Leveraging Molecular Signatures of Cancer for Dynamic Network Modeling

by

Enes Sefa Ayar

A Dissertation Submitted to the
Graduate School of Sciences and Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Master of Science

in

Computational Sciences and Engineering

August 17, 2023

Leveraging Molecular Signatures of Cancer for Dynamic Network Modeling

Koç University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Enes Sefa Ayar

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

_____

Assoc. Prof. Nurcan Tunçbağ (Advisor)

_____

Prof. Attila Gürsoy

_____

Assist. Prof. Tuğba Süzek

Date: _____

To My Brother,

ABSTRACT

Leveraging the Molecular Signatures of Cancer for Dynamic Network

Enes Sefa Ayar
Master of Science in Computational Sciences and Engineering
August 17, 2023

Drug resistance poses a significant challenge to the effectiveness of therapies, driven by accumulation of molecular alterations within dynamic cellular networks. In this thesis, we used a discrete dynamic model, Graph-based Cellular Automata (GCA), to reveal the network-based history of tumor progression and causal association between network modules and drug resistance by data integration. GCA can simulate dynamic systems using initial static information, set of states and simple transition rules. The reference graph is a tissue-specific interactome that consists of both protein-protein interactions and the regulatory network of the transcription factor to gene interactions composed of 8,228 nodes and 63,574 edges. By incorporating known biology and statistical rules of molecular alterations, including stimulations, repressions, and (non)-linear pairwise molecular correlations, GCA simulates molecular signalling and propagates mutation effects downstream of signalling pathways and complexes. Eventually, GCA gives a trajectory of subnetwork models for each context. In comparisons of simulations with and without mutations at the node level, we detected functional subnetworks within the dynamic network structure. We used publicly available omics data from a well-established cancer cell line repository to optimize the GCA model and construct dynamic networks for each cell-line-drug pair for interpreting drug resistance mechanisms at the pathway level. The accuracy of these drug representative networks were evaluated by cross-validation and on an independent test from Patient-Derived Xenografts (PDX). Notably, we found context-specific pathways (e.g. MAPK signalling) involving proteins from drug-resistant cell lines and PDX samples, thereby linking them to investigated drug resistance mechanisms. Overall, this approach, from molecular alterations to dynamic networks, transforms already available large datasets to gain new clinically relevant insights about drug resistance, offering potential implications for cancer therapy.

ÖZETÇE

Kanserin Moleküler İzlerinin Dinamik Ağ Modellemede Kullanılması

Enes Sefa Ayar

Hesaplamalı Bilimler ve Mühendislik, Yüksek Lisans

17 Ağustos 2023

İlaç direnci, dinamik hücresel ağlarda moleküler değişikliklerin birikmesinden kaynaklanan tedavilerin etkinliği için önemli bir sorun teşkil etmektedir. Bu tezde, tümör ilerlemesinin ağ tabanlı geçmişini ve ağ modülleri ile ilaç direnci arasındaki nedensel ilişkiyi veri entegrasyonu yoluyla ortaya çıkarmak için ayrık bir dinamik model olan çizge tabanlı hücresel otomata (GCA) kullandık. GCA, statik verileri, durum kümelerini ve basit geçiş kurallarını kullanarak dinamik sistemleri simüle edebilir. Referans ağ, hem protein-protein etkileşimlerinden hem de transkripsiyon faktörlerinden, 8.228 düğüm ve 63.574 kenardan, ve gen etkileşimlerini düzenleyen etkileşimlerden oluşan dokuya özgü bir interaktomdur. GCA aktive edici, baskılayıcı ve doğrusal olan yada olmayan ikili moleküler korelasyonlar dahil olmak üzere bilinen biyoloji ve moleküler değişikliklerin istatistiksel kurallarını birleştirerek moleküler sinyalizasyonu simüle etmek ve sinyal yolaklarının ve komplekslerinin akış aşağısında mutasyon etkilerini yaymak için kullanılmıştır. Sonuç olarak GCA, her durum için bir alt ağ modelleri gezingesi vermiştir. Düğüm seviyesinde mutasyonlu ve mutasyonsuz simülasyonların karşılaştırılmasıyla, dinamik ağ yapısı içinde işlevsel alt ağlar tespit ettik. GCA modelini optimize etmek ve yolak düzeyinde ilaç direnç mekanizmalarını yorumlamak için her bir hücre hattı-ilaç çifti için dinamik ağlar oluşturmak üzere kapsamlı bir kanser hücre hattı veri tabanından halka açık omik verilerini kullandık. Bu ilaç temsili ağların doğruluğu, çapraz doğrulama ve Hastadan Türetilmiş Ksenograft (PDX) içeren bağımsız bir testle değerlendirdik. Özellikle, ilaca dirençli hücre hatlarından ve PDX örneklerinden proteinleri içeren bağlama özgü yolaklar (örn. MAPK sinyali) bulduk ve böylece bunları araştırılan ilaç direnci mekanizmalarına bağladık. Genel olarak, moleküler değişikliklerden dinamik ağlara kadar bu yaklaşım, halihazırda mevcut olan büyük veri kümelerini dönüştürerek ilaç direnci hakkında klinik olarak ilgili yeni öngörüler elde etmiş ve kanser tedavisi için potansiyel çıkarımlar sunmuştur.

ACKNOWLEDGEMENTS

First and foremost, I sincerely thank my advisor, Assoc. Prof. Nurcan Tunçbağ for her guidance and kindness throughout this work. Working with her has been a privilege, and I appreciate the chances she gave me to sharpen my research skills. Her expertise, understanding, and professionalism have been priceless in shaping the direction of this thesis, and I am truly grateful for her mentorship.

Furthermore, I would like to thank my thesis committee members: Prof. Dr. Attila Gürsoy, and Assistant Professor Tuğba Süzek, for their invaluable feedbacks and accepting to be part of my thesis committee.

I thank the TÜBİTAK ARDEB 1001 program for their financial support to project 121E245. Thank you for empowering us to make an impact in our research field.

I am incredibly grateful to my dear colleagues, Uğur Şahin, Bengi Ruken Yavuz, Kaan Arıcı, Cansu Demirel, Sina Dadmand, Sertan Ali Dikli, Aslı Dansık, Yıldız Aydın, Yiğit Şibal, İdil İlayda Duran, and Ekin Su Erdem for making my academic journey more manageable and the achievements more joyful. I also want to thank my close friends, Rıdvan Balamur, Barışcan Bozkurt, Mert Gayretli, Ali Sertan Yaka, and Ülkem Kasapoğlu, who have been a source of encouragement and companionship.

Finally, I extend my heartfelt gratitude to my family for their support and encouragement in the completion of this thesis. To my parents, Meryem Ayar and Süleyman Ayar, thank you for believing in my abilities and providing constant motivation. To my siblings, Emre Ayar, Hülya Çin, and Melek Tuna, your presence strengthened me during challenging times. I owe a heartfelt thank-you to my beloved Ceren Uzun. Our deep connection has led us on an extraordinary journey of learning and exploration. Your constant presence has been a source of unwavering support and boundless patience.

TABLE OF CONTENTS

List of Tables ix

List of Figures x

Abbreviations xi

Chapter 1: Introduction 1

Chapter 2: Literature Review 5

2.1 Prominent Role of Mutations in Cancer Progression 5

2.2 Molecular Landscape of Cancer 6

2.3 Drug Resistance in Cancer 7

2.4 Network Medicine 9

2.5 Dynamic Network Modelling 10

Chapter 3: Materials and methods 13

3.1 Reference Network Construction 13

3.2 Graph-based Cellular Automata Model 14

3.2.1 Overview of the GCA Approach 14

3.2.2 Transition Rules of GCA 15

3.2.3 Pairwise Estimation Function 16

3.2.4 Use of Mutations in the GCA Model 17

3.3 Application of GCA Model on DepMap Cell Lines 18

3.3.1 Mutation Profiles of DepMap Cell Lines 18

3.3.2 Optimizations on GCA Model 19

3.3.3 Synthetic Reference Networks 19

3.3.4 Specificity Analysis 20

3.3.5 DepMap Cell Line Simulations 20

3.3.6 Construction of Drug Modules 21

3.3.7 Cell Line Blind Cross Validation of Drug Modules 21

3.3.8 Network Proximity Analysis 23

3.3.9 Drug Module Visualization and Investigating Resistance Mechanisms 23

3.4 Application of the GCA Model to Patient Derived Xenografts.....24

3.4.1 PDX Samples.....24

3.4.2 Drug Resistance Predictions on PDX Samples.....24

3.4.3 Investigation of Resistance Mechanisms in PDX Modules 25

Chapter 4: RESULTS.....27

4.1 Data Statistics.....27

4.2 Optimization of the Model 29

4.2.1 Sensitivity of Model to Differentiate Mutations.....29

4.2.2 Synthetic Reference Tests 31

4.2.3 Specificity Results.....33

4.3 Drug Modules 34

4.3.1 Validation of Drug Modules.....34

4.3.2 Investigation of Drug Resistance Mechanisms on Cell Lines 35

4.4 Investigation of Drug Resistance Mechanisms on PDX Samples 38

Chapter 5: DISCUSSION.....43

Bibliography.....47

LIST OF TABLES

Table 4.1: Number of protein groups in total and in reference interactome.....27
Table 4.2: Applicable number of entities in transition rules.....28

LIST OF FIGURES

Figure 2.1: Conceptual representation of the GCA approach. 12
Figure 3.1: Overview of the GCA model on the molecular simulations 15
Figure 3.2: Scatter plot of expression values for gene RNF11 and ITCH1 16
Figure 3.3: Scatter plot of expression values for gene RNF11 and ITCH1 with smoothened estimation function. 17
Figure 3.4: Graphical representation of mutation applications 18
Figure 3.5: Graphical representation of drug module construction 21
Figure 3.6: Illustration of drug modules leveraged for predictions on test cell-lines. 22
Figure 3.7: Conceptual representation of network proximity 23
Figure 4.1: Enrichment of gene sets obtained from oncogene tumor suppressor mutations. 29
Figure 4.2: Distribution of similarity ratios. 31
Figure 4.3: Performance of model in synthetic networks 32
Figure 4.4: Results of specificity analysis performed on the model 33
Figure 4.5: Evaluation of model performance using 10-fold cross-validation 35
Figure 4.6: Representation of merged drug modules together with drug targets 37
Figure 4.7: Representation of merged drug module resistant exclusive gene enrichments 37
Figure 4.8: Evaluation of model performance on PDX samples. 38
Figure 4.9: Representation of merged PDX samples and intersection with resistant and sensitive drug modules of buparlisib. 39
Figure 4.10: Evaluation of model performance on PDX samples. 40

ABBREVIATIONS

AKT	A serine/threonine protein kinase
BIOGRID	Biological General Repository for Interaction Datasets
CA	Cellular Automata
CCLE	Cancer Cell Line Encyclopedia
CDKN2A	Cyclin dependent kinase inhibitor 2A
cGMP-PKG	cGMP-dependent protein kinase
CORUM	The Comprehensive Resource of Mammalian Protein Complexes
COSMIC	Catalogue Of Somatic Mutations In Cancer
CPTAC	Clinical Proteomic Tumor Analysis Consortium
DepMap	Dependency Map
DNA	Deoxyribonucleic Acid
ERK	Extracellular Signal Regulated Kinase
FGF1	Fibroblast Growth Factor 1
GCA	Graph-based Cellular Automata
GDSC	Genomics of Drug Sensitivity in Cancer
GNGT1	Guanine nucleotide-binding protein G(T) subunit gamma-T1
HMS LINCS	The Harvard Medical School Library of Integrated Network-based Cellular Signatures
IL3	Interleukin-3
Interactome INSIDER	INtegrated Structural Interactome and genomic Data browsER
iRefWeb	Interaction Reference Index
ITCH1	Itchy E3 Ubiquitin Protein Ligase 1
JAK/STAT	Janus Kinase/Signal Transducers and Activators of Transcription
KEGG	Kyoto Encyclopedia of Genes and Genomes
LCK	Lymphocyte-Specific Protein Tyrosine Kinase
MAPK	Mitogen-Activated Protein Kinase
MCC	Matthew's correlation coefficient
MTOR	Mammalian target of rapamycin
PARP1	Poly (ADP-ribose) polymerase 1
PDGFRA	Platelet-derived growth factor receptor alpha
PDX	Patient-drived Xenografts

PI3K	Phosphatidylinositol-3-kinase
PIK3C3	Phosphatidylinositol 3-kinase catalytic subunit type 3
PIK3CA	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha
PIK3CB	Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit beta
PIK3CD	Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit delta
PIK3CG	Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit gamma
PKIs	Protein kinase inhibitors
RNF11	RING Finger Protein 11
SOS2	Drosophila Son of sevenless 2
TCGA	The Cancer Genome Atlas
TFDP2	Transcription factor Dp-2
TRRUST	Transcriptional Regulatory Relationships Unraveled by Sentence-based Text mining
VEGFC	vascular endothelial growth factor C

Chapter 1:
INTRODUCTION

Mutations are genetic changes that occur in the DNA sequence of an organism. They arise spontaneously or be induced by external factors. While they play a crucial role in the evolution of biological organisms and serve a purpose, some of these changes lead to abnormalities (Fitzgerald & Rosenberg, 2019; Loewe & Hill, 2010). Some mutations occur in non-coding DNA regions known as 'Junk DNA,' which are thought to lack functionality (Palazzo & Gregory, 2014). Despite this assumption, these seemingly non-functional parts play a significant role either by acting as a buffer to reduce the likelihood of mutations occurring in the functional segments or having unknown regulatory functions (Palazzo & Gregory, 2014).

Beyond the specific type and location of a mutation on a gene, the effects of mutations on the intricate cellular interactions remain largely unpredictable (Loewe & Hill, 2010). In the context of cancer and its causal relationship with mutations, it becomes crucial to identify common patterns among cancer cells and compare them to normal cells. However, this task is highly challenging due to the diverse effects of mutations, the complexity of the cellular organization, and the accumulation of a wide range of mutation combinations (Salk, Fox, & Loeb, 2010).

Leveraging a wide range of omics data, genomics, transcriptomics, proteomics, and metabolomics is equally important to consider all aspects of cancer. These complementary layers offer valuable insights into cellular processes, elucidating the impact of mutations on gene expression, protein function, and overall pathways/complexes. In the last two decades, we have seen an explosion of high-throughput studies, resulting in vast amounts of omics data that present a puzzle to integrate and fully comprehend all intermediate steps from DNA to cellular activity. The challenge arises to integrating these omics data for identifying disease-related genes/proteins and exploring pathway perturbations in cancer. Despite its importance, this task remains challenging and requires further attention.(Demirel, Arici, & Tuncbag, 2022).

Targeted drug therapies aim to eliminate cancer cells from a healthy population of cells by leveraging their distinctive molecular characteristics (Padma, 2015). Unlike traditional

treatments like chemotherapy and radiation, side effects are mostly minimized since healthy cells are distinguished from cancer cells. These therapies achieve that by disrupting specific signaling pathways and molecular targets responsible for cancer growth and survival. For instance, Protein kinase inhibitors (PKIs) are a class of molecules that specifically bind to active sites of protein kinases, obstructing the process of phosphorylation, a vital step in initiating intracellular signaling cascades (Lemmon & Schlessinger, 2010). By inhibiting the phosphorylation of specific protein residues, these inhibitors promote cell cycle arrest at certain checkpoints, effectively restraining uncontrolled cell division, which is particularly relevant in the context of cancer cells (Otto & Sicinski, 2017).

A significant challenge arises as further molecular alterations accumulate in cancer cells, increasing heterogeneity and triggering drug resistance. This resistance can occur through various mechanisms, such as activating alternative signaling pathways, mutations in the drug's target site, increased drug efflux, or alterations in the tumor microenvironment (Housman et al., 2014; Mansoori, Mohammadi, Davudian, Shirjang, & Baradaran, 2017). The effects of these molecular alterations extend beyond the neighborhood of the altered gene or protein, propagating through the network of interactions and affecting the functionality of regulation mechanisms.

One promising computational approach is network medicine, which utilizes network science to represent these interactions (Sonawane, Weiss, Glass, & Sharma, 2019). The approach involves constructing and analyzing networks of molecular and cellular components from various layers of omics data (Barabasi, Gulbahce, & Loscalzo, 2011). The network construction process comprises several crucial steps, including reference construction, seed selection, propagation, and module detection. Detected modules represent meaningful subsets of the reference network determined through seed selection and propagation steps. Subsequently, these modules are further explored using network and learning-based methods (Sharma et al., 2015).

The current network medicine literature mainly focuses on obtaining valuable insights from static network models, but they overlook the dynamic nature of biological systems (Chaudhuri & Srivastava, 2022; Kostic, Hilgetag, & Tittgemeyer, 2020). The dynamic nature of biological systems arises from the fact that they constantly change over time due to various internal and external influences. This dynamic aspect is a crucial area that needs adequate attention (Abou-Jaoude et al., 2016; Braha, 2020). By examining the

interactions between nodes at different time points in the system's evolution, it can be possible to identify key patterns and processes that shape the system's dynamic behavior (Battiston et al., 2020). However, the need for real-time series data or its limited availability is a significant challenge. Alternatively, discrete dynamic models emerge as promising solutions. Discrete dynamic models leverage static data snapshots to simulate the dynamics of biological systems in discrete time steps. They use transition functions to represent the behavior of complex systems. One promising example is the cellular automata model (CA). CA models are simple yet powerful simulations of complex systems. They consist of a grid of cells that evolve based on predefined rules and their neighboring cell states. These models reveal patterns and emergent behaviors, aiding understanding self-organization and chaos in various fields (Ishida, 2018). Furthermore, the Graph-based Cellular Automata (GCA) model is an exciting variation of CA that uses a graph-based non-uniform neighborhood. It has been applied successfully in various fields but not in biology. Applying GCA to molecular networks could simulate complex cellular signaling pathways more flexibly and realistically, capturing dynamic interactions and providing insights into drug resistance.

In Chapter 2, we emphasized the prominent role of mutations and introduced the multi-omics efforts to comprehend different aspects of cancer progression. We also introduced the biological background of drug resistance and the computational background of network medicine, as well as the computational foundations of network medicine and dynamic network modeling, which we believe hold the potential to illuminate the mechanisms underlying drug resistance.

In Chapter 3, we provided the followed methodology in detail with proper references for databases utilized. We first described the construction of the reference network, then described the details of the GCA model. Afterward, we provided optimizations and how drug modules related to drug resistance were constructed, validated, and analyzed.

In Chapter 4, we summarized our key findings and compared our method with previous relevant studies. We also discussed our method's significance, limitations, strengths, and weaknesses. Additionally, we provide perspectives and future directions for this study.

