LEVERAGING THE HUMAN KINOME FOR ANTICANCER AGENT CYTOTOXICITY POTENCY PREDICTION

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS OF
THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

MERİÇ KINALI

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE
OF MASTER OF SCIENCE
IN
BIOINFORMATICS

SEPTEMBER 2019

Approval of the thesis:

LEVERAGING THE HUMAN KINOME FOR PREDICTION OF ANTICANCER AGENT CYTOTOXICITY

Submitted by MERİÇ KINALI in partial fulfillment of the requirements for the degree of **Master of Science in the Health Informatics Program, Middle East Technical University** by,

Prof. Dr. Deniz Zeyrek Bozşahin
Dean, **Graduate School of Informatics, METU** _____

Assoc. Prof. Dr. Yeşim Aydın Son
Head of Department, **Health Informatics, METU** _____

Prof. Dr. Rengül Çetin Atalay
Supervisor, **Health Informatics, METU** _____

Asst. Prof. Dr. Aybar Can Acar
Co-Supervisor, **Health Informatics, METU** _____

Examining Committee Members:

Assoc. Prof. Dr. Yeşim Aydın Son
Health Informatics Dept., METU _____

Prof. Dr. Rengül Çetin Atalay
Health Informatics Dept., METU _____

Asst. Prof. Dr. Aybar Acar
Health Informatics Dept., METU _____

Assoc. Prof. Dr. Tunca Doğan
Health Informatics Dept., Hacettepe University _____

Assoc. Prof. Dr. Özlen Konu
Molecular Biology and Genetics Dept., Bilkent University _____

Date: **06.09.2019**

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : MERİÇ KINALI

Signature : _____

ABSTRACT

LEVERAGING THE HUMAN KINOME FOR ANTICANCER AGENT CYTOTOXICITY POTENCY PREDICTION

Kınalı, Meriç

MSc., Department of Bioinformatics

Supervisor: Prof. Dr. Rengül Çetin Atalay
Cosupervisor: Asst. Prof. Dr. Aybar Can Acar

September 2019, 68 pages

Cancer is the second deadly disease globally. Cell signaling cascades with altered protein kinase activities induce the majority of the hallmarks in cancer such as proliferation, angiogenesis, invasion, and metastasis. The major subtype of primary liver cancer, Hepatocellular carcinoma (HCC) has limited therapeutic options. In this study, we presented a regression model, which was applied initially on cytotoxic bioactivity data obtained from HCC cells treated with 120 kinase inhibitors called CanSyL dataset. The model then extended on publicly available datasets. The model uses human kinome tree topology-based classes of protein kinases. Small-molecule kinase inhibitors can act on other pathways by "off-target" or "pathway cross-talk" effects in addition to their previously reported targets. Our objective in this study was to predict these off-target effects as potential new targets by regularizing the regression space based on the kinome tree topology. Our regression model was tested on the CanSyL dataset by applying leave-one-out cross-validation and achieved promising predictions (median RMSE between 2.5-4 %) for the kinase inhibitor vulnerability matrix based on the regularization of the human kinome tree, with no bias in the estimates. Then we scaled up our approach to the public datasets (CCLE and GDSC). Some of the kinase inhibitors were identified as outliers based on their individual RMSE. They were significantly different from the kinase inhibitor groups that they belong to, according to the Mann-Whitney U test ($p<0.05$). This difference in specificity suggests that outlier inhibitors are more specific inhibitors while non-outlier inhibitors are mostly general multi-kinase inhibitors.

Keywords: Hepatocellular Carcinoma, Kinome Tree, Regression Model, Cancer Cell-lines, Kinase Inhibitors

ÖZ

İNSAN KİNOMUNUN HEDEF DIŞI İLAÇ ETKİLEŞİMİ TAHMİNİNDE KULLANILMASI

Kınalı, Meriç
Yüksek Lisans, Biyoenformatik Bölümü
Tez Yöneticisi: Prof. Dr. Rengül Çetin Atalay
Tez Eşyöneticisi: Dr. Öğr. Üyesi Aybar Can Acar

Eylül 2019, 68 sayfa

Kanser ölümcül hastalıklar arasında dünyada ikinci sırada yer almaktadır. Çesitli nedenlerle işlevleri değişen protein kinazlar görev aldıkları proliferasyon, anjiyojenez, invazyon, and metastaz gibi karsinogeneze temel oluşturan hücre aktivitelerini kontrol ederler. Karaciğer kanserinin en çok rastlanan türü olan Hepatosellüler karsinom (HSK) için tedavi olanakları sınırlıdır. Bu çalışmada, kinaz ailelerinin hassasiyetini belirlememizi sağlayan farklı HSK hücre hatlarından elde edilen sitotoksik bioaktivite değerlerini kullanarak bir regresyon modeli geliştirdik. Bu regresyon modelini kinazların filogenetik ağaç dallanma bilgisine göre düzenleyebilmek için insan kinom ağacındaki topolojisini kullandık. Kinaz inhibitörlerinin asıl hedefleri dışındaki yolakları da etkiledikleri bilinmektedir. Bu çalışmadaki amacımız regresyon modeli ile hesapladığımız hassasiyet değerlerini kullanarak kinaz inhibitörlerinin hedef-dışı etkilerini hesaplayabilmektir. Oluşturduğumuz regresyon modelinin performansını, KanSiL veri seti kullanarak test ettiğimizde, başarılı sonuçlar gözlemledik (medyan RMSE 2.5-4 %). Bir sonraki adımda, geliştirdiğimiz metodu diğer veri setlerine uyguladık (CCLE ve GDSC). Önerdiğimiz model ile yapılan analiz sonucunda kinaz inhibitörleri sahip oldukları RMSE değerine göre aykırı ve aykırı olmayanlar şeklinde gruplanmıştır. Son adım olarak aykırı olan kinaz inhibitör grubunun Mann-Whitney U test sonuçları, aykırı olmayan gruba göre anlamlı şekilde farklı bulunmuştur ($p<0.05$). Bu farklılık bize ayrıca bu metod ile seçici ve seçici olamayan kinaz inhibitörlerini de belirleyebileceğimizi göstermektedir.

Anahtar Sözcükler: Hepatosellüler Karsinom, Regresyon Modeli, Kanser Hücre Hattı, Kinaz İnhibitörleri, Kinom Ağacı

To My Family

ACKNOWLEDGEMENTS

Firstly, I would like to thank my advisors Prof. Dr. Rengül Çetin-Atalay and Dr. Aybar Can Acar for their support, assistance, and patience throughout my master's studies.

Besides my supervisors, I would like to thank the rest of my thesis committee: Dr. Yeşim Aydın Son, Dr. Özlen Konu and Dr. Tunca Doğan for their time and valuable suggestions for this thesis.

I would like to thank my loveliest colleagues Cansu Demirel, Cansu Dinçer, Elif Bozlak, Evrim Fer, Gökçe Sanger, Muazzez Çelebi Çınar and Ali Çınar for their endless support and valuable friendships. I would also like to thank my precious previous colleagues Melek Umay Tüz, Elif Tuğce Karoğlu, and Dilara Yılmaz. I would like to thank the Health Informatics department and especially department secretary Hakan Güler for his help and patience.

I would like to express my very great appreciation to Dr. Ayça Arslan Ergül for her enthusiastic encouragement for years. I would like to offer my special thanks to Dr. Augustin Luna for his valuable time and suggestions for this thesis.

Last but not least, I would like to express my deepest gratitude to my family Abdullah Kınalı, Sema Kınalı and my lovely cat Şirin for supporting me spiritually throughout writing this thesis and my life in general.

This thesis is supported by KanSiL project (2016K121540).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AGC	Containing protein kinases A, G, and C
aPK	Atypical Protein Kinase
CAMK	Calcium/Calmodulin-dependent Kinase
CCLE	Cancer Cell Line Encyclopedia
CK1	Casein Kinase I
CMGC	Containing Cyclin-dependent Kinase, MAPK, glycogen synthase kinase 3 and CDC2-like
EGFR	Epidermal Growth Factor Receptor
ePK	Eukaryotic Protein Kinase
FDA	U.S. Food and Drug Administration
FGFR	Fibroblast Growth Factor Receptor
GDSC	Genomics of Drug Sensitivity in Cancer
HCC	Hepatocellular Carcinoma
HGF	Hepatocyte Growth Factor
IGF	Insulin-like growth factor
JNK	Jun N-terminal Kinase
LIMK1	LIM Domain Kinase 1
LOOCV	Leave-one-out Cross-validation
MLCK	Myosin light-chain Kinase
PI3K	Phosphoinositide-3-kinase
PDGF	Platelet-derived Growth Factor
PIKK	Phosphatidylinositol 3-Kinase-Related Kinases
PLGF	Placental growth factor
ROCK1	Rho-associated protein kinase 1
STE	Homologues of yeast Sterile 7, sterile 11 and sterile 20
TK	Tyrosine Kinase
TKL	Tyrosine Kinase-like
VEGFR	Vascular Endothelial Growth Factor Receptor

CHAPTER 1

1.INTRODUCTION

Cancer is the world's second major cause of death, and the prevalence of cancer has risen significantly (Jemal et al., 2019). At the cellular level, cancer development is a multi-stage process that includes mutation and accelerating capacity for hallmarks such as proliferation, survival, invasion, and metastasis (Cooper & Hausman, 2007). According to the latest GLOBOCAN statistics liver cancer is the fourth leading cause of death (782 000 deaths, 8.2%) worldwide (Bray et al., 2018). The most prevalent and malignant type of primary liver cancer is Hepatocellular carcinoma (HCC) with limited therapeutic options (Balogh et al., 2016). Many studies have shown tumor heterogeneity of HCC which makes identification of the biomarkers or molecular subtypes challenging and critical for identification of new and effective therapeutics (Lu, Hsu, & Cheng, 2016). It has been proved that cell line models maintain the molecular background of cancer types in breast cancer and HCC derived cell-lines which will enable to identify potential biomarkers of different molecular subtypes (R. S. Finn et al., 2013).

Kinases catalyze the transfer of the phosphoryl groups of ATP onto the hydroxyl group of target proteins, a process known as phosphorylation, through their well-conserved eukaryotic protein kinase (ePK) catalytic domain (Cohen, 2001). Phosphorylation of the proteins modulates fundamental cellular activities such as proliferation, survival, apoptosis, metabolism, transcription, differentiation, and migration (Reimand, Wagih, & Bader, 2013). Since kinases are enzymes that responsible for the phosphorylation process, dysregulated protein kinases due to mutations, chromosomal rearrangements, gene amplification or epigenetic changes are associated with cancer and many other diseases. Therefore, they are essential clinical targets for cancer studies.

Cell signaling cascades with altered protein kinase activities induce the majority of the hallmarks of HCC such as proliferation, angiogenesis, invasion, and metastasis. Known altered pathways of HCC such as proliferation and

angiogenesis comprise growth factors and their receptors which are epidermal growth factor receptor (EGFR), platelet-derived growth factor (PDGF), insulin-like growth factor (IGF), hepatocyte growth factor (HGF), fibroblast growth factor receptor (FGFR), and vascular endothelial growth factor receptor (VEGFR) (Moeini, Cornellà, & Villanueva, 2012). Aberrant activation of these growth factors and their receptors also activates Ras, Raf, MAPK, ERK, PI3K/AKT, and mTOR which have been found as critical in the formation of hepatocellular carcinoma (Muntane, J. De la Rosa, Docobo, Garcia-Carbonero, & J. Padillo, 2013). Pathways related to differentiation are also deregulated in HCC, such as Wnt signaling in which protein kinases are key regulators in many steps (Verheyen & Gottardi, 2010). Therefore, human protein kinases are important targets and complete classification of them is essential for cancer studies. Human genome encodes 518 protein kinases and 20 lipid kinases which correspond approximately 2% of the genome have been grouped into groups, families, and subfamilies and called as human kinome based on their similarity in the amino acid sequence of the ePK catalytic domain (Manning, Whyte, Martinez, Hunter, & Sudarsanam, 2002). The ePK catalytic domain composes of N-terminal lobe (N-lobe) and C-terminal lobe (C-lobe) which are connected by cleft where ATP binds and most kinase inhibitors are designed to interrupt the cleft in between N-lobe and C-lobe (P. Wu, Nielsen, & Clausen, 2015). Human kinome is divided into eight main kinase groups based on their sequence similarity in their ePK domains as AGC (containing protein kinases A, G and C), CAMK (calcium/calmodulin-dependent kinase), CK1 (Casein Kinase 1), CMGC (containing cyclin-dependent kinase, MAPK, glycogen synthase kinase 3 and CDC2-like), STE (homologues of yeast Sterile 7, sterile 11 and sterile 20), TK (tyrosine kinase), TKL (tyrosine kinase-like) and atypical protein kinases (aPKs) which lack sequence similarity to ePK catalytic domain. This classification has promoted the study of the role of kinases in specific diseases in parallel with the development of kinase inhibitors. Several previous studies have indicated that defining off-target effects of small-molecule kinase inhibitors is critical for effective cancer treatments and rational drug development strategies. There is a need for the development of different kinome profiling tools in order to study the selectivity of kinase inhibitors toward their targets. After the human kinome tree was originally classified by Manning et al., the resulting phylogenetic tree was used to develop several kinome tree viewer tools, disease associated kinase resources, or quantitative selectivity methods. To date, the kinome tree topology was used to develop many kinome tree viewers such as KinMap (Eid, Turk, Volkamer, Rippmann, & Fulle, 2017), Coral (Metz et al., 2018), TREEspot (Davis et al., 2011), and the NCGC Kinome Viewer (tripod.nih.gov). The kinome tree topology also was used to develop kinase and disease-associated resources such as KIDFamMap (Chiu et al., 2013) and KinMutBase (Ortutay, Väliaho, Stenberg, & Vihinen, 2005).

Cancer-derived cell lines are essential and widely used models in cancer biology studies. In order to test the efficacy of therapeutic agents, cancer cell lines have been used by several studies. Although the clinical relevance of cell lines has been questioned, these models provide the development of anticancer inhibitors such as bortezomib for the treatment of multiple myeloma (Gillet, Varma, & Gottesman, 2013) as well as other lead molecules. Large-scale pharmacogenomic studies provide a collection of human cancer-derived cell lines tested for different inhibitors. The Genomics of Drug Sensitivity in Cancer (GDSC) [CITE: PMID 23180760] study has provided drug response data of 138 anti-cancer therapeutics across 700 human cancer-derived cell lines, from 29 tissue types (Garnett et al., 2012) and the Cancer Cell Line Encyclopedia (CCLE) [CITE: PMID 22460905] has provided a collection of drug response data of 24 anti-cancer therapeutics across 500 cell lines (Barretina et al., 2012). We have used publicly available GDSC and CCLE datasets additional to our in house CanSyL data to compare and examine bioactivity results of the available HCC cell lines treated with kinase inhibitors comprehensively.

In this study, we built a regression model to predict the efficacy of novel kinase inhibitors using kinase tree topology through the human kinome tree. It is well known that in addition to their targeted signaling pathway, small-molecule kinase inhibitors can affect other pathways by "off-target" or "pathway cross-talk" effects. Our objective in this study was to predict these off-target effects of kinase inhibitors by regularizing the regression space based on the kinome tree.

The specific aim and concepts of this thesis were described with respect to the latest studies about underlying biology and the hallmarks of hepatocellular carcinoma and the roles of kinases on them in Chapter 2. The off-target effect concept for the small molecule kinase inhibitors and previously proposed quantitative selectivity measurement methods were described in detail. Chapter 3 presents the overall methodology and the regression model and its individual steps such as matrix multiplication, cross-validation, and cosine similarity. The application of the regression model on CanSyL, GDSC and CCL data and their results were presented in Chapter 4. Finally, our model and its eventual use for off-target effect predictions to be exploited in drug repositioning or drug repurposing were discussed in Chapter 5.

CHAPTER 2

2.LITERATURE REVIEW

2.1. Molecular Cellular Biology of Primary Liver Cancer

2.1.1. Impact of Protein Kinases on the Hallmarks of Cancer

HCC is a complex disease which often develops over a background of chronic liver disease or cirrhosis. Chronic cellular injury due to viral hepatitis B or C infections, genotoxic or metabolic stress, induces carcinogenic events. Studies have shown that aberrant cell signaling pathways caused during these chronic events are the main reasons for this carcinogenesis (Whittaker, Marais, & Zhu, 2010). Therefore, the identification of the key players of the signaling pathways involved in hallmarks of cancer is critical for the developing novel molecular targeted therapies for HCC. Therefore, the importance of protein kinases which are considered as molecular switches is discussed for their involvement in proliferation and neovascularization which are the most significant hallmarks of hepatocellular cancer.

Aberrations in protein kinase signaling pathways drive multiple hallmarks of cancer including survival, motility, proliferation, metabolism, angiogenesis, genomic instability, and evading immunity (Gross, Rahal, Stransky, Lengauer, & Hoeflich, 2015). Kinase pathways regulate survival by controlling apoptosis and necroptosis regulators or changing their expression. Receptor tyrosine kinases with integrins regulate cytoskeletal dynamics and, hence motility by increasing the activation of Focal adhesion kinase (FAK), ROCK1 (Rho-associated protein kinase 1), MLCK (Myosin light-chain kinase), PAK1, LIMK1 (LIM domain kinase 1). Also, MAPK pathway components affect cell cycle progression and proliferation. Moreover, cancer cells secrete angiogenic ligands for VEGFR, FGFR, and TIE2 to increase vascularization and angiogenesis. Alterations the tumor suppressor genes PTEN and INPP4B lead to the lack of their inhibitory action on Phosphoinositide-3-kinases (PI3Ks), and in turn, result in increased glucose uptake and metabolism in primary liver cancer cells. Hence PI3K/Akt the cell survival pathway is activated (Pavlova & Thompson, 2016). More importantly, loss of heterozygosity mutations in tumor suppressor TP53,

ATM kinase, and CDKN2A result in oncogene-induced DNA replication stress which causes genomic instability (Negrini, Gorgoulis, & Halazonetis, 2010).

Cell signaling cascades with altered protein kinase signaling cause the majority of the hallmarks of HCC such as proliferation, angiogenesis, invasion, and metastasis as shown in table 2.1 (Moeini et al., 2012; Niwa et al., 2005). Known dysregulated pathways of HCC such as proliferation and angiogenesis comprise growth factors and their receptors which are EGFR, PDGF, IGF, HGF, FGFR, and VEGFR (Moeini et al., 2012). Pathways related to differentiation are also deregulated, such as Wnt signaling in which protein kinases are key regulators in many steps (Verheyen & Gottardi, 2010). Taking into account all of these, the information of altered kinase signaling pathways of HCC, kinase inhibitors, and kinome classification offers an important opportunity for drug discovery.

Table 2.1: Altered Molecular Pathways in HCC pathogenesis.

Altered Pathway	Gene/target	Alteration	Molecular Therapies
Differentiation & Development			
Wnt/beta-catenin	CTNNB1	Activating mutation/ overexpression	-
	AXIN1	Inactivating mutation/LOH	
	APC	Inactivating mutation	
Notch	NOTCH1	Overexpression	-
	NOTCH3	Overexpression	
Hedgehog	SHH	Activating overexpression	
	SMO	Activating overexpression	
	HHIP	Downregulated by LOH	
Hippo	MST1/2	Down-regulated	-
	pYAP	Down-regulated	
Cell Cycle Regulation			
p53/cell cycle	TP53	Inactivating mutation/LOH	Gene-therapy/ Ad5CMV-p53 gene
	CDKN2A (p16)	Inactivating mutation /Hypermethylation	Flavopiridol
	IRF2	Inactivating mutation	-
Proliferation			
EGF	EGF/EGFR	Upregulated	Erlotinib, Gefinib, Cetuximap, Lapatinib
HGF/MET	HGF	Upregulated	SU5416/11274
	HGFR (MET)	Upregulated	Cabozantinib (XL184), Foretinib
IGF	IGF2	Overexpression	IMC-A12/Cixutumumab
	IGF-2R	Down-regulating mutation/LOH	
PI3K/AKT/mTOR	PIK3CA	Activating mutation	BKM120
	PTEN	Down-regulating mutation/LOH	-
	MTORC1	Upregulated	Everolimus, Rapamycin
RAS/MAPK	KRAS	Activating mutation	Sorafenib
	RPS6K3	Inactivating mutation	-
Angiogenesis			
FGF	FGF19	Upregulated	-
	FGFR1/2	Upregulated/ activating mutations	Brivanib
PDGF	PDGFRA	Upregulated	Sorafenib, Sunitinib, Imatinib
VEGF	VEGF	Upregulated	
	VEGFR2	Upregulating amplifications	Sorafenib, Brivanib, Suninib
Cell Growth &Migration			
JAK/STAT	SOCS3	Methylation associated silencing	AG490

HCC oncogenesis involves various dysregulated signaling pathways. Among these, proliferation and survival pathways provide many opportunities for the development of molecularly targeted therapies (R. Finn, 2013).

