2022

Identifying Network Biomarkers for Each Breast Cancer Subtypes Along with Their Effective Single and Paired Repurposed Drugs Using Network-Based Machine Learning Techniques

Forough Firoozbakht
University of Windsor

IDENTIFYING NETWORK BIOMARKERS FOR EACH BREAST CANCER SUBTYPES ALONG WITH THEIR EFFECTIVE SINGLE AND PAIRED REPURPOSED DRUGS USING NETWORK-BASED MACHINE LEARNING TECHNIQUES

by
Forough Firoozbakht

A Dissertation
Submitted to the Faculty of Graduate Studies
through the School of Computer Science
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy at the
University of Windsor

Windsor, Ontario, Canada
2022

IDENTIFYING NETWORK BIOMARKERS FOR EACH BREAST CANCER SUBTYPES ALONG WITH THEIR EFFECTIVE SINGLE AND PAIRED REPURPOSED DRUGS USING NETWORK-BASED MACHINE LEARNING TECHNIQUES

by

FOROUGH FIROOZBAKHT

APPROVED BY:

A. B. Diallo, External Examiner
University of Quebec in Montreal

E. Abdel-Raheem
Department of Electrical and Computer Engineering

J. Lu
School of Computer Science

D. Wu
School of Computer Science

A. Ngom, Co-Advisor
School of Computer Science

L. Rueda, Co-Advisor
School of Computer Science

September 6, 2022

Declaration of Co-Authorship and Previous Publication

I. Co-Authorship

I hereby declare that this Dissertation incorporates the outcome of a joint research undertaken in collaboration with Michele D'agnillo, Dr. Iman Rezaeian, Dr. Alioune Ngom, Dr. Luis Rueda and Dr. Lisa Porter.

- Chapter 2 of the thesis includes the outcome of publication which have the following other co-authors: I. Rezaeian, M. D'agnillo, L. Porter, L. Rueda, and A. Ngom. Only my primary contributions towards this publication are included in this thesis, and the contribution of co-authors was as follows: M. D'agnillo: some wet-lab analysis and biological validations. I. Rezaeian: helping with design of the pipeline, L.Porter, L. Rueda and A. Ngom: mentoring and consultation on the research.

- Chapter 3 of the thesis includes the outcome of publication which have the following other co-authors: I. Rezaeian, L. Rueda, & A. Ngom. Only my primary contributions towards this publication are included in this thesis, and the contribution of co-authors was as follows: I. Rezaeian: helping with design of the pipeline, L. Rueda and A. Ngom: mentoring and consultation on the research.

The collaboration is covered in Chapters 2 and 3 of the Dissertation. In this research, experimental designs, applying and optimizing different machine learning methods for prediction, numerical and visual analysis were performed by the author.

I am aware of the University of Windsor Senate Policy on Authorship and I certify that I have properly acknowledged the contribution of other researchers to my Dissertation, and have obtained written permission from each of the co-author(s) to include the above material(s) in my Dissertation. I certify that, with the above qualification, this Dissertation, and the research to which it refers, is the product of my own work.

II. Previous Publications

This Dissertation includes two original papers that have been previously published/submitted for publication in conferences and peer reviewed journals, as follows:

Dissertation chapter	Publication title
Chapter 2	Firoozbakht, F., Rezaeian, I., D'agnillo, M., Porter, L., Rueda, L., & Ngom, A. (2017). An integrative approach for identifying network biomarkers of breast cancer subtypes using genomic, interactomic, and transcriptomic data. Journal of Computational Biology, 24(8), 756-766.
Chapter 3	Firoozbakht, F., Rezaeian, I., Rueda, L., & Ngom, A. (2022). Computationally repurposing drugs for breast cancer subtypes using a network-based approach. BMC bioinformatics, 23(1), 1-36.

I certify that I have obtained a written permission from the copyright owner(s) to include the above published material(s) in my Dissertation. I certify that the above material describes work completed during my registration as graduate student at the University of

Windsor.

I declare that, to the best of my knowledge, my Dissertation does not infringe upon anyone copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my Dissertation, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my Dissertation. I declare that this is a true copy of my Dissertation, including any final revisions, as approved by my Dissertation committee and the Graduate Studies office, and that this Dissertation has not been submitted for a higher degree to any other University or Institution.

Abstract

Breast cancer is a complex disease that can be classified into at least 10 different molecular subtypes. Appropriate diagnosis of specific subtypes is critical for ensuring the best possible patient treatment and response to therapy. Current computational methods for determining the subtypes are based on identifying differentially expressed genes (i.e., biomarkers) that can best discriminate the subtypes. Such approaches, however, are known to be unreliable since they yield different biomarker sets when applied to data sets from different studies. Gathering knowledge about the functional relationship among genes will identify "network biomarkers" that will enrich the criteria for biomarker selection. Cancer network biomarkers are subnetworks of functionally related genes that "work in concert" to perform functions associated with a tumorigenic. We propose a machine learning framework that can be used to identify network biomarkers and driver genes for each specific breast cancer subtype. Our results show that the resulting network biomarkers can separate one subtype from the others with very high accuracy. We also propose an integrated approach that can best capture knowledge (and complex relationships) contained within and between drugs,genes and disease data. A network-based machine learning approach is applied thereafter by using the extracted knowledge and relationships in order to identify single and pair of approved or experimental drugs with potential therapeutic effects on different breast cancer subtypes.

Dedication

To my beloved husband and son

Acknowledgements

I would like to take this opportunity to express my sincere gratitude to Dr. Luis Rueda and Dr. Alioune Ngom, my supervisor and my co-supervisor, for their steady encouragement, patient guidance and enlightening discussions throughout my graduate studies. Without their help, the work presented here could not have been possible.

I also would like to express my gratitude and appreciation to Dr. Lisa porter for her constructive suggestions and feedbacks.

I also want to express my gratitude to my colleague and husband, Dr. Iman Rezaeian, who shared many works with me and gave me many useful suggestions during my research.

I also wish to express my appreciation to Dr. Dr. Abdoulaye Banire Diallo, Department of Computer Science, University of Quebec in Montreal, Dr. Esam Abdel-Raheem, Department of Electrical and Computer Engineering, Dr. Jianguo Lu and Dr. Dan Wu, School of Computer Science for being in the committee and spending their valuable time.

Finally, I would like to thank all of my colleagues in pattern recognition and bioinformatics lab for their consistent support and help.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Over the years, researchers have gathered large quantities of data to better understand the molecular dynamics behind complicated diseases such as cancer. However, the complexity of these data makes it difficult for researchers to carry out in-depth analysis and extract the pertinent information. Here is where the need for high-performance computers, combined with appropriate algorithms, models and programs, can become extremely helpful. In the past few years, with the help of technological advancements in computer hardware and software, computer scientists made a tremendous progress in developing machine learning algorithms that are able to quickly analyze huge amounts of data and provide models that are useful for identifying relevant pieces of information and making predictions about status and progress of various diseases. As these models become increasingly advanced and sophisticated, they are opening up a new world of possibilities for big data analysis in many domains including healthcare, by transforming the way these conducted studies can provide important discoveries much quicker.

1.1 Cancer and Its Origin

Cancer is one of the leading causes of death worldwide, and a perfect example of where Artificial Intelligence (AI) and machine learning can be a huge help. Cancer is the result of an uncontrolled division of abnormal cells in a particular part of the body that can invade and kill normal tissue and organs around it [37]. Despite tremendous efforts, scientists have not yet been able to identify a systematic treatment for many types of cancer. One of the main reasons for that is an extremely complex nature of this disease, with many subtypes, each often can have a different diagnosis and treatment procedure. Cancer is not limited to a specific geographical location either. Figure 1.1 shows the share of population with any types of cancer. As seen in the figure, United states and Canada are on the lead and we do not see in decline in the disease across various geographical locations.

Cancer is not a new disease either. Conducted studies on fossilized bones and mummified tissues have shown that malignant transformations have been targeting humans and animals for a long time [69].

Although cancer is a complex disease, it has been suggested that only 10% of cancer cases might originates from inherited mutations [16] and the majority of cases has been suspected to come from either high penetrance genes or polymorphisms [22, 214]. For example, specific inherited mutations in BRCA1 and BRCA2 genes account for only 5 to 10 percent of all breast cancer cases [41]. While the majority of human cancer cases has been related to age [56] and environmental factors [16, 69].

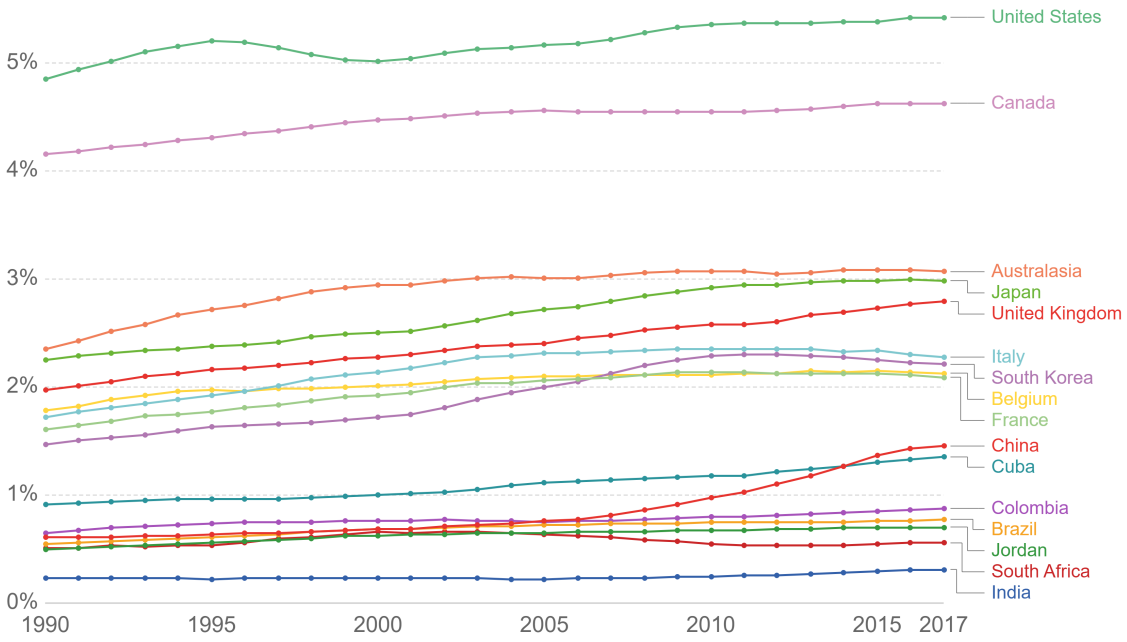


Figure 1.1: Share of total population with any form of cancer, measured as the age-standardized percentage. This share has been age-standardized assuming a constant age structure to compare prevalence between countries and through time (source: https://ourworldindata.org/cancer)

1.2 Breast Cancer

Breast cancer is the leading cause of death among women in most developed countries including Canada [36]. Figure 1.2 shows the share of Canadian population with different types of cancer in 2017.

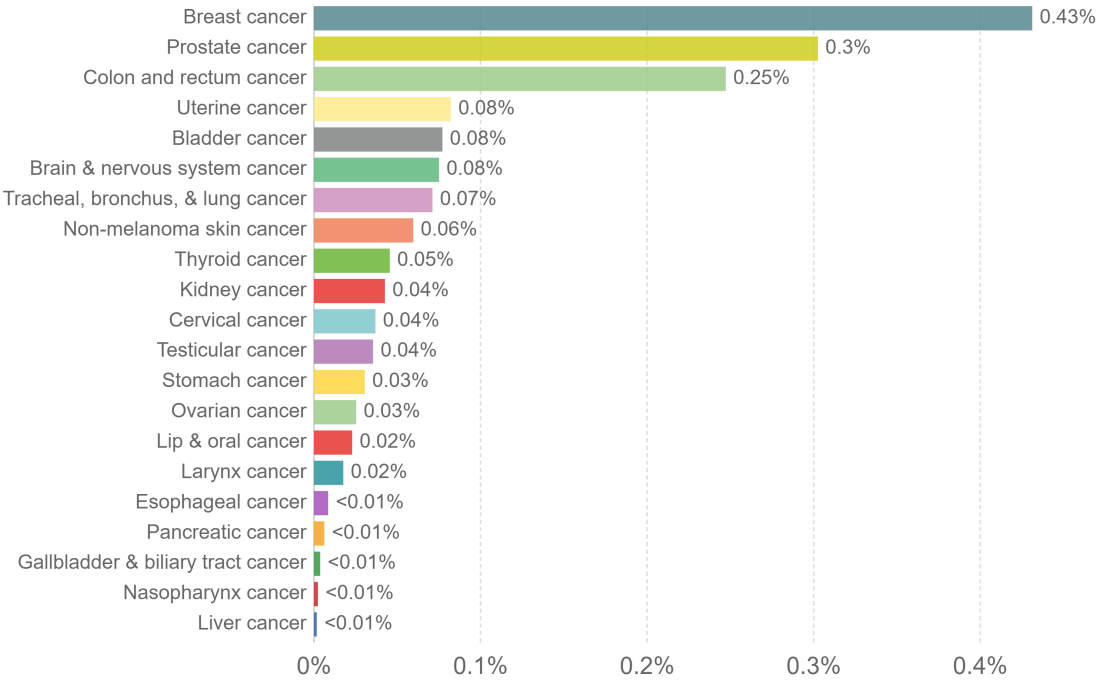


Figure 1.2: Share of total population with different types of cancer in 2017, measured as the age-standardized percentage. (source: https://ourworldindata.org/cancer)

Breast cancer is not a single disease, but a heterogeneous disease comprising different entities with distinct pathological and clinical properties [59, 89, 190, 199]. In next sections, we take a look at development of this disease as well as various existing categorizations.

1.3 Breast Cancer Development

In regard to disease development, breast cancer can be categorized into several stages, spanning from stage zero to stage four [30]. Except stage zero, each stage consists of the Roman numbers I, II, III, or IV often followed by A, B, or C. In general, the higher the number, the more advanced the breast cancer. In stage zero, there is no evidence of cancer cells. Also, if there exists non-cancerous but abnormal cells, they have remained within the same part of the breast that they have originally initiated and there is no sign of getting through to or invading neighboring normal tissues. In stage I, cancer cells are breaking through to or invading normal surrounding breast tissues. The size of the cancerous region at this stage is no more than 2 centimeters normally. in Stage IA, the cancer has spread into the fatty breast tissue. The tumor itself is no larger than 2 centimeters, or there may be no tumor observed in fatty breast tissue. In stage IB a tiny amounts of cancer cells have been found in a few lymph nodes. In stage II, the cancer might have been grown, spread, or both. In stage IIA, the tumor in the breast is still small and the cancer has not spread to more than three lymph nodes. In stage IIB, breast tumor is grown bigger (up to 5 centimeters). In stage III, it's considered advanced, and it's harder to fight. In stage IIIA, the cancer has been found in up to nine of the lymph nodes that usually form a chain or it has spread to or enlarged the lymph nodes deep in the breast. In stage IIIB, the tumor has grown into the chest wall or skin around the breast. In stage IIIC, cancer has been found in ten or more lymph nodes, or has spread above or below the collarbone. In Stage IV, breast cancer cells have spread far away from the breast and lymph nodes around it. This stage is described as "metastatic," meaning it has spread beyond the region of the body where it was first found into the bones, lungs, liver, brain or other organs. The 5-year survival rate (survival rate of patients 5 years after their first diagnosis) of patients can decrease dramatically from 100%

in stage I to as low as 23% in Stage IV [36]. Thus, early detection of the cancer is vital for proper treatment and increasing the survival rate of the patients.

1.4 Breast Cancer Subtypes

It has been shown that breast cancer tumours consist of various pathological and biological features and exhibit distinctive behaviors that eventually lead to different responses during the treatment and hence should not be treated with the same therapeutic strategy [31]. So, accurate clustering of breast cancer tumours into clinically relevant subtypes is vital for further therapeutic decisions and effective treatment of the disease.

1.4.1 Classical Breast Cancer Subtypes

Sørlie et al. grouped more than 500 breast cancer samples into five intrinsic subtypes with distinct clinical outcomes: luminal A, luminal B, HER2 , Basal and normal-like tumors [157, 186]. He reported a distinctive gene signature for those breast cancer subtypes. the idea was that the differences underlying the gene expression patterns among cancer subtypes can effectively reflect the differences of the tumors at the molecular level [187]. These subtypes have been repeated by other studies with different numbers of signature genes for each subtype. Onitilo et al. classified breast cancer into four groups based on Immunohistochemistry (IHC) profile on Estrogen receptors (ER), Progesterone receptors (PR) and Human epidermal growth factor receptor-2 (HER2) expressions, into four groups and claimed that his IHC-based classification correlated well with intrinsic gene expression microarray categorization [149]. Hu et al. reported a 306 gene signature with the ability of distinguishing these subtypes [88]. Parker et al. reported a 50-gene classifier (PAM50),

with significant prognostic performance on breast tumors [64, 79, 155] with the ability of being used in clinical trials [8]. Rezaeian et al. proposed a hierarchical classification model consisting of only 18 genes that was able to classify these 5 breast cancer subtypes with more than 95% accuracy [170]. In this section, we discuss about each of these subtypes in more details.

Luminal Subtypes

The luminal-like tumors express hormone receptors, with expression profiles reminiscent of the luminal epithelial component of the breast [157]. At least two subtypes exist within luminal-like tumors: luminal A and luminal B. In short, luminal A represents the [ER+—PR+,HER2-] group (tumors with ER or PR positivity and HER2 negativity) and luminal B represents the [ER+—PR+, HER2+] group (tumors with ER or PR positivity and HER2 positivity) [206]. Luminal tumors are the most common subtypes among breast cancer, with luminal A being the majority. In general, the luminal subtypes carry a good prognosis, and luminal A tumors have a significantly better prognosis than the luminal B subtype [187].

HER-2 Subtype

The intrinsic HER2 over-expression tumors refer to those identified using gene expression array, which is similar to the [ER-,PR-,HER2+] subgroup by IHC [149] or fluorescence in situ hybridization (FISH) [206]. There has not been any known relationship between race, age or known risk factors and HER2 up-regulation [38,51,147] and this subtype has usually a poor prognosis [186–188], which seems to be related to higher risk of early relapse among those patients without complete removal of tumor cells [38].

Basal Subtype

Basal subtype accounts for 60% to 90% of triple negative (ER-,PR-,HER2-]) breast cancer tumors [71, 198]. There has been a lot of research in recent years regarding Basal subtype, because tumors belonging to this subtype tend to follow aggressive clinical course and currently there is no standard and systematic therapy for this subtype. Compared with the other subtypes, Basal subtype is associated with younger patients and is also more common to develop in African-American women [42]. The size of tumors in this subtype is normally larger than the other subtypes and tend to grow more rapidly [86, 165]. Given the nature of triple negative receptor status, basal tumors are not responsive to conventional targeted breast cancer therapies such as hormone therapy, chemotherapy is the only main option for patients with this breast cancer subtype [29].

1.4.2 Multigenomic Breast cancer subtypes

Traditional classification of breast cancer relies solely on gene expression (GE) as the main driver behind distinction of breast cancer subtypes. However, to be able to take into account all possible molecular drivers for each subtype, not only the gene expression has to be taken into account, but also other genomic information such as copy number abberation (CNA), copy number variation (CNV) and single nucleotide polymorphisms (SNP) have to be considered as well. Curtis et al. [53] proposed an integrated analysis of copy number and gene expression data conducted on 2,000 breast cancer patients with long-term clinical followup and shown that genomic variants such as CNA, CNV and SNP were associated with expression of close to 40% of the genes, with CNAs having the dominant role in that effect. By incorporating the information corresponding to both CNA and GE, they discovered 10 distinctive breast cancer sub-groups. In a followup study, Ali et al. validated

their previous 10 subtypes (known as IntClust subtypes) and shown that the results are reproducible in a larger analysis consisting of 7500 patients [15].

1.5 Breast Cancer Bioinformatics

Breast cancer (BC) is a complex disease consisting of five subtypes [185] to ten subtypes [53], each arising from a distinct molecular mechanism and having a distinct clinical progression, and occurring in sites that can be distinguished based in part on characteristic gene expression signatures. Recent studies have revealed extensive diversity both between and within BC tumors, and that most tumors present unique characteristics. This heterogeneity poses significant challenges to BC diagnosis and treatment, with many BC patients undergoing over-treatment [65]. The development of BC is caused by multiple somatic mutations of a small number of genes, called driver genes (or drivers), whose mutation changes deregulate many biological processes or cellular pathways, and therefore leading to initiation and progression of BC as well as resistance to treatment [82, 146]. The passenger genes (or passengers) are those genes whose deregulations or expression changes are the by-products of the drivers. Thus, the drivers are the genes which provoke the disease via somatic mutations. Together, the drivers and their passengers are called gene biomarkers and an important task is to discover them. A more important task, however, is to find the drivers in order to understand/characterize the disease and develop better therapies [20]. Finding the drivers is, therefore, a challenging problem due to the heterogeneity of BC tumors [92]. A biomarker is a biomolecule (e.g. gene, RNA, protein, metabolite) found in body fluids or tissues that is a sign of a normal or abnormal process, or of a condition or disease [40]. Different types of biomarkers are identified in cancer research and used in cancer medicine to make different predictions: 1) risk biomarker: to predict predisposition to cancer [205];

2) diagnostic biomarker: to predict cancer subtype [182]; 3) prognostic biomarker: to predict cancer outcome [209]; 4) predictive biomarker: to predict response to therapy [163]; 5) treatment biomarker: to predict an effective therapy [34]; 6) progression biomarker: to predict cancer stage [28]; 7) monitoring biomarker: to predict if therapy is working [212]; and 8) recurrence biomarker: to predict if cancer will return [232]. BC is the most frequently diagnosed type of cancer and one of the leading causes of cancer death in women [183]. The correct diagnosis of a patient's BC subtype is critical for ensuring the best possible therapy and care. Methods such as MRI, mammography, or CT scan examine phenotypical mammary change but provide little effective information to guide therapy. Other risk factors such as tumor size and lymph node status are also insufficient to accurately predict tumor classes [12].

Since the ten breast cancer subtypes are driven from much larger sample size and also have been shown promising with distinct clinical outcomes in recent years [13, 138], our downstream analysis in this thesis is based on ten subtypes. In Chapter 2, we use different machine learning techniques in order to leverage the knowledge about the functional relationship among genes and identify network biomarkers and driver genes for each breast cancer subtype. In Chapter 3, we will leverage pathway information and also the network biomarkers identified in Chapter 2 in order to find the best combination of re-positioned drugs for each breast cancer subtype.

1.6 Introduction to network-based Machine Learning Techniques

The field of network research has seen an increase in interest over the past decade, with emphasis shifting away from analysis of small graphs to consideration of large-scale graphs, as well as complex networks. Many branches of science have adopted such networks as a way of representing complex systems. These models are typically used to represent systems that include complex topologies and are very voluminous. Complex networks areas have emerged as unifying topics in complex systems due to the technological advances as well as the amount of data being collected and analyzed [68]. An analysis of random networks resulting from that investigation led to the development of a new area of study termed the theory of random networks, which involves combining graph theory and probability theory to generate and analyze large-scale graphs. Complex network areas have emerged as unifying topics in complex systems due to technological advances as well as the amount of data being collected and analyzed [24].

There is ample evidence that complex networks exist in real-world settings. Here are some examples of real-world network representations: biological neural networks [191, 213], financial networks [44,161,189], information networks [224], social networks among individuals [97, 177] and between companies and organizations [135], food webs [136], metabolic networks [54] and distribution as the bloodstream [215], protein-protein interaction networks [208], postal delivery and electricity distribution networks [14], to name a few.

In general, we can categorize network-based machine learning techniques into three groups: Unsupervised, Supervised, and semi-supervised. In the next few sections we dis-

cuss about each of these groups.

1.6.1 Network-based Unsupervised Learning

Clustering data is one of the main tasks of unsupervised learning. Network-based methods are well suited for data clustering tasks, as we do not know how the clusters will be formed or how many will exist in unsupervised learning methods. Data clustering is essentially a community detection problem once the original dataset is constructed. During transformation, each vertex represents a data item, and connections are established according to similarity measures of the vertex. Clusters are often referred to as communities when performing a community detection task. A community is described as a sub-graph whose vertices are highly connected internally, but relatively sparsely connected with others. Network-based methods are particularly useful when dealing with clusters that have varying shapes, spatial proximity, orientation, and density [100]. For a better understanding of various phenomena in complex networks, community detection can be very useful [83]. Complex networks are characterized by a modular structure. Some modules may have many connections, whereas others may be sparse [144]. Global statistics may be misleading when there is a lot of variation between communities. The modular structure of a network may also influence the way in which dynamical processes are conducted (e.g., spreading processes and synchronizations [17]) on the network. Functional modules in biological networks are communities whose members perform essential cellular tasks in concert to form coherent units. For example, modules are common in metabolic networks [167] and protein phosphorylation networks [98].

In biological networks, the identification of functional modules may be a promising computational paradigm for discovering functions of genes and proteins. In the case of

proteins with unknown functions, the modules to which they belong can be used to classify them [153]. Modules are sets of genes or proteins that work together to perform biological processes. Biotechnology and drug design have benefited a great deal from the identification of functional modules. The removal of a whole functional module can be used to delete a certain function in many cases. Modules in complex networks can be detected in a number of different ways [74]. some of the popular approaches consider communities as groups of adjacent motifs [153], while others are influenced by information theory [171], message passing [76], or Bayesian principle [32]. Modularity, a quantity that is typically optimized, is a widely used class of algorithms [145].

It has been a long and hard struggle to develop accurate and efficient solutions to the NP-complete problem of community detection. Some of these solutions include the spectral method [210], the betweenness-based technique [143], modularity greedy optimization [142], detection of communities based on the Potts model [169], synchronization [17], information theory [75], and random walks [85]. There is a comprehensive discussion of this topic in [74].

1.6.2 Network-based Supervised Learning

Although network-based unsupervised and semi-supervised learning techniques have been studied more extensively in the literature [48,91], network-based supervised learning methods have not been studied that much and there is still a big space for discovering new ways to utilize network models for supervised learning.

Classification is one of the most used tasks in supervised learning domain. Network-based classification techniques would be preferable than regular classification methods in cases such as relational classification, where the class label of a sample might not de-

pend solely on its own attributes, but also on the labels of its neighbor samples [128]. Among the many problems that can be solved by relational classification techniques are the finding of molecular pathways in gene expression [178], link prediction in social networks [11, 25, 118], and classification of linked scientific publications [111]. A similar approach is suitable for various other applications, including recommendation systems, the identification of probable associations in e-commerce sites and scientific collaboration networks, as well as the analysis of criminal networks and the structural analysis of bacteria or other biological organisms. Consequently, all of these applications require more efficient and versatile methods for link prediction, making it an important and scientifically attractive research topic. Similarly, relational classification can be used to identify small connected subgraphs in a social network that best reflect the relationship between two vertices. Researchers in [70] have proposed an efficient algorithm based on electrical circuit laws for identifying the connected subgraphs from large social networks. Additionally, it has been demonstrated that a connected subgraph can be used to effectively compute several topological feature values for the supervised link prediction problem when the network is large [11].

In one sense, we can categorize existing classification algorithms into two groups. Local classifiers are those that employ collective inference only at specific stages of the learning process. One may employ a local classifier, such as Naïve Bayes or relational probability trees, to predict labels for each unlabeled vertex and further use a collective inference algorithm, such as ICA [125] or Gibbs sampling [96], to restate the class labels of vertices that are used in the next iteration. Secondly, there are so-called global formulation-based methods, which do not use a separate local classifier, but rather use the entire algorithm for training and inference. Using this method, training aims to optimize a global objec-

tive function. The loopy belief propagation algorithm and relaxation labeling are examples of these algorithms [179]. A supervised learning network-based framework for relational data classification in networks was proposed in [129] as a solution to the unification problem. Three components are considered in the model: a local classifier that uses the training set to determine the probability distribution for the classes; a relational classifier that does the same but now considers the nearby relations in the network; and a collective inference component that refines the prediction further.

Chapter 2

An Integrative Approach for Identifying Network Biomarkers of Breast Cancer Subtypes Using Genomic, Interactomic, and Transcriptomic Data

2.1 Introduction

Most bioinformatics methods have focused on identifying BC biomarkers as small subsets of differentially expressed genes. However, differentially expressed genes have limited predictive performance due to (i) the heterogeneity within tissues and across patients and (ii) the dependence among genes, gene products, or pathways. To accurately identify effective BC biomarkers, new bioinformatics methods integrating additional biological information with gene expression data has become necessary. Within the last five years, new

classes of biomarkers called *cancer network biomarkers* (NBs) have been defined and studied [121, 122, 228]. A cancer network biomarker (NB) is a disease-related sub-network of interacting genes identified by an appropriate integration of a secondary network (e.g. protein interaction network or cellular pathway network) data with the primary gene expression data, thus taking into account the dependencies among genes.

In this chapter, we propose a framework that can be used to identify differential NBs specific to each breast cancer subtype. First, we select and combine relevant features using CNV, CNA and GE data, in order to obtain a set of candidate genes for each breast cancer subtype consisting of (i) genes that are differentially expressed in the subtype and (ii) genes that have significant copy numbers in the subtype. Then, each gene in the candidate set is used to seed the search for discriminative NBs in an input protein-protein interaction (PPI) network.

2.2 Materials and Methods

We have used the METABRIC dataset [53], which contains the copy number values and gene expression levels of 2000 primary breast tumors with long-term clinical follow-up. It can be accessed from the European Genome-Phenome Archive using the accession number EGAS00000000083. In [53], the copy number aberrations and copy number variations generated using Affymetrix SNP 6.0 arrays and gene expression data were obtained using Illumina HT 12 technology. The dataset contains two sets of data, *validation* set and *discovery* set. Due to the lack of class labels in the validation set, in this paper. we only use the discovery set, which contains 997 samples from ten subtypes of breast cancer. Each sample contains expression data for 48,803 probe IDs. The expression of all probes corresponding to the same gene have been merged based on the median expression of those probes, which

maps all the probes to 24,351 unigenes. The number of samples corresponding to each subtype are listed in Table 2.1.

Table 2.1: Number of samples corresponding to each of ten subtypes.

