IDENTIFYING POTENTIAL THERAPEUTIC MOLECULES FOR HEPATOCELLULAR CARCINOMA THROUGH MACHINE LEARNING-BASED DRUG REPURPOSING

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

TUĞÇE BAŞER

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY
IN
THE DEPARTMENT OF MEDICAL INFORMATICS

SEPTEMBER 2024

Approval of the thesis:

IDENTIFYING POTENTIAL THERAPEUTIC MOLECULES FOR HEPATOCELLULAR CARCINOMA THROUGH MACHINE LEARNING-BASED DRUG REPURPOSING

Submitted by TUĞÇE BAŞER in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Health Informatics Department, Middle East Technical University** by,

Prof. Dr. Banu Günel Kılıç
Dean, **Graduate School of Informatics** _____

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

Asst. Prof. Dr. Burçak Otlu Sarıtaş
Supervisor, **Health Informatics** _____

Prof. Dr. Rengül Çetin-Atalay
Co-Supervisor, **Department of Medicine, University of Chicago** _____

Examining Committee Members:

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

Asst. Prof. Dr. Burçak Otlu Sarıtaş
Health Informatics Dept., METU _____

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

Prof. Dr. Erden Banoğlu
Pharmaceutical Chemistry, Gazi University _____

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

Date: *06.09.2024*

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: Tuğçe BAŞER

Signature : _____

ABSTRACT

IDENTIFYING POTENTIAL THERAPEUTIC MOLECULES FOR HEPATOCELLULAR CARCINOMA THROUGH MACHINE LEARNING-BASED DRUG REPURPOSING

BAŞER, Tuğçe

Ph.D., Department of Health Informatics

Supervisor: Asst. Prof. Dr. Burçak OTLU SARITAŞ

Co-Supervisor: Prof. Dr. Rengül ÇETİN ATALAY

September 2024, 179 pages

Hepatocellular carcinoma (HCC) is the most common primary liver cancer with a high mortality rate due to limited treatment options. Systemic drug treatments increase patient survival and often extend life by several months. The development of new small molecule chemotherapeutics is both time consuming and costly. Therefore, drug repurposing is being used as an effective strategy to identify and implement new treatment options for this mortal disease. The aim of this study is to identify potential drug candidates for the treatment of HCC through reuse of existing compounds using the machine learning tool MDeePred. The open target platform, UniProt, ChEMBL, and Expasy databases were used to create a dataset for MDeePred to predict drug-target interactions (DTIs). Enrichment analyses of DTIs were conducted, leading to the selection of 6 out of 380 DTIs identified by MDeePred for further analyses. The physicochemical properties, lipophilicity, water solubility, drug-likeness and medicinal chemistry properties of the drug candidates and approved drugs for advanced stage HCC (lenvatinib, regorafenib, and sorafenib) were meticulously and curated. The majority of drug candidates fell within conventional ranges in terms of drug properties and demonstrated target docking abilities. Our findings revealed the binding efficiency of selected drug compounds to identified targets associated with

HCC. As a result, small molecules were identified that can be further evaluated experimentally as potential drug candidates in HCC. This study also highlights the importance of the MDeePred deep learning tool in *in silico* drug repurposing studies in cancer treatment.

Keywords: Drug candidate, drug repurposing, hepatocellular carcinoma, machine learning, MDeePred

ÖZ

MAKİNE ÖĞRENİMİ TABANLI İLAÇ YENİDEN KULLANIMI YOLUYLA HEPATOSELÜLER KARSİNOM İÇİN POTANSİYEL TERAPÖTİK MOLEKÜLLERİN BELİRLENMESİ

