each other. Previous computational methods have clustered networks in general but the focus is generally on the entire proteome, not the kinome interaction network, when applied to a biological context [41, 89, 90].

Our focus on the kinome provides a unique perspective of cancer biology. The goal of this work is to define the human kinome interaction network from public databases, to characterize the architecture of this network, and to determine enriched biological functions. Here, we present a method for clustering the kinome interaction network into subnetworks and analyzing enrichment of function, understudied kinases, and targeted kinases within these subnetworks.

## 3.2 Biological Data

### 3.2.1 Kinome Network Databases

Human protein interactions from Molecular INTeraction Database (MINT) [91], BioGRID [92], Database of Interacting Proteins (DIP) [93], Human Protein Reference Database (HPRD) [82], Interologous Interaction Database (I2D) [94], Reactome [86, 95], Human Integrated Protein-Protein Interaction rEference (HIPPIE) [81], and PhosphoSite [85] were compiled to create the Kinome Interaction Network (KIN). Only kinase-kinase interactions were considered for the KIN, duplicate edges were only counted one time, and self-loops/self-self interactions are allowed. For example, if the connection between ERBB2 and EGFR appeared in HPRD and PhosphoSite, the connection between these two kinases would only appear once in the final network. All compiled interactions are undirected since the direction of interaction may not always be specified in the databases. The compiled KIN contains 453 kinases and 3690 interactions (including the 355 self-self interactions).

### 3.2.2 Inhibitor Databases

Kinase inhibitor target information is used from the KinomeSCAN assay in the LINCS database. This database tested 144 kinase inhibitors and over 380 kinases [96]. The KinomeSCAN assay measures the amount of binding between purified kinases and inhibitor compounds relative to a control sample with no inhibitor compound added. The data provided by this assay comes in two forms: percent inhibition and $K_d$. Percent inhibition shows, for each kinase tested, how much of the kinase is inhibited relative to control. The $K_d$ values in this database indicate the binding dissociation constant between the inhibitor and each kinase tested. For all of the inhibitors tested in this database, the concentrations were high so a very strict threshold for indicating on- or off-targets of the drugs is used ($>$90% inhibition or $K_d < 1\mu$M ). For this characterization, inhibitor targets (on- and off-targets) are all treated equally. This means that if a specific inhibitor is designed to inhibit/bind to ERBB2 but it also inhibits/binds to EGFR, then ERBB2 and EGFR are both considered targets of that inhibitor.

### 3.2.3 Defining Understudied and Targeted Kinases and Subnetworks in the Kinome

Understudied kinases are defined using the criteria discussed in Chapter 2. Targeted kinases are specifically defined using the LINCS data in section 3.2.2. A kinase is considered "targeted" if there is at least one drug that significantly binds to it and/or inhibits its function. Conversely, a kinase is labeled "untargeted" if there are no inhibitors that bind to it or inhibit its function, as defined in section 3.2.2.

## 3.3 Methods

### 3.3.1 Subnetwork Identification

A network constructed using publicly available data is clustered using the IGraph package in R to find subnetworks in the overall network [97]. Many methods exist for clustering networks into modules, or subnetworks. Specifically, the FastGreedy and SpinGlass algorithms from the IGraph R package [97] and the MCODE clustering algorithm (also an R package, "ProNet") [90, 98] are discussed here. Many network clustering algorithms exist and have been explored previously but these three are chosen for this specific context to handle an undirected, unweighted network [99].

The MCODE and FastGreedy algorithms are deterministic, meaning the clustering of a network is the same every time. The MCODE algorithm is one standard way to cluster networks and has been implemented in many software packages (R and Cytoscape, for example) [90]. This algorithm is deterministic but does have many parameters that can be adjusted to fit the particular network being clustered [90].

FastGreedy is a bottom-up algorithm, meaning all nodes start as individual clusters, that uses a modularity score to determine whether or not clusters should be merged [38]. When the modularity score is higher for the network when two clusters are merged than when they are separate, the two clusters are merged. This process is repeated until the modularity score cannot be improved by merging clusters. This algorithm has detection limitations that do not allow it to detect smaller subnetworks [38]. The outline of this method is shown in Figure 3.1(A).

The SpinGlass algorithm clusters the curated network using a model based in statistical physics [39, 40, 41]. Each node starts in its own subnetwork (or spin state) and subnetworks are collapsed

together based on the connections (i.e. edges) between the two subnetworks being analyzed. This clustering method is nondeterministic, which means that each run of the SpinGlass algorithm will provide a slightly different clustering of the same network. The reason for varying results after each run is due to the fact that the SpinGlass algorithm randomly chooses seed nodes to begin collapsing subnetworks, as seen in Figure 3.1(B) [40]. As shown in Figure 3.1(B), to provide a more accurate clustering of the network, a consensus network is compiled by running the SpinGlass algorithm 1000 times and determining which kinases cluster together over 90% of the time. If two kinases cluster together over 90% of the time, the kinases are assigned to the same cluster.

### 3.3.2 Comparison of Subnetwork Identification Across Algorithms

To assess the similarity of various network clustering algorithms, multiple metrics are calculated between all pairs of algorithms (FastGreedy to consensus SpinGlass, FastGreedy to MCODE, and consensus SpinGlass to MCODE). The metrics used here are Variation of Information (VI) [100], Normalized Mutual Information (NMI) [101], Split-Join Distance (SJD) [102], and Unadjusted/Adjusted Rand Index (URI/ARI) [103].

VI is the amount of information lost and gained when going from one clustering to another and is always positive [100]. The VI metric describes what is different between two clusterings, so a smaller value of VI indicates a similar clustering and vice versa. NMI is a similarity measure based on a mutual information score, normalized based on the sizes of the clusters, and is bounded between 0 and 1. A value of 0 indicates no mutual information in the clustering while a value closer to 1 indicates that the information in the two clusterings is very similar [101]. The SJD is a measure that shows how many "moves" are needed to go from one clustering to another, where a single "move" consists of splitting off a single element off of one cluster and then either attaching it to another cluster or starting a new cluster. A large value of SJD indicates that there are many "moves" required to achieve the same clustering, which means the clusterings are not very similar [102]. The URI and ARI are similarity measures based on how much the clusterings agree and disagree. The difference between URI and ARI is that ARI adjusts for the expected overlap based on the sizes of the clusters. The URI is bounded between 0 and 1 but the ARI can be negative if the similarity is less than expected [103].

### 3.3.3 Statistical Assessment of Understudied and Targeted Subnetworks in the KIN

Functional support for the KIN is defined by statistically significant ($p < 0.05$) GO terms ("GO biological process complete") found using the online PANTHER tool [104]. The statistical significance is determined with a modified background of only the 453 kinases in the compiled KIN to determine enrichment of GO terms for each subnetwork. The p-values are also corrected in PANTHER for multiple testing with the Bonferroni correction [104].

A hypergeometric test is used to assess (over- or under-) enrichment of understudied (or targeted) kinases in each subnetwork identified using methods discussed previously ($p < 0.05$). An assumption for this statistical test is that the labeling of a kinase as understudied or targeted is independent from the network architecture compiled from the public databases.

## 3.4 Biological Assumptions

This method takes advantage of a biologic assumption. The assumption is that the KIN functions with modules (or subnetworks) that have closely related nodes. The modules are found using the consensus clusters from running a clustering algorithm (SpinGlass) many times. The identified subnetworks rely on accuracy (and completeness) of the mined data that informs the underlying network connections.

## 3.5 Results & Discussion

### 3.5.1 Clustering the KIN

Figure 3.2 shows the clustering of the same network ($n = 453$, $m = 3960$) with three different algorithms (Figure 3.2(A) MCODE, 3.2(B) FastGreedy, and 3.2(C) consensus SpinGlass) discussed in the Methods section. Visually these clustering results look different and Table 3.1 provides a table of metrics to quantify the similarities and differences among the three algorithms. MCODE does not perform well on this network and the various metrics in Table 3.1 quantitatively support the dissimilarity between the MCODE clusters and those found by the FastGreedy or consensus SpinGlass algorithms. MCODE compared to FastGreedy and SpinGlass have an ARI of 0.03 and 0.02, respectively. This means that the clustering found by MCODE and the clustering found by

FastGreedy (or SpinGlass) are no more similar than expected by chance. Similarly, the SJD distances for these two comparisons are 351 (MCODE to FastGreedy) and 358 (MCODE to SpinGlass), meaning that a node would have to be removed from a cluster and joined to another cluster around 350 times to make the clusterings the same. One reason for this poor performance may be due to the highly interconnected clusters within this network. MCODE may not be able to detect smaller clusters of highly connected nodes in a very dense network.

With a VIM of 1.50, ARI of 0.55, and NMI of 0.64, the FastGreedy and consensus SpinGlass algorithms are similar in their final clustering results. The consensus SpinGlass algorithm does detect more clusters (or subnetworks) that contain fewer kinases than the FastGreedy algorithm. As discussed in the Methods section, the FastGreedy algorithm is known to have a detection limit with smaller cluster sizes in networks.

While the SpinGlass algorithm is not deterministic and can, therefore, have varied clustering results, the consensus SpinGlass algorithm combines the results of many runs of the SpinGlass algorithm to provide a clustering that is more comprehensive. The SpinGlass algorithm does not have the detection limitation that the FastGreedy algorithm does but the consensus algorithm provides a similar clustering while still detecting smaller subnetworks in the overall network.

### 3.5.2 Functional Annotation of Subnetworks in the KIN

Tables 3.2 and 3.3 list the statistically significant GO biological processes terms as determined by PANTHER, as described in Methods. Some of the clusters between the two algorithm results have overlapping significant GO terms (for example, FastGreedy cluster 1 and consensus SpinGlass cluster 1 both have “Positive regulation of NF-kappaB transcription factor activity”, among others). In Table 3.2, it is clear that some of the clusters are large and, therefore, have fairly diverse significant terms (clusters 2 and 3, for example). The statistically significant GO terms for the consensus SpinGlass algorithm results (Table 3.3) are more specific than those for the FastGreedy algorithm results (Table 3.2).

The significant terms for the consensus SpinGlass algorithm (Table 3.3) are more focused due to the smaller size of most of the clusters. In particular, clusters 1, 2, 3, 9, and 10 have very focused functions. Cluster 9 has significant functions related to bone formation, osteoblast differentiation, and signaling pathways that are directly related to bone formation. Similarly, cluster 10 is focused on ion

transport and pancreatic cell function. These examples show that the consensus SpinGlass algorithm identify meaningful groups of kinases in the KIN and is corroborated with GO term enrichment.

### 3.5.3 Understudied Kinase Enrichment in Subnetworks in the KIN

Figure 3.3 shows the clustering of (A) FastGreedy and (B) consensus SpinGlass with understudied labels as node shape (understudied: triangles, characterized: circles). Given the results from the previous sections, the statistical results for the consensus SpinGlass algorithm are the only ones presented here. Table 3.4 shows the hypergeometric p-values for understudied subnetworks from consensus SpinGlass clustering. The subnetworks can be over- or under-enriched for understudied kinases and these are noted in Table 3.4 next to the significant p-values ($<0.05$). Consensus SpinGlass subnetworks 2, 3, and 4 are under-enriched for understudied kinases, meaning they are enriched for characterized kinases. These three clusters (2, 3, and 4) are larger, containing 51, 45, and 64 kinases, respectively, and are all enriched in GO functional annotations (Table 3.3).

Subnetworks 6 and 13 are over-enriched for understudied kinases so these subnetworks are understudied. Subnetwork 6 is the largest subnetwork in the consensus SpinGlass results with 144 kinases and 68 of these being understudied. This subnetwork is enriched for GO functional annotations that mostly revolve around cell cycle processes and cell growth (Table 3.3). Subnetwork 13 contains four understudied kinases: TTN, NEK9, NEK6, and NEK7. Not much is known about these kinases and, unfortunately but not unexpectedly, this subnetwork is not enriched for GO functional annotations.

### 3.5.4 Targeted Kinase Enrichment in Subnetworks in the KIN

The number of inhibitors that target each node in Figure 3.3 is indicated by node color (blue = zero inhibitors, purple = moderate number of inhibitors, red = high number of inhibitors). Table 3.4 also show the p-values for the targeted subnetworks from the consensus SpinGlass clustering. Again, these subnetworks can be over- or under-enriched for targeted kinases and this is noted in Table 3.4 next to the significant p-values ($<0.05$).

Under-enriched subnetworks for targeted kinases (i.e. undertargeted subnetworks) for the consensus SpinGlass clustering are subnetworks 6, 21, and 23. Subnetworks 6 and 23 are undertargeted but are enriched for GO functional annotations (Table 3.3), while subnetwork 21 is not enriched for

functional annotations. The subnetworks that are over-enriched for targeted kinases are subnetworks 1 and 4 so they are overtargeted subnetworks. Both of these subnetworks are enriched for GO functional annotations (Table 3.3), which is expected.

### 3.5.5 Integration of Functional Annotation with Understudied and Targeted Enrichment

The consensus SpinGlass clusters 6 and 13 are over-enriched for understudied kinases (i.e. they are "understudied subnetworks"), while clusters 2, 3, and 4 are under-enriched for understudied kinases (i.e. these clusters are "characterized subnetworks"). Clusters 1 and 4 from the consensus SpinGlass algorithm results are over-enriched for targeted kinases (i.e. they are "targeted subnetworks") and clusters 6, 21, and 23 are under-enriched for targeted kinases (i.e. they are "undertargeted subnetworks"). An important point to note is that the majority of the KIN (besides clusters 2, 3, and 4) has understudied kinases distributed throughout the network.

In particular, clusters 4 and 6 have interesting enrichment results. Subnetwork 4 is significantly under-enriched for understudied kinases AND significantly over-enriched for targeted kinases, which means this subnetwork is significantly characterized and targeted (Table 3.4). Conversely, the kinases in subnetwork 6 are significantly over-enriched for understudied kinases AND significantly under-enriched for targeted kinases, meaning this subnetwork is significantly understudied and undertargeted (Table 3.4). The results for these two subnetworks suggest an inverse correlation between a kinase being understudied and targeted. This inverse relationship is expected because inhibitors are typically designed to targeted specific proteins and we tend to know a lot about intended targets. For example, inhibitors are designed to target MEK1/2 because of its role in regulating cell division.