Subtypes	1	2	3	4	5	6	7	8	9	10
# of Samples	76	45	156	167	94	44	109	143	67	96

To obtain a NB corresponding to each subtype, we consider each subtype as positive class and the remaining subtypes as negative class. Thus, by performing a one-against-all classification scheme, separately for each subtype, we can obtain the specific NB that best discriminates that subtype from the other subtypes. Figure 2.1 illustrates the proposed framework for finding NBs corresponding to each subtype.

2.2.1 Obtaining Candidate Genes

In the first step, we use CNA, CNV and GE data to find the most informative genes, separately for each subtype, which are used later as seeds to find the best separating NBs of a given subtype. To do so, we first use CNA/CNV information to find those genes that have very high genotypic aberration in each subtype based on their GISTIC score [27]. GISTIC identifies significant aberrations using two steps. In the first step, it calculates the G-score statistic, which involves both the frequency of occurrence and the amplitude of the aberration. In the second step, it assesses the significance of each aberration using Fisher's Exact test [168]. To make sure that we only target aberrations in the copy number and not common variations across different populations, we use the HapMap database [52]. HapMap is a catalog of common genetic variants that occur in human. We only consider those genes for a significant test that have CNA but no CNV. We also use gene expression data to identify the top differentially expressed genes for each subtype. For this, we used Chi2 [120] to rank

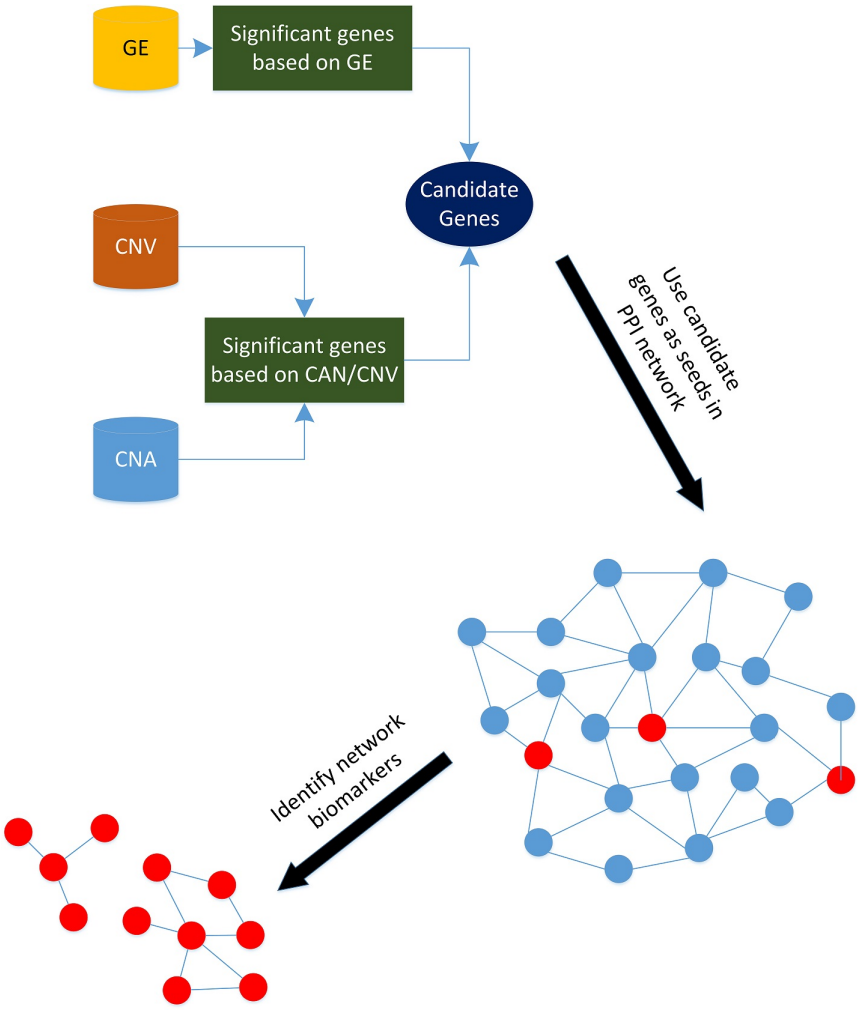


Figure 2.1: The proposed framework for finding NBs corresponding to each subtype.

genes based on their ability to separate each subtype from the remaining subtypes. At the end, after obtaining the top genes using CNA/CNV and GE data separately, if CNA/CNV analysis determined N genes as significant in terms of their genomic aberrations, we select the top N genes from GE data; then out of these two gene sets, we take the intersection as candidate genes, which will be used as seeds in our PPI network data.

2.2.2 Obtaining NB for Each Subtype

In this step, we use the candidate genes obtained from the previous step as seeds in the PPI network data. First, we combined the human PPI network data obtained from BioGrid [192], HPRD [162], Intact [103], DIP [221] and MINT [46] into a single unified large PPI network consisting of 230,000 protein-protein interactions and 15,823 proteins as the union of all aforementioned databases. We only included those PPIs that are verified in two or more of aforementioned databases.

Second, we mapped all candidate genes onto our PPI network in order to be used as seeds for finding the NBs Starting from a given seed node v, the search for the best separating NB proceeds as follows. We iteratively aggregate its neighboring nodes u in a greedy manner, using breath-first search algorithm. A neighbor u is inserted into the current aggregate N if and only if its inclusion (i.e., the new aggregate $N+u$) increases the correlation between the expression of the genes in the aggregate and the given subtype; that is, when $|correlation(N+u, subtype) - correlation(N, subtype)| > \Delta$, where Δ is 0.001. Then, the same process is repeated on the new aggregate $N+u$. This process continues until all possible neighbors (with any distance) from the new aggregate are evaluated, resulting in a subnetwork, S_v, obtained from seed v. The same process is also applied to all the seeds obtained for a given subtype, and the union of all the subnetworks is considered as the final NB of that given subtype.

Since the order of candidate genes may alter the expansion of subnetworks, depending on which candidate gene reaches a certain gene first, we shuffle the candidate genes 100 times and obtain the network for each case individually. At the end, we merge all 100 networks. In this case, each individual interaction have a confidence score from 1 to 100, which represents the number of times each interaction appeared in all 100 networks. We

categorize interactions in three groups; low, medium and high confidence, which contain those interactions that present in less that 30%, between 30% and 70%, and more than 70% of the networks, respectively. Table 2.2 shows the distribution of the interactions in each subtype NB.

2.2.3 Evaluating the Predictive Performance of each NB

The following measures are used for evaluating the predictive performance of each NB.

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN}, \tag{2.1}$$

F-measure uses both precision and recall measures to compute the score as follows:

$$F\text{-}measure = 2 \times \frac{Percision \times Recall}{Precision + Recall}, \tag{2.2}$$

where

$$Precision = \frac{TP}{TP+FP}, \tag{2.3}$$

$$Recall = \frac{TP}{TP+FN}, \tag{2.4}$$

Another measure, the area under the receiving operating characteristics (ROC) curve, AUC, shows the trade-off between Specificity and Sensitivity (Recall), where:

$$Sensitivity\ (Recall) = \frac{TP}{TP+FN}, \tag{2.5}$$

$$Specifity = \frac{TN}{TN+FP}, \tag{2.6}$$

Above, TP, TN, FP, FN means true positive, true negative, false positive, and false neg-

Table 2.2: Number of interactions in NBs corresponding to each subtype. Interactions have been categorized in three groups: low, medium and high confidence, which contain interactions that are present in less that 30%, between 30% and 70%, and more than 70% of the networks, respectively.

Subtype	Total # of Interactions	Low Confidence	Medium Confidence	High Confidence
1	2,389	2,230	126	33
2	3,013	2,890	91	32
3	2,524	2,260	177	87
4	1,444	1,170	184	90
5	1,999	1,866	104	29
6	2,900	2,608	211	81
7	2,294	2,102	118	74
8	2,750	2,585	106	59
9	3,000	2,787	161	52
10	936	814	94	28

ative, respectively.

2.3 Results

Table 2.3 shows the number of selected genes and interactions in the obtained NB corresponding to each of the ten breast cancer subtypes. Since the classes are highly imbalanced, using a more robust performance measure such as AUC provides less bias insight regarding the performance of the NBs for each subtype. As shown in the table, the AUC of the NBs for almost all of the subtypes are more than 0.95, which indicates the excellent predictive performance of each NB.

We trained a random forest classifier containing 50 trees along with 10-fold cross-validation scheme to evaluate the effectiveness of candidate and high confidence genes involved in each subtype's NB in discriminating each subtype individually. Tables 2.4, and 2.5 show the performance of candidate genes and high confidence genes in each subtype,

Table 2.3: Comparison between the number of genes, interactions and the performance of NBs for ten breast cancer subtypes.

Subtype	# of genes	# of interactions	Phenotype correlation	Accuracy	F-measure	AUC
1	2385	2120	-0.947	94.1%	0.928	0.970
2	2948	2432	-0.913	96.6%	0.959	0.966
3	3309	3089	0.916	93.7	0.941	0.939
4	4557	3846	0.929	95.6	0.922	0.952
5	1541	1382	-0.96	97.1	0.964	0.993
6	5382	3987	-0.902	95.7	0.939	0.961
7	3111	2879	-0.947	94.8	0.934	0.952
8	4343	3622	-0.923	93.84	0.943	0.971
9	2266	2151	-0.949	95.6	0.935	0.982
10	2921	2662	0.951	96.1	0.963	0.975

respectively. As shown in the tables, though candidate genes themselves can provide an accurate gene signature for each subtype of breast cancer, adding high confidence genes to candidate gene sets increase the classification performance.

Table 2.4: Using candidate genes corresponding to each subtype for classification.

Subtype	Candidate Genes	Accuracy (%)	F-measure	AUC	MCC
1	42	93.78	0.935	0.950	0.521
2	16	95.08	0.945	0.832	0.314
3	32	85.55	0.853	0.854	0.436
4	96	87.96	0.873	0.891	0.531
5	18	91.07	0.908	0.897	0.449
6	69	95.78	0.949	0.868	0.338
7	16	88.66	0.883	0.840	0.382
8	27	86.86	0.866	0.881	0.448
9	59	94.48	0.932	0.904	0.423
10	75	95.68	0.957	0.965	0.758

Figure 2.2 shows the genes with medium and high confidence in Subtype-1 NB. As shown in the figure, some of the hub genes in the subnetwork such as Cyclin Dependent Kinase 1 (CDK1) are known indicators in breast cancer prognosis [106] and further investigations for determining their possible roles in Subtype-1 of breast cancer is in progress.

Table 2.5: Using high confidence genes and candidate genes corresponding to each subtype for classification. High confidence genes are those that are present in more than 70% of the networks.

Subtype	High Confidence Genes	Accuracy (%)	F-measure	AUC	MCC
1	103	94.98	0.947	0.964	0.607
2	59	97.59	0.975	0.990	0.709
3	148	88.56	0.877	0.893	0.516
4	225	87.96	0.869	0.904	0.516
5	68	97.09	0.972	0.992	0.840
6	187	98.99	0.990	0.997	0.875
7	104	92.17	0.908	0.906	0.515
8	100	91.37	0.907	0.930	0.610
9	144	95.38	0.948	0.962	0.562
10	115	95.88	0.959	0.968	0.765

We used Intogen's mutational breast cancer driver genes [2] and compared them with the genes that we identified in the NBs of each breast cancer subtype. Table 2.6 shows all mutational driver genes and their overlap with NB corresponding to one of the subtypes. As shown in the table, out of 184 mutational driver genes, our model covered 125 of them as part of NBs in different breast cancer subtypes. This is impressive since our model covered these genes without having access to any mutational data corresponding to METABRIC dataset.

We also computed the *odds ratio* [176] of having a deletion or amplification in each candidate gene and compared their relation to the expression of that specific gene across different subtypes. Figure 2.3 shows the odds ratio and gene expression of candidate genes for Subtype-1 of breast cancer. Odds ratio show how a deletion/amplification in a specific gene is likely to separate a subtype from the others; the higher the ratio, the more effective is that aberration in separating one subtype from the rest. In most of the cases, we found the copy number aberration and gene expression as two independent factors, which means having a high odds ratio for a gene does not necessarily mean that the gene expression

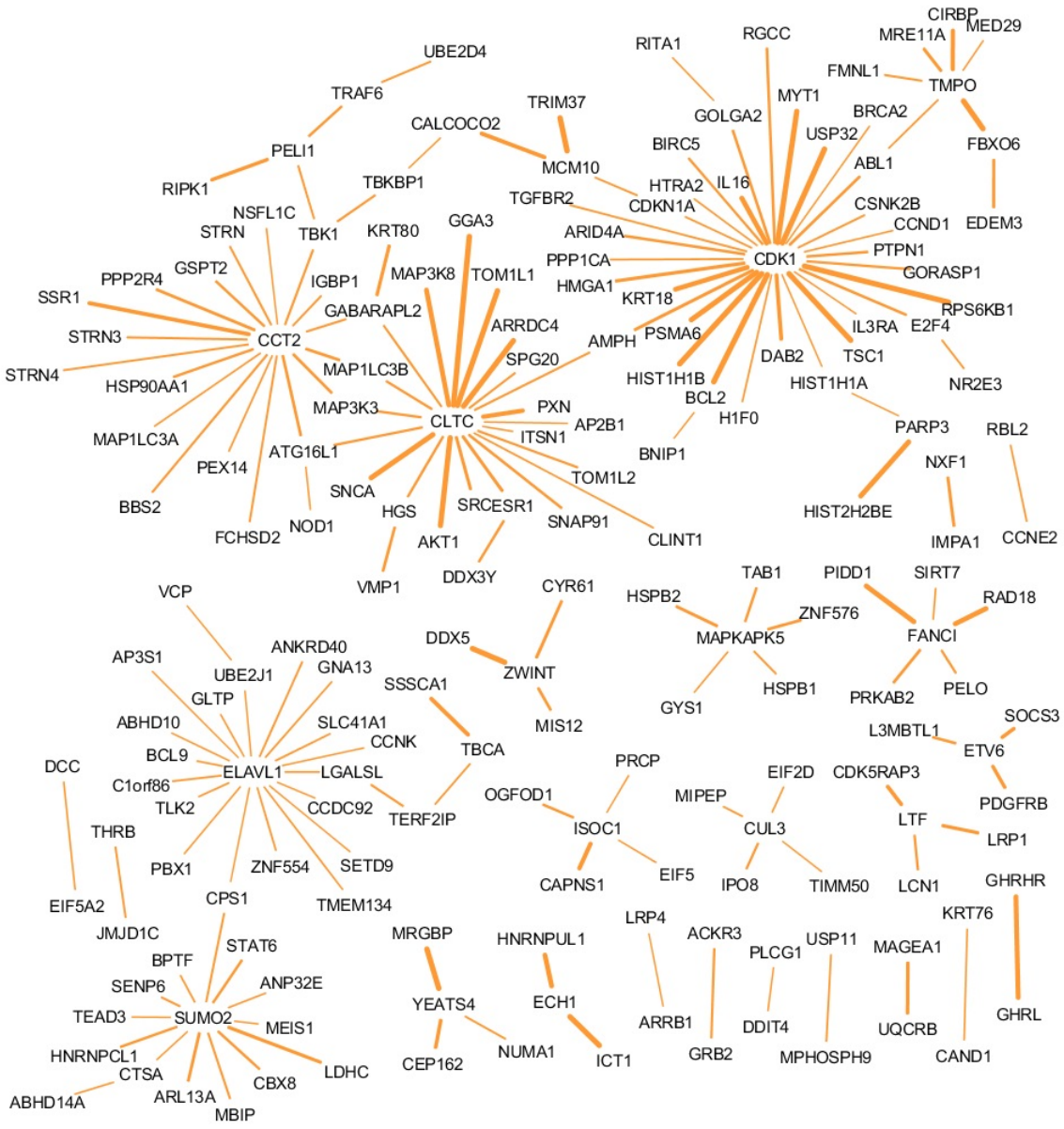


Figure 2.2: The NB of Subtype1 including medium and high confidence interactions.

pattern for that gene has totally different patterns in one subtype against the other subtypes.

Some of genes appeared in more than one subtype as a high confidence gene. Figure 2.4 depicts these genes along with the subtypes these genes belong to. As shown in the

Table 2.6: Mutational driver genes identified in the NBs of breast cancer subtypes.

Gene	Present	Gene	Present	Gene	Present	Gene	Present	Gene	Present	Gene	Present
PIK3CA		LPHN2		KDM5C	✓	KALRN	✓	ERCC2	✓	MAX	
TP53	✓	CDKN1B	✓	APC	✓	EIF4A2	✓	HSPA8	✓	EIF2C3	
PTEN	✓	TBL1XR1	✓	ARID2	✓	MGA	✓	NUP107		ARNTL	✓
AKT1	✓	BRCA2	✓	CIC	✓	MECOM	✓	ERBB2IP		KLF4	✓
SF3B1	✓	BRCA1	✓	SMAD4	✓	ARHGAP35	✓	BMPR2	✓	G3BP2	✓
KRAS		ANK3		BAP1	✓	NUP98	✓	MLH1		TCF12	✓
GATA3	✓	ERBB2	✓	PBRM1		STAG1	✓	CLTC	✓	CARM1	✓
MAP3K1	✓	MYH9	✓	DDX5	✓	SMARCA4		NOTCH1	✓	TCF7L2	✓
MLL3		MED23	✓	KEAP1	✓	BCOR	✓	SUZ12		SEC24D	
CDH1	✓	MLLT4	✓	STK11	✓	PTPRU		HLA-A	✓	ZFP36L2	✓
NCOR1	✓	ARID4B	✓	RPL5	✓	FLT3	✓	CNOT3		CAST	✓
MAP2K4	✓	RPGR		PHF6		ARFGEF2	✓	SOS2	✓	CLASP2	✓
RUNX1	✓	HCFC1		FUBP1	✓	BPTF	✓	HLF	✓	ACSL6	
NF1		MYH14		EIF1AX		FOXP1	✓	DHX15	✓	MUC20	
RB1	✓	NOTCH2	✓	MACF1	✓	CEP290		EIF4G1	✓	NF2	✓
ATM	✓	SPTAN1	✓	AHNAK	✓	MED24	✓	ACO1	✓	ITSN1	✓
ARID1A		PRKAR1A	✓	MED12	✓	CSDE1	✓	LCP1	✓	RBM5	✓
TBX3		CCAR1	✓	AKAP9	✓	EP300	✓	PIP5K1A		AQR	✓
MLL2		RFC4		TAF1		FN1	✓	NR4A2	✓	MSR1	
CBFB		CAD	✓	SVEP1	✓	BNC2	✓	CHEK2	✓	THRAP3	
CTCF		SRGAP1	✓	ASPM		CHD9	✓	MKL1	✓	GOLGA5	✓
CHD4		ACVR1B	✓	ATR	✓	POLR2B	✓	CUL1	✓	ACTB	✓
PIK3R1	✓	GPS2		MLL		PIK3CB	✓	TNPO1		RHEB	
STAG2		PRKCZ	✓	MTOR		LRP6		DIS3		ATF1	✓
CASP8	✓	FBXW7	✓	ASH1L	✓	FMR1		FUS	✓	ATIC	✓
FOXA1	✓	BRAF		NSD1		SOS1	✓	CLSPN	✓	PCSK6	
MYB	✓	NRAS	✓	CDK12	✓	PCDH18		STK4	✓	CCT5	✓
ZFP36L1		IDH1	✓	MYH11	✓	DDX3X	✓	RBBP7	✓	HNRPDL	
PAX5		SETD2		TRIO	✓	AFF4	✓	SFPQ	✓	TGFBR2	✓
TFDP1		EGFR	✓	SETDB1	✓	TOM1	✓	ELF1		STIP1	✓
PTGS1		NDRG1	✓	PIK3R3		CSNK1G3					

figure, some of the genes such as *COPS*5, *GRB*2 and *MAP*1*LC*3*B* appeared in the network of four subtypes, despite of not being among the candidate genes in any of the subtypes. This implies that these genes, in spite of having non-significant copy number aberration and gene expression simultaneously, actively participate in differentiation of several breast cancer subtypes.

2.4 Discussion

We used the IPAD pathway analysis database and tool [229] to determine the diseases associated with the hub nodes in the NBs obtained for each subtype. For example, out of

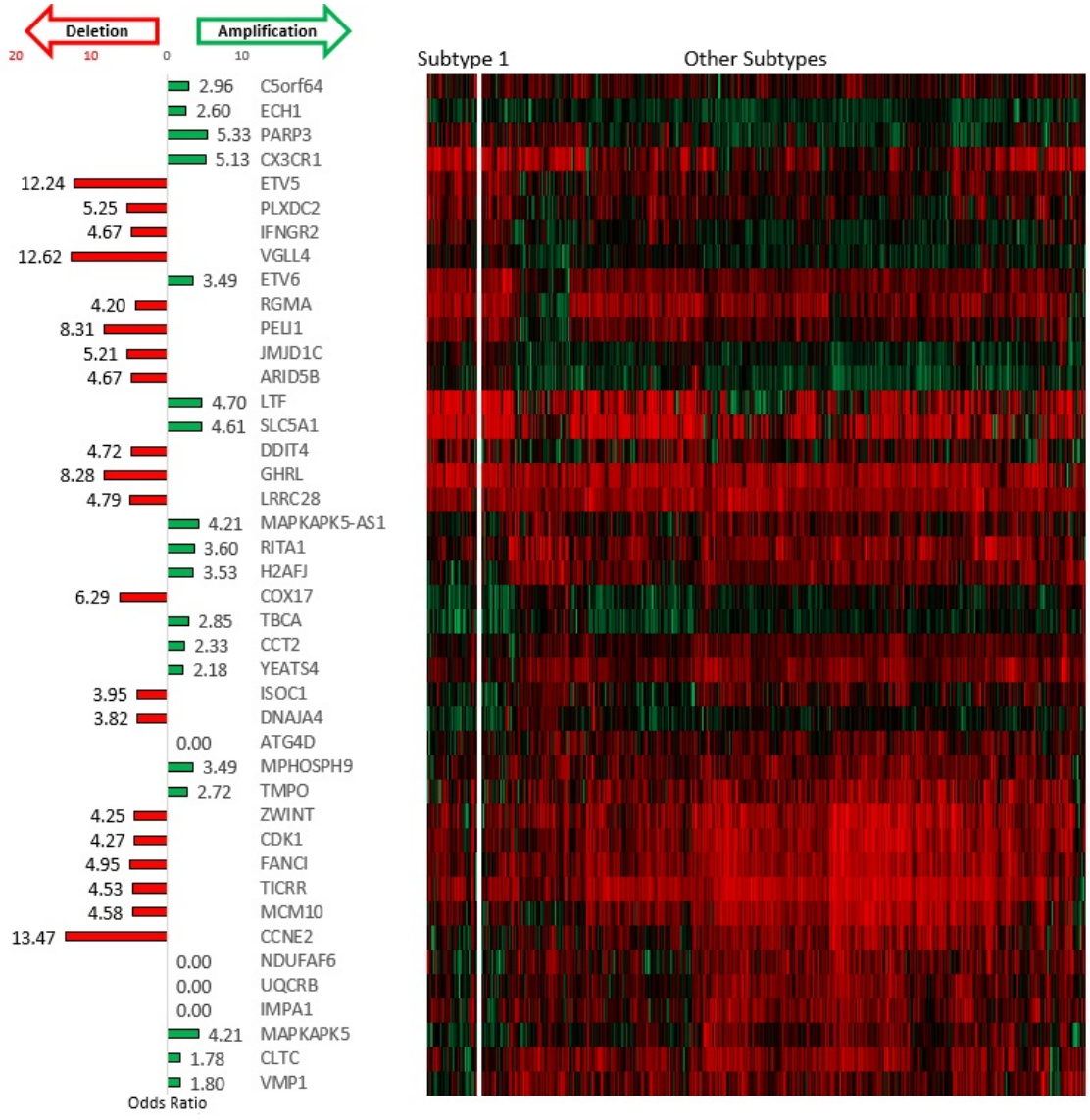


Figure 2.3: The odds ratio and gene expression of candidate genes for Subtype-1

nine genes that had more than 10 connections in the NB corresponding to Subtype-1, at least three of them have been related to breast cancer, in the literature. Table 2.7 shows the involvement of these hub genes in breast cancer.

Table 2.7: Breast related diseases corresponding to candidate genes of each subtype NB with p-value of less than 0.05.

Disease ID	Disease Name	Involved genes in the disease	Subtype	P-value
MESH:D058922	Inflammatory Breast Neoplasms	DDIT4 , ECH1, COX17, UQCRB, TMPO, CCNE2, MAPKAPK5, CDK1, FANCI, GHRL, ETV6, CX3CR1, ZWINT, H2AFJ	1	2.8e-2
MESH:D018270	Carcinoma, Ductal, Breast	CX3CR1, ETV6, LTF, CDK1, FANCI, GHRL, SLC5A1, ZWINT, IFNGR2, ECH1, DDIT4, CLTC, UQCRB, COX17, VGLL4, ARID5B, MAPKAPK5, CCNE2, TMPO, H2AFJ, CCT2	1	4.49e-2

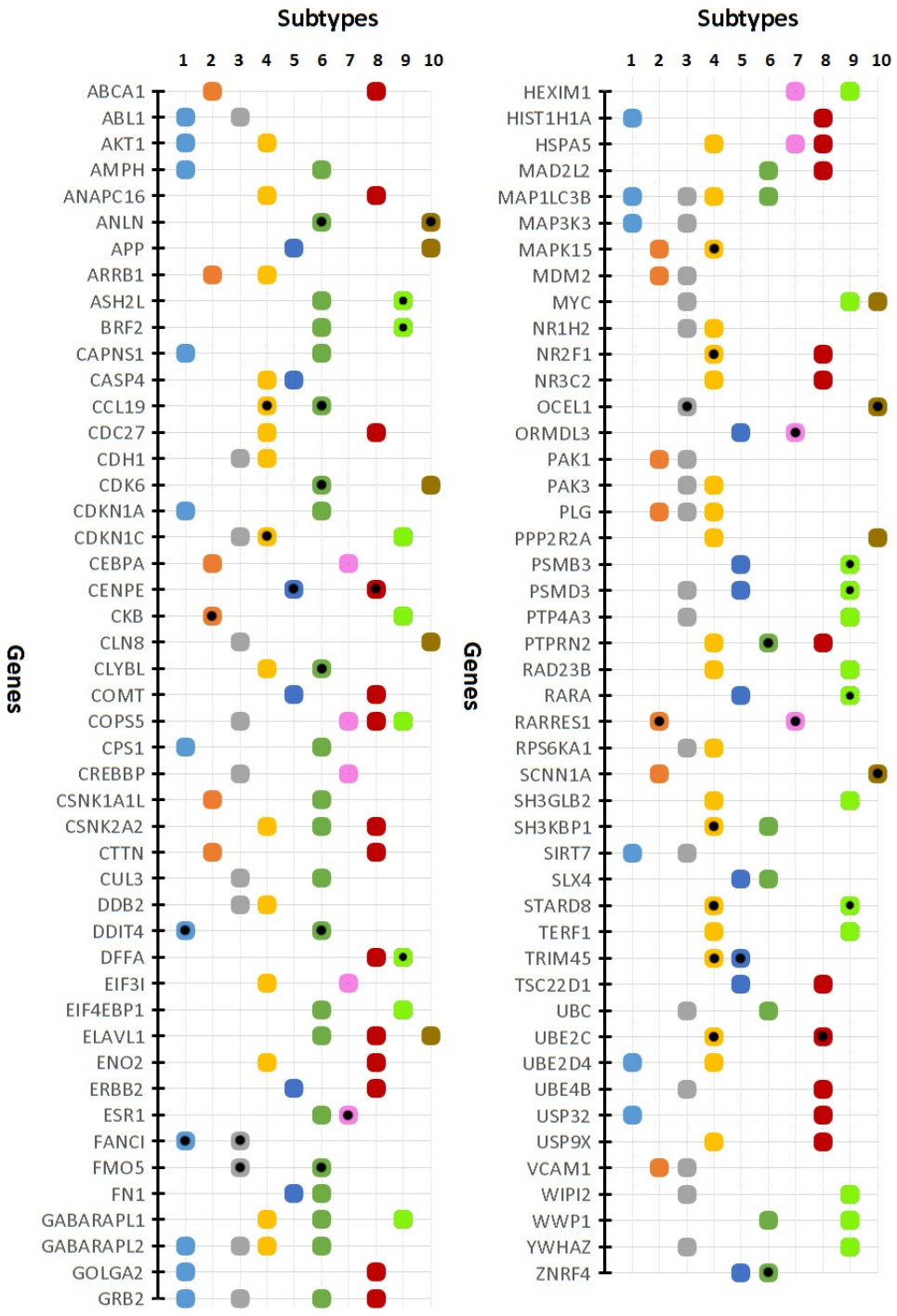


Figure 2.4: List of High confidence genes participating in NB of more than one subtype. Markers with black circle depicts the candidate genes.

Chapter 3

Computationally Repurposing Drugs for Breast Cancer Subtypes Using a Network-Based Approach

3.1 Introduction

Discovery of a new drug, especially for cancer can be a very time-consuming and costly process. It normally takes between 10 to 15 years to develop a new drug [181] and can cost upward of tens of billion dollars [7, 58, 61]. On the other hand, the success rate of developing a new drug for cancer is very low [217], and the number of new FDA-approved drugs has been declining since past 25 years [63].

Drug repositioning and repurposing are effective alternative strategies to find new uses of existing drugs. Both drug repositioning and repurposing processes consist of using an existing drug for treatment of a disease other than its primary or initial purpose. If the drug is already FDA-approved, the process is called drug repurposing, while if the drug is in trial

or experimental phase, the process is called drug repositioning. Since in this work we use the same methodology for all existing drugs, irrespective of their FDA-approval status, for simplicity, we refer to the methodology as ”drug repurposing” in a general sense.

In case of repurposed drugs (and to a lesser degree for repositioned drugs, depending on their trial or experimental stage), the overall cost and time associated with using it for treatment of other diseases is significantly lower than developing a new drug [58].