BAŞER, Tuğçe

Doktora, Sağlık Bilişimi Bölümü

Tez Yöneticisi: Dr. Öğr. Üyesi Burçak OTLU SARITAŞ

Tez Eş Yöneticisi: Prof. Dr. Rengül ÇETİN ATALAY

Eylül 2024, 179 sayfa

Hepatoselüler karsinom (HCC), sınırlı tedavi seçenekleri nedeniyle yüksek mortalite oranına sahip, en sık görülen primer karaciğer kanseridir. Sistemik ilaç tedavileri hastanın hayatta kalma oranını arttırır ve sıklıkla yaşam süresini birkaç ay uzatır. Yeni küçük moleküllü kemoterapötiklerin geliştirilmesi hem zaman alıcı hem de maliyetlidir. Bu nedenle ilacın yeniden kullanılması, bu ölümcül hastalık için yeni tedavi seçeneklerinin belirlenmesi ve uygulanmasında etkili bir strateji olarak kullanılıyor. Bu çalışmanın amacı, MDeePred makine öğrenme aracını kullanarak mevcut bileşiklerin yeniden kullanılması yoluyla HCC tedavisi için potansiyel ilaç adaylarını belirlemektir. Açık hedef platformu, UniProt, ChEMBL ve Expasy veritabanları, MDeePred'in ilaç-hedef etkileşimlerini (DTI'ler) tahmin etmesi için bir veri kümesi oluşturmak üzere kullanıldı. DTI'ların zenginleştirme analizleri gerçekleştirildi ve MDeePred tarafından tanımlanan 380 DTI'dan 6'sının daha ileri analizler için seçilmesi sağlandı. İlaç adaylarının ve ileri evre HCC için onaylı ilaçların (lenvatinib, regorafenib ve sorafenib) fizikokimyasal özellikleri, lipofilikliği, suda çözünürlüğü, ilaca benzerliği ve tıbbi kimya özellikleri titizlikle belirlendi ve küratörlendi. İlaç adaylarının çoğunluğu, ilaç özellikleri açısından geleneksel aralıklara düştü ve hedef yerleştirme yetenekleri gösterdi. Bulgularımız, seçilen ilaç

bileşiklerinin HCC ile ilişkili belirlenen hedeflere bağlanma etkinliğini ortaya çıkardı. Sonuç olarak HCC'de potansiyel ilaç adayı olarak deneysel olarak daha fazla değerlendirilebilecek küçük moleküller belirlendi. Bu çalışma aynı zamanda kanser tedavisinde ilaç yeniden kullanım çalışmalarında MDeePred derin öğrenme aracının önemini de vurgulamaktadır.

Anahtar Sözcükler: İlaç aday molekülleri, ilaç yeniden konumlandırma, hepatoselüler karsinoma, makine öğrenimi, MDeePred

To My Mom and Dad

ACKNOWLEDGMENTS

First of all, I would like to express my deepest gratitude to my advisors Asst. Prof. Dr. Burçak Otlu Sarıtaş, Prof. Dr. Rengül Çetin-Atalay, and Prof. Dr. Mehmet Volkan Atalay, who supported and guided me throughout this thesis and made this study possible, for their patience and support.

I would like to express my special thanks to my understanding thesis committee members, Prof. Dr. Erden Banoğlu ve Assistant Prof. Dr. Aybar Can Acar, who supported me throughout my thesis process.

I would like to thank to jury members of the examination committee, Assoc. Prof. Dr. Yeşim Aydın Son and Assoc. Prof. Dr. Özlen Konu Karakayalı, for their help.

I would like to thank Assist. Prof. Dr. Ahmet Süreyya Rifaioğlu for being with me with his patience, endless support, knowledge and experience.

I would like to thank Orçun Yılmaz and Batıhan Konuk for their help.

I would like to express my endless gratitude to Dr. Seda Şirin for her endless love, support and help. At the same time, I would like to express my sincere gratitude for being more than just a friend to me, but for being like a sister to me and for her never-ending encouragement when I was feeling down. I could not have achieved this success without her tremendous support and friendship.

Throughout this journey, I would like to express my deepest gratitude to my beloved mother Şenay Başer and my beloved father Ufuk Başer for their unconditional love, support and endless belief in me in my academic life and in all the decisions I have made throughout my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADME	Absorption, Distribution, Metabolism, Excretion
AFB1	Aflatoxin B1
AFP	Alpha-fetoprotein
AKT	Protein Kinase B
AKT1	Serine/Threonine Protein Kinase AKT
ALK	ALK Tyrosine Kinase Receptor
ATP	Adenosine Triphosphate
BBB	Blood Brain Barrier
BCLC	Barcelona Clinic Liver Cancer
BOILED Egg	Brain or Intestinal Estimated Egg
cccDNA	Covalently Closed Circular DNA
CEUS	Contrast-enhanced Ultrasound
CI	Concordance Index
CNNs	Convolutional Neural Networks
CNS	Central Nervous System
CpG	Cytosine-phosphate-guanosine Dinucleotide
CT	Computed Tomography
DTIs	Drug Target Interactions
FGFR1	Fibroblast Growth Factor Receptor 1
FGFs	Fibroblast Growth Factors
FLEX	Flexibility
FLT3	Tyrosine-protein Kinase Receptor FLT3
GIA	Gastrointestinal Absorption
GO	Gene Ontology
GSEA	Gene Set Enrichment Analysis
HBV	Hepatitis B Virus
HCC	Hepatocellular Carcinoma
HCV	Hepatitis C Virus
IC_{50}	Half-maximal Inhibitory Concentration
INSATU	Saturation
INSOLU	Solubility
LIPO	Lipophilicity
Log P	Partition Coefficient
M	Molar
ML	Machine Learning
MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error
mTOR	Mammalian Target of the Rapamycin