## 3.6 Conclusions

This work provides a comprehensive KIN that has been compiled from 8 different protein interaction databases. We cluster this comprehensive network into functional modules/clusters, or subnetworks, using a consensus SpinGlass algorithm. The functional subnetworks identified in the KIN have a statistically significant enrichment for GO biological processes that are functionally

Table 3.1: Comparison of clustering methods (FastGreedy, SpinGlass, and MCODE) using variation of information (VI), split-join distance (SJD), unadjusted rand index (URI), adjusted rand index (ARI), and normalized mutual information (NMI).

| | | **Variation Of Infor. (VIM)** | **Split-Join Distance (SJD)** | **Unadj. Rand Index (URI)** | **Adj. Rand Index (ARI)** | **Norm. Mutual Infor. (NMI)** |
|---|---|---|---|---|---|---|
| **FastGreedy** | **SpinGlass** | 1.50 | 228 | 0.85 | 0.55 | 0.64 |
| **FastGreedy** | **MCODE** | 2.07 | 351 | 0.39 | 0.03 | 0.06 |
| **SpinGlass** | **MCODE** | 2.58 | 358 | 0.34 | 0.02 | 0.07 |

specific. We also presented a characterization of the understudied and targeted kinases in relation to the architecture of the KIN. We found there are specific subnetworks that are significantly understudied/characterized and those that are significantly (under)targeted.

The value this work provides to the biological field is that it can help elucidate functions of understudied kinases in the KIN where there are statistically significant GO biological process terms. Additionally, this work can direct kinase inhibitor development to provide potential subnetworks that need more study.

Table 3.2: Significantly Enriched GO Biological Processes in FastGreedy Clusters.

| FastGreedy Cluster | GO biological process complete |
|---|---|
| **1** | Activation of MAP kinase activity<br>JNK cascade<br>Regulation of NF-kappaB transcription factor activity<br>Posttranscriptional regulation of gene expression |
| **2** | Cell division<br>Regulation of cell cycle<br>Immune system process<br>Positive regulation of multicellular organismal process<br>Regulation of immune response<br>Cell migration<br>Positive regulation of hydrolase activity<br>Transmembrane receptor protein kinase signaling pathway<br>Animal organ morphogenesis<br>MAPK cascade<br>Regulation of GTPase activity |
| **3** | Positive regulation of PI3K signaling<br>Ephrin receptor signaling pathway<br>Positive regulation of GTPase activity<br>Leukocyte migration<br>Single organism cell adhesion<br>Angiogenesis<br>Regulation of actin filament-based process<br>Regulation of leukocyte activation<br>Regulation of ERK1/2 cascade<br>Innate immune response<br>Positive regulation of cell migration<br>Tube development<br>Chemotaxis<br>Cell morphogenesis involved in differentiation<br>Regulation of cell adhesion<br>Cellular response to endogenous stimulus<br>Positive regulation of cell proliferation<br>Positive regulation of developmental process<br>Regulation of anatomical structure morphogenesis<br>Tissue development<br>Regulation of multicellular organismal development<br>Animal organ development |
| **7** | Wnt signaling pathway, planar cell polarity pathway |
| **8** | Regulation of PI3K signaling<br>Cellular macromolecule metabolic process |
| **11** | Inositol phosphate biosynthetic process<br>Biological regulation<br>Cellular macromolecule metabolic process |
| **15** | CDP-choline pathway |

Table 3.3: Significantly Enriched GO Biological Processes in SpinGlass Clusters.

| SpinGlass Cluster | GO biological process complete |
|---|---|
| **1** | Nucleotide-binding oligomerization domain containing signaling pathway<br>Activation of JUN kinase activity<br>Activation of MAPKK activity<br>Toll-like receptor signaling pathway<br>Positive regulation of NF-kappaB transcription factor activity |
| **2** | Platelet activation<br>Posttranscriptional regulation of gene expression<br>Intracellular signal transduction |
| **3** | Ephrin receptor signaling pathway<br>Regulation of cell projection organization |
| **4** | Positive regulation of PI3K signaling<br>Positive regulation of GTPase activity<br>Leukocyte migration<br>Innate immune response<br>Regulation of ERK1/2 cascade<br>Transmembrane receptor protein tyrosine kinase signaling pathway<br>Positive regulation of cell migration<br>Positive regulation of cell proliferation<br>Cell differentiation |
| **6** | Mitotic cell cycle process<br>Regulation of cell cycle process<br>Regulation of immune response<br>Positive regulation of hydrolase activity<br>Positive regulation of MAP kinase activity<br>Cell migration<br>MAPK cascade<br>Cellular response to growth factor stimulus<br>Transmembrane receptor protein tyrosine kinase signaling pathway<br>Animal organ morphogenesis |
| **9** | Activin receptor signaling pathway<br>Positive regulation of bone mineralization<br>Positive regulation of pathway-restricted SMAD protein phosphorylation<br>BMP signaling pathway<br>Positive regulation of osteoblast differentiation<br>Artery development<br>Regionalization |
| **10** | Positive regulation of ion transmembrane transporter activity<br>Negative regulation of pancreatic juice secretion<br>Regulation of sodium ion transmembrane transporter activity<br>Negative regulation of sodium ion transport |
| **23** | Inositol phosphate biosynthetic process<br>Biological regulation<br>Cellular macromolecule metabolic process |

Table 3.4: Enrichment for understudied and targeted kinases in the KIN clustered with the consensus SpinGlass algorithm. P-values are calculated from the hypergeometric distribution, where the population size (N) is 453 (total number of kinases in the KIN); the number of "successes" (M) in the KIN are 164 and 327 for understudied and targeted kinases, respectively; the sample size (s) is the number of kinases in the cluster; and the number of observed "successes" (k) is the number of understudied or targeted kinases in a particular subnetwork.

| **Cluster Number** | **Total Number (s)** | **Number Understudied (k)** | **P-value Understudied** | **Number Targeted (k)** | **P-value Targeted** |
|---|---|---|---|---|---|
| **1** | 51 | 19 | 0.491 | 42 | 0.056 (over) |
| **2** | 51 | 10 | 5.4e-3 (under) | 33 | 0.136 |
| **3** | 45 | 10 | 0.026 (under) | 34 | 0.368 |
| **4** | 64 | 9 | 2.47e-5 (under) | 57 | 4.6e-4 (over) |
| **5** | 4 | 3 | 0.137 | 2 | 0.310 |
| **6** | 144 | 68 | 6.8e-4 (over) | 96 | 0.048 (under) |
| **8** | 6 | 4 | 0.129 | 4 | 0.532 |
| **9** | 22 | 10 | 0.240 | 19 | 0.096 |
| **10** | 6 | 2 | 0.623 | 3 | 0.215 |
| **11** | 4 | 2 | 0.459 | 4 | 0.270 |
| **12** | 2 | 0 | 0.406 | 2 | 0.521 |
| **13** | 4 | 4 | 0.017 (over) | 3 | 0.690 |
| **15** | 2 | 2 | 0.131 | 2 | 0.521 |
| **16** | 2 | 0 | 0.406 | 2 | 0.521 |
| **17** | 3 | 1 | 0.702 | 2 | 0.625 |
| **18** | 2 | 2 | 0.131 | 2 | 0.521 |
| **19** | 2 | 1 | 0.594 | 2 | 0.521 |
| **21** | 5 | 2 | 0.595 | 0 | 0.002 (under) |
| **22** | 2 | 0 | 0.406 | 2 | 0.521 |
| **23** | 6 | 0 | 0.066 | 0 | 4.2e-4 (under) |
| **24** | 2 | 2 | 0.131 | 1 | 0.479 |
| **25** | 2 | 2 | 0.131 | 1 | 0.479 |
| **33** | 2 | 1 | 0.362 | 2 | 0.521 |

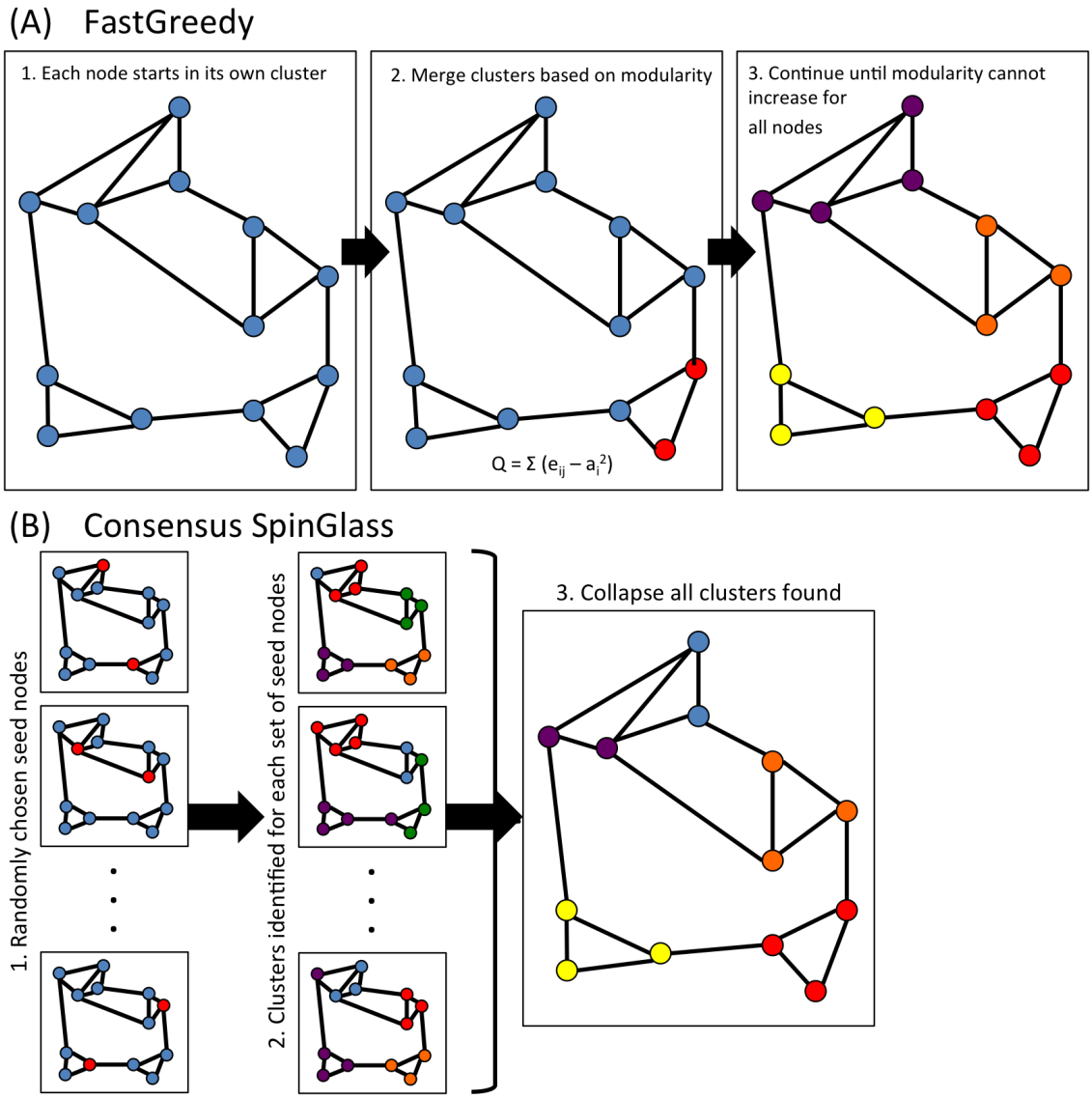



Figure 3.1: Methods for clustering networks into modules, or subnetworks. (A) 1. The FastGreedy algorithm starts with all nodes as individual clusters. 2. Clusters are merged based on modularity ($Q$) scores ($Q = \sum e_{ij} - a_i^2$) for all clusters in the network. 3. This process is continued until the modularity for all clusters cannot be improved by merging clusters. Colors of nodes in the network indicate cluster identity. (B) 1. The SpinGlass algorithm is initialized many times (1000) separately on the same network. For each run, the algorithm chooses seed nodes randomly (red nodes). 2. The SpinGlass algorithm is run for each set of seed nodes chosen. Node color indicates cluster identify, as in (A). 3. Consensus clusters are defined by incorporating all variations in cluster identification from step 2.

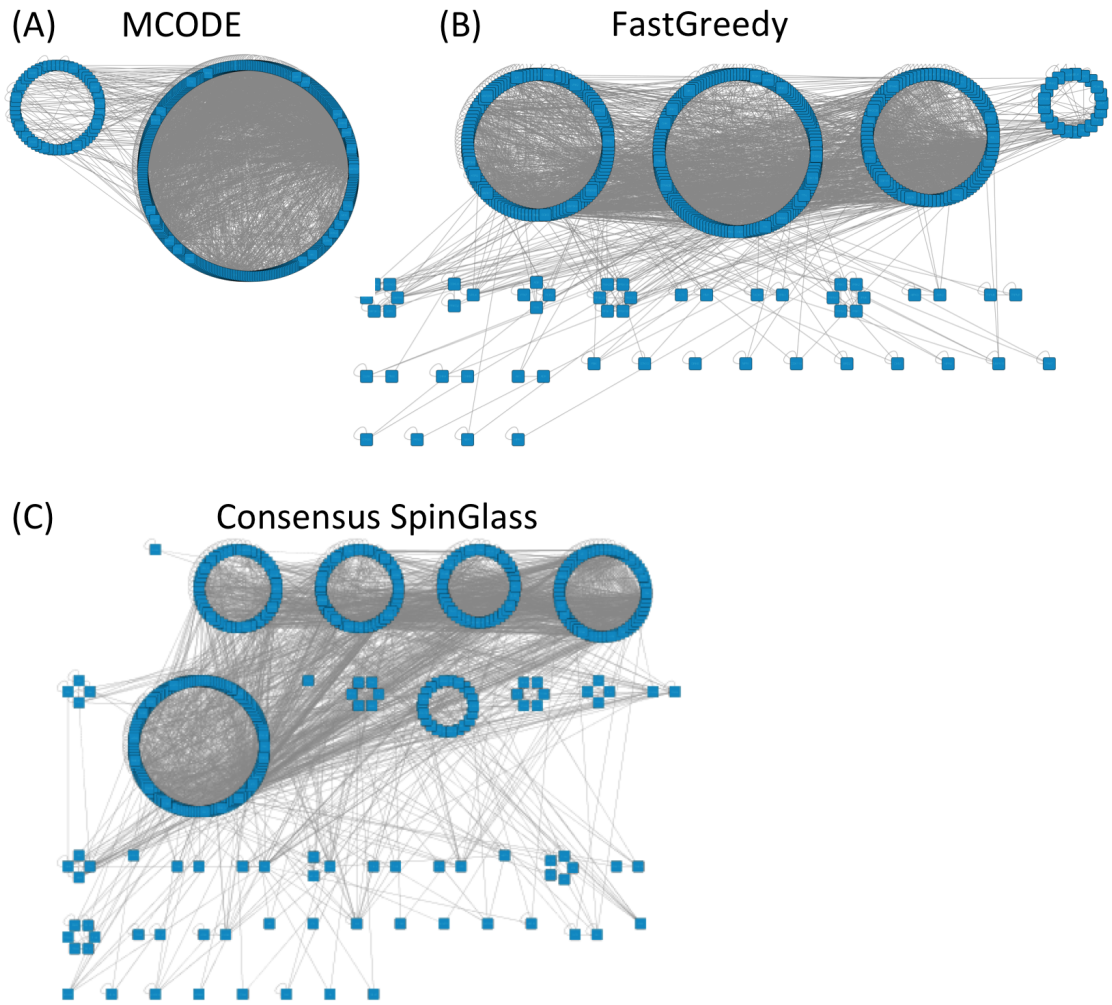



Figure 3.2: Results for Clustering Kinome into Subnetworks. Clustering of network (n = 453, m = 3690) using (A) MCODE, (B) FastGreedy, and (C) consensus SpinGlass algorithms.