In order to repurpose an existing drug for a new disease, the main challenge is to identify new relationships between drugs and diseases. To overcome this challenge, a variety of approaches have been introduced including computational, biological and experimental approaches, as well as hybrid schemes that combine both computational and biological techniques. Computational approaches for drug repurposing bear much lower cost and other barriers in comparison to biological experimental approaches, which makes it a more appealing strategy and a very good starting point for further clinical trials and biological validations [150].

The majority of existing computational methods for drug repurposing are based on the comparison between gene expression response of various cell lines before and after treatment or a combination of several types of data corresponding to various aspects of disease-drug relationships [80, 123, 141, 234]. For example, Lotfi et al. grouped drug repurposing methods based on their principle source of biological data and core methodology, including gene regulatory networks, metabolic networks and molecular interaction networks [123], while Zou et al. categorized drug repurposing methods into two groups of data-driven and hypothesis-driven approaches [234]. Xue et al., on the other hand, focused on the underlying methodology used in drug repurposing, when it regards to categorizing those methods [223]. Luo et al. used Singular Value Thresholding (SVT) to predict scores for un-

known drug–disease pairs based on known relationship between drugs and diseases [126]. Zhang et al. utilized a drug similarity network, a disease similarity network, and known drug-disease associations to explore the potential associations among unrelated pairs of drugs and diseases [231].

Generally speaking, we can group drug repurposing approaches into three distinct groups: text-mining approaches [81, 94, 108, 109, 114, 166, 230], semantics-based approaches [139, 154, 233], and finally network-based approaches [33, 67, 107, 130, 159, 194–196, 207, 218, 219, 227]. The latter takes into the account the relationship and interactions between genes in their corresponding pathways. For example Bourdakou et al. used statistical co-expression networks to highlight and prioritize genes for breast cancer subtypes and leveraging them for drug repurposing [33]. One of the biggest difference between the proposed framework and the previous network-based methods is the ability of the proposed framework to identify not only single drugs, but also pairs of combined drugs (and theoretically unlimited number of drug combinations) for a given disease with a reasonable computational overhead, which enables it to find combinations of drugs that could far outreach the therapeutic effects of single drugs for a given breast cancer subtype (or any other disease in general).

This paper introduces a novel network-based approach to identify drugs with the highest repurposability with respect to each of ten breast cancer subtypes. This goal is achieved by first finding driver genes responsible for each subtype using genomic and transcriptomic data, which are then used along with pathway data in order to find those drugs that have the highest repurposing scores for each of ten breast cancer subtypes. The results show that the proposed method is able to identify potential effective known and experimental drugs developed for other diseases to be repurposed for various breast cancer subtypes. Indeed,

further wet lab analysis is needed to determine the therapeutic level of identified drugs on each breast cancer subtype. For reference, what we refer to here as ten breast cancer subtypes are ten distinctive sub-groups identified in [53].

Moreover, we used the proposed method to identify single and pairs of drugs for Triple Negative (TN) breast cancer tumors. Between 10% to 15% of breast cancer cases are considered as TN, where they lack any hormone epidermal growth factor receptor 2 (HER-2), estrogen receptors (ER), and progesterone receptors (PR) in the tumor [131]. Thus, the traditional targeted (often hormone) therapy that targets one of these hormones are ineffective in these cases. This lack of targeted therapies has intensified the interest in this group of patients. Our results show that the proposed method were able to computationally identify single and paired repurposed drugs that could have therapeutic effect on this this group.

3.2 Materials and Methods

For drug expression data, we used level-5 data of the LINCS dataset (from Gene Expression Omnibus with the reference number GSE70138), which consists of z-score values of more than 118,000 drug/ concentration/ treatment_time for more than 12,000 genes. In order to make the process more computationally manageable, we used only the lowest and highest dosages of each drug (generally 0.04 and 10.0 μmol, correspondingly) and a default 24-hour time-point frame for the analysis, in case of having more than one time-point frame. If a drug does not have a 24-hour time-point frame, we use the default time-point frame indicated in LINCS database.

3.2.1 Obtaining Candidate Genes

In the first step, we use CNA, CNV and GE data to find the most informative genes, separately for each subtype. To do so, we first use CNA/CNV information to find those genes that have very high genotypic aberration in each subtype based on their GISTIC score [27]. GISTIC identifies significant aberrations using two steps. In the first step, it calculates the G-score statistic, which involves both the frequency of occurrence and the amplitude of the aberration. In the second step, it assesses the significance of each aberration using Fisher's Exact test [168]. These two steps take place in 3.1(a). To make sure that we only target aberrations in the copy number and not common variations across different populations, we use the HapMap database [52] (shown in Figure 3.1(b)). HapMap is a catalog of common genetic variants that occur in human. We only consider those genes for a significant test that have CNA but no CNV. We also use gene expression data to identify the top differentially expressed genes for each subtype. For this, we used Chi2 [120] to rank genes based on their ability to separate each subtype from the remaining subtypes. At the end, after obtaining the top genes using CNA/CNV and GE data separately, if CNA/CNV analysis determined N genes as significant in terms of their genomic aberrations, we select the top N genes from GE data; then out of these two gene sets, we take the intersection as candidate genes. These candidate genes are those genes that have both significant differences in terms of gene expression and copy number aberrations.

3.2.2 Obtaining gene scores

The measurement is a normalized z-score value for each replicate of a given gene treated with the same perturbation agent (i.e., perturbagen: either drugs or small molecule compounds or others) based on 95% confidence level [202]. Thus, for each pair of gene and

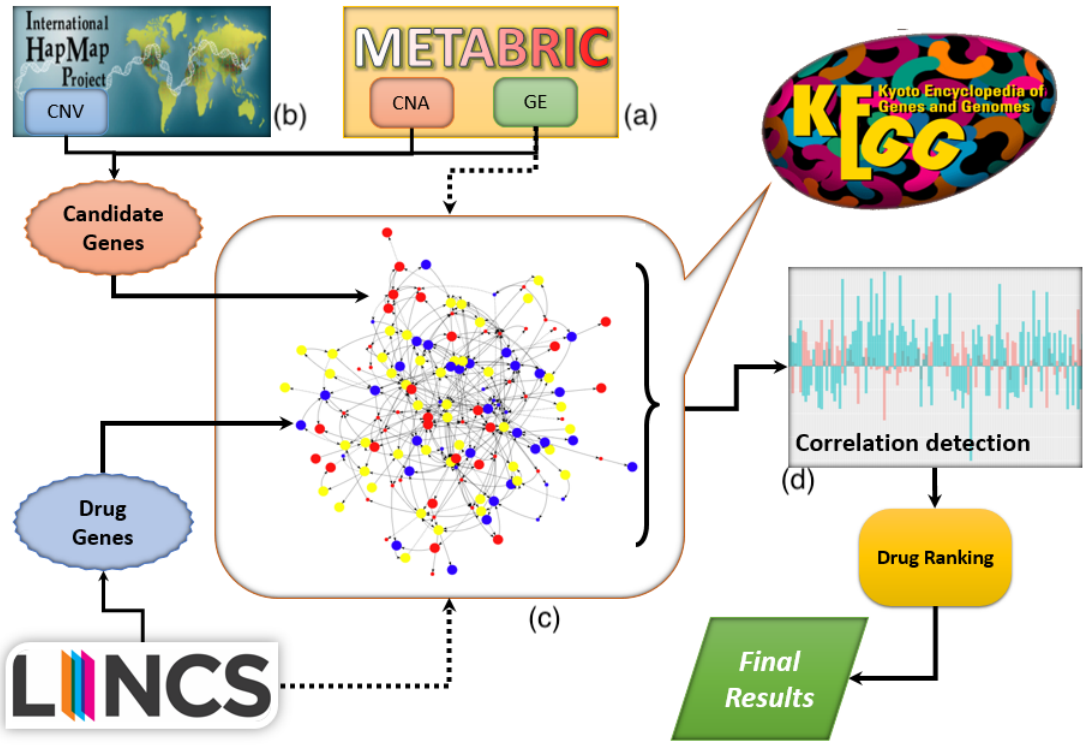


Figure 3.1: Schematic view of the proposed framework for identification of best repurposing drugs for each breast cancer subtype. METABRIC dataset is used to obtain copy number aberration and gene expression data for breast cancer subtypes. HapMap data is used to obtain copy number variation information. Lincs dataset is used to obtain the effect of different drug compounds on gene expression of cancer samples. KEGG dataset is used to create a universal pathway network [99] .

drug agent, we consider a value between -10 and 10. A value close to zero shows that the expression of a given gene will not be affected by the drug agent. In comparison, based on the concepts of gene expression inhibition and induction, a negative or positive score shows that the expression of the given gene decreases or increases, respectively, because of the effect of the drug agent.

We have used the METABRIC dataset [53], which contains the copy number values and gene expression levels of 2000 primary breast tumors with long-term clinical follow-up. It can be accessed from the European Genome-Phenome Archive using the accession number EGAS00000000083. In [53], the copy number aberrations and copy number variations generated using Affymetrix SNP 6.0 arrays and gene expression data were obtained using Illumina HT 12 technology. The dataset contains two sets of data, *validation* set and *discovery* set. Due to the lack of class labels in the validation set, in this paper we only use the discovery set, which contains 997 samples from ten subtypes of breast cancer. Each sample contains expression data for 48,803 probe IDs. The expression of all probes corresponding to the same gene have been merged based on the median expression of those probes, which maps all the probes to 24,351 unigenes. We calculate the same normalized z-score values for each of the ten breast cancer subtypes in the METABRIC dataset, (which can be accessed from European Genome-Phenome Archive with the study id EGAS00000000083), in such a way that the normalized z-score of each gene is a value between -10 and 10. A value close to zero shows that the expression of a given gene will not be affected by the disease, while a negative or positive score shows that the expression of the given gene decreases or increases, respectively, because of the effect of the disease.

3.2.3 Creating unified global human pathway

In the next step, we use the KEGG Pathway database to find all possible paths between genes [99]. A biological pathway is a series of actions among molecules in a cell that leads to a certain product or a change in the cell. It can trigger the assembly of new molecules, such as a fat or protein, turn genes on and off, or spur a cell to move [5]. The version of the KEGG Pathway database we used contains 265 human pathways. So, by taking union of all genes and also all existing direct relations between each pair of genes, we create a unified global human pathway (UGHP). The UGHP contains interaction between 4985 genes in all 265 human pathways in KEGG as a matrix, where $UGHP_{ij}$ represents signaling interaction type between gene i and gene j. The values of the matrix could be -1, 1 or 0 representing activation, suppression, or no direct signal from gene i to gene j.

3.2.4 Calculating drug-disease repurposing score

At this point, for each drug D_i and breast cancer subtype S_j, we perform the following steps:

1. Select the top 50 affected genes by the drug D_i from the LINCS dataset by ranking the genes based on their absolute z-score values and call them *drug genes*. Note that at the end of the process, the pipeline focuses only on negative correlations between drug and disease.

2. Use the candidate genes corresponding to subtype S_j that we identified in using copy number alteration (CNA), copy number variation (CNV) and gene expression (GE) data [73]. We call these candidate genes *disease genes*.

3. Map back these drug and disease genes to UGHP to create a drug-disease network,

D_iS_j, which contains the shortest paths between each pair of drug-disease genes (shown in Figure 3.1(c)). Thus, the maximum number of nodes in D_iS_j network, N, is given as follows:

$$N = G_{dr} + G_{di} + G_i \tag{3.1}$$

where G_{dr} is the number of drug genes, G_{di} is the number of disease genes, and G_i is the number of intermediate genes in the shortest path between each pair of drug and disease genes.

4. Since for each gene in this drug-disease network we have two z-score values (one for the effect of drug and one for the effect of disease), we construct two arrays, one consists of drug z-score values while the other consists of disease z-score values with identical gene order.

5. We compute Pearson correlation [110, 184] between the above arrays using the following formula (shown in Figure 3.1(d)):

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{3.2}$$

where x_i and $\bar{x}$ are the z-score value of gene i and average of all z-score values for the drug group in drug-disease network, and y_i and $\bar{y}$ are the z-score value of gene i and average of all z-score values for the disease group in the drug-disease network, respectively.

Figure 3.1 depicts the proposed framework for the identification of best repurposing drugs for each breast cancer subtype.

Obtaining a positive correlation between a given drug genes and the disease genes

means that the drug and the disease have similar effect on the genes in the drug-disease network. In contrast, obtaining a negative correlation implies that the drug's effect on the genes in the drug-disease network is opposite to the effect of the disease on the genes in that network.

Obtaining a negative correlation is a favorable case in this context, because we are looking for drugs that could have a potential reverse effect on the genes affected by the disease.

3.2.5 Identifying Combinations of Drugs for Repurposing

In the previous section, we solely focused on effects of each individual perturbation agent on each subtype of breast cancer. In this section, we test the hypothesis that combination of two or more drugs might be more effective in terms of reversing the effect of the disease, i.e., by generating a more negative correlation with the disease than each drug independently. For simplicity, in this step, we assume that the z-score value of a given pair of drugs to be additive with respect to the z-score value of each of those drugs independently. In other words, if we assume that the z-score value of drug D_i on gene G is X_i, and the z-score value of drug D_j on the same gene G is X_j, we can then assume that the z-score value for the given pair of drugs [D_i,D_j] is X, where $X = X_i + X_j$.

Thus, in order to find the best repurposed pair of drugs for a given subtype of breast cancer, first, we calculate the combined z-score value of all genes for every pair of drugs, and then we pick the top genes with the highest absolute value of their combined z-score. Figure 3.2 depicts the proposed framework for identification of the best pair of repurposed drugs for each breast cancer subtype.

3.2.6 Calculating drug-disease repurposing score for a pair of drugs

For pair of drug D_i and D_j and breast cancer subtype S_k, we perform the following steps:

1. Calculate the combined z-score value of all genes for pair of drugs D_{ij}, and selecting the top 50 genes with the highest absolute value of their combined z-score values as *paired drug genes*.

2. Use the candidate genes corresponding to subtype S_k that we identified in using copy number alteration (CNA), copy number variation (CNV) and gene expression (GE) data [73]. We call these candidate genes *disease genes*.

3. Map back these drug and disease genes to UGHP to create a drug-disease network, $D_{ij}S_k$, which contains the shortest paths between each pair of drug-disease genes (shown in Figure 3.2(c)). Thus, the maximum number of nodes in $D_{ij}S_k$ network, N, is given as follows:

$$N = G_{dr} + G_{di} + G_{ij} \tag{3.3}$$

   where G_{dr} is the number of drug genes, G_{di} is the number of disease genes, and G_{ij} is the number of intermediate genes in the shortest path between each pair of combined drug and disease gene.

4. Since for each gene in this drug-disease network we have two z-score values (one for the effect of drug and one for the effect of disease), we construct two arrays, one consists of drug z-score values while the other consists of disease z-score values with identical gene order.

5. We compute Pearson correlation [110, 184] between the above arrays using the fol-

lowing formula (shown in Figure 3.2(d)):

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{3.4}$$

where x_i and $\bar{x}$ are the z-score value of gene i and average of all z-score values for the drug group in drug-disease network, and y_i and $\bar{y}$ are the z-score value of gene i and average of all z-score values for the disease group in the drug-disease network, respectively.

6. Finally, we perform a post-verification analysis on drug interference for the identified pairs of drugs using *DrugBank's Interaction Checker* tool, in order to confirm if there is any known interference between any of those identified pairs. (shown in Figure 3.2(e)).

   A drug-drug interference is a situation in which one drug affects the activity of another. Drugs may interact with each other to cause side effects that are unexpected or unintended. If any pair of drugs have known drug-drug interference, we remove them from the analysis. For example, the combination of Tadalafil and palbociclib generated a negative correlation of -0.65 with subtype 3, which put them in the top 10 list of paired-drugs for this subtype. But given the fact that they have a known moderate interaction with each other, this pair has been removed from the analysis [3]. The reason for doing a post-verification analysis instead of checking it as a pre-process step, is that the post-verification approach gives us the flexibility of updating the results with newly discovered drug interference in the future without a need to rerun the analysis. Also, using a post-verification approach gives us the ability to deal with interferences between a given pair of drugs at different levels depending on the type

or level of interference.

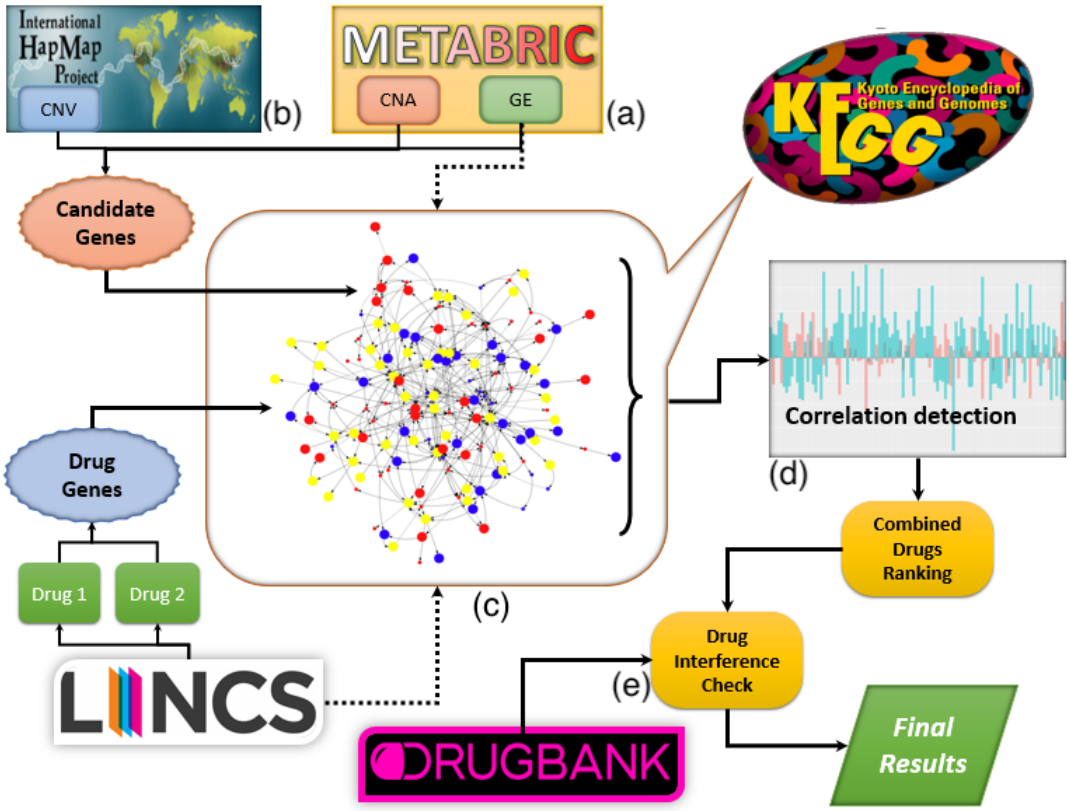


Figure 3.2: Schematic view of the proposed framework for identification of best pair of repurposing drugs for each breast cancer subtype. METABRIC dataset is used to obtain copy number aberration and gene expression data for breast cancer subtypes. HapMap data is used to obtain copy number variation information. Lincs dataset is used to obtain the effect of different drug compounds on gene expression of cancer samples. KEGG dataset is used to create a universal pathway network [99]. And finally, DrugBank's drug interference checker is used to check any possible interference between each pair of drugs.

3.2.7 Extension to Triple-Negative Breast Cancer Tumors

In this section, we leverage the proposed pipeline for the ten breast cancer subtypes considered in the previous section, to identify potential repurposable drugs specifically for

triple negative beast cancer tumors. In order to do so for identifying candidate genes, we treat triple negative samples in METABRIC as one group and all the remaining samples as another group. By running the pipeline introduced in [73], we identify the most discriminative genes in terms of gene expression and copy number aberration between TN and non-TN groups. Then, using the pipelines depicted in Figures 3.1 and 3.2, we identify the top repurposing single and paired drugs for the TNBC subtype.

3.3 Results and Discussion

For reference, Table 3.1 shows the list of drugs that have been approved by FDA to date for breast cancer treatment [1]. The results show that the proposed model is able to identify highly negative correlated drugs corresponding to each of ten breast cancer subtypes, both when used in single drug mode or for identifying pairs of drugs. Some of the well-known and widely used breast cancer drugs have been identified among the top drugs, which again shows that the proposed approach was able to pick up current drugs with high accuracy. For example, Goserelin (Zoladex) is a well known and FDA approved hormone therapy drug for treatment of BC that showed up in top ten drugs for subtypes 2 and 8. Also, Palbociclib (Ibrance) is another well known and FDA approved chemo therapy drug for treatment of BC that showed up in top ten drugs for subtype 4. Moreover, Ruxolitinib, which showed up in top ten drugs for 9 out of 10 subtypes (table 3.2) has been under several trials and studies regarding its potential inhibiting effects on BC [105, 193].

Table 3.1: List of FDA-approved drugs for breast cancer treatment.

Abemaciclib	Kadcyla (Ado-Trastuzumab Emtansine)
Abitrexate (Methotrexate)	Kisqali (Ribociclib)
Abraxane	Lapatinib Ditosylate
Ado-Trastuzumab Emtansine	Letrozole
Afinitor (Everolimus)	Lynparza (Olaparib)
Anastrozole	Megestrol Acetate
Aredia (Pamidronate Disodium)	Methotrexate
Arimidex (Anastrozole)	Methotrexate LPF (Methotrexate)
Aromasin (Exemestane)	Mexate (Methotrexate)
Capecitabine	Mexate-AQ (Methotrexate)
Clafen (Cyclophosphamide)	Neosar (Cyclophosphamide)
Cyclophosphamide	Neratinib Maleate
Cytoxan (Cyclophosphamide)	Nerlynx (Neratinib Maleate)
Docetaxel	Nolvadex (Tamoxifen Citrate)
Doxorubicin Hydrochloride	Olaparib
Ellence (Epirubicin Hydrochloride)	Paclitaxel
Epirubicin Hydrochloride	Ixempra (Ixabepilone)
Eribulin Mesylate	Palbociclib
Everolimus	Pamidronate Disodium
Exemestane	Perjeta (Pertuzumab)
5-FU (Fluorouracil Injection)	Pertuzumab
Fareston (Toremifene)	Ribociclib
Faslodex (Fulvestrant)	Tamoxifen Citrate
Femara (Letrozole)	Taxol (Paclitaxel)
Fluorouracil Injection	Taxotere (Docetaxel)
Folex (Methotrexate)	Thiotepa
Folex PFS (Methotrexate)	Toremifene
Fulvestrant	Trastuzumab
Gemcitabine Hydrochloride	Tykerb (Lapatinib Ditosylate)
Gemzar (Gemcitabine Hydrochloride)	Velban (Vinblastine Sulfate)
Goserelin Acetate	Velsar (Vinblastine Sulfate)
Halaven (Eribulin Mesylate)	Verzenio (Abemaciclib)
Herceptin (Trastuzumab)	Vinblastine Sulfate
Ibrance (Palbociclib)	Xeloda (Capecitabine)
Ixabepilone	Zoladex (Goserelin Acetate)

Table 3.2: Rank comparison among the top 30 drugs across all 10 breast cancer subtypes

Overall Rank	Drugs	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Median
1	Ruxolitinib	1	3	2	43	1	1	2	2	2	1	2
2	Tranilast	4	12	11	193	7	11	4	43	32	4	11
3	Rupatadine	22	44	6	1	31	2	15	11	1	51	13
4	Ribavirin	2	10	7	2	9	20	60	66	21	17	13.5
5	Deferiprone	3	14	3	15	27	4	20	6	183	20	14.5
6	Etofylline-Clofibrate	16	29	9	20	15	45	5	15	45	9	15.5
7	Fingolimod	36	6	58	137	16	9	3	109	20	13	18
8	ICI-185282	61	4	21	46	17	7	14	39	35	3	19
9	PF-04217903	10	25	16	398	10	40	10	107	25	10	20.5
10	Raloxifene	30	43	26	7	25	31	8	4	3	75	25.5
11	EDTA	13	55	10	8	18	127	112	196	38	6	28
12	Amiprilose	23	46	57	13	42	10	61	14	7	34	28.5
13	Bafilomycin A1	101	28	33	6	21	34	21	30	191	5	29
14	Dexamethasone	27	21	91	171	32	68	23	87	17	23	29.5
15	Dofequidar	11	88	24	14	106	29	104	29	31	56	30
16	MK-1775	14	49	12	61	24	37	13	46	41	7	30.5
17	TG-100801	121	50	44	60	6	17	11	18	8	53	31
18	Swainsonine	15	58	17	34	90	48	63	28	23	27	31
19	Raclopride	24	54	13	4	30	36	19	71	137	36	33
20	L-690330	37	7	38	35	61	32	12	314	5	133	36
21	Phentermine	57	22	43	57	135	5	36	12	26	55	39.5
22	PHA-767491	64	8	8	153	13	257	9	325	379	15	39.5
23	PD-173074	59	18	42	108	37	27	17	356	15	142	39.5
24	Lidocaine	122	17	103	212	12	88	25	17	51	30	40.5
25	Mibampator	98	15	20	22	93	12	46	60	85	37	41.5
26	PD-153035	26	108	214	18	35	128	33	47	235	39	43
27	AKT-inhibitor-1-2	21	24	40	110	48	57	34	150	282	22	44
28	Maraviroc	41	77	25	9	41	3	48	149	132	415	44.5
29	SDZ-NKT-343	104	5	65	203	5	39	6	51	241	24	45
30	Clomipramine	48	198	37	177	22	46	44	155	153	32	47

3.3.1 Single drug repurposing

Figure 3.3 shows the distribution of drug repurposing scores across the ten breast cancer subtypes. There are a few interesting observations. First, the response level of different BC subtypes to tested drugs are different. While the distribution of correlation scores among the tested drugs versus some of the subtypes such as subtypes 1, 4 and 6 are relatively narrow (which implies relatively lower response level of the aforementioned subtypes to the tested drugs), in some other subtypes, such as subtypes 2 and 8, we observe a wider distribution of these scores. This shows that effects of tested drugs could be widely different across subtypes. The second observation is regarding the median of these scores. As shown in the figure, in all subtypes, we observe a slight distribution bias toward negative repurposing scores, which implies that the tested drugs tend to exhibit more of a therapeutic effect than adverse effect.

Tables 3.3 to 3.12 show the top 20 inhibiting drugs corresponding to each of the ten subtypes. These drugs fall into three categories. *Experimental* drugs are those that are at the pre-clinical or at an animal testing stage. *Investigational* drugs are those that are in stage I, II or III of human clinical trials. Finally, *Approved* drugs are those drugs that have already been approved by FDA to be used for treatment of various diseases. Drugs that are FDA approved to be used for BC treatment (i.e. those listed in 3.1) have been highlighted in bold. Also, reference column lists any publication that suggested usage of that drug for BC treatment.

Some of the drugs in these lists are well-known and have been used extensively for either breast cancer or other types of cancer. For example, *Raloxifene* is among the top ten drugs in most, if not all, of the ten subtypes. It was originally approved by FDA in 1997 for the management and prevention of osteoporosis in postmenopausal women and reduction

Figure 3.3: Distribution of drug-disease correlation for 3,742 drugs across 10 breast cancer subtypes.