Chapter 2:
LITERATURE REVIEW

2.1 Prominent Role of Mutations in Cancer Progression

In cancer biology, a fundamental hallmark distinguishing tumor cells from their normal counterparts is the loss of regulation over cellular signaling, resulting in uncontrollable proliferation (Hanahan & Weinberg, 2011). Cellular signaling is a tightly regulated process that governs the behavior of cells in response to internal and external cues (Nair, Chauhan, Saha, & Kubatzky, 2019). It involves molecular events, where proteins and genes interact with each other through signaling pathways that control essential cellular processes such as growth, differentiation, cell cycle, apoptosis (programmed cell death), and DNA repair (Campos & Clemente-Blanco, 2020; Duronio & Xiong, 2013; Elmore, 2007). This signaling is precisely regulated in normal cells, ensuring that cells respond appropriately to their environment and maintain homeostasis. However, this control is disrupted in cancer due to mutations that affect key genes and proteins involved in cellular signaling. These mutations may cause specific signaling proteins to become hyperactive, resulting in relentless cellular growth and division (Zenonos & Kyprianou, 2013). Alternatively, mutations might impair signaling pathways responsible for apoptosis, eliminating abnormal or damaged cells (Elmore, 2007). Moreover, mutations can compromise DNA repair pathways, contributing to genomic instability and the accumulation of further genetic alterations that fuel cancer progression (Torgovnick & Schumacher, 2015).

Cancer progression can be defined as a process with the accumulation of mutations that progressively worsen the malignancy of cancer cells. Uncontrolled division signals begin with initial mutations, and tens or hundreds of mutations accumulate from the beginning of this process. Distinguishing the effects of these mutations is crucial to understanding the nature of mutations in cancer progression to treat patients better. These mutations in cancer progression can be categorized into three groups based on their impact: drivers, passengers, and latent drivers (Kumar et al., 2020; McFarland et al., 2017). Driver mutations are genetic alterations that contribute to the development and progression of cancer (Ostroverkhova, Przytycka, & Panchenko, 2023). These mutations are called “drivers” because they provide a selective growth advantage to cells that carry them,

enabling them to outcompete normal cells and proliferate more rapidly. In contrast, passenger mutations are genetic changes that occur in cancer cells but do not directly contribute to the development or progression of the disease (Stratton, Campbell, & Futreal, 2009). Their primary characteristic lies in their frequent co-occurrence with the driver mutations. While passenger mutations are thought to have non-proliferative effects on disease phenotypes, their properties and role are poorly understood (Wodarz, Newell, & Komarova, 2018). However, accumulating slightly deleterious passenger mutations slow cancer progression, presenting a potential therapeutic target (McFarland et al., 2017). Finally, latent mutations can potentially drive cancer development, even though they have not yet been recognized as cancer hallmarks (Nussinov & Tsai, 2015; Yavuz, Tsai, Nussinov, & Tuncbag, 2023). Nonetheless, when they synergize with other mutations, they are believed to play a significant role in promoting cancer development and drug resistance (Nussinov & Tsai, 2015).

2.2 Molecular Landscape of Cancer

Cancer is a profoundly intricate disease, with mutation profiles being only one aspect of the investigation. It necessitates a comprehensive study of genetic, genomic, and molecular alterations within cancer cells. To address this necessity, significant efforts were made, and comprehensive initiatives and projects emerged, such as The Cancer Genome Atlas (TCGA) (Tomczak, Czerwinska, & Wiznerowicz, 2015), Cancer Cell Line Encyclopedia (CCLE) (Ghandi et al., 2019), and Clinical Proteomic Tumor Analysis Consortium (CPTAC) (Rudnick et al., 2016). These efforts involve the analysis of thousands of tumor samples using advanced genomic sequencing and molecular profiling technologies. The data they have deposited serve as invaluable resources for exploring the molecular landscape of cancer, ranging from cell lines to patient-derived samples, encompassing proteomics data, immunological landscapes, and drug responses.

These diverse resources have proven invaluable for providing powerful models to study mechanisms of cancer biology and personalized, targeted therapies. One major contribution of these efforts is the recognition that cancer is not a single disease but rather a collection of distinct subtypes (Zhang, Chen, & Creighton, 2023). In addressing this heterogeneity problem, Network-based stratification has emerged as a compelling method. To categorize tumor samples according to their specific characteristics, this

integrative method uses the genomic landscape of tumor samples and molecular gene networks (Hofree, Shen, Carter, Gross, & Ideker, 2013; C. Liu, Han, Zhang, Nussinov, & Cheng, 2021; S. Wang et al., 2018). Furthermore, several studies have leveraged perturbation-based methods to rewire cellular networks of cancer cells and integrate diverse omics data types to identify susceptible proteins and pathways (Acuner-Ozbabacan et al., 2014; Drake et al., 2016).

In addition to high-throughput studies, there is also an increase in public datasets enhancing our understanding of biological systems. Resources like KEGG and Wikipathways outline biological pathways. KEGG uses manual curation for pathway construction (Kanehisa, Furumichi, Sato, Kawashima, & Ishiguro-Watanabe, 2023), while WikiPathways relies on community inputs (Pico et al., 2008). The CORUM database catalogs verified mammal protein complexes from literature, experiments, and other databases (Giurgiu et al., 2019). The OmniPath integrates various molecular interactions—protein, signaling, and regulatory (Türei, Korcsmáros, & Saez-Rodriguez, 2016). The BIOGRID compiles diverse interaction types (e.g., physical association, direct interaction) (Oughtred et al., 2021). The TRRUST focuses on transcriptional networks composed of transcription factors and target genes (Han et al., 2018). These sources and many other offer complementary datasets, each capturing different facets of biological processes.

2.3 Drug Resistance in Cancer

Resistance to cancer drugs is one of the major obstacles to the successful treatment of cancer patients (X. Wang, Zhang, & Chen, 2019). The emergence of drug-resistant cancer cells can make therapies ineffective, leading to reduced patient survival rates (Housman et al., 2014). Several mechanisms proposed how drug resistance developed at the molecular level (Mashouri et al., 2019). As cancer cells rapidly divide and are prone to genetic instability, mutations can arise at a faster rate compared to normal cells (Yao & Dai, 2014). Each cell accumulates different combinations of mutations, leading to the formation of heterogeneous tumor cell populations. Furthermore, administrated drugs create selective pressure for the survival of resistant clones, leading to the expansion of resistant populations within the tumor (Friedman, 2016). An equally significant aspect of drug resistance concerns the timeframe of its emergence. Cancer samples can be

inherently resistant to drugs or acquire resistance after the drug treatment (Hata et al., 2016). Some drugs can significantly increase mutation rates, giving rise to resistant subpopulations that emerge and dominate the tumor sample (Kuczynski, Sargent, Grothey, & Kerbel, 2013; Kuosmanen et al., 2021). On the other hand, resistance may already be present in certain subpopulations (Greaves & Maley, 2012). To develop a thorough comprehension of the mechanisms driving cancer drug resistance, it is essential to consider these factors, including the time frame, order of accumulation, types of mutations, and whether these mutations are drug-induced or inherent.

Cancer drugs mainly interfere with the nature of these highly proliferative cells through different mechanisms, including alterations in drug targets, enhanced drug efflux, DNA repair pathways, and dysregulation of apoptosis (Housman et al., 2014). Resistant cells are believed to escape from these interventions through alternative mechanisms, allowing them to bypass the effects of the drugs (Boumahdi & de Sauvage, 2020). To counteract drug resistance, several approaches have been developed. Combinatorial drug therapies, for instance, have demonstrated success in overcoming drug resistance in several studies (Al-Lazikani, Banerji, & Workman, 2012; Jaaks et al., 2022; X. Wang et al., 2019). By targeting multiple pathways simultaneously or consecutively, these therapies aim to prevent the escape of resistant cells. Nonetheless, it is essential to acknowledge that this strategy may come with certain side effects and can be only effective for a short time, potentially leading to increased toxicity for the patient (Bozic et al., 2013; Felson, Anderson, & Meenan, 1994). To mitigate this issue, another elegant study suggested using multi drugs at lower concentrations to reduce the toxicity while increasing the efficacy (Fernandes Neto et al., 2020). Even though there are improvements, effective combinatorial drug therapies still require extensive knowledge of molecular mechanisms in specific cancer types.

Studying the complex nature of molecular alterations in the genome and proteome, it can be possible to reveal the molecular signatures of drug resistance (Le Tourneau, Borcoman, & Kamal, 2019; Malone, Oliva, Sabatini, Stockley, & Siu, 2020). Genetic mutations and their effects on protein level play a significant role in shaping this molecular signature since it is believed to have a causal relationship with drug resistance. Several studies have shown that mutations are critical in predicting drug resistance (Chapman et al., 2011; Lievre et al., 2006; Rodes et al., 2000; Z. Yang et al., 2017). Further identifying and understanding mutations' effects could help better forecast cancer

progression. To this end, computational models have also been successfully applied to understand the nature of mutations in cancer better (Y. Li et al., 2021; Mularoni, Sabarinathan, Deu-Pons, Gonzalez-Perez, & Lopez-Bigas, 2016; Sherman et al., 2022).

2.4 Network Medicine

Network medicine is an innovative and interdisciplinary medical research and systems biology approach. It is rooted in the understanding that many diseases and biological processes are not isolated entities but rather intricate systems governed by many factors and components. Traditional reductionist methods, which study individual elements of a system in isolation, may struggle to capture the complexity and interconnectedness of these systems (Ahn, Tewari, Poon, & Phillips, 2006). Diseases are modeled as interconnected molecular and cellular networks rather than individual isolated entities. The integration of high-throughput molecular omics datasets obtains this comprehensiveness. Integration makes it possible to construct maps of disease-specific networks called disease modules, which can be used to identify key regulators, pathways, and modules that contribute to the disease's development and progression.

Since the introduction of network medicine, numerous research studies have further validated its importance and utility (H. Chen et al., 2015; Choobdar et al., 2019; Hasin, Seldin, & Lusis, 2017). These studies have demonstrated how network medicine has been applied to various diseases, especially cancer. It gave remarkable results, particularly in biomarker identification for patients or subgroups. These biomarkers, representing unique molecular signatures, are essential for predicting resistant and sensitive drugs specific to patient groups (patient stratification) (Garnett et al., 2012). Identifying biomarkers, enhancing targeted therapy results, and reducing the risk of side effects (Cheng, Kovacs, & Barabasi, 2019a; Goetz & Schork, 2018). Drugs targeting the same or similar pathways can be repurposed for different diseases using disease-specific network modules, enabling fast and practical therapy alternatives (Morselli Gysi et al., 2021; Zhou et al., 2021). Moreover, network medicine can propose improvements in combinational drug therapies to enhance therapeutic efficacy (J. Li, Xu, & McIndoe, 2022). Network medicine can quantitatively assess drug-disease relations to offer successful combinatorial candidates and narrow the number of required experimental studies. This approach evaluates the proximity of drug targets and disease modules at the network level, assessing whether

two drugs targeting a common disease module while not targeting the same sub-modules or pathways, which indicates a higher chance of success in combinatorial therapy (Cheng, Kovacs, & Barabasi, 2019b)

2.5 Dynamic Network Modelling

Mutations and their relationship with drug resistance were mostly investigated with static models in network medicine. However, cells are dynamic systems, and static models do not entirely represent their dynamic nature. Dynamic network modeling can help identify the underlying mechanisms that drive the system's behavior, providing insights into the fundamental principles that govern complex biological systems (Budak, Eren Ozsoy, AYDIN SON, Can, & Tuncbag, 2015; Somvanshi & Venkatesh, 2014). Moreover, dynamic network modeling allows investigation of how perturbations or changes in the system affect its dynamic behavior (Albert, 2007; Di Cara, Garg, De Micheli, Xenarios, & Mendoza, 2007). This property is significant in biological systems, where gene mutations, drug treatments, or environmental changes can profoundly affect phenotypical outcomes. Dynamic modeling facilitates predicting better how it will respond to these perturbations and identifying potential targets for therapeutic intervention (Hemedan, Schneider, & Ostaszewski, 2023; Pappalardo et al., 2020).

High-throughput technologies comprise a broad spectrum of measurements that provide insights into the molecular mechanisms of cells, yielding a vast amount of multi-omics data. Our understanding of the cells is limited to these multi-omics datasets that represent snapshots of cellular states, often accompanied by a significant amount of noise. Although experimental methods for directly measuring dynamicity are lacking, we can enhance our insights by incorporating the time series dimension into the data, enabling the application of dynamic modeling approaches. Dynamic modeling enables the identification of critical transitional points, which helps to understand the sequence of events and uncover hidden patterns in cellular behavior. Discrete dynamic approaches have emerged as valuable tools in this context. By reconstructing the temporal trajectory of cells in pseudo-time, it can be possible to simulate cellular behavior together with perturbations and identify key regulatory points that might be critical for drug resistance, disease progression, or therapeutic interventions. Moreover, integrating different omics layers, such as genomics,

transcriptomics, and proteomics, to build dynamic models facilitates filling the gaps resulting from missing dynamical data.

Discrete dynamic models are mathematical models used to describe the behavior of systems that change over time in discrete time steps. Unlike continuous dynamic models, which use differential equations to represent continuous changes, discrete dynamic models divide time into pseudo time points and use difference equations or recursive formulas to capture changes after each step. These models involve state variables that define the system's current state, and a transition function determines how these variables change from one time step to the next. Discrete dynamic models find application in ecological population dynamics, computational algorithms, social network analysis, and more, providing valuable insights into complex systems and aiding in problem-solving across diverse domains (Erguler, 2018; Hunter, Krivitsky, & Schweinberger, 2012).

Cellular automata (CA) is a powerful class of discrete dynamic models used to simulate the behavior of complex dynamical systems, including biological systems. In CA models, a grid of cells represents the spatial environment, and each cell can take on different states. The state of a cell is updated over iterations based on transition rules. This inherent simplicity and the ability to capture complex behaviors make CA a suitable choice for dynamic network modeling in biology (Bandini, Mauri, & Serra, 2001). CA has been proven in biological applications such as tumor growth, metastasis, infectious disease spread simulations, the modeling of intercellular interactions, and even the evolution of skin scales (Dupin, Eyraud, Maurat, Sac-Epee, & Vallois, 2023; Manukyan, Montandon, Fofonjka, Smirnov, & Milinkovitch, 2017; Monteagudo & Santos, 2015; Pfeifer et al., 2008; Poleszczuk & Enderling, 2014; Prieto-Langarica, Kojouharov, Chen-Charpentier, & Tang, 2011; Reher, Klink, Deutsch, & Voss-Bohme, 2017; Valentim, Rabi, & David, 2023). Another similar modeling approach Petri nets used to represent and analyze the behavior of concurrent systems. They consist of places (representing states), transitions (representing events), and arcs (representing relationships between them). By depicting the flow of tokens (representing resources or events) through these elements, Petri nets help visualize, verify, and optimize the behavior of complex systems (Pinney, Westhead, & McConkey, 2003).

An exciting variation of CA is the graph-based CA (GCA) model, which employs a non-uniform neighborhood in a network instead of a uniform neighborhood in a lattice, as

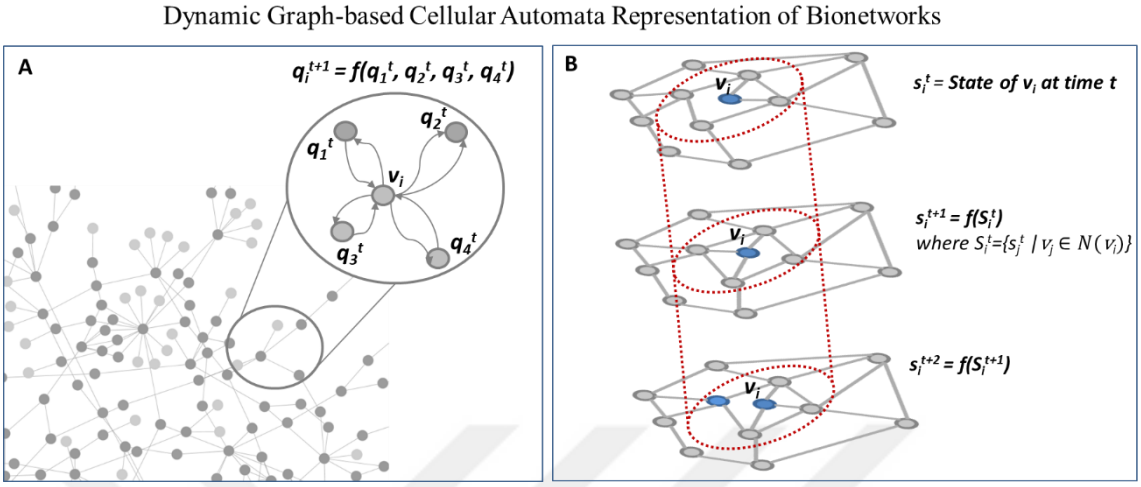

Figure 2.1: Conceptual representation of the GCA approach. (A) The state of a node in the next iteration is determined by the state of the nodes in its first neighborhood within the interactome. (B) Demonstrating how the activity of nodes in the first neighborhood of a selected node can evolve across iterations.

shown in Figure 2.1. GCA can simulate dynamic systems starting from initial static information and simple transition rules. These rules are intuitively defined based on the local relations among components in the system. GCA has been used in many different research fields, such as urban modeling (Barreira-González, Gómez-Delgado, & Aguilera-Benavente, 2015; Małecki, 2017; Nowak & Lewenstein, 1996), simulation of surface flows (Rinaldi, Dalponte, Vénere, & Clausse, 2012), and social network analysis (Małecki, Jankowski, & Rokita, 2013). However, it has not been successfully implemented for biological systems or molecular simulations. By applying this theory to molecular networks, it is likely to simulate and analyze complex cellular signaling pathways. The network structure of GCA allows for the representation of complex molecular interactions more flexibly than traditional lattice-based models. This flexibility is crucial in capturing biological systems' diverse and intricate nature, where molecules can interact in non-uniform and dynamic ways. Additionally, transition rules can capture the temporal dynamics of molecular interactions, providing a more realistic representation of cellular behavior. Perturbations on simulations may provide insights into how dynamic alterations in the genome or proteome affect signaling pathways and lead to drug resistance.