2.1.2. Role of Protein Kinases in Hepatocellular Carcinoma Vascularization

As one of the hallmarks of cancer, vascularization is regulated by complex interactions between growth factors, endothelial cells and secreted ligands (Hanahan & Weinberg, 2011). VEGF, FGF, and Angiopoietins (Ang family of proteins) are major regulators of neovascularization (Chao et al., 2003). VEGF's angiogenic activity is tightly regulated by gene dosage (Carmeliet & Jain, 2000). Previous studies have reported the formation and growth of HCC depends heavily on the formation of new blood vessels in which VEGF members and their receptors are critical (Figure 2.1) (Zhu, Duda, Sahani, & Jain, 2011). Moon et al. (2003) have shown that the overexpression of the VEGF increases vascularity and tumor growth in samples obtained from 49 patients with HCC. VEGF members stimulate cellular responses by binding receptor tyrosine kinases (VEGFRs). Additionally, VEGF expression can be induced independently by hypoxia apart from oncogenic mutations, hormones, cytokines and signaling molecules such as nitric oxide and MAP kinases.

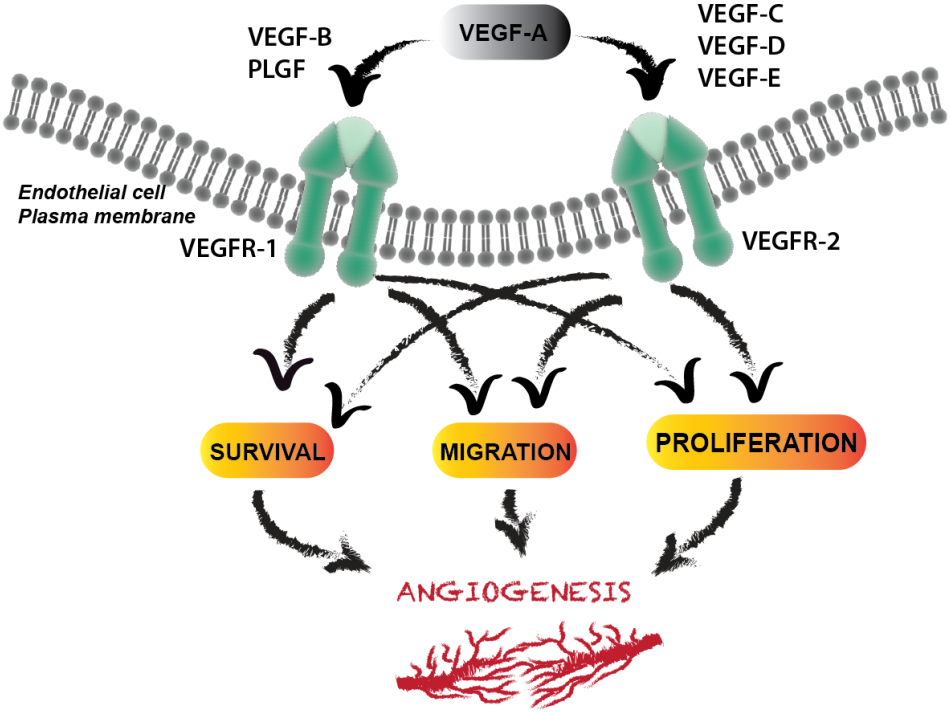


Figure 2.1: VEGF pathway representing the effect of tumor-secreted VEGF on endothelial cells. VEGF ligands (VEGF-A, VEGF-B, VEGF-C, VEGF-D,

VEGF-E, and PLGF) bind to their receptor tyrosine kinases specifically which activates signal-transduction pathways leading the survival, proliferation, and migration of endothelial cells to form new vessels.

Taken together, targeting angiogenesis is an effective therapy for HCC treatment such as approved anti-VEGF therapy with multi-kinase inhibitor sorafenib (Zhu et al., 2011). The biological impacts of anti-angiogenic agents on HCC treatment have been tested in many other studies as well. These studies mostly investigated multi-targeted small-molecule tyrosine kinase inhibitors, including Sunitinib (Zhu et al., 2009), Brivanib (Huynh et al., 2008), Linifanib (Cainap et al., 2015), GW786034 (Yau et al., 2011), PTK787 (Wood et al., 2000), and AZD2171 (Alberts et al., 2012).

2.1.3. Role of Protein Kinases in Hepatocellular Carcinoma Proliferation

PI3K/Akt/mTOR pathway is one of the important pathways have been shown to be involved in HCC due to PTEN tumor suppressor deletion (Buontempo et al., 2011). This major intracellular pathway regulates cell growth, proliferation, and survival. The Ras/MAPK pathway is another significant pathway in the formation of HCC. MAPKs integrates extracellular stimulations such as growth factors, cytokines, and extracellular stress signals into intracellular responses. Many studies have shown that MAPK pathways are deregulated in tumors, including HCC. MAPK family belongs to CMGC kinase group and this family contains extracellular signal-regulated kinases (ERKs), Jun N-terminal kinases (JNKs), and p38 MAPKs (Min, He, & Hui, 2011).

2.1.4. Importance of Tyrosine Kinase Inhibitors in Hepatocellular Carcinoma Therapy

As described above, critical signaling pathways in HCC formation, mostly involve tyrosine kinases. Up to now, 38 kinase inhibitors have been approved for cancer treatment in general by the FDA. Sorafenib and Regorafenib are the only FDA approved small molecule for HCC and they are multi-kinase inhibitors acting on receptor tyrosine kinases (Ferguson & Gray, 2018). There are two classes of tyrosine kinases as receptor tyrosine kinases (RTK) which are both cell surface receptors, and kinase enzymes and non-receptor tyrosine kinase (NRTK) (M. K. Paul & Mukhopadhyay, 2004). Tyrosine kinases are essential enzymes and mediators for the cell signaling process because the majority of them are cell surface receptors and they are responsible for the initiation of cell signaling cascades leading to carcinogenesis.

2.2. Small molecule Kinase Inhibitors' Selectivity Toward Their Targets

The catalytic activity of the protein kinases is dependent on the ATP molecule, which binds the ATP-binding pocket located in between a β sheet containing N-terminal lobe (N-lobe), an α helix dominated C-terminal lobe (C-lobe), and a connecting hinge region. Although variety in their main amino acid sequence, the human kinases share a high degree of similarity in their 3D structures, particularly in their catalytically active kinase domain where the ATP-binding pocket is positioned. The ePK catalytic domain, consisting N-lobe and C-lobe, is linked through a cleft where ATP binds and most kinase inhibitors are intended to interrupt the cleft between N-lobe and C-lobe (P. Wu, Nielsen, & Clausen, 2016). The access to the active site where the ATP-binding pocket is located is controlled by a flexible (conformationally mobile) activation loop beginning with a conserved sequence of three amino acids, aspartate (D), phenylalanine (F) and glycine (G) which is known as DFG motif (Knighton et al., 1991). On activation, kinase structures undergo conformational changes in the active (DFG-in) and inactive (DFG-out) form (Kooistra & Volkamer, 2017). Understanding these conformational changes is a key concept to reveal disease-causing roles of kinases (Möbitz, 2015). Kinase inhibitors are classified according to their binding mode as "irreversible" or "reversible". Based on their binding site, reversible small-molecule kinase inhibitors are categorized into four main groups: *Type I*, *Type II*, *Type III* (Allosteric Inhibitors) and *Type IV* (Substrate Directed Inhibitors) as listed in Table 2.1 (Bhullar et al., 2018). Since *Type I* inhibitors occupy ATP-binding pocket of the kinase in its active DFG-in state, they show a low kinase selectivity as the targeted ATP pocket is conserved through the human kinome (Bhullar et al., 2018; Zhang, Yang, & Gray, 2009). Despite their large-scale clinical achievement, type I kinase inhibitors (e.g., erlotinib, gefitinib) come with potential off-target side effects due to their low selectivity (Bhullar et al., 2018). In contrast to the majority of inhibitors, *type II* inhibitors (e.g., imatinib and sorafenib) recognize the inactive DFG-out conformation of the kinases and occupy an additional hydrophobic binding site adjacent to the ATP binding site (Kufareva & Abagyan, 2008; Zhang et al., 2009). *Allosteric (type III) inhibitors* (e.g., CI-1040, GNF-2) bind in an allosteric pocket other than an active site without interacting with the ATP binding pocket and allosterically modulates kinase activity (P. Wu et al., 2015; Zhang et al., 2009). These inhibitors display the highest degree of target kinase selectivity since they exploit binding sites and physiological mechanisms unique to a specific kinase (Zhang et al., 2009). *Substrate-directed or type IV inhibitors* (e.g., ON012380*)* bind to an allosteric site reversibly (Blanc, Geney, & Menet, 2013; Cox, Shomin, & Ghosh, 2011; P. Wu et al., 2015). These kinase inhibitors do not compete with ATP and therefore they tend to show a higher degree of selectivity against kinases (Blanc et al., 2013). Irreversible, *Covalent or type V* (e.g., afatinib), kinase inhibitors bind covalently to the ATP-binding site with a

reactive nucleophilic cysteine residue, leading in ATP site blockage and irreversible inhibition (P. Wu et al., 2016).

Table 2.2: Types of small molecule kinase inhibitors.

Class of Inhibitor	**Binding Site**	**Selectivity**	**Binding Type**
Type I	ATP site	Low	Reversible
Type II	ATP site and DFG pocket	High	Reversible
Type III (Allosteric)	Allosteric (ATP pocket vicinity)	High	Reversible
Type IV (Substrate-directed)	Allosteric (substrate-binding domain)	High	Reversible
Type V (Covalent)	ATP site	Low	Irreversible

The ATP-binding site is evolutionarily conserved among kinases (Manning et al., 2002). Hence, most kinase inhibitors targeting ATP-binding site promiscuously inhibit multiple kinases (Anastassiadis, Deacon, Devarajan, Ma, & Peterson, 2011). Previous studies have demonstrated that defining off-target effects of small-molecule kinase inhibitors is critical for effective cancer treatments and rational drug development strategies. The need to study the selectivity of kinase inhibitors has led to the development of novel quantitative methods to measure the selectivity of the kinase inhibitors. Karaman et al., and his colleagues introduced "Selectivity Score" (S) which is calculated for each compound by dividing the number of kinases bound by an inhibitor with a specific affinity score to the total number of kinases experimented (Karaman et al., 2008). Graczyk et al. proposed another selectivity measurement metric for kinase selectivity which is a novel application of the Gini coefficient (G) (Graczyk, 2007). In this compound concentration-dependent selectivity evaluation method, the magnitude of inhibition is measured for each kinase at a single point for a specific ATP concentration. In this method first, the sum of magnitudes of inhibition for all kinases is calculated to find total inhibition. Second, kinases are sorted in increasing order. After a cumulative fraction of total inhibition is plotted against the cumulative fraction of kinases, the Gini coefficient is calculated through the Lorenz Curve. As a result of this metric, nonselective inhibitors have scores close to zero whereas selective compounds have scores close to one. Consequently, selectivity patterns of the kinase inhibitors give useful insight into potential interactions in which an inhibitor may be involved. However, the significance of these selectivity patterns must be understood in the context of cell and tissue biology (Smyth & Collins, 2009).

CHAPTER 3

3.DATASETS AND METHODS

In this chapter, the methodology of this study is explained then the concepts integrated into our model such as matrix multiplication, cross-validation, and cosine similarity are presented with respect to the specific aims of this thesis.

3.1. Overview of the Methodology

In this study, a regression model was built through matrix multiplication of selectivity and bioactivity matrices based on cosine similarity using the kinome tree feature space.

First, to determine compatible *selectivity score vector (**s**)* and *vulnerability score vector (**v**)*, for each cell line we calculated the similarity between cell lines' responses to kinase inhibitors through cosine similarity. Second, we computed the similarity between the kinome tree regularized *selectivity score vector (**s**)* and the *family vulnerability score vector (**fv**)* for each cell line also through cosine similarity. Then we calculated the error percentage between these cosine similarities using root mean square error (RMSE) in the Leave-one-out cross-validation (LOOCV) process. As a result of this method, we aimed to predict a *vulnerability score (**v**)* for each kinase family, its kinome tree regularized version *(**fv**)* and the error rate between them to be able to make off-target predictions. We applied this procedure to all three datasets (CanSyL, CCLE, and GDSC) exploited in this thesis. Then we used the regularized and the non-regularized vulnerability scores for the kinase genes belonging to their families in the enrichment analysis to see whether kinome-tree based regularization reveals better prediction for off-target effects of kinase inhibitors.

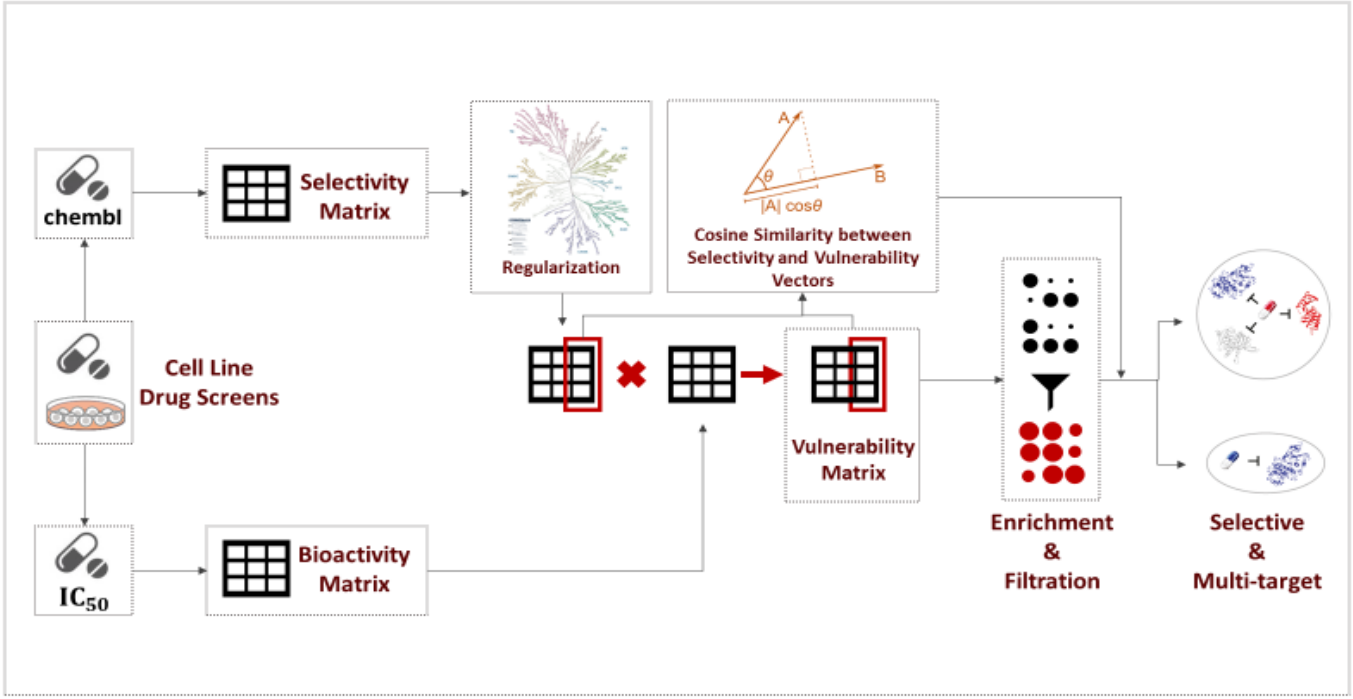


Figure 3.1: The workflow of the thesis work. Drug screen data were retrieved from databases to build "selectivity" and "bioactivity" matrices. Vulnerability matrix was obtained by multiplying these two matrices. Selectivity matrix was regularized based on the kinome tree topology. Regularized and non-regularized vectors were compared based on cosine similarity. Targetable pathways have found through enrichment analysis using vulnerability scores and inhibitor enrichment analysis was performed using these pathways and grouped inhibitors.

3.2. Datasets

3.2.1. Data from CanSyL

Hepatocellular carcinoma cancer cell lines (Huh7, HepG2, Mahlavu, FOCUS) were grown in Dulbecco's Modified Eagles Medium (DMEM) supplemented with 10% fetal bovine serum (Gibco, Invitrogen), 1% non-essential amino acids (Gibco, Invitrogen) and 100 units/ml penicillin/streptomycin (Gibco, Invitrogen). Cells were maintained at 37 °C in a humidified incubator under 5% CO_2. Kinase inhibitors were purchased from Calbiochem/MERCK. NCI-SRB assay was initially performed to define the inhibitory concentrations (IC50 values) of compounds in HCC cell-lines Huh7, HepG2, FOCUS, and MAHLAVU. From this experiment, we obtained HCC cell-line data which contains 4 cancer cell lines and 120 inhibitors.