MWA	Microwave Ablation
NAFLD	Non-alcoholic Fatty Liver Disease
NASH	Non-alcoholic Steatohepatitis
PAINS	Pan-assay Interference Compounds
PI3K	Phosphatidylinositol 3-kinase
PIK3CA	PI3-kinase p110-alpha Subunit
POLAR	Polarity
PTEN	Phosphatase and Tensin Homolog
rcDNA	Relaxed Circular DNA
RFA	Radiofrequency Ablation
ROS	Reactive Oxygen Species
SIZE	Size
SMILES	Simplified Molecular Input Line Entry System
TACE	Transarterial Chemoembolization
TARE	Transarterial Radiation
TKIs	Tyrosine Kinase Inhibitors
TPSA	Topological Polar Surface Area
US	Ultrasound

CHAPTER 1

INTRODUCTION

1.1. Cancer

Cancer, the largest cause of mortality worldwide, is defined by uncontrolled cell proliferation and the tendency to spread to other parts of the body. It is caused by both hereditary factors and exogenous carcinogens, which might be physical, chemical, or biological (Emaduldeen et al., 2022). Worldwide, cancer incidence and mortality rates vary. Because of lifestyle and demographic factors, developed regions have high rates. For example, according to the most recent GLOBOCAN forecasts, there will be a high global cancer burden in 2020, with 19.3 million novel cases and 10.0 million deaths, with breast cancer emerging as the most common cancer worldwide in terms of diagnosis (Sung et al., 2021). Lung cancer was the primary cause of cancer death, accounting for around 1.8 million fatalities (18.7% of total), followed by colorectal cancer (9.3% of total), liver cancer (7.8% of total), female breast cancer (6.9% of total), and stomach cancer (6.8% of total) (Bray et al., 2024). One of the most significant factors in cancer prevention and early detection strategies are lifestyle factors, including factors such as smoking, obesity and alcohol intake (Jazieh et al., 2023). Considering factors such as cancer development, epidemiology and risk factors; since cancer has a complex structure, the importance and necessity of increasing special public health strategies and conducting continuous research to reduce the occurrence and impact of cancer all over the world are emphasized.

1.2. Liver Cancer

Liver cancer, which originates in liver cells (Figure 1), is a serious global health issue due to its increasing prevalence and high mortality rates. Liver cancer is the eighth most common cancer type and ranks third in cancer deaths. Moreover, hepatocellular carcinoma (HCC) is the most common type of liver cancer (Chidambaranathan-Reghupaty et al., 2021; Wang et al., 2024a). Research indicates an increasing trend in the frequency of liver cancer in various regions, which can be related to several reasons such as the hepatitis B viruses and hepatitis C viruses and lifestyle-related risks including obesity and alcoholism (Huang et al., 2021). North Africa and East and Southeast Asia have the highest rates of liver cancer (Parra et al., 2023).