Chapter 3:
MATERIALS AND METHODS

3.1 Reference Network Construction

Since the aim of the project was to model intracellular molecular signaling, we constructed a reference network with mostly directed interactions annotated as stimulation and repression. First, we merged all edges in the OmniPath and TRRUST databases containing stimulation and repression interactions. We decided on conflicting cases for the same interaction based on the consensus column in the OmniPath database. Second, we determined the confidence scores of the edges according to the confidence scores from the HIPPIE (version 2.3) (Alanis-Lobato, Andrade-Navarro, & Schaefer, 2017), iRefWeb (Turner et al., 2010), STRING (Szklarczyk et al., 2021) and Intact databases (Orchard et al., 2014). In cases where there is more than one score for the same edge, we prioritized the HIPPIE database. The HIPPIE database was prioritized because the HIPPIE interactome scores have consistently demonstrated better performance compared to other alternatives (Arici & Tuncbag, 2021). We assigned the highest score in the other databases for the interactions not found in the HIPPIE. We gave them the highest value of 1.0 for the edges with no score information from any databases. We applied the filter only for the interactions in the STRING database in which we used "experimental" and "database" channels (conf > 0.7).

In addition to the directed interactions in the reference interactome, we added some high-confidence undirected interactions as bi-directed interactions. We also added interactions with a high confidence score (conf > 0.83) in the HIPPIE database, which is not directed but transition rules can be applied. The application of transition rules relies on two conditions: the coexistence within the same complex or pathway, and the presence of an edge in the BIOGRID database. We selected edges found in BIOGRID (version 4.4.212), KEGG (version 103.0), WikiPathways or CORUM (version 4.0) databases for compliance with transition rules. We considered only three interaction types: "Physical Association," "Direct Interaction," and "Association." from the BIOGRID database.

In the GCA model, since the state update of each node over the neighborhoods is applied according to the transition rules, we removed the isolated nodes and the self-edges from the reference network. We have also counted the number of protein/gene groups as

oncogenes, tumor suppressors, kinase, and transcription factors. For this purpose, we took the list of oncogenes from the "Oncogene Database" (Y. Liu, Sun, & Zhao, 2017), and the list of tumor suppressors from the "Tumor Suppressor Gene Database" (Zhao, Sun, & Zhao, 2013). We obtained tumor suppressors from the dataset of multiple databases such as Network of Cancer Genes (Repana et al., 2019), COSMIC (Sondka et al., 2018). We obtained kinases from the "KinHub" database (Eid, Turk, Volkamer, Rippmann, & Fulle, 2017). We retrieved the list of transcription factors from "The Human Transcription Factors" database (Lambert et al., 2018).

3.2 Graph-based Cellular Automata Model

3.2.1 Overview of the GCA Approach

We synchronously updated the activity values of genes based on weighted estimates of their neighbors through iterative nonlinear regression analysis. We used the DepMap cell line expression values as log(1+TPM), reflecting the respective activities, but they can be substituted with any other numeric measure that represents gene activities. We performed iterations 30 times by deactivating the edges according to mutations at the 15^{th} iteration. We detected the effect of mutations on the dynamic network structure in comparisons of simulations with and without mutations at the node level. We chose three sets of nodes that form subnetworks from the reference network. These subnetworks contain the nodes of effective gene mutations that lead to edge deactivation, isolated genes that lack any active edges, and genes that have partially lost their edges while displaying a significant (activity change > 0.585). Overview of the GCA model is represented in Figure 3.1.

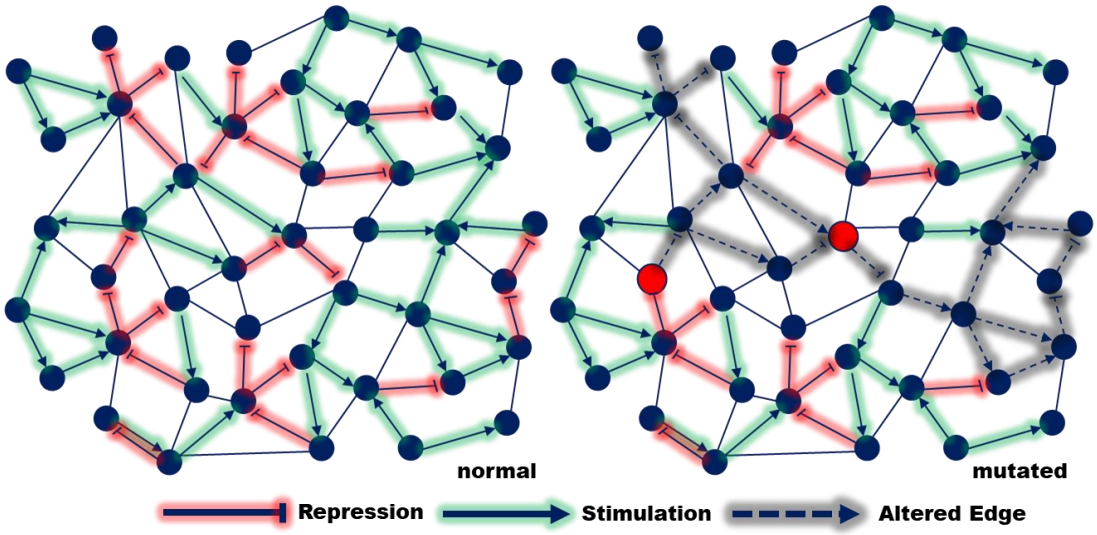

Figure 3.1: Overview of the GCA model on the molecular simulations. The left panel illustrates a mutation-free simulation of the network, where stimulation and repression edges are depicted in green and red, respectively. In the right panel, nodes with marked mutations are represented in red, and the downstream related edges are deactivated, indicated by the color gray.

3.2.2 Transition Rules of GCA

We used a regression analysis that can detect both linear and nonlinear relationships in the developed approach. We used activity values of genes, and directionality of interactions in OmniPath, and TRRUST databases in regression analysis. We selected at least 10% of the cell lines closest to any source gene activity value by a symmetric sliding window. We made estimation based on the distribution of activity values for the target gene within these selected cell lines by calculating the mean and standard deviation of the distribution. We estimated the activity value of the target gene as μ+2σ (μ=mean, σ=standard deviation) for stimulations and the μ-2σ for repressions.

$$y_{t+1} = (y_t + [\sum_1^m (x_i c_i e_i) / \sum_1^m (x_i c_i)])/2 \qquad (3.1)$$

In Equation (3.1), y_{t+1} is the activity value of the target node at time t+1, y_t is the amount of value at time t, m is the number of effective edges to the target node, x_i is the activity amount of the nodes affecting the target node, c_i is the confidence score of the edge, e_i is the predicted regression result. As a result, using this equation, we estimated a gene's activity value at time t+1 according to the activity value of neighboring genes at time t that stimulate or repress it.

3.2.3 Pairwise Estimation Function

In Figure 3.2, the estimation function generates estimations (green line) for the target node (ITCH1) at every activity value of the source node (RNF11). However, it is oversensitive since there is intrinsic noise and insignificant relative changes in the transcription profiles. We smoothened estimation function to improve efficiency and prevent oversensitivity issues by retaining estimations only for integer values, as shown in Figure 3.3. Target node state estimated by interpolating activity values between two adjacent integer points.

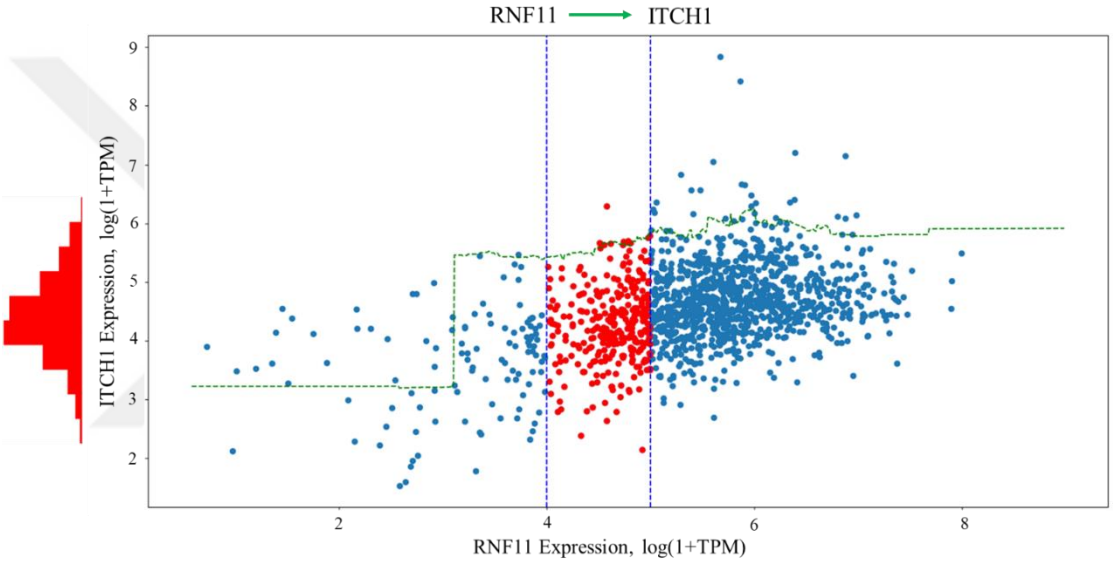

Figure 3.2: Scatter plot of expression values for gene RNF11 and ITCH1. Every dot represents a DepMap cell line. X-axis shows the expression value of source node (RNF11) while y-axis shows the expression value of target node (ITCH1). Vertical dashed blue lines showing the limits of symmetric sliding window which encapsulates %10 of the data points (red dotes). On the left, distribution of ITCH1 expression values of red dots represented. Green dashed line represents the estimation at each point based on the estimation function and stimulation information.

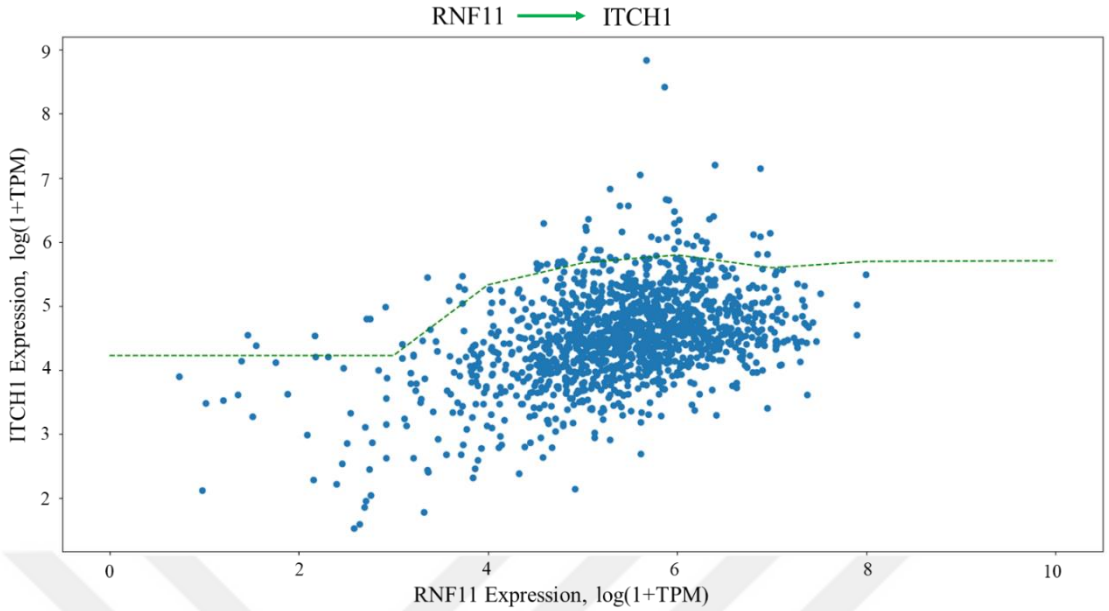

Figure 3.3: Scatter plot of expression values for gene RNF11 and ITCH1 with smoothened estimation function. Every dot represents a DepMap cell line. X-axis shows the expression value of source node (RNF1) while y-axis shows the expression value of target node (ITCH1). Green dashed line represents the estimation at each point based on the smoothened estimation function.

3.2.4 Use of Mutations in the GCA Model

Mutations change the primary interactions of proteins and propagate their effects through interaction networks. Therefore, from the immediate interaction partners of mutated proteins, we modified the downstream edges along the biological complex and pathways in which they are found. We determined effected pathways and complexes by considering the mutated node and its first neighbors. In each iteration, we progressively deactivate the next neighboring edges until no downstream edges are remained. We selected the first edges affected in the reference network according to the type of mutations. We propagated the effects of deleterious mutations (Frameshift, Early Stop Codon, Deleterious, etc.) over the reference network starting from all the edges where the protein interacts. On the other hand, we evaluated mutations that are not deleterious but cause a change in a single amino acid. In this case, instead of propagating from all interactions, we propagated over the affected edges according to the binding sites of the proteins. We used Interactome INSIDER database (Meyer et al., 2018) for binding site information. The application of mutations across the reference network is summarized in Figure 3.4.

The model identifies three distinct sets of nodes: 1) mutated nodes, 2) partially isolated nodes, and 3) completely isolated nodes. Mutated nodes correspond to the nodes where mutations have occurred, and edges altered. Partially isolated nodes refer to nodes that have lost some active edges and showed significant activity changes compared to not mutated simulations. Completely isolated nodes are nodes that have lost all their edges due to deactivation, and activity states cannot be updated anymore.

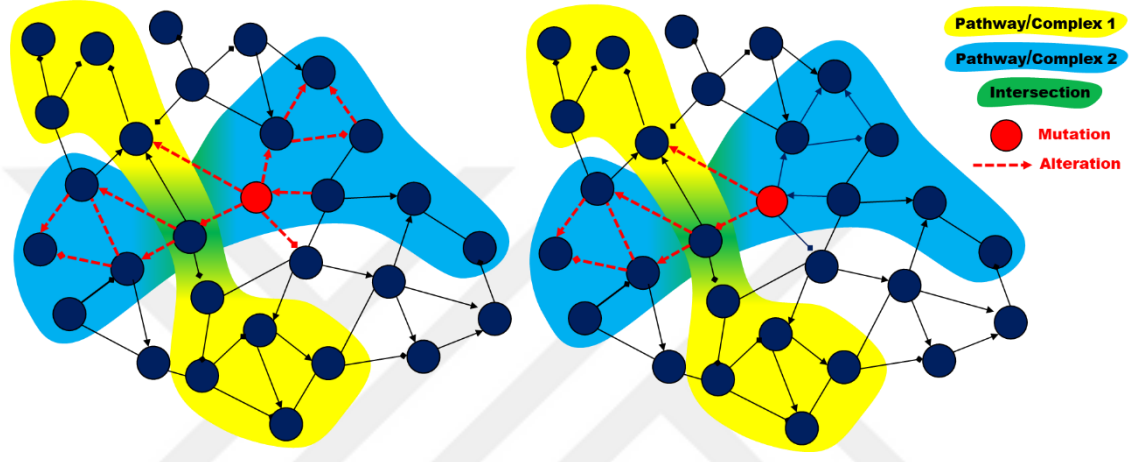

Figure 3.4: Graphical representation of mutation applications: Mutated genes are depicted in red, and the downstream effects are illustrated by dashed red lines. The left panel demonstrates deleterious mutation propagation, while the right panel represents point mutations. Mutations are constrained within their respective pathways or complexes and do not diffuse across different pathways or complexes.

3.3 Application of GCA Model on DepMap Cell Lines

In this section, we applied GCA Model to DepMap cell lines by using mutation profiles, transcription profiles, and drug responses from the DepMap database (Tsherniak et al., 2017)

3.3.1 Mutation Profiles of DepMap Cell Lines

We filtered out mutations that are not deleterious effect or happening at the interacting residues of proteins. Then, we applied deleterious and point mutations in the cell lines and run the simulations both with and without mutations for each cell line. Thus, we

measured the effect of mutations on signaling by looking at the difference between the two simulations at the gene expression level for each node.

3.3.2 Optimizations on GCA Model

We optimized the GCA model across multiple scenarios to examine its ability to differentiate mutation classes and individual mutations. To achieve this, we selected the cell line with the most mutations in the oncogene and tumor suppressor nodes (ACH-000998). Subsequently, we employed three different simulation approaches by applying mutations on different simulations. We considered only oncogene mutations in the first simulations, while in the second simulations, we examined all tumor suppressor mutations. Lastly, we individually applied all mutations for oncogenes, tumor suppressors, and all together. As a result, the optimization process focused on 61 tumor suppressors and 23 oncogene mutations. Enrichments of the different sets of mutations performed in EnrichR. For each mutation application, we assessed the similarity between resulting subnetworks by calculating the 'Jaccard Index' for subnetworks. We represented the distribution of values in a histogram after subtracting from 1, called the similarity score.

3.3.3 Synthetic Reference Networks

We generated synthetic networks using two different methods: the random Erdös-Renyi method and the Barabasi-Albert method. The Erdös-Renyi model is a random graph algorithm that generates random networks by adding edges with a constant probability 'p' between all pairs of nodes. To maintain consistency with the actual reference network, we set the value of 'p' to the number of edges divided by ten times the number of nodes in the reference network. This way, we obtained random networks with similar sizes of actual reference. On the other hand, the Barabasi-Albert algorithm follows a preferential attachment mechanism, meaning nodes with more connections are more likely to attract new links. The algorithm takes two numbers as input 'n' (number of nodes) and 'm' (Number of edges to attach from a new node to existing nodes). We set 'm' value by considering the actual reference by calculating the average degree in the reference.
Based on these two models, networks with different node numbers were incorporated, ranging from 200 to 20,000. We maintained the pathway, biological complex, and co-

expression information in the networks at ratios consistent with the actual data but fully randomized. We repeated each test hundred times with random variations and completely random 15 destructive and 5-point mutations. We examined the outputted number of isolated and partially isolated node numbers of the model.

3.3.4 Specificity Analysis

We performed the specificity analysis of the model across three scenarios to ensure the resulting subnodes are specific. We chose one of the ("ACH-000001") cell line, and assesed the occurrence of each resulting node in randomly generated scenarios. In the first scenario, we randomly selected mutated genes (initial seed nodes) from the entire reference network. We shuffled the node names within the reference network for the second scenario. For the third scenario, we swapped the edges connecting the nodes. Edge swapping keeps the original topology the same and only changes the neighbors of the proteins. In each scenario, we conducted 100 trials and analyzed the occurrence frequency of nodes in the obtained resulting subnetworks separately for each scenario.

3.3.5 DepMap Cell Line Simulations

The DepMap database contains mutation profiles of over 1000 cell lines, their responses to more than 4000 molecules, and various other omics datasets. We identified 951 cell lines suitable for simulations by intersecting the mutation and transcription profiles. We used this data to perform GCA simulations for cell lines by conducting 30 iterations per cell line. We determined the number of iterations by considering the activity states reached equilibrium before and after we applied the mutations. The increased number of iterations did not cause different results but increased the computational time cost, so we chose the minimum possible iteration number. We set the initial state of each gene in the simulation by computing the average expression values of cell lines within its lineage. We also conducted 30 simulations for each lineage type without mutations. Simulations with and without mutations were compared, resulting in a comparison of 681 cell lines because not all mutation profiles cause alteration at the edge level. In this node-level comparison, we created subnetworks for each cell line, which included a highly affected set of nodes regarding activity states.