Table 3.1: Small molecule kinase inhibitors and their IC_{50} values four HCC cell lines used in thesis from Calbiochem/MERCK (*NI: No inhibition)

INHIBITORS	FOCUS	Mahlavu	Huh7	HepG2
MeSAdo	NI*	NI	NI	59.04
AG1296	NI	NI	9.54	26.76
AG1478	12.23	10.92	15.22	7.60
AG82	NI	NI	NI	NI
Akti-1/2	5.21	5.58	8.82	1.88
Aloisine-A	9.75	7.63	9.45	2.59
Alsterpaullone	<0.01	<0.01	<0.01	<0.01
Aminopurvalanol-A	2.24	0.39	2.20	0.32
ATM-Kinase Inhibitor	14.17	17.10	8.21	9.57
Aurora Kinase Cdk Inhibitor	0.10	0.09	0.47	<0.01
Aurora Kinase Inhibitor II	8.39	14.85	6.33	15.37
Bcr-Abl Inhibitor	10.02	8.18	8.04	8.98
Ro-31-8220	<0.01	<0.01	<0.01	<0.01
Ro-31-8425	0.69	4.31	0.02	1.78
Bohemine	35.02	26.89	24.99	24.53
BPDQ	10.29	11.75	11.24	5.84
BPIQ-I	27.67	36.39	15.92	18.41
Casein Kinase II Inhibitor I	33.64	18.90	55.16	29.99
CK2 Inhibitor DMAT	15.16	23.62	4.06	9.22
Cdk Inhibitor p35	6.59	7.63	7.94	4.45

Cdk1 Inhibitor CGP74514A	<0.01	<0.01	<0.01	<0.01
CDK 1/2 Inhibitor II NU6102	10.82	9.62	10.66	4.62
Cdk2 Inhibitor II Compound3	NI	NI	NI	NI
Cdk2 Inhibitor III	5.72	4.10	2.03	<0.01
Cdk2/5 Inhibitor	NI	NI	NI	NI
Cdk2/9 Inhibitor	<0.01	<0.01	<0.01	<0.01
Chk2 Inhibitor II	12.03	12.66	9.34	10.82
CLK Inhibitor TG003	NI	NI	42.62	NI
Curcumin, Curcuma Ionga L.	13.36	19.54	9.71	16.08
Daphnetin	NI	NI	NI	NI
DNA-PK Inhibitor	NI	NI	NI	NI
EGFR Inhibitor	<0.01	<0.01	<0.01	<0.01
EGFR Inhibitor II, BIBX1382	6.69	10.67	8.50	7.75
EGFR / ErbB-2 Inhibitor	14.42	20.03	14.72	15.87
EGF / ErbB2 / ErbB-4 Inhibitor	24.63	49.08	22.01	27.18
Epigallocatechin Gallate	NI	NI	NI	NI
Erk Inhibitor	32.85	28.55	20.74	31.43
Erk Inhibitor II FR 180204	NI	NI	82.10	42.62
Fascaplysin, Synthetic	<0.01	<0.01	<0.01	<0.01
FGF / VEGF RTK Inhibitor	1.85	8.90	0.01	2.85
Genistein	30.28	40.02	74.70	18.88
Gö 7874, Hydrochloride	3.22	8.53	0.67	0.73
H89, Dihydrochloride	4.63	4.45	2.94	2.16
HA1100, Hydroxyfasudil	NI	NI	NI	18.83
Hypericin	14.65	15.20	19.34	23.66
IC261	<0.01	<0.01	<0.01	<0.01
Indirubin Derivative E804	<0.01	1.50	3.92	0.00
IRAK-1/4 Inhibitor	NI	NI	NI	24.85
Isogranulatimide	14.82	37.89	40.43	6.63
JAK Inhibitor I	NI	NI	NI	14.86
JNK Inhibitor V	0.59	5.40	2.05	1.84
Kenpaullone	19.37	26.65	16.07	8.67
LY 294002	10.56	14.15	4.00	6.48
MEK Inhibitor II	33.37	35.32	36.73	28.95
MEK1/2 Inhibitor	15.43	15.86	16.54	2.19

ML-7, Hydrochloride	9.20	17.78	12.71	9.19
MNK1 Inhibitor	NI	NI	43.61	3.56
Olomoucine	NI	NI	NI	NI
PD 98059	NI	NI	68.20	23.10
PKCßII / EGFR Inhibitor	9.54	9.96	7.40	4.69
PP1 Analog	11.75	52.71	10.77	9.46
PP1 Analog II, 1NM-PP1	4.51	4.58	4.75	16.82
H-7, Dihydrochloride	NI	52.25	10.45	16.92
H-8-Dihydrochloride	NI	NI	38.08	14.05
KN-93	0.69	0.15	0.45	0.13
Purvalanol-A	2.11	6.91	6.03	4.70
Quercetin dihydrate	67.78	NI	26.33	53.17
ROCK Inhibitor	74.29	NI	NI	12.17
R-Roscovitine	7.01	10.99	10.38	9.51
SB-203580, Iodo-	9.64	17.29	12.84	7.38
SB-202190	27.12	28.90	11.94	6.71
SB-218078	34.82	34.82	4.62	0.00
SCY	28.45	66.75	38.81	24.68
SKF-86002	15.94	14.30	13.19	21.07
ST-638	NI	NI	NI	NI
Staurosporine	<0.01	<0.01	<0.01	<0.01
STO-609	NI	NI	NI	22.03
SU-5402	NI	NI	8.96	28.03
SU-9516	6.78	7.68	5.63	4.77
TGFß-R1-Inhibitor	NI	73.94	NI	NI
TGFß-R1-Inhibitor II	NI	NI	NI	NI
TX-1123	6.62	2.74	3.15	1.77
TX-1918	9.74	6.09	10.75	6.04
Tyrene CR4	11.31	12.45	3.05	7.15
Tyrophostin AGL-2043	44.72	126.96	11.24	16.22
W-5, Hydrochloride	38.80	57.37	NI	NI
W-7, Hydrochloride	31.91	13.95	34.26	21.81
Wee1/Chk1 Inhibitor	0.64	1.35	0.00	0.03
Wortmannin	32.26	NI	30.49	20.62
ZM-336372	NI	NI	NI	24.89

2-thioadenosine	NI	NI	NI	NI
A3-hydrochloride	19.82	19.76	20.81	7.15
AG-17	<0.01	<0.01	<0.01	<0.01
AG-18	NI	NI	NI	NI
AG-30	NI	NI	NI	NI
AG-99	NI	NI	NI	NI
AG-112	NI	NI	NI	NI
AG-126	95.69	8.21	38.24	50.76
AG-213	NI	NI	38.73	28.07
AG-490	63.78	49.35	23.13	13.90
AG-527	NI	46.62	54.97	15.36
AG-825	NI	NI	NI	0.05
AG-879	8.51	4.82	9.78	4.20
AG-1295	NI	NI	NI	0.00
Butein	15.07	18.36	12.45	11.60
Emodin	11.76	6.30	12.16	12.71
Piceatannol	23.09	NI	22.67	23.02
Lavendustin C	NI	NI	NI	NI
Tamoxifen, 4-Hydroxy-(Z)-	12.76	19.96	13.18	11.15
Et-18-OCH3	0.99	7.18	16.68	7.73
Tamoxifen Citrate	9.14	7.68	6.61	5.15
HA 1077, dihydrochloride	NI	NI	NI	NI
Geldanamycin, S	<0.01	0.24	<0.01	<0.01
Herbimycin A	4.07	3.36	1.55	0.16
K252-a	0.63	0.43	0.00	0.00
Met Kinase Inhibitor	1.68	4.90	1.55	1.34
Calphostin C	0.25	0.22	0.18	0.04
JNK Inhibitor II	17.34	NI	14.04	25.66
K-252b, Nocardiopsis sp.	11.13	NI	2.34	5.38
SB 239063	NI	NI	NI	NI

3.2.2. Data from CCLE

Bioactivity file (CCLE_NP24.2009_Drug_data_2015.02.24.csv) was downloaded from the CCLE website (https://portals.broadinstitute.org/ccle) and the bioactivity data were retrieved from the file for only the liver cancer cell lines (Hep3B2.1-7, Hep-G2, HLE, HLF, Huh-1, JHH-2, JHH-4, JHH-6, PLC/PRF/5-Alexander cells, SK-HEP-1, SNU-182, SNU-423, SNU-449).

3.2.3. Data from GDSC

Bioactivity file (v17.3_fitted_dose_response.xlsx) was downloaded from the GDSC website (https://www.cancerrxgene.org/) and the bioactivity data were retrieved from the file for only the liver cancer cell lines (C3A, Hep3B2-1-7, Huh-1, Huh7, SK-HEP-1, SNU-449). IC_{50} values were filtered based on the area under the curve (AUC) values greater than 0.80.

3.2.4. Target Set

ChEMBL is an open large-scale database (https://www.ebi.ac.uk/chembl/) which was mostly manually curated from medicinal chemistry literature (Gaulton et al., 2012). CheMBL database provides a large variety of resources for the drug discovery studies and contains 1.8 million compounds, 11000 drugs, and 12000 targets in total (Mendez et al., 2019). Kinase Targets of the purchased small molecule kinase (Table 3.1) inhibitors in CanSyL, CCLE and GDSC datasets were retrieved from ChEMBL database (version 23) through chembl_webresource_client package separately (https://github.com/chembl/chembl_webresource_client). ChEMBL web services offer 20 distinct types of resources and they can be listed by invoking new_client through chembl_webresource_client package. From 20 different resources, we have used “activity” and “target” to retrieve target information (ChEMBL_ID and target name) of the kinase inhibitors in all three datasets using ChEMBL_IDs of the inhibitors (Davies et al., 2015). Targets of the specific inhibitors were filtered with respect to taxonomy (i.e. human) and pChEMBL (4.4) value. Activity point with pChEMBL value indicates that the corresponding record has been curated. Family, group and subgroup information of the targets were retrieved from the Human Kinome Tree (http://kinase.com/web/current/) based on Manning et al.

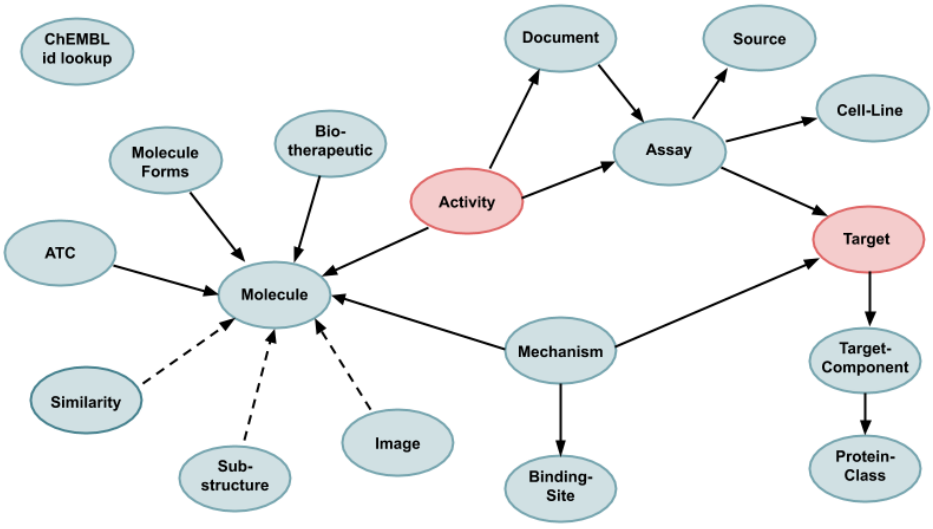


Figure 3.2: The diagram of the ChEMBL web service resources and resources used in this study. Oval shapes in red indicate resources used to retrieve the targets in this study and line between them indicate that they share a common attribute.

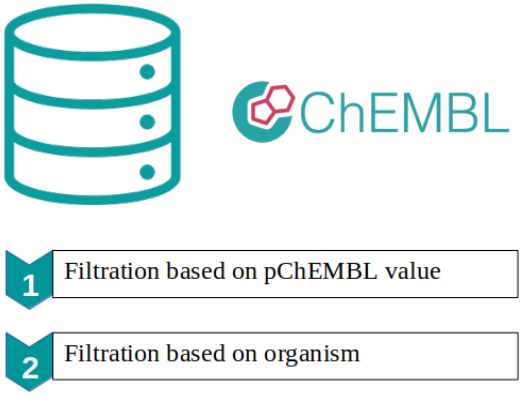


Figure 3.3: Filtration parameters for the chembl_websource_client package. Targets of the kinase inhibitors were filtered with respect to taxonomy (human) and pChEMBL (4.4) value.

Table 3.2: The information of the dataset used in the thesis work.

Data	**CanSyL**	**CCLE**	**GDSC**	**Retrieved From**
Inhibitor	81	19	61	CanSyL, CCLE, GDSC
Cell-line	4	13	6	CanSyL, CCLE, GDSC
Kinase Family	92	89	94	Manning et al., 2002
Kinase Group	9	9	9	Manning et al., 2002
Target	359	334	349	ChEMBL Database

3.3. Kinome Tree-based IC_{50} Regression Model

3.3.1. Selectivity and Bioactivity Matrices

In order to calculate the selectivity score, the number of targeted kinases in a specific family was divided by the number of total targets for each inhibitor (Equation 2.1). Family selectivity score calculation modified from Kahraman et. al. selectivity score formula. Then, bioactivity matrix was obtained by scaling IC_{50} value between 0 and 1 where a higher value indicates a stronger effect (Equation 2.2).

$$Selectivity = \frac{\#\ targeted\ kinases\ in\ a\ family}{\#all\ targets\ of\ inhibitor} \quad (2.1)$$

$$Bioactivity\ Score = \frac{Max_{IC_{50}} - IC_{50}}{Max_{IC_{50}} - Min_{IC_{50}}} \quad (2.2)$$

3.3.2. Implementation of the Matrix Multiplication

Matrix multiplication is a basic tool in linear algebra which has various applications in bioinformatics. Matrix product contains a record of all information of the two vectors which corresponds to the composition of each term in the two matrices. Namely, if A is an *r x s* matrix and B is an *s x t* matrix, product matrix *r x t* equal to multiplication of each term in a row of A and each term in a column of B as illustrated in Figure 2.2. (Klein 2013). In our case, we obtained the product matrix, called a vulnerability matrix, by multiplying each term in a row of selectivity matrix which contains selectivity scores of all inhibitors for a kinase family and column of bioactivity matrix which contains IC50 values of each inhibitor for a specific cell line. As a result, we obtained a combination of each row of selectivity and column of bioactivity results which we called the vulnerability score of each kinase family in a specific cell line (Figure 2.3). Matrix multiplication steps in the methodology were performed using the matrix multiplication function in R (>A %*% B #A and B are two matrices).

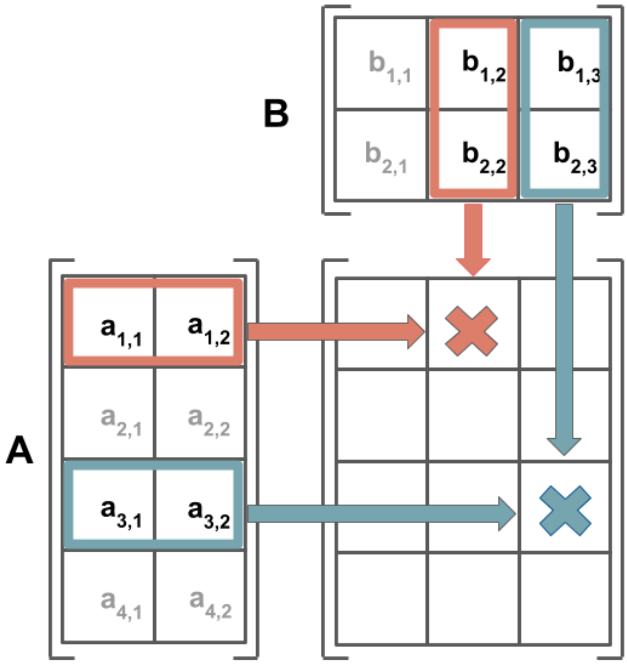


Figure 3.4: Demonstration of the matrix multiplication. Each intersection in the product of two matrices A and B equal to the multiplication of each term in a row of A and each term in a column of B. (Notation: Capital letters represent matrices and lowercase letters represent vectors.

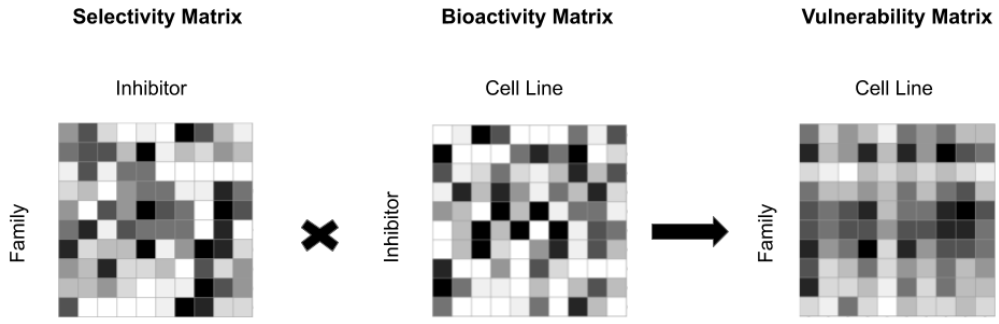


Figure 3.5: The first step of the methodology. The first step is the kinase family vulnerability score calculation based on the selectivity matrix and bioactivity matrix multiplication.

In the first step of our methodology, we obtained the product matrix which is vulnerability matrix by multiplying each term in a row of selectivity matrix containing selectivity scores of all inhibitors for a kinase family and column of bioactivity matrix containing normalized IC_{50} values of each inhibitor for a specific cell line (Figure 3.5). As a result, we obtained a combination of each row of selectivity and column of bioactivity results which we called the vulnerability score of each kinase family in a specific cell line

3.3.3. Regularization

The feature space was regularized on the human kinome by grouping kinases based on their families and groups based on the assumption that off-target interactions are more likely to occur for the closely related kinases.

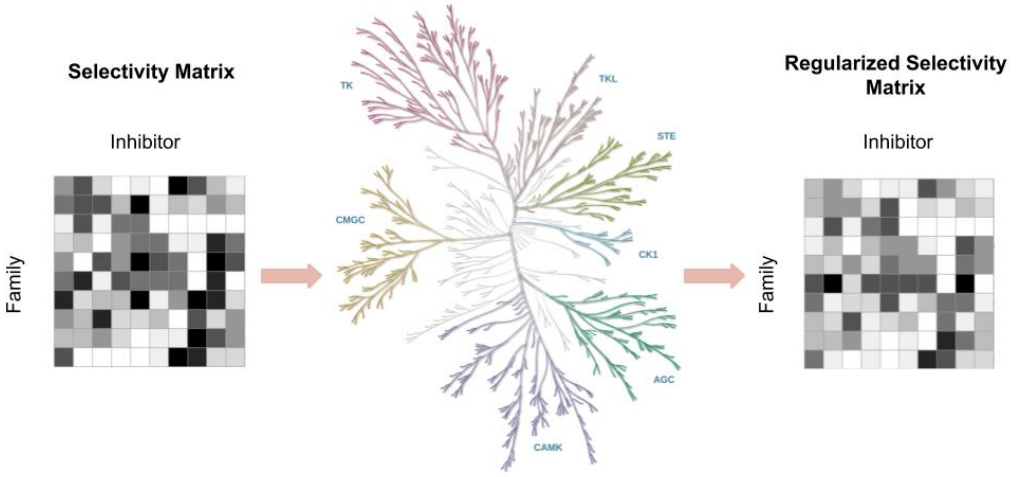


Figure 3.6: The second step of the methodology**.** The second step of the methodology is regularization of the selectivity matrix by distributing the selectivity score of the inhibitors for the kinase families between the kinase groups based on human kinome tree.

In the regularization step, the selectivity score of the novel inhibitors for the kinase families is distributed between the kinase groups based on human kinome tree. To do so, half of the score assigned to its real family. The remaining half of the score is equally shared between the other families in that kinase group as illustrated in Figure 3. We decided the value of regularization parameter empirically as 0.5 based on the minimum RMSE result. After regularization of the selectivity scores based on the human kinome tree, we obtained a vulnerability matrix with regularized selectivity matrix and the same bioactivity matrix. We calculated the error rate between regularized and non-regularized vulnerability matrices through RMSE.

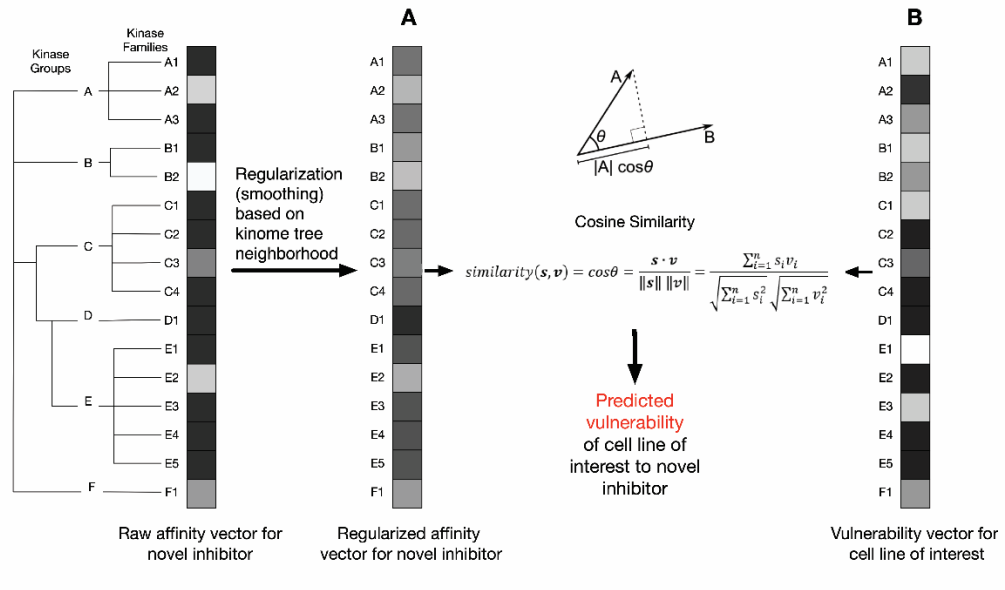


Figure 3.7: The detailed figure of the vulnerability prediction method based on regularization (*selectivity score vector (**s**)* and *vulnerability score vector (**v**)*).