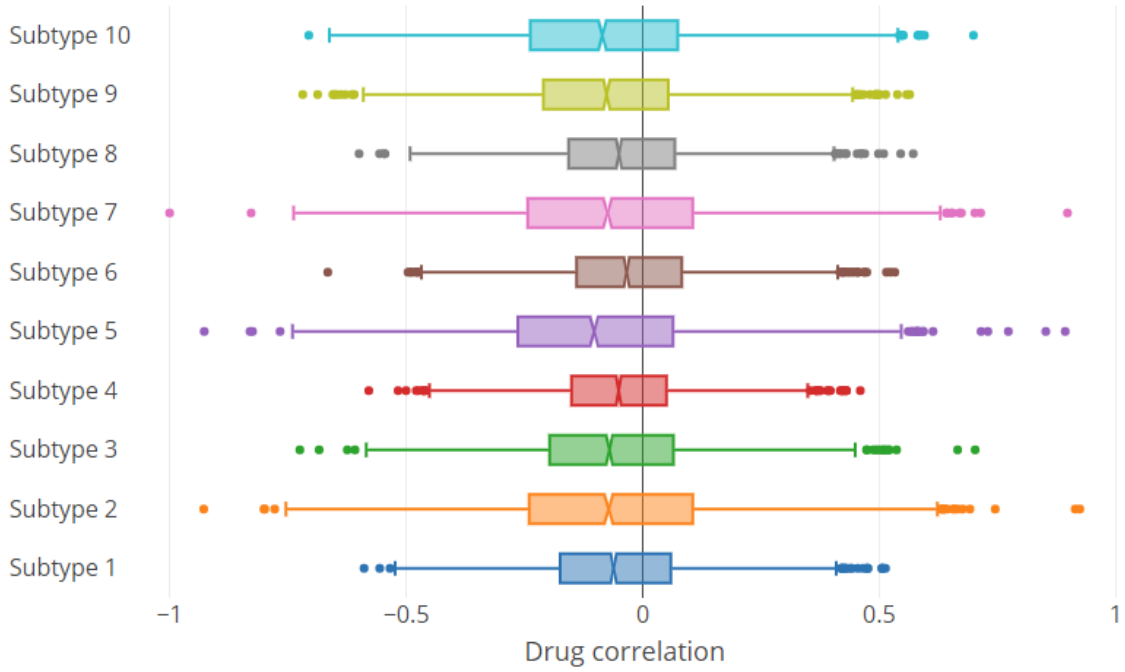


Table 3.3: Top 20 drugs corresponding to subtype 1

Rank	Drug	Dosage (μm)	Treatment Time (hours)	Score	Drug Type	**References**
1	Ruxolitinib	10	24	-0.589	Approved	**[105, 193]**
2	Ribavirin	10	24	-0.556	Approved	**[43]**
3	Deferiprone	0.04	24	-0.534	Approved	**[57, 72]**
4	Tranilast	0.04	24	-0.523	Investigational	**[152]**
5	Tadalafil	10	24	-0.521	Approved	
6	Rimexolone	0.04	24	-0.515	Approved	
7	Bardoxolone methyl	0.04	24	-0.511	Investigational	**[104, 220]**
8	Sirolimus	0.04	24	-0.507	Approved	**[90, 140]**
9	Crizotinib	0.04	24	-0.503	Approved	**[19, 116]**
10	PF-04217903	0.04	24	-0.499	Investigational	
11	Dofequidar	10	24	-0.498	Experimental	**[164, 175]**
12	GSK-2636771	10	24	-0.495	Investigational	
13	Edetic acid	10	24	-0.494	Approved	
14	MK-1775	10	24	-0.490	Investigational	**[18]**
15	MF-101	10	24	-0.478	Investigational	
16	Ranolazine	0.04	24	-0.476	Approved	**[112]**
17	Semaxanib	0.04	24	-0.475	Experimental	**[101]**
18	Iniparib	10	24	-0.474	Investigational	**[62]**
19	AKT-inhibitor 1/2	10	24	-0.473	Experimental	
20	Rupatadine	0.04	24	-0.473	Approved	

Table 3.4: Top 20 drugs corresponding to subtype 2

Rank	Drug	Dosage (μm)	Treatment Time (hours)	Score	Drug Type	**References**
1	Bromocriptine	0.04	24 h	-0.93	Approved	**[180]**
2	**Goserelin-Acetate**	0.04	24 h	-0.8	Approved	**[226]**
3	Ruxolitinib	10	24 h	-0.8	Approved	**[105, 193]**
4	ICI-185,282	0.04	24 h	-0.78	Experimental	
5	SDZ-NKT-343	10	24 h	-0.75	Experimental	
6	Fingolimod	0.04	24 h	-0.72	Approved	**[172]**
7	L-690330	10	24 h	-0.71	Investigational	
8	PHA-767491	10	24 h	-0.68	Experimental	**[132]**
9	Tolvaptan	10	24 h	-0.67	Approved	
10	Ribavirin	10	24 h	-0.65	Approved	**[43]**
11	Hymecromone	0.04	24 h	-0.64	Investigational	**[6]**
12	Tranilast	0.04	24 h	-0.64	Investigational	**[152]**
13	Sapitinib	10	24 h	-0.64	Investigational	**[78]**
14	Deferiprone	0.04	24 h	-0.63	Approved	**[57, 72]**
15	Mibampator	0.04	24 h	-0.63	Investigational	
16	Citrulline	0.04	24 h	-0.63	Investigational	
17	Lidocaine	0.04	24 h	-0.63	Approved	**[47, 119]**
18	PD-173074	10	24 h	-0.62	Experimental	**[50]**
19	Ibuprofen	10	24 h	-0.62	Approved	**[156, 222]**
20	Garcinol	10	24 h	-0.62	Experimental	**[10, 174]**

Table 3.5: Top 20 drugs corresponding to subtype 3

Rank	Drug	Dosage (μm)	Treatment Time (hours)	Score	Drug Type	**References**
1	Bromocriptine	0.04	24 h	-0.72	Approved	**[180]**
2	Ruxolitinib	10	24 h	-0.68	Approved	**[105, 193]**
3	Deferiprone	0.04	24 h	-0.62	Approved	**[57, 72]**
4	Tadalafil	10	24 h	-0.61	Approved	
5	Sirolimus	0.04	24 h	-0.58	Approved	**[90, 140]**
6	Rupatadine	0.04	24 h	-0.58	Approved	
7	Ribavirin	10	24 h	-0.57	Approved	**[43]**
8	Pha-767491	10	24 h	-0.57	Experimental	**[132]**
9	Etofylline-Clofibrate	10	24 h	-0.57	Approved	
10	EDTA	10	24 h	-0.56	Approved	**[21]**
11	Tranilast	0.04	24 h	-0.56	Investigational	**[152]**
12	MK-1775	10	24 h	-0.56	Investigational	**[18]**
13	Raclopride	10	24 h	-0.54	Investigational	
14	Iniparib	10	24 h	-0.54	Investigational	**[62]**
15	Bexarotene	10	24 h	-0.54	Approved	**[49]**
16	PF-04217903	0.04	24 h	-0.54	Investigational	
17	Swainsonine	0.05	24 h	-0.53	Experimental	**[148]**
18	JTC-801	0.04	24 h	-0.53	Experimental	**[115]**
19	Ofloxacin	10	24 h	-0.53	Approved	**[137]**
20	Mibampator	0.04	24 h	-0.52	Investigational	

Table 3.6: Top 20 drugs corresponding to subtype 4

Rank	Drug	Dosage (μm)	Treatment Time (hours)	Score	Drug Type	**References**
1	Rupatadine	0.04	24 h	-0.58	Approved	
2	Ribavirin	10	24 h	-0.52	Approved	**[43]**
3	**Palbociclib**	0.04	24 h	-0.5	Approved	**[204]**
4	Raclopride	10	24 h	-0.48	Investigational	
5	PHA-793887	10	24 h	-0.47	Investigational	
6	Bafilomycin A1	0.05	6 h	-0.46	Experimental	
7	Raloxifene	0.04	24 h	-0.46	Approved	**[45]**
8	EDTA	0.04	24 h	-0.46	Approved	**[21]**
9	Maraviroc	10	24 h	-0.46	Approved	**[158]**
10	Ebselen	10	24 h	-0.45	Investigational	**[200]**
11	Dasatinib	10	24 h	-0.45	Approved	**[203]**
12	Labetalol	0.04	24 h	-0.45	Approved	
13	Amiprilose	0.04	24 h	-0.44	Experimental	
14	Dofequidar	10	24 h	-0.44	Investigational	**[173]**
15	Deferiprone	0.04	24 h	-0.44	Approved	**[57,72]**
16	Sapitinib	10	24 h	-0.44	Investigational	**[78]**
17	Calcitriol	0.04	24 h	-0.44	Approved	**[197]**
18	PD-153035	0.04	24 h	-0.43	Investigational	**[77]**
19	Finasteride	10	24 h	-0.43	Approved	**[133]**
20	Etofylline-Clofibrate	10	24 h	-0.43	Approved	

Table 3.7: Top 20 drugs corresponding to subtype 5

Rank	Drug	Dosage (μm)	Treatment Time (hours)	Score	Drug Type	**References**
1	Ruxolitinib	10	24 h	-0.93	Approved	**[105, 193]**
2	AMG-837	0.04	24 h	-0.83	Experimental	
3	Citrulline	0.04	24 h	-0.82	Investigational	
4	Loperamide	10	24 h	-0.77	Approved	**[66]**
5	SDZ-NKT-343	10	24 h	-0.74	Experimental	
6	TG-100801	0.04	24 h	-0.72	Investigational	
7	Tranilast	0.04	24 h	-0.72	Investigational	**[152]**
8	Bisoprolol	10	24 h	-0.71	Approved	**[216]**
9	Ribavirin	10	24 h	-0.7	Approved	**[43]**
10	PF-04217903	0.04	24 h	-0.7	Investigational	
11	Sirolimus	0.04	24 h	-0.69	Approved	**[90, 140]**
12	Lidocaine	0.04	24 h	-0.69	Approved	**[47, 119]**
13	PHA-767491	10	24 h	-0.68	Experimental	**[132]**
14	WZ-4-145	0.04	3 h	-0.68	Experimental	
15	Etofylline-Clofibrate	10	24 h	-0.67	Approved	
16	Fingolimod	0.04	24 h	-0.67	Approved	**[172]**
17	ICI-185,282	0.04	24 h	-0.67	Experimental	
18	EDTA	10	24 h	-0.67	Approved	**[21]**
19	Labetalol	0.04	24 h	-0.66	Approved	
20	MF-101	10	24 h	-0.66	Experimental	

Table 3.8: Top 20 drugs corresponding to subtype 6

Rank	Drug	Dosage (μm)	Treatment Time (hours)	Score	Drug Type	**References**
1	Ruxolitinib	10	24 h	-0.67	Approved	**[105, 193]**
2	Rupatadine	0.04	24 h	-0.5	Approved	
3	Maraviroc	10	24 h	-0.49	Approved	**[158]**
4	Deferiprone	0.04	24 h	-0.49	Approved	**[57, 72]**
5	Phentermine	0.04	24 h	-0.49	Approved	
6	Iniparib	10	24 h	-0.48	Investigational	**[151]**
7	ICI-185,282	0.04	24 h	-0.48	Experimental	
8	Racecadotril	0.04	24 h	-0.48	Investigational	
9	Fingolimod	0.04	24 h	-0.47	Approved	**[172]**
10	Amiprilose	0.04	24 h	-0.47	Experimental	
11	Tranilast	0.04	24 h	-0.46	Investigational	**[152]**
12	Mibampator	0.04	24 h	-0.46	Investigational	
13	Favipiravir	0.04	24 h	-0.45	Approved	
14	Selisistat	0.04	24 h	-0.45	Experimental	
15	ZD-7288	10	24 h	-0.45	Experimental	
16	Proglumide	10	24 h	-0.44	Experimental	
17	TG-100801	0.04	24 h	-0.44	Investigational	
18	Ranolazine	0.04	24 h	-0.44	Approved	**[112]**
19	Semaxanib	0.04	24 h	-0.44	Investigational	**[101]**
20	Ribavirin	10	24 h	-0.44	Approved	**[43]**

Table 3.9: Top 20 drugs corresponding to subtype 7

Rank	Drug	Dosage (μm)	Treatment Time (hours)	Score	Drug Type	**References**
1	Bromocriptine	0.04	24 h	-1	Approved	**[180]**
2	Ruxolitinib	10	24 h	-0.83	Approved	**[105, 193]**
3	Fingolimod	0.04	24 h	-0.74	Approved	**[172]**
4	Tranilast	0.04	24 h	-0.72	Investigational	**[152]**
5	Etofylline-Clofibrate	10	24 h	-0.71	Approved	
6	SDZ-NKT-343	10	24 h	-0.71	Experimental	
7	Isbufylline	0.04	24 h	-0.71	Experimental	
8	Raloxifene	0.04	24 h	-0.7	Approved	**[45]**
9	PHA-767491	10	24 h	-0.69	Experimental	**[132]**
10	PF-04217903	0.04	24 h	-0.69	Investigational	
11	TG-100801	0.04	24 h	-0.69	Investigational	
12	L-690330	10	24 h	-0.69	Investigational	
13	MK-1775	10	24 h	-0.68	Investigational	**[26, 55]**
14	ICI-185,282	0.04	24 h	-0.68	Experimental	
15	Rupatadine	0.04	24 h	-0.67	Approved	
16	Hymecromone	0.04	24 h	-0.66	Investigational	**[6]**
17	PD-173074	10	24 h	-0.66	Experimental	**[50]**
18	MG-132	20	24 h	-0.66	Experimental	**[23]**
19	Raclopride	10	24 h	-0.65	Investigational	
20	Deferiprone	0.04	24 h	-0.65	Approved	**[57, 72]**

Table 3.10: Top 20 drugs corresponding to subtype 8

Rank	Drug	Dosage (μm)	Treatment Time (hours)	Score	Drug Type	**References**
1	Semaxanib	0.04	24 h	-0.6	Investigational	**[101]**
2	Ruxolitinib	10	24 h	-0.56	Approved	**[105, 193]**
3	**Goserelin-Acetate**	0.04	24 h	-0.55	Approved	**[226]**
4	Raloxifene	0.04	24 h	-0.54	Approved	**[45]**
5	XMD11-85h	0.04	3 h	-0.49	Experimental	
6	Deferiprone	0.04	24 h	-0.49	Approved	**[57, 72]**
7	Cinepazide	0.04	24 h	-0.48	Investigational	
8	Ebselen	10	24 h	-0.48	Investigational	**[200]**
9	WH-4-025	0.04	24 h	-0.48	Experimental	
10	Nimesulide	0.04	24 h	-0.47	Approved	**[95]**
11	Rupatadine	0.04	24 h	-0.47	Approved	
12	Phentermine	0.04	24 h	-0.47	Approved	
13	MF-101	10	24 h	-0.47	Experimental	
14	Amiprilose	0.04	24 h	-0.47	Experimental	
15	Etofylline-Clofibrate	10	24 h	-0.47	Approved	
16	XMD-1150	10	3 h	-0.46	Experimental	
17	Lidocaine	10	24 h	-0.46	Approved	**[47, 119]**
18	TG-100801	0.04	24 h	-0.46	Investigational	
19	Dasatinib	10	24 h	-0.45	Approved	**[203]**
20	Apitolisib	10	24 h	-0.45	Investigational	**[93]**

Table 3.11: Top 20 drugs corresponding to subtype 9

Rank	Drug	Dosage (μm)	Treatment Time (hours)	Score	Drug Type	**References**
1	Rupatadine	0.04	24 h	-0.72	Approved	
2	Ruxolitinib	10	24 h	-0.69	Approved	**[105, 193]**
3	Raloxifene	0.04	24 h	-0.65	Approved	**[45]**
4	Emtricitabine	10	24 h	-0.65	Approved	**[201]**
5	L-690330	10	24 h	-0.65	Investigational	
6	Tepotinib	10	24 h	-0.64	Approved	**[87]**
7	Amiprilose	0.04	24 h	-0.63	Experimental	
8	TG-100801	0.04	24 h	-0.61	Investigational	
9	MG-132	20	24 h	-0.61	Experimental	**[23]**
10	Belinostat	0.04	24 h	-0.59	Approved	**[124, 235]**
11	Bromocriptine	0.04	24 h	-0.59	Approved	**[180]**
12	Vidarabine	0.04	24 h	-0.59	Approved	
13	Ranolazine	0.04	24 h	-0.58	Approved	**[112]**
14	Lisinopril	0.04	24 h	-0.58	Approved	**[216]**
15	PD-173074	10	24 h	-0.57	Experimental	**[50]**
16	Vilazodone	10	24 h	-0.57	Approved	**[84]**
17	Dexamethasone	0.04	24 h	-0.57	Approved	**[39, 225]**
18	Semaxanib	0.04	24 h	-0.57	Investigational	**[101]**
19	Mocetinostat	0.04	24 h	-0.57	Investigational	**[102]**
20	Fingolimod	0.04	24 h	-0.56	Approved	**[172]**

Table 3.12: Top 20 drugs corresponding to subtype 10

Rank	Drug	Dosage (μm)	Treatment Time (hours)	Score	Drug Type	**References**
1	Ruxolitinib	10	24 h	-0.71	Approved	**[105, 193]**
2	Bitopertin	10	24 h	-0.66	Investigational	
3	ICI-185,282	0.04	24 h	-0.66	Experimental	
4	Tranilast	0.04	24 h	-0.65	Investigational	**[152]**
5	Bafilomycin A1	0.05	6 h	-0.65	Experimental	
6	EDTA	10	24 h	-0.64	Approved	**[21]**
7	MK-1775	10	24 h	-0.63	Investigational	**[26, 55]**
8	Emtricitabine	10	24 h	-0.63	Approved	**[201]**
9	Etofylline-Clofibrate	10	24 h	-0.63	Approved	
10	PF-04217903	0.04	24 h	-0.61	Investigational	
11	Sapitinib	10	24 h	-0.61	Investigational	**[78]**
12	Bisoprolol	10	24 h	-0.61	Approved	**[216]**
13	Fingolimod	0.04	24 h	-0.61	Approved	**[172]**
14	XMD11-85h	0.04	3 h	-0.61	Experimental	
15	PHA-767491	10	24 h	-0.6	Experimental	**[132]**
16	Finasteride	10	24 h	-0.6	Approved	**[133]**
17	Ribavirin	10	24 h	-0.6	Approved	**[43]**
18	Labetalol	0.04	24 h	-0.6	Approved	
19	MG-132	20	24 h	-0.6	Experimental	**[23]**
20	Deferiprone	0.04	24 h	-0.59	Approved	**[57, 72]**

in risk for invasive breast cancer. However, recent studies have shown that this drug might be effective for breast cancer treatments [9, 160]. Also, *Ruxolitinib*, which is among the top three drugs for all but subtype 4, was approved by the FDA for the treatment of patients with intermediate or high-risk myelofibrosis [134], though it is currently used in multiple clinical trials in patients with metastatic breast cancer as well [127, 193].

The findings discussed above show that the proposed method is able to correctly identify *Raloxifene* and *Ruxolitinib* drugs as very good candidates for most of the BC subtypes. We also observe investigational and experimental drugs in the list for each of the subtypes that could have therapeutic effects on each BC subtype. For example, *PHA-793887* is a potent inhibitor of multiple cyclin-dependent kinases such as CDK2, CDK5 and CDK7, and has been shown to possess the ability to affect the differentiation of melanoma cells. [35, 60]. This drug is currently in a clinical trial phase [4].

In another comparison, Table 3.2 shows the top 30 drugs ranked by their median score across all ten subtypes. As shown in the table, some drugs such as *Palbociclib* and *PHA-793887* demonstrate potential effectiveness across all of the subtypes by being ranked among the top drugs. In contrast, some others such as *Silmitasertib* and *Proglumide* demonstrate potential effectiveness in some of the subtypes, while being less effective in others.

Also, Figures 3.4 and 3.5 show the perturbation scores and drug-disease network of one of the top identified drugs, *Ruxolitinib*, for Subtype 1. *Ruxolitinib*, as mentioned earlier in this paper, is a small-molecule kinase inhibitor that is selective for the Janus Associated Kinases (JAK) 1 and 2, which are responsible for the mediation of cytokine and growth factor signaling, which, in turn, affects the immune function and hematopoiesis [211].

Figure 3.4: Perturbation scores across all genes involved in drug-disease network of top repurposed drug (Ruxolitinib) corresponding to subtype 1. Red bars depict the scores of subtype 1, while green bars depict the scores for the repurposed drug.

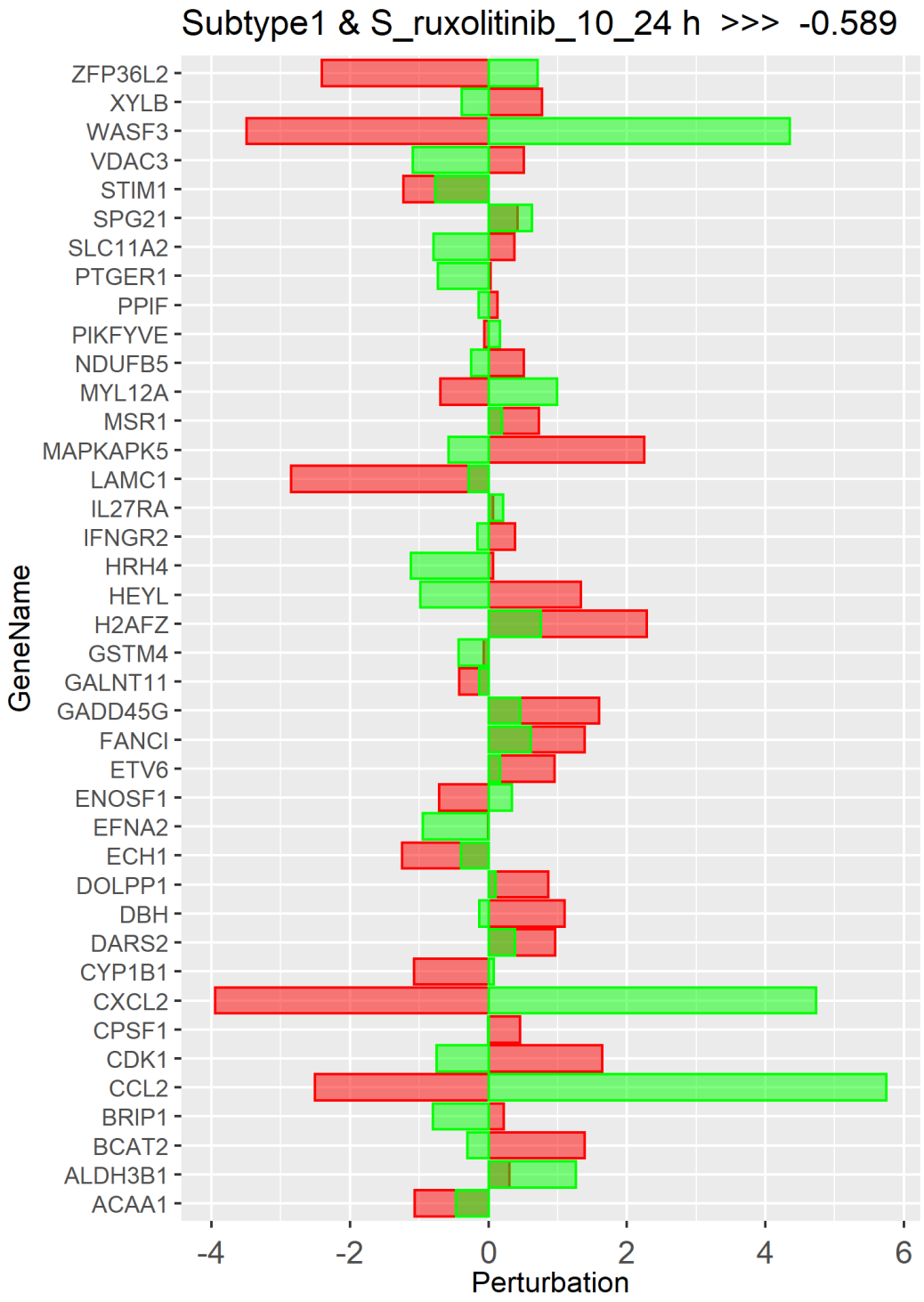
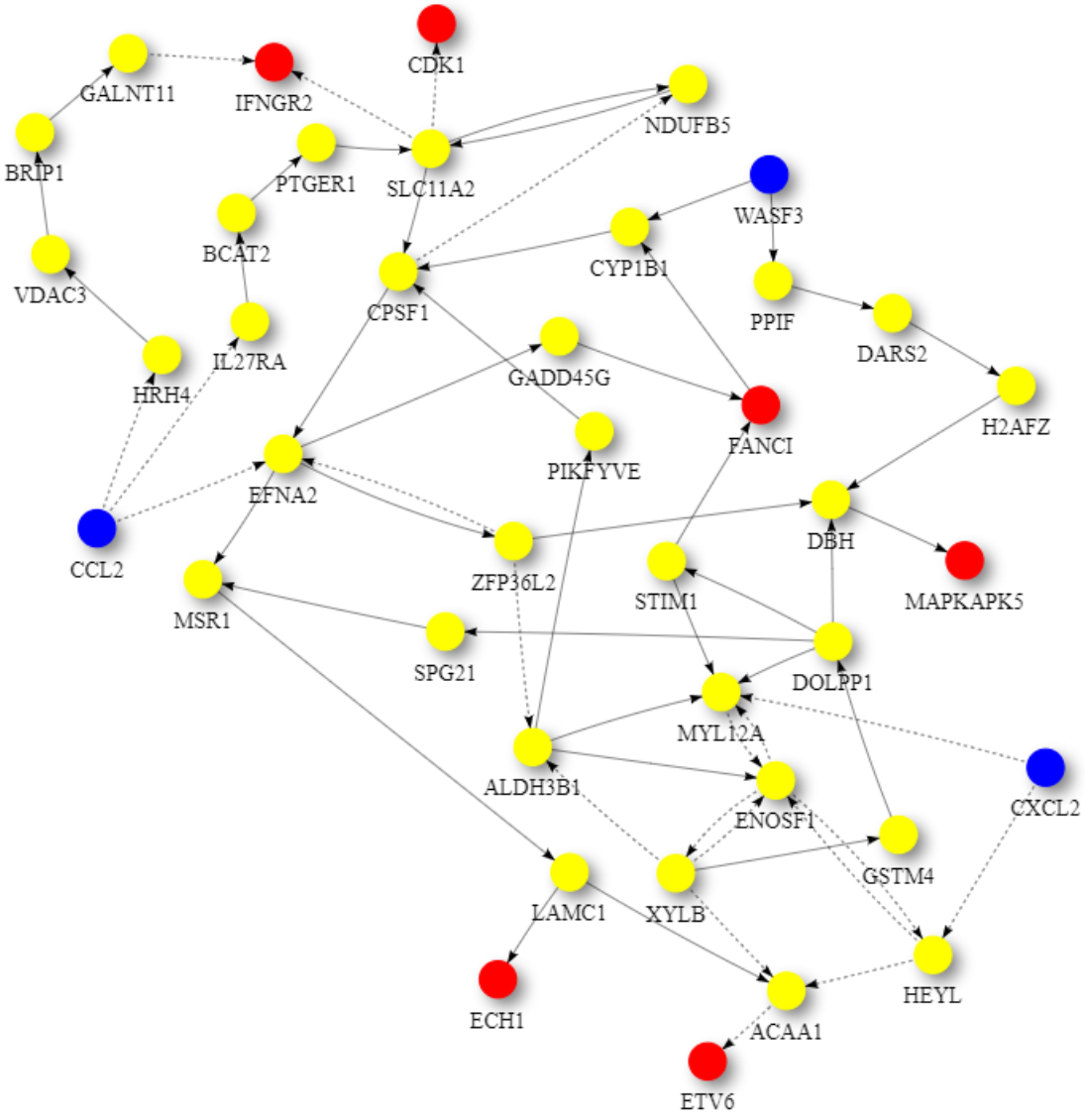


Figure 3.5: Unified Global Human Pathway (UGHP) subnetwork corresponding to the top repurposed drug (Ruxolitinib) and subtype 1. Blue nodes depict Drug related genes, while red nodes depict Subtype 1 related candidate genes involved in this drug-disease pathway.



3.3.2 Paired Drug Repurposing

Figures 3.6 to 3.15 depict the top pairs of drugs with the highest anti-correlation scores with Subtypes 1-10 of breast cancer. Here, we show the pairs of drugs that have a better anti-correlation score than the best single drug for each of these subtypes. Also, we limit the number of pairs to a maximum of 100 top pairs of such drugs, if there is more than 100 pairs with better score than the best single drug.

Moreover, Tables 3.13 to 3.22 show top ten pairs of drugs for each of the ten breast cancer subtypes. Observing these tables, we infer that many of the top ranked pairs of drugs contain at least one individual top ranked drug, though there are some notable exceptions. For example, drugs *TG-100801* and *Phensuximide* are not even among the top 50 repurposed drugs corresponding to Subtype 2 when administered independently with mere correlation scores of -0.57 and -0.55 to subtype 2, respectively. However, when administered together, the correlation between that pair and subtype 2 grows to a noticeable -0.97 range, which places the pair in the second spot among the top repurposed pairs for that subtype. We observe a similar catalyzing effect in combination of *Pregnenolone* and *Bromocriptine* with respect to subtype 4, and combination of *Amikacin* and *Tadalafil* with respect to subtype 5. Also, Figure 3.16 depicts drug-disease network (DDN) of two perturbation agents (Tadalafil and PF-04620110) and subtype 1 of breast cancer. Blue nodes depict drug related genes, while red nodes depict candidate genes related to subtype 1. Also solid arrows depict activating relationship between involved genes, while dotted arrow depicts a suppressing relationship.

Goserelin-Acetate, which is sold under brand name *Zoladex* among others, is as a sex hormone suppression drug approved by FDA intended for use in the treatment of breast and prostate cancer [113]. As shown in Tables 3.4 and 3.14, *Goserelin-Acetate* as a single

Figure 3.6: Top pairs of drugs with highest anti-correlation corresponding to subtype 1.

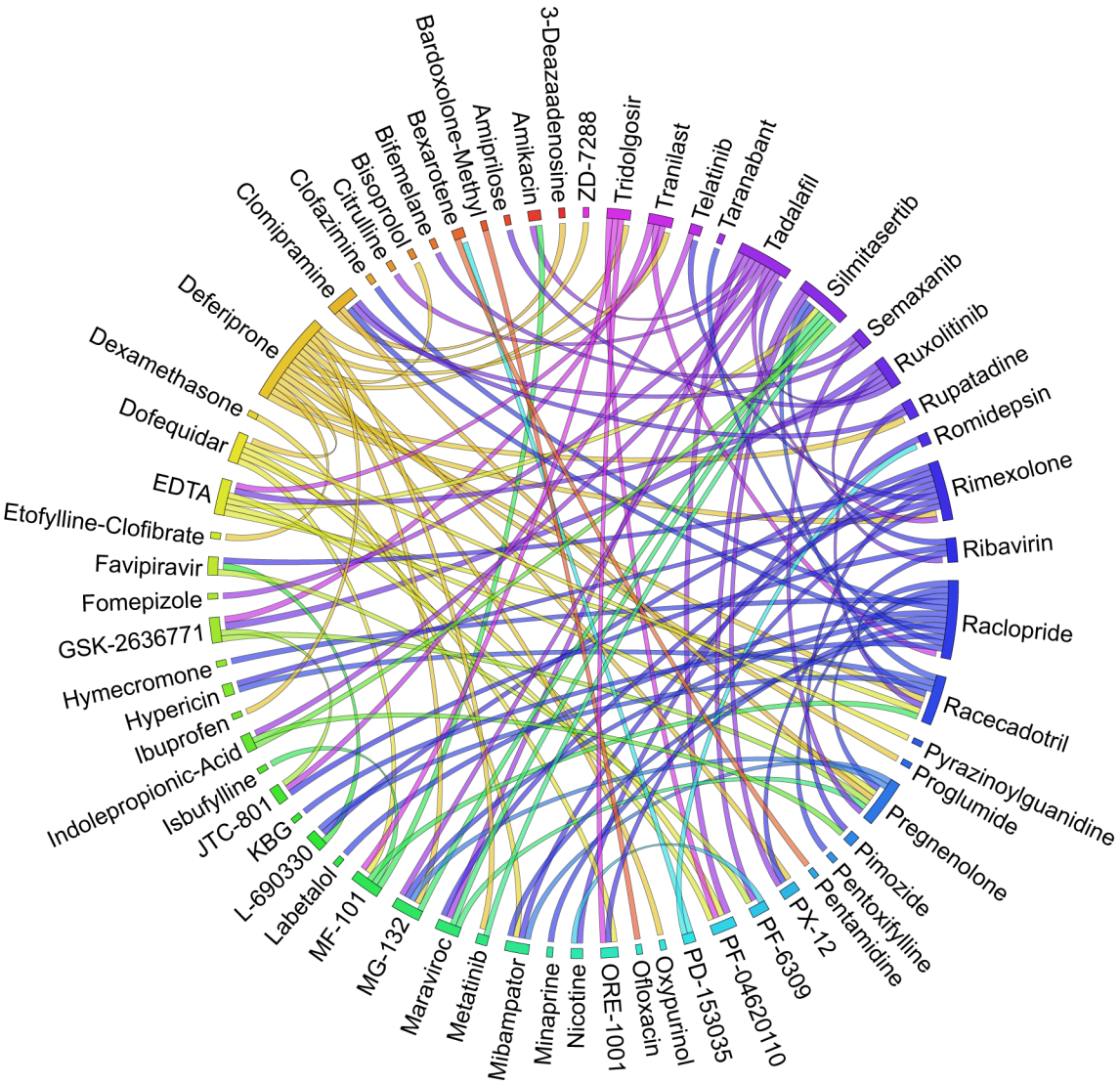


Figure 3.7: Top pairs of drugs with highest anti-correlation corresponding to subtype 2.