3.3.6 Construction of Drug Modules

We used drug data from GDSC1 and GDSC2 datasets to determine drug resistance in cell lines. Drug-resistant cell lines had a z-value above 2.0, while sensitive cell lines had a z-value below -2.0 as suggested. After filtering out cell lines without sensitive and resistant counterparts for the same drug, we obtained 15,441 drug-cell line pairs (402 drugs and 956 cell lines). We combined the resistant and sensitive cell line subnetworks into consensus subnetworks by majority voting and removed common genes. This resulted in

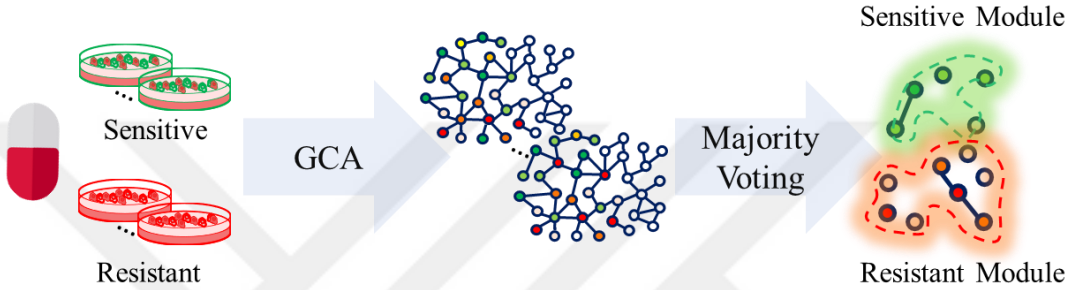

Figure 3.5: Graphical representation of drug module construction. Drug modules are built for each drug based on networks comprising both resistant and sensitive cell lines. Through majority voting at the node level, a consensus subnetwork is determined, while the differing subsets are extracted as modules.

two drug modules representing resistant and sensitive modules for each drug, as depicted in Figure 3.5.

3.3.7 Cell Line Blind Cross Validation of Drug Modules

We employed a systematic approach to validate the relationship between drug modules and drug resistance. We divided the complete dataset of cell lines into 'k' subsets. To construct drug modules, we repeatedly used the 'k-1' of these subsets 'k' times, while the remaining subset was reserved as the test set for cross-validation. For the case where 'k' was set to 10, the test set consisted of 68 cell lines, and the training set included 613. The number of drug modules generated also varied since the lack of test cell lines may be the only ones resistant or sensitive to a drug. Notably, we excluded the test cell line modules from generating any drug modules, ensuring an unbiased evaluation.

We considered network proximity measures to determine whether a test cell line exhibits resistance or sensitivity to a specific drug (Figure 3.6). These measures assess the closeness of a test cell line subnetwork to both the sensitive and resistant drug modules associated with the drug of interest. We predicted it to be resistant to the drug if a test cell line network demonstrated a closer distance to the resistant module. If it displayed a closer distance to the sensitive module, we predicted it to be sensitive to the drug. Additionally, we put lower than threshold for the z-scores to assess the significance of the closeness. Lower z-scores suggest significant proximity measures, so predictions made below determined z-score thresholds. We considered resistance as positive class and sensitives as negative class in the confusion matrix. We plotted the accuracy, precision, recall and MCC values as the measure drug modules prediction performance and reliability.

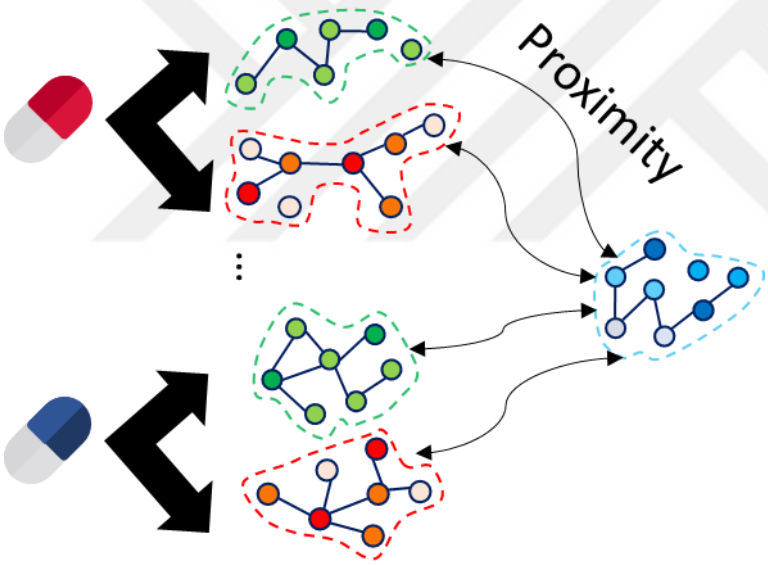

Figure 3.6: Illustration of drug modules leveraged for predictions on test cell lines. Predictions made for each drug cell line pair considering network proximity.

3.3.8 Network Proximity Analysis

In network proximity analysis, three numbers (distance, z-score, and p-value) are outputted for a given two subnetworks. There are different distance measures, but we used the best-performing 'closest' measure in this study. It represents the average shortest path length between the subnetworks and the nearest module protein. In simpler terms, it measures the average number of steps required to navigate from one subnetwork to the other through the nodes in the reference network. A shorter distance suggests a closer functional relationship or stronger connectivity between the two subnetworks, while a longer distance indicates a more distant or weaker connection. The z-score represents whether the distance between two node sets significantly differs from expectation in selected random subnetworks of similar size and topology. A negative z-score shows that the observed distance is smaller than the random expectation, whereas a positive z-score implies the opposite. The proximity approach is depicted in Figure 3.7.

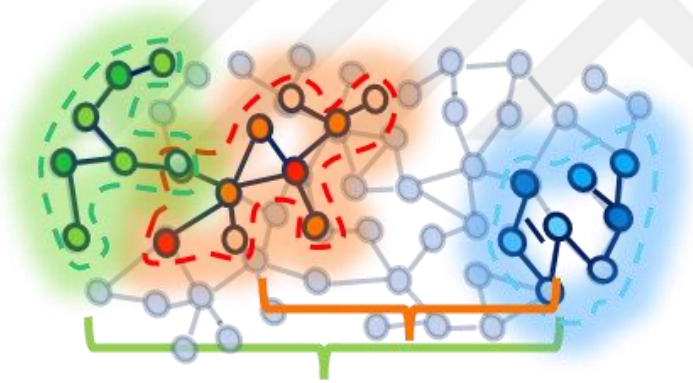

Figure 3.7: Conceptual representation of network proximity. Network proximity measures how two subnetworks close to each other within a reference network.

3.3.9 Drug Module Visualization and Investigating Resistance Mechanisms

We merged the most divergent sensitive and resistant drug modules and visualized to make observations. Drug modules normally do not contain common genes between sensitive and drug-resistant genes, but we included them for visualization purposes. Additionally, we expanded the gene sets with the drug's specific targets. The visualization focused on the giant component, representing the connected network with the most genes. Common genes, different genes, and drug targets were visualized in different colors in Cytoscape. We represented directional edges from the reference network to provide more

details. To further gain insights into generated drug modules, we performed enrichment analysis on resistant and sensitive drug modules by excluding common genes using EnrichR. We plotted pathway enrichment against -log10(adjusted p-values).

3.4 Application of the GCA Model to Patient Derived Xenografts

This section evaluates the optimized GCA model on an independent dataset focused on Patient-Derived Xenografts (PDX) as a case study.

3.4.1 PDX Samples

We utilized mutation profiles and drug responses of PDX samples from (Gao et al., 2015). We filtered the mutation data of the PDX samples and kept only deleterious 'Del0.8' and point mutations 'MutNovel'. We applied these mutations in the PDX samples in the simulations, conducting 30 iterations per PDX sample. Thus, we measured the effect of mutations on signaling by looking at the difference between these simulations and simulations without mutations at node level. We determined the initial state of each node in the simulation by calculating the average expression values of DepMap cell lines within its lineage. Comparison of simulations with and without mutations resulted in 369 PDX samples out of 399 because not all mutation profiles cause alteration at the edge level. In this node-level comparison, we created subnetworks for each PDX sample, which included a highly affected set of genes regarding activity state.

3.4.2 Drug Resistance Predictions on PDX Samples

This data set also includes drug responses for 37 drugs on PDX samples. The responses were categorized into four types: Complete Responses (CR), Partial Responses (PR), Stable Diseases (SD), and Progressive Diseases (PD). In this study, we only considered PD as resistant, while the rest were considered sensitive to the drug. Of the 37 drugs, five matched (Buparlisib, Ruxolitinib, Trametinib, LGK974, and Tamoxifen) with the drug modules generated using DepMap cell line datasets. We calculated proximity measures between PDX simulations and drug modules to make predictions. If a PDX module was closer to the resistant module than the sensitive module of a drug, we predicted it as resistant, and vice versa. We used the responses of PDX samples as ground truth and

evaluated the predictions accordingly. We made the predictions for 168 PDX samples across five drugs.

3.4.3 Investigation of Resistance Mechanisms in PDX Modules

We selected the best-predicted drug was buparlisib, and PDX samples correctly identified as resistant to buparlisib for visualization. The majority voting between the 82 PDX samples resulted in a consensus network and merged with resistance and sensitive modules of the buparlisib. We expanded the gene sets with the drug's specific targets named MTOR, PIK3CA, PIK3CB, PIK3CD, PIK3CG, and PIK3C3. The visualization focused on the giant component, representing the connected network with the most genes. We visualized PDX module genes, resistant module genes, sensitive module genes, and drug targets in different colors by using CytoScape. We also represented directional edges from the reference network to provide more details.

Chapter 4:
RESULTS

4.1 Data Statistics

We constructed a reference network consisting of 8,455 nodes and 62,649 edges with high reliability, reduced false-positive effect, and mostly from directional interactions. The reference consists of 43,123 stimulations, 12,715 repressions and 6,811 bi-directed edges. The number of proteins contained in each group of oncogenes, tumor suppressors, kinase, and transcription factors in the reference interactome are given in Table 4.1.

Table 4.1: Number of protein groups in total and in reference interactome

Protein Group	**Total Number**	**Number of Found in the Reference Interactome**
Oncogenes	803	651
Tumor Suppressors	1217	941
Transcription Factors	1639	1288
Kinases	518	491

Another piece of information that can be included in the transition rules in the GCA method is that their nodes are located in known biological pathways and complexes. We annotated pathways for 4,795 proteins and complexes for 4,252 proteins out of the 8,455 proteins in the reference interactome. In addition, we included 14,953 edges from the BIOGRID database. We also detected binding sites for 46.492 protein pairs out of 63,574 edges in the reference. Lastly, we obtained mutations that co-exist or are mutually exclusive (Tsherniak et al. 2017) using the mutation profiles of 1392 cell lines, a total of 953 (803 genes) mutations. We found 25,009 of 453,628 mutation pairs are co-occurring while 1,266 couples are mutually exclusive (Fisher's exact test; p-value <0.05). We find that this information was applicable in a total of 26,285 mutations. Mutual exclusivity and cooccurrence are currently not employed in the project, but they will be used to infer mutational orders. Transition rules-related statistics are represented in Table 4.2.

Table 4.2: Applicable number of entities in transition rules.

Transition Rule Name	**Applicable Number**
Same pathways (950)	4,795 proteins
Same complexes (956)	4,252 proteins
BIOGRID	14,953 interactions
Protein binding site	46.492 protein pairs
Co-occurring mutations	25,009 mutation pairs
Mutually exclusive mutations	1,266 mutation pairs

We filtered the mutation data of the cell lines in the DepMap database, and as a result, 589,736 mutations that caused changes in the amino acid level in 17,576 proteins were detected. While 60,250 mutations have deleterious effects, 18,399 correspond to the proteins' binding region. In addition, 189,839 mutations caused changes in the amino acid level, affecting 4,682 proteins across 399 samples in PDX samples.

4.2 Optimization of the Model

4.2.1 Sensitivity of Model to Differentiate Mutations

We examined the effect of mutations from different categories to test the ability of the model to distinguish different mutation classes, tumor suppressors, and oncogenes. A comparison between mutations on the tumor suppressor genes and oncogenes revealed that 77 genes present in the output set of the tumor suppressor simulations were absent in the oncogene simulations. While 83 genes found in the output of the oncogene simulations did not appear in the tumor suppressor simulations. Nonetheless, 119 nodes were common to both categories. To gain deeper insights, we analyzed the pathways where genes from different sets were enriched in EnrichR by excluding common genes observed in both (Figure 4.1). The effect of mutations on tumor suppressors or oncogenes varies, and the developed GCA model can represent this difference.

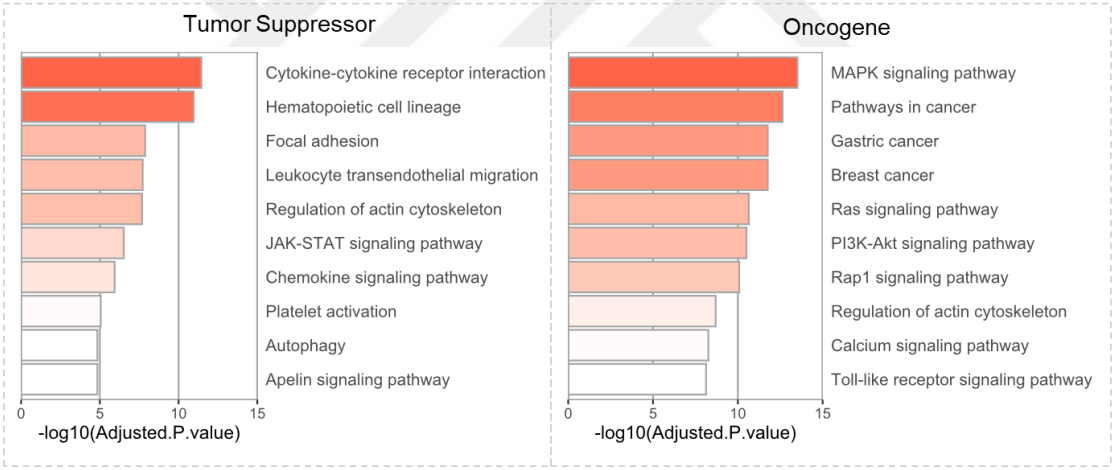

Figure 4.1: Enrichment of gene sets obtained from oncogene tumor suppressor mutations. Mutations on tumor suppressor genes simulated and resulting gene sets from the model enriched on the left. Mutations on oncogenes simulated and resulting gene sets enriched on the right. Significant and cancer related pathways were plotted with -log10 of adjusted p-values. Only unique gene sets of simulations were used in the enrichment analysis.

When we evaluate the enriched pathways involved in tumor suppressors, we can find relevant evidences that support these enrichments. Tumor suppressors are known to inhibit cell migration, and their reduced expression enhances migration. They interact with focal adhesion kinases and reduce tyrosine phosphorylation (Maziveyi, Dong,

Baranwal, & Alahari, 2018; Tamura et al., 1998). Tumor suppressor phosphatases negatively regulate cell-extracellular matrix interactions, exerting their tumor-suppressive function in this way (Pickup, Mouw, & Weaver, 2014; Sever & Brugge, 2015; Tamura et al., 1998). The JAK/STAT pathway is also crucial in human cancer, and it is activated by specific molecules and controlled by tumor suppressor genes (Amoyel, Anderson, & Bach, 2014).

In addition to tumor suppressor enrichments, oncogenic enrichments are also quite relevant. RAS genes encode small GTPase proteins in signal transduction pathways (Colicelli, 2004). Mutations in RAS genes, known as "activating mutations" or oncogenic mutations (Fernandez-Medarde & Santos, 2011; Prior, Lewis, & Mattos, 2012), cause constant activation of RAS proteins, leading to uncontrolled cell signaling, proliferation, and cancer growth. RAS specifically regulates the ERK/MAPK pathway (McCain, 2013). Activated ERK proteins translocate to the nucleus, triggering genes involved in cell division and survival (Santarpia, Lippman, & El-Naggar, 2012).

As an additional high-resolution test, we examined individual mutations from different classes. We applied the same mutations one by one, and the effect of each mutation was evaluated independently. We simulated a total of 84 mutations (61 tumor suppressors and 23 oncogenes), leading to the generation of 84 distinct subnetworks. Then, we compared these networks at the node level to assess how well the model reflected differences at the single mutation level (Figure 4.2). Despite having similarities, some mutations demonstrated distinct impacts on different regions within the interactome. This implies that certain mutations could have broader-reaching effects, simultaneously influencing multiple pathways or cellular processes. On the other hand, some mutations might be more specific, affecting only localized regions of the network.

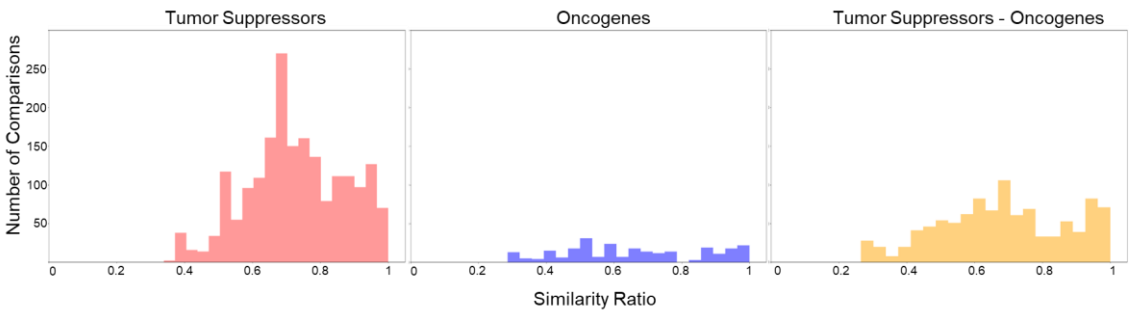

Figure 4.2: Distribution of similarity ratios. The histograms illustrate the similarity ratio distributions of GCA model networks resulting from simulations of individual mutations. The analysis compares networks generated from mutations on tumor suppressor genes (left), mutations on oncogenes (middle), and all mutations combined (right). Each mutation was applied separately, and the similarity ratio was calculated by subtracting the Jaccard distance from 1. A similarity ratio of 1 indicates that one network contains all nodes of the other subnetwork, while a ratio of 0 means there are no common nodes between the two subnetworks.