3.3.4. Implementation of the Cosine Similarity

The cosine similarity is used to measure the similarity between two vectors by calculating the cosine angle between them to determine if two vectors point in the same direction (Han, Kamber, & Pei, 2012). We used cosine similarity after the matrix multiplication step to measure the similarity between the non-regularized selectivity vector and bioactivity vector and then to measure the similarity between the regularized selectivity vector and bioactivity vector, as follows:

$$similarity(\boldsymbol{s},\boldsymbol{v}) = cos\theta = \frac{\boldsymbol{s}\cdot\boldsymbol{v}}{\|\boldsymbol{s}\|\,\|\boldsymbol{v}\|} = \frac{\sum_{i=1}^{n} s_i v_i}{\sqrt{\sum_{i=1}^{n} s_i^2}\sqrt{\sum_{i=1}^{n} v_i^2}} \qquad (2.3)$$

*(selectivity score vector (**s**)* and *vulnerability score vector (**v**)*)

We calculated the error percentage between these cosine similarities using RMSE through LOOCV method. We applied this procedure to all three datasets. As a result of this method, we obtained higher error rates when the regularized

version of the selectivity scores of inhibitors is not correlated with a vulnerability score of kinase families in the cell lines based on target affection of the inhibitors.

3.3.5. Cross-Validation

Cross-validation is a statistical model evaluation method. One of the distinctive cases of k-fold cross-validation methods is Leave-one-out cross-validation (LOOCV). In this method, k equals to the number of occurrences in the data where in each iteration except for one observation all the data are used for training and the model is tested on that particular observation (Refaeilzadeh, Tang, & Liu, 2009)□. We used a LOOCV method to assess the error percentage between regularized and non-regularized regression models, wherein each fold one inhibitor is completely removed and the model is trained on the remaining inhibitors.

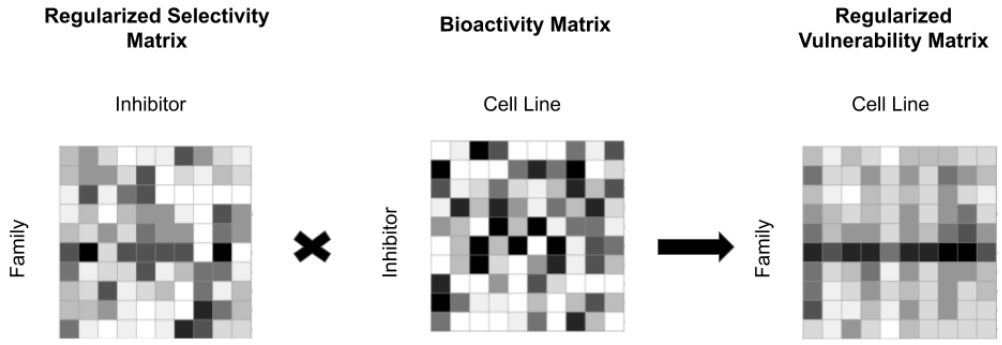


Figure 3.8: Representation of the kinome tree-based regularized vulnerability score prediction. The third step of the methodology is the regularized vulnerability score prediction method based on the kinome tree regularized selectivity matrix and bioactivity matrix multiplication.

3.4. Pathway Enrichment Analysis and Target Specificity in Enriched Pathways

Vulnerability scores obtained from the kinase family regularization or scores from generated using the kinome tree topology were used to perform the enrichment analysis to identify the pathways or cellular processes which are similarly inhibited by small-molecule kinase inhibitors. The vulnerability scores were truncated from 0.05 for the noise reduction then normalized for the enrichment analysis. Enrichr tool (Chen et al., 2013) was used through "enrichR" R package which provides an interface to several databases of 'Enrichr web-based enrichment analysis tool'. Enrichment analysis was performed with several gene sets which are Panther_2016, GO_Biological_Process_2018, NCI-Nature_2016, and KEGG_2019_Human to examine whether enrichment results are correlated between the results from enrichment analysis with different gene sets. Enrichr tool reports four different scores which are p-value, q-value, Z-score (rank score), and the combined score of the enrichment result. Enrichr implements *Fisher's exact test* to compute p-value of the enrichment, which is later corrected using *Benjamini-Hochberg method* to obtain *adjusted p-value* or *q-value* for multiple hypothesis testing. *Z-score* or *rank score* is applied by Enrichr to correct the Fisher's exact test p-value by considering deviation from an expected rank since it produces lower p-values for longer gene lists (Chen et al., 2013). Enrichr *combined score* is computed by the logarithm of multiplied p-value and z-scores. *combined score* ranking is then compared with other scoring methods (Kuleshov et al., 2016). Enriched pathways were determined with the cutoff p-adj < 0.05 and positive combined scores. Based on the enrichment analysis some of the inhibitors were found to be distributed as outliers. The distinctly aligned inhibitors were identified to test the difference between the outlier and non-outlier inhibitors according to RMSE result in respect to their selectivity to kinase families we used Mann-Whitney U test. As a result of this analysis, we observed those outlier inhibitors tend to be more selective while non-outliers tend to have multiple targets although some of the inhibitors do not follow this trend. To identify enriched outlier and non-outlier inhibitors in significantly targeted pathways, the hypergeometric test which takes the size of the overlap between the inhibitor set and the list of all inhibitors (the background) as parameters, and without replacement, was applied using "phyper" function in R.

Figure 3.9: Implementation of the pathway enrichment analysis and defining target specificity in enriched pathways. The last step of the methodology is defining enriched pathways and targeted kinase genes in these pathways by their adjusted p-value and the combined score to determine overrepresented specific or multi-kinase inhibitors targeting these genes through the hypergeometric test.

CHAPTER 4

4.RESULTS

4.1.Vulnerability Scores of the Kinase Families

The regression model that we applied in thesis allowed us to obtain first the small molecule kinase inhibitor "Selectivity matrix" and the "Bioactivity matrix" and then the "vulnerability matrix" through matrix-matrix multiplication which finally predicts the vulnerability score of the kinase families specific to primary liver cancer cell lines for CanSyL, CCLE, and GDSC datasets. These scores are then allowed us to analyze the effect and the efficiency of the kinase inhibitors on HCC cells. Hence, we aimed to prioritize kinase families and the cellular events that they involved through enrichment tools. We obtained the cell-kinase vulnerability scores with or without using kinome tree topology and compared their efficiency for this prioritization. When we used the kinome tree family topology we were able to regularize the kinome-tree based selectivity matrix by distributing half of the selectivity score of the kinase family between kinase group in which that family belongs to, based on the assumption that off-target interactions are more likely to occur for the closely related kinases. Kinome tree topology-based the vulnerability matrix scores showed that the most sensitive kinase families are CDK, EGFR, MAPK, PIKK, PKC, Src, CK2, CK1, Aur, Abl in ranking order in HCC cells. Moreover, regularization step reveals that GSK and CLK are effectively sensitive in the CanSyL dataset (Figure 4.1). For the CCLE dataset, the most vulnerable kinase families are EGFR, BRD, STE7, Src, MAPK, STE20, RAF, PDGFR, Eph, and FGFR. Furthermore, regularization step uncovers that RIO, BCR and Alpha kinase families are vulnerable in some of the poorly differentiated liver cancer cell-lines while Axl is vulnerable in non-aggressive liver cancer cell lines (Figure 4.2). In GDSC dataset, PIKK, MAPK, BRD, Src, EGFR, PDGFR, VEGFR, STE7, CDK are the most sensitive kinase families based on vulnerability matrix and Aur, STE20 kinase families are efficiently sensitive kinase families based on regularization step (Figure 4.3).

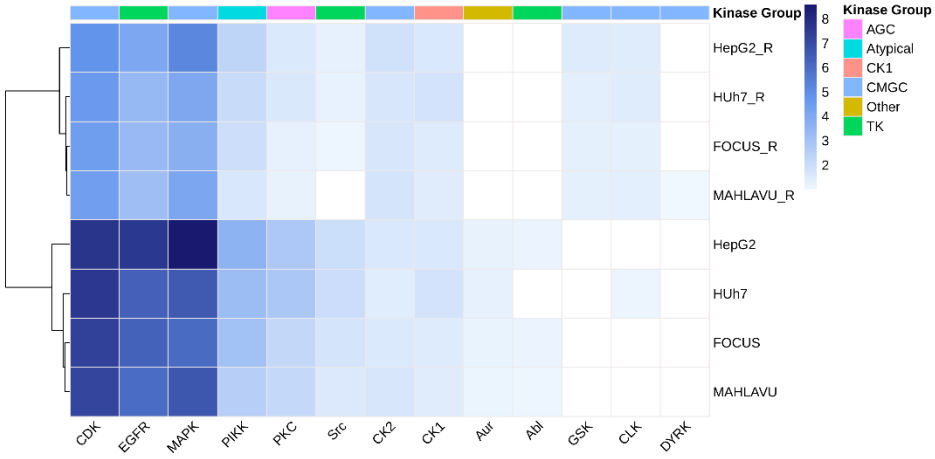


Figure 4.1: The most vulnerable ten kinase families in CanSyL dataset for both kinome tree regularized and non-regularized vulnerability scores. Kinome tree regularized scores were represented with "_R" for each cell line.

The vulnerability matrix scores of the CanSyL dataset indicates the most sensitive kinase families which are CDK, EGFR, MAPK, PIKK, PKC, Src, CK2, CK1, Aur, Abl in ranking order.

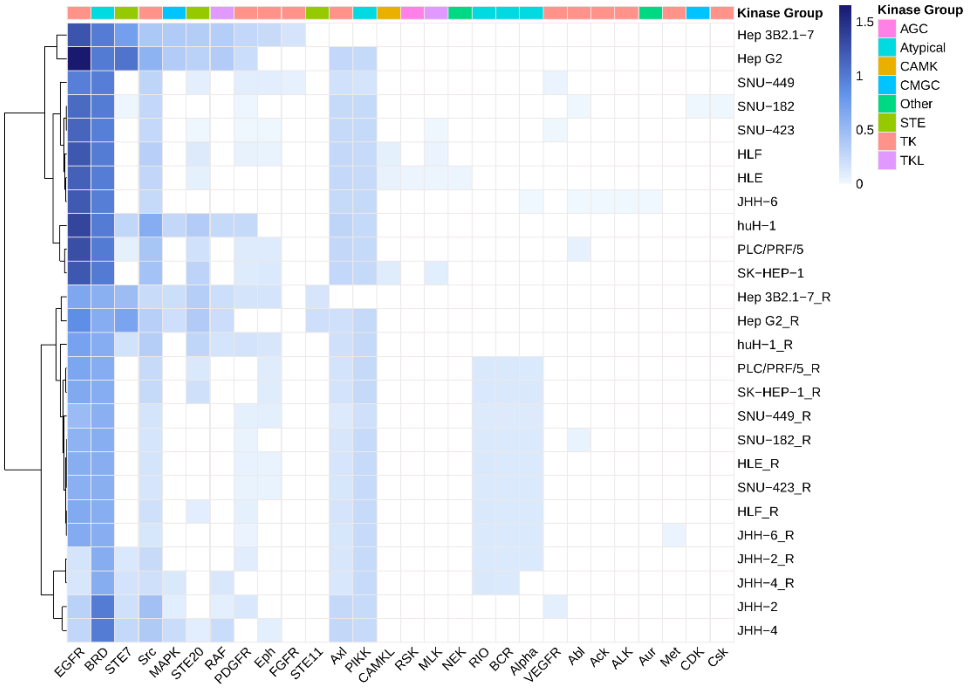


Figure 4.2: The most vulnerable ten kinase families in CCLE dataset for both kinome tree regularized and non-regularized vulnerability scores. Kinome tree regularized scores were represented with "_R" for each cell line.

For the CCLE dataset, the most vulnerable kinase families are EGFR, BRD, STE7, Src, MAPK, STE20, RAF, PDGFR, Eph, and FGFR. Furthermore, regularization step uncovers that RIO, BCR and Alpha kinase families are vulnerable in aggressive liver cancer cell-lines while Axl is vulnerable in non-aggressive liver cancer cell lines except Hep3B2.1-7 (Figure 4.2).

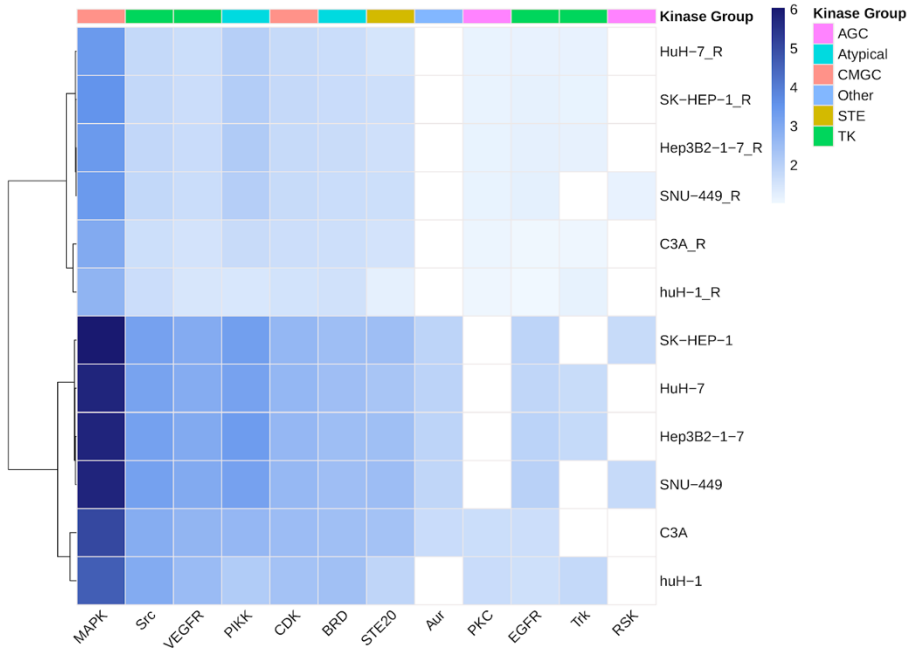


Figure 4.3: The most vulnerable ten kinase families in GDSC dataset for both kinome tree regularized and non-regularized vulnerability scores. Kinome tree regularized scores were represented with "_R" for each cell line.

We selected the most vulnerable ten kinase families for all three datasets to examine small molecule kinase inhibitor target prioritization. According to vulnerability scores of all three datasets, most kinase families showed similar trends in CanSyL and GDSC datasets with kinome tree topology non-regularized and regularized scores. The most efficiently targeted kinase group was CMGC for CanSyL and TK for both CCLE and GDSC datasets. In CanSyL dataset Aur, Abl, GSK and CLK kinase families had different patterns in the regularized and non-regularized scores. Kinases belonging to GSK and CLK families can be targeted efficiently with the inhibitors in the dataset while Aur and Abl cannot be targeted within the top 10 vulnerable kinase families in the dataset (Figure 4.1). Regularization reveals kinases belonging to RIO, BCR and Alpha kinase families may be affected by inhibitors in the CCLE dataset in aggressive cell lines specifically (Figure 4.2).

PKC, Trk, and Aur kinase families also have different patterns for their regularized and non-regularized vulnerability scores in GDSC dataset. Kinases belonging to PKC and Trk families can be targeted efficiently with the inhibitors in the dataset while Aur family cannot be targeted as efficiently as families within the top 10 vulnerable kinase families in the dataset (Figure 4.3). EGFR, PIKK, and Src are the most vulnerable kinase families according to both regularized and non-regularized vulnerability matrices for all three datasets. This may suggest that the most targetable kinase families with kinase inhibitors in HCC cell-lines are EGFR, PIKK, and Src.

4.2. Error Rate of the Human Kinome Based on Regularized Selectivity and Non-regularized Vulnerability Scores

Results with this methodology using the CanSyL dataset applying LOOCV have achieved promising predictions (median RMSE between 2.5-4 %) for the vulnerability matrix based on regularization of the human kinome tree, with no bias in the estimates (Figure 4.4). When we scaled up the approach to the CCLE and GDSC datasets, our method achieved good cross-validation performance for most drugs in GDSC (median RMSE within 4%) and in CCLE (median RMSE between 2-5%) (Figure 4.5-4.6). Outlier and non-outlier inhibitors, according to RMSE result, and with respect to their specificity to kinase families, are significantly different from each other in all datasets according to the Mann-Whitney U test ($p<0.05$). This difference in specificity suggests that outlier inhibitors are more specific inhibitors and non-outlier inhibitors are mostly multi-kinase inhibitors. As a result of this analysis, we observed those outlier inhibitors tend to target specific kinase families while non-outliers tend to have multiple targets although some of the inhibitors do not follow this trend.

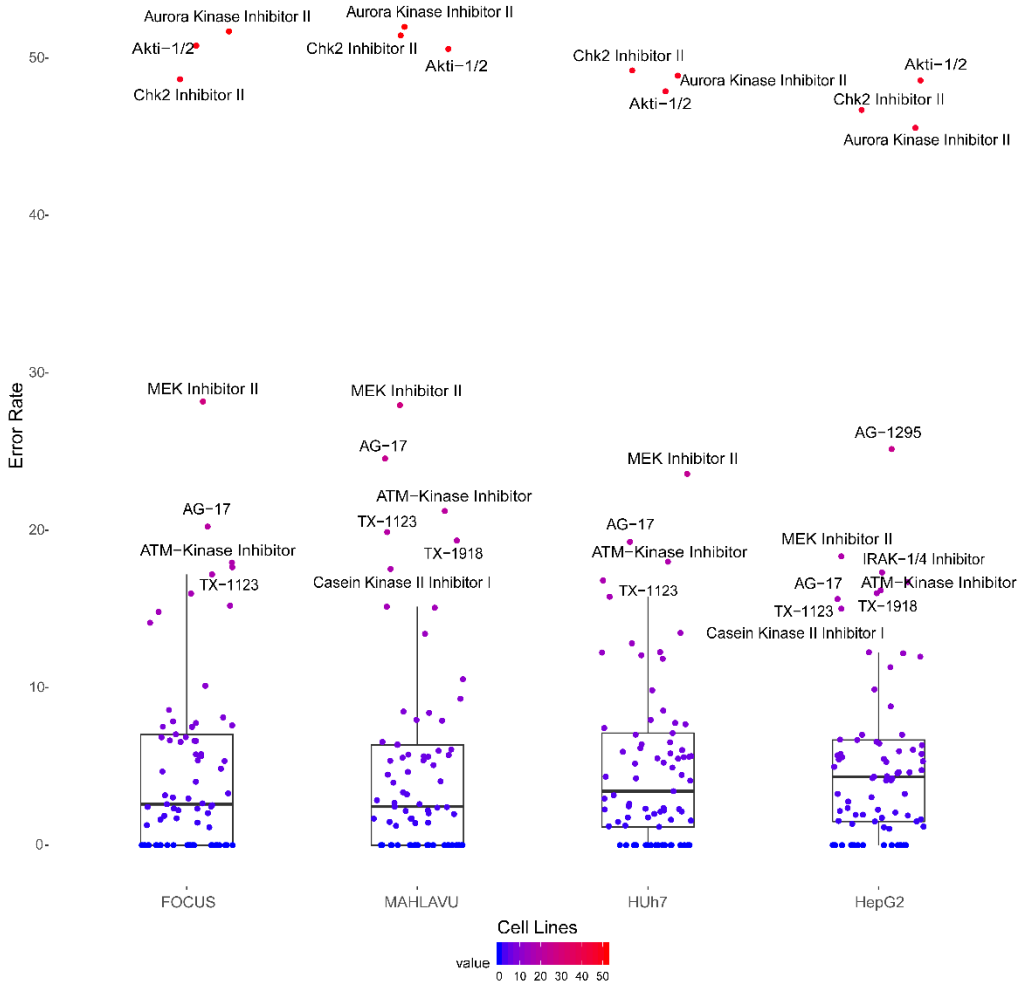


Figure 4.4: Error rate of the model based on CanSyL dataset. The error rate of the similarity between kinome tree-based regularized selectivity score and the non-regularized vulnerability score was calculated using RMSE.

As a result of the LOOCV test, outliers indicate inhibitors with the highest error rates with respect to their individual RMSE values in percentage. Outlier kinase inhibitors are Aurora Kinase Inhibitor II, Akti-1/2, Chk Inhibitor II, MEK Inhibitor II, AG-17, ATM Kinase Inhibitor, and TX-1123 for the poorly differentiated FOCUS cell-line in the CanSyL dataset. In addition to previous outliers of the FOCUS cell-line, TX-1918, and Casein Kinase II Inhibitor I are outliers for the poorly differentiated Mahlavu cell-line. For the well-differentiated Huh7 cell-line, outlier inhibitors are MEK Inhibitor II, AG-17, ATM Kinase Inhibitor, and TX-1123. In addition to outliers of the Huh-7 cell-

line, IRAK-1/4, TX-1918, and Casein Kinase II Inhibitor I are outliers for the well-differentiated HepG2 cell-line.