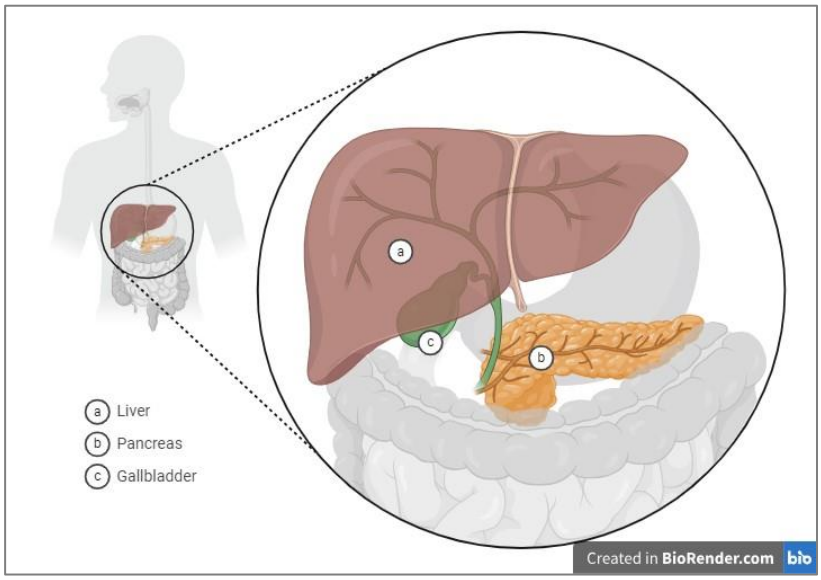


Figure 1: Location of the liver within the body

1.3. Hepatocellular Carcinoma (HCC)

HCC is the most prevalent form of primary liver cancer (accounting for 80–90% of cases) and is the most common cause of cancer-related deaths globally (Kim et al., 2024). In Western populations, recent studies have revealed a significant global variation in HCC incidence, because of primarily to varying risk factors such as aflatoxin exposure, alcohol consumption, chronic viral hepatitis (B and C), and the rising prevalence of non-alcoholic fatty liver disease (NAFLD) (Lampimukhi et al., 2023; Llovet et al., 2021a).

Due to the high frequency of hepatitis B and hepatitis C infections as well as dietary aflatoxins, the incidence of HCC is especially high in sub-Saharan Africa and East Asia. On the other hand, the number of cases in the US has been comparatively lower, despite the fact that it is rising as a result of increased rates of metabolic syndrome and obesity, both of which are major risk factors for HCC linked to NAFLD (Antwi et al., 2022; Kew, 2013; Michelotti et al., 2021).

In addition, it has been shown that there are gender variations in the prevalence of HCC, with males having higher rates than women. These differences may be caused by biological variables as well as differences in susceptibility to risk factors including tobacco and alcohol (Manieri et al., 2019; Zhang et al., 2021).

However, detecting HCC remains difficult since it has a complicated pathophysiology and a range of genetic, environmental, and viral causes (Parra et al., 2023).

1.4. Risk Factors

HCC is predominantly caused by a variety of well-defined risk factors. It has long been known that chronic viral infections, for example, hepatitis B (HBV) and hepatitis C (HCV), have a significant role in the development of HCC, especially in areas where these diseases are endemic (Yang et al., 2023).

Drinking alcohol is another important risk factor that can exacerbate the hepatocarcinogenic process, particularly when paired with viral hepatitis. Alcohol is a major factor in HCC risk because of its detrimental effects and interactions with other liver illnesses, such as hepatitis and aflatoxin exposure (Jacob et al., 2023).

Similarly, the risk of HCC is greatly increased by exposure to aflatoxins, strong liver carcinogens present in contaminated food sources, especially in developing nations (Dhandayuthapani et al., 2022).

NAFLD and more serious type of NAFLD, non-alcoholic steatohepatitis (NASH), have recently become an important risk factors for HCC, especially in Western nations where type 2 diabetes and obesity are common (Llovet et al., 2021a).

Genetic predispositions are also important in HCC susceptibility, with particular genetic mutations and family clustering identified as major risk factors. Advances in genetic and molecular profiling are helping to understand the complicated relationships between hereditary variables and environmental exposures in the formation of HCC (Yang et al., 2023).

In general, identifying the risk factors is critical for establishing focused prevention and treatment methods for HCC, with the goal of reducing the worldwide burden.