4.2.2 Synthetic Reference Tests

We tested the model on different-sized synthetic networks to ensure the scalability and limitations of the model. In network models, it's essential to assess their scalability to perform in smaller or larger networks. This ensures that the model is applicable in various real-world scenarios regarding network size and properties. Random Erdös-Renyi method does not guarantee to obtain connected networks, leading to the formation of isolated nodes or components forming. Thus, results generated by GCA in smaller networks are misleading and do not reflect the signaling abnormalities. As the network grows, the number of components decreases, and the probability of obtaining a connected network increase. The number of partially isolated is observed to be relatively high initially, eventually reaching equilibrium and converging to much lower rates. In networks with an Erdos Renyi topology, the model is inconsistent and cannot even be used in networks larger than 20,000 nodes. On the other hand, resulting network sizes are more consistent in the random networks generated by the Barabasi-Albert method. Interestingly, the propagation of mutations does not lead to any nodes becoming isolated, indicating the significance of the pathway and complex information in the system. The Barabasi method is also known for generating random networks having closer topologies to actual

biological networks. From these tests, three conclusions can be drawn: firstly, the GCA model is not reliable for disconnected networks or non-biological topology networks; secondly, scale-free networks, which are represented by the Barabási-Albert method, are better suited for measuring distortions in signaling by GCA model, and thirdly, pathway and complex structure throughout the reference networks plays a significant role in the model.

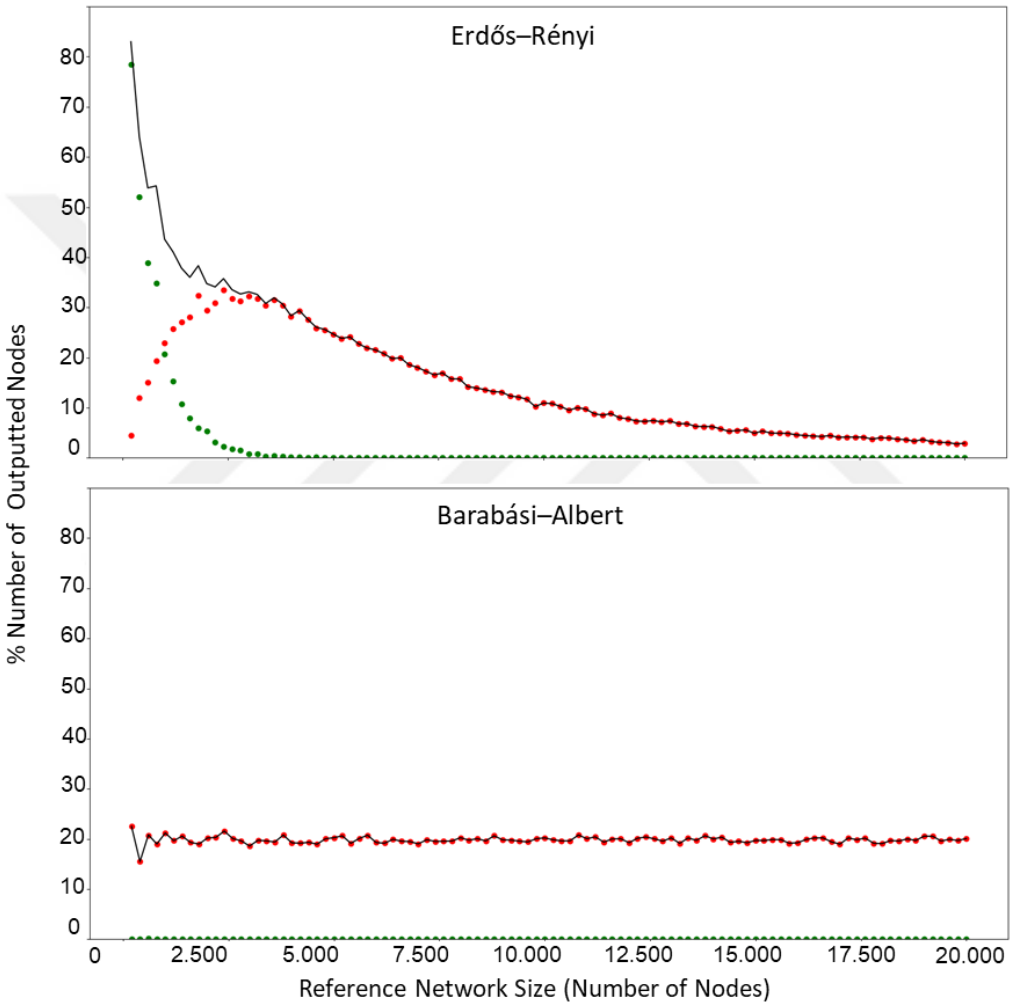

Figure 4.3: Performance of model in synthetic networks. Two different random network generation methods used to evaluate performance as Erdos-Renyi and Barabasi-Albert. Performance evaluated as the capability of producing subnetworks enriched by partially isolated (red) and isolated nodes (green). Black line represents the sum of red and green lines. X-axis represents number of nodes in generated random networks while y-axis represents percentage of output nodes relative to reference network nodes.

4.2.3 Specificity Results

The objective of specificity analysis is to evaluate the model's ability to produce distinct results. The accumulation of data points near lower occurrence values in Figure 4.4 indicates that the nodes in the resulting network are not commonly observed in random scenarios. Instead, they yield specific outcomes based on the initial nodes and the reference interactome utilized in the background. Specificity test results are summarized in Figure 4.4.

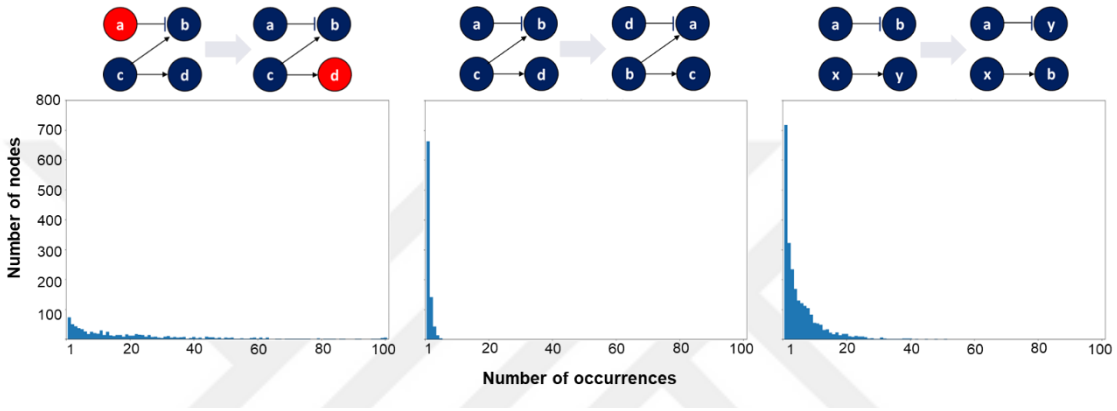

Figure 4.4: Results of specificity analysis performed on the model. In the first test, starting nodes (mutations) were randomly selected from the reference network. The second test involved randomly mixing the node names. In the third test, four nodes were randomly selected, and the edges were swapped. Each case was repeated 100 times, and the frequency of the nodes in the model's output was presented as a histogram.

In the first randomized seed selection scenario, small portion of genes happen to occur at most of the random simulations. These genes are not specific for a given seed gene set and happen to be in the result set independent of seed selection. In the second randomized label shuffling scenario, the resulting subset of nodes is quite different from each other, which implies that the reference network is quite significant and the place of genes in the reference network directly affects the results. In the third randomization of the edge swap scenario, generated resulting subnetworks can be considered distinct. Edge swap is a more controlled way of label shuffling, which preserves the topology and degree of each node in the graph. These randomization results suggest that the neighborhood of each gene is significant in each reference network. To efficiently use these specificity results, the frequency of occurrence within the network is normalized between 0 and 1. Then, the frequency is subtracted from 1 to calculate a specificity score for each node. Using this

score, the specific parts of the output of the model were determined. We used specificity scores to filter out non-specific results in actual simulations.

4.3 Drug Modules

We constructed drug modules from the DepMap cell lines by using GCA simulations and drug responses. In this section, we evaluated these graphs in detail.

4.3.1 Validation of Drug Modules

Cross-validation is crucial for developing and evaluating models. Its main purposes are to assess performance, generalize the model, and aid in hyperparameter tuning. By averaging evaluation over different data subsets, it reduces random variations, leading to a more reliable estimate of the model's performance. We used cross-validation to validate drug modules through their relationship with drug resistance.

We employed a 10-fold cross-validation approach to address the impact of majority voting between cell lines in the training set on the generation of drug modules. By adopting 10-fold cross-validation, we ensured that approximately 90% of the available cell lines were sufficient, leading to more accurate and specific drug modules. This approach effectively mitigated the risk of non-specific drug modules and provided comprehensive coverage of the resistance aspect.

The Matthew Correlation Coefficient (MCC) is a crucial metric for evaluating classification performance , particularly in imbalanced datasets (Chicco, Totsch, & Jurman, 2021). It considers all elements of the confusion matrix, including true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). MCC provides a more accurate measure of a model's ability to classify both positive and negative instances correctly. The MCC score ranges from -1 to +1, where +1 signifies a perfect classifier, 0 represents a random classifier, and -1 indicates a classifier that performs worse than random.

Our findings demonstrate a noteworthy trend: the model's performance improves as the z-score of proximity measurements decreases (Figure 4.5). Additionally, we observed interesting relationships across different folds: weak positive relationship (+.20 to +.29), moderate positive relationship (+.30 to +.39), and strong positive relationship (+.40 to

+.69) in terms of MCC value. Metrics, as depicted in Figure 4.5, underscore the critical impact of the z-score of the proximity measure on the model.

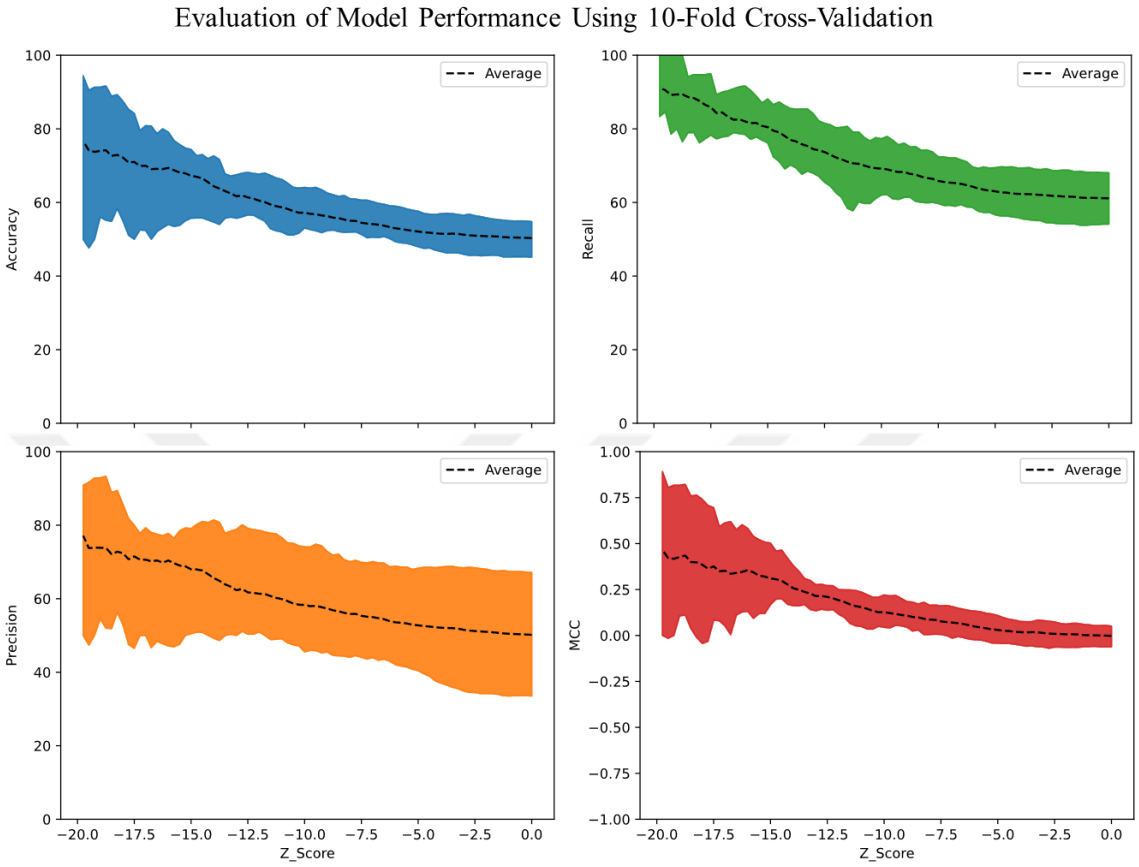

Figure 4.5: Evaluation of model performance using 10-fold cross-validation. The plot displays the performance metrics against the z-score of proximity measures. Accuracy is represented in blue on the top left, recall in green on the top right, precision in orange on the bottom left, and MCC (Matthews Correlation Coefficient) in red on the bottom right. The dashed lines indicate the average of 10 folds.

4.3.2 Investigation of Drug Resistance Mechanisms on Cell Lines

To conduct a comprehensive comparison analysis of a specific drug, we visualized a consensus network for the sensitive/resistant categories by using the GCA models of 39 drug-resistant and 8 sensitive cell lines in the simulation of the selected LCK (Lymphocyte-Specific Protein Tyrosine Kinase) inhibitor (JW-7-24-1). Proteins that are targets of drugs have also been used in visualization. The HMS LINCS (Vempati et al., 2014) (https://lincs.hms.harvard.edu/db/) database was used to identify drug targets.

Figure 4.6 shows that some nodes are grouped in the subnetworks of resistant cell lines. Drug target proteins were found to be closer to the nodes specific to sensitive cell lines. These nodes grouped in resistant cell lines may play an important role in drug resistance mechanisms. In addition, we analyzed pathways with only proteins from resistant cell lines and are shown in Figure 4.7.

One of the statistically significant enriching biological pathways, the MAPK (Mitogen-Activated Protein Kinase) signaling pathway is a protein network that regulates various cellular processes, including cell proliferation, differentiation, and survival. Abnormal activation of the MAPK pathway plays a role in many diseases, including cancer, and may contribute to the development of resistance to certain drugs or treatments by promoting tumor growth. Therefore, drug resistance and the MAPK signaling pathway have an important relationship with each other in cancer and other diseases, as activation of this pathway can lead to the development of resistance to certain drugs or treatments (Lee, Rauch, & Kolch, 2020). LCK inhibitors are important in the context of drug resistance and the MAPK signaling pathway. LCK is a tyrosine kinase that plays a critical role in the activation of the MAPK pathway (Bommhardt, Schraven, & Simeoni, 2019). Another pathway that is enriched in resistant cells is the apelin signaling pathway which is also important in the context of drug resistance (Uribesalgo et al., 2019). Apelin is a peptide hormone that regulates various biological processes, including cell proliferation, survival, and drug resistance. Apelin is a signaling molecule that binds to its receptor, APJ, activating various signaling pathways within the cell, including the PI3K/AKT and MAPK signaling pathways. Apelin has been shown to increase drug resistance in various types of cancer, such as breast cancer, lung cancer, and hepatocellular carcinoma (Y. Yang, Lv, Ye, & Zhang, 2016). Inhibition of the apelin/APJ signaling pathway has been proposed as a potential strategy to overcome drug resistance in cancer cells (L. Liu et al., 2021; Uribesalgo et al., 2019). Therefore, the Apelin signaling pathway is an important target for the development of new anti-cancer therapies. Finally, activation of the cGMP-PKG signaling pathway can inhibit PARP1, an enzyme that responds to DNA damage in cancer cells and induces cell death (Ba & Garg, 2011). Therefore, inhibition of the cGMP-PKG signaling pathway can be considered a potential strategy to reduce drug resistance in cancer cells and increase the efficacy of cancer treatments.

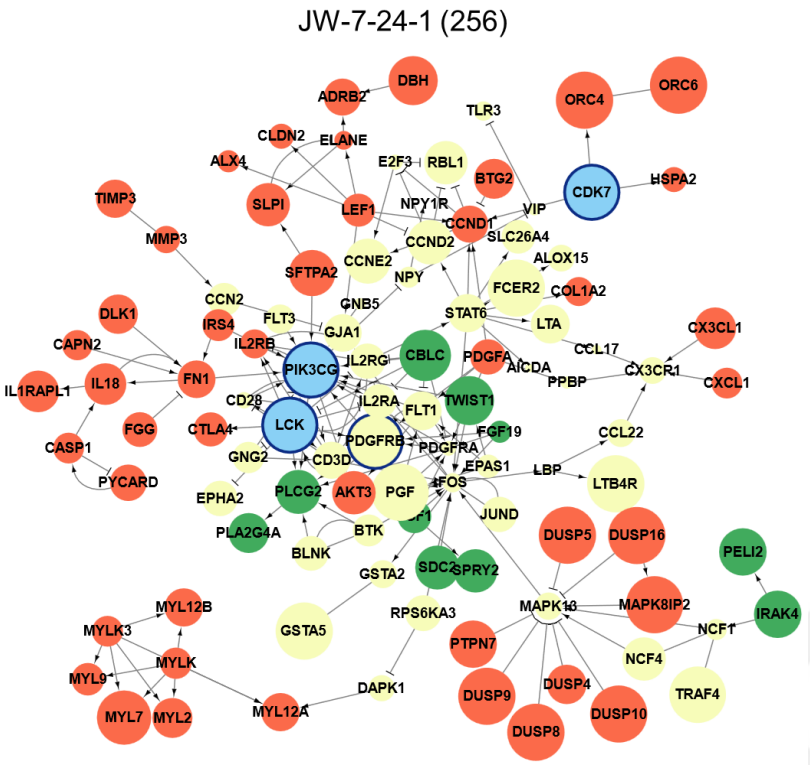

Figure 4.6: Representation of merged drug modules together with drug targets. Red nodes are unique to resistant cell lines, green ones are unique to sensitive cell lines and yellow ones are the common in both. Blue nodes are the drug targets of the drug JW-7-24-1 with the drug ID 256.

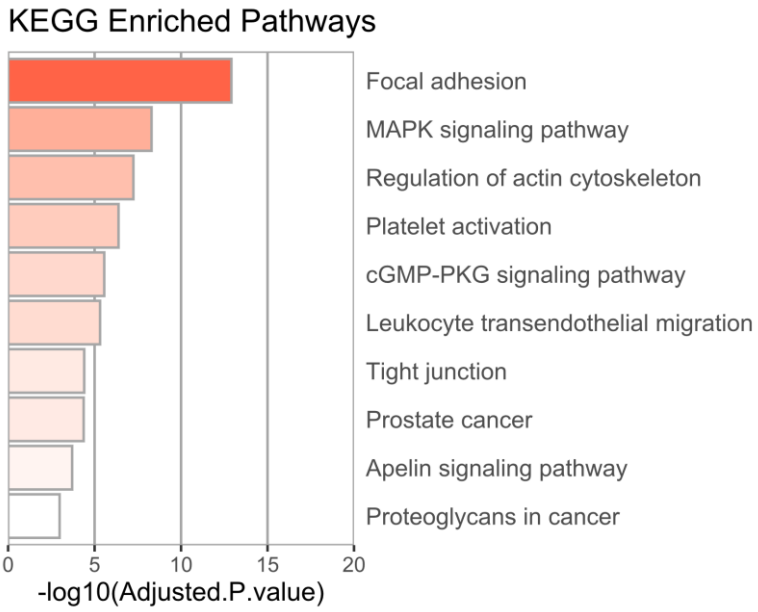

Figure 4.7: Representation of merged drug module resistant exclusive gene enrichments. Enrichment analysis for the red nodes showing the related pathways. Enrichment performed using EnrichR.