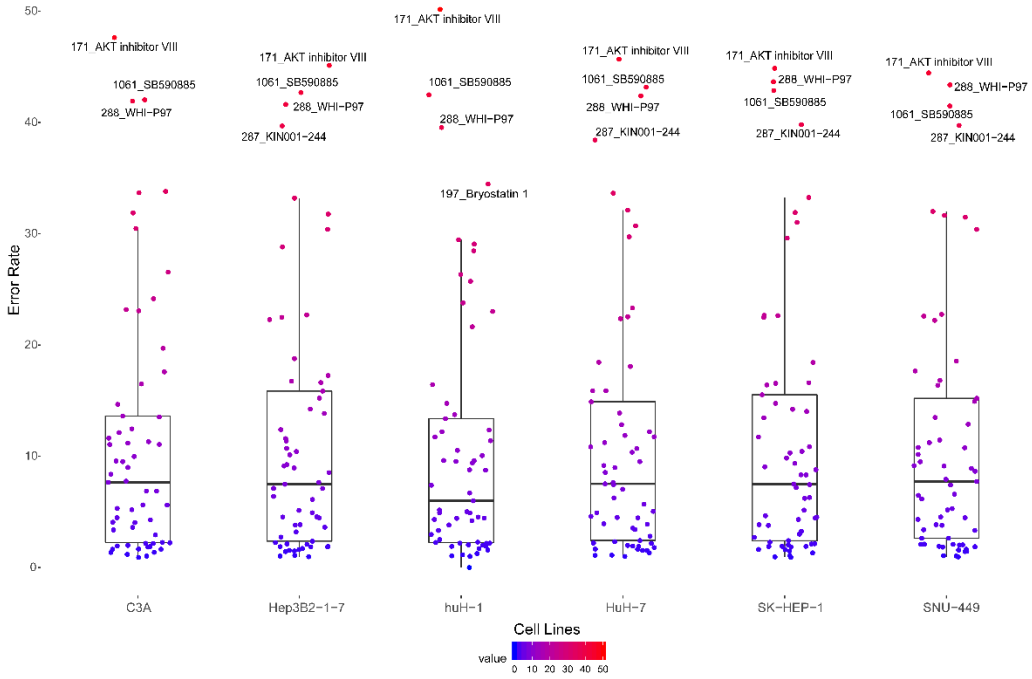


Figure 4.5: Error rate of the model based on GDSC dataset. The error rate of the similarity between kinome tree-based regularized selectivity score and the non-regularized vulnerability score was calculated using RMSE.

Outliers indicate inhibitors with the highest error rates among all inhibitors in GDSC dataset in figure 4.6. Akt Inhibitor VIII, SB590885, and WHI-P97 inhibitors are outliers for all cell lines in GDSC dataset. Additionally, Hep3B2-1-7, Huh-7, SK-HEP-1, and SNU-449 have KIN001-224 as an outlier. Moreover, Bryostatin-1 is an outlier for the only hUH-1 cell line.

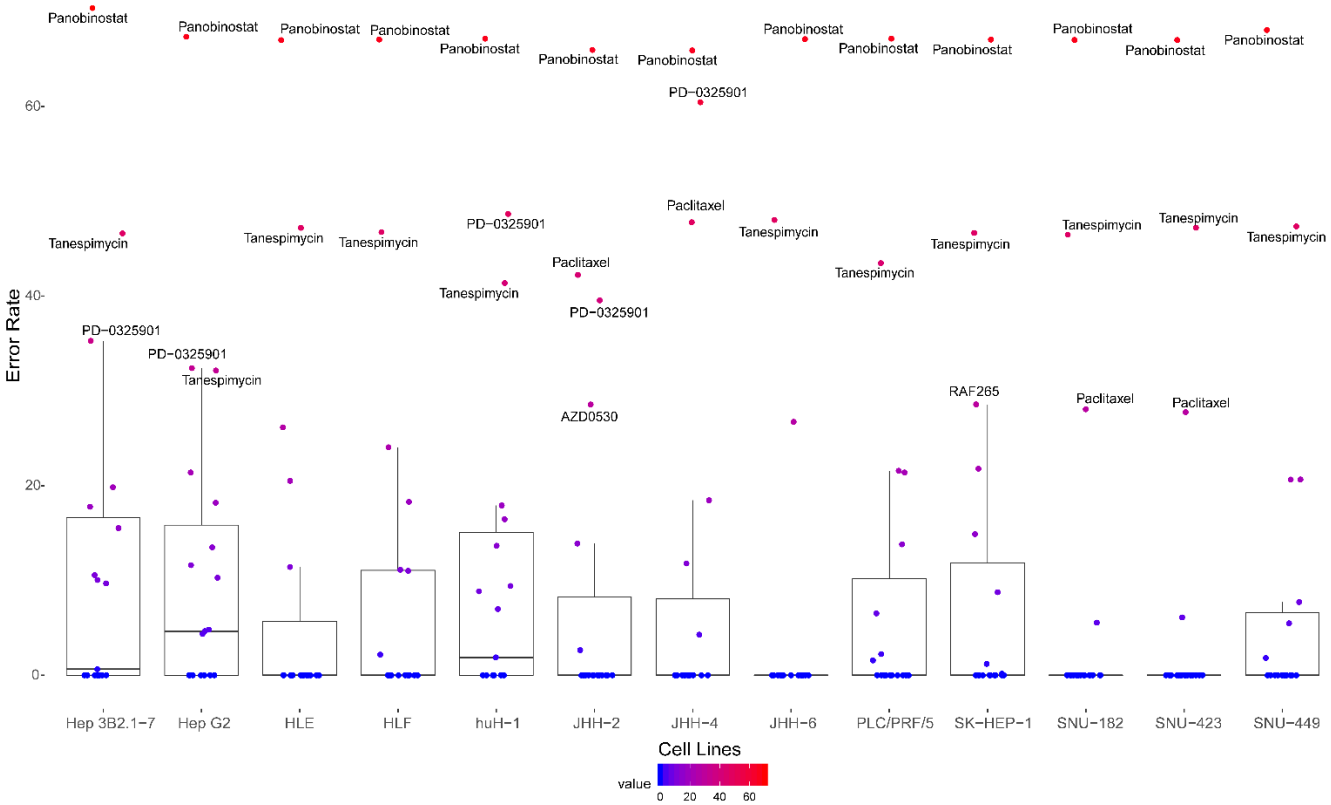


Figure 4.6: Error rate of the model based on CCLE dataset. Error rate of the similarity between kinome tree-based regularized selectivity score and the non-regularized vulnerability score was calculated using RMSE.

Outliers indicate inhibitors with the highest error rates among all inhibitors in CCLE dataset in figure 4.5. Panobinostat has the highest error rate for all cell lines. Another inhibitor Tanespimycin is an outlier for all cell lines except for the JHH-2 and JHH-4.

4.3. Pathway Enrichment Analysis with Vulnerability Scores of Kinase Families

The vulnerability scores and their ranked data is also used to identify enrich cellular events and pathways upon treatment with the small molecule kinase inhibitors. The enrichment results allowed us to effectively predict how and through which cellular events these inhibitors achieve their bioactivities.

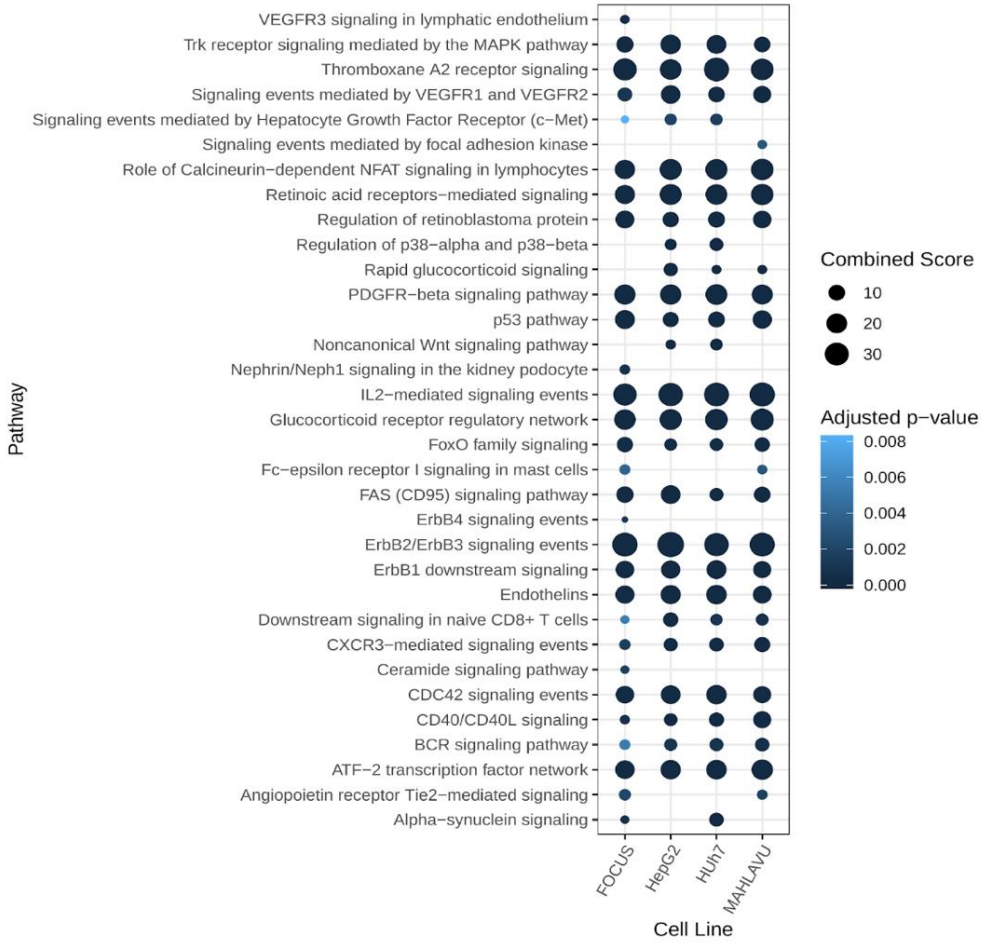


Figure 4.7: Pathway enrichment results for the CanSyL dataset using the NCI-Nature gene set. Enrichment analysis was performed using non-regularized vulnerability scores.

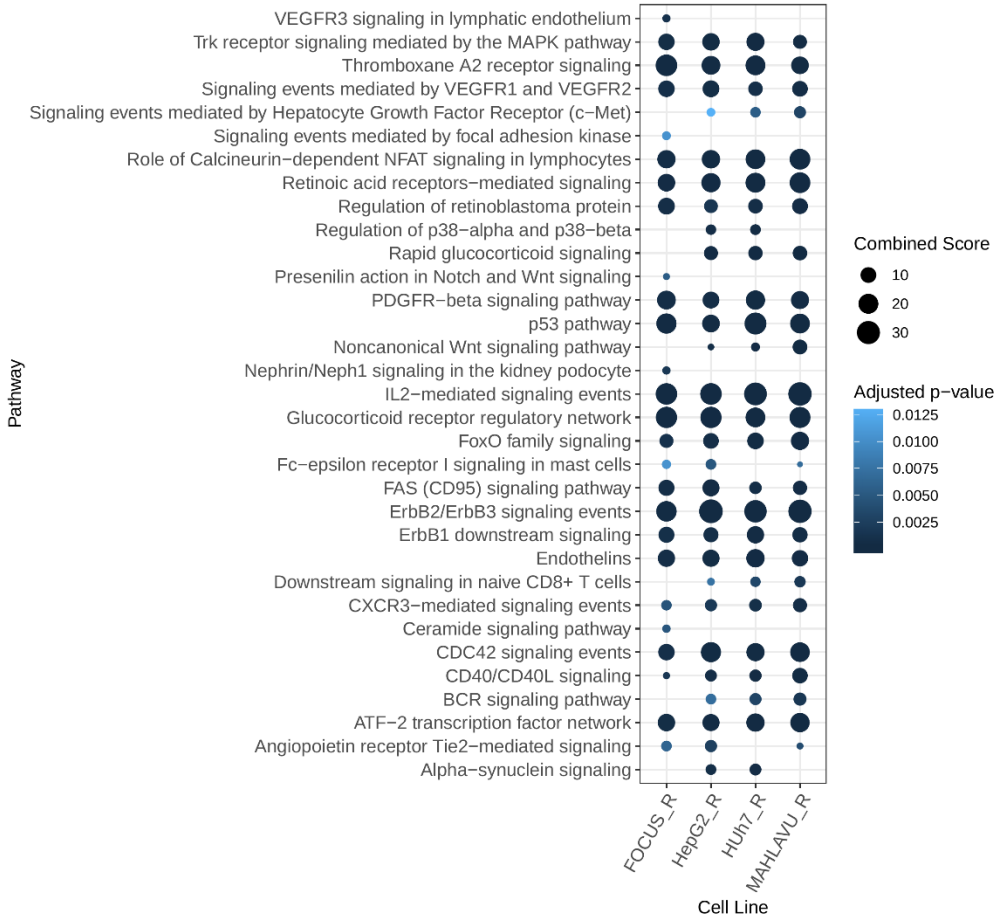


Figure 4.8: Pathway enrichment results for the CanSyL dataset using the NCI-Nature gene set. Enrichment analysis was performed using kinome tree regularized vulnerability scores.

We also observed that when kinome tree topology used for the calculation of vulnerability matrix, the specific effect of the kinase inhibitors on individual cell lines become more significantly predicted. For example “Presenilin action in Notch and Wnt signaling” enrichment term is specifically associated (Adjusted p-value < 0.05 and combined Score > 0) with FOCUS cell-line with kinome tree regularized vulnerability scores (Figure 4.8), although it is not significantly targeted with non-regularized vulnerability scores (Figure 4.7). According to enrichment results performed with kinome tree-based regularized vulnerability scores, possible dysregulated growth factors such as Hepatocyte growth factor (c-Met), VEGFR1 and VEGFR2 related signaling, PDGFR-beta signaling which causes HCC formation, can be targeted in both well and poorly differentiated HCC cell lines with the inhibitors in the CanSyL dataset.

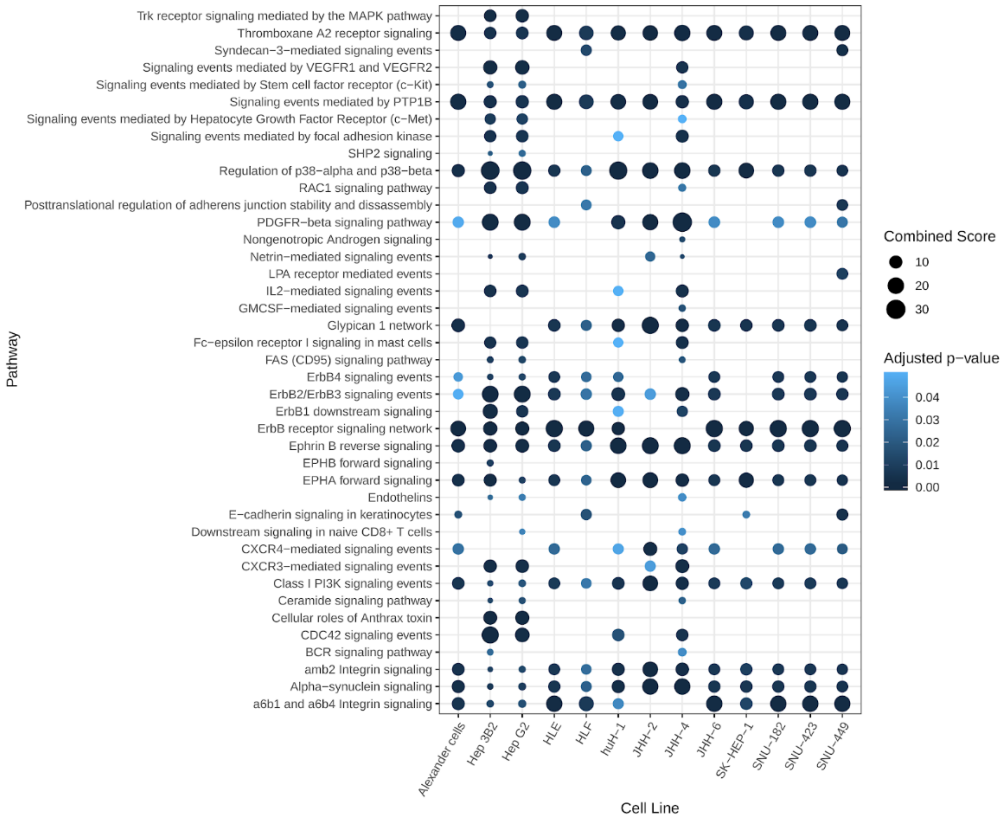


Figure 4.9: Pathway enrichment results for the CCLE dataset using the NCI-Nature gene set. Enrichment analysis results with non-regularized vulnerability scores.

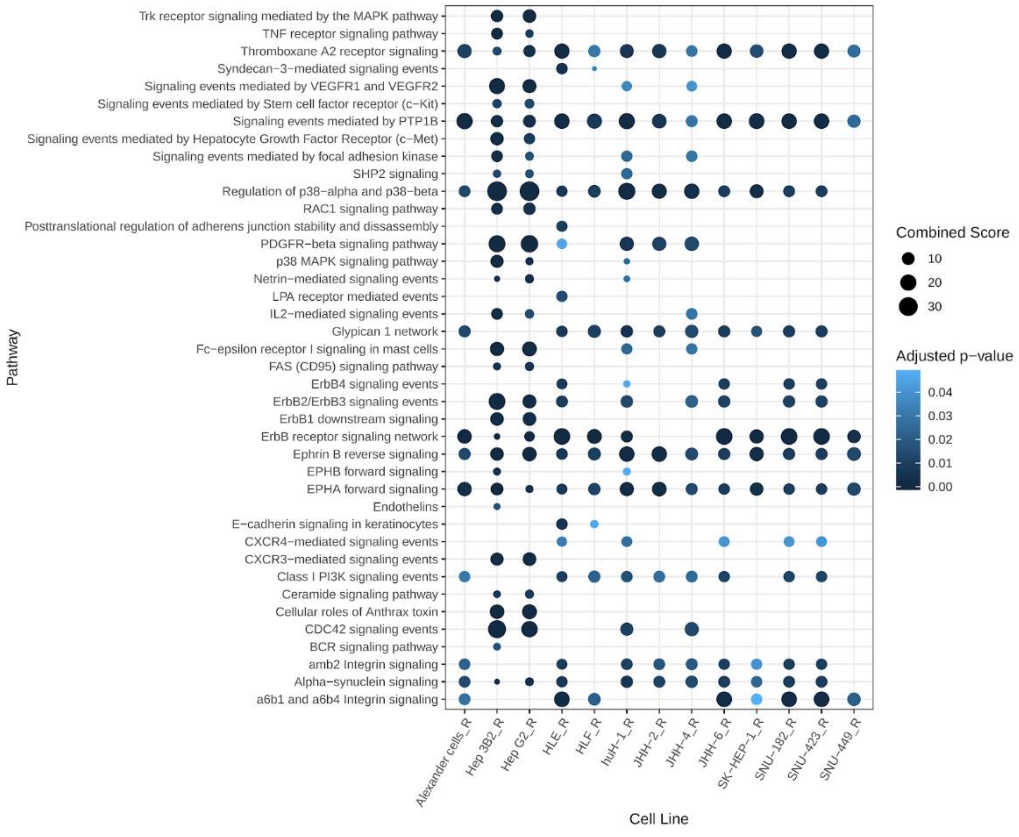


Figure 4.10: Pathway enrichment results for the CCLE dataset using the NCI-Nature gene set. Enrichment analysis was performed using kinome tree regularized vulnerability scores.