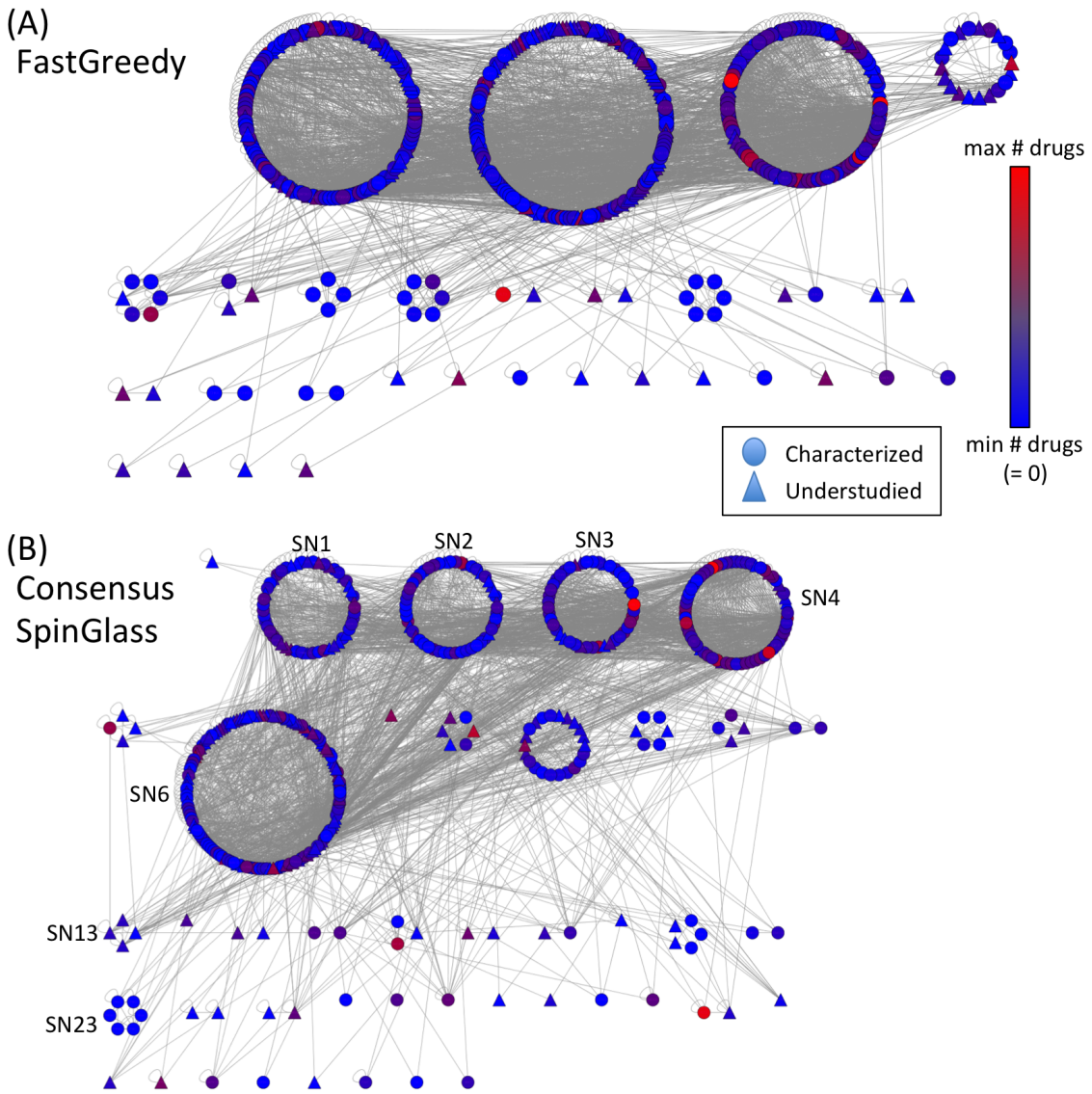



Figure 3.3: Understudied and targeted subnetworks in the KIN using the (A) FastGreedy and (B) consensus SpinGlass algorithms. The shape of each node denotes if the kinase is characterized (circles) or understudied (triangles). The color of the nodes range from blue to purple to red for zero inhibitors targeting to moderate number of inhibitors targeting to maximum number of inhibitors targeting each kinase.

# CHAPTER 4

# OPTIMIZATION FRAMEWORK FOR RATIONALLY CHOOSING COMBINATION DRUG THERAPIES

## 4.1 Introduction

The success of the kinase inhibitor imatinib for the treatment of chronic myeloid leukemia has provided a roadmap for the more general use of small-molecule compounds in targeted cancer therapies [105, 106, 107]. However, while additional anti-cancer compounds continue to be developed, their therapeutic success has been dampened by the development of drug resistance within patients. This resistance can occur through multiple mechanisms including enhanced drug efflux from target cells, altered regulation/over-expression of the drug target, mutations within the target kinase that reduce binding efficacy, the activation of secondary/compensatory pathways, as well as other adaptations [108, 109, 110] .These resistance mechanisms can exist within the tumor de novo, before treatment is given, as well as arise during the course of treatment through selection.

The use of combination therapies has been long recognized as a potential means to overcome resistance to kinase inhibitors, with modeling efforts beginning to provide support for particular treatment strategies [16, 26, 27, 109, 111, 112]. In fact, recent modeling of the evolutionary dynamics of cancer suggests that if there are any possible mutations that confer resistance, sequential administration of targeted compounds will always result in failure, while the use of two or more compounds maintains the opportunity for a successful cure [22].

While the potential value of combination therapies in cancer is recognized, the design of therapies is poorly developed. Clinically, combinations are commonly identified retrospectively, through observation of tumor response profiles and up-regulated pathways, which is exactly the road to sequential treatment discussed previously. High throughput screening methods, though essential, are confronted with exponential growth with the number of drug combinations, doses,

cellular/tissue context and other parameters such that these efforts must be limited to a narrower window of combination possibilities.

In this work, we describe a basic framework for modeling combination therapy as a network optimization problem. Previously, [56] proposed a similar methodology but lacks incorporation of practical costs for potential drug therapies. Specifically, we focus on modeling kinase inhibitors as perturbations that act on the broad network of protein kinases (the kinome) that underlies cell signaling and broadly regulates cell growth, death and other key processes. Given a desired pattern of network inhibition and a set of drugs and their targets, this framework provides the optimum drug combination providing the desired network effect. As part of an optimization framework, this approach allows for the integration of multiple constraints into the final solution. Example constraints include the exclusion or inclusion of specific molecular targets or the prevention of (or requirement for) specific drug combinations. Incorporation of drug dosages and associated side effects, costs and other practical constraints in treatment can be similarly incorporated through extensions to this framework. While conceptually simple, this optimization-based approach provides a quantitative and extendable basis for the rational design of combination therapy regimes.

## 4.2 Methods

### 4.2.1 Kinome Network Databases

Data from the Human Protein Reference Database (HPRD) [82], Interologous Interaction Database (I2D) [94], Reactome [86, 95], Human Integrated Protein-Protein Interaction rEference (HIPPIE) [81], and PhosphoSite [85] were compiled to create an integrated Kinase Interaction Network (KIN) describing known physical interactions between kinases. Only kinase-kinase interactions were considered for the KIN and duplicate interactions (edges) were only counted once.

### 4.2.2 Kinome Subnetwork Identification

We utilized a deterministic community detection method to identify potential modules within the larger KIN. Specifically, we utilized the FastGreedy algorithm implemented within the IGraph package in R to cluster the KIN, which was compiled from multiple databases [97, 113]. Using this algorithm, individual kinases (singletons) are merged into larger communities based on a modularity

score. Merging is iteratively performed until the modularity of the communities can no longer be improved [114]. Network plots are rendered using IGraph in Mathematica [115].

### 4.2.3 Inhibitor Databases

We utilized public data from the Library of Integrated Network-based Cellular Signatures (LINCS) database to determine which kinases are targeted by a given kinase inhibitor [96]. This data is based on the KinomeSCAN assay, which utilizes an active site-directed competition binding assay that measures interactions between the compound and the kinase [116]. As this assay tests compound binding to over 450 kinases, a comprehensive picture of inhibitor binding is achieved, including so-called "off-target" binding to kinases that the inhibitor was not originally designed to target. From LINCS, we utilized data on 144 kinase inhibitors and their binding to over 380 kinases. We required that inhibitors must induce a greater than 90% inhibition of kinase activity or have a $K_d$ (binding dissociation constant) of less than 1$\mu$M in order for them to be considered a drug that targets that particular kinase. A listing of inhibitors and their associated kinases is provided in the software download described later in Methods.

### 4.2.4 Optimization of Targeted Drug Therapy Combination

In this work, we approach problem of identifying an optimal drug combination as an optimization problem, a very general schematic of which is shown in Figure 4.1a. In this example, the goal is to maximize the sum of two variables ($x$ and $y$) subject to several other constraints. Within these constraints, a region of acceptable solutions is defined (green region within the figure) upon which the optimal solution that satisfies the constraints as well as the cost or objective function can be determined.

For our particular application, the optimization being performed is a binary version of the Diet Problem, where each network inhibitor fulfills a particular need - the inhibition of one or more specified kinases [117]. It should be noted that multiple inhibitors are typically able to fulfill a given need. In addition, each inhibitor will usually also target kinases beyond the list of those specified. A central premise of this work is the minimization of off-target inhibition (inhibition of kinases that are not specified by the user); a goal that forms the basis of the cost/objective function used here and described further below.

#### 4.2.4.1 Costs and Objective Function for Optimization

As described previously, we wish to minimize the degree of off-target inhibition occuring through treatment with a given set of drugs. The simplest formulation of this cost is shown in Equation 4.1, where the cost ($c$) for a given inhibitor ($i$), $c_i$, is minimized when there are no off-targets and increases with greater amounts of off-target binding. Again, off-target kinases are simply those not specifically targeted for inhibition.

$$c_i = \frac{\text{\# off-targets}_i}{\text{\# on-targets}_i} \tag{4.1}$$

While we use Equation 4.1 in our objective function, the cost function used can be made much broader and incorporate numerous other goals. For illustration, Equation 4.2 is an example of a cost function, defined for each inhibitor, that incorporates side effects ($SideEffect_i$), the concentration of inhibitor that reduces the response by 50% ($IC50_i$), degree of subnetwork inhibition ($SubNetCoverage_i$), and patient toxicity concentration ($[Toxicity]_i$) into the optimization.

$$c_i = SideEffect_i + SubNetCoverage_i + IC50_i + [Toxicity]_i \tag{4.2}$$

To define notation for this optimization problem, let $x_i$ be the binary value that the optimization problem is trying to solve for. This means that a $x_i$ value of "1" indicates that network inhibitor *i* should be used to obtain the optimal solution (i.e. the optimal combination of inhibitors) and a $x_i$ value of "0" indicates that network inhibitor *i* should not be used in the optimal solution. Also, let $c_i$ be the cost associated with using network inhibitor *i*. The objective function for this optimization problem is to minimize the cost of all of the network inhibitors chosen by the solver that maximally targets the subnetworks. The mathematical representation of this objective function is shown in Equation 4.3.

$$min\{\sum_{i=|drugs|} c_i x_i\} \tag{4.3}$$

#### 4.2.4.2 Constraints for Optimization: Subnetworks

We further utilize network relationships, defined through physical interactions that exist between kinases, as potential constraints for the optimization. The definition of subnetworks is particularly relevant in determining regions over which drugs may have greater (or lesser) inhibition. As an example, a toy network depicting kinases as nodes and edges representing physical interactions is shown in Figure 4.1b. These kinases are then clustered into smaller subnetwork groupings where no singletons exist and all kinases can only belong to a single subnetwork (Figure 4.1c). If we want to ensure that one or more members of a particular subnetwork is being inhibited, we then relate inhibitors and their subnetworks as constraints. Looking at Figure 4.1d, if we want to target subnetwork 2 (SN2 in Figure 4.1d), it is observed that only two drugs ($D1$ and $D2$) target members of this subnetwork. This can then be incorporated as the constraint in Equation 4.4 and indicates that either inhibitor $D1$ or $D2$ or both must be provided in the final solution.

$$\text{Hit SN2: } x_{D1} + x_{D2} \geq 1 \tag{4.4}$$

To avoid inhibiting members of a given subnetwork, similar constraints as above can be constructed. Suppose hitting subnetwork SN2 is not beneficial, with its inhibition associated with strong toxicity responses. Equation 4.5 is the mathematical representation of the constraint to avoid inhibition of members of this subnetwork.