4.4 Investigation of Drug Resistance Mechanisms on PDX Samples

First, we predict the drug responses of the PDX samples and present the metrics in Figure 4.8. The metrics reveal the promising performance of the models in relating resistance with PDX samples. However, it is important to note that significant performance is mainly observed in predicting drug-resistant samples, as approximately 92% of the PDX samples are labeled as drug-resistant. This dataset imbalance and the models' limited ability to predict sensitive samples can be also attributed to the larger size of the predicted drug-resistant modules compared to sensitive modules. Additionally, number of predicted drug responses were 5 (Buparlisib, Ruxolitinib, Trametinib, LGK974, and Tamoxifen) and PDX were 168 and it gradually decreases as z-score threshold lowered.

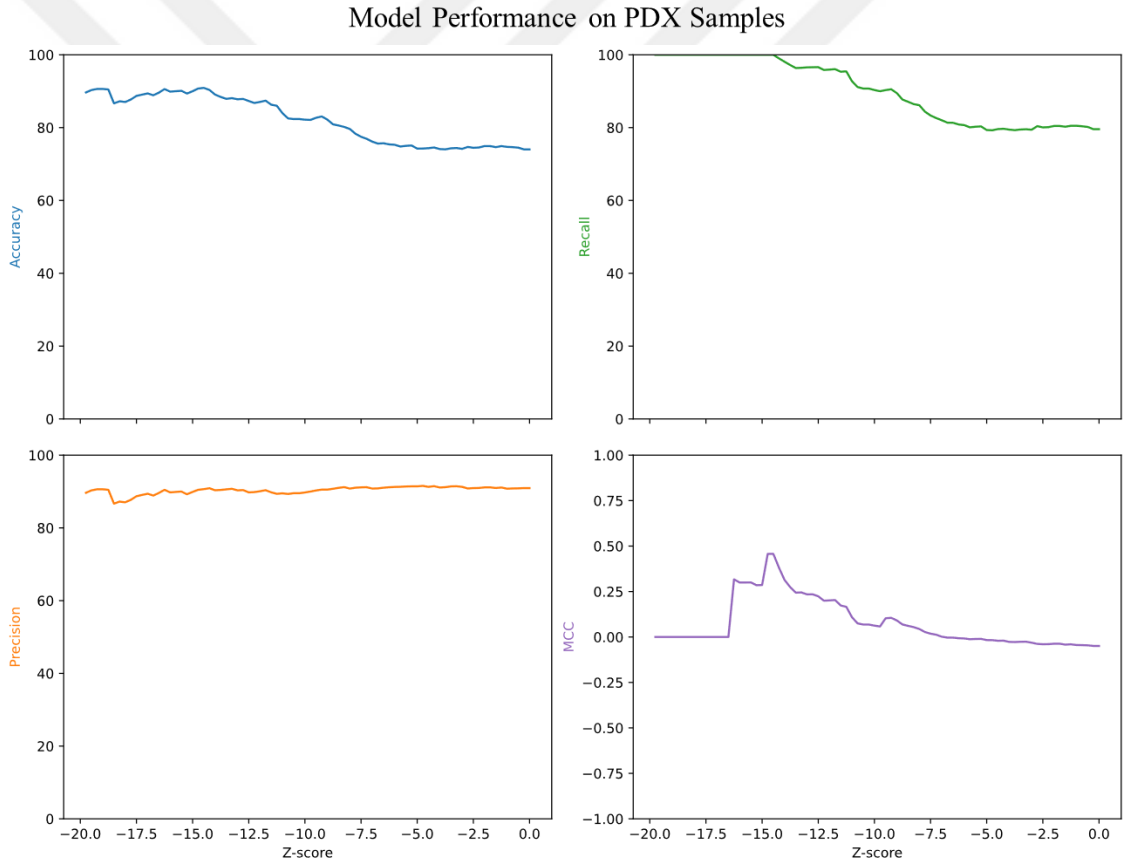

Figure 4.8: Evaluation of model performance on PDX samples. The plot displays the performance metrics against the z-score of proximity measures. Accuracy is represented in blue on the top left, recall in green on the top right, precision in orange on the bottom left, and MCC (Matthews Correlation Coefficient) in red on the bottom right.

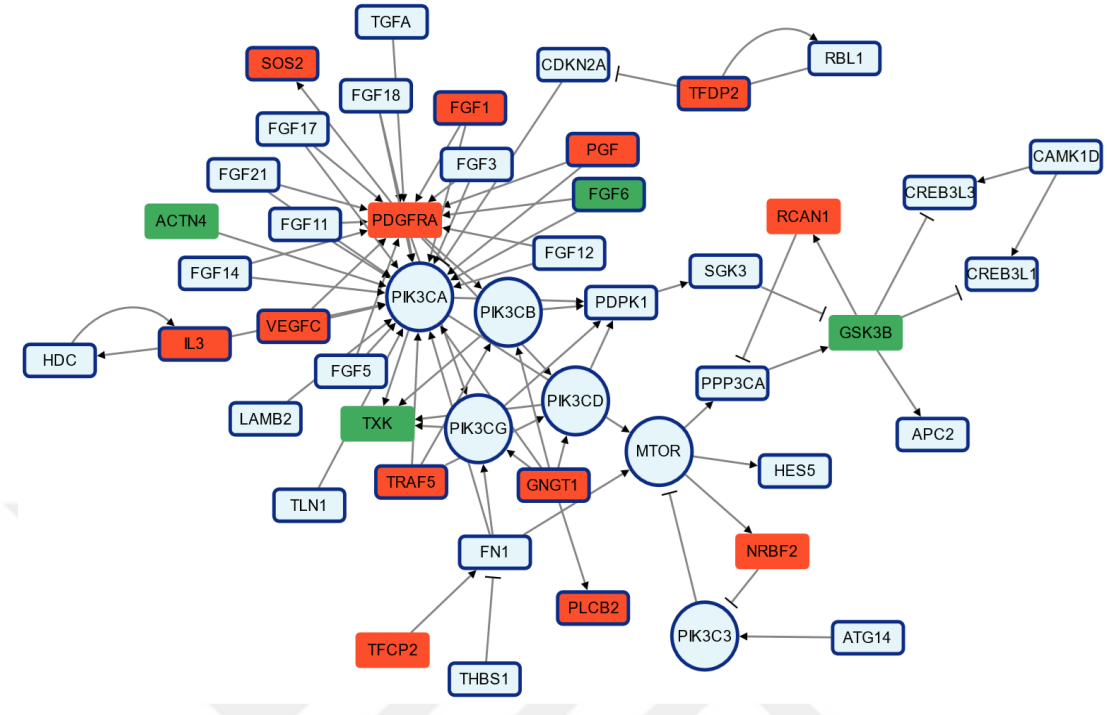

Figure 4.9: Representation of merged PDX samples and intersection with resistant and sensitive drug modules of buparlisib. Red nodes are unique to resistant module, green ones are unique to sensitive module and blue nodes are the ones seen in the majority of PDX samples. Sphere-shaped nodes are the drug targets of the drug buparlisib with the drug ID 1873.

We further examined the which subset of nodes found to be related to drug resistance. Buparlisib was chosen as a test case to study resistance mechanisms. To conduct a thorough comparison analysis of buparlisib, we created a consensus network for the 82 PDX samples accurately identified as drug-resistant by the GCA model. Figure 4.9 reveals that certain nodes in the resistant modules are enriched compared to the sensitive modules. We further examined these resistant module genes and analyzed pathways with proteins exclusively from the resistant modules intersected with PDX modules, as depicted in Figure 4.9.

These 7 out of 9 genes (SOS2, FGF1, PGF, TFDP2, VEGFC, IL3, and GNGT1) observed in the majority of PDX sample simulations and drug-resistant modules of buparlisib are known to be in PI3K-Akt signaling pathway. The activation of the PI3K/AKT pathway is related to drug resistance against buparlisib (He et al., 2021). Furthermore, we checked the activity states of these genes in the model through iterations (Figure 4.10). In comparison of mutated and not mutated states in 83 PDX samples, some of these genes

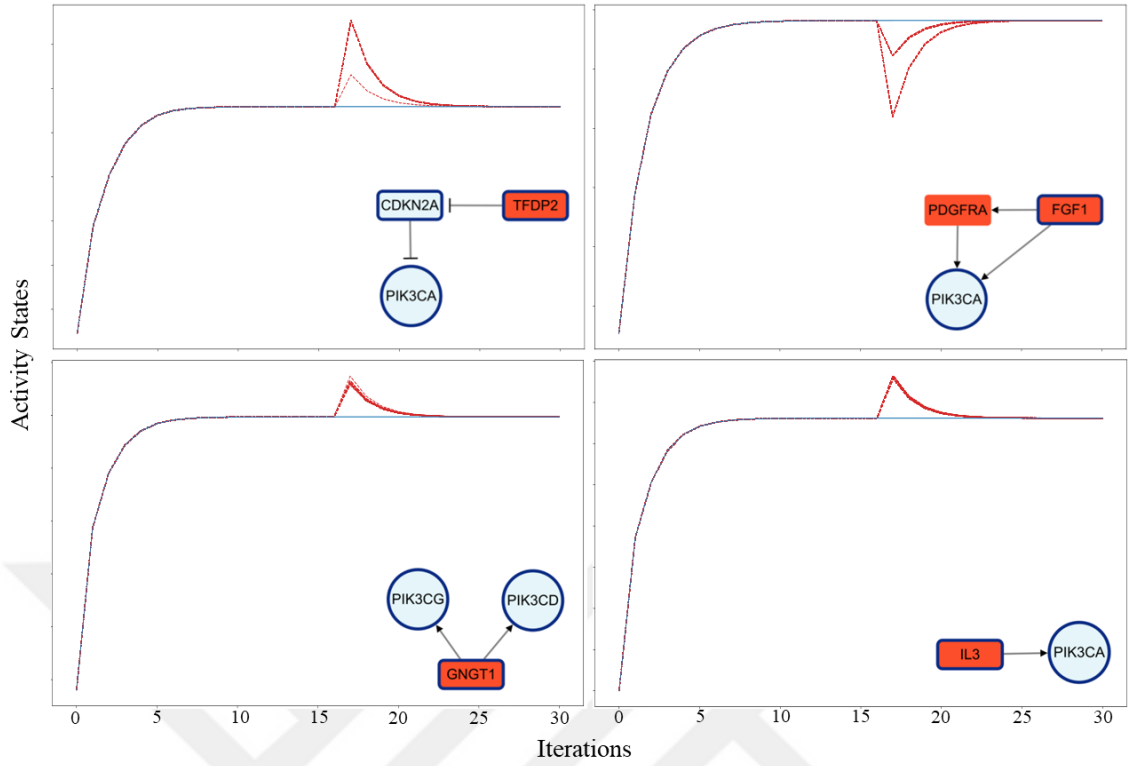

Figure 4.10: Activity changes of some detected genes from merged PDX samples and intersection with resistant and sensitive drug modules of buparlisib. Red nodes are unique to resistant module, and light-blue nodes are the ones seen in the majority of PDX samples. Sphere-shaped nodes are the drug targets of the drug buparlisib with the drug ID 1873. Blue lines represent the activity values without mutations and red line represent activity values at simulations with mutations. Values belong to the identified genes in both resistance modules and PDX simulations. On the right of each panel, association of each identified gene with its closest drug target depicted.

are shown to be upregulated or downregulated. One important observation is that they are not conflicting, and effects may vary between simulations. Associations of these affected proteins and targets of the buparlisib are also visualized in Figure 4.10. One such example is the upregulation of TFDP2 activity, it shows varying activity among the PDX simulations. Additionally, we can infer that it suppresses the tumor suppressor gene CDKN2A, thereby promoting the activity of PIK3CA. Conversely, the downregulation of FGF1, an activator of PIK3CA, contributes to its decreased activity. However, a secondary interaction with PDGFRA may still stimulate PIK3CA. Further activations of other proteins GNGT1, IL3 may suggest that there is an overall activation in the PIK3/AKT pathway. An additional insight reveals that the mutation effects (peaks)

diminish within five iterations for each gene in the context of this regression model. The heights of these peaks are constrained by the maximum activity value derived from expression profiles. Through averaging the prior temporal activity state and future state estimation, the fading-out phenomenon occurs within the span of 5 iterations.

Chapter 5:
DISCUSSION

Drug resistance is a significant challenge in cancer, hindering the effectiveness of targeted therapies due to accumulating mutations in dynamic cellular networks(X. Wang et al., 2019). To address this dynamicity, we adopted the GCA model, relying on a high-confidence reference interactome and using pairwise regression analysis on edges to simulate molecular signaling and perturbations. We propagated mutation effects downstream of signaling pathways and complexes. The model was optimized by using mutation landscapes, transcription profiles, and drug responses of well-characterized cell lines. We can differentiate individual mutation effects and different mutation classes (oncogenes, tumor suppressors) in the simulations. We created drug modules for each drug based on commonalities among simulation networks of drug-resistant and sensitive cell lines. We confirmed the reliability and accuracy of these drug modules by cross-validation and showed a case study in an independent PDX dataset. We identified specific pathways involving proteins from drug-resistant cell lines and PDX samples using these drug modules. These pathways were found to be closely associated with the resistance mechanisms of investigated drugs. We further analyzed the identified genes at a single-gene resolution, which is a crucial layer of detail to our comprehension of drug resistance. Single gene analysis further narrows down the set of possible targets that can be used in targeted therapies.

In the computational modeling of cancer, the focus lies on modeling cancer within the framework of the cancer microenvironment (Frieboes et al., 2009; Owen et al., 2011; Sun et al., 2012). These models incorporate cellular behaviors through mathematical equations. The major concern is that these approaches often neglect the intricate molecular-level complexity of cells and oversimplify their behaviors to just a handful of functions. In contrast, our proposed model considers the role of molecular networks in driving cancer cell behaviors. Instead of employing highly complex mathematical equations, we adopted simple local transition rules and extensive molecular networks. Additionally, there has been a growing popularity of learning-based drug resistance prediction algorithms (Daemen et al., 2013; Dong et al., 2015; Dorman et al., 2016; Menden et al., 2013). These algorithms primarily prioritize prediction efficiency, but they are often unable to provide a comprehensive understanding of the underlying mechanisms

driving drug resistance. Although our model may not achieve the same level of predictive metrics as these algorithms, our findings are more interpretable and provide more details of the underlying mechanisms.

Through comprehensive enrichment analysis of drug modules, we successfully identified several crucial pathways, including MAPK, apelin, and cGMP-PKG signaling pathways, known for their relevance to drug resistance. Moreover, our investigation has revealed specific associations between the investigated drug, LCK inhibitor (JW-7-24-1), and the enriched pathways. Further analysis of PDX simulations provided additional insights for another drug buparlisib and similarly, we have found strong evidence supporting the identified PI3K/AKT pathway and resistance to buparlisib (He et al., 2021). These pathways have already been recognized as targets to overcome drug resistance. Our novelty lies in narrowing down potential targets, enabling us to assess the impact of individual proteins. These findings guide the well-reasoned treatment strategies rooted in understanding the intricacies of drug resistance and pathway interactions.

While our developed approach exhibits notable strengths, we also recognize certain weaknesses that need to be addressed. Transcription profiles of cell lines were used as activity states, and they only represent one facet of the activity states of the proteins (Haider & Pal, 2013). We acknowledge the role of epigenetic regulations in influencing protein activity, so we plan to integrate additional omics data sets, particularly proteomics and phosphoproteomics, to better capture cellular signaling activities (C. Chen et al., 2023). We observed that conventional dimensionality reduction methods were insufficient in obtaining such improved activity profiles; instead, they compromised the model's robustness. Another important limitation of the model is to indirectly consider the factors outside of the molecular interactions like intercellular signaling, tumor microenvironment parameters, and metabolic activity. Addressing molecular aspects is, in fact, the opposite of general trends in the field.

Our findings also highlighted that the model not only distinguishes individual mutations or mutational profiles but also indicates that the order of mutation applications can yield slightly different results. Therefore, our future research will analyze the impact of mutation orders and delve deeper into these subtle differences. Additionally, we will refine the study's activity representation by integrating diverse omics layers, particularly by incorporating single-cell omics. The RNA velocity concept is relevant and applicable in this context since it extrapolates the present cellular state toward the future state (La

Manno et al., 2018). RNA velocity adds another dynamic perspective to our understanding of cellular processes and molecular dynamics. As a continuation of this study, we will also stratify patients using TCGA datasets based on molecular alterations and the interactome data of cancer patients. These simulations will be tailored to individual patients. While prior research categorized cancer types and tissue origins using TCGA data, we'll take a different approach (Hoadley et al., 2018; Hoadley et al., 2014). Instead of relying on isolated omics data and static pathways, we will classify the tumors based on dynamic networks. This way, we will construct a network-based taxonomy of the tumors within each cancer and cross-cancer type. We'll predict optimal treatment strategies by linking drug response patterns from cell line models with our network-based tumor classification.

Although we focused on drug resistance, extending this study to investigate molecular mechanisms driving metastasis or relevant biological questions believed to underlie molecular signaling abnormalities is possible. This methodology might even be generalized for identifying vulnerabilities across different types of networks having a scale-free topology. Successful adaptation requires in-depth domain knowledge and simplified transition rules to formulate the dynamicity of complex systems.

In conclusion, we adapted the GCA model to simulate molecular signaling and perturbations to reveal drug resistance mechanisms. Our study introduces a new perspective by integrating multi-omics data, network medicine, and discrete dynamic network modeling, which can lead to significant advancements in understanding and addressing the dynamicity of drug resistance and other medical challenges. The source code and materials can be accessed on GitHub at https://github.com/EnesSefaAyar/Graph-based_Cellular_Automata.