Human kinome tree-based regularization reveals that "p38 MAPK signaling pathway" can be targeted in Hep3B, HepG2 and hUH1 cell-lines and "TNF receptor signaling pathway" can be targeted in all cell-lines significantly (Figure 4.10) whereas these pathways were not significantly identified according to enrichment analysis performed with non-regularized scores (Figure 4.9). Since CCLE dataset does not contain inhibitor which specifically acts on p38 MAPK, there is no significant enrichment result for "p38 MAPK signaling pathway" in the analysis result performed with non-regularized scores. However, through regularization, we can see the p38 MAPK signaling pathway is affected by off-target effects of some of the kinase inhibitors. According to this analysis, tyrosine kinase inhibitors Erlotinib, Lapatinib, Sorafenib, Dovitinib, Vandetanib, and Nilotinib, CDK4 and CDK6 inhibitor Palbociclib, receptor tyrosine Kinase ALK, and HGFR inhibitor Crizotinib and c-Met tyrosine kinase

inhibitor PHA-665752 (Kim et al., 2019) have off-target effects on p38 MAPK signaling pathway.

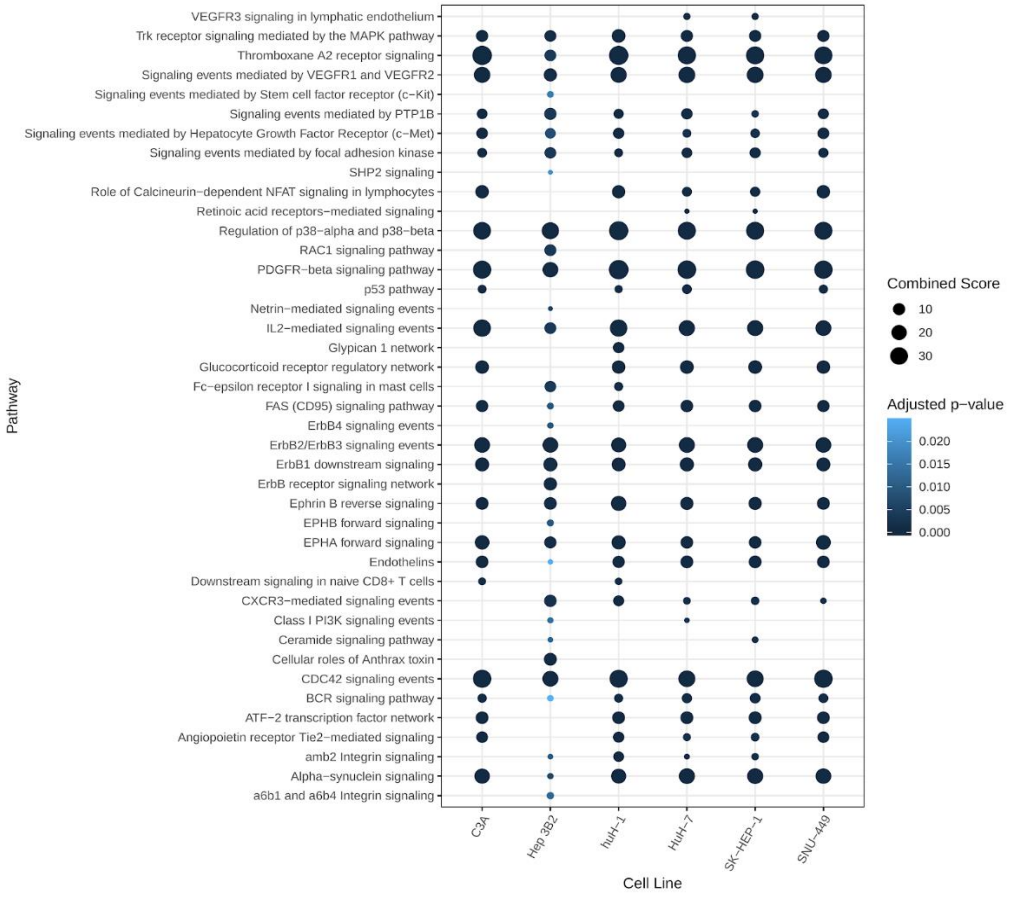


Figure 4.11: Pathway enrichment results for the GDSC dataset using the NCI-Nature gene set. Enrichment analysis was performed using non-regularized vulnerability scores.

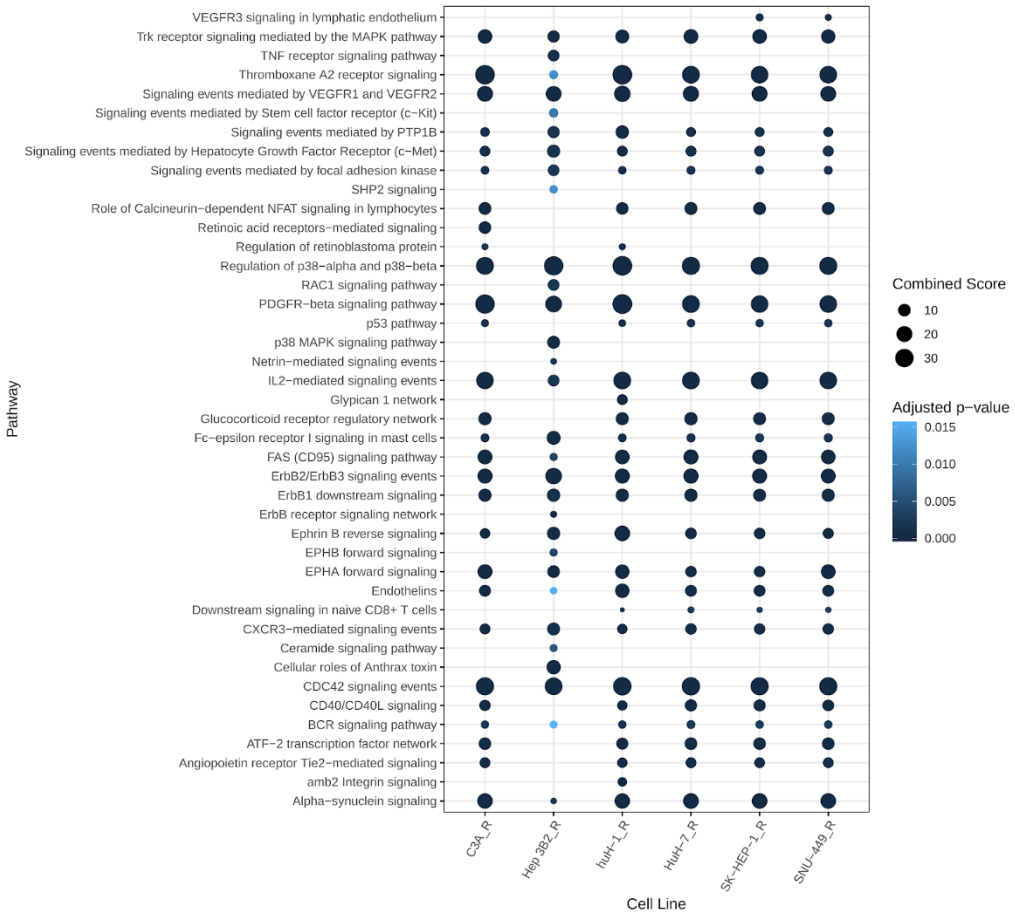


Figure 4.12: Pathway enrichment results for the GDSC dataset using the NCI-Nature gene set. Enrichment analysis was performed using kinome tree regularized vulnerability scores.

For the GDSC dataset, through our methodology, we have found the pathways that are identified by off-target effects of the kinase inhibitors are “CD40/CD40L signaling”, “Regulation of retinoblastoma protein”, p38 MAPK signaling pathway”, and “TNF receptor signaling pathway” (Figure 4.12).

4.4. Significance of the Specific and Multi-Kinase Inhibitors in the Targeted Pathways

The p-values, for the overrepresentation of the outlier and non-outlier inhibitor sets among all inhibitors used in all datasets, were calculated using the Hypergeometric test. As a result of this method, pathways that are targeted by selective or multi-kinase inhibitors were identified. In the CanSyl dataset "Regulation of Rb protein" and "Role of Calcineurin-dependent NFAT signaling in lymphocytes" are targeted effectively by selective kinase inhibitor MEK Kinase inhibitor II (Figure 4.13).

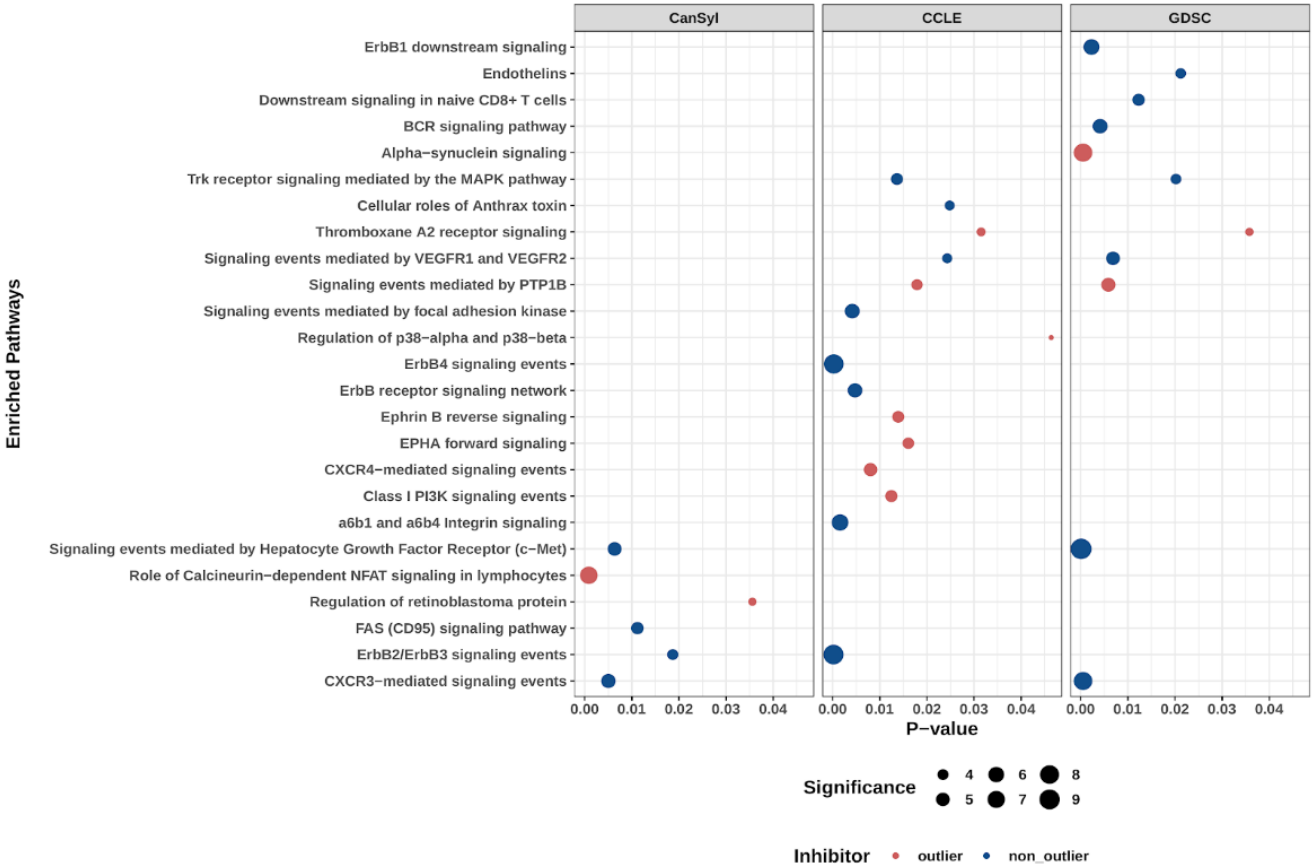


Figure 4.13: Overrepresented outlier and non-outlier inhibitor set. The p-values for the overrepresented outlier and non-outlier inhibitors calculated using hypergeometric test among all inhibitors used in all three datasets.

CHAPTER 5

5. DISCUSSION

In this study, we present a regression model to predict the effectiveness of a new inhibitor in a family-based manner using the human kinome tree. It is well known that in addition to their targeted signaling pathway small molecule kinase inhibitors can affect other pathways by "off-target" or "pathway cross-talk" effects. Our objective in this study was to predict these off-target effects of kinase inhibitors by regularizing the regression space based on the kinome tree family topology.

The need to study the selectivity of kinase inhibitors has led to the development of novel quantitative methods to measure the selectivity of these inhibitors. Examples of these quantitative methods are "Selectivity Score" and "Gini coefficient" selectivity measurement metrics. Selectivity Score is calculated for each compound by dividing the number of kinases bound by an inhibitor with a specific affinity score to the total number of kinases experimented. Gini coefficient selectivity measurement method considers the magnitude of inhibition which is based on kinase activity measured at a specific ATP concentration. In this method first, to find total inhibition, the sum of magnitudes of inhibition for all kinases is calculated. Second, kinase activities are sorted in increasing order. After a cumulative fraction of total inhibition is plotted against the cumulative fraction of kinase activities, the Gini coefficient is calculated through the Lorenz Curve. The method we presented in this study provides ranking between the catalog of inhibitors in terms of the combination of their selectivity and bioactivity values. As opposed to the selectivity score introduced by Karaman et. al., our methodology does not require a cut-off value which enables scoring of small molecule kinase inhibitors without depending on defining any threshold. Moreover, in contrast with the Gini coefficient selectivity method explained in chapter 2.2, our methodology is population dependent. In another study, a novel quantitative method has been developed to predict the adverse effects of drugs related to the inhibition of the kinase targets using linear algebra. They have predicted associations between kinase targets and adverse event frequencies in human patients through matrix multiplication using publicly available kinome-wide experimental data (inhibitor-target dissociation constant) (Yang et al., 2010). We have employed matrix multiplication, as in this study, to predict associations between kinase families and HCC cell lines.

Results with this methodology using the CanSyL dataset applying LOOCV have achieved promising predictions (median RMSE between 2.5-4 %) for the vulnerability matrix based on regularization of the human kinome tree family topology. Our approach also accomplished promising cross-validation performance most drugs in GDSC (median RMSE within 6-7.5%) and in CCLE (median RMSE between 2-5%) as well. Outlier and non-outlier inhibitors, according to RMSE result, and with respect to their specificity to kinase families, are significantly different from each other in all datasets according to the Mann-Whitney U test ($p<0.05$). This difference in specificity may assist to distinguish specific inhibitors (outlier) and multi-kinase inhibitors (non-outlier inhibitors) through our method for future studies.

After applying our methodology, we found that EGFR, PIKK, and Src are the most vulnerable kinase families in HCC cell-lines according to both regularized and non-regularized vulnerability matrices in all three datasets. This suggests that the most efficiently targeted kinase families in Hepatocellular Carcinoma cell-lines with kinase inhibitors targeting EGFR, PIKK, and Src (Figure 4.1). However, these inhibitors did not show cell line specificity, hence indicating targeting these kinases may not provide tumor or cell-specific actions which cannot be used for personized therapies.

In a previous study, based on activity results, it has been found that Aurora Kinase Inhibitor II between four Aurora kinase inhibitors in the EMD Millipore collection was the most potent and selective pan-Aurora inhibitor (CAS 331770-21-9, S = 0.04). According to the same study, Chk2 inhibitor II (CAS 516480-79-8, s = 0.01) was highly specific for Chk2 (Gao et al., 2013). Based on prior findings, Aurora kinase inhibitor II and Chk Inhibitor II are selective and specific inhibitors for their targets. Likewise, we obtained a high error rate for these inhibitors after regularization based on the kinome tree which indicates the specificity of the kinase inhibitor in our model (Figure 4.4). In another study, it has been found that type II kinase inhibitors tend to be more selective since they are non-ATP competitors (Blanc et al., 2013). Accordingly, we obtained a high error rate for the RAF265, one of the type II inhibitor, in the CCLE dataset (Figure 4.5). In addition, for PD0325901, which is an extremely selective type III (allosteric) kinase inhibitor (P. K. Wu & Park, 2015) in CCLE dataset, we acquired a high error rate as well. Similarly, we obtained a high error rate for the Akti-1/2 inhibitor in the CanSyL dataset. The Akti-1/2 inhibitor is an allosteric (type III) and highly selective inhibitor which blocks Akt1 and Akt2 but not Akt3 (Gilot, Giudicelli, Lagadic-Gossmann, & Fardel, 2010). These findings indicate that with our methodology selective small-molecule kinase inhibitors can be identified.

Enrichment results of this thesis show targetable pathways through inhibitors in the dataset for each HCC cell line. Wnt signaling is one of the key signaling pathways activated in HCC leading transformation of the normal liver to HCC (Zaret & Grompe, 2008). Aihara et al. showed that primary liver cancer tumor growth can be inhibited using small molecule inhibitor MO-I-1100 which reduces activation of Notch signaling in FOCUS cells (Aihara et al., 2014). In accordance with these, we found that "Presenilin action in Notch and Wnt signaling" enrichment term is specifically associated with FOCUS cell-line with kinome tree regularized vulnerability scores, although it is not significantly targeted with non-regularized vulnerability scores. This finding suggests that EGFR inhibitors tested in CanSyL dataset such as LY294002, AG1478, and PD98059 or Curcumin Curcuma Longa L may be used to target Wnt signaling to reduce aggressive primary liver cancer growth in accordance with previous studies (Mimeault & Batra, 2011; I. Paul, Bhattacharya, Chatterjee, & Ghosh, 2013; Tan et al., 2005) . Additionally, regularization method provides a distinction between some tumor cell lines as can be seen from poorly differentiated (FOCUS and MAHLAVU) and well-differentiated (HepG2 and Huh-7) cell-lines in the CanSyL dataset (Figure 4.8).

For future work, different cut-off values can be used to increase the model performance and reduce the noise in the data. The method may be applied to a different set of target proteins with biological metadata such as nuclear receptor inhibitors which has a classification similar to the kinome tree topology.