$$\text{Avoid SN2: } x_{D1} + x_{D2} = 0 \tag{4.5}$$

#### 4.2.4.3 Constraints for Optimization: Targeting of Specific Nodes

Similar to earlier constraints, a particular node can be guaranteed to be targeted. Suppose we want to, at a minimum, target node F in Figure 4.1d. This requires that at least one drug targeting F is provided as part of the solution. $D4$ and $D5$ both target node F in this example so either one (or both) of these drugs should be used so that node F is inhibited. This constraint is mathematically represented as Equation 4.6.

$$\text{Hit F: } x_{D4} + x_{D5} \geq 1 \tag{4.6}$$

Similarly, inhibition of a specific kinase may not be desirable for any number of reasons. As before, Equation 4.7 provides the constraint needed so as to avoid targeting node A in the network of Figure 4.1d.

$$\text{Avoid A: } x_{D1} = 0 \tag{4.7}$$

#### 4.2.4.4 Constraints for Optimization: Inhibitor Interactions

A variety of other constraints can be incorporated into the determination of an optimum therapy. Another important constraint commonly encountered within the clinic is the need to take into consideration drug-drug interactions that can occur, often having significant detrimental effects to patients. If adverse effects between two network inhibitors are known, we can avoid their combination as shown in Equation 4.8, where inhibitor $D2$ and $D3$ are not to be used together.

$$\text{Adverse Drug Interaction: } x_{D2} + x_{D3} \leq 1 \tag{4.8}$$

Along these lines, another type of inhibitor constraint is to only allow a certain number of inhibitors in the final cocktail chosen. Equation 4.9 shows this constraint mathematically, where $M$ is the maximum number of inhibitors allowed in the final solution.

$$\text{Limit Number Of Inhibitors Chosen: } \sum_i x_i \leq M \tag{4.9}$$

#### 4.2.4.5 Solving with Linear Programming

We solve the objective function given by Equation 4.3 along with any associated constraints using linear programming (LP) [118, 119]. More specifically, we utilize the dual simplex method implemented with branch-and-bound algorithm within Gurobi [120]. As has been described above, the objective function and all constraints are linear and LP will reliably find optimal solutions to these problems [121, 122].

### 4.2.5 Access to Software/Code and Example Datasets

The method can be run using code provided at *https://github.com/gomezlab/TBD*, which also provides all necessary data sets. R code, which clusters a network into subnetworks and writes the appropriate files for NeTOPT, is also included. NeTOPT can be run from the MATLAB command line by running the NeTOPT_withExamples.m file in the /src/ directory of the repository. The subnetwork files for the examples presented here are included in the repository, which means that NeTOPT can be run without using the R code for the examples presented here. If trying to run using a new input network, the R code will need to be run before employing NeTOPT in MATLAB.

## 4.3 Results

We now provide examples of the application of NeTOPT to the identification of drug combinations under several scenarios, including exploring drug combinations currently being investigated in clinical trials.

### 4.3.1 Example #1: Targeting MTOR, TESK1 and ERBB2

As a first demonstration of this approach, we apply this methodology to a simple example where we require that three kinases, MTOR, TESK1, and ERBB2 (located in SN1, SN4, and SN2 in Figure 4.2(a), respectively) are to be targeted by one or more of the drugs in our inventory (data provided in GitHub repository). In addition to inhibiting these specific targets, we assume that there is an adverse interaction between the drugs Afatinib and HG99101 such that one cannot be used with the other if they were to be chosen for the combination therapy. We note that, built into our objective function, we also want to minimize off-target effects and avoid inhibition of as many other kinases as possible.

With these constraints and the goal of minimizing non-specific targeting codified as the cost function of Equation 4.1, application of NeTOPT provides a drug combination solution consisting of Foretinib (a EGFR, ERBB2, and TESK1 inhibitor) and WYE125132 (a MTOR inhibitor). This solution satisfies all of our constraints while minimizing the number of off-target effects. The degree to which a kinome subnetwork is targeted as described by percent coverage for each drug is also shown in Table 4.1 and simultaneously provides a picture of off-target effects. In this example, while 3 kinases were targeted for inhibition, a total of 128 kinases across six subnetworks were inhibited.

While the results of this simple example could likely be determined by a manual approach, this problem becomes rapidly intractable as the size of the network/number of targets, size of drug library (with their on- and off-target profiles), and number/complexity of constraints grows.

### 4.3.2 Example #2: Subnetwork Targeting

Another simple, but practical example is for the design of combination therapies that maximally inhibit one subnetwork, while minimizing the inhibition of all other subnetworks. For this example, we choose to inhibit subnetwork 2 (SN2) as shown Figure 4.2a. Solving for a combination therapy, we find the set of drugs capable of maximally suppressing this subnetwork that includes A443654, AZD5438, AZD7762, CGP60474, R406, XMD1499, and XMD1527 (Table 4.2). This combination is able to target nearly 58% of SN2. However, while minimization of off-target inhibition is specified, large numbers of kinases are inhibited that lie outside of SN2, highlighting the challenges of obtaining subnetwork-specific inhibition.

### 4.3.3 Example #3: HER2+ Breast Cancer

In HER2-enriched breast cancer, treatment with Lapatinib, a HER2/ERBB2 inhibitor, is a common and effective therapy. However, it has been found that some patients will also have up-regulation of VEGF in addition to the characteristic amplification of HER2/ERBB2, associated with poorer clinical outcomes [123]. Similarly, cross-talk between HER2 and VEGF has been shown to promote tumor growth and angiogenesis, suggesting that inhibition of HER2/ERBB2 along with administration of a VEGFR1/2/3 inhibitor such as Sunitinib or Pazopanib could prove more efficacious [124]. To test this, a combination of Lapatinib plus Pazopanib or Lapatinib alone was evaluated in a clinical trial with patients having advanced/metastatic breast cancer [123]. For the combination, the progressive disease rate (PDR) was found to not be improved relative to monotherapy and toxicity was higher with the combination, though the response rate (RR) was increased.

We ran NeTOPT with the general objective of targeting ERBB2/HER2 as well as VEGFR1/2/3 while minimizing off-targets. Optimization indicated that Lapatinib and AC220 were an optimal combination of kinase inhibitors under the given constraints, with the targeting profile of Lapatinib + AC220 and that of Lapatinib + Pazopanib shown in Figure 4.2(b) and (c), respectively. While

targeting is similar, AC220 does not target two subnetworks that are hit relatively strongly by Pazopanib and has 73% less off-target binding overall (17 kinases inhibited for Lapatinib + AC220 vs 63 for Lapatinib + Pazopanib). Of note is that the AC220 kinase inhibitor is currently in Phase II clinical trials to test efficacy and safety for use as a single agent [125].

### 4.3.4 Example #4: Flavopiridol and Leukemia

Flavopiridol is commonly used in the clinic as a cyclin-dependent kinase inhibitor for chronic lymphocytic leukemia (CLL), acute myeloid leukemia (AML), and chronic myeloid leukemia (CML) [126]. While Flavopiridol targets cell-cycle related kinases, recent clinical trials have focused on trying to assess the effectiveness of inhibiting anti-apoptotic pathways, with inhibitors such as Imatinib in CML and ABL1+ cancers. In particular, ABL1 and ABL2 have been found to be up-regulated in these cancer types and, along with PGFRB, are thought to be important for the cancer cells' ability to reprogram around certain therapies including Flavopiridol [126].

In this example, we wished to see what other inhibitors are identified in combination with Flavopiridol when requiring the concomitant inhibition of ABL1, ABL2 and PGFRB due to their anti-apoptotic function and associated up-regulation. Optimization with these constraints and the cost function for all drugs being that of Equation 4.1, we found that Flavopiridol and Imatinib were identified as the inhibitors best able to satisfy the described constraints while minimizing off-target effects. Imatinib is an already FDA approved drug, while Flavopiridol has completed Phase II trials [125, 127]. In fact, the combination of Flavopiridol and Imatinib is in Phase I clinical trial for ABL+ leukemias [128]. Designated targets for this drug combination are shown in Figure 4.2(d) as are the subnetworks targeted. Associated on- and off-targets of these drugs, alone and in combination, are shown based on the subnetworks targeted in Table 4.4.

We also performed a less constrained prediction where CDK4 and CDK9 were specifically designated as targets in the optimization, while not requiring Flavopiridol to be the inhibitor for these kinases. As before, ABL1/2 and PGFRB were also designated as inhibitor targets. Results of this optimization are shown in Figure 4.2(e) where the designated targets are inhibited by the identified combination of Imatinib and AT7519. This is further detailed in Table 4.4 that shows a 39% decrease (56 vs 34 kinases) in the number of kinases targeted in this new combination versus that using Flavopiridol. The inhibitor AT7519 is currently in Phase I clinical trial for use on its own [125].

These results suggest that NeTOPT has the potential to improve on current combination therapies by identifying those, for example, with fewer off-targets and therefore, potentially fewer negative side-effects within patients.

### 4.4 Discussion

While significant strides have been made in our understanding of cancer pathogenesis, many challenges exist in translating this understanding into novel therapeutics that greatly improve patient outcomes. Resistance to therapies, toxicity effects, and individual patient variation are just a few hurdles to overcome. This work provides a quantitative way to pursue potential drug combinations experimentally.

The mathematical framework presented here provides a quantitative, rational methodology for choosing combination drug therapies for cancer treatment using biological context information. The optimization method allows for a set of drugs to obtain the optimal solution, not just a pair or a single drug. The methodology presented here is flexible to be able to consider many types of constraints in addition to the ones presented here to choose the set of drugs. Notably, constraints directly associated with kinase function in the PPIN can be written to incorporate these into the final solution. For example, all nodes associated with a particular KEGG pathway (i.e. MEK/ERK pathway or another important pathway) or GO annotation (i.e. "apoptosis" or similar) could be added to the "targets" list as input to NeTOPT. These types of constraints could potentially help choose combinations that can prevent or reduce reprogramming in cancer tissue.

For this method, inhibitor targets (on- and off-targets) are all treated equally. Targets for each inhibitor are determined from the KinomeSCAN assay in the LINCS database [96]. This means that if a specific inhibitor is designed to inhibit/bind to ERBB2 but it also inhibits/binds to EGFR, then ERBB2 and EGFR are both considered targets of that inhibitor. Another major assumption for the optimization method involves the cost functions associated with each inhibitor in the LINCS database. We assume that the cost functions accurately capture the "true" cost of using a drug therapy. The kinase inhibitors were used at a concentration of 10 $\mu$M, which is much higher than most of the inhibitors' IC50. This could skew the determination of on- and off-targets determined from this

dataset but this is only for a proof of concept to test the methodology and the high threshold for determining inhibition (over 90%) is used to help offset this bias.

Though the current implementation of NeTOPT provides only a non-binary output, extensions to this approach could incorporate continuous outputs. For instance, while dosage in the form of IC50 can be incorporated. Currently, NeTOPT can incorporate dosing information, such as IC50, into the cost associated with using a particular drug but the result of the optimization is a "yes/no" answer detailing which drugs should be used based on their targeting profiles and costs. A solution that provides two answers (which drugs and their concentrations) would be ideal but this framework could easily result in degeneracy of the LP solve and/or multiple optimal solutions. To account for these issues, further study in the framing of the problem to produce this particular output needs to happen.

For instance, a recent study found that treatment of HER2-enriched breast cancer with lapatinib induced a broad response in kinase activity within 48 hours that compensated for the inhibition of ERBB2 and revealed numerous potential mechanisms for drug resistance [21]. In combination with lapatinib treatment, targeted inhibition of other induced kinases led to improved growth inhibition, but the heterogeneity of responses presented significant challenges for the identification of broadly appropriate drug combinations. In an attempt to make the response to inhibitor more durable, inhibition of the adaptive kinome response was approached through the suppression of BET-bromodomain chromatin readers BRD2/3/4 with the small molecule, JQ1.

Combination of lapatinib with JQ1 reduced lapatinib-induced kinome reprogramming and arrested growth, while either drug alone was unable to do so. In context of the kinome network utilized as part of NeTOPT, JQ1 kinase targets are found to be exclusively localized to a single network (SN8), with three of five kinase within this subnetwork (BRD2/3/4) being inhibited (see Figure 4.2c). JQ1 and other BET-bromodomain inhibitors are currently in a number of human clinical trials to determine their efficacy and safety in combination with other therapies [129]. While requiring additional investigation, the design of therapies that specifically target a significant fraction of one or more well-defined functional subnetworks appears to be a promising approach to combination therapy design. More generally, computational strategies such as the optimization approach described here has the potential to provide a useful framework through which the rational design of combination therapies can be achieved.