BIBLIOGRAPHY

Abou-Jaoude, W., Traynard, P., Monteiro, P. T., Saez-Rodriguez, J., Helikar, T., Thieffry, D., & Chaouiya, C. (2016). Logical Modeling and Dynamical Analysis of Cellular Networks. *Front Genet, 7*, 94. doi:10.3389/fgene.2016.00094

Acuner-Ozbabacan, E. S., Engin, B. H., Guven-Maiorov, E., Kuzu, G., Muratcioglu, S., Baspinar, A., . . . Nussinov, R. (2014). The structural network of Interleukin-10 and its implications in inflammation and cancer. *BMC Genomics, 15 Suppl 4*(Suppl 4), S2. doi:10.1186/1471-2164-15-S4-S2

Ahn, A. C., Tewari, M., Poon, C. S., & Phillips, R. S. (2006). The limits of reductionism in medicine: could systems biology offer an alternative? *PLoS Med, 3*(6), e208. doi:10.1371/journal.pmed.0030208

Al-Lazikani, B., Banerji, U., & Workman, P. (2012). Combinatorial drug therapy for cancer in the post-genomic era. *Nat Biotechnol, 30*(7), 679-692. doi:10.1038/nbt.2284

Alanis-Lobato, G., Andrade-Navarro, M. A., & Schaefer, M. H. (2017). HIPPIE v2.0: enhancing meaningfulness and reliability of protein-protein interaction networks. *Nucleic Acids Res, 45*(D1), D408-D414. doi:10.1093/nar/gkw985

Albert, R. (2007). Network inference, analysis, and modeling in systems biology. *Plant Cell, 19*(11), 3327-3338. doi:10.1105/tpc.107.054700

Amoyel, M., Anderson, A. M., & Bach, E. A. (2014). JAK/STAT pathway dysregulation in tumors: a Drosophila perspective. *Semin Cell Dev Biol, 28*, 96-103. doi:10.1016/j.semcdb.2014.03.023

Arici, M. K., & Tuncbag, N. (2021). Performance Assessment of the Network Reconstruction Approaches on Various Interactomes. *Front Mol Biosci, 8*, 666705. doi:10.3389/fmolb.2021.666705

Ba, X., & Garg, N. J. (2011). Signaling mechanism of poly(ADP-ribose) polymerase-1 (PARP-1) in inflammatory diseases. *Am J Pathol, 178*(3), 946-955. doi:10.1016/j.ajpath.2010.12.004

Bandini, S., Mauri, G., & Serra, R. (2001). Cellular automata: From a theoretical parallel computational model to its application to complex systems. *Parallel Computing, 27*(5), 539-553. doi:https://doi.org/10.1016/S0167-8191(00)00076-4

Barabasi, A. L., Gulbahce, N., & Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. *Nat Rev Genet, 12*(1), 56-68. doi:10.1038/nrg2918

Barreira-González, P., Gómez-Delgado, M., & Aguilera-Benavente, F. (2015). From raster to vector cellular automata models: A new approach to simulate urban growth with the help of graph theory. *Computers, Environment and Urban Systems, 54*, 119-131. doi:https://doi.org/10.1016/j.compenvurbsys.2015.07.004

Battiston, F., Cencetti, G., Iacopini, I., Latora, V., Lucas, M., Patania, A., . . . Petri, G. (2020). Networks beyond pairwise interactions: Structure and dynamics. *Physics Reports, 874*, 1-92. doi:https://doi.org/10.1016/j.physrep.2020.05.004

Bommhardt, U., Schraven, B., & Simeoni, L. (2019). Beyond TCR Signaling: Emerging Functions of Lck in Cancer and Immunotherapy. *Int J Mol Sci, 20*(14). doi:10.3390/ijms20143500

Boumahdi, S., & de Sauvage, F. J. (2020). The great escape: tumour cell plasticity in resistance to targeted therapy. *Nat Rev Drug Discov, 19*(1), 39-56. doi:10.1038/s41573-019-0044-1

Bozic, I., Reiter, J. G., Allen, B., Antal, T., Chatterjee, K., Shah, P., . . . Nowak, M. A. (2013). Evolutionary dynamics of cancer in response to targeted combination therapy. *Elife, 2*, e00747. doi:10.7554/eLife.00747

Braha, D. (2020). Patterns of ties in problem-solving networks and their dynamic properties. *Sci Rep, 10*(1), 18137. doi:10.1038/s41598-020-75221-3

Budak, G., Eren Ozsoy, O., AYDIN SON, Y., Can, T., & Tuncbag, N. (2015). Reconstruction of the temporal signaling network in Salmonella-infected human cells. *Frontiers in Microbiology, 6*. doi:10.3389/fmicb.2015.00730

Campos, A., & Clemente-Blanco, A. (2020). Cell Cycle and DNA Repair Regulation in the Damage Response: Protein Phosphatases Take Over the Reins. *Int J Mol Sci, 21*(2). doi:10.3390/ijms21020446

Chapman, P. B., Hauschild, A., Robert, C., Haanen, J. B., Ascierto, P., Larkin, J., . . . Group, B.-S. (2011). Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N Engl J Med, 364*(26), 2507-2516. doi:10.1056/NEJMoa1103782

Chaudhuri, S., & Srivastava, A. (2022). Network approach to understand biological systems: From single to multilayer networks. *J Biosci, 47*. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/36222127

Chen, C., Wang, J., Pan, D., Wang, X., Xu, Y., Yan, J., . . . Liu, G. P. (2023). Applications of multi-omics analysis in human diseases. *MedComm (2020), 4*(4), e315. doi:10.1002/mco2.315

Chen, H., Zhu, Z., Zhu, Y., Wang, J., Mei, Y., & Cheng, Y. (2015). Pathway mapping and development of disease-specific biomarkers: protein-based network biomarkers. *J Cell Mol Med, 19*(2), 297-314. doi:10.1111/jcmm.12447

Cheng, F., Kovacs, I. A., & Barabasi, A. L. (2019a). Network-based prediction of drug combinations. *Nat Commun, 10*(1), 1197. doi:10.1038/s41467-019-09186-x

Cheng, F., Kovacs, I. A., & Barabasi, A. L. (2019b). Publisher Correction: Network-based prediction of drug combinations. *Nat Commun, 10*(1), 1806. doi:10.1038/s41467-019-09692-y

Chicco, D., Totsch, N., & Jurman, G. (2021). The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Min, 14*(1), 13. doi:10.1186/s13040-021-00244-z

Choobdar, S., Ahsen, M. E., Crawford, J., Tomasoni, M., Fang, T., Lamparter, D., . . . Marbach, D. (2019). Assessment of network module identification across complex diseases. *Nat Methods, 16*(9), 843-852. doi:10.1038/s41592-019-0509-5

Colicelli, J. (2004). Human RAS superfamily proteins and related GTPases. *Sci STKE, 2004*(250), RE13. doi:10.1126/stke.2502004re13

Daemen, A., Griffith, O. L., Heiser, L. M., Wang, N. J., Enache, O. M., Sanborn, Z., . . . Gray, J. W. (2013). Modeling precision treatment of breast cancer. *Genome Biol, 14*(10), R110. doi:10.1186/gb-2013-14-10-r110

Demirel, H. C., Arici, M. K., & Tuncbag, N. (2022). Computational approaches leveraging integrated connections of multi-omic data toward clinical applications. *Mol Omics, 18*(1), 7-18. doi:10.1039/d1mo00158b

Di Cara, A., Garg, A., De Micheli, G., Xenarios, I., & Mendoza, L. (2007). Dynamic simulation of regulatory networks using SQUAD. *BMC Bioinformatics, 8*, 462. doi:10.1186/1471-2105-8-462

Dong, Z., Zhang, N., Li, C., Wang, H., Fang, Y., Wang, J., & Zheng, X. (2015). Anticancer drug sensitivity prediction in cell lines from baseline gene expression through recursive feature selection. *BMC Cancer, 15*, 489. doi:10.1186/s12885-015-1492-6

Dorman, S. N., Baranova, K., Knoll, J. H., Urquhart, B. L., Mariani, G., Carcangiu, M. L., & Rogan, P. K. (2016). Genomic signatures for paclitaxel and gemcitabine resistance in breast cancer derived by machine learning. *Mol Oncol, 10*(1), 85-100. doi:10.1016/j.molonc.2015.07.006

Drake, J. M., Paull, E. O., Graham, N. A., Lee, J. K., Smith, B. A., Titz, B., . . . Stuart, J. M. (2016). Phosphoproteome Integration Reveals Patient-Specific Networks in Prostate Cancer. *Cell, 166*(4), 1041-1054. doi:10.1016/j.cell.2016.07.007

Dupin, I., Eyraud, E., Maurat, E., Sac-Epee, J. M., & Vallois, P. (2023). Probabilistic cellular automata modelling of intercellular interactions in airways: complex pattern formation in patients with chronic obstructive pulmonary disease. *J Theor Biol, 564*, 111448. doi:10.1016/j.jtbi.2023.111448

Duronio, R. J., & Xiong, Y. (2013). Signaling pathways that control cell proliferation. *Cold Spring Harb Perspect Biol, 5*(3), a008904. doi:10.1101/cshperspect.a008904

Eid, S., Turk, S., Volkamer, A., Rippmann, F., & Fulle, S. (2017). KinMap: a web-based tool for interactive navigation through human kinome data. *BMC Bioinformatics, 18*(1), 16. doi:10.1186/s12859-016-1433-7

Elmore, S. (2007). Apoptosis: a review of programmed cell death. *Toxicol Pathol, 35*(4), 495-516. doi:10.1080/01926230701320337

Erguler, K. (2018). sPop: Age-structured discrete-time population dynamics model in C, Python, and R. *F1000Res, 7*, 1220. doi:10.12688/f1000research.15824.3

Felson, D. T., Anderson, J. J., & Meenan, R. F. (1994). The efficacy and toxicity of combination therapy in rheumatoid arthritis. A meta-analysis. *Arthritis Rheum, 37*(10), 1487-1491. doi:10.1002/art.1780371012

Fernandes Neto, J. M., Nadal, E., Bosdriesz, E., Ooft, S. N., Farre, L., McLean, C., . . . Bernards, R. (2020). Multiple low dose therapy as an effective strategy to treat EGFR inhibitor-resistant NSCLC tumours. *Nat Commun, 11*(1), 3157. doi:10.1038/s41467-020-16952-9

Fernandez-Medarde, A., & Santos, E. (2011). Ras in cancer and developmental diseases. *Genes Cancer, 2*(3), 344-358. doi:10.1177/1947601911411084

Fitzgerald, D. M., & Rosenberg, S. M. (2019). What is mutation? A chapter in the series: How microbes "jeopardize" the modern synthesis. *PLoS Genet, 15*(4), e1007995. doi:10.1371/journal.pgen.1007995

Frieboes, H. B., Edgerton, M. E., Fruehauf, J. P., Rose, F. R., Worrall, L. K., Gatenby, R. A., . . . Cristini, V. (2009). Prediction of drug response in breast cancer using integrative experimental/computational modeling. *Cancer Res, 69*(10), 4484-4492. doi:10.1158/0008-5472.CAN-08-3740

Friedman, R. (2016). Drug resistance in cancer: molecular evolution and compensatory proliferation. *Oncotarget, 7*(11), 11746-11755. doi:10.18632/oncotarget.7459

Gao, H., Korn, J. M., Ferretti, S., Monahan, J. E., Wang, Y., Singh, M., . . . Sellers, W. R. (2015). High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nat Med, 21*(11), 1318-1325. doi:10.1038/nm.3954

Garnett, M. J., Edelman, E. J., Heidorn, S. J., Greenman, C. D., Dastur, A., Lau, K. W., . . . Benes, C. H. (2012). Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature, 483*(7391), 570-575. doi:10.1038/nature11005

Ghandi, M., Huang, F. W., Jane-Valbuena, J., Kryukov, G. V., Lo, C. C., McDonald, E. R., 3rd, . . . Sellers, W. R. (2019). Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature, 569*(7757), 503-508. doi:10.1038/s41586-019-1186-3

Giurgiu, M., Reinhard, J., Brauner, B., Dunger-Kaltenbach, I., Fobo, G., Frishman, G., . . . Ruepp, A. (2019). CORUM: the comprehensive resource of mammalian protein complexes-2019. *Nucleic Acids Res, 47*(D1), D559-D563. doi:10.1093/nar/gky973

Goetz, L. H., & Schork, N. J. (2018). Personalized medicine: motivation, challenges, and progress. *Fertil Steril, 109*(6), 952-963. doi:10.1016/j.fertnstert.2018.05.006

Greaves, M., & Maley, C. C. (2012). Clonal evolution in cancer. *Nature, 481*(7381), 306-313. doi:10.1038/nature10762

Haider, S., & Pal, R. (2013). Integrated analysis of transcriptomic and proteomic data. *Curr Genomics, 14*(2), 91-110. doi:10.2174/1389202911314020003

Han, H., Cho, J. W., Lee, S., Yun, A., Kim, H., Bae, D., . . . Lee, I. (2018). TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res, 46*(D1), D380-D386. doi:10.1093/nar/gkx1013

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell, 144*(5), 646-674. doi:10.1016/j.cell.2011.02.013

Hasin, Y., Seldin, M., & Lusis, A. (2017). Multi-omics approaches to disease. *Genome Biol, 18*(1), 83. doi:10.1186/s13059-017-1215-1

Hata, A. N., Niederst, M. J., Archibald, H. L., Gomez-Caraballo, M., Siddiqui, F. M., Mulvey, H. E., . . . Engelman, J. A. (2016). Tumor cells can follow distinct evolutionary paths to become resistant to epidermal growth factor receptor inhibition. *Nat Med, 22*(3), 262-269. doi:10.1038/nm.4040

He, Y., Sun, M. M., Zhang, G. G., Yang, J., Chen, K. S., Xu, W. W., & Li, B. (2021). Targeting PI3K/Akt signal transduction for cancer therapy. *Signal Transduct Target Ther, 6*(1), 425. doi:10.1038/s41392-021-00828-5

Hemedan, A. A., Schneider, R., & Ostaszewski, M. (2023). Applications of Boolean modeling to study the dynamics of a complex disease and therapeutics responses. *Front Bioinform, 3*, 1189723. doi:10.3389/fbinf.2023.1189723

Hoadley, K. A., Yau, C., Hinoue, T., Wolf, D. M., Lazar, A. J., Drill, E., . . . Laird, P. W. (2018). Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell, 173*(2), 291-304 e296. doi:10.1016/j.cell.2018.03.022

Hoadley, K. A., Yau, C., Wolf, D. M., Cherniack, A. D., Tamborero, D., Ng, S., . . . Stuart, J. M. (2014). Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell, 158*(4), 929-944. doi:10.1016/j.cell.2014.06.049

Hofree, M., Shen, J. P., Carter, H., Gross, A., & Ideker, T. (2013). Network-based stratification of tumor mutations. *Nat Methods, 10*(11), 1108-1115. doi:10.1038/nmeth.2651

Housman, G., Byler, S., Heerboth, S., Lapinska, K., Longacre, M., Snyder, N., & Sarkar, S. (2014). Drug resistance in cancer: an overview. *Cancers (Basel), 6*(3), 1769-1792. doi:10.3390/cancers6031769

Hunter, D. R., Krivitsky, P. N., & Schweinberger, M. (2012). Computational Statistical Methods for Social Network Models. *J Comput Graph Stat, 21*(4), 856-882. doi:10.1080/10618600.2012.732921

Ishida, T. (2018). Possibility of Controlling Self-Organized Patterns with Totalistic Cellular Automata Consisting of Both Rules like Game of Life and Rules Producing Turing Patterns. *Micromachines (Basel), 9*(7). doi:10.3390/mi9070339

Jaaks, P., Coker, E. A., Vis, D. J., Edwards, O., Carpenter, E. F., Leto, S. M., . . . Garnett, M. J. (2022). Effective drug combinations in breast, colon and pancreatic cancer cells. *Nature, 603*(7899), 166-173. doi:10.1038/s41586-022-04437-2

Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M., & Ishiguro-Watanabe, M. (2023). KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res, 51*(D1), D587-D592. doi:10.1093/nar/gkac963

Kostic, D., Hilgetag, C. C., & Tittgemeyer, M. (2020). Unifying the essential concepts of biological networks: biological insights and philosophical foundations. *Philos Trans R Soc Lond B Biol Sci, 375*(1796), 20190314. doi:10.1098/rstb.2019.0314

Kuczynski, E. A., Sargent, D. J., Grothey, A., & Kerbel, R. S. (2013). Drug rechallenge and treatment beyond progression--implications for drug resistance. *Nat Rev Clin Oncol, 10*(10), 571-587. doi:10.1038/nrclinonc.2013.158

Kumar, S., Warrell, J., Li, S., McGillivray, P. D., Meyerson, W., Salichos, L., . . . Gerstein, M. B. (2020). Passenger Mutations in More Than 2,500 Cancer Genomes: Overall Molecular Functional Impact and Consequences. *Cell, 180*(5), 915-927 e916. doi:10.1016/j.cell.2020.01.032

Kuosmanen, T., Cairns, J., Noble, R., Beerenwinkel, N., Mononen, T., & Mustonen, V. (2021). Drug-induced resistance evolution necessitates less aggressive treatment. *PLoS Comput Biol, 17*(9), e1009418. doi:10.1371/journal.pcbi.1009418

La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., . . . Kharchenko, P. V. (2018). RNA velocity of single cells. *Nature, 560*(7719), 494-498. doi:10.1038/s41586-018-0414-6

Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., . . . Weirauch, M. T. (2018). The Human Transcription Factors. *Cell, 172*(4), 650-665. doi:10.1016/j.cell.2018.01.029

Le Tourneau, C., Borcoman, E., & Kamal, M. (2019). Molecular profiling in precision medicine oncology. *Nat Med, 25*(5), 711-712. doi:10.1038/s41591-019-0442-2

Lee, S., Rauch, J., & Kolch, W. (2020). Targeting MAPK Signaling in Cancer: Mechanisms of Drug Resistance and Sensitivity. *Int J Mol Sci, 21*(3). doi:10.3390/ijms21031102

Lemmon, M. A., & Schlessinger, J. (2010). Cell signaling by receptor tyrosine kinases. *Cell, 141*(7), 1117-1134. doi:10.1016/j.cell.2010.06.011

Li, J., Xu, H., & McIndoe, R. A. (2022). A novel network based linear model for prioritization of synergistic drug combinations. *PLoS One, 17*(4), e0266382. doi:10.1371/journal.pone.0266382

Li, Y., Burgman, B., Khatri, I. S., Pentaparthi, S. R., Su, Z., McGrail, D. J., . . . Yi, S. S. (2021). e-MutPath: computational modeling reveals the functional landscape of genetic mutations rewiring interactome networks. *Nucleic Acids Res, 49*(1), e2. doi:10.1093/nar/gkaa1015

Lievre, A., Bachet, J. B., Le Corre, D., Boige, V., Landi, B., Emile, J. F., . . . Laurent-Puig, P. (2006). KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer. *Cancer Res, 66*(8), 3992-3995. doi:10.1158/0008-5472.CAN-06-0191

Liu, C., Han, Z., Zhang, Z. K., Nussinov, R., & Cheng, F. (2021). A network-based deep learning methodology for stratification of tumor mutations. *Bioinformatics, 37*(1), 82-88. doi:10.1093/bioinformatics/btaa1099

Liu, L., Yi, X., Lu, C., Wang, Y., Xiao, Q., Zhang, L., . . . Guan, X. (2021). Study Progression of Apelin/APJ Signaling and Apela in Different Types of Cancer. *Front Oncol, 11*, 658253. doi:10.3389/fonc.2021.658253

Liu, Y., Sun, J., & Zhao, M. (2017). ONGene: A literature-based database for human oncogenes. *J Genet Genomics, 44*(2), 119-121. doi:10.1016/j.jgg.2016.12.004

Loewe, L., & Hill, W. G. (2010). The population genetics of mutations: good, bad and indifferent. *Philos Trans R Soc Lond B Biol Sci, 365*(1544), 1153-1167. doi:10.1098/rstb.2009.0317

Małecki, K. (2017). Graph Cellular Automata with Relation-Based Neighbourhoods of Cells for Complex Systems Modelling: A Case of Traffic Simulation. *Symmetry, 9*(12). doi:10.3390/sym9120322

Małecki, K., Jankowski, J., & Rokita, M. (2013, 2013//). *Application of Graph Cellular Automata in Social Network Based Recommender System.* Paper presented at the Computational Collective Intelligence. Technologies and Applications, Berlin, Heidelberg.