REFERENCES

Aihara, A., Huang, C. K., Olsen, M. J., Lin, Q., Chung, W., Tang, Q., … Wands, J. R. (2014). A cell-surface β-hydroxylase is a biomarker and therapeutic target for hepatocellular carcinoma. *Hepatology*. https://doi.org/10.1002/hep.27275

Alberts, S. R., Fitch, T. R., Kim, G. P., Morlan, B. W., Dakhil, S. R., Gross, H. M., & Nair, S. (2012). Cediranib (AZD2171) in patients with advanced hepatocellular carcinoma: A phase II north central cancer treatment group clinical trial. *American Journal of Clinical Oncology: Cancer Clinical Trials*. https://doi.org/10.1097/COC.0b013e3182118cdf

Anastassiadis, T., Deacon, S. W., Devarajan, K., Ma, H., & Peterson, J. R. (2011). Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity. *Nature Biotechnology*. https://doi.org/10.1038/nbt.2017

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., … Garraway, L. A. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature, 483*(7391), 603–607. https://doi.org/10.1038/nature11003

Bhullar, K. S., Lagarón, N. O., McGowan, E. M., Parmar, I., Jha, A., Hubbard, B. P., & Rupasinghe, H. P. V. (2018). Kinase-targeted cancer therapies: progress, challenges and future directions. *Molecular Cancer, 17*(1), 48. https://doi.org/10.1186/s12943-018-0804-2

Blanc, J., Geney, R., & Menet, C. (2013). Type II Kinase Inhibitors: An Opportunity in Cancer for Rational Design. *Anti-Cancer Agents in Medicinal Chemistry*. https://doi.org/10.2174/1871520611313050008

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*. https://doi.org/10.3322/caac.21492

Buontempo, F., Ersahin, T., Missiroli, S., Senturk, S., Etro, D., Ozturk, M., … Neri, M. L. (2011). Inhibition of Akt signaling in hepatoma cells induces apoptotic cell death independent of Akt activation status. *Investigational New Drugs*. https://doi.org/10.1007/s10637-010-9486-3

Cainap, C., Qin, S., Huang, W. T., Chung, I. J., Pan, H., Cheng, Y., … El-Nowiem, S. (2015). Linifanib versus sorafenib in patients with advanced hepatocellular carcinoma: Results of a randomized phase III trial. *Journal of Clinical Oncology*. https://doi.org/10.1200/JCO.2013.54.3298

Carmeliet, P., & Jain, R. K. (2000). Angiogenesis in cancer and other diseases. *Nature*. https://doi.org/10.1038/35025220

Chao, Y., Li, C. P., Chau, G. Y., Chen, C. P., King, K. L., Lui, W. Y., … Lee, S. D. (2003). Prognostic significance of vascular endothelial growth factor, basic fibroblast growth factor, and angiogenin in patients with resectable hepatocellular carcinoma after surgery. *Annals of Surgical Oncology*. https://doi.org/10.1245/ASO.2003.10.002

Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., … Ma'ayan, A. (2013). Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. https://doi.org/10.1186/1471-2105-14-128

Chiu, Y. Y., Lin, C. T., Huang, J. W., Hsu, K. C., Tseng, J. H., You, S. R., & Yang, J. M. (2013). KIDFamMap: A database of kinase-inhibitor-disease family maps for kinase inhibitor selectivity and binding mechanisms. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gks1218

Cohen, P. (2001). The role of protein phosphorylation in human health and disease. . *European Journal of Biochemistry*, *268*(19), 5001–5010. https://doi.org/10.1046/j.0014-2956.2001.02473.x

Cooper, G. M., & Hausman, R. E. (2007). The Development and Causes of Cancer. *The Cell: A Molecular Approach*. https://doi.org/10.1016/j.pbb.2009.09.001

Cox, K. J., Shomin, C. D., & Ghosh, I. (2011). Tinkering outside the kinase ATP box: Allosteric (type IV) and bivalent (type V) inhibitors of protein kinases. *Future Medicinal Chemistry*. https://doi.org/10.4155/fmc.10.272

Davies, M., Nowotka Michałand Papadatos, G., Dedman, N., Gaulton, A., Atkinson, F., Bellis, L., & Overington, J. P. (2015). ChEMBL web services: Streamlining access to drug discovery data and utilities. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkv352

Davis, M. I., Hunt, J. P., Herrgard, S., Ciceri, P., Wodicka, L. M., Pallares, G., … Zarrinkar, P. P. (2011). Comprehensive analysis of kinase inhibitor selectivity. *Nature Biotechnology*. https://doi.org/10.1038/nbt.1990

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*. https://doi.org/10.1186/s12859-016-1433-7

Ferguson, F. M., & Gray, N. S. (2018). Kinase inhibitors: The road ahead. *Nature Reviews Drug Discovery*. https://doi.org/10.1038/nrd.2018.21

Finn, R. (2013). Emerging targeted strategies in advanced hepatocellular carcinoma. *Seminars in Liver Disease*. https://doi.org/10.1055/s-0033-1333632

Finn, R. S., Aleshin, A., Dering, J., Yang, P., Ginther, C., Desai, A., … Slamon, D. J. (2013). Molecular subtype and response to dasatinib, an Src/Abl small molecule kinase inhibitor, in hepatocellular carcinoma cell lines in vitro. *Hepatology*, *57*(5), 1838–1846. https://doi.org/10.1002/hep.26223

Gao, Y., Davies, S. P., Augustin, M., Woodward, A., Patel, U. A., Kovelman, R., & Harvey, K. J. (2013). A broad activity screen in support of a chemogenomic map for kinase signalling research and drug discovery. *Biochemical Journal*. https://doi.org/10.1042/BJ20121418

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. https://doi.org/10.1038/nature11005

Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., … Overington, J. P. (2012). ChEMBL: A large-scale bioactivity database for drug discovery. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkr777

Gillet, J. P., Varma, S., & Gottesman, M. M. (2013). The clinical relevance of cancer cell lines. *Journal of the National Cancer Institute*, *105*(7), 452–458. https://doi.org/10.1093/jnci/djt007

Gilot, D., Giudicelli, F., Lagadic-Gossmann, D., & Fardel, O. (2010). Akti-1/2, an allosteric inhibitor of Akt 1 and 2, efficiently inhibits CaMKIα activity and aryl hydrocarbon receptor pathway. *Chemico-Biological Interactions*. https://doi.org/10.1016/j.cbi.2010.08.011

Graczyk, P. P. (2007). Gini coefficient: A new way to express selectivity of kinase inhibitors against a family of kinases. *Journal of Medicinal Chemistry*. https://doi.org/10.1021/jm070562u

Gross, S., Rahal, R., Stransky, N., Lengauer, C., & Hoeflich, K. P. (2015, May 1). Targeting cancer with kinase inhibitors. *Journal of Clinical Investigation*. American Society for Clinical Investigation. https://doi.org/10.1172/JCI76094

Han, J., Kamber, M., & Pei, J. (2012). *02. Getting to Know Your Data*. *Data Mining*. https://doi.org/10.1016/B978-0-12-381479-1.00002-2

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: The next generation. *Cell*. https://doi.org/10.1016/j.cell.2011.02.013

Huynh, H., Ngo, V. C., Fargnoli, J., Ayers, M., Khee, C. S., Heng, N. K., … Tran, E. (2008). Brivanib alaninate, a dual inhibitor of vascular endothelial growth factor receptor and fibroblast growth factor receptor tyrosine kinases, induces growth inhibition in mouse models of human hepatocellular carcinoma. *Clinical Cancer Research*. https://doi.org/10.1158/1078-0432.CCR-08-0509

Jemal, A., Siegel, R., Ward, E., Murray, T., Xu, J., & Thun, M. J. (2019). Cancer Statistics, 2019. *CA: A Cancer Journal for Clinicians*, *57*(1), 43–66. https://doi.org/10.3322/canjclin.57.1.43

Karaman, M. W., Herrgard, S., Treiber, D. K., Gallant, P., Atteridge, C. E., Campbell, B. T., … Zarrinkar, P. P. (2008). A quantitative analysis of kinase inhibitor selectivity. *Nature Biotechnology*. https://doi.org/10.1038/nbt1358

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., … Bolton, E. E. (2019). PubChem 2019 update: Improved access to chemical data. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gky1033

Knighton, D. R., Zheng, J., Ten Eyck, L. F., Ashford, V. A., Xuong, N. H., Taylor, S. S., & Sowadski, J. M. (1991). Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science*. https://doi.org/10.1126/science.1862342

Kooistra, A. J., & Volkamer, A. (2017). Kinase-Centric Computational Drug Development. In *Annual Reports in Medicinal Chemistry*. https://doi.org/10.1016/bs.armc.2017.08.001

Kufareva, I., & Abagyan, R. (2008). Type-II kinase inhibitor docking, screening, and profiling using modified structures of active kinase states. *Journal of Medicinal Chemistry*. https://doi.org/10.1021/jm8010299

Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., … Ma'ayan, A. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkw377

Lu, L., Hsu, C., & Cheng, A. (2016). Tumor Heterogeneity in Hepatocellular Carcinoma : Facing the Challenges, 128–138. https://doi.org/10.1159/000367754

Manning, G., Whyte, D. B., Martinez, R., Hunter, T., & Sudarsanam, S. (2002). The protein kinase complement of the human genome. *Science*, *298*(5600), 1912–1934. https://doi.org/10.1126/science.1075762

Mendez, D., Gaulton, A., Bento, A. P., Chambers, J., De Veij, M., Félix, E., … Leach, A. R. (2019). ChEMBL: Towards direct deposition of bioassay data. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gky1075

Metz, K. S., Deoudes, E. M., Berginski, M. E., Jimenez-Ruiz, I., Aksoy, B. A., Hammerbacher, J., … Phanstiel, D. H. (2018). Coral: Clear and Customizable Visualization of Human Kinome Data. *Cell Systems*. https://doi.org/10.1016/j.cels.2018.07.001

Mimeault, M., & Batra, S. K. (2011). Potential applications of curcumin and its novel synthetic analogs and nanotechnology-based formulations in cancer prevention and therapy. *Chinese Medicine*. https://doi.org/10.1186/1749-8546-6-31

Min, L., He, B., & Hui, L. (2011). Mitogen-activated protein kinases in hepatocellular carcinoma development. *Seminars in Cancer Biology*. https://doi.org/10.1016/j.semcancer.2010.10.011

Möbitz, H. (2015). The ABC of protein kinase conformations. *Biochimica et Biophysica Acta - Proteins and Proteomics*. https://doi.org/10.1016/j.bbapap.2015.03.009

Moeini, A., Cornellà, H., & Villanueva, A. (2012). Emerging Signaling Pathways in Hepatocellular Carcinoma. *Liver Cancer*, *1*(2), 83–93. https://doi.org/10.1159/000342405

Moon, W. S., Rhyu, K. H., Kang, M. J., Lee, D. G., Yu, H. C., Yeum, J. H., … Tarnawski, A. S. (2003). Overexpression of VEGF and angiopoietin 2: A key to high vascularity of hepatocellular carcinoma? *Modern Pathology*. https://doi.org/10.1097/01.MP.0000071841.17900.69

Muntane, J., J. De la Rosa, A., Docobo, F., Garcia-Carbonero, R., & J. Padillo, F. (2013). Targeting Tyrosine Kinase Receptors in Hepatocellular Carcinoma. *Current Cancer Drug Targets*. https://doi.org/10.2174/15680096113139990075

Negrini, S., Gorgoulis, V. G., & Halazonetis, T. D. (2010). Genomic instability an evolving hallmark of cancer. *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/nrm2858

Niwa, Y., Kanda, H., Shikauchi, Y., Saiura, A., Matsubara, K., Kitagawa, T., … Yoshikawa, H. (2005). Methylation silencing of SOCS-3 promotes cell growth and migration by enhancing JAK/STAT and FAK signalings in human hepatocellular carcinoma. *Oncogene*. https://doi.org/10.1038/sj.onc.1208788

Ortutay, C., Väliaho, J., Stenberg, K., & Vihinen, M. (2005). KinMutBase: A registry of disease-causing mutations in protein kinase domains. *Human Mutation*. https://doi.org/10.1002/humu.20166

Paul, I., Bhattacharya, S., Chatterjee, A., & Ghosh, M. K. (2013). Current Understanding on EGFR and Wnt/β-Catenin Signaling in Glioma and Their Possible Crosstalk. *Genes and Cancer*. https://doi.org/10.1177/1947601913503341

Paul, M. K., & Mukhopadhyay, A. K. (2004). Tyrosine kinase - Role and significance in Cancer. *International Journal of Medical Sciences*.

Pavlova, N. N., & Thompson, C. B. (2016). The Emerging Hallmarks of Cancer Metabolism. *Cell Metabolism*. https://doi.org/10.1016/j.cmet.2015.12.006

Reimand, J., Wagih, O., & Bader, G. D. (2013). The mutational landscape of phosphorylation signaling in cancer. *Scientific Reports*, *3*. https://doi.org/10.1038/srep02651

Smyth, L. A., & Collins, I. (2009). Measuring and interpreting the selectivity of protein kinase inhibitors. *Journal of Chemical Biology*. https://doi.org/10.1007/s12154-009-0023-9

Tan, X., Apte, U., Micsenyi, A., Kotsagrelos, E., Luo, J. H., Ranganathan, S., … Monga, S. P. S. (2005). Epidermal growth factor receptor: A novel target of the Wnt/β-catenin pathway in liver. *Gastroenterology*. https://doi.org/10.1053/j.gastro.2005.04.013

Verheyen, E. M., & Gottardi, C. J. (2010). Regulation of Wnt/β-catenin signaling by protein kinases. *Developmental Dynamics*. https://doi.org/10.1002/dvdy.22019

Whittaker, S., Marais, R., & Zhu, A. X. (2010). The role of signaling pathways in the development and treatment of hepatocellular carcinoma. *Oncogene*, *29*(36), 4989–5005. https://doi.org/10.1038/onc.2010.236

Wood, J. M., Bold, G., Buchdunger, E., Cozens, R., Ferrari, S., Frei, J., … Marmé, D. (2000). PTK787/ZK 222584, a novel and potent inhibitor of vascular endothelial growth factor receptor tyrosine kinases, impairs vascular endothelial growth factor-induced responses and tumor growth after oral administration. *Cancer Research*.

Wu, P. K., & Park, J. I. (2015). MEK1/2 Inhibitors: Molecular Activity and Resistance Mechanisms. *Seminars in Oncology*. https://doi.org/10.1053/j.seminoncol.2015.09.023

Wu, P., Nielsen, T. E., & Clausen, M. H. (2015). FDA-approved small-molecule kinase inhibitors. *Trends in Pharmacological Sciences*. https://doi.org/10.1016/j.tips.2015.04.005

Wu, P., Nielsen, T. E., & Clausen, M. H. (2016, January 1). Small-molecule kinase inhibitors: An analysis of FDA-approved drugs. *Drug Discovery Today*. Elsevier Ltd. https://doi.org/10.1016/j.drudis.2015.07.008

Yang, X., Huang, Y., Crowson, M., Li, J., Maitland, M. L., & Lussier, Y. A. (2010). Kinase inhibition-related adverse events predicted from in vitro kinome and clinical trial data. *Journal of Biomedical Informatics*, *43*(3), 376–384. https://doi.org/https://doi.org/10.1016/j.jbi.2010.04.006

Yau, T., Chen, P. J., Chan, P., Curtis, C. M., Murphy, P. S., Suttle, A. B., … Poon, R. T. (2011). Phase I dose-finding study of pazopanib in hepatocellular carcinoma: Evaluation of early efficacy, pharmacokinetics, and pharmacodynamics. *Clinical Cancer Research*. https://doi.org/10.1158/1078-0432.CCR-11-0793

Zaret, K. S., & Grompe, M. (2008). Generation and regeneration of cells of the liver and pancreas. *Science*. https://doi.org/10.1126/science.1161431

Zhang, J., Yang, P. L., & Gray, N. S. (2009). Targeting cancer with small molecule kinase inhibitors. *Nature Reviews Cancer*. https://doi.org/10.1038/nrc2559

Zhu, A. X., Duda, D. G., Sahani, D. V., & Jain, R. K. (2011). HCC and angiogenesis: Possible targets and future directions. *Nature Reviews Clinical Oncology*. https://doi.org/10.1038/nrclinonc.2011.30

Zhu, A. X., Sahani, D. V., Duda, D. G., Di Tomaso, E., Ancukiewicz, M., Catalano, O. A., … Jain, R. K. (2009). Efficacy, safety, and potential biomarkers of sunitinib monotherapy in advanced hepatocellular carcinoma: A phase II study. *Journal of Clinical Oncology*. https://doi.org/10.1200/JCO.2008.20.9908

APPENDICES

APPENDIX A

Table A.A.1: The inhibitor list of CanSyL data their reference numbers

Inhibitors:	**Pubchem Names:**	**Pubchem CID:**	**CAS #:**	**CATALOG #**
MeSAdo	Methylthioadenosine	439176	2457-80-9	260585
AG1296	Tyrphostin AG 1296	2049	146535-11-7	658551
AG1478	Tyrphostin AG 1478	2051	175178-82-2	658552
AG82	Tyrphostin A25	2061	118409-58-8	658400
Akti-1/2	Akt Inhibitor VIII	10196499	612847-09-3	124018
Aloisine-A	ALOISINE A	5326843	496864-16-5	128125
Alsterpaullone	Alsterpaullone	5005498	237430-03-4	126870
Aminopurvalanol-A	Aminopurvalanol A	6604931	220792-57-4	164640
ATM-Kinase Inhibitor	KU-55933	5278396	587871-26-9	118500
Aurora Kinase/Cdk Inhibitor	JNJ-7706621	5330790	443797-96-4	189406
Aurora Kinase Inhibitor II	Aurora Kinase Inhibitor II	6610278	331770-21-9	189404
Bcr-Abl Inhibitor	GNF-2	5311510	778270-11-4	197221
Ro-31-8220	Ro 31-8220 mesylate	11628205	138489-18-6	557520
Ro-31-8425	Ro-31-8425	2404	131848-97-0	557514
Bohemine	Bohemine	2422	189232-42-6	203600
BPDQ	Bpdq	2426	169205-87-2	203697
BPIQ-I	bpiq-i	2427	174709-30-9	203696
Casein Kinase II Inhibitor I	4,5,6,7-tetrabromobenzotriazole	1694	17374-26-4	218697
CK2 Inhibitor	DMAT	5326976	749234-11-5	218699

Cdk Inhibitor p35	2-Hydroxybohemine	4155347	471270-60-7	219457
Cdk1 Inhibitor	CGP74514A	2794188	190653-73-7	217696
CDK 1/2 Inhibitor II	Cdk1/2 Inhibitor III	5330812	443798-55-8	217713
Cdk2 Inhibitor II	Cdk2 Inhibitor II	5858639	222035-13-4	219445
Cdk2 Inhibitor III	CVT-313	6918386	199986-75-9	238803
Cdk2/5 Inhibitor	Cdk2/5 Inhibitor	16760362	21886-12-4	219448
Cdk2/9 Inhibitor	Cdk2/9 Inhibitor	447961	507487-89-0	238806
Chk2 Inhibitor II	BML-277	9969021	516480-79-8	220486
CLK Inhibitor TG003	TG003	1893668	300801-52-9	219479
Curcumin, Curcuma Ionga L.	Curcumin	969516	458-37-7	239802
Daphnetin	Daphnetin	5280569	486-35-1	268295
DNA-PK Inhibitor	6-Nitroveratraldehyde	88505	20357-25-9	260960
EGFR Inhibitor	EGFR Inhibitor	9549299	879127-07-8	324674
EGFR Inhibitor II, BIBX1382	Falnidamol	6918508	196612-93-8	324832
EGFR / ErbB-2 Inhibitor	EGFR/ErbB-2 Inhibitor	9843206	179248-61-4	324673
EGF / ErbB2 / ErbB-4 Inhibitor	HDS 029	11566580	881001-19-0	324840
Epigallocatechin Gallate	(-)-Epigallocatechin gallate	65064	989-51-5	324880
Erk Inhibitor	SCHEMBL15021964	16218944	1049738-54-6	328006
Erk Inhibitor II FR 180204	FR 180204	11493598	865362-74-9	328007
Fascaplysin, Synthetic	Fascaplysin	73292	114719-57-2	341251
FGF / VEGF RTK Inhibitor				341607
Genistein	Genistein	5280961	446-72-0	345834
Gö 7874, Hydrochloride	Gö 7874, Hydrochloride	11540703	-	365252
H89, Dihydrochloride	h-89	449241	127243-85-0	371963
HA1100, Hydroxyfasudil	Hydroxyfasudil Hydrochloride	11371328	155558-32-0	390602
Hypericin	Hypericin	5281051	548-04-9	400076
IC261	Ic261	5288600	186611-52-9	400090
Indirubin Derivative E804	Indirubin Derivative E804	6419764	854171-35-0	402081
IRAK-1/4 Inhibitor	IRAK-1-4 Inhibitor I	11983295	509093-47-4	407601

Isogranulatimide	Isogranulatimide	6419741	219829-00-2	371957
JAK Inhibitor I	Pyridone 6	5494425	457081-03-7	420099
JNK Inhibitor V	AS601245	10109823	345987-15-7	420129
Kenpaullone	Kenpaullone	3820	142273-20-9	422000
LY 294002	154447-36-6	3973	154447-36-6	440202
MEK Inhibitor II	MEK Inhibitor II	389898	623163-52-0	444938
MEK1/2 Inhibitor	SL327	9549284	305350-87-2	444939
ML-7, Hydrochloride	110448-33-4	9803932	110448-33-4	475880
MNK1 Inhibitor	CGP 57380	11644425	522629-08-9	454861
Olomoucine	Olomoucine	4592	101622-51-9	495620
PD 98059	PD 98059	4713	167869-21-8	513000
PKCßII / EGFR Inhibitor	DAPH 2	6711154	145915-60-2	539652
PP1 Analog	221243-82-9	4877	221243-82-9	529579
PP1 Analog II, 1NM-PP1	1-NM-PP1	5154691	221244-14-0	529581
H-7-Dihydrochloride	H-7 dihydrochloride	73332	108930-17-2	371955
H-8-Dihydrochloride	H-8 dihydrochloride	150584	113276-94-1	371958
KN-93	kn-93	5312122	139298-40-1	422711
Purvalanol-A	Purvalanol A	456214	212844-53-6	540500
Quercetin dihydrate	Quercetin Dihydrate	5284452	6151-25-3	551600
ROCK Inhibitor	y-27632	448042	146986-50-7	688000
R-Roscovitine	Roscovitine	160355	186692-46-6	557360
SB-203580, Iodo-				559400
SB-202190	SB 202190	5353940	152121-30-7	559388
SB-218078	SB 218078	3387354	135897-06-2	559402
SCY	Scytonemin	5486761	152075-98-4	565715
SKF-86002	skf-86002	5228	72873-74-6	567305
ST-638	ST638	5353962	107761-24-0	567790
Staurosporine	Staurosporine	44259	62996-74-1	569397