1.4.1. Chronic viral infections: Hepatitis B virus (HBV) infections

HBV is a severe worldwide health concern, characterized as a DNA virus that primarily targets liver cells (hepatocytes) and can cause chronic infection. Globally, it is forecasted that nearly 257 million individuals have HBV infection chronically. The life cycle of virus includes a unique replication process involving the transformation of its covalently closed circular DNA (cccDNA) from its relaxed circular DNA (rcDNA), which serves as a template for viral replication and protein production (Wei & Ploss, 2021). Surface antigens (S-HBsAg, M-HBsAg, and L-HBsAg), the small, medium, and large size surface antigens, are incorporated as transmembrane proteins in a lipid bilayer that forms the envelope of HBV. The envelope encases a nucleocapsid composed of 120 core proteins dimers (HBcAg), which contains the viral genome and polymerase (HBVPol). HBV is divided into 9 genotypes (A-I), one hypothetical genotype (J), and at least 35 subgroups (Campos-Valdez et al., 2021). HBV has the ability to interact with tumor suppression genes and oncogenes by integrating into host chromosomes. Consequently, severe fibrosis is not an essential requirement for HBV-induced HCC to arise (Kaur et al., 2022).

Hepatocellular carcinogenesis can be induced by HBV infection through indirect or direct mechanisms. In addition to this, through integration into host genes or inducing mutations, HBV can promote epigenetic remodeling of host DNA, which can lead to chromosomal remodeling and the aberrant expression of tumor suppressor and oncogene genes. It can also enhance the instability of the genome of the host cell (Li et al., 2021). Through the regulation of cell metabolism, activation of many cancer-related signaling pathways, and the regulation of other mechanisms, it can also lead to the malignant transformation of liver cells. On the other hand, HBV infection-induced chronic inflammation alters the liver microenvironment through interactions between the virus and innate immune cells and adaptive immune cells. These interactions help the virus evade immune surveillance and progress the disease from inflammation to tumor formation (Jiang et al., 2021).

The global incidence of HBV varies significantly, with high rates in parts of Asia and sub-Saharan Africa, where the virus is often transferred from mother to child during birth or in early childhood. In these areas, the incidence of chronic HBV infection can lead to a substantial burden of liver disease and cancer (Chuang et al., 2022).

1.4.2. Chronic viral infections: Hepatitis C virus (HCV) infections

HCV is surrounded by envelope, positive-sense, 9.6 kb single-stranded RNA virus with highly conserved 5' and 3' untranslated sequences. HCV mainly infects hepatocytes, as the expression of essential entry receptors and liver-specific cellular host factors (miRNA-122) required for viral replication occurs in hepatocytes (D'souza et al., 2020). HCV is a bloodborne disease that affects only the liver specifically. The majority of people infected with HCV are unable to remove the infection by itself, resulting in a lifelong chronic condition. Long-term liver inflammation caused by HCV leads to the development of advanced liver diseases such as cirrhosis, liver fibrosis, and HCC, resulting in death (Dash et al., 2020).

HCV proteins can directly activate mitogenic pathways, prevent cell death and increase ROS production. Furthermore, HCV initiates a persistent inflammatory process by causing the accumulation of lymphocytes in the liver and the production of various cytokines such as LTα and LTβ, which are closely associated with the development of HCC. Chronic inflammation increases ROS generation, which is thought to be a major cause of genetic alterations (Ivanov et al., 2017). ROS causes activation of the TGF-β pathway, leading to activation of hepatic stellate cells and fibrogenesis. TGF-β, together with TLR4, has a critical role in epithelial-mesenchymal transition. HCV disrupts host lipid metabolism, causing fat accumulation in the liver, and this condition is associated with HCC in many patients. Additionally, HCV can also trigger angiogenic and metastatic pathways. It has recently been determined that polymorphisms in the DEPDC5 and MICA genes increase the risk of developing HCC (Vescova et al., 2016).

The global incidence of HCV varies, with high prevalence rates in some areas of Eastern Europe, Sub-Saharan Africa and East Asia, primarily due to differences in

exposure to risk factors such as unsafe injection practices and blood transfusions. The burden of disease is further compounded by HCV's association with increased risks of development of liver cirrhosis and HCC, particularly in individuals with longstanding chronic infections (Roudot-Thoraval et al., 2021). Roughly, 71 million individiuals have currently HCV infections. Only 20-30% of these people develop liver cirrhosis. In addition, 1-4% of patients with cirrhosis develop HCC each year (Dash et al., 2020).

1.4.3. Aflatoxin B1

Exposure to toxins such as aflatoxin, liver viral infections, and chronic hepatitis are linked to the progression of HCC. Aflatoxin, a metabolite of mold in the environment, has been found to be widespread in nature (Mungamuri et al., 2020). Its potential for causing harm is obvious. There are around 18 known aflatoxins, but those of particular concern are AFB 1, AFB 2, AFG 1, AFG 2, AFM 1, and AFM 2 (Awuchi et al., 2020).