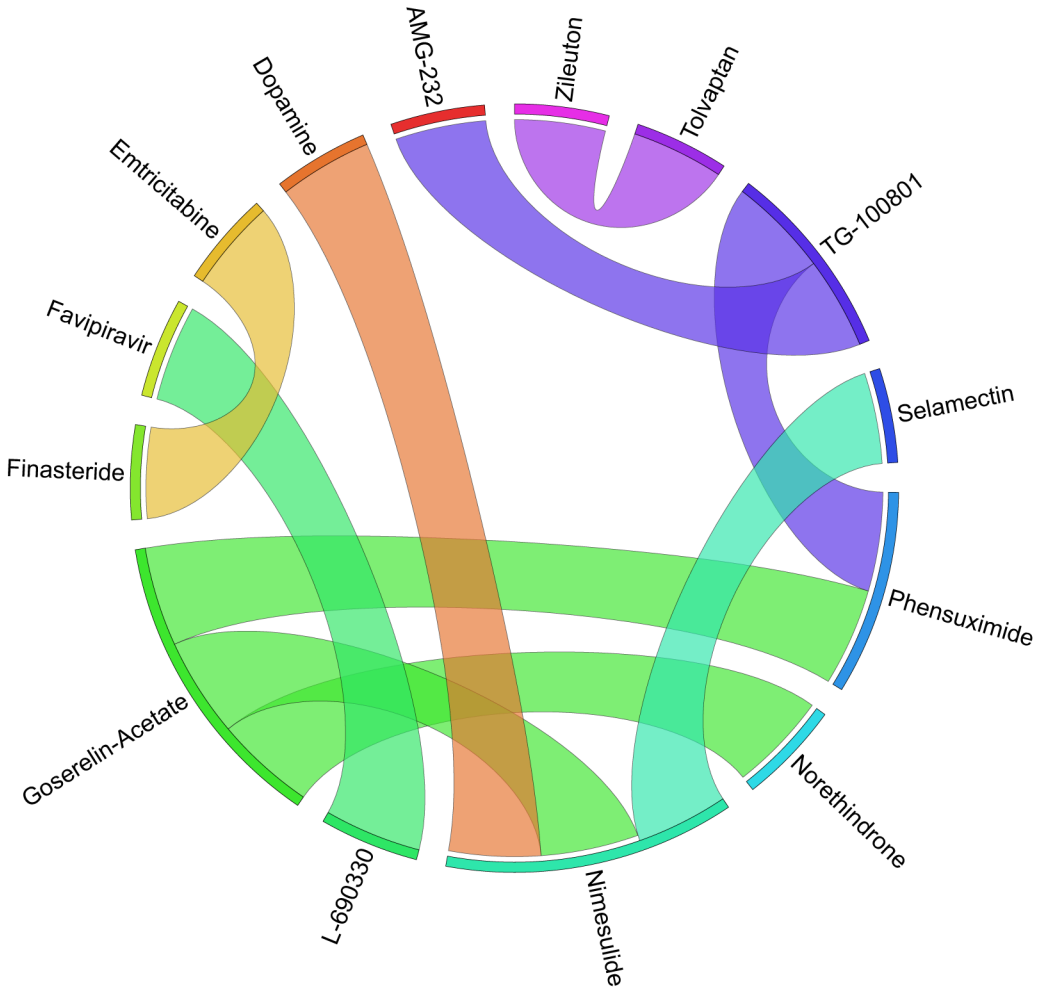


Figure 3.8: Top pairs of drugs with highest anti-correlation corresponding to subtype 3.

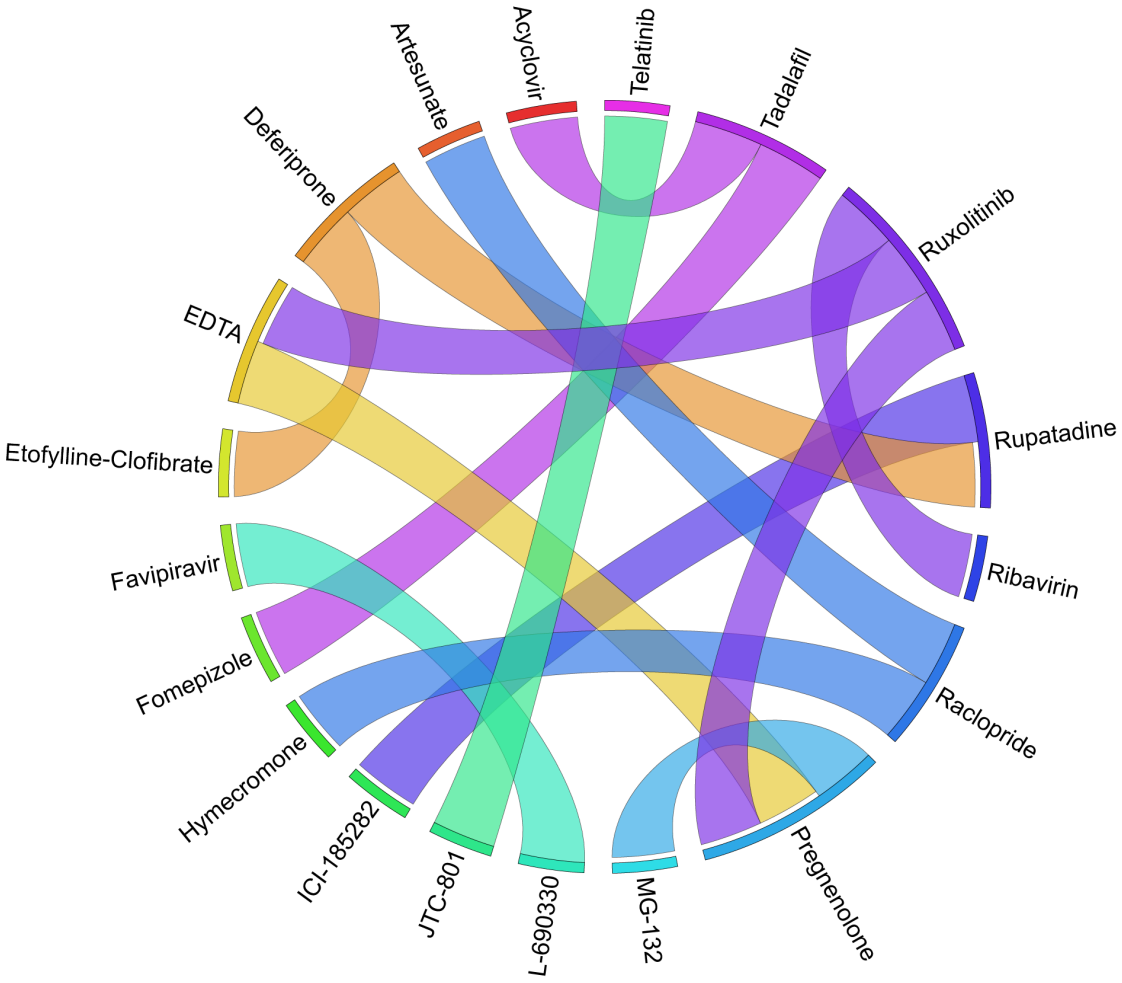


Figure 3.9: Top pairs of drugs with highest anti-correlation corresponding to subtype 4.

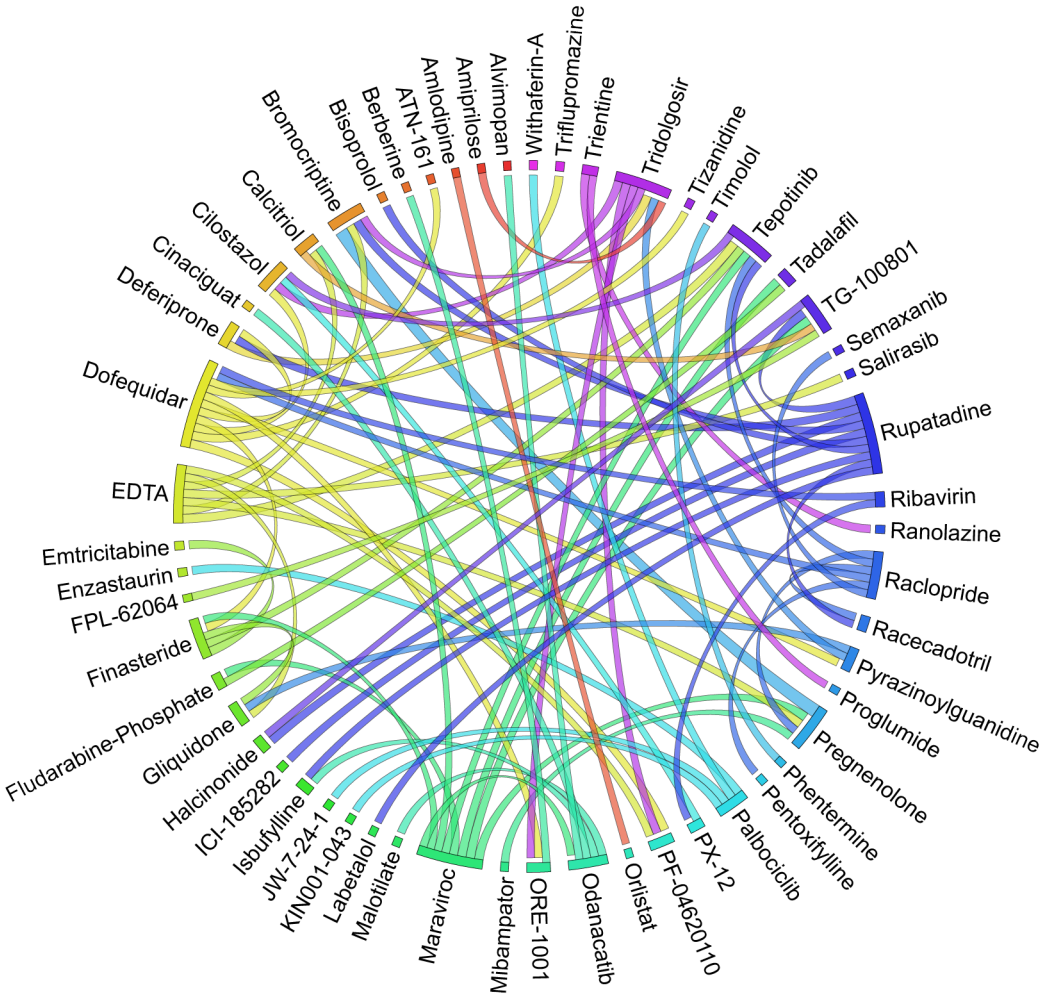


Figure 3.10: Top pairs of drugs with highest anti-correlation corresponding to subtype 5.

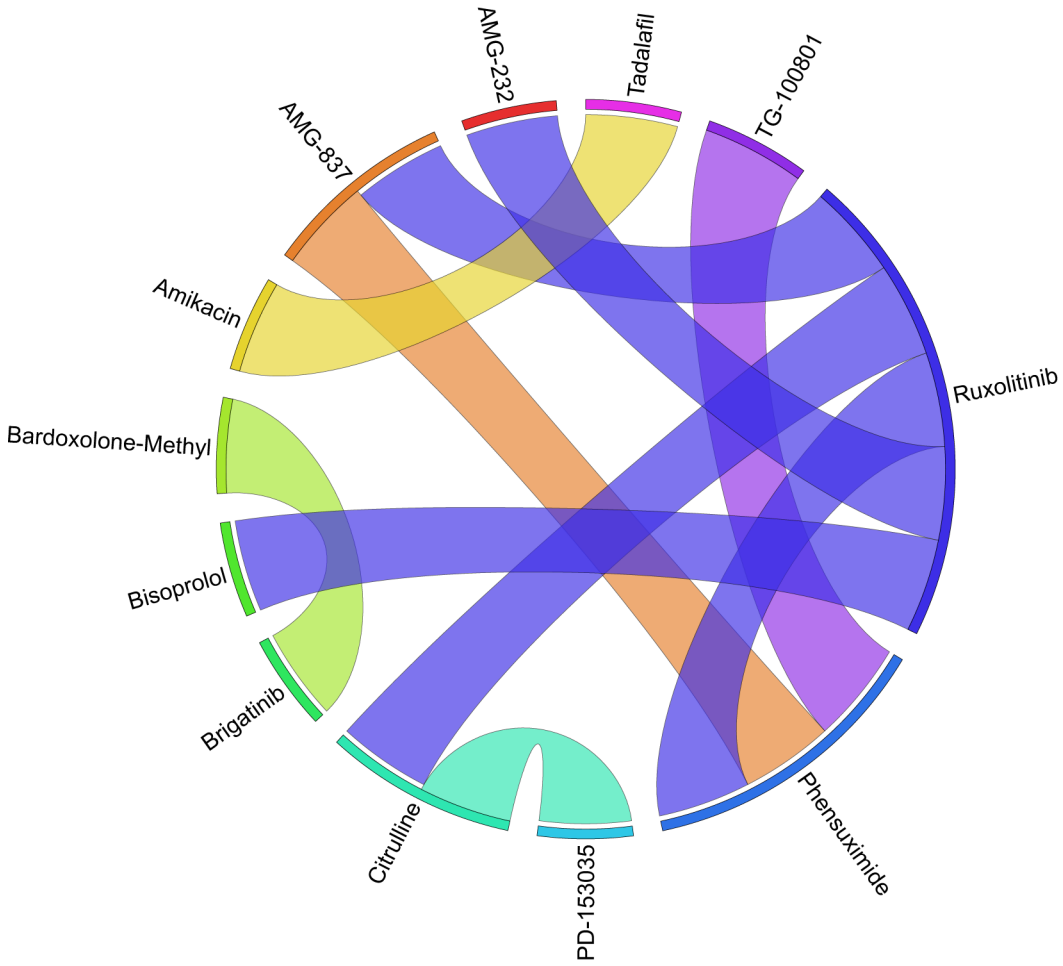


Figure 3.11: Top pairs of drugs with highest anti-correlation corresponding to subtype 6.

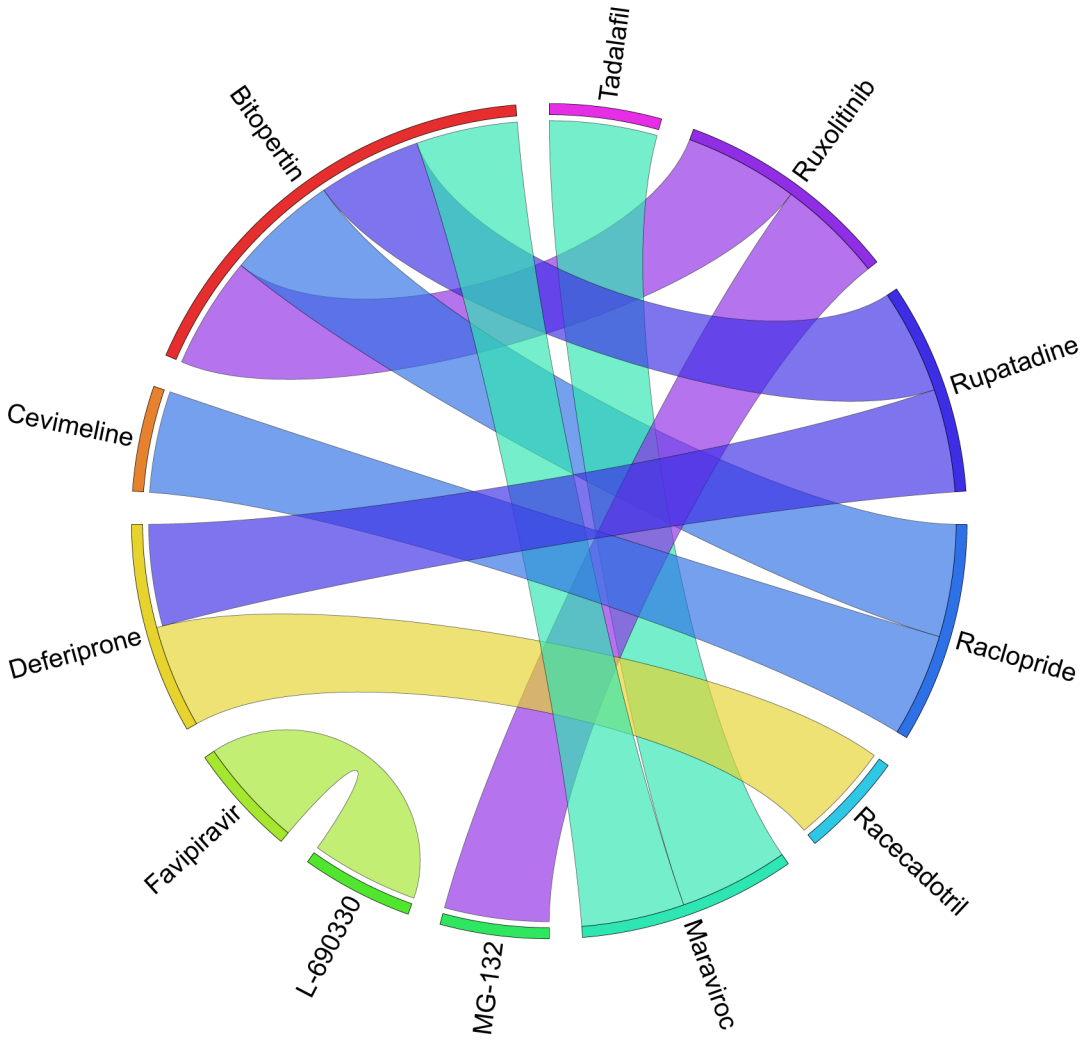


Figure 3.12: Top pairs of drugs with highest anti-correlation corresponding to subtype 7.

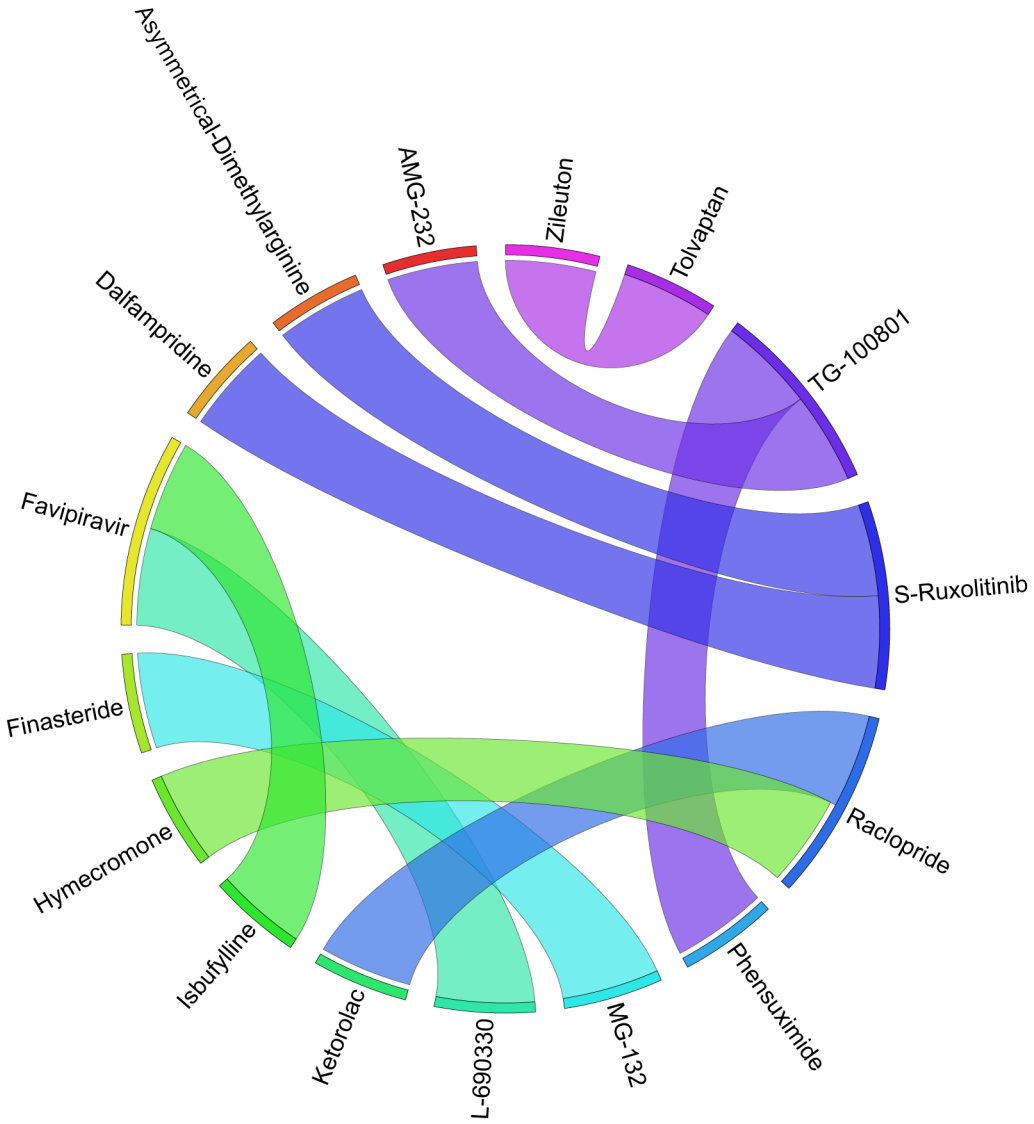


Figure 3.13: Top pairs of drugs with highest anti-correlation corresponding to subtype 8.

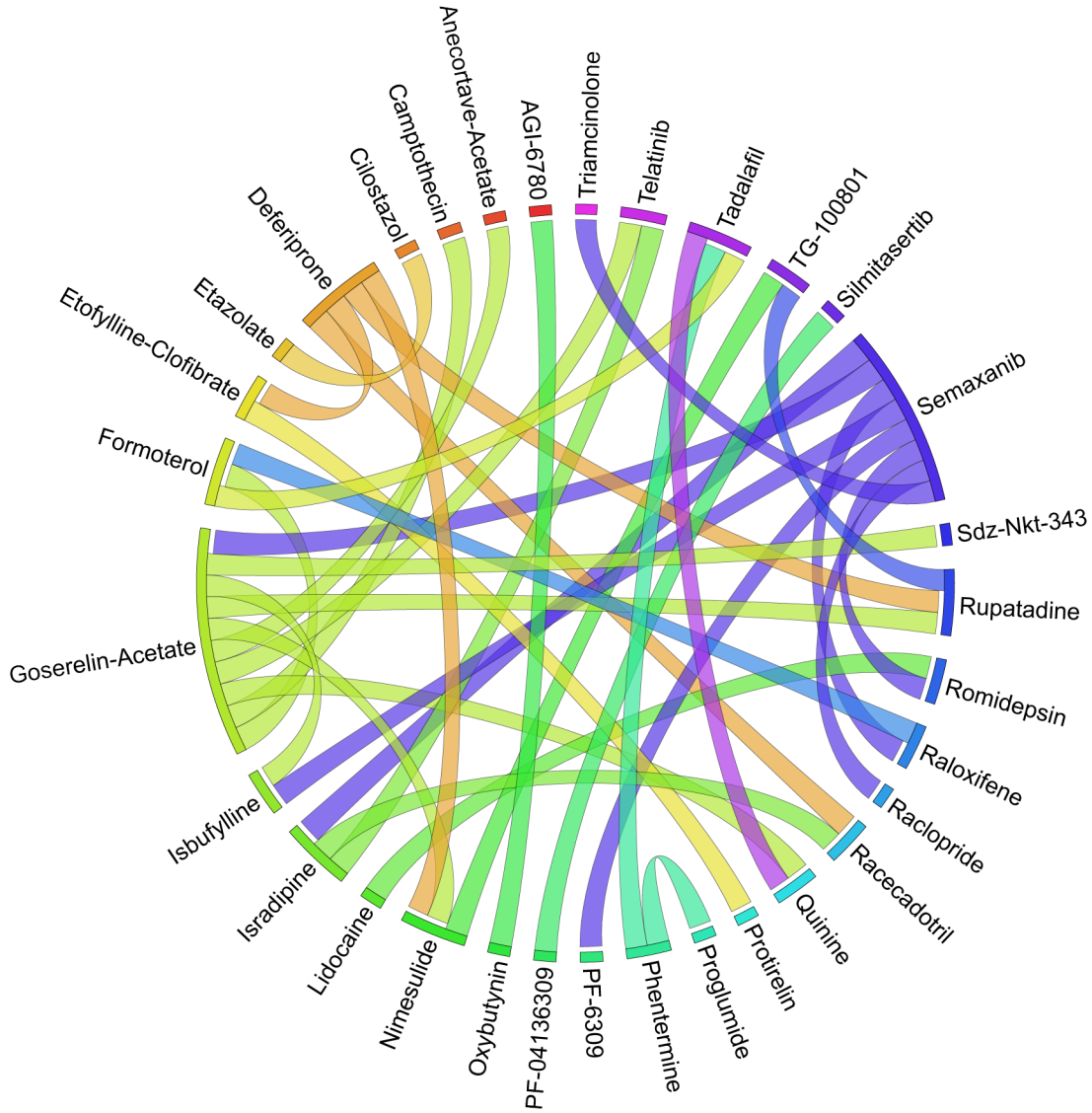


Figure 3.14: Top pairs of drugs with highest anti-correlation corresponding to subtype 9.

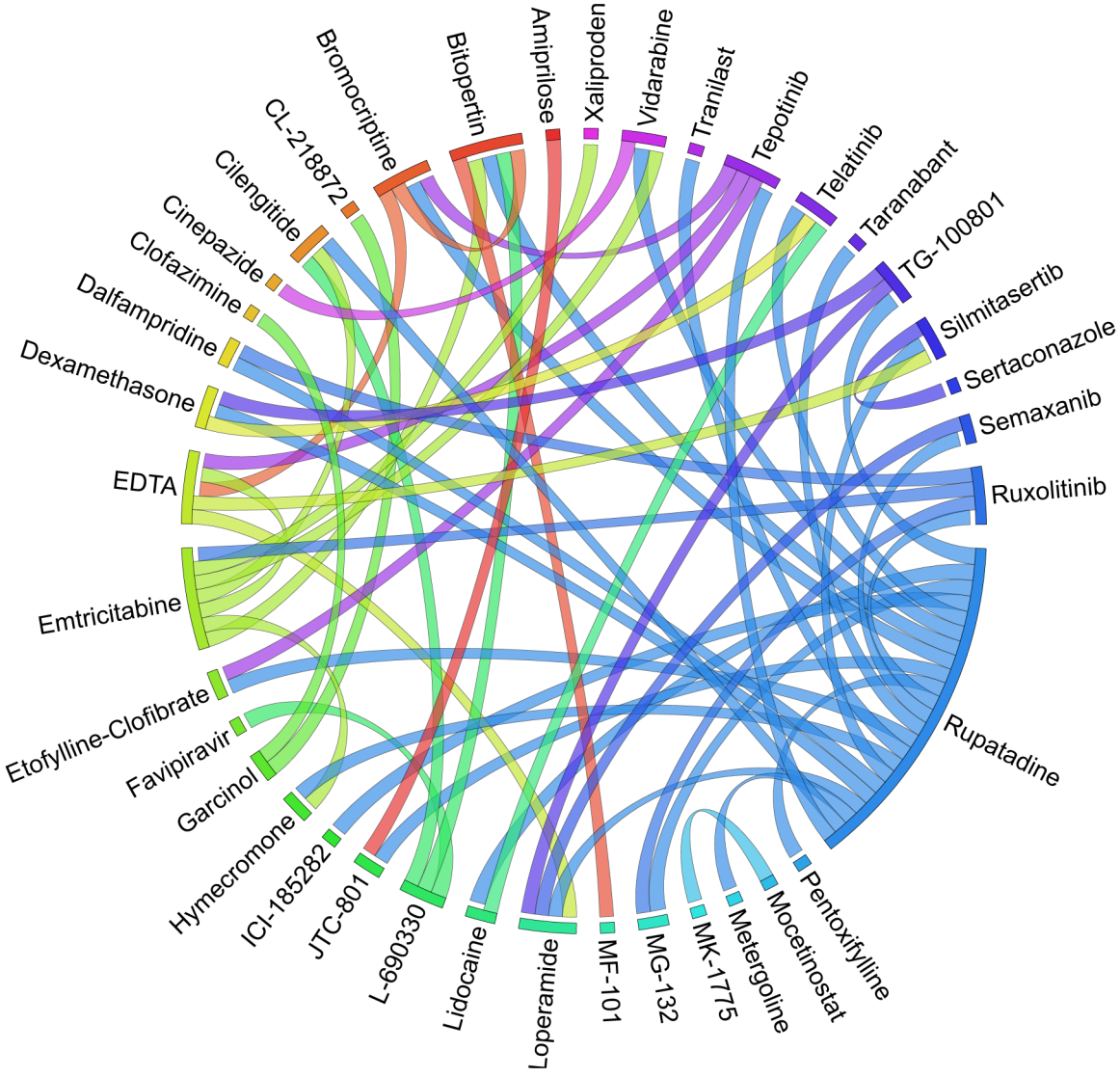


Figure 3.15: Top pairs of drugs with highest anti-correlation corresponding to subtype 10.