Table 4.1: Subnetwork coverage (percentage) of each drug chosen by optimization method in Example #1: Targeting MTOR, TESK1 and ERBB2. Bolded percentages represent the subnetworks of the specific targets for the example presented (MTOR, ERBB2, TESK1).

| | **Foretinib** (149 targets) | **WYE125132** (3 targets) | **Foretinib + WYE125132** |
|---|---|---|---|
| **SN1** | 23.4% (30 targets) | – | **23.4% (30 targets)** |
| **SN2** | 13.4% (24 targets) | 0.1% (1 target) | **14.0% (25 targets)** |
| **SN3** | 51.6% (63 targets) | 1.6% (2 targets) | 53.2% (65 targets) |
| **SN4** | 19.1% (4 targets) | – | **19.1% (4 targets)** |
| **SN5** | 10.0% (1 target) | – | 10.0% (1 target) |
| **SN6** | – | – | – |
| **SN7** | 75.0% (3 targets) | – | 75.0% (3 targets) |
| **SN8** | – | – | – |

Table 4.2: Subnetwork coverage (percentage) of each drug chosen by optimization method (A443654, AZD5438, AZD7762, CGP60474, R406, XMD1499, XMD1527) for Example #2: Subnetwork Targeting.

| | **A443654** (81 targets) | **AZD5438** (127 targets) | **AZD7762** (189 targets) | **CGP60474** (11 targets) | **R406** (221 targets) | **XMD1499** (17 targets) | **XMD1527** (53 targets) | **Combination Of All** |
|---|---|---|---|---|---|---|---|---|
| **SN1** | 16.4% (21 targets) | 20.3% (26 targets) | 24.2% (31 targets) | 1.6% (2 targets) | 32.8% (42 targets) | 1.6% (2 targets) | 8.6% (11 targets) | 50.8% (65 targets) |
| **SN2** | 15.6% (28 targets) | 24.6% (44 targets) | 30.7% (55 targets) | 3.9% (7 targets) | 29.1% (52 targets) | 3.9% (7 targets) | 12.3% (22 targets) | **57.5% (103 targets)** |
| **SN3** | 6.6% (8 targets) | 17.2% (21 targets) | 45.9% (56 targets) | – | 60.7% (74 targets) | – | 3.3% (4 targets) | 70.5% (86 targets) |
| **SN4** | 4.8% (1 target) | 9.5% (2 targets) | 19.1% (4 targets) | – | 28.6% (6 targets) | 4.8% (1 target) | 4.8% (1 target) | 33.3% (7 targets) |
| **SN5** | 30.0% (3 targets) | 30.0% (3 targets) | 30.0% (3 targets) | 10.0% (1 target) | 30.0% (3 targets) | – | 20.0% (2 targets) | 60.0% (6 targets) |
| **SN6** | – | – | – | – | – | – | – | – |
| **SN7** | 50.0% (2 targets) | 25.0% (1 target) | – | – | 25.0% (1 target) | – | – | 100% (4 targets) |
| **SN8** | – | – | – | – | – | – | – | – |

Table 4.3: Subnetwork coverage (percentage) of each drug chosen by optimization method in Example #3: HER2+ Breast Cancer. Bolded percentages represent the subnetworks of the specific targets for the example presented (ERBB2 and VGFR1/2/3).

| | **Lapatinib** (7 targets) | **AC220** (17 targets) | **Pazopanib** (49 targets) | **Lapatinib + AC220** | **Lapatinib + Pazopanib** |
|---|---|---|---|---|---|
| **SN1** | 0.1% (1 target) | 1.6% (2 targets) | 9.4% (12 targets) | 2.3% (3 targets) | 9.4% (12 targets) |
| **SN2** | – | 0.1% (1 target) | 5.0% (9 targets) | 0.1% (1 target) | 5.0% (9 targets) |
| **SN3** | 2.5% (3 targets) | 7.4% (9 targets) | 14.8% (18 targets) | **9.8% (12 targets)** | **17.2% (21 targets)** |
| **SN4** | – | 4.8% (1 target) | 9.5% (2 targets) | 4.8% (1 target) | 9.5% (2 targets) |
| **SN5** | – | – | 10.0% (1 target) | – | 10.0% (1 target) |
| **SN6** | – | – | – | – | – |
| **SN7** | – | – | 25.0% (1 target) | – | 25.0% (1 target) |
| **SN8** | – | – | – | – | – |

Table 4.4: Subnetwork coverage (percentage) of each drug chosen by optimization method in Example #4: Flavopiridol and Leukemia. Bolded percentages represent the subnetworks of the specific targets for the example presented (Flavopiridol plus targeting ABL1, ABL2, PGFRB).

| | **Flavopiridol** (58 targets) | **AT7519** (30 targets) | **Imatinib** (13 targets) | **Flavopiridol + Imatinib** | **AT7519 + Imatinib** |
|---|---|---|---|---|---|
| **SN1** | 10.2% (13 targets) | 1.6% (2 targets) | – | 10.2% (13 targets) | 1.6% (2 targets) |
| **SN2** | 13.4% (24 targets) | 10.1% (18 targets) | 0.1% (1 target) | **14.0% (25 targets)** | **10.6% (19 targets)** |
| **SN3** | 4.9% (6 targets) | – | 9.0% (11 targets) | **12.3% (15 targets)** | **9.0% (11 targets)** |
| **SN4** | 4.8% (1 target) | – | 4.8% (1 target) | 9.5% (2 targets) | 4.8% (1 target) |
| **SN5** | 10.0% (1 target) | 10.0% (1 target) | – | 10.0% (1 target) | 10.0% (1 target) |
| **SN6** | – | – | – | – | – |
| **SN7** | – | – | – | – | – |
| **SN8** | – | – | – | – | – |

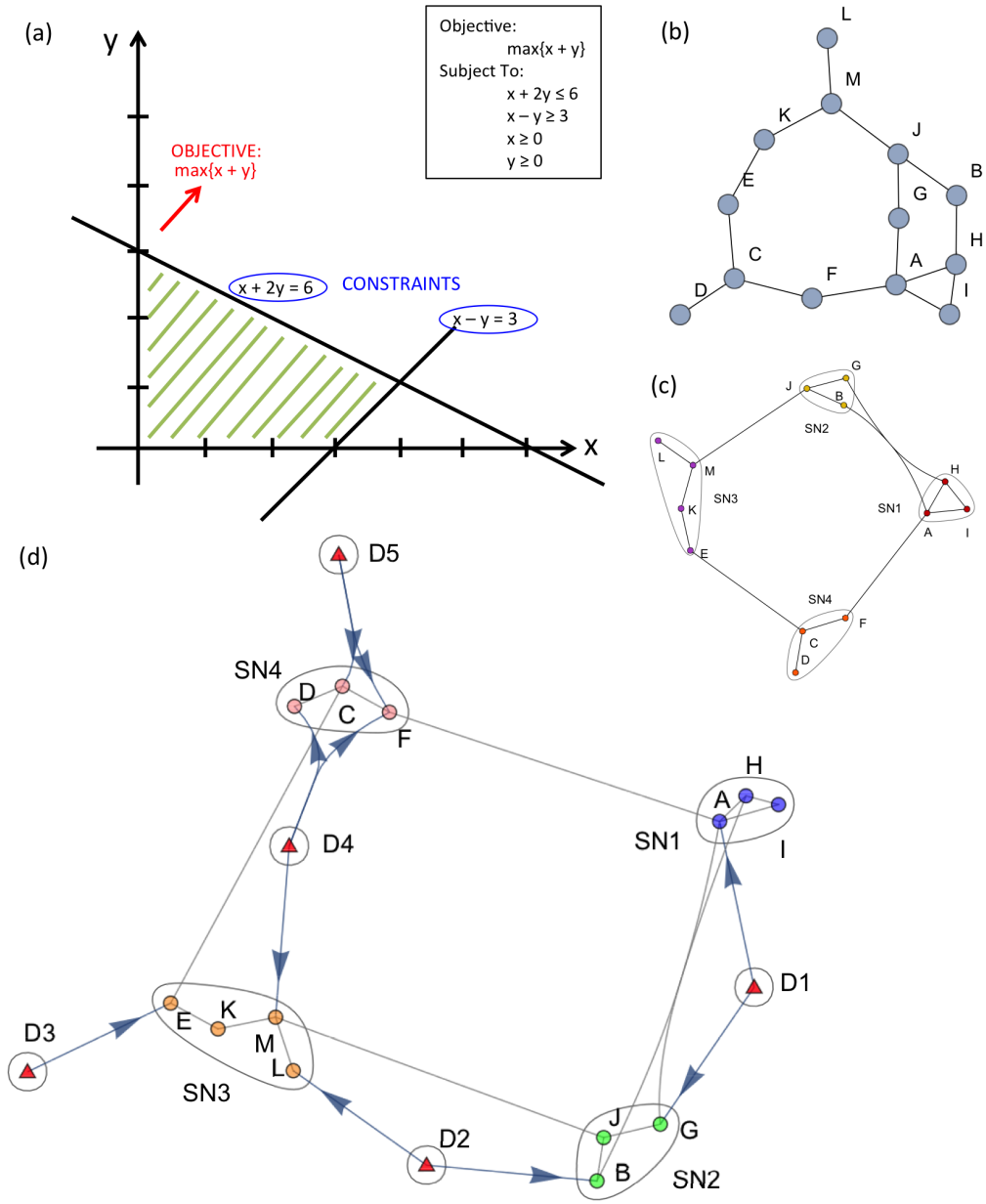



Figure 4.1: (a) Graphical explanation of linear optimization with objective function and constraints. The mathematical constraints shown represent a biological "rule". (b) Network depiction of an example network. (c) Network of the results of clustering the example network in (b) using FastGreedy. (d) Network depiction of the relationships of the inhibitors (red triangles–D1 through D5) to the nodes and subnetworks in the example network.

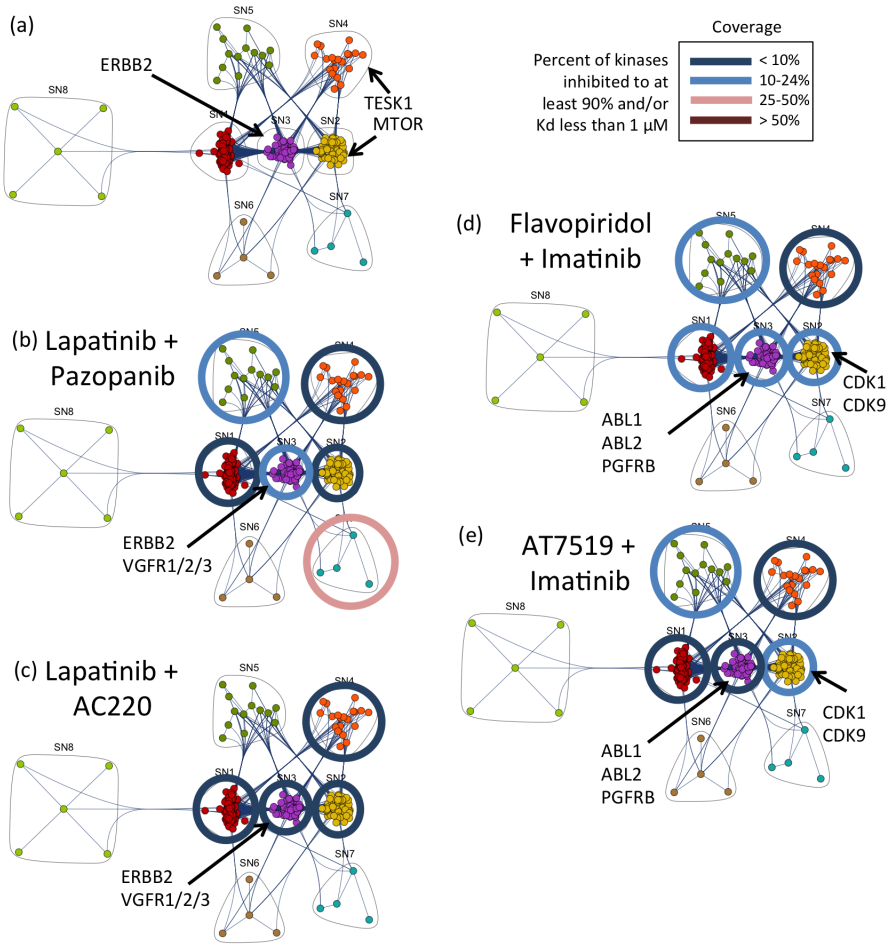



Figure 4.2: (a) Graphical depiction of Examples #1 and #2 targets as input for NeTOPT. (b) Targeting profile of Lapatinib plus Pazopanib with location of the targets noted, from Example #3, in the PPIN. The color of the circle around a subnetwork indicates how many kinases are targeted by the inhibitor (dark blue: $<10\% \rightarrow$ dark red: $>50\%$). (c) Targeting profile of Lapatinib plus AC220 with location of the targets noted, from Example #3, in the PPIN. Colors of the circles around subnetworks is the same as in (b). (d) Targeting profile of Flavopiridol + Imatinib with the location of the targets, from Example #4, in the PPIN. Colors of the circles around subnetworks is the same as in (b). (e) Targeting profile of AT7519 + Imatinib with the location of the targets, from Example #4, in the PPIN. Color or the circles around subnetworks are the same as in (b).

# CHAPTER 5
# CONCLUSIONS

## 5.1 Summary of Thesis Contributions

In Chapter 2 of this thesis, functional kinome profiles of characterized and understudied kinases distinguish breast cancer subtypes. Additionally, the functional profiles provide complementary information to gene expression levels about the breast cancer subtypes. The kinases that best describe the separation between the subtypes in the functional kinome are spread throughout the kinome interaction network, are responsible for a variety of functions in cells, and include characterized and understudied kinases.

Chapter 3 focuses on characterizing the structure of the kinome and relating the architecture of the network to biological function. Understudied and undertargeted kinases are very important in the signaling in the cell. Being understudied or undertargeted does not equal unimportant–it simply means research has not focused on them thus far. Some subnetworks of kinases are overrepresented with understudied/targeted kinases, while some are underrepresented in understudied/targeted kinases. Kinase inhibitors are generally designed for "easy-to-hit" kinases, such as MEK1/2. Some kinases, like cMyc or Ras, are untargettable and therefore have to be targeted indirectly. Kinase inhibitors hit across the subnetworks identified but are enriched in some subnetworks in the KIN.

The model presented in Chapter 4 provides a quantitative and unbiased method to choose a set of drugs to administer when certain targets need to be hit. This allows biologists to set their own target list to combat reprogramming and incorporate realistic clinical considerations (toxicity, number of inhibitors in combination therapy, adverse drug interactions, etc.).

## 5.2 Impact on Cancer Biology

This work contributes computational and quantitative approaches to further the study of kinases, which are primary targets for many cancer therapeutics, by shedding light on potential functions of

understudied kinases and identifying undertargeted subnetworks, or groups, of kinases. The potential to repurpose drugs for new biological/clinical contexts is also an important contribution to cancer biology.

### 5.3 Challenges

The lack of comprehensive data set that are completely consistent across data types is a challenge in quantitative studies, like the ones presented in this thesis. This challenge will be less and less impactful as the amount of public data, across experimental platforms, increases in the near future. Another main challenge for this work is the completeness of the biological annotations of protein functions. Many proteins have very little information (if any) on their biological role, which can make directing therapeutic development/repurposing challenging. The work presented in this thesis will help with this challenge to provide some insight into potential biological functions for these understudied proteins but experimental testing and validation is still needed.

### 5.4 Future Directions

Further development of analyses and techniques to characterize the understudied kinome should be explored. Regression methods, such as lasso and elastic net regression, can be used to associate patterns of MIB binding of characterized kinases to understudied kinases [130, 131]. The characterized kinases that tend to be associated with the functional levels of the understudied kinases could provide insight into the biological function of the understudied kinase. MIB binding profiles of cell lines and tumors could be used in conjunction with proteomic data on tumors from public sources to expand the repertoire of samples [132]. Network clustering of a compiled protein-protein interaction network (PPIN) can also indicate the function of understudied kinases and could direct drug studies as well. Similarly, regression methods can help determine which kinases are responsible for reprogramming after drug treatment, which can inform drug treatment decisions [130, 131].