Aflatoxin B1 (AFB1), produced primarily by Aspergillus species, is a potent carcinogen known to contaminate various food products, posing serious health risks, particularly liver cancer in humans and animals. AFB1 undergoes metabolic activation in the liver, where it is converted by cytochrome P450 enzymes into a reactive epoxide form that can form adducts with DNA, leading to mutations and the initiation of cancer (Min et al., 2021; Song et al., 2022). Researches show that AFB1 metabolites bind to DNA by alkylation of bases, leading to changes in the cell cycle and the tumor suppressor gene p53 mutation (Cai et al. 2020). The arginine to serine (G to T) mutation occurring at codon 249 of the p53 tumor suppressor gene (R249S; 249ser mutation) is specific to aflatoxin exposure and has been detected in 64% of HCC patients (Kew, 2023; Villar et al., 2012). AFB1 also induces oxidative stress and inflammatory responses, which further contribute to its toxicity and carcinogenicity. The generation of reactive oxygen species (ROS) and consequent oxidative damage are key mechanisms by which AFB1 exerts its harmful effects, leading to a cascade of cellular events that compromise cellular integrity and function (Ma et al., 2021).

The threat of AFB1 is globally recognized, with higher exposure rates in regions with hot and humid climates that promote fungal growth on crops like grains and nuts (Alameri et al., 2023). The incidence of AFB1-related health issues, such as HCC, is particularly high in parts of Africa and Asia due to dietary exposure to contaminated foodstuffs. Despite efforts to mitigate AFB1 contamination, it remains a significant challenge due to its stability and prevalence in staple foods (Hamid et al. 2013).

1.4.4. Alcohol

Alcohol is categorized by the International Agency for Research on Cancer as a category 1 carcinogen (Matsushita and Takaki, 2019). Alcohol consumption is a significant risk factor for HCC, the most frequent type of liver cancer. Chronic alcohol consumption leads to liver damage, including cirrhosis, fibrosis, acute/chronic hepatitis, and fatty liver, which are significant precursors to HCC. Alcohol metabolites, particularly acetaldehyde, are directly hepatotoxic and contribute to

oxidative stress and inflammation, which promote mutagenesis and tumor development in the liver (Jacob et al., 2023). It has been claimed that alcohol misuse raises the relative risk of HCC by three to ten fold (Matsushita and Takaki, 2019).

The mechanism by which alcohol exacerbates liver disease involves the induction of chronic inflammation, impairment of the liver's natural defenses, and the promotion of a fibrogenic response in hepatic cells (Subramaniyan et al., 2021; Yu et al., 2022). Alcohol activates various mechanisms that lead to hepatocyte proliferation and cancer formation. These mechanisms include ethanol metabolism pathways that include acetaldehyde production and CYP2E1 induction. Both promote the formation of ROS, lipid peroxidation, and DNA strand breaks. Additionally, alcohol causes DNA hypomethylation and damage to the retinoic acid pathway, promoting oncogene activation and inducing cancer formation (Sidharthan and Kottillil, 2014; Taniai, 2020). Various host factors, such as gene polymorphisms (C282Y heterozygosity and CYP2E1 c2), lead to increased carcinogenesis. Iron deposition and neovascularization are associated with both alcoholic liver disease and HCC. In the end, the immune system plays a critical role; TLR4/NANOG stimulates the growth of alcohol-induced HCC, whereas NK cells provide protection (Sidharthan and Kottillil, 2014). These effects are magnified by alcohol's interaction with other risk factors such as viral hepatitis, obesity, and smoking (Huang et al., 2023; Yu et al., 2022).

Globally, the incidence of alcohol-related HCC varies, with higher rates in regions where heavy drinking is more prevalent. For instance, in Eastern Europe and parts of Asia, where alcohol consumption is significant, there is a correspondingly high rate of HCC (Cho et al., 2023; Lampimukhi et al., 2023). Genetic factors also contribute to the susceptibility to alcohol-related liver disease, with variations in genes related to alcohol metabolism and liver function affecting the risk of developing HCC (Buch et al., 2022).