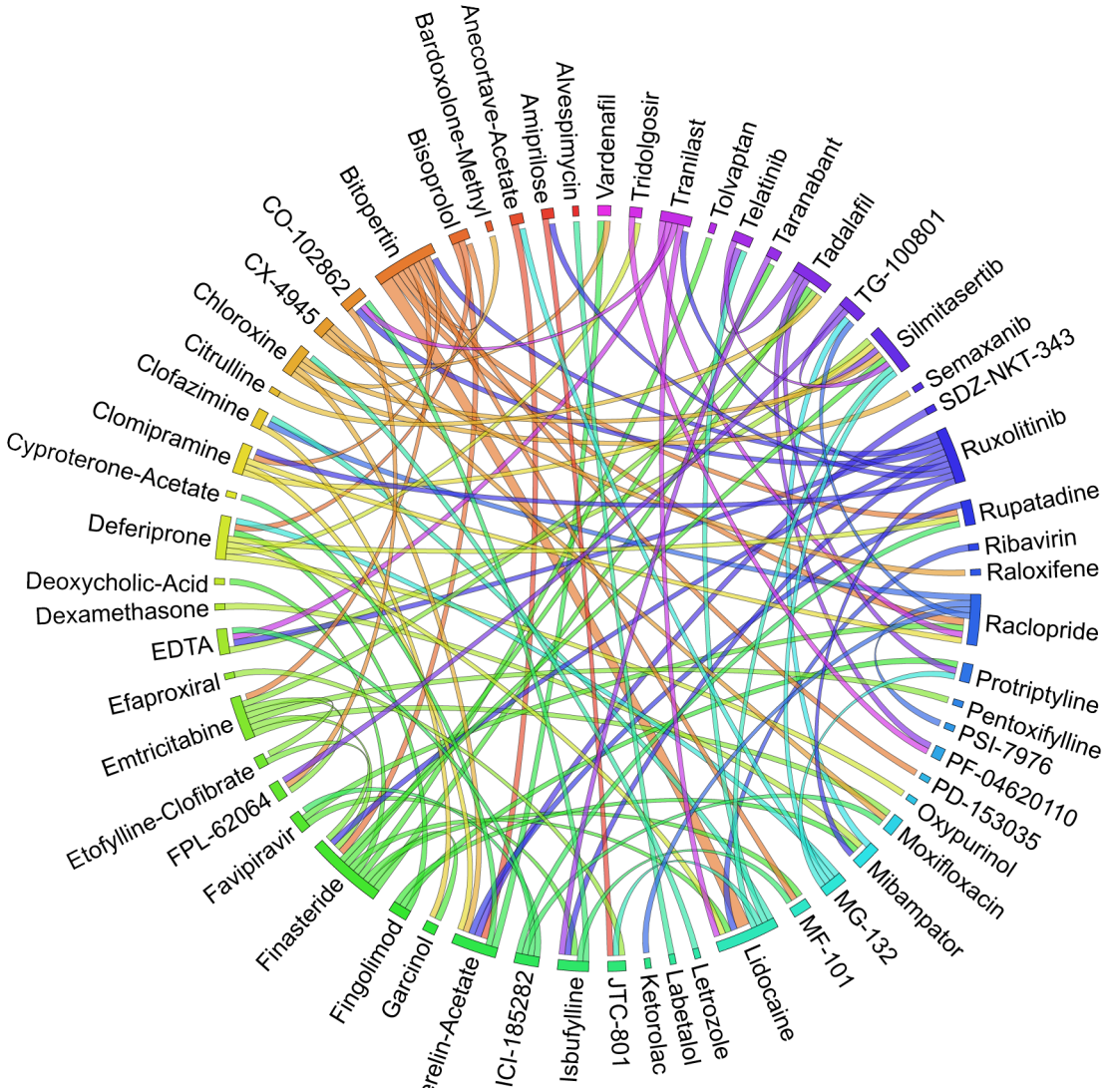


Figure 3.16: Drug-disease network (DDN) of Tadalafil and PF-04620110 drugs with subtype 1. Blue nodes depict drug related genes, while red nodes depict candidate genes related to subtype 1.

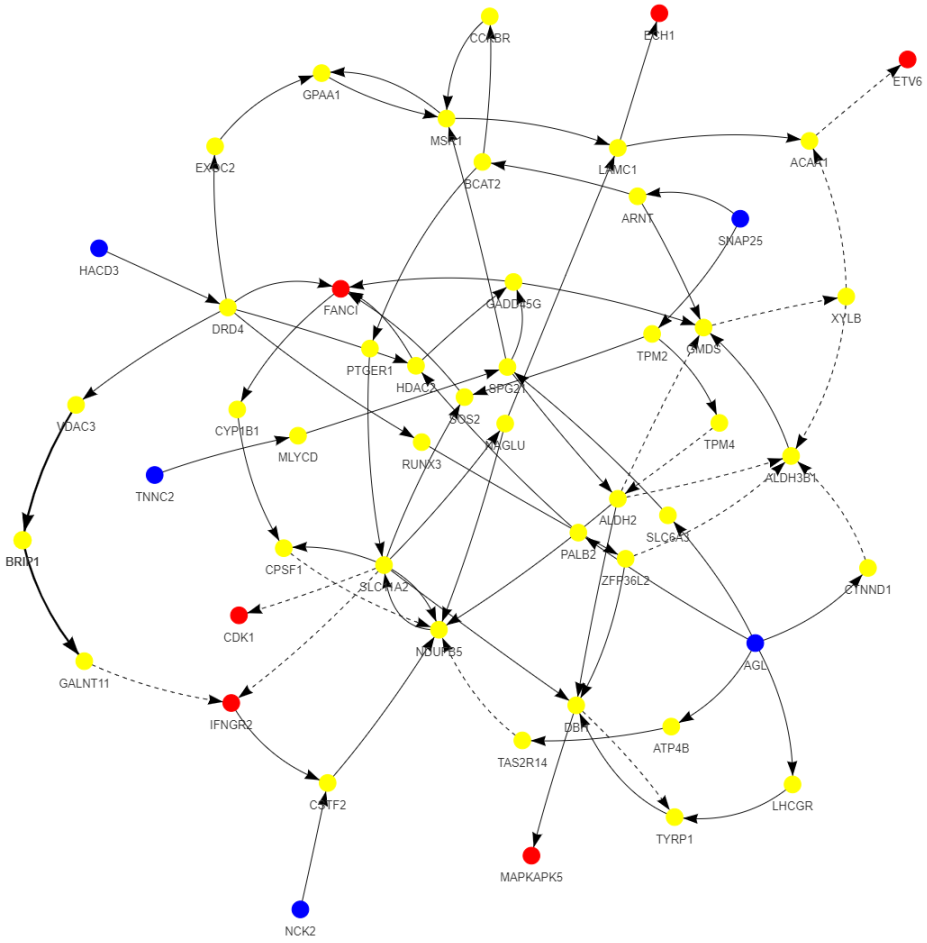


Table 3.13: Top 10 pairs of drugs with their correlation with subtype 1 of breast cancer, considering both combined and individually.

Drug1	Drug2	C. Cor.	Cor. 1	Cor. 2	C. Rank	Rank 1	Rank 2
Ruxolitinib	GSK-2636771	-0.76	-0.59	-0.50	1	1	12
Tadalafil	PF-04620110	-0.70	-0.52	-0.42	2	5	66
Deferiprone	Rimexolone	-0.69	-0.53	-0.51	3	3	6
Deferiprone	Rupatadine	-0.69	-0.53	-0.47	4	3	22
Raclopride	Racecadotril	-0.69	-0.47	-0.45	5	24	35
Tranilast	GSK-2636771	-0.68	-0.52	-0.50	6	4	12
Tadalafil	Amikacin	-0.68	-0.52	-0.41	7	5	71
L-690330	Favipiravir	-0.67	-0.45	-0.44	8	37	44
Deferiprone	Etofylline-Clofibrate	-0.67	-0.53	-0.48	9	3	16
Ruxolitinib	Ribavirin	-0.67	-0.59	-0.56	10	1	2

Table 3.14: Top 10 pairs of drugs with their correlation with subtype 2 of breast cancer, considering both combined and individually.

Drug1	Drug2	C. Cor.	Cor. 1	Cor. 2	C. Rank	Rank 1	Rank 2
Goserelin-Acetate	Norethindrone	-0.97	-0.80	-0.54	1	2	75
TG-100801	Phensuximide	-0.97	-0.57	-0.55	2	50	64
Goserelin-acetate	Phensuximide	-0.96	-0.80	-0.55	3	2	64
Goserelin-acetate	Nimesulide	-0.96	-0.80	-0.54	4	2	66
L-690330	Favipiravir	-0.94	-0.71	-0.53	5	7	83
Dopamine	Nimesulide	-0.90	-0.57	-0.54	6	51	66
TG-100801	AMG-232	-0.89	-0.57	-0.56	7	50	59
Tolvaptan	Zileuton	-0.89	-0.67	-0.53	8	9	87
Emtricitabine	Finasteride	-0.89	-0.60	-0.57	9	33	53
Nimesulide	Selamectin	-0.89	-0.54	-0.54	10	66	72

Table 3.15: Top 10 pairs of drugs with their correlation with subtype 3 of breast cancer, considering both combined and individually.

Drug1	Drug2	C. Cor.	Cor. 1	Cor. 2	C. Rank	Rank 1	Rank 2
Tadalafil	Acyclovir	-0.79	-0.61	-0.45	1	4	73
Tadalafil	Fomepizole	-0.78	-0.61	-0.43	2	4	100
Rupatadine	ICI-185282	-0.77	-0.58	-0.52	3	6	21
Deferiprone	Etofylline-Clofibrate	-0.76	-0.62	-0.57	4	3	9
Deferiprone	Rupatadine	-0.75	-0.62	-0.58	5	3	6
Ruxolitinib	Ribavirin	-0.74	-0.68	-0.57	6	2	7
L-690330	Favipiravir	-0.74	-0.49	-0.47	7	38	54
Raclopride	Hymecromone	-0.74	-0.54	-0.48	8	13	48
EDTA	pregnenolone	-0.74	-0.56	-0.51	9	10	28
Raclopride	Artesunate	-0.74	-0.54	-0.47	10	13	59

Table 3.16: Top 10 pairs of drugs with their correlation with subtype 4 of breast cancer, considering both combined and individually.

Drug1	Drug2	C. Cor.	Cor. 1	Cor. 2	C. Rank	Rank 1	Rank 2
Pregnenolone	Bromocriptine	-0.99	-0.42	-0.40	1	24	38
EDTA	EMD-1214063	-0.69	-0.46	-0.35	2	8	81
Maraviroc	Pregnenolone	-0.68	-0.46	-0.42	3	9	24
Rupatadine	TG-100801	-0.68	-0.58	-0.37	4	1	60
Maraviroc	TG-100801	-0.67	-0.46	-0.37	5	9	60
Finasteride	Emtricitabine	-0.66	-0.43	-0.37	6	19	68
Maraviroc	Tadalafil	-0.66	-0.46	-0.35	7	9	89
Dofequidar	Triflupromazine	-0.66	-0.44	-0.36	8	14	78
Dofequidar	PF-04620110	-0.66	-0.44	-0.38	9	14	55
Dofequidar	Finasteride	-0.65	-0.44	-0.43	10	14	19

Table 3.17: Top 10 pairs of drugs with their correlation with subtype 5 of breast cancer, considering both combined and individually.

Drug1	Drug2	C. Cor.	Cor. 1	Cor. 2	C. Rank	Rank 1	Rank 2
TG-100801	Phensuximide	-0.97	-0.72	-0.63	1	6	28
Ruxolitinib	Phensuximide	-0.89	-0.93	-0.63	2	1	28
Ruxolitinib	AMG-837	-0.89	-0.93	-0.83	3	1	2
AMG-837	Phensuximide	-0.89	-0.83	-0.63	4	2	28
Ruxolitinib	Citrulline	-0.89	-0.93	-0.82	5	1	3
Amikacin	Tadalafil	-0.88	-0.59	-0.57	6	47	65
L-citrulline	PD-153035	-0.88	-0.82	-0.62	7	3	35
Ruxolitinib	AMG-232	-0.88	-0.93	-0.56	8	1	73
Bardoxolone-methyl	AP-26113	-0.88	-0.63	-0.58	9	33	52
Ruxolitinib	Bisoprolol	-0.88	-0.93	-0.71	10	1	8

Table 3.18: Top 10 pairs of drugs with their correlation with subtype 6 of breast cancer, considering both combined and individually.

Drug1	Drug2	C. Cor.	Cor. 1	Cor. 2	C. Rank	Rank 1	Rank 2
Ruxolitinib	Bitopertin	-0.73	-0.67	-0.40	1	1	44
Raclopride	Bitopertin	-0.72	-0.41	-0.40	2	36	44
Maraviroc	Tadalafil	-0.70	-0.49	-0.43	3	3	23
Favipiravir	L-690330	-0.70	-0.45	-0.42	4	13	32
Ruxolitinib	MG-132	-0.68	-0.67	-0.38	5	1	55
Rupatadine	Bitopertin	-0.67	-0.50	-0.40	6	2	44
Maraviroc	Bitopertin	-0.66	-0.49	-0.40	7	3	44
Raclopride	Cevimeline	-0.66	-0.41	-0.36	8	36	74
Deferiprone	Racecadotril	-0.66	-0.49	-0.48	9	4	8
Rupatadine	Deferiprone	-0.66	-0.50	-0.49	10	2	4

Table 3.19: Top 10 pairs of drugs with their correlation with subtype 7 of breast cancer, considering both combined and individually.

Drug1	Drug2	C. Cor.	Cor. 1	Cor. 2	C. Rank	Rank 1	Rank 2
L-690330	Favipiravir	-0.93	-0.69	-0.54	1	12	87
MG-132	Finasteride	-0.92	-0.66	-0.55	2	18	83
TG-100801	Phensuximide	-0.91	-0.69	-0.61	3	11	37
Raclopride	Ketorolac	-0.89	-0.65	-0.57	4	19	57
Hymecromone	Raclopride	-0.88	-0.66	-0.65	5	16	19
Ruxolitinib	ADMA	-0.87	-0.83	-0.57	6	2	56
Ruxolitinib	Dalfampridine	-0.87	-0.83	-0.63	7	2	24
TG-100801	AMG-232	-0.87	-0.69	-0.59	8	11	51
Tolvaptan	Zileuton	-0.87	-0.64	-0.59	9	22	49
Isbufylline	Favipiravir	-0.86	-0.71	-0.54	10	7	87

Table 3.20: Top 10 pairs of drugs with their correlation with subtype 8 of breast cancer, considering both combined and individually.

Drug1	Drug2	C. Cor.	Cor. 1	Cor. 2	C. Rank	Rank 1	Rank 2
Emaxanib	Isradipine	-0.71	-0.60	-0.42	1	1	34
Semaxanib	Goserelin-Acetate	-0.68	-0.60	-0.55	2	1	3
Goserelin-Acetate	Formoterol	-0.68	-0.55	-0.43	3	3	31
Goserelin-Acetate	Camptothecin	-0.66	-0.55	-0.38	4	3	72
Deferiprone	Racecadotril	-0.66	-0.49	-0.40	5	6	56
Deferiprone	Etofylline-Clofibrate	-0.65	-0.49	-0.47	6	6	15
Semaxanib	Raloxifene	-0.65	-0.60	-0.54	7	1	4
Lidocaine	Romidepsin	-0.64	-0.46	-0.41	8	17	49
Goserelin-Acetate	Telatinib	-0.64	-0.55	-0.41	9	3	44
Isradipine	Telatinib	-0.64	-0.42	-0.41	10	34	44

Table 3.21: Top 10 pairs of drugs with their correlation with subtype 9 of breast cancer, considering both combined and individually.

Drug1	Drug2	C. Cor.	Cor. 1	Cor. 2	C. Rank	Rank 1	Rank 2
Rupatadine	TG-100801	-0.85	-0.72	-0.61	1	1	8
L-690330	Favipiravir	-0.83	-0.65	-0.48	2	5	76
Rupatadine	Lidocaine	-0.82	-0.72	-0.51	3	1	51
Rupatadine	MG-132	-0.82	-0.72	-0.61	4	1	9
Rupatadine	ICI-185,282	-0.81	-0.72	-0.52	5	1	35
Rupatadine	Telatinib	-0.79	-0.72	-0.47	6	1	82
Rupatadine	Vidarabine	-0.79	-0.72	-0.59	7	1	12
Garcinol	CL-218872	-0.78	-0.56	-0.49	8	24	60
Rupatadine	Bitopertin	-0.77	-0.72	-0.48	9	1	71
Emtricitabine	Bitopertin	-0.77	-0.65	-0.48	10	4	71

Table 3.22: Top 10 pairs of drugs with their correlation with subtype 10 of breast cancer, considering both combined and individually.

Drug1	Drug2	C. Cor.	Cor. 1	Cor. 2	C. Rank	Rank 1	Rank 2
Tadalafil	Telatinib	-0.86	-0.55	-0.53	1	47	61
Bitopertin	Lidocaine	-0.85	-0.66	-0.57	2	2	35
Anecortave-Acetate	Goserelin-Acetate	-0.84	-0.56	-0.52	3	38	64
Finasteride	Tadalafil	-0.84	-0.60	-0.55	4	16	47
Bitopertin	Raclopride	-0.83	-0.66	-0.57	5	2	36
SDZ-NKT-343	Goserelin-acetate	-0.83	-0.59	-0.52	6	24	64
Goserelin-Acetate	Vardenafil	-0.81	-0.52	-0.48	7	64	96
Favipiravir	Isbufylline	-0.81	-0.59	-0.58	8	25	28
Bisoprolol	Finasteride	-0.80	-0.61	-0.60	9	12	16
Finasteride	Garcinol	-0.80	-0.60	-0.55	10	16	48

drug produces an correlation score of -0.8 with respect to subtype 2 of BC, which places it in the top spot among the single drugs for this subtype. However, if combined with either *Norethindrone*, *Phensuximide* or *Nimesulide*, the correlation score decreases to almost -0.97. This means that combining either of the aforementioned drugs with *Goserelin-Acetate* can result in a more effective therapeutic drug for this BC subtype.

3.3.3 Triple Negative Breast Cancer Subtype

Table 3.23 shows the identified driver genes corresponding to the TN group. Moreover, Table 3.24 shows the top 20 single repurposed drugs for the triple negative breast cancer (TNBC) subtype. As shown in the table, *Ruxolitinib* is by far the most negatively correlated drug for TNBC subtype and can be investigated further for its effectiveness on this particular type of breast cancer [105, 117, 193].

Moreover, Table 3.25 and Figure 3.17 show the top 10 pairs of repurposed drugs and their corresponding scores with respect to TNBC subtype. We observe that despite being the tenth repurposed single drug, *Bromocriptine* is managed to become one of the most effective repurposed drugs when paired with *Isradipin*, *Emtricitabine* and *Etofylline-Clofibrate*. Although *Bromocriptine* has been suggested in earlier studies as a potential repurposed drug in cancer therapy [180], these new combinations have not seem to be evaluated so far for breast cancer treatment, which can be investigated further both computationally and clinically. Another interesting observation from Tables 3.24, 3.25 is that there are only four pairs of drugs with anti-correlation scores better than *Ruxolitinib* as the best single drug identified for TNBC subtype.

Based on these observations, we infer that from both single drug and paired drug experiments, there are some promising drugs that can be repurposed either individually or

Figure 3.17: Top pairs of drugs with highest anti-correlation corresponding to TNBC subtype.

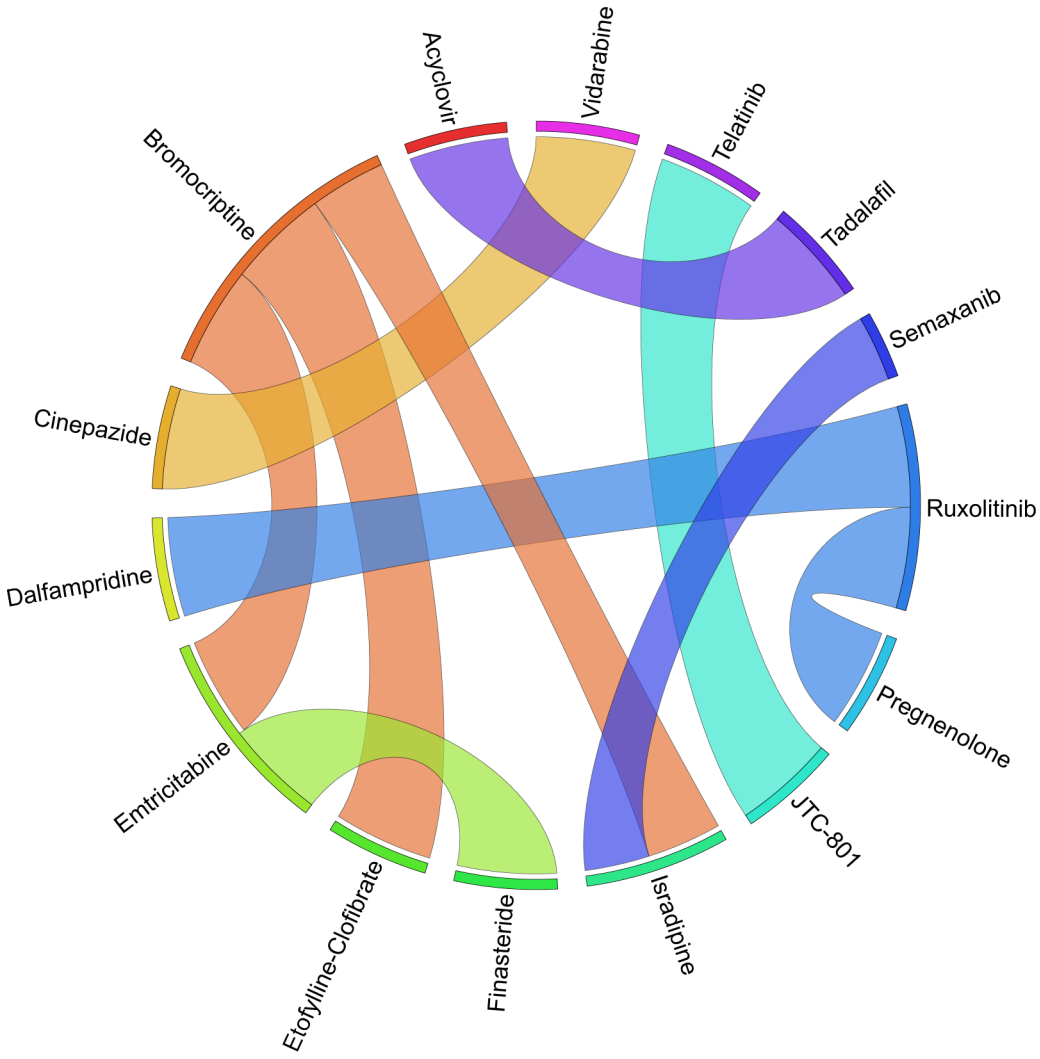


Table 3.23: Identified driver genes associated with triple negative breast cancer subtype.

ACRV1	ADCY9	AKT2	ALOX12B	APTX	C17orf100
C1QBP	CALM2	CARD18	CLPSL1	DARS2	DEFB136
DHX33	EEF1E1	EHHADH	ELAC2	EPPIN	FAXDC2
FOXO3	GAB2	GAL3ST3	GFER	GJA10	GUCA2A
HACE1	HP	HTR3D	IFNA21	KLHDC8A	LDOC1L
LINC00628	LINC00919	MFAP4	MPRIP	MRGPRF	MRPL13
MUC21	OR1S1	OR3A1	PLEKHA8	PMCHL1	PNPLA3
POGK	POLR3G	PRPH2	RFPL4B	SDHC	SIRT5
SLC1A4	SLC25A11	SLC35F2	SLFN12L	SNX29	SRPK1
STOML2	SUV39H2	TAS2R20	TATDN1	THOC1	TOMM22
TRIM72	TRMT12	TWIST2	TXNDC17	URB2	VDAC3
WFDC10A	ZC3H7B	ZNF23			

Table 3.24: Top 20 drugs along with their ranking corresponding to TNBC subtype.

Rank	Drug Name	Pearson Correlation	Dosage	Time
1	Ruxolitinib	-0.765	10	24 h
2	Raloxifene	-0.652	0.04	24 h
3	PF-04217903	-0.649	0.04	24 h
4	PD-173074	-0.603	10	24 h
5	TG-100801	-0.591	0.04	24 h
6	Dexamethasone	-0.589	0.04	24 h
7	Semaxanib	-0.588	0.04	24 h
8	Rupatadine	-0.575	0.04	24 h
9	Tranilast	-0.574	0.04	24 h
10	Bromocriptine	-0.565	0.04	24 h
11	ICI-185282	-0.562	0.04	24 h
12	Hymecromone	-0.558	0.04	24 h
13	Emtricitabine	-0.556	10	24 h
14	Phentermine	-0.554	0.04	24 h
15	Fludarabine-Phosphate	-0.548	10	24 h
16	JTC-801	-0.547	0.04	24 h
17	AMG-232	-0.541	10	24 h
18	Lidocaine	-0.525	0.04	24 h
19	Amiprilose	-0.520	0.04	24 h
20	Labetalol	-0.515	0.04	24 h

Table 3.25: Top 10 pairs of drugs with their correlation with respect to TN breast cancer subtype, both when they are combined and individually.

Drug1	Drug2	C. Cor.	Cor. 1	Cor. 2	C. Rank	Rank 1	Rank 2
Bromocriptine	Isradipine	-0.85	-0.57	-0.47	1	10	51
Cinepazide	Vidarabine	-0.80	-0.45	-0.45	2	74	75
Bromocriptine	Emtricitabine	-0.79	-0.57	-0.56	3	10	13
Bromocriptine	Etofylline-Clofibrate	-0.77	-0.57	-0.51	4	10	22
Emtricitabine	Finasteride	-0.76	-0.56	-0.50	5	13	26
JTC-801	Telatinib	-0.75	-0.55	-0.47	6	16	47
Ruxolitinib	Dalfampridine	-0.74	-0.77	-0.50	7	1	31
Semaxanib	Isradipine	-0.74	-0.59	-0.47	8	7	51
Ruxolitinib	Pregnenolone	-0.74	-0.77	-0.48	9	1	40
Tadalafil	Acyclovir	-0.74	-0.47	-0.46	10	43	58

in combination with another drug (as a pair) with potential therapeutic effects for each of the ten breast cancer subtypes. Some of these drugs such as *Ruxolitinib* have a high anti-correlation score for most of the subtypes, while some of the drugs such as *Maraviroc* [158] seem to be more effective on a particular subtype rather than on others. The fact that the top single drug, *Ruxolitinib*, is currently in multiple clinical trials in patients with metastatic breast cancer [127, 193] shows that the proposed method is able to computationally predict the potential therapeutic effect of this drug on multiple breast cancer subtypes, as well as on TNBC subtype. Indeed, further wet lab analysis is needed to determine the therapeutic level of identified drugs on each breast cancer subtype.

Chapter 4

Conclusion and Future Work

4.1 Conclusion

In Chapter 2 We have introduced a novel framework for identifying NBs related to each of the ten breast cancer subtypes. In the proposed framework, we are :

- Using CNA/CNV information along with GE data to determine a set of candidate genes for each breast cancer subtype.
- Using identified seeds from the previous step to find the differential NBs of a given subtype with the candidate genes already generated for the subtype.
- Training a random forest classifier for each subtype using biomarkers in NB of the corresponding BC subtype to measure the performance of each NB in separating one BC subtype from the rest using different performance measures.

Our results show that NBs can separate one subtype from others with very high degree of accuracy. This may provide great utility in properly stratifying patients for treatment.

Moreover, the obtained NBs may also allow breast cancer researchers to gain insight into the mechanisms driving different breast cancer subtypes.

Also, using the identified candidate genes in Chapters 2 and 3, we proposed a network-based computational drug repurposing framework where we are :

- Creating a drug-disease network for each drug and BC subtype where drug genes obtained from LINCS database and the disease genes coming from previously identified candidate genes.
- Creating a universal pathway network by combining all available pathways in KEGG database and super-imposing those drug and disease genes on the pathway network and then finding all the shortest paths between each drug gene to each disease gene.
- Calculating the correlation of all the genes in the that network when we use drug induced gene expression versus using disease induced gene expression data.
- Finding those drugs that have a highly negative correlation with each subtype as a potential drug with therapeutic effects.
- Extending this analysis to use a combination of two drugs and measuring the effect of each combination on the disease.

Some of the top identified drugs are either known (breast) cancer drugs or in different trial phases to be repurposed for breast or other types of cancer, while some of the identified single or paired drugs have not been used for breast cancer treatment yet, which provides opportunity for further clinical experiments and trials. Using genomic and transcriptomic data as well as both copy number variations and copy number aberrations would help the initial process to identify the driver genes more effectively and hence finding the final set of repourposed drugs that can be highly effective for treatment of other types of cancer.

4.2 Future Work

Some of the potential future work related to Chapter 2 is as follows:

- Taking into the account other types of biological information such as pathway data could help finding better NBs with ability of including that information during network building process.

- While some of the candidate genes identified here are well known, there are some other candidate genes that have not been explored generally in cancer and specifically in breast cancer research. Some of those genes might act as one of the central hubs in the network, which makes their impact even more prominent. Further research is needed on those genes to understand the mechanism of action and how can they be activated/inhibited.

Moreover, Some of the potential future work related to Chapter 3 is as follows:

- The proposed drug repurposing framework has the potential of identifying a combination of more than two drugs, which could help identifying new and enhanced sets of drugs for various types of cancer as well as other types of diseases.

- The current framework can be extended to leverage more complex and nonlinear combinations of drugs in order to find the most suitable sets of drugs for each disease.

- Since for the simplicity of the process we excluded pairs of drugs with known drug-drug interference, potential drug-drug interactions have not been considered yet. Thus, another possible future work could be extending the framework to include such pairs of drugs with known interactions and their effect on the disease.

Bibliography

[1] Drugs approved for breast cancer. https://www.cancer.gov/about-cancer/treatment/drugs/breast, Last accessed on 2020-08-15.

[2] Intogen's mutational breast cancer driver genes. https://www.intogen.org/search?cancer=BRCA, Last accessed on 2019-07-05.

[3] Palbociclib and tadalafil interactions.

[4] Pha-793887 — drug bank. https://go.drugbank.com/drugs/DB12686. Accessed: 2021-02-20.

[5] Biological pathways fact sheet, 2022.

[6] Noora Abdulaziz and Yasser Fakri Mustafa. Anticancer properties of hymecromone-derived compounds: A review. *International Journal of Pharmaceutical Research*, 13(1), 2021.

[7] Christopher P Adams and Van V Brantner. Estimating the cost of new drug development: is it really $802 million? *Health Affairs*, 25(2):420–428, 2006.