Extension of the optimization methodology presented in this thesis is also an area for further study. The output of the current optimization method is binary: 1 for choosing a drug, 0 for not choosing a drug. It would be ideal to have the output of this method to be continuous instead of binary. To do this, it would require dosing information for each drug it is possible to choose, which

is a large experimental undertaking since the drugs would have to be tested in multiple contexts to make the computational approach feasible. Another expansion of the optimization method would be to incorporate network and subnetwork information into the input. As more experimental data is generated to solidify protein-protein interactions and the actions of drugs in the PPIN, the optimization method will become "smarter" by incorporating different data types to make a more informed decision about the "best" drugs to include in a combination therapy.

## REFERENCES


[1] The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*, advance on, Sep 2012.

[2] Antoinette Hollestelle, Jord H A Nagel, Marcel Smid, Suzanne Lam, Fons Elstrodt, Marijke Wasielewski, Ser Sue Ng, Pim J French, Justine K Peeters, Marieke J Rozendaal, Muhammad Riaz, Daphne G Koopman, Timo L M Ten Hagen, Bertie H C G M de Leeuw, Ellen C Zwarthoff, Amina Teunisse, Peter J van der Spek, Jan G M Klijn, Winand N M Dinjens, Stephen P Ethier, Hans Clevers, Aart G Jochemsen, Michael A den Bakker, John A Foekens, John W M Martens, and Mieke Schutte. Distinct gene mutation profiles among luminal-type and basal-type breast cancer cell lines. *Breast cancer research and treatment*, 121(1):53–64, May 2010.

[3] Yuan Tang, Yue Wang, Mohammad F. Kiani, and Bin Wang. Classification, Treatment Strategy, and Associated Drug Resistance in Breast Cancer. *Clinical Breast Cancer*, 16(5):335–343, Oct 2016.

[4] Stuart J Schnitt. Classification and prognosis of invasive breast cancer: from morphology to molecular taxonomy. *Modern Pathology*, 23:S60–S64, May 2010.

[5] M. C.U. Cheang, D. Voduc, C. Bajdik, S. Leung, S. McKinney, S. K. Chia, C. M. Perou, and T. O. Nielsen. Basal-Like Breast Cancer Defined by Five Biomarkers Has Superior Prognostic Value than Triple-Negative Phenotype. *Clinical Cancer Research*, 14(5):1368–1376, Mar 2008.

[6] Maggie C U Cheang, Stephen K Chia, David Voduc, Dongxia Gao, Samuel Leung, Jacqueline Snider, Mark Watson, Sherri Davies, Philip S Bernard, Joel S Parker, Charles M Perou, Matthew J Ellis, and Torsten O Nielsen. Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. *Journal of the National Cancer Institute*, 101(10):736–50, May 2009.

[7] Michail Ignatiadis and Christos Sotiriou. Luminal breast cancer: from biology to treatment. *Nature reviews. Clinical oncology*, 10(9):494–506, Sep 2013.

[8] C M Perou, T Sø rlie, M B Eisen, M van de Rijn, S S Jeffrey, C a Rees, J R Pollack, D T Ross, H Johnsen, L a Akslen, O Fluge, a Pergamenschikov, C Williams, S X Zhu, P E Lø nning, a L Bø rresen Dale, P O Brown, and D Botstein. Molecular portraits of human breast tumours. *Nature*, 406(6797):747–52, Aug 2000.

[9] Chad Creighton. The molecular profile of luminal B breast cancer. *Biologics: Targets and Therapy*, 6:289, Aug 2012.

[10] Christos Sotiriou and Lajos Pusztai. Gene-expression signatures in breast cancer. *The New England journal of medicine*, 360(8):790–800, Feb 2009.

[11] Aleix Prat and Charles M. Perou. Deconstructing the molecular portraits of breast cancer. *Molecular Oncology*, 5(1):5–23, Feb 2011.

[12] Clifford A Hudis and Luca Gianni. Triple-negative breast cancer: an unmet medical need. *The oncologist*, 16 Suppl 1(suppl 1):1–11, Jan 2011.

[13] Fares Al-Ejeh, Wei Shi, Mariska Miranda, Peter T Simpson, Ana Cristina Vargas, Sarah Song, Adrian P Wiegmans, Alex Swarbrick, Alana L Welm, Michael P Brown, Georgia Chenevix-Trench, Sunil R Lakhani, and Kum Kum Khanna. Treatment of triple-negative breast cancer using anti-EGFR-directed radioimmunotherapy combined with radiosensitizing chemotherapy and PARP inhibitor. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, 54(6):913–21, Jun 2013.

[14] Shi Ma. Estrogen receptor-positive breast cancer molecular signatures and therapeutic potentials (Review). *Biomedical Reports*, Oct 2013.

[15] P Blume-Jensen and T Hunter. Oncogenic kinase signalling. *Nature*, 411(6835):355–65, May 2001.

[16] Jianming Zhang, Priscilla L Yang, and Nathanael S Gray. Targeting cancer with small molecule kinase inhibitors. *Nature reviews. Cancer*, 9(1):28–39, Jan 2009.

[17] James S Duncan, Martin C Whittle, Kazuhiro Nakamura, Amy N Abell, Alicia A Midland, Jon S Zawistowski, Nancy L Johnson, Deborah A Granger, Nicole Vincent Jordan, David B Darr, Jerry Usary, Pei-Fen Kuan, David M Smalley, Ben Major, Xiaping He, Katherine A Hoadley, Bing Zhou, Norman E Sharpless, Charles M Perou, William Y Kim, Shawn M Gomez, Xin Chen, Jian Jin, Stephen V Frye, H Shelton Earp, Lee M Graves, and Gary L Johnson. Dynamic reprogramming of the kinome in response to targeted MEK inhibition in triple-negative breast cancer. *Cell*, 149(2):307–21, Apr 2012.

[18] Oleg Fedorov, Susanne Müller, and Stefan Knapp. The (un)targeted cancer kinome. *Nature Chemical Biology*, 6(3):166–169, Mar 2010.

[19] Tahrin Mahmood and Ping-Chang Yang. Western blot: technique, theory, and trouble shooting. *North American journal of medical sciences*, 4(9):429–34, Sep 2012.

[20] Brett Spurrier, Sundhar Ramalingam, and Satoshi Nishizuka. Reverse-phase protein lysate microarrays for cell signaling analysis. *Nature Protocols*, 3(11):1796–1808, oct 2008.

[21] Timothy J Stuhlmiller, Samantha M Miller, Jon S Zawistowski, Kazuhiro Nakamura, Adriana S Beltran, James S Duncan, Steven P Angus, Kyla A L Collins, Deborah A Granger, Rachel A Reuther, Lee M Graves, Shawn M Gomez, Pei-Fen Kuan, Joel S Parker, Xin Chen, Noah Sciaky, Lisa A Carey, H Shelton Earp, Jian Jin, and Gary L Johnson. Inhibition of Lapatinib-Induced Kinome Reprogramming in ERBB2-Positive Breast Cancer by Targeting BET Family Bromodomains. *Cell reports*, 11(3):390–404, Apr 2015.

[22] Ivana Bozic, Johannes G Reiter, Benjamin Allen, Tibor Antal, Krishnendu Chatterjee, Preya Shah, Yo Sup Moon, Amin Yaqubie, Nicole Kelly, Dung T Le, Evan J Lipson, Paul B Chapman, Luis A Diaz, Bert Vogelstein, and Martin A Nowak. Evolutionary dynamics of cancer in response to targeted combination therapy. *eLife*, 2:e00747, Jun 2013.

[23] Sandra M. Swain, José Baselga, Sung-Bae Kim, Jungsil Ro, Vladimir Semiglazov, Mario Campone, Eva Ciruelos, Jean-Marc Ferrero, Andreas Schneeweiss, Sarah Heeson, Emma Clark, Graham Ross, Mark C. Benyunes, and Javier Cortés. Pertuzumab, Trastuzumab, and Docetaxel in HER2-Positive Metastatic Breast Cancer. *New England Journal of Medicine*, 372(8):724–734, Feb 2015.

[24] Richard S Finn, John P Crown, Istvan Lang, Katalin Boer, Igor M Bondarenko, Sergey O Kulyk, Johannes Ettl, Ravindranath Patel, Tamas Pinter, Marcus Schmidt, Yaroslav Shparyk, Anu R Thummala, Nataliya L Voytko, Camilla Fowst, Xin Huang, Sindy T Kim, Sophia Randolph, and Dennis J Slamon. The cyclin-dependent kinase 4/6 inhibitor palbociclib in combination with letrozole versus letrozole alone as first-line treatment of oestrogen receptor-positive, HER2-negative, advanced breast cancer (PALOMA-1/TRIO-18): a randomised phase 2 study. *The Lancet Oncology*, 16(1):25–35, Jan 2015.

[25] A Srinivas Reddy and Shuxing Zhang. Polypharmacology: drug discovery for the future. *Expert review of clinical pharmacology*, 6(1):41–7, Jan 2013.

[26] Zachary A. Knight, Henry Lin, and Kevan M. Shokat. Targeting the cancer kinome through polypharmacology. *Nature Reviews Cancer*, 10(2):130–137, Feb 2010.

[27] Dimitrios Zardavas, José Baselga, and Martine Piccart. Emerging targeted agents in metastatic breast cancer. *Nature reviews. Clinical oncology*, 10(4):191–210, Apr 2013.

[28] Heiko Hoffmann. Kernel PCA for novelty detection. *Pattern Recognition*, 40(3):863–874, Mar 2007.

[29] Ronald Jansen, Haiyuan Yu, Dov Greenbaum, Yuval Kluger, Nevan J Krogan, Sambath Chung, Andrew Emili, Michael Snyder, Jack F Greenblatt, and Mark Gerstein. A Bayesian networks approach for predicting protein-protein interactions from genomic data. *Science (New York, N.Y.)*, 302(5644):449–53, October 2003.

[30] Svante Wold, Paul Geladi, Kim Esbensen, and Jerker Öhman. Multi-way principal components-and PLS-analysis. *Journal of Chemometrics*, 1(1):41–56, Jan 1987.

[31] Yudi Pawitan, Judith Bjöhle, Lukas Amler, Anna-Lena Borg, Suzanne Egyhazi, Per Hall, Xia Han, Lars Holmberg, Fei Huang, Sigrid Klaar, Edison T Liu, Lance Miller, Hans Nordgren, Alexander Ploner, Kerstin Sandelin, Peter M Shaw, Johanna Smeds, Lambert Skoog, Sara Wedrén, and Jonas Bergh. Gene expression profiling of breast cancer survivability by pooled cDNA microarray analysis using logistic regression, artificial neural networks and decision trees. *Breast Cancer Research*, 7(6):R953, 2005.

[32] David W Mount, Charles W Putnam, Sara M Centouri, Ann M Manziello, Ritu Pandey, Linda L Garland, and Jesse D Martinez. Using logistic regression to improve the prognostic value of microarray gene expression data sets: application to early-stage squamous cell carcinoma of the lung and triple negative breast carcinoma. *BMC medical genomics*, 7:33, Jun 2014.

[33] Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *The Journal of Machine Learning Research*, 3:1157–1182, Mar 2003.

[34] T S Furey, N Cristianini, N Duffy, D W Bednarski, M Schummer, and D Haussler. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics (Oxford, England)*, 16(10):906–14, Oct 2000.

[35] Ron Kohavi. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *International Joint Conference on Artificial Intelligence*, 1995.

[36] D.M.W. Powers. Evaluation: From precision, recall and f-factor to rox informedness, markedness & correlation. *Journal of Machine Learning Technologies*, 2(1):37–63, 2011.

[37] Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3):75–174, 2010.

[38] Aaron Clauset, M. E. J. Newman, and Cristopher Moore. Finding community structure in very large networks. *arXiv*, Aug 2004.

[39] Joerg Reichardt and Stefan Bornholdt. Statistical Mechanics of Community Detection. *arXiv*, Mar 2006.

[40] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69(2):026113, Feb 2004.

[41] V. A. Traag and Jeroen Bruggeman. Community detection in networks with positive and negative links. *arXiv*, Nov 2008.

[42] Di Chen, Xi Liu, Yiping Yang, Hongjun Yang, and Peng Lu. Systematic synergy modeling: understanding drug synergy from a systems biology perspective. *BMC Systems Biology*, 9(1):56, Dec 2015.

[43] Karen A Ryall and Aik Choon Tan. Systems biology approaches for advancing the discovery of effective drug combinations. *Journal of cheminformatics*, 7(1):7, Jan 2015.

[44] Xiaochen Sun, Santiago Vilar, and Nicholas P. Tatonetti. High-Throughput Methods for Combinatorial Drug Discovery. *Science Translational Medicine*, 5(205), 2013.

[45] Jing Tang and Tero Aittokallio. Network pharmacology strategies toward multi-target anticancer therapies: from computational models to experimental design principles. *Current pharmaceutical design*, 20(1):23–36, 2014.

[46] Laura J van 't Veer, Hongyue Dai, Marc J van de Vijver, Yudong D He, Augustinus A M Hart, Mao Mao, Hans L Peterse, Karin van der Kooy, Matthew J Marton, Anke T Witteveen, George J Schreiber, Ron M Kerkhoven, Chris Roberts, Peter S Linsley, René Bernards, and Stephen H Friend. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, 415(6871):530–6, Jan 2002.

[47] Jihye Kim, Vihas T Vasu, Rangnath Mishra, Katherine R Singleton, Minjae Yoo, Sonia M Leach, Eveline Farias-Hesson, Robert J Mason, Jaewoo Kang, Preveen Ramamoorthy, Jeffrey A Kern, Lynn E Heasley, James H Finigan, and Aik Choon Tan. Bioinformatics-driven discovery of rational combination for overcoming EGFR-mutant lung cancer resistance to EGFR therapy. *Bioinformatics (Oxford, England)*, page 323, May 2014.