STO-609	STO-609	3467590	52029-86-4	570250
SU-5402	Su5402	5289418	215543-92-3	572630
SU-9516	SU9516	5289419	377090-84-1	572650
TGFß-R1-Inhibitor	396129-53-6	447966	396129-53-6	616451
TGFß-R1-Inhibitor II	RepSox	449054	446859-33-2	616452
TX-1123	TX-1123	403661	157397-06-3	655200
TX-1918	TX-1918	6419746	503473-32-3	655203
Tyrene CR4	LS-104	9861871	-	655230
Tyrophostin AGL-2043	AGL 2043	9817165	226717-28-8	121790
W-5, Hydrochloride	61714-25-8	173829	61714-25-8	681625
W-7, Hydrochloride	61714-27-0	124887	61714-27-0	681629
Wee1/Chk1 Inhibitor	Wee1/Chk1 Inhibitor	16760707	1177150-89-8	681637
Wortmannin	Wortmannin	312145	19545-26-7	681675
ZM-336372	ZM 336372	5730	208260-29-1	692000
2-thioadenosine	2-Thioadenosine	6451965	43157-50-2	589400
A3-hydrochloride	A3 Hydrochloride	9861903	78957-85-4	100122
Tyrphostin AG-17	Tyrphostin A9	5614	10537-47-0	658425
AG-18	Tyrphostin 23	2052	118409-57-7	658395
AG-30	Tyrphostin AG 30	5328793	122520-79-0	121760
AG-99	Tyrphostin 46	5328768	118409-59-9	658430
AG-112	Tyrphostin AG 112	5328804	-	658440
AG-126	AG 126	2046	118409-62-4	658452
AG-213	Tyrphostin 47	6809674	122520-86-9	658405
AG-490	Tyrphostin B42	5328779	133550-30-8	658401
AG-527	Tyrphostin B44	5328772	133550-32-0	658402
AG-825	AG 825	6091659	149092-50-2	121765
AG-879	Tyrphostin AG 879	5487525	148741-30-4	658460

AG-1295	6,7-Dimethyl-2-phenylquinoxaline	2048	71897-07-9	658550
Butein	Butein	5281222	487-52-5	203987
Emodin	Emodin	3220	518-82-1	324694
Piceatannol	Piceatannol	667639	10083-24-6	527948
Lavendustin C	Lavendustin C	3896	125697-93-0	234450
Tamoxifen, 4-Hydroxy-(Z)-isomer	4-Hydroxytamoxifen	449459	68047-06-3	579002
Et-18-OCH3	ET-18-OCH3	6918215	70641-51-9	341207
Tamoxifen Citrate	Tamoxifen Citrate	2733525	54965-24-1	579000
HA 1077	Fasudil	3547	103745-39-7	371970
Geldanamycin	Geldanamycin	5288382	30562-34-6	345805
Herbimycin A	Herbimycin A	5311102	70563-58-5	375670
K252-a	k-252a	3813	97161-97-2	420298
Met Kinase Inhibitor	SU11274	9549297	658084-23-2	448101
Calphostin C	Calphostin C	2533	121263-19-2	208725
JNK Inhibitor II	1,9-Pyrazoloanthrone	8515	129-56-6	420119
K-252b, Nocardiopsis sp.	K-252a, Nocardiopsis sp.	490561	99533-80-9	420319
SB 239063	SB 239063	5166	193551-21-2	559404

APPENDIX B

Pathway Enrichment Results for All Three Datasets Using Go Biological Process Gene Set

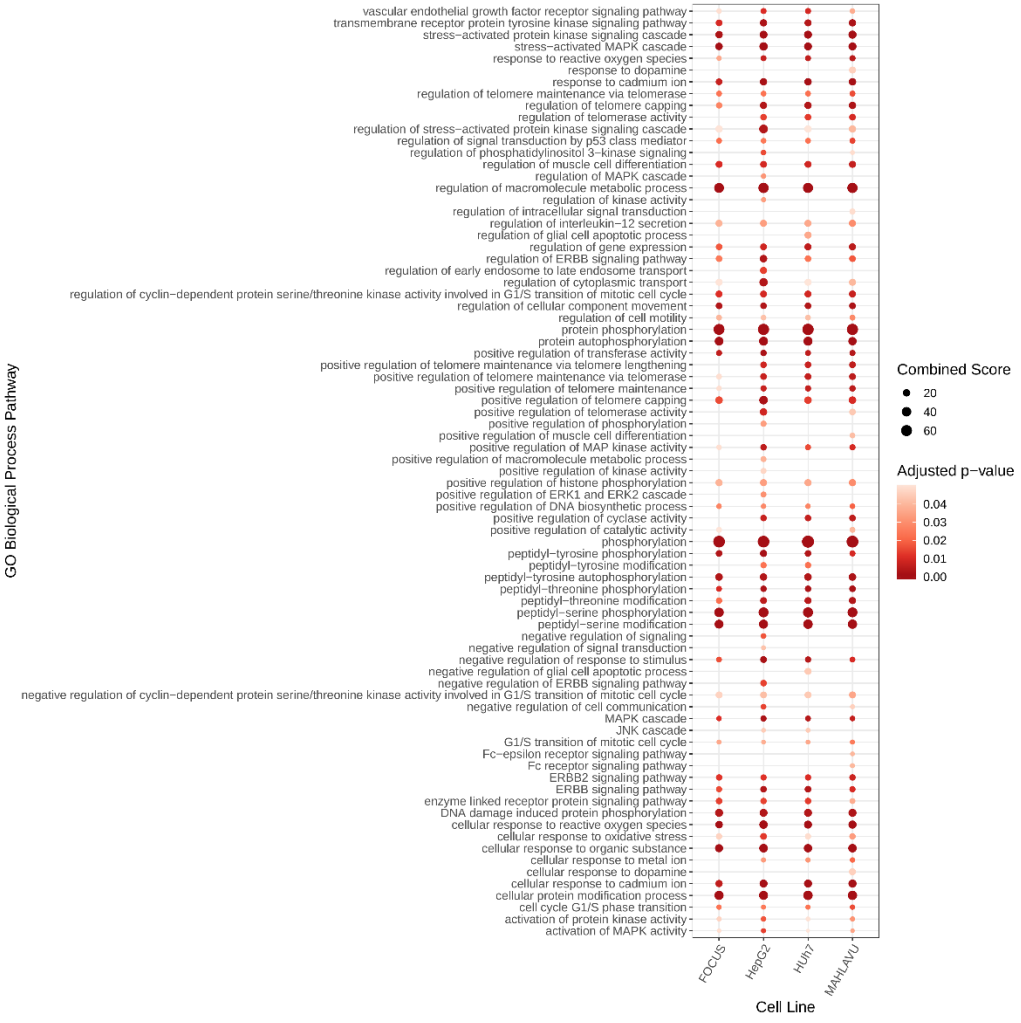


Figure A.B.1: Pathway enrichment results for the CanSyL dataset using the Go Biological Process gene set. Enrichment analysis was performed using non-regularized vulnerability scores.

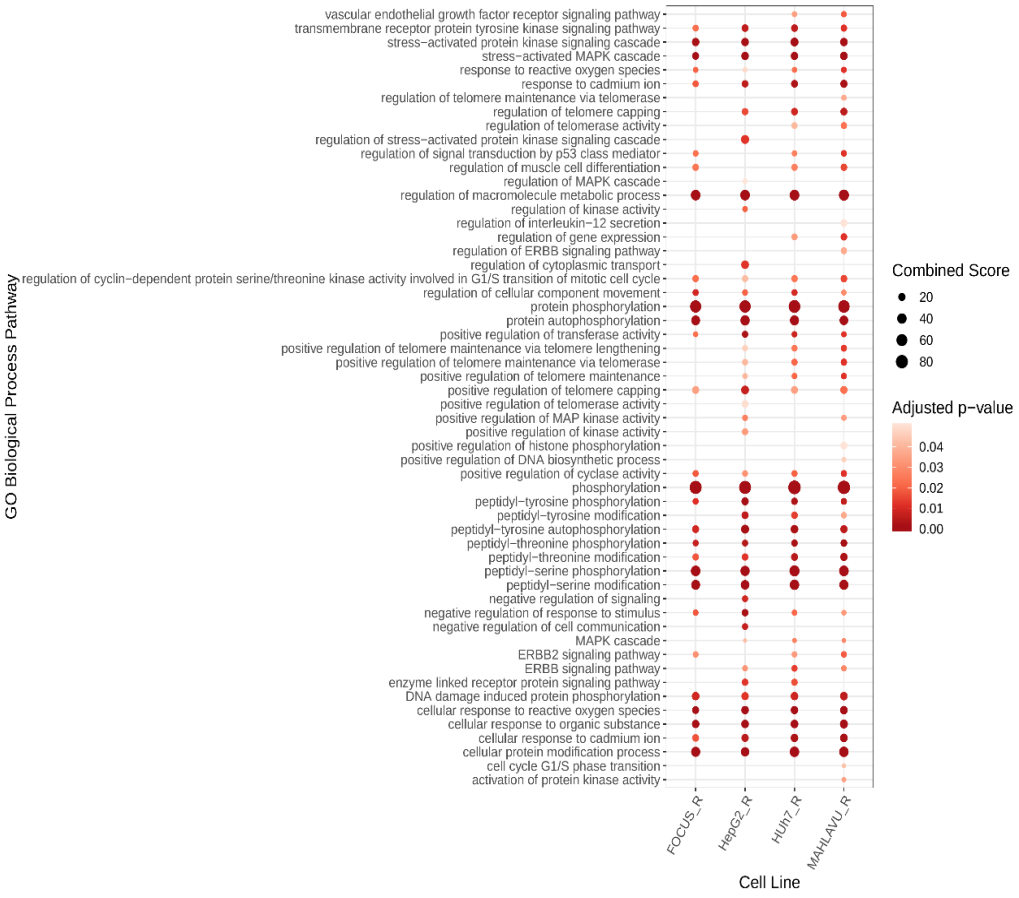
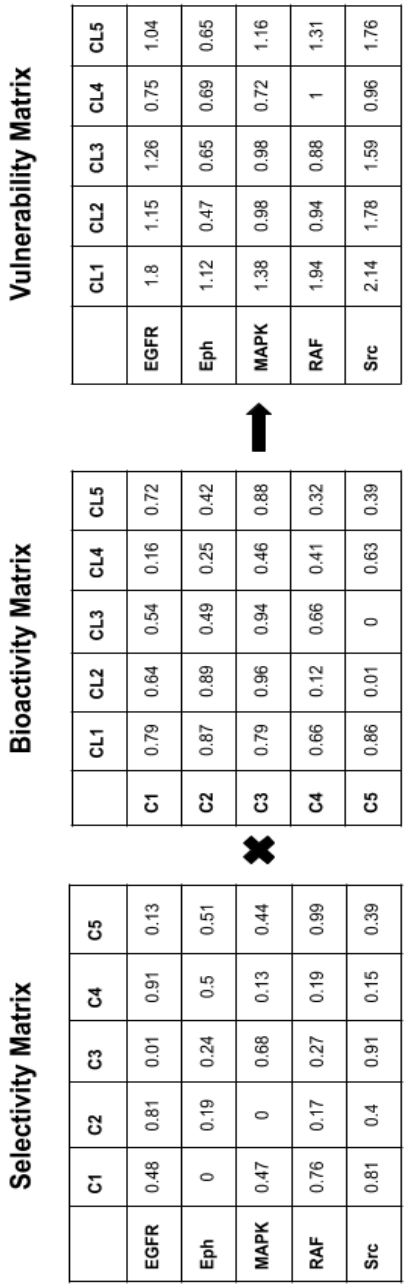


Figure A.B.2: Pathway enrichment results for the CanSyL dataset using the Go Biological Process gene set. Enrichment analysis was performed using kinome tree regularized vulnerability scores.

APPENDIX C

Representation of the Analysis with the Toy Data

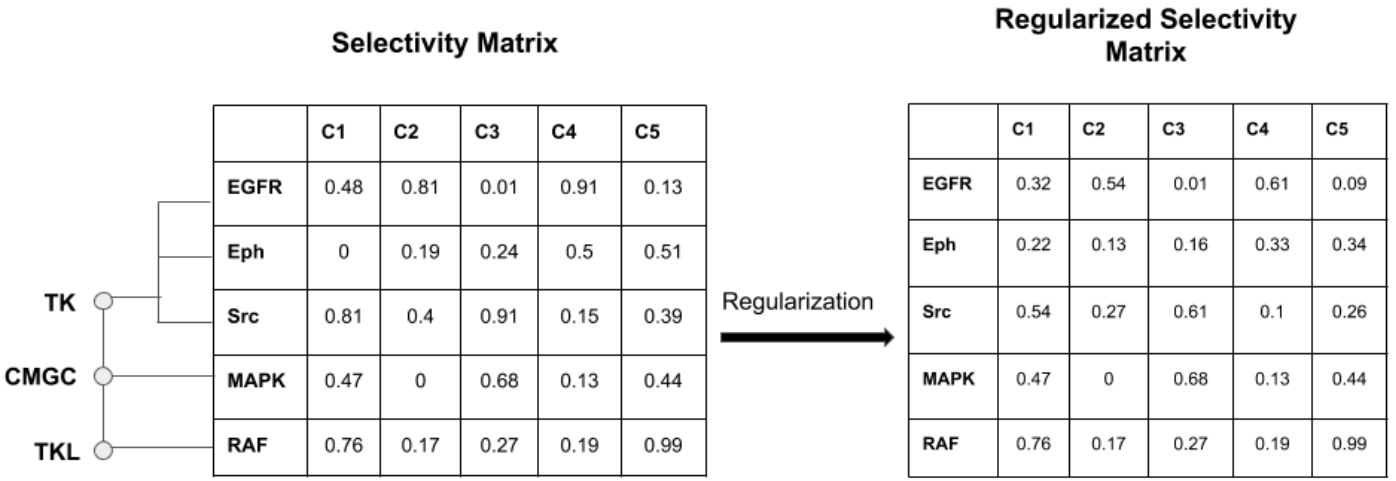


	C1	C2	C3	C4	C5
EGFR	0.48	0.81	0.01	0.91	0.13
Eph	0	0.19	0.24	0.5	0.51
MAPK	0.47	0	0.68	0.13	0.44
RAF	0.76	0.17	0.27	0.19	0.99
Src	0.81	0.4	0.91	0.15	0.39

	CL1	CL2	CL3	CL4	CL5
C1	0.79	0.64	0.54	0.16	0.72
C2	0.87	0.89	0.49	0.25	0.42
C3	0.79	0.96	0.94	0.46	0.88
C4	0.66	0.12	0.66	0.41	0.32
C5	0.86	0.01	0	0.63	0.39

	CL1	CL2	CL3	CL4	CL5
EGFR	1.8	1.15	1.26	0.75	1.04
Eph	1.12	0.47	0.65	0.69	0.65
MAPK	1.38	0.98	0.98	0.72	1.16
RAF	1.94	0.94	0.88	1	1.31
Src	2.14	1.78	1.59	0.96	1.76

Figure A.C.1: Kinase family vulnerability score calculation based on the selectivity matrix and bioactivity matrix multiplication. C, Compound; CL, Cell Line.



	C1	C2	C3	C4	C5
EGFR	0.48	0.81	0.01	0.91	0.13
Eph	0	0.19	0.24	0.5	0.51
Src	0.81	0.4	0.91	0.15	0.39
MAPK	0.47	0	0.68	0.13	0.44
RAF	0.76	0.17	0.27	0.19	0.99

	C1	C2	C3	C4	C5
EGFR	0.32	0.54	0.01	0.61	0.09
Eph	0.22	0.13	0.16	0.33	0.34
Src	0.54	0.27	0.61	0.1	0.26
MAPK	0.47	0	0.68	0.13	0.44
RAF	0.76	0.17	0.27	0.19	0.99

Figure A.C.2: The second step of the methodology is regularization of the selectivity matrix by distributing the selectivity score of the inhibitors for the kinase families between the kinase groups based on human kinome tree. C, Compound.

Regularized Selectivity Matrix

	C1	C2	C3	C4	C5
EGFR	0.32	0.54	0.01	0.61	0.09
Eph	0.22	0.13	0.16	0.33	0.34
MAPK	0.47	0	0.68	0.13	0.44
RAF	0.76	0.17	0.27	0.19	0.99
Src	0.54	0.27	0.61	0.1	0.26

✖

Bioactivity Matrix

	CL1	CL2	CL3	CL4	CL5
C1	0.79	0.64	0.54	0.16	0.72
C2	0.87	0.89	0.49	0.25	0.42
C3	0.79	0.96	0.94	0.46	0.88
C4	0.66	0.12	0.66	0.41	0.32
C5	0.86	0.01	0	0.63	0.39

➡

Regularized Vulnerability Matrix

	CL1	CL2	CL3	CL4	CL5
EGFR	1.2	0.76	0.84	0.5	0.69
Eph	0.92	0.45	0.55	0.49	0.59
MAPK	1.38	0.98	0.98	0.72	1.16
RAF	1.94	0.94	0.88	1	1.31
Src	1.43	1.18	1.06	0.64	1.17

Figure A.C.3: The third step of the methodology is the regularized vulnerability score prediction method based on the kinome tree regularized selectivity matrix and bioactivity matrix multiplication. C, Compound; CL, Cell Line.

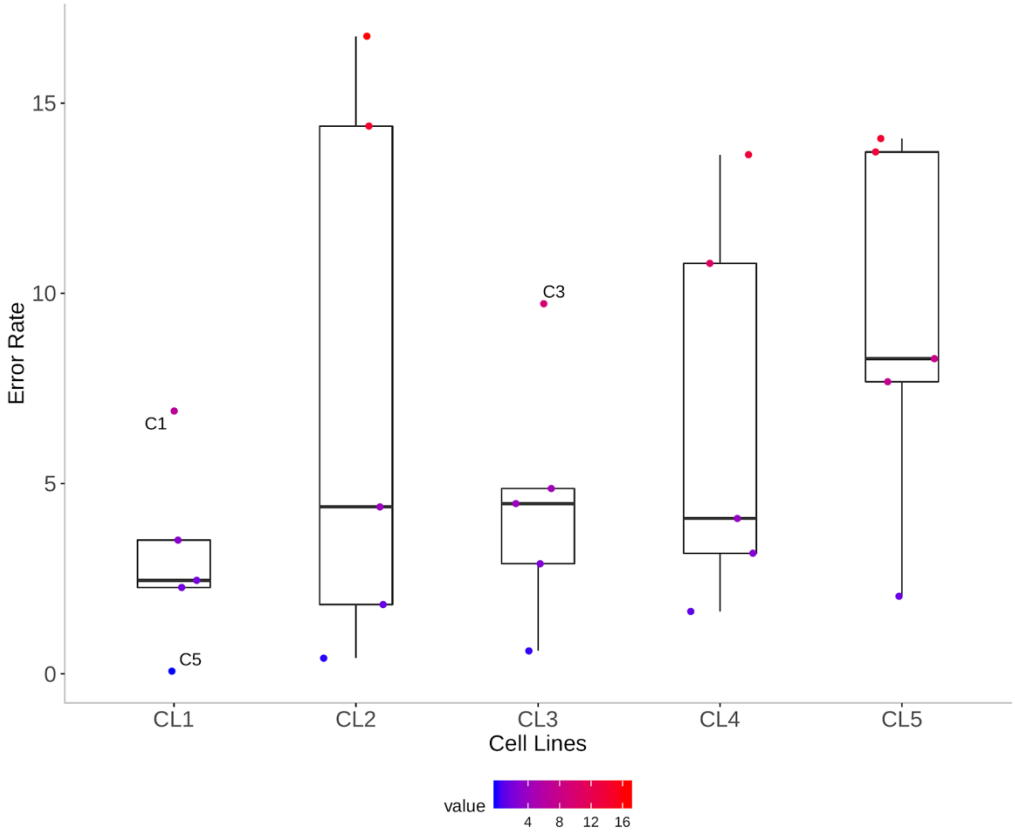


Figure A.C.4: Error rate of the model based on toy data. The error rate of the similarity between kinome tree-based regularized selectivity score and the non-regularized vulnerability score was calculated using RMSE. C, Compound; CL, Cell Line.