[8] Felipe Ades, Dimitrios Zardavas, Ivana Bozovic-Spasojevic, Lina Pugliano, Debora Fumagalli, Evandro De Azambuja, Giuseppe Viale, Christos Sotiriou, and Martine

Piccart. Luminal b breast cancer: molecular characterization, clinical management, and future perspectives. *Journal of Clinical Oncology*, 32(25):2794–2803, 2014.

[9] NB Mutlu Ağardan, Z Değim, Şükran Yılmaz, Levent Altıntaş, and Turgut Topal. Tamoxifen/raloxifene loaded liposomes for oral treatment of breast cancer. *Journal of Drug Delivery Science and Technology*, 57:101612, 2020.

[10] Vaishali Aggarwal, Hardeep Singh Tuli, Jagjit Kaur, Diwakar Aggarwal, Gaurav Parashar, Nidarshana Chaturvedi Parashar, Samruddhi Kulkarni, Ginpreet Kaur, Katrin Sak, Manoj Kumar, et al. Garcinol exhibits anti-neoplastic effects by targeting diverse oncogenic factors in tumor cells. *Biomedicines*, 8(5):103, 2020.

[11] Mohammad Al Hasan, Vineet Chaoji, Saeed Salem, and Mohammed Zaki. Link prediction using supervised learning. In *SDM06: Workshop on Link Analysis, Counter-terrorism and Security*, volume 30, pages 798–805, 2006.

[12] Sumayah Al-Mahmood, Justin Sapiezynski, Olga B Garbuzenko, and Tamara Minko. Metastatic and triple-negative breast cancer: challenges and treatment options. *Drug Delivery and Translational Research*, 8(5):1483–1507, 2018.

[13] Emilio Alba, Oscar M Rueda, Ana Lluch, Joan Albanell, Suet-Feung Chin, Jose Ignacio Chacon, Lourdes Calvo, Juan De La Haba, Begona Bermejo, Nuria Ribelles, et al. Integrative cluster classification to predict pathological complete response to neoadjuvant chemotherapy in early breast cancer., 2018.

[14] Réka Albert, István Albert, and Gary L Nakarado. Structural vulnerability of the north american power grid. *Physical Review Letters*, 69(2):025103, 2004.

[15] H Raza Ali, Oscar M Rueda, Suet-Feung Chin, Christina Curtis, Mark J Dunning, Samuel AJR Aparicio, and Carlos Caldas. Genome-driven integrated classification of breast cancer validated in over 7,500 samples. *Genome Biology*, 15(8):431, 2014.

[16] Preetha Anand, Ajaikumar B Kunnumakara, Chitra Sundaram, Kuzhuvelil B Harikumar, Sheeja T Tharakan, Oiki S Lai, Bokyung Sung, and Bharat B Aggarwal. Cancer is a preventable disease that requires major lifestyle changes. *Pharmaceutical research*, 25(9):2097–2116, 2008.

[17] Alex Arenas, Albert Diaz-Guilera, and Conrad J Pérez-Vicente. Synchronization reveals topological scales in complex networks. *Physical Review Letters*, 96(11):114102, 2006.

[18] Nicole Marie Nieves Aviles, Noel Mayol, Maria T Gonzalez, Adriana Diaz, Iris K Salgado, Walter Silva, and Hector Maldonado. Anticancer and chemosensitizing effects of wee-1 kinase inhibitor mk-1775 in triple negative (mb-231) breast cancer cells. *The FASEB Journal*, 33(S1):674–17, 2019.

[19] Nehad M Ayoub, Dalia R Ibrahim, Amer E Alkhalifa, and Belal A Al-Husein. Crizotinib induced antitumor activity and synergized with chemotherapy and hormonal drugs in breast cancer cells via downregulating met and estrogen receptor levels. *Investigational New Drugs*, 39(1):77–88, 2021.

[20] Matthew H Bailey, Collin Tokheim, Eduard Porta-Pardo, Sohini Sengupta, Denis Bertrand, Amila Weerasinghe, Antonio Colaprico, Michael C Wendl, Jaegil Kim, Brendan Reardon, et al. Comprehensive characterization of cancer driver genes and mutations. *Cell*, 173(2):371–385, 2018.

[21] Irving Balbuena-Rebolledo, Itzia Irene Padilla-Martínez, Martha Cecilia Rosales-Hernández, and Martiniano Bello. Repurposing fda drug compounds against breast cancer by targeting egfr/her2. *Pharmaceuticals*, 14(8):791, 2021.

[22] Allan Balmain, Joe Gray, and Bruce Ponder. The genetics and genomics of cancer. *Nature Genetics*, 33(3s):238, 2003.

[23] Priyanka Baloni, Wikum Dinalankara, John C Earls, Theo A Knijnenburg, Donald Geman, Luigi Marchionni, and Nathan D Price. Identifying personalized metabolic signatures in breast cancer. *Metabolites*, 11(1):20, 2021.

[24] Albert-László Barabási et al. Emergence of scaling in complex networks. *Handbook of Graphs and Networks: From the Genome to the Internet. Berlin: Wiley-VCH*, 2003.

[25] Albert-Laszlo Barabâsi, Hawoong Jeong, Zoltan Néda, Erzsebet Ravasz, Andras Schubert, and Tamas Vicsek. Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications*, 311(3-4):590–614, 2002.

[26] Rebecca SS Barbosa, Paola M Dantonio, Taís Guimarães, Mariana B de Oliveira, Veruska L Fook Alves, Alex Freire Sandes, Rodrigo Carlini Fernando, and Gisele WB Colleoni. Sequential combination of bortezomib and wee1 inhibitor, mk-1775, induced apoptosis in multiple myeloma cell lines. *Biochemical and Biophysical Research Communications*, 519(3):597–604, 2019.

[27] Rameen Beroukhim, Gad Getz, Leia Nghiemphu, Jordi Barretina, Teli Hsueh, David Linhart, Igor Vivanco, Jeffrey C Lee, Julie H Huang, Sethu Alexander, et al. Assess-

ing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proceedings of the National Academy of Sciences*, 104(50):20007–20012, 2007.

[28] Myron G Best, Pieter Wesseling, and Thomas Wurdinger. Tumor-educated platelets as a noninvasive biomarker source for cancer detection and progression monitoring. *Cancer Research*, 78(13):3407–3412, 2018.

[29] Giampaolo Bianchini, Justin M Balko, Ingrid A Mayer, Melinda E Sanders, and Luca Gianni. Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. *Nature Reviews Clinical Oncology*, 13(11):674, 2016.

[30] Kirby I Bland, Edward M Copeland, V Suzanne Klimberg, and William J Gradishar. *The breast: comprehensive management of benign and malignant diseases*. Elsevier Inc, 2017.

[31] Fiona M Blows, Kristy E Driver, Marjanka K Schmidt, Annegien Broeks, Flora E Van Leeuwen, Jelle Wesseling, Maggie C Cheang, Karen Gelmon, Torsten O Nielsen, Carl Blomqvist, et al. Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short and long term survival: a collaborative analysis of data for 10,159 cases from 12 studies. *PLoS Medicine*, 7(5):e1000279, 2010.

[32] Mikael Bodén, Graham Dellaire, Kevin Burrage, and Timothy L Bailey. A bayesian network model of proteins' association with promyelocytic leukemia (pml) nuclear bodies. *Journal of Computational Biology*, 17(4):617–630, 2010.

[33] Marilena M Bourdakou, Emmanouil I Athanasiadis, and George M Spyrou. Discovering gene re-ranking efficiency and conserved gene-gene relationships derived from gene co-expression network analysis on breast cancer data. *Scientific Reports*, 6(1):1–29, 2016.

[34] Richard J Bownes, Arran K Turnbull, Carlos Martinez-Perez, David A Cameron, Andrew H Sims, and Olga Oikonomidou. On-treatment biomarkers can improve prediction of response to neoadjuvant chemotherapy in breast cancer. *Breast Cancer Research*, 21(1):1–12, 2019.

[35] Maria Gabriella Brasca, Clara Albanese, Rachele Alzani, Raffaella Amici, Nilla Avanzi, Dario Ballinari, James Bischoff, Daniela Borghi, Elena Casale, Valter Croci, et al. Optimization of 6, 6-dimethyl pyrrolo [3, 4-c] pyrazoles: Identification of pha-793887, a potent cdk inhibitor suitable for intravenous dosing. *Bioorganic & Medicinal Chemistry*, 18(5):1844–1853, 2010.

[36] Freddie Bray, Jacques Ferlay, Isabelle Soerjomataram, Rebecca L Siegel, Lindsey A Torre, and Ahmedin Jemal. Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a Cancer Journal for Clinicians*, 2018.

[37] Darren R Brenner, Abbey Poirier, Ryan R Woods, Larry F Ellison, Jean-Michel Billette, Alain A Demers, Shary Xinyu Zhang, Chunhe Yao, Christian Finley, Natalie Fitzgerald, et al. Projected estimates of cancer in canada in 2022. *CMAJ*, 194(17):E601–E607, 2022.

[38] James D Brenton, Lisa A Carey, Ahmed Ashour Ahmed, and Carlos Caldas. Molecular classification and molecular forecasting of breast cancer: ready for clinical application? *Journal of Clinical Oncology*, 23(29):7350–7360, 2005.

[39] Frederic Buxant, Nadège Kindt, Guy Laurent, Jean-Christophe Noël, and Sven Saussez. Antiproliferative effect of dexamethasone in the mcf-7 breast cancer cell line. *Molecular Medicine Reports*, 12(3):4051–4054, 2015.

[40] Robert M Califf. Biomarker definitions and their applications. *Experimental Biology and Medicine*, 243(3):213–221, 2018.

[41] Philippe M Campeau, William D Foulkes, and Marc D Tischkowitz. Hereditary breast cancer: new genetic developments, new therapeutic avenues. *Human Genetics*, 124(1):31–42, 2008.

[42] Lisa A Carey, Charles M Perou, Chad A Livasy, Lynn G Dressler, David Cowan, Kathleen Conway, Gamze Karaca, Melissa A Troester, Chiu Kit Tse, Sharon Edmiston, et al. Race, breast cancer subtypes, and survival in the carolina breast cancer study. *Jama*, 295(21):2492–2502, 2006.

[43] Joshua Casaos, Noah L Gorelick, Sakibul Huq, John Choi, Yuanxuan Xia, Riccardo Serra, Raphael Felder, Tarik Lott, Richard E Kast, Ian Suk, et al. The use of ribavirin as an anticancer therapeutic: will it go viral? *Molecular Cancer Therapeutics*, 18(7):1185–1194, 2019.

[44] Rodrigo C Castro Miranda, Sergio R Stancato de Souza, Thiago Christiano Silva, and Benjamin Miranda Tabak. Connectivity and systemic risk in the brazilian national payments system. *Journal of Complex Networks*, 2(4):585–613, 2014.

[45] Jane A Cauley, Larry Norton, Marc E Lippman, Stephen Eckert, Kathryn A Krueger, David W Purdie, Jordi Farrerons, Avraham Karasik, Dan Mellstrom, Kong Wah Ng, et al. Continued breast cancer risk reduction in postmenopausal women treated with raloxifene: 4-year results from the more trial. *Breast cancer Research and Treatment*, 65(2):125–134, 2001.

[46] Arnaud Ceol, Andrew Chatr Aryamontri, Luana Licata, Daniele Peluso, Leonardo Briganti, Livia Perfetto, Luisa Castagnoli, and Gianni Cesareni. Mint, the molecular interaction database: 2009 update. *Nucleic Acids Research*, page gkp983, 2009.

[47] Thien-Nga Chamaraux-Tran, Carole Mathelin, Marc Aprahamian, Girish P Joshi, Catherine Tomasetto, Pierre Diemunsch, and Cherif Akladios. Antitumor effects of lidocaine on human breast cancer cells: an in vitro and in vivo experimental trial. *Anticancer Research*, 38(1):95–105, 2018.

[48] Olivier Chapelle, Bernhard Schölkopf, and Alexander Zien. Semi-supervised learning. adaptive computation and machine learning. *MIT Press, Cambridge, MA, USA. Cited in page (s)*, 21(1):2, 2010.

[49] Luxi Chen, Chao Long, Jennifer Nguyen, Dhiraj Kumar, and Jiyong Lee. Discovering alkylamide derivatives of bexarotene as new therapeutic agents against triple-negative breast cancer. *Bioorganic & Medicinal Chemistry Letters*, 28(3):420–424, 2018.

[50] Nicole J Chew, Elizabeth V Nguyen, Shih-Ping Su, Karel Novy, Howard C Chan, Lan K Nguyen, Jennii Luu, Kaylene J Simpson, Rachel S Lee, and Roger J Daly. Fgfr3 signaling and function in triple negative breast cancer. *Cell Communication and Signaling*, 18(1):1–17, 2020.

[51] Graham A Colditz, Bernard A Rosner, Wendy Y Chen, Michelle D Holmes, and Susan E Hankinson. Risk factors for breast cancer according to estrogen and progesterone receptor status. *Journal of the National Cancer Institute*, 96(3):218–228, 2004.

[52] International HapMap 3 Consortium et al. Integrating common and rare genetic variation in diverse human populations. *Nature*, 467(7311):52–58, 2010.

[53] Christina Curtis, Sohrab P Shah, Suet-Feung Chin, Gulisa Turashvili, Oscar M Rueda, Mark J Dunning, Doug Speed, Andy G Lynch, Shamith Samarajiwa, Yinyin Yuan, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, 486(7403):346, 2012.

[54] Marcio Rosa Da Silva, Hongwu Ma, and An-Ping Zeng. Centrality, network capacity, and modularity as parameters to analyze the core-periphery structure in metabolic networks. *Proceedings of the IEEE*, 96(8):1411–1420, 2008.

[55] Alexandre de Nonneville, Pascal Finetti, Daniel Birnbaum, Emilie Mamessier, and François Bertucci. Wee1 dependency and pejorative prognostic value in triple-negative breast cancer. *Advanced Science*, 8(17):2101030, 2021.

[56] James DeGregori. Challenging the axiom: does the occurrence of oncogenic mutations truly limit cancer development with age? *Oncogene*, 32(15):1869, 2013.

[57] Fatemeh Delavar Mendi, Amir Sh. Saljooghi, Mohammad Ramezani, Rafal Kruszynski, Morgane Poupon, Monika Kucerakova, Volker Huch, Pawel Socha, Maryam Babaei, and Mona Alibolandi. Five new complexes with deferiprone and n, n-donor ligands: evaluation of cytotoxicity against breast cancer mcf-7 cell line

and hsa-binding determination. *Journal of Biomolecular Structure and Dynamics*, 39(13):4845–4858, 2021.

[58] PP Deotarse, AS Jain, MB Baile, NS Kolhe, and AA Kulkarni. Drug repositioning: a review. *International Journal of Pharmaceutical Research Review*, 4:51–8, 2015.

[59] Christine Desmedt, Christos Sotiriou, and Martine J Piccart-Gebhart. Development and validation of gene expression profile signatures in early-stage breast cancer. *Cancer Investigation*, 27(1):1–10, 2009.

[60] Christopher Dextras, Myagmarjav Dashnyam, Lesley A Mathews Griner, Janani Sundaresan, Bryan Chim, Zhiya Yu, Suman Vodnala, Chyi-Chia Richard Lee, Xin Hu, Noel Southall, et al. Identification of small molecule enhancers of immunotherapy for melanoma. *Scientific Reports*, 10(1):1–11, 2020.

[61] Michael Dickson and Jean Paul Gagnon. The cost of new drug discovery and development. *Discovery Medicine*, 4(22):172–179, 2009.

[62] Véronique Diéras, Hervé Bonnefoi, Emilio Alba, Ahmad Awada, Bruno Coudert, Xavier Pivot, Joseph Gligorov, Agnes Jager, Stefania Zambelli, Geoffrey J Lindeman, et al. Iniparib administered weekly or twice-weekly in combination with gemcitabine/carboplatin in patients with metastatic triple-negative breast cancer: a phase ii randomized open-label study with pharmacokinetics. *Breast Cancer Research and Treatment*, 177(2):383–393, 2019.

[63] Joseph A DiMasi, Lanna Feldman, A Seckler, and A Wilson. Trends in risks associated with new drug development: success rates for investigational drugs. *Clinical Pharmacology & Therapeutics*, 87(3):272–277, 2010.

[64] Mitch Dowsett, Ivana Sestak, Elena Lopez-Knowles, Kalvinder Sidhu, Anita K Dunbier, J Wayne Cowens, Sean Ferree, James Storhoff, Carl Schaper, and Jack Cuzick. Comparison of pam50 risk of recurrence score with onco type dx and ihc4 for predicting risk of distant recurrence after endocrine therapy. *Journal of Clinical Oncology*, 31(22):2783–2790, 2013.

[65] Stephen W Duffy, Olorunsola Agbaje, Laszlo Tabar, Bedrich Vitak, Nils Bjurstam, Lena Björneld, Jonathan P Myles, and Jane Warwick. Overdiagnosis and overtreatment of breast cancer: estimates of overdiagnosis from two trials of mammographic screening for breast cancer. *Breast Cancer Research*, 7(6):1–8, 2005.

[66] Shenouda G Elia, Ahmed A Al-Karmalawy, Mohamed Y Nasr, and Mohamed F Elshal. Loperamide potentiates doxorubicin sensitivity in triple-negative breast cancer cells by targeting mdr1 and jnk and suppressing mtor and bcl-2: In vitro and molecular docking study. *Journal of Biochemical and Molecular Toxicology*, page e22938, 2021.

[67] Dorothea Emig, Alexander Ivliev, Olga Pustovalova, Lee Lancashire, Svetlana Bureeva, Yuri Nikolsky, and Marina Bessarabova. Drug target prediction and repositioning using an integrated network-based approach. *PLoS One*, 8(4):e60618, 2013.

[68] Paul Erdős, Alfréd Rényi, et al. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1):17–60, 1960.

[69] Paul W Ewald. Ancient cancers and infection-induced oncogenesis. *International Journal of Paleopathology*, 21:178–185, 2018.

[70] Christos Faloutsos, Kevin S McCurley, and Andrew Tomkins. Fast discovery of connection subgraphs. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 118–127, 2004.

[71] Cheng Fan, Daniel S Oh, Lodewyk Wessels, Britta Weigelt, Dimitry SA Nuyten, Andrew B Nobel, Laura J Van't Veer, and Charles M Perou. Concordance among gene-expression–based predictors for breast cancer. *New England Journal of Medicine*, 355(6):560–569, 2006.

[72] Marco Fiorillo, Fanni Tóth, Matteo Brindisi, Federica Sotgia, and Michael P Lisanti. Deferiprone (dfp) targets cancer stem cell (csc) propagation by inhibiting mitochondrial metabolism and inducing ros production. *Cells*, 9(6):1529, 2020.

[73] Forough Firoozbakht, Iman Rezaeian, Michele D'agnillo, Lisa Porter, Luis Rueda, and Alioune Ngom. An integrative approach for identifying network biomarkers of breast cancer subtypes using genomic, interactomic, and transcriptomic data. *Journal of Computational Biology*, 24(8):756–766, 2017.

[74] Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5):75–174, 2010.

[75] Santo Fortunato, Vito Latora, and Massimo Marchiori. Method to find community structures based on information centrality. *Physical Review Letters*, 70(5):056104, 2004.

[76] Brendan J Frey and Delbert Dueck. Clustering by passing messages between data points. *Science*, 315(5814):972–976, 2007.

[77] David W Fry, Alan J Kraker, Amy McMichael, Linda A Ambroso, James M Nelson, Wilbur R Leopold, Richard W Connors, and Alexander J Bridges. A specific inhibitor of the epidermal growth factor receptor tyrosine kinase. *Science*, 265(5175):1093–1095, 1994.

[78] Hai-Ling Gao, Pranav Gupta, Qingbin Cui, Yunali V Ashar, Zhuo-Xun Wu, Leli Zeng, Zi-Ning Lei, Qiu-Xu Teng, Charles R Ashby Jr, Yingjun Guan, et al. Sapitinib reverses anticancer drug resistance in colon cancer cells overexpressing the abcb1 transporter. *Frontiers in Oncology*, 10:2258, 2020.

[79] M Gnant, M Filipits, R Greil, H Stoeger, M Rudas, Z Bago-Horvath, B Mlineritsch, W Kwasny, M Knauer, C Singer, et al. Predicting distant recurrence in receptor-positive breast cancer patients with limited clinicopathological risk: using the pam50 risk of recurrence score in 1478 postmenopausal patients of the abcsg-8 trial treated with adjuvant endocrine therapy alone. *Annals of Oncology*, 25(2):339–345, 2013.

[80] Mehmet Gönen. Predicting drug-target interactions from chemical and genomic kernels using bayesian matrix factorization. *Bioinformatics*, 28(18):2304–2310, 2012.

[81] Ruggero Gramatica, Tiziana Di Matteo, Stefano Giorgetti, Massimo Barbiani, Dorian Bevec, and Tomaso Aste. Graph theory enables drug repurposing–how a mathematical model can drive the discovery of hidden mechanisms of action. *PloS One*, 9(1):e84912, 2014.

[82] Guowei Gu, Derek Dustin, and Suzanne AW Fuqua. Targeted therapy for breast cancer and molecular mechanisms of resistance to treatment. *Current Opinion in Pharmacology*, 31:97–103, 2016.

[83] Natali Gulbahce and Sune Lehmann. The art of community detection. *BioEssays*, 30(10):934–938, 2008.

[84] William D Gwynne, Mirza S Shakeel, Adele Girgis-Gabardo, and John A Hassell. The role of serotonin in breast cancer stem cells. *Molecules*, 26(11):3171, 2021.

[85] Zahid Halim, Hussain Mahmood Sargana, Muhammad Waqas, et al. Clustering of graphs using pseudo-guided random walk. *Journal of Computational Science*, 51:101281, 2021.

[86] Colan Ho-Yen, Rebecca L Bowen, and JL Jones. Characterization of basal-like breast cancer: an update. *Diagnostic Histopathology*, 18(3):104–111, 2012.

[87] Evthokia Hobbs, Eric Lindquist, Brian Sullivan, Esther Yoon, Ying Yuan, Angela Marx, Jie Willey, Huiming Sun, Rachel Layman, Daniel Stover, et al. Abstract ot-28-04: Neratinib and tepotinib combination in advanced breast cancer and inflammatory breast cancer patients with abnormal her2 and c-met pathway activity as measured by the celsignia signaling pathway activity test, 2021.

[88] Zhiyuan Hu, Cheng Fan, Daniel S Oh, JS Marron, Xiaping He, Bahjat F Qaqish, Chad Livasy, Lisa A Carey, Evangeline Reynolds, Lynn Dressler, et al. The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics*, 7(1):96, 2006.

[89] Takayuki Iwamoto and Lajos Pusztai. Predicting prognosis of breast cancer with gene signatures: are we lost in a sea of data? *Genome Medicine*, 2(11):81, 2010.

[90] Nurul Fathiyatul Nabila Jaffar, Muhammad Shahidan Muhammad Sakri, Hasnan Jaafar, Wan Faiziah Wan Abdul Rahman, Tengku Ahmad Damitri Al-Astani Tengku,

et al. Evaluation of nmu-induced breast cancer treated with sirolimus and sunitinib on breast cancer growth. *Asian Pacific Journal of Cancer Prevention: APJCP*, 21(10):2919, 2020.

[91] Anil K Jain. Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, 31(8):651–666, 2010.

[92] Mariam Jamal-Hanjani, Sergio A Quezada, James Larkin, and Charles Swanton. Translational implications of tumor heterogeneity. *Clinical Cancer Research*, 21(6):1258–1266, 2015.

[93] Dong Kee Jang, Yu Geon Lee, Young Chan Chae, Jun Kyu Lee, Woo Hyun Paik, Sang Hyub Lee, Yong-Tae Kim, and Ji Kon Ryu. Gdc-0980 (apitolisib) treatment with gemcitabine and/or cisplatin synergistically reduces cholangiocarcinoma cell growth by suppressing the pi3k/akt/mtor pathway. *Biochemical and Biophysical Research Communications*, 529(4):1242–1248, 2020.

[94] Giup Jang, Taekeon Lee, Byung Mun Lee, and Youngmi Yoon. Literature-based prediction of novel drug indications considering relationships between entities. *Molecular BioSystems*, 13(7):1399–1405, 2017.

[95] Laila A Jaragh-Alhadad, Gamaleldin I Harisa, and Fars K Alanazi. Development of nimesulide analogs as a dual inhibitor targeting tubulin and hsp27 for treatment of female cancers. *Journal of Molecular Structure*, 1248:131479, 2022.

[96] David Jensen, Jennifer Neville, and Brian Gallagher. Why collective inference improves relational classification. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 593–598, 2004.

[97] Jing Jiang, Christo Wilson, Xiao Wang, Wenpeng Sha, Peng Huang, Yafei Dai, and Ben Y Zhao. Understanding latent interactions in online social networks. *ACM Transactions on the Web (TWEB)*, 7(4):1–39, 2013.

[98] Jing Jin and Tony Pawson. Modular evolution of phosphorylation-based signalling systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1602):2540–2555, 2012.

[99] Minoru Kanehisa, Miho Furumichi, Yoko Sato, Mari Ishiguro-Watanabe, and Mao Tanabe. Kegg: integrating viruses and cellular organisms. *Nucleic Acids Research*, 49(D1):D545–D551, 2021.

[100] George Karypis, Eui-Hong Han, and Vipin Kumar. Chameleon: Hierarchical clustering using dynamic modeling. *Computer*, 32(8):68–75, 1999.

[101] Jagroop Kaur, Baljit Kaur, and Palwinder Singh. Rational modification of semaxanib and sunitinib for developing a tumor growth inhibitor targeting atp binding site of tyrosine kinase. *Bioorganic & Medicinal Chemistry Letters*, 28(2):129–133, 2018.

[102] Hacer Kaya Cakır and Onur Eroglu. In vitro anti-proliferative effect of capecitabine (xeloda) combined with mocetinostat (mgcd0103) in 4t1 breast cancer cell line by immunoblotting. *Iranian Journal of Basic Medical Sciences*, 24(11):1515–1522, 2021.

[103] Samuel Kerrien, Bruno Aranda, Lionel Breuza, Alan Bridge, Fiona Broackes-Carter, Carol Chen, Margaret Duesbury, Marine Dumousseau, Marc Feuermann, Ursula Hinz, et al. The intact molecular interaction database in 2012. *Nucleic Acids Research*, page gkr1088, 2011.

[104] Namrata Khurana, Partha K Chandra, Hogyoung Kim, Asim B Abdel-Mageed, Debasis Mondal, and Suresh C Sikka. Bardoxolone-methyl (cddo-me) suppresses androgen receptor and its splice-variant ar-v7 and enhances efficacy of enzalutamide in prostate cancer cells. *Antioxidants*, 9(1):68, 2020.

[105] Ji Won Kim, Jaya Gautam, Ji Eun Kim, Jung Kim, and Keon Wook Kang. Inhibition of tumor growth and angiogenesis of tamoxifen-resistant breast cancer cells by ruxolitinib, a selective jak2 inhibitor. *Oncology Letters*, 17(4):3981–3989, 2019.

[106] SJ Kim, S Nakayama, Y Miyoshi, T Taguchi, Y Tamaki, T Matsushima, Y Torikoshi, S Tanaka, T Yoshida, H Ishihara, et al. Determination of the specific activity of cdk1 and cdk2 as a novel prognostic indicator for early breast cancer. *Annals of Oncology*, 19(1):68–72, 2008.

[107] Sebastian Köhler, Sebastian Bauer, Denise Horn, and Peter N Robinson. Walking the interactome for prioritization of candidate disease genes. *The American Journal of Human Genetics*, 82(4):949–958, 2008.

[108] Martin Krallinger, Ramon Alonso-Allende Erhardt, and Alfonso Valencia. Text-mining approaches in molecular biology and biomedicine. *Drug Discovery Today*, 10(6):439–445, 2005.

[109] Finn Kuusisto, John Steill, Zhaobin Kuang, James Thomson, David Page, and Ron Stewart. A simple text mining approach for ranking pairwise associations in biomedical applications. *AMIA Summits on Translational Science Proceedings*, 2017:166, 2017.

[110] Justin Lamb, Emily D Crawford, David Peck, Joshua W Modell, Irene C Blat, Matthew J Wrobel, Jim Lerner, Jean-Philippe Brunet, Aravind Subramanian, Kenneth N Ross, et al. The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Science*, 313(5795):1929–1935, 2006.

[111] Olga Lavrenova and Vasily Pavlov. The publication of a knowledge organization classification system as a linked data vocabulary. pages 62–73, 2017.

[112] Alice Lee, Scott P Fraser, and Mustafa BA Djamgoz. Propranolol inhibits neonatal nav1. 5 activity and invasiveness of mda-mb-231 breast cancer cells: Effects of combination with ranolazine. *Journal of Cellular Physiology*, 234(12):23066–23081, 2019.

[113] Chandra P Leo, Bettina Hentschel, Thomas D Szucs, and Cornelia Leo. Fda and ema approvals of new breast cancer drugs—a comparative regulatory analysis. *Cancers*, 12(2):437, 2020.

[114] Jiao Li, Xiaoyan Zhu, and Jake Yue Chen. Building disease-specific drug-protein connectivity maps from molecular interaction networks and pubmed abstracts. *PLoS Computational Biology*, 5(7):e1000450, 2009.

[115] Jun-Xia Li, Yan-Ping Bi, Jing Wang, Xue Yang, Yan-Fang Tian, and Zhong-Fang Sun. Jtc-801 inhibits the proliferation and metastasis of ovarian cancer cell skov3 through inhibition of the pi3k-akt signaling pathway. *Die Pharmazie-An International Journal of Pharmaceutical Sciences*, 73(5):283–287, 2018.

[116] Mingyuan Li, Yuan Li, Shiqin Li, Lin Jia, Chunyang Du, Meng Li, Shuangshuang Li, Hervé Galons, Na Guo, and Peng Yu. Co-delivery of f7 and crizotinib by ther-

mosensitive liposome for breast cancer treatment. *Journal of Liposome Research*, pages 1–11, 2021.

[117] Wenlong Li, Hongmei Yang, Xiaoli Li, Lou Han, Ning Xu, and Aiping Shi. Signaling pathway inhibitors target breast cancer stem cells in triple-negative breast cancer. *Oncology Reports*, 41(1):437–446, 2019.