[48] Liang Zhao, M Guillaume Wientjes, and Jessie L-S Au. Evaluation of combination chemotherapy: integration of nonlinear regression, curve shift, isobologram, and combination index analyses. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 10(23):7994–8004, Dec 2004.

[49] Ting-Chao Chou. Theoretical Basis, Experimental Design, and Computerized Simulation of Synergism and Antagonism in Drug Combination Studies. *Pharmacological Reviews*, 58(3), 2006.

[50] R. J. Tallarida. Quantitative Methods for Assessing Drug Synergism. *Genes & Cancer*, 2(11):1003–1008, Nov 2011.

[51] Luca Grieco, Laurence Calzone, Isabelle Bernard-Pierrot, François Radvanyi, Brigitte Kahn-Perlès, and Denis Thieffry. Integrative Modelling of the Influence of MAPK Network on Cancer Cell Fate Decision. *PLoS Computational Biology*, 9(10):e1003286, Oct 2013.

[52] Oliver E Sturm, Richard Orton, Joan Grindlay, Marc Birtwistle, Vladislav Vyshemirsky, David Gilbert, Muffy Calder, Andrew Pitt, Boris Kholodenko, and Walter Kolch. The mammalian MAPK/ERK pathway exhibits properties of a negative feedback amplifier. *Science signaling*, 3(153):ra90, Jan 2010.

[53] F. Cheng and Z. Zhao. Machine learning-based prediction of drug-drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties. *Journal of the American Medical Informatics Association*, 21(e2):e278–e286, Oct 2014.

[54] Sergio Iadevaia, Yiling Lu, Fabiana C. Morales, Gordon B. Mills, and Prahlad T. Ram. Identification of optimal drug combinations targeting cellular networks: Integrating phospho-proteomics and computational network analysis. *Cancer Research*, 2010.

[55] Alexei Vazquez, AA Borisy, PJ Elliott, NW Hurst, MS Lee, J Lehár, ER Price, G Serbedzija, GR Zimmermann, MA Foley, BR Stockwell, CT Keith, RH Shoemaker, PK Wong, F Yu, A Shahangian, G Cheng, R Sun, C Ho, MA Yildirim, KL Goh, M Cusick, AL Barabási, M Vidal, S Nelander, W Wang, B Nilsson, QB She, C Pratilas, N Rosen, P Gennemark, C Sander, M Campillo, M Kuhn, A Gavin, LJ Jensen, P Bork, MR Garey, DS Johnson, V Vazirani, DS Johnson, DT Teachey, C Sheen, J Hall, T Ryan, VI Brown, J Fish, GS Reid, AE Seif, R Norris, YJ Chang, M Carroll, SA Grupp, M Mézard, and M Tarzia. Optimal drug combinations and minimal hitting sets. *BMC Systems Biology*, 3(1):81, 2009.

[56] K. Pang, Y.-W. Wan, W. T. Choi, L. A. Donehower, J. Sun, D. Pant, and Z. Liu. Combinatorial therapy discovery using mixed integer linear programming. *Bioinformatics*, 30(10):1456–1463, May 2014.

[57] Joel S Parker, Michael Mullins, Maggie C U Cheang, Samuel Leung, David Voduc, Tammi Vickery, Sherri Davies, Christiane Fauron, Xiaping He, Zhiyuan Hu, John F Quackenbush, Inge J Stijleman, Juan Palazzo, J S Marron, Andrew B Nobel, Elaine Mardis, Torsten O Nielsen, Matthew J Ellis, Charles M Perou, and Philip S Bernard. Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 27(8):1160–7, Mar 2009.

[58] Aleix Prat, Barbara Adamo, Cheng Fan, Vicente Peg, Maria Vidal, Patricia Galván, Ana Vivancos, Paolo Nuciforo, Héctor G Palmer, Shaheenah Dawood, Jordi Rodón, Santiago Ramony Cajal, Josep Maria Del Campo, Enriqueta Felip, Josep Tabernero, and Javier Cortés. Genomic analyses across six cancer types identify basal-like breast cancer as a unique molecular entity. *Scientific reports*, 3:3544, Jan 2013.

[59] Aleix Prat, Maggie Chon U Cheang, Miguel Martín, Joel S Parker, Eva Carrasco, Rosalía Caballero, Scott Tyldesley, Karen Gelmon, Philip S Bernard, Torsten O Nielsen, and Charles M Perou. Prognostic significance of progesterone receptor-positive tumor cells within immunohistochemically defined luminal A breast cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 31(2):203–9, Jan 2013.

[60] Katherine A. Hoadley, Christina Yau, Denise M. Wolf, Andrew D. Cherniack, David Tamborero, Sam Ng, Max D.M. Leiserson, Beifang Niu, Michael D. McLellan, Vladislav Uzunangelov, Jiashan Zhang, Cyriac Kandoth, Rehan Akbani, Hui Shen, Larsson Omberg, Andy

Chu, Adam A. Margolin, Laura J. van't Veer, Nuria Lopez-Bigas, Peter W. Laird, Benjamin J. Raphael, Li Ding, A. Gordon Robertson, Lauren A. Byers, Gordon B. Mills, John N. Weinstein, Carter Van Waes, Zhong Chen, Eric A. Collisson, Christopher C. Benz, Charles M. Perou, and Joshua M Stuart. Multiplatform Analysis of 12 Cancer Types Reveals Molecular Classification within and across Tissues of Origin. *Cell*, 158(4):929–944, Aug 2014.

[61] Alicia A Midland, Martin C Whittle, James S Duncan, Amy N Abell, Kazuhiro Nakamura, Jon S Zawistowski, Lisa A Carey, H Shelton Earp III, Lee M Graves, Shawn M Gomez, and Gary L Johnson. Defining the expressed breast cancer kinome. *Cell Research*, 22(4):620–623, Apr 2012.

[62] Henrik Daub, Jesper V. Olsen, Michaela Bairlein, Florian Gnad, Felix S. Oppermann, Roman Körner, Zoltán Greff, György Kéri, Olaf Stemmann, and Matthias Mann. Kinase-Selective Enrichment Enables Quantitative Phosphoproteomics of the Kinome across the Cell Cycle. *Molecular Cell*, 31(3):438–448, Aug 2008.

[63] Jake E Delmore, Ghayas C Issa, Madeleine E Lemieux, Peter B Rahl, Junwei Shi, Hannah M Jacobs, Efstathios Kastritis, Timothy Gilpatrick, Ronald M Paranal, Jun Qi, Marta Chesi, Anna C Schinzel, Michael R McKeown, Timothy P Heffernan, Christopher R Vakoc, P Leif Bergsagel, Irene M Ghobrial, Paul G Richardson, Richard A Young, William C Hahn, Kenneth C Anderson, Andrew L Kung, James E Bradner, and Constantine S Mitsiades. BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell*, 146(6):904–17, Sep 2011.

[64] Richard Marcotte, Kevin R Brown, Fernando Suarez, Azin Sayad, Konstantina Karamboulas, Paul M Krzyzanowski, Fabrice Sircoulomb, Mauricio Medrano, Yaroslav Fedyshyn, Judice L Y Koh, Dewald van Dyk, Bohdana Fedyshyn, Marianna Luhova, Glauber C Brito, Franco J Vizeacoumar, Frederick S Vizeacoumar, Alessandro Datti, Dahlia Kasimer, Alla Buzina, Patricia Mero, Christine Misquitta, Josee Normand, Maliha Haider, Troy Ketela, Jeffrey L Wrana, Robert Rottapel, Benjamin G Neel, and Jason Moffat. Essential gene profiles in breast, pancreatic, and ovarian cancer cells. *Cancer discovery*, 2(2):172–89, Feb 2012.

[65] Nicole Vincent Jordan, Aleix Prat, Amy N Abell, Jon S Zawistowski, Noah Sciaky, Olga A Karginova, Bingying Zhou, Brian T Golitz, Charles M Perou, and Gary L Johnson. SWI/SNF chromatin-remodeling factor Smarcd3/Baf60c controls epithelial-mesenchymal transition by inducing Wnt5a signaling. *Molecular and cellular biology*, 33(15):3011–25, Aug 2013.

[66] I Hedenfalk, D Duggan, Y Chen, M Radmacher, M Bittner, R Simon, P Meltzer, B Gusterson, M Esteller, O P Kallioniemi, B Wilfond, A Borg, J Trent, M Raffeld, Z Yakhini, A Ben-Dor, E Dougherty, J Kononen, L Bubendorf, W Fehrle, S Pittaluga, S Gruvberger, N Loman, O Johannsson, H Olsson, and G Sauter. Gene-expression profiles in hereditary breast cancer. *The New England journal of medicine*, 344(8):539–48, Feb 2001.

[67] Christos Sotiriou, Soek-Ying Neo, Lisa M McShane, Edward L Korn, Philip M Long, Amir Jazaeri, Philippe Martiat, Steve B Fox, Adrian L Harris, and Edison T Liu. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proceedings of the National Academy of Sciences of the United States of America*, 100(18):10393–8, Sep 2003.

[68] Bing Zhang, Jing Wang, XiaojingWang, Jing Zhu, Qi Liu, Zhiao Shi, Matthew C Chambers, Lisa J Zimmerman, Kent F Shaddox, Sangtae Kim, Sherri R Davies, Sean Wang, PeiWang,

Christopher R Kinsinger, Robert C Rivers, Henry Rodriguez, R Reid Townsend, Matthew J C Ellis, Steven a Carr, David L Tabb, Robert J Coffey, Robbert J C Slebos, Daniel C Liebler, and the N C I Cptac. Proteogenomic characterization of human colon and rectal cancer. *Nature*, pages 1–21, 2014.

[69] Ataman Sendoel, Joshua G. Dunn, Edwin H. Rodriguez, Shruti Naik, Nicholas C. Gomez, Brian Hurwitz, John Levorse, Brian D. Dill, Daniel Schramek, Henrik Molina, Jonathan S. Weissman, and Elaine Fuchs. Translation from unconventional 5' start sites drives tumour initiation. *Nature*, 2017.

[70] Raquel de Sousa Abreu, Luiz O. Penalva, Edward M. Marcotte, and Christine Vogel. Global signatures of protein and mRNA expression levels. *Molecular BioSystems*, 5(12):1512–26, Dec 2009.

[71] Mazen W Karaman, Sanna Herrgard, Daniel K Treiber, Paul Gallant, Corey E Atteridge, Brian T Campbell, Katrina W Chan, Pietro Ciceri, Mindy I Davis, Philip T Edeen, Raffaella Faraoni, Mark Floyd, Jeremy P Hunt, Daniel J Lockhart, Zdravko V Milanov, Michael J Morrison, Gabriel Pallares, Hitesh K Patel, Stephanie Pritchard, Lisa M Wodicka, and Patrick P Zarrinkar. A quantitative analysis of kinase inhibitor selectivity. *Nature Biotechnology*, 26(1):127–132, Jan 2008.

[72] Mindy I Davis, Jeremy P Hunt, Sanna Herrgard, Pietro Ciceri, Lisa M Wodicka, Gabriel Pallares, Michael Hocker, Daniel K Treiber, and Patrick P Zarrinkar. Comprehensive analysis of kinase inhibitor selectivity. *Nature Biotechnology*, 29(11):1046–1051, Oct 2011.

[73] A Bhattacharyya. On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematical Society*, 35:99–109, 1943.

[74] Jon S. Zawistowski, Samantha M. Bevill, Daniel R. Goulet, Timothy J. Stuhlmiller, Adriana S. Beltran, Jose F. Olivares-Quintero, Darshan Singh, Noah Sciaky, Joel S. Parker, Naim U. Rashid, Xin Chen, James S. Duncan, Martin C. Whittle, Steven P. Angus, Sara Hanna Velarde, Brian T. Golitz, Xiaping He, Charlene Santos, David B. Darr, Kristalyn Gallagher, Lee M. Graves, Charles M. Perou, Lisa A. Carey, H. Shelton Earp, and Gary L. Johnson. Enhancer Remodeling During Adaptive Bypass to MEK Inhibition Is Attenuated by Pharmacological Targeting of the P-TEFb Complex. *Cancer Discovery*, 2017.

[75] Matthieu Chartier, Thierry Chénard, Jonathan Barker, and Rafael Najmanovich. Kinome Render: a stand-alone and web-accessible tool to annotate the human protein kinome tree. *PeerJ*, 1:e126, Jan 2013.

[76] Kai Wang, Darshan Singh, Zheng Zeng, Stephen J Coleman, Yan Huang, Gleb L Savich, Xiaping He, Piotr Mieczkowski, Sara A Grimm, Charles M Perou, James N MacLeod, Derek Y Chiang, Jan F Prins, and Jinze Liu. MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic acids research*, 38(18):e178, Oct 2010.

[77] Shao Li, Bo Zhang, and Ningbo Zhang. Network target for screening synergistic drug combinations with application to traditional Chinese medicine. *BMC Systems Biology*, 5(Suppl 1):S10, 2011.

[78] Anneleen Daemen, Obi L Griffith, Laura M Heiser, Nicholas J Wang, Oana M Enache, Zachary Sanborn, Francois Pepin, Steffen Durinck, James E Korkola, Malachi Griffith, Joe S Hur, Nam Huh, Jongsuk Chung, Leslie Cope, Mary Fackler, Christopher Umbricht, Saraswati Sukumar, Pankaj Seth, Vikas P Sukhatme, Lakshmi R Jakkula, Yiling Lu, Gordon B Mills, Raymond J Cho, Eric A Collisson, Laura J van't Veer, Paul T Spellman, and Joe W Gray. Modeling precision treatment of breast cancer. *Genome Biology*, 14(10):R110, 2013.

[79] L. M. Heiser, A. Sadanandam, W.-L. Kuo, S. C. Benz, T. C. Goldstein, S. Ng, W. J. Gibb, N. J. Wang, S. Ziyad, F. Tong, N. Bayani, Z. Hu, J. I. Billig, A. Dueregger, S. Lewis, L. Jakkula, J. E. Korkola, S. Durinck, F. Pepin, Y. Guan, E. Purdom, P. Neuvial, H. Bengtsson, K. W. Wood, P. G. Smith, L. T. Vassilev, B. T. Hennessy, J. Greshock, K. E. Bachman, M. A. Hardwicke, J. W. Park, L. J. Marton, D. M. Wolf, E. A. Collisson, R. M. Neve, G. B. Mills, T. P. Speed, H. S. Feiler, R. F. Wooster, D. Haussler, J. M. Stuart, J. W. Gray, and P. T. Spellman. Subtype and pathway specific responses to anticancer compounds in breast cancer. *Proceedings of the National Academy of Sciences*, 109(8):2724–2729, Feb 2012.