Malone, E. R., Oliva, M., Sabatini, P. J. B., Stockley, T. L., & Siu, L. L. (2020). Molecular profiling for precision cancer therapies. *Genome Med, 12*(1), 8. doi:10.1186/s13073-019-0703-1

Mansoori, B., Mohammadi, A., Davudian, S., Shirjang, S., & Baradaran, B. (2017). The Different Mechanisms of Cancer Drug Resistance: A Brief Review. *Adv Pharm Bull, 7*(3), 339-348. doi:10.15171/apb.2017.041

Manukyan, L., Montandon, S. A., Fofonjka, A., Smirnov, S., & Milinkovitch, M. C. (2017). A living mesoscopic cellular automaton made of skin scales. *Nature, 544*(7649), 173-179. doi:10.1038/nature22031

Mashouri, L., Yousefi, H., Aref, A. R., Ahadi, A. M., Molaei, F., & Alahari, S. K. (2019). Exosomes: composition, biogenesis, and mechanisms in cancer metastasis and drug resistance. *Mol Cancer, 18*(1), 75. doi:10.1186/s12943-019-0991-5

Maziveyi, M., Dong, S., Baranwal, S., & Alahari, S. K. (2018). Nischarin regulates focal adhesion and Invadopodia formation in breast cancer cells. *Mol Cancer, 17*(1), 21. doi:10.1186/s12943-018-0764-6

McCain, J. (2013). The MAPK (ERK) Pathway: Investigational Combinations for the Treatment Of BRAF-Mutated Metastatic Melanoma. *P T, 38*(2), 96-108. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/23599677

McFarland, C. D., Yaglom, J. A., Wojtkowiak, J. W., Scott, J. G., Morse, D. L., Sherman, M. Y., & Mirny, L. A. (2017). The Damaging Effect of Passenger Mutations on Cancer Progression. *Cancer Res, 77*(18), 4763-4772. doi:10.1158/0008-5472.CAN-15-3283-T

Menden, M. P., Iorio, F., Garnett, M., McDermott, U., Benes, C. H., Ballester, P. J., & Saez-Rodriguez, J. (2013). Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties. *PLoS One, 8*(4), e61318. doi:10.1371/journal.pone.0061318

Meyer, M. J., Beltran, J. F., Liang, S., Fragoza, R., Rumack, A., Liang, J., . . . Yu, H. (2018). Interactome INSIDER: a structural interactome browser for genomic studies. *Nat Methods, 15*(2), 107-114. doi:10.1038/nmeth.4540

Monteagudo, A., & Santos, J. (2015). Treatment Analysis in a Cancer Stem Cell Context Using a Tumor Growth Model Based on Cellular Automata. *PLoS One, 10*(7), e0132306. doi:10.1371/journal.pone.0132306

Morselli Gysi, D., do Valle, I., Zitnik, M., Ameli, A., Gan, X., Varol, O., . . . Barabasi, A. L. (2021). Network medicine framework for identifying drug-repurposing opportunities for COVID-19. *Proc Natl Acad Sci U S A, 118*(19). doi:10.1073/pnas.2025581118

Mularoni, L., Sabarinathan, R., Deu-Pons, J., Gonzalez-Perez, A., & Lopez-Bigas, N. (2016). OncodriveFML: a general framework to identify coding and non-coding regions with cancer driver mutations. *Genome Biol, 17*(1), 128. doi:10.1186/s13059-016-0994-0

Nair, A., Chauhan, P., Saha, B., & Kubatzky, K. F. (2019). Conceptual Evolution of Cell Signaling. *Int J Mol Sci, 20*(13). doi:10.3390/ijms20133292

Nowak, A., & Lewenstein, M. (1996). Modeling Social Change with Cellular Automata. In R. Hegselmann, U. Mueller, & K. G. Troitzsch (Eds.), *Modelling and Simulation in the Social Sciences from the Philosophy of Science Point of View* (pp. 249-285). Dordrecht: Springer Netherlands.

Nussinov, R., & Tsai, C. J. (2015). 'Latent drivers' expand the cancer mutational landscape. *Curr Opin Struct Biol, 32*, 25-32. doi:10.1016/j.sbi.2015.01.004

Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., . . . Hermjakob, H. (2014). The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res, 42*(Database issue), D358-363. doi:10.1093/nar/gkt1115

Ostroverkhova, D., Przytycka, T. M., & Panchenko, A. R. (2023). Cancer driver mutations: predictions and reality. *Trends Mol Med, 29*(7), 554-566. doi:10.1016/j.molmed.2023.03.007

Otto, T., & Sicinski, P. (2017). Cell cycle proteins as promising targets in cancer therapy. *Nat Rev Cancer, 17*(2), 93-115. doi:10.1038/nrc.2016.138

Oughtred, R., Rust, J., Chang, C., Breitkreutz, B. J., Stark, C., Willems, A., . . . Tyers, M. (2021). The BioGRID database: A comprehensive biomedical resource of

curated protein, genetic, and chemical interactions. *Protein Sci, 30*(1), 187-200. doi:10.1002/pro.3978

Owen, M. R., Stamper, I. J., Muthana, M., Richardson, G. W., Dobson, J., Lewis, C. E., & Byrne, H. M. (2011). Mathematical modeling predicts synergistic antitumor effects of combining a macrophage-based, hypoxia-targeted gene therapy with chemotherapy. *Cancer Res, 71*(8), 2826-2837. doi:10.1158/0008-5472.CAN-10-2834

Padma, V. V. (2015). An overview of targeted cancer therapy. *Biomedicine (Taipei), 5*(4), 19. doi:10.7603/s40681-015-0019-4

Palazzo, A. F., & Gregory, T. R. (2014). The case for junk DNA. *PLoS Genet, 10*(5), e1004351. doi:10.1371/journal.pgen.1004351

Pappalardo, F., Russo, G., Pennisi, M., Parasiliti Palumbo, G. A., Sgroi, G., Motta, S., & Maimone, D. (2020). The Potential of Computational Modeling to Predict Disease Course and Treatment Response in Patients with Relapsing Multiple Sclerosis. *Cells, 9*(3). doi:10.3390/cells9030586

Pfeifer, B., Kugler, K., Tejada, M. M., Baumgartner, C., Seger, M., Osl, M., . . . Tilg, B. (2008). A cellular automaton framework for infectious disease spread simulation. *Open Med Inform J, 2*, 70-81. doi:10.2174/1874431100802010070

Pickup, M. W., Mouw, J. K., & Weaver, V. M. (2014). The extracellular matrix modulates the hallmarks of cancer. *EMBO Rep, 15*(12), 1243-1253. doi:10.15252/embr.201439246

Pico, A. R., Kelder, T., van Iersel, M. P., Hanspers, K., Conklin, B. R., & Evelo, C. (2008). WikiPathways: pathway editing for the people. *PLoS Biol, 6*(7), e184. doi:10.1371/journal.pbio.0060184

Pinney, J. W., Westhead, D. R., & McConkey, G. A. (2003). Petri Net representations in systems biology. *Biochem Soc Trans, 31*(Pt 6), 1513-1515. doi:10.1042/bst0311513

Poleszczuk, J., & Enderling, H. (2014). A High-Performance Cellular Automaton Model of Tumor Growth with Dynamically Growing Domains. *Appl Math (Irvine), 5*(1), 144-152. doi:10.4236/am.2014.51017

Prieto-Langarica, A., Kojouharov, H., Chen-Charpentier, B., & Tang, L. (2011). A Cellular Automata Model of Infection Control on Medical Implants. *Appl Appl*

Math, 6(1), 1-10. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/23543851

Prior, I. A., Lewis, P. D., & Mattos, C. (2012). A comprehensive survey of Ras mutations in cancer. *Cancer Res, 72*(10), 2457-2467. doi:10.1158/0008-5472.CAN-11-2612

Reher, D., Klink, B., Deutsch, A., & Voss-Bohme, A. (2017). Cell adhesion heterogeneity reinforces tumour cell dissemination: novel insights from a mathematical model. *Biol Direct, 12*(1), 18. doi:10.1186/s13062-017-0188-z

Repana, D., Nulsen, J., Dressler, L., Bortolomeazzi, M., Venkata, S. K., Tourna, A., . . . Ciccarelli, F. D. (2019). The Network of Cancer Genes (NCG): a comprehensive catalogue of known and candidate cancer genes from cancer sequencing screens. *Genome Biol, 20*(1), 1. doi:10.1186/s13059-018-1612-0

Rinaldi, P., Dalponte, D., Vénere, M., & Clausse, A. (2012). Graph-based cellular automata for simulation of surface flows in large plains. *Asian Journal of Applied Science, 5*(4), 224-231.

Rodes, B., Holguin, A., Soriano, V., Dourana, M., Mansinho, K., Antunes, F., & Gonzalez-Lahoz, J. (2000). Emergence of drug resistance mutations in human immunodeficiency virus type 2-infected subjects undergoing antiretroviral therapy. *J Clin Microbiol, 38*(4), 1370-1374. doi:10.1128/JCM.38.4.1370-1374.2000

Rudnick, P. A., Markey, S. P., Roth, J., Mirokhin, Y., Yan, X., Tchekhovskoi, D. V., . . . Stein, S. E. (2016). A Description of the Clinical Proteomic Tumor Analysis Consortium (CPTAC) Common Data Analysis Pipeline. *J Proteome Res, 15*(3), 1023-1032. doi:10.1021/acs.jproteome.5b01091

Salk, J. J., Fox, E. J., & Loeb, L. A. (2010). Mutational heterogeneity in human cancers: origin and consequences. *Annu Rev Pathol, 5*, 51-75. doi:10.1146/annurev-pathol-121808-102113

Santarpia, L., Lippman, S. M., & El-Naggar, A. K. (2012). Targeting the MAPK-RAS-RAF signaling pathway in cancer therapy. *Expert Opin Ther Targets, 16*(1), 103-119. doi:10.1517/14728222.2011.645805

Sever, R., & Brugge, J. S. (2015). Signal transduction in cancer. *Cold Spring Harb Perspect Med, 5*(4). doi:10.1101/cshperspect.a006098

Sharma, A., Menche, J., Huang, C. C., Ort, T., Zhou, X., Kitsak, M., . . . Barabasi, A. L. (2015). A disease module in the interactome explains disease heterogeneity, drug response and captures novel pathways and genes in asthma. *Hum Mol Genet, 24*(11), 3005-3020. doi:10.1093/hmg/ddv001

Sherman, M. A., Yaari, A. U., Priebe, O., Dietlein, F., Loh, P. R., & Berger, B. (2022). Genome-wide mapping of somatic mutation rates uncovers drivers of cancer. *Nat Biotechnol, 40*(11), 1634-1643. doi:10.1038/s41587-022-01353-8

Somvanshi, P. R., & Venkatesh, K. V. (2014). A conceptual review on systems biology in health and diseases: from biological networks to modern therapeutics. *Syst Synth Biol, 8*(1), 99-116. doi:10.1007/s11693-013-9125-3

Sonawane, A. R., Weiss, S. T., Glass, K., & Sharma, A. (2019). Network Medicine in the Age of Biomedical Big Data. *Front Genet, 10*, 294. doi:10.3389/fgene.2019.00294

Sondka, Z., Bamford, S., Cole, C. G., Ward, S. A., Dunham, I., & Forbes, S. A. (2018). The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat Rev Cancer, 18*(11), 696-705. doi:10.1038/s41568-018-0060-1

Stratton, M. R., Campbell, P. J., & Futreal, P. A. (2009). The cancer genome. *Nature, 458*(7239), 719-724. doi:10.1038/nature07943

Sun, X., Zhang, L., Tan, H., Bao, J., Strouthos, C., & Zhou, X. (2012). Multi-scale agent-based brain cancer modeling and prediction of TKI treatment response: incorporating EGFR signaling pathway and angiogenesis. *BMC Bioinformatics, 13*, 218. doi:10.1186/1471-2105-13-218

Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., . . . von Mering, C. (2021). The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res, 49*(D1), D605-D612. doi:10.1093/nar/gkaa1074

Tamura, M., Gu, J., Matsumoto, K., Aota, S., Parsons, R., & Yamada, K. M. (1998). Inhibition of cell migration, spreading, and focal adhesions by tumor suppressor PTEN. *Science, 280*(5369), 1614-1617. doi:10.1126/science.280.5369.1614

Tomczak, K., Czerwinska, P., & Wiznerowicz, M. (2015). The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Pozn), 19*(1A), A68-77. doi:10.5114/wo.2014.47136

Torgovnick, A., & Schumacher, B. (2015). DNA repair mechanisms in cancer development and therapy. *Front Genet, 6*, 157. doi:10.3389/fgene.2015.00157

Tsherniak, A., Vazquez, F., Montgomery, P. G., Weir, B. A., Kryukov, G., Cowley, G. S., . . . Hahn, W. C. (2017). Defining a Cancer Dependency Map. *Cell, 170*(3), 564-576 e516. doi:10.1016/j.cell.2017.06.010

Türei, D., Korcsmáros, T., & Saez-Rodriguez, J. (2016). OmniPath: guidelines and gateway for literature-curated signaling pathway resources. *Nature Methods, 13*(12), 966-967. doi:10.1038/nmeth.4077

Turner, B., Razick, S., Turinsky, A. L., Vlasblom, J., Crowdy, E. K., Cho, E., . . . Wodak, S. J. (2010). iRefWeb: interactive analysis of consolidated protein interaction data and their supporting evidence. *Database (Oxford), 2010*, baq023. doi:10.1093/database/baq023

Uribesalgo, I., Hoffmann, D., Zhang, Y., Kavirayani, A., Lazovic, J., Berta, J., . . . Penninger, J. M. (2019). Apelin inhibition prevents resistance and metastasis associated with anti-angiogenic therapy. *EMBO Mol Med, 11*(8), e9266. doi:10.15252/emmm.201809266

Valentim, C. A., Rabi, J. A., & David, S. A. (2023). Cellular-automaton model for tumor growth dynamics: Virtualization of different scenarios. *Comput Biol Med, 153*, 106481. doi:10.1016/j.compbiomed.2022.106481

Vempati, U. D., Chung, C., Mader, C., Koleti, A., Datar, N., Vidovic, D., . . . Schurer, S. C. (2014). Metadata Standard and Data Exchange Specifications to Describe, Model, and Integrate Complex and Diverse High-Throughput Screening Data from the Library of Integrated Network-based Cellular Signatures (LINCS). *J Biomol Screen, 19*(5), 803-816. doi:10.1177/1087057114522514

Wang, S., Ma, J., Zhang, W., Shen, J. P., Huang, J., Peng, J., & Ideker, T. (2018). Typing tumors using pathways selected by somatic evolution. *Nat Commun, 9*(1), 4159. doi:10.1038/s41467-018-06464-y

Wang, X., Zhang, H., & Chen, X. (2019). Drug resistance and combating drug resistance in cancer. *Cancer Drug Resist, 2*(2), 141-160. doi:10.20517/cdr.2019.10

Wodarz, D., Newell, A. C., & Komarova, N. L. (2018). Passenger mutations can accelerate tumour suppressor gene inactivation in cancer evolution. *J R Soc Interface, 15*(143). doi:10.1098/rsif.2017.0967

Yang, Y., Lv, S. Y., Ye, W., & Zhang, L. (2016). Apelin/APJ system and cancer. *Clin Chim Acta, 457*, 112-116. doi:10.1016/j.cca.2016.04.001

Yang, Z., Hackshaw, A., Feng, Q., Fu, X., Zhang, Y., Mao, C., & Tang, J. (2017). Comparison of gefitinib, erlotinib and afatinib in non-small cell lung cancer: A meta-analysis. *Int J Cancer, 140*(12), 2805-2819. doi:10.1002/ijc.30691

Yao, Y., & Dai, W. (2014). Genomic Instability and Cancer. *J Carcinog Mutagen, 5*. doi:10.4172/2157-2518.1000165

Yavuz, B. R., Tsai, C. J., Nussinov, R., & Tuncbag, N. (2023). Pan-cancer clinical impact of latent drivers from double mutations. *Commun Biol, 6*(1), 202. doi:10.1038/s42003-023-04519-5

Zenonos, K., & Kyprianou, K. (2013). RAS signaling pathways, mutations and their role in colorectal cancer. *World J Gastrointest Oncol, 5*(5), 97-101. doi:10.4251/wjgo.v5.i5.97

Zhang, Y., Chen, F., & Creighton, C. J. (2023). Pan-cancer molecular subtypes of metastasis reveal distinct and evolving transcriptional programs. *Cell Rep Med, 4*(2), 100932. doi:10.1016/j.xcrm.2023.100932

Zhao, M., Sun, J., & Zhao, Z. (2013). TSGene: a web resource for tumor suppressor genes. *Nucleic Acids Res, 41*(Database issue), D970-976. doi:10.1093/nar/gks937

Zhou, Y., Xu, J., Hou, Y., Leverenz, J. B., Kallianpur, A., Mehra, R., . . . Cheng, F. (2021). Network medicine links SARS-CoV-2/COVID-19 infection to brain microvascular injury and neuroinflammation in dementia-like cognitive impairment. *bioRxiv*. doi:10.1101/2021.03.15.435423