[118] David Liben-Nowell and Jon Kleinberg. The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology*, 58(7):1019–1031, 2007.

[119] Hengrui Liu, James P Dilger, and Jun Lin. Lidocaine suppresses viability and migration of human breast cancer cells: Trpm7 as a target for some breast cancer cell lines. *Cancers*, 13(2):234, 2021.

[120] Huan Liu and Rudy Setiono. Chi2: Feature selection and discretization of numeric attributes. In *2012 IEEE 24th International Conference on Tools with Artificial Intelligence*, pages 388–388. IEEE Computer Society, 1995.

[121] Ke-Qin Liu, Zhi-Ping Liu, Jin-Kao Hao, Luonan Chen, and Xing-Ming Zhao. Identifying dysregulated pathways in cancers from pathway interaction networks. *BMC Bioinformatics*, 13(1):126, 2012.

[122] Xiaoping Liu, Zhi-Ping Liu, Xing-Ming Zhao, and Luonan Chen. Identifying disease genes and module biomarkers by differential interactions. *Journal of the American Medical Informatics Association*, 19(2):241–248, 2012.

[123] Maryam Lotfi Shahreza, Nasser Ghadiri, Sayed Rasoul Mousavi, Jaleh Varshosaz, and James R Green. A review of network-based approaches to drug repositioning. *Briefings in bioinformatics*, 19(5):878–892, 2017.

[124] Pengwei Lu, Yuanting Gu, Lin Li, Fang Wang, Xue Yang, and Yunqing Yang. Retracted article: Belinostat suppresses cell proliferation by inactivating wnt/β-catenin pathway and promotes apoptosis through regulating pkc pathway in breast cancer. *Artificial cells, Nanomedicine, and Biotechnology*, 47(1):3955–3960, 2019.

[125] Wei Lu and Jagath C Rajapakse. Eliminating indeterminacy in ica. *Neurocomputing*, 50:271–290, 2003.

[126] Huimin Luo, Min Li, Shaokai Wang, Quan Liu, Yaohang Li, and Jianxin Wang. Computational drug repositioning using low-rank matrix approximation and randomized algorithms. *Bioinformatics*, 34(11):1904–1912, 2018.

[127] Filipa Lynce, James T Williams, Meredith M Regan, Craig A Bunnell, Rachel A Freedman, Sara M Tolaney, Wendy Y Chen, Erica L Mayer, Ann H Partridge, Eric P Winer, et al. Phase i study of jak1/2 inhibitor ruxolitinib with weekly paclitaxel for the treatment of her2-negative metastatic breast cancer. *Cancer Chemotherapy and Pharmacology*, pages 1–7, 2021.

[128] Sofus A Macskassy and Foster Provost. A simple relational classifier. Technical report, New York Univ, Stern School of Business, 2003.

[129] Sofus A Macskassy and Foster Provost. Classification in networked data: A toolkit and a univariate case study. *Journal of Machine Learning Research*, 8(5), 2007.

[130] Víctor Martínez, Carmen Navarro, Carlos Cano, Waldo Fajardo, and Armando Blanco. Drugnet: Network-based drug–disease prioritization by integrating heterogeneous data. *Artificial Intelligence in Medicine*, 63(1):41–49, 2015.

[131] Kelly E McCann, Sara A Hurvitz, and Nicholas McAndrew. Advances in targeted therapies for triple-negative breast cancer. *Drugs*, pages 1–14, 2019.

[132] Ronan P McLaughlin, Jichao He, Vera E Van Der Noord, Jevin Redel, John A Foekens, John WM Martens, Marcel Smid, Yinghui Zhang, and Bob Van de Water. A kinase inhibitor screen identifies a dual cdc7/cdk9 inhibitor to sensitise triple-negative breast cancer to egfr-targeted therapy. *Breast Cancer Research*, 21(1):1–15, 2019.

[133] Mathias Meijer, Lau Caspar Thygesen, Anders Green, Martha Emneus, Klaus Brasso, Peter Iversen, Eero Pukkala, Kristian Bolin, Knut Stavem, and Annette K Ersbøll. Finasteride treatment and male breast cancer: a register-based cohort study in four nordic countries. *Cancer Medicine*, 7(1):254–260, 2018.

[134] Ruben A Mesa, Uma Yasothan, and Peter Kirkpatrick. Ruxolitinib. *Nature Reviews Drug discovery*, 11(2):103–104, 2012.

[135] Mark S Mizruchi et al. *The American corporate network 1904-1974*, volume 138. SAGE Publications, Incorporated, 1982.

[136] Jose M Montoya and Ricard V Solé. Small world patterns in food webs. *Journal of Theoretical Biology*, 214(3):405–412, 2002.

[137] Ken Morimoto and Hiroaki Kinoshita. Once-daily use of ofloxacin for prophylaxis in breast cancer surgery. *Chemotherapy*, 44(2):135–141, 1998.

[138] Abhik Mukherjee, Roslin Russell, Suet-Feung Chin, B Liu, OM Rueda, HR Ali, G Turashvili, B Mahler-Araujo, IO Ellis, S Aparicio, et al. Associations between genomic stratification of breast cancer and centrally reviewed tumour pathology in the metabric cohort. *NPJ Breast Cancer*, 4(1):5, 2018.

[139] Joseph Mullen, Simon J Cockell, Peter Woollard, and Anil Wipat. An integrated data driven approach to drug repositioning using gene-disease associations. *PloS One*, 11(5):e0155811, 2016.

[140] Uttom Nandi, Ichioma Onyesom, and Dennis Douroumis. An in vitro evaluation of antitumor activity of sirolimus-encapsulated liposomes in breast cancer cells. *Journal of Pharmacy and Pharmacology*, 73(3):300–309, 2021.

[141] Francesco Napolitano, Yan Zhao, Vânia M Moreira, Roberto Tagliaferri, Juha Kere, Mauro D'Amato, and Dario Greco. Drug repositioning: a machine-learning approach through data integration. *Journal of Cheminformatics*, 5(1):30, 2013.

[142] Mark EJ Newman. Fast algorithm for detecting community structure in networks. *Physical Review Letters*, 69(6):066133, 2004.

[143] Mark EJ Newman. Finding community structure in networks using the eigenvectors of matrices. *Physical Review Letters*, 74(3):036104, 2006.

[144] Mark EJ Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, 2006.

[145] Mark EJ Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Physical review Letters*, 69(2):026113, 2004.

[146] Serena Nik-Zainal, Helen Davies, Johan Staaf, Manasa Ramakrishna, Dominik Glodzik, Xueqing Zou, Inigo Martincorena, Ludmil B Alexandrov, Sancha Martin, David C Wedge, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*, 534(7605):47–54, 2016.

[147] Katie M O'Brien, Stephen R Cole, Chiu-Kit Tse, Charles M Perou, Lisa A Carey, William D Foulkes, Lynn G Dressler, Joseph Geradts, and Robert C Millikan. Intrinsic breast tumor subtypes, race, and long-term survival in the carolina breast cancer study. *Clinical Cancer Research*, 16(24):6100–6110, 2010.

[148] Kenneth Olden, Pascal Breton, Krzysztof Grzegorzewski, Yoshiaki Yasuda, Barry L Gause, Oladipo A Oredipe, Sheila A Newton, and Sandra L White. The potential importance of swainsonine in therapy for cancers and immunology. *Pharmacology and Therapeutics*, 50(3):285–290, 1991.

[149] Adedayo A Onitilo, Jessica M Engel, Robert T Greenlee, and Bickol N Mukesh. Breast cancer subtypes based on er/pr and her2 expression: comparison of clinicopathologic features and survival. *Clinical Medicine & Research*, pages cmr–2009, 2009.

[150] Tudor I Oprea and John P Overington. Computational and practical aspects of drug repositioning. *Assay and Drug Development Technologies*, 13(6):299–306, 2015.

[151] Joyce O'Shaughnessy, Cynthia Osborne, John E Pippen, Mark Yoffe, Debra Patt, Christine Rocha, Ingrid Chou Koo, Barry M Sherman, and Charles Bradley. Iniparib plus chemotherapy in metastatic triple-negative breast cancer. *New England Journal of Medicine*, 364(3):205–214, 2011.

[152] Soha Osman, Afsheen Raza, Lobna Al-Zaidan, Varghese Philipose Inchakalody, Maysaloun Merhi, Kirti S Prabhu, Nouha Abdelaziz, Shereena Hydrose, Shahab Uddin, and Said Dermime. Anti-cancer effects of tranilast: An update. *Biomedicine & Pharmacotherapy*, 141:111844, 2021.

[153] Gergely Palla, Imre Derényi, Illés Farkas, and Tamás Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–818, 2005.

[154] Guillermo Palma, Maria-Esther Vidal, and Louiqa Raschid. Drug-target interaction prediction using semantic similarity and edge partitioning. In *International Semantic Web Conference*, pages 131–146. Springer, 2014.

[155] Joel S Parker, Michael Mullins, Maggie CU Cheang, Samuel Leung, David Voduc, Tammi Vickery, Sherri Davies, Christiane Fauron, Xiaping He, Zhiyuan Hu, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of Clinical Oncology*, 27(8):1160, 2009.

[156] Nathan D Pennock, Holly A Martinson, Qiuchen Guo, Courtney B Betts, Sonali Jindal, Takahiro Tsujikawa, Lisa M Coussens, Virginia F Borges, and Pepper Schedin. Ibuprofen supports macrophage differentiation, t cell recruitment, and tumor suppression in a model of postpartum breast cancer. *Journal for Immunotherapy of Cancer*, 6(1):1–23, 2018.

[157] Charles M Perou, Therese Sørlie, Michael B Eisen, Matt Van De Rijn, Stefanie S Jeffrey, Christian A Rees, Jonathan R Pollack, Douglas T Ross, Hilde Johnsen, Lars A Akslen, et al. Molecular portraits of human breast tumours. *Nature*, 406(6797):747, 2000.

[158] Asim Pervaiz, Michael Zepp, Saqib Mahmood, Doaa Mohamed Ali, Martin R Berger, and Hassan Adwan. Ccr5 blockage by maraviroc: a potential therapeutic option for metastatic breast cancer. *Cellular Oncology*, 42(1):93–106, 2019.

[159] Azam Peyvandipour, Nafiseh Saberian, Adib Shafi, Michele Donato, and Sorin Draghici. A novel computational approach for drug repurposing using systems biology. *Bioinformatics*, 34(16):2817–2825, 2018.

[160] Paul F Pinsky, Eric A Miller, Brandy M Heckman-Stoddard, and Lori Minasian. Breast cancer characteristics and survival among users versus nonusers of raloxifene. *Cancer Prevention Research*, 13(1):83–90, 2020.

[161] Sebastian Poledna, José Luis Molina-Borboa, Serafín Martínez-Jaramillo, Marco Van Der Leij, and Stefan Thurner. The multi-layer network nature of systemic risk and its implications for the costs of financial crises. *Journal of Financial Stability*, 20:70–81, 2015.

[162] TS Keshava Prasad, Renu Goel, Kumaran Kandasamy, Shivakumar Keerthikumar, Sameer Kumar, Suresh Mathivanan, Deepthi Telikicherla, Rajesh Raju, Beema Shafreen, Abhilash Venugopal, et al. Human protein reference database—2009 update. *Nucleic Acids Research*, 37(suppl 1):D767–D772, 2009.

[163] Arsela Prelaj, Rebecca Tay, Roberto Ferrara, Nathalie Chaput, Benjamin Besse, and Raffaele Califano. Predictive biomarkers of response for immune checkpoint inhibitors in non–small-cell lung cancer. *European Journal of Cancer*, 106:144–159, 2019.

[164] Sarjana Raikwar and Sanjay K Jain. Opportunities and challenges in breast cancer. *International Journal of Pharmacy & Life Sciences*, 11(7):6858, 2020.

[165] EA Rakha, TC Putti, DM Abd El-Rehim, C Paish, AR Green, DG Powe, AH Lee, JF Robertson, and IO Ellis. Morphological and immunophenotypic analysis of breast carcinomas with basal and myoepithelial differentiation. *The Journal of Pathology*, 208(4):495–506, 2006.

[166] Majid Rastegar-Mojarad, Ravikumar Komandur Elayavilli, Dingcheng Li, Rashmi Prasad, and Hongfang Liu. A new method for prioritizing drug repositioning candidates extracted by literature-based discovery. In *2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 669–674. IEEE, 2015.

[167] Erzsébet Ravasz, Anna Lisa Somera, Dale A Mongru, Zoltán N Oltvai, and A-L Barabási. Hierarchical organization of modularity in metabolic networks. *Science*, 297(5586):1551–1555, 2002.

[168] Michel Raymond and Francois Rousset. An exact test for population differentiation. *Evolution*, 49(6):1280–1283, 1995.

[169] Jörg Reichardt and Stefan Bornholdt. Detecting fuzzy community structures in complex networks with a potts model. *Physical Review Letters*, 93(21):218701, 2004.

[170] Iman Rezaeian, Yifeng Li, Martin Crozier, Eran Andrechek, Alioune Ngom, Luis Rueda, and Lisa Porter. Identifying informative genes for prediction of breast cancer subtypes. In *IAPR International Conference on Pattern Recognition in Bioinformatics*, pages 138–148. Springer, 2013.

[171] Martin Rosvall and Carl T Bergstrom. An information-theoretic framework for resolving community structure in complex networks. *Proceedings of the National Academy of Sciences*, 104(18):7327–7331, 2007.

[172] Tristan Rupp, Océane Pelouin, Laurie Genest, Christophe Legrand, Guillaume Froget, and Vincent Castagné. Therapeutic potential of fingolimod in triple negative breast cancer preclinical models. *Translational Oncology*, 14(1):100926, 2021.

[173] Toshiaki Saeki, Tadashi Nomizu, Masakazu Toi, Yoshinori Ito, Shinzaburo Noguchi, Tadashi Kobayashi, Taro Asaga, Hironobu Minami, Naohito Yamamoto, Kenjiro Aogi, et al. Dofequidar fumarate (ms-209) in combination with cyclophosphamide, doxorubicin, and fluorouracil for patients with advanced or recurrent breast cancer. *Journal of Clinical Oncology*, 25(4):411–417, 2007.

[174] Rainer Schobert and Bernhard Biersack. Chemical and biological aspects of garcinol and isogarcinol: Recent developments. *Chemistry & Biodiversity*, 16(9):e1900366, 2019.

[175] Maria Giovanna Scioli, Gabriele Storti, Federico D'Amico, Pietro Gentile, Giulia Fabbri, Valerio Cervelli, and Augusto Orlandi. The role of breast cancer stem cells as a prognostic marker and a target to improve the efficacy of breast cancer therapy. *Cancers*, 11(7):1021, 2019.

[176] Nova Scotia. Explaining odds ratios. *J Can Acad Child Adolesc Psychiatry*, 19:227, 2010.

[177] Elizabeth S Scott. Social norms and the legal regulation of marriage. *Virginia Law Review*, pages 1901–1970, 2000.

[178] Eran Segal, Haidong Wang, and Daphne Koller. Discovering molecular pathways from protein interaction and gene expression data. *Bioinformatics*, 19(suppl_1):i264–i272, 2003.

[179] Prithviraj Sen, Galileo Namata, Mustafa Bilgic, Lise Getoor, Brian Galligher, and Tina Eliassi-Rad. Collective classification in network data. *AI Magazine*, 29(3):93–93, 2008.

[180] Ean-Jeong Seo, Yoshikazu Sugimoto, Henry Johannes Greten, and Thomas Efferth. Repurposing of bromocriptine for cancer therapy. *Frontiers in Pharmacology*, 9:1030, 2018.

[181] Aylin Sertkaya, Anna Birkenbach, Ayesha Berlind, and John Eyraud. Examination of clinical trial costs and barriers for drug development. *Report, US Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation, Washington, DC*, pages 1–92, 2014.

[182] Masih Sherafatian. Tree-based machine learning algorithms identified minimal set of mirna biomarkers for breast cancer diagnosis and molecular subtyping. *Gene*, 677:111–118, 2018.

[183] Rebecca L Siegel, Kimberly D Miller, Hannah E Fuchs, and Ahmedin Jemal. Cancer statistics, 2022. *CA: A Cancer Journal for Clinicians*, 2022.

[184] Marina Sirota, Joel T Dudley, Jeewon Kim, Annie P Chiang, Alex A Morgan, Alejandro Sweet-Cordero, Julien Sage, and Atul J Butte. Discovery and preclinical validation of drug indications using compendia of public gene expression data. *Science Translational Medicine*, 3(96):96ra77–96ra77, 2011.

[185] Therese Sørlie. Molecular portraits of breast cancer: tumour subtypes as distinct disease entities. *European Journal of Cancer*, 40(18):2667–2675, 2004.

[186] Therese Sørlie, Charles M Perou, Robert Tibshirani, Turid Aas, Stephanie Geisler, Hilde Johnsen, Trevor Hastie, Michael B Eisen, Matt Van De Rijn, Stefanie S Jeffrey, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences*, 98(19):10869–10874, 2001.

[187] Therese Sørlie, Robert Tibshirani, Joel Parker, Trevor Hastie, James Stephen Marron, Andrew Nobel, Shibing Deng, Hilde Johnsen, Robert Pesich, Stephanie Geisler, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proceedings of the National Academy of Sciences*, 100(14):8418–8423, 2003.

[188] 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*, 100(18):10393–10398, 2003.

[189] Sergio RS Souza, Benjamin M Tabak, Thiago C Silva, and Solange M Guerra. Insolvency and contagion in the brazilian interbank market. *Physica A: Statistical Mechanics and its Applications*, 431:140–151, 2015.

[190] A Spitale, P Mazzola, D Soldini, L Mazzucchelli, and A Bordoni. Breast cancer classification according to immunohistochemical markers: clinicopathologic features and short-term survival analysis in a population-based study from the south of switzerland. *Annals of oncology*, 20(4):628–635, 2008.

[191] Olaf Sporns. Network analysis, complexity, and brain function. *Complexity*, 8(1):56–60, 2002.

[192] Chris Stark, Bobby-Joe Breitkreutz, Andrew Chatr-Aryamontri, Lorrie Boucher, Rose Oughtred, Michael S Livstone, Julie Nixon, Kimberly Van Auken, Xiaodong Wang, Xiaoqi Shi, et al. The biogrid interaction database: 2011 update. *Nucleic Acids Research*, 39(suppl 1):D698–D704, 2011.

[193] Daniel G Stover, Carlos R Gil Del Alcazar, Jane Brock, Hao Guo, Beth Overmoyer, Justin Balko, Qiong Xu, Aditya Bardia, Sara M Tolaney, Rebecca Gelman, et al. Phase ii study of ruxolitinib, a selective jak1/2 inhibitor, in patients with metastatic triple-negative breast cancer. *NPJ Breast Cancer*, 4(1):1–9, 2018.

[194] Lovro Šubelj and Marko Bajec. Unfolding communities in large complex networks: Combining defensive and offensive label propagation for core extraction. *Physical Review Letters*, 83(3):036103, 2011.

[195] Mengying Sun, Sendong Zhao, Coryandar Gilvary, Olivier Elemento, Jiayu Zhou, and Fei Wang. Graph convolutional networks for computational drug development and discovery. *Briefings in Bioinformatics*, 21(3):919–935, 2020.

[196] Yi Sun, Zhen Sheng, Chao Ma, Kailin Tang, Ruixin Zhu, Zhuanbin Wu, Ruling Shen, Jun Feng, Dingfeng Wu, Danyi Huang, et al. Combining genomic and network characteristics for extended capability in predicting synergistic drugs for cancer. *Nature Communications*, 6:8481, 2015.

[197] Srilatha Swami, Aruna V Krishnan, Jennifer Y Wang, Kristin Jensen, Lihong Peng, Megan A Albertelli, and David Feldman. Inhibitory effects of calcitriol on the growth

of mcf-7 breast cancer xenografts in nude mice: selective modulation of aromatase expression in vivo. *Hormones and Cancer*, 2(3):190–202, 2011.

[198] Rebecca R Swenson, Christie J Rizzo, Larry K Brown, Nanetta Payne, Ralph J DiClemente, Laura F Salazar, Peter A Vanable, Michael P Carey, Robert F Valois, Daniel Romer, et al. Prevalence and correlates of hiv testing among sexually active african american adolescents in four us cities. *Sexually Transmitted Diseases*, 36(9):584, 2009.

[199] Ping Tang, Jianmin Wang, and Patria Bourne. Molecular classifications of breast carcinoma with similar terminology and different definitions: are they the same? *Human Pathology*, 39(4):506–513, 2008.

[200] Noura M Thabet and Enas M Moustafa. Synergistic effect of ebselen and gamma radiation on breast cancer cells. *International Journal of Radiation Biology*, 93(8):784–792, 2017.

[201] Kutlwano Thabethe, Gbenga Adefolaju, and Margot Hosie. An in vitro study of the effects of emtricitabine, tenofovir disoproxil fumarate and efavirenz on a breast cancer cell line, mcf-7. *Journal of Basic and Applied Scientific Research*, 3:444–452, 2013.

[202] Jeffrey G Thomas, James M Olson, Stephen J Tapscott, and Lue Ping Zhao. An efficient and robust statistical modeling approach to discover differentially expressed genes using genomic expression profiles. *Genome Research*, 11(7):1227–1236, 2001.

[203] Jun Tian, Fatmah Al Raffa, Meiou Dai, Alaa Moamer, Baharak Khadang, Ibrahim Y Hachim, Khldoun Bakdounes, Suhad Ali, Bertrand Jean-Claude, and Jean-Jacques Lebrun. Dasatinib sensitises triple negative breast cancer cells to chemotherapy by targeting breast cancer stem cells. *British Journal of Cancer*, 119(12):1495–1507, 2018.

[204] Nicholas C Turner, Jungsil Ro, Fabrice André, Sherene Loi, Sunil Verma, Hiroji Iwata, Nadia Harbeck, Sibylle Loibl, Cynthia Huang Bartlett, Ke Zhang, et al. Palbociclib in hormone-receptor–positive advanced breast cancer. *New England Journal of Medicine*, 373(3):209–219, 2015.

[205] Ravi Kant Upadhyay. Emerging risk biomarkers in cardiovascular diseases and disorders. *Journal of Lipids*, 2015, 2015.

[206] Carlos S Vallejos, Henry L Gómez, Wilder R Cruz, Joseph A Pinto, Richard R Dyer, Raúl Velarde, Juan F Suazo, Silvia P Neciosup, Mauricio León, A Miguel, et al. Breast cancer classification according to immunohistochemistry markers: subtypes and association with clinicopathologic variables in a peruvian hospital database. *Clinical Breast Cancer*, 10(4):294–300, 2010.

[207] Oron Vanunu, Oded Magger, Eytan Ruppin, Tomer Shlomi, and Roded Sharan. Associating genes and protein complexes with disease via network propagation. *PLoS Computational Biology*, 6(1):e1000641, 2010.

[208] Jia Wang, Sheng-kai Huang, Mei Zhao, Mei Yang, Jia-ling Zhong, Yu-yu Gu, Hua Peng, Yi-qun Che, and Chang-zhi Huang. Identification of a circulating microrna signature for colorectal cancer detection. *PloS One*, 9(4):e87451, 2014.

[209] Kevin K Wang, Zhihui Yang, Tian Zhu, Yuan Shi, Richard Rubenstein, J Adrian Tyndall, and Geoff T Manley. An update on diagnostic and prognostic biomarkers for traumatic brain injury. *Expert Review of Molecular Diagnostics*, 18(2):165–180, 2018.

[210] Yongyu Wang. *High Performance Spectral Methods for Graph-Based Machine Learning*. PhD thesis, Michigan Technological University, 2021.

[211] Barbara L Wehde, Patrick D Rädler, Hridaya Shrestha, Stevi J Johnson, Aleata A Triplett, and Kay-Uwe Wagner. Janus kinase 1 plays a critical role in mammary cancer progression. *Cell Reports*, 25(8):2192–2207, 2018.

[212] Tao Wei, Xiaoyu Zhang, Qi Zhang, Jiaqi Yang, Qi Chen, Jianxin Wang, Xiang Li, Junye Chen, Tao Ma, Guogang Li, et al. Vimentin-positive circulating tumor cells as a biomarker for diagnosis and treatment monitoring in patients with pancreatic cancer. *Cancer Letters*, 452:237–243, 2019.

[213] Juyang Weng and Matthew D Luciw. Brain-inspired concept networks: Learning concepts from cluttered scenes. *IEEE Intelligent Systems*, 29(6):14–22, 2014.

[214] Nicolas Wentzensen and Christine D Berg. Population testing for high penetrance genes: are we there yet?, 2018.

[215] Geoffrey B West, James H Brown, and Brian J Enquist. A general model for the structure and allometry of plant vascular systems. *Nature*, 400(6745):664–667, 1999.

[216] Asdi Wihandono, Yohana Azhar, Maman Abdurahman, and Sjarief Hidayat. The role of lisinopril and bisoprolol to prevent anthracycline induced cardiotoxicity in

locally advanced breast cancer patients. *Asian Pacific Journal of Cancer Prevention*, 22(9):2847–2853, 2021.

[217] Chi Heem Wong, Kien Wei Siah, and Andrew W Lo. Estimation of clinical trial success rates and related parameters. *Biostatistics*, 20(2):273–286, 2019.

[218] Chao Wu, Ranga C Gudivada, Bruce J Aronow, and Anil G Jegga. Computational drug repositioning through heterogeneous network clustering. *BMC Systems Biology*, 7(5):S6, 2013.

[219] Hao Wu, Lin Gao, Jihua Dong, and Xiaofei Yang. Detecting overlapping protein complexes by rough-fuzzy clustering in protein-protein interaction networks. *PloS One*, 9(3):e91856, 2014.

[220] Xinghan Wu, Chuanliang Liu, Zihaoran Li, Chengcheng Gai, Dejun Ding, Weijuan Chen, Fengyun Hao, and Wentong Li. Regulation of gsk3β/nrf2 signaling pathway modulated erastin-induced ferroptosis in breast cancer. *Molecular and Cellular Biochemistry*, 473(1):217–228, 2020.

[221] Ioannis Xenarios, Lukasz Salwinski, Xiaoqun Joyce Duan, Patrick Higney, Sul-Min Kim, and David Eisenberg. Dip, the database of interacting proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Research*, 30(1):303–305, 2002.

[222] Wenxi Xia, Shuangzhe Zhang, Yueqing Li, Jiangli Fan, Bin Liu, Lei Wang, and Xiaojun Peng. Ibuprofen-derived fluorescence inhibitor of cox-2 for breast cancer imaging, prevention and treatment. *Dyes and Pigments*, 190:109326, 2021.

[223] Hanqing Xue, Jie Li, Haozhe Xie, and Yadong Wang. Review of drug repositioning approaches and resources. *International Journal of Biological Sciences*, 14(10):1232, 2018.

[224] Jaewon Yang and Jure Leskovec. Overlapping communities explain core–periphery organization of networks. *Proceedings of the IEEE*, 102(12):1892–1902, 2014.

[225] Qing-yu Yao, Jian Li, Rong Chen, Ye Yao, Jun-sheng Xue, Wen-jun Chen, Wei Lu, and Tian-yan Zhou. Preclinical pk/pd model for the combinatorial use of dexamethasone and sulpiride in the treatment of breast cancer. *Acta Pharmacologica Sinica*, 40(12):1596–1602, 2019.

[226] I Nyoman Wawan Tirtha Yasa. The role of goserelin acetate in the management of pre-menopausal breast cancer. *School of Surgical Oncology for General Surgeons XV*, page 49, 2018.

[227] Liang Yu, Jianbin Huang, Zhixin Ma, Jing Zhang, Yapeng Zou, and Lin Gao. Inferring drug-disease associations based on known protein complexes. *BMC Medical Genomics*, 8(2):S2, 2015.

[228] Fan Zhang and Jake Y Chen. Breast cancer subtyping from plasma proteins. *BMC Medical Genomics*, 6(Suppl 1):S6, 2013.

[229] Fan Zhang and Renee Drabier. Ipad: the integrated pathway analysis database for systematic enrichment analysis. *BMC Bioinformatics*, 13(Suppl 15):S7, 2012.

[230] Ming Zhang, Gerold Schmitt-Ulms, Christine Sato, Zhengrui Xi, Yalun Zhang, Ye Zhou, Peter St George-Hyslop, and Ekaterina Rogaeva. Drug repositioning

for alzheimer's disease based on systematic 'omics' data mining. *PloS One*, 11(12):e0168812, 2016.

[231] Ping Zhang, Fei Wang, and Jianying Hu. Towards drug repositioning: a unified computational framework for integrating multiple aspects of drug similarity and disease similarity. In *AMIA Annual Symposium Proceedings*, volume 2014, page 1258. American Medical Informatics Association, 2014.

[232] Shujun Zhang, Lutao Du, Lishui Wang, Xiumei Jiang, Yao Zhan, Juan Li, Keqiang Yan, Weili Duan, Yinghui Zhao, Lili Wang, et al. Evaluation of serum exosomal lnc rna-based biomarker panel for diagnosis and recurrence prediction of bladder cancer. *Journal of Cellular and Molecular Medicine*, 23(2):1396–1405, 2019.

[233] Qian Zhu, Cui Tao, Feichen Shen, and Christopher G Chute. Exploring the pharmacogenomics knowledge base (pharmgkb) for repositioning breast cancer drugs by leveraging web ontology language (owl) and cheminformatics approaches. pages 172–182, 2014.

[234] Jun Zou, Ming-Wu Zheng, Gen Li, and Zhi-Guang Su. Advanced systems biology methods in drug discovery and translational biomedicine. *BioMed Research International*, 2013, 2013.

[235] Yu Zuo, Heng Xu, Zhifeng Chen, Fengmin Xiong, Bei Zhang, Kaixian Chen, Hualiang Jiang, Cheng Luo, and Hao Zhang. 17-aag synergizes with belinostat to exhibit a negative effect on the proliferation and invasion of mda-mb-231 breast cancer cells. *Oncology Reports*, 43(6):1928–1944, 2020.

Vita Auctoris

Forough Firoozbakht was born in 1981 in Shiraz, Iran. She received her Bachelors degree in Computer Engineering from Azad University, Shiraz, Iran, in 2008, and her Master in Computer Science from University of Windsor in 2013. Her research interests include machine learning, bioinformatics, data mining, cancer research specially regarding breast cancer.