[80] Michael D Slack, Elisabeth D Martinez, Lani F Wu, and Steven J Altschuler. Characterizing heterogeneous cellular responses to perturbations. *Proceedings of the National Academy of Sciences of the United States of America*, 105(49):19306–11, Dec 2008.

[81] Martin H. Schaefer, Jean-Fred Fontaine, Arunachalam Vinayagam, Pablo Porras, Erich E. Wanker, and Miguel A. Andrade-Navarro. HIPPIE: Integrating Protein Interaction Networks with Experiment Based Quality Scores. *PLoS ONE*, 7(2):e31826, Feb 2012.

[82] T S Keshava Prasad, Renu Goel, Kumaran Kandasamy, Shivakumar Keerthikumar, Sameer Kumar, Suresh Mathivanan, Deepthi Telikicherla, Rajesh Raju, Beema Shafreen, Abhilash Venugopal, Lavanya Balakrishnan, Arivusudar Marimuthu, Sutopa Banerjee, Devi S Somanathan, Aimy Sebastian, Sandhya Rani, Somak Ray, C J Harrys Kishore, Sashi Kanth, Mukhtar Ahmed, Manoj K Kashyap, Riaz Mohmood, Y L Ramachandra, V Krishna, B Abdul Rahiman, Sujatha Mohan, Prathibha Ranganathan, Subhashri Ramabadran, Raghothama Chaerkady, and Akhilesh Pandey. Human Protein Reference Database–2009 update. *Nucleic acids research*, 37(Database issue):D767–72, Jan 2009.

[83] K. R. Brown and I. Jurisica. Online Predicted Human Interaction Database. *Bioinformatics*, 21(9):2076–2082, May 2005.

[84] Kevin R Brown and Igor Jurisica. Unequal evolutionary conservation of human protein interactions in interologous networks. *Genome Biology*, 8(5):R95, 2007.

[85] Peter V Hornbeck, Bin Zhang, Beth Murray, Jon M Kornhauser, Vaughan Latham, and Elzbieta Skrzypek. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic acids research*, 43(Database issue):D512–20, Jan 2015.

[86] David Croft, Antonio Fabregat Mundo, Robin Haw, Marija Milacic, Joel Weiser, Guanming Wu, Michael Caudy, Phani Garapati, Marc Gillespie, Maulik R Kamdar, Bijay Jassal, Steven Jupe, Lisa Matthews, Bruce May, Stanislav Palatnik, Karen Rothfels, Veronica Shamovsky, Heeyeon Song, Mark Williams, Ewan Birney, Henning Hermjakob, Lincoln Stein, and Peter D'Eustachio. The Reactome pathway knowledgebase. *Nucleic acids research*, 42(Database issue):D472–7, Jan 2014.

[87] Emmy D. G. Fleuren, Luxi Zhang, Jianmin Wu, and Roger J. Daly. The kinome 'at large' in cancer. *Nature Reviews Cancer*, 16(2):83, 2016.

[88] Fumi Shima, Shigeyuki Matsumoto, Yoko Yoshikawa, Takashi Kawamura, Masayuki Isa, and Tohru Kataoka. Current status of the development of Ras inhibitors. *Journal of Biochemistry*, 158(2):91–99, Aug 2015.

[89] Pall F Jonsson, Tamara Cavanna, Daniel Zicha, and Paul A Bates. Cluster analysis of networks generated through homology: automatic identification of important protein communities involved in cancer metastasis. *BMC bioinformatics*, 7:2, 2006.

[90] Gary Bader and Christopher Hogue. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*, 4(1):2, 2003.

[91] L. Licata, L. Briganti, D. Peluso, L. Perfetto, M. Iannuccelli, E. Galeota, F. Sacco, A. Palma, A. P. Nardozza, E. Santonico, L. Castagnoli, and G. Cesareni. MINT, the molecular interaction database: 2012 update. *Nucleic Acids Research*, 40(D1):D857–D861, Jan 2012.

[92] C. Stark, Bobby-Joe Breitkreutz, Teresa Reguly, Lorrie Boucher, Ashton Breitkreutz, and Mike Tyers. BioGRID: a general repository for interaction datasets. *Nucleic Acids Research*, 34(90001):D535–D539, Jan 2006.

[93] L. Salwinski, Christopher S Miller, Adam J Smith, Frank K Pettit, James U Bowie, and David Eisenberg. The Database of Interacting Proteins: 2004 update. *Nucleic Acids Research*, 32(90001):449D–451, Jan 2004.

[94] Max Kotlyar, Chiara Pastrello, Nicholas Sheahan, and Igor Jurisica. Integrated interactions database: tissue-specific view of the human and model organism interactomes. *Nucleic acids research*, 44(D1):D536–41, Jan 2016.

[95] Marija Milacic, Robin Haw, Karen Rothfels, Guanming Wu, David Croft, Henning Hermjakob, Peter D'Eustachio, and Lincoln Stein. Annotating Cancer Variants and Anti-Cancer Therapeutics in Reactome. *Cancers*, 4(4):1180–1211, Nov 2012.

[96] Harvard Medical School. Library of Integrated Network-based Cellular Signatures (LINCS), 2015. Accessed: 2016-05-04.

[97] Gabor Csardi and Tamas Nepusz. The igraph software package for complex network research. *InterJournal*, Complex Systems:1695, 2006.

[98] Xiang-Yun Wu and Xia-Yu Xia. *ProNet: Biological Network Construction, Visualization and Analyses*, 2015. R package version 1.0.0.

[99] Zhao Yang, René Algesheimer, Claudio J. Tessone, A. Sánchez, and Y. Moreno. A Comparative Analysis of Community Detection Algorithms on Artificial Networks. *Scientific Reports*, 6(1):30750, Nov 2016.

[100] Marina Meila. Comparing clusterings: an information based distance. *Journal of Multivariate Analysis*, 98(5):873–895, May 2007.

[101] Aaron F. McDaid, Derek Greene, and Neil Hurley. Normalized Mutual Information to evaluate overlapping community finding algorithms. *arXiv*, Oct 2011.

[102] Qi Zhang, Eric Yi Liu, Abhishek Sarkar, and Wei Wang. Split-Order Distance for Clustering and Classification Hierarchies. In *Scientific and Statistical Database Management*, pages 517–534. Springer, Berlin, Heidelberg, 2009.

[103] William M. Rand. Objective Criteria for the Evaluation of Clustering Methods. *Journal of the American Statistical Association*, 66(336):846–850, Dec 1971.

[104] Huaiyu Mi, Anushya Muruganujan, John T Casagrande, and Paul D Thomas. Large-scale gene function analysis with the PANTHER classification system. *Nature Protocols*, 8(8):1551–1566, Jul 2013.

[105] Elisabeth Buchdunger, Terence O'Reilly, and Jeanette Wood. Pharmacology of imatinib (STI571). *European journal of cancer (Oxford, England : 1990)*, 38 Suppl 5:S28–36, Sep 2002.

[106] Renaud Capdeville, Elisabeth Buchdunger, Juerg Zimmermann, and Alex Matter. Glivec (STI571, imatinib), a rationally developed, targeted anticancer drug. *Nature Reviews Drug Discovery*, 1(7):493–502, Jul 2002.

[107] Lixian Wu, Xianling Chen, Lisen Huang, Jue Tian, Fang Ke, Jianhua Xu, Yuanzhong Chen, and Ming Zheng. A Novobiocin Derivative, XN4, Inhibits the Proliferation of Chronic Myeloid Leukemia Cells by Inducing Oxidative DNA Damage. *PLOS ONE*, 10(4):e0123314, Apr 2015.

[108] Henrik Daub, Katja Specht, and Axel Ullrich. Strategies to overcome resistance to targeted protein kinase inhibitors. *Nature reviews. Drug discovery*, 3(12):1001–10, Dec 2004.

[109] Natalia L Komarova and Dominik Wodarz. Drug resistance in cancer: principles of emergence and prevention. *Proceedings of the National Academy of Sciences of the United States of America*, 102(27):9714–9, Jul 2005.

[110] Caitriona Holohan, Sandra Van Schaeybroeck, Daniel B. Longley, and Patrick G. Johnston. Cancer drug resistance: an evolving paradigm. *Nature Reviews Cancer*, 13(10):714–726, Sep 2013.

[111] Juan Carlos Montero, Samuel Seoane, Alberto Ocaña, and Atanasio Pandiella. Inhibition of SRC family kinases and receptor tyrosine kinases by dasatinib: possible combinations in solid tumors. *Clinical cancer research : an official journal of the American Association for Cancer Research*, 17(17):5546–52, Sep 2011.

[112] Ivana Bozic, Benjamin Allen, and Martin A. Nowak. Dynamics of targeted cancer therapy. *Trends in Molecular Medicine*, 18(6):311–316, Jun 2012.

[113] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015.

[114] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008.

[115] Wolfram Research, Inc. *Mathematica Version 11.0*. MathWorks, Champaign, Illinois, 2016.

[116] DiscoverX. Kinomescan - world's largest kinase assay panel. https://www.discoverx.com/services/drug-discovery-development-services/kinase-profiling/kinomescan, 2016.

[117] H. Ronald Pulliam. Diet optimization with nutrient constraints. *The American Naturalist*, 109(970):765–768, 1975.

[118] Gurobi Optimization Inc. Gurobi Optimizer Reference Manual, 2015.

[119] MATLAB. *version 8.1.0.604 (R2013a)*. The MathWorks Inc., Natick, Massachusetts, 2010.

[120] Manfred Padberg. *Linear Optimization and Extensions*, volume 12 of *Algorithms and Combinatorics*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1999.

[121] A. H. Land and A. G. Doig. An Automatic Method of Solving Discrete Programming Problems. *Econometrica*, 28(3):497, Jul 1960.

[122] Silvano. Martello and Paolo. Toth. *Knapsack problems : algorithms and computer implementations*. J. Wiley & Sons, 1990.

[123] Thomas Bachelot, Jose A Garcia-Saenz, Sunil Verma, Maya Gutierrez, Xavier Pivot, Mark F Kozloff, Catherine Prady, Xin Huang, Reza Khosravan, Zhixiao Wang, Rossano Cesari, Vanessa Tassell, Kenneth A Kern, Jean-Yves Blay, and Ana Lluch. Sunitinib in combination with trastuzumab for the treatment of advanced breast cancer: activity and safety results from a phase II study. *BMC Cancer*, 14(1):166, Dec 2014.

[124] Lily Yen, Xiao-Li You, Ala-Eddin Al Moustafa, Gerald Batist, Nancy E Hynes, Sylvie Mader, Sylvain Meloche, and Moulay A Alaoui-Jamali. Heregulin selectively upregulates vascular endothelial growth factor secretion in cancer cells and stimulates angiogenesis. *Oncogene*, 19(31):3460–3469, Jul 2000.

[125] U.S. National Institutes of Health. Clinicaltrials.gov. clinicaltrials.gov, 2015.

[126] Angelika Berger, Veronika Sexl, Peter Valent, and Richard Moriggl. Inhibition of STAT5: a therapeutic option in BCR-ABL1-driven leukemia. *Oncotarget*, 5(20):9564–76, Oct 2014.

[127] U.S. Food & Drug Administration Center for Drug Evaluation and Research. Drug Approvals and Databases - Approved Drug Products with Therapeutic Equivalence Evaluations (Orange Book). *FDA Orange Book*, 2016.

[128] Prithviraj Bose, Edward B. Perkins, Connie Honeycut, Martha D. Wellons, Tammy Stefan, James W. Jacobberger, Emmanouil Kontopodis, Jan H. Beumer, Merrill J. Egorin, Chiyo K. Imamura, W. Douglas Figg, Judith E. Karp, Omer N. Koc, Brenda W. Cooper, Selina M. Luger, A. Dimitrios Colevas, John D. Roberts, and Steven Grant. Phase I trial of the combination of flavopiridol and imatinib mesylate in patients with Bcr-Abl+ hematological malignancies. *Cancer Chemotherapy and Pharmacology*, 69(6):1657–1667, Jun 2012.

[129] Aristeidis Chaidos, Valentina Caputo, and Anastasios Karadimitris. Inhibition of bromodomain and extra-terminal proteins (BET) as a potential therapeutic approach in haematological malignancies: emerging preclinical and clinical evidence. *Therapeutic advances in hematology*, 6(3):128–41, Jun 2015.

[130] T Tran, E Ong, and AP Hodges. Prediction of kinase inhibitor response using activity profiling, in-vitro screening, and elastic net regression. *arXiv*, pages 1–20, 2013.

[131] Joshua D Kangas, Armaghan W Naik, and Robert F Murphy. Efficient discovery of responses of proteins to compounds using active learning. *BMC bioinformatics*, 15(1):143, Jan 2014.

[132] Kuan-lin Huang, Shunqiang Li, Philipp Mertins, Song Cao, Harsha P. Gunawardena, Kelly V. Ruggles, D. R. Mani, Karl R. Clauser, Maki Tanioka, Jerry Usary, Shyam M. Kavuri, Ling Xie, Christopher Yoon, Jana W Qiao, John Wrobel, Matthew A. Wyczalkowski, Petra Erdmann-Gilmore, Jacqueline E. Snider, Jeremy Hoog, Purba Singh, Beifung Niu, Zhanfang Guo, Sam Qiancheng Sun, Souzan Sanati, Emily Kawaler, Xuya Wang, Adam Scott, Kai Ye, Michael D. McLellan, Michael C. Wendl, Anna Malovannaya, Jason M. Held, Michael A. Gillette, David Fenyö, Christopher R. Kinsinger, Mehdi Mesri, Henry Rodriguez, Sherri R. Davies, Charles M. Perou, Cynthia Ma, R. Reid Townsend, Xian Chen, Steven A. Carr, Matthew J. Ellis, and Li Ding. Proteogenomic integration reveals therapeutic targets in breast cancer xenografts. *Nature Communications*, 8:14864, Mar 2017.