1.4.5. Non-alcoholic fatty liver disease (NAFLD)

NAFLD is a common disease defined by excessive accumulation of fat in the liver without significant alcohol consumption. Metabolic diseases such as central obesity, excess blood sugar, high blood pressure, dyslipidemia, and liver dysfunction lead to NAFLD (Pouwels et al., 2022). It affects approximately 25% of the global population and is a significant contributor to liver-related morbidity and mortality (Powell et al., 2021). In recent decades, NAFLD has become known as one of the most common chronic liver disorders globally (Pouwels et al., 2022). NAFLD has become one of the leading causes of liver transplants, greatly increasing the risk of liver cancer, heart disease, and death.

The pathophysiology of NAFLD involves multiple interrelated processes, including insulin resistance, oxidative stress, lipotoxicity, and chronic inflammation. These factors contribute to the progression of NAFLD to more severe liver damage (Juanola et al., 2021). Environmental factors, particularly diet and lifestyle, play very significant roles in the incidence and progression of the disease.

NAFLD is defined as accumulation of 5% or more triglycerides in liver tissue (i.e. steatosis) in the absence of extreme alcohol intake and other liver problems (for example, chronic viral hepatitis or steatogenic drug use). This disease, along with steatosis and liver inflammation, can progress to a more serious condition called NASH, with or without fibrosis (Chrysavgis et al., 2022). NASH is a more advanced stage of the disease and, over time, causes inflammation and scar tissue to form in the liver. This condition can eventually result in cirrhosis and hepatocellular carcinoma (Samy et al., 2024) (Figure 2).

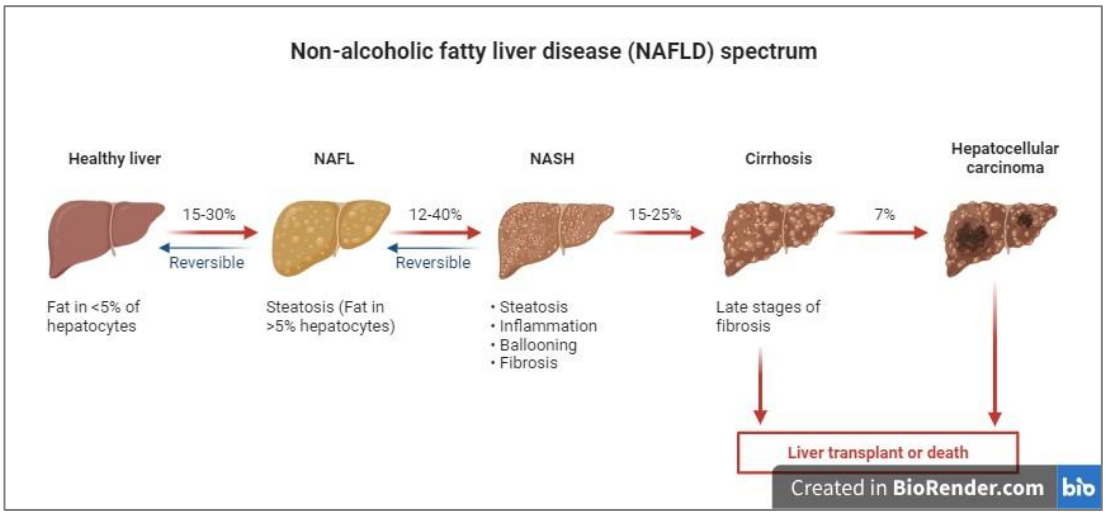


Figure 2: Non-alcoholic fatty liver disease spectrum

NAFLD was linked to 19.2% of HCC cases among Medicare patients in the United States (32.07% in inpatients and 20.22% outpatients), but HCV infection was linked to just 9.75%. Compared to HBV-associated HCC patients, NAFLD-associated HCC patients had a higher mortality rate (odds ratio 1.87; $p < 0.001$) (Younossi and Henry, 2021).

Globally, the incidence of NAFLD is increasing, paralleling the rise in type 2 diabetes and obesity. This trend is particularly pronounced in Western countries but is also a growing concern in developing regions where Western diets are becoming more common (Gangopadhyay et al., 2022; Goshnell et al., 2024). Despite its prevalence, NAFLD often remains undiagnosed until serious liver damage has occurred, underscoring the need for better diagnostic and therapeutic strategies.

1.5. Diagnosis

HCC screening is vital for early detection and improving treatment outcomes, particularly in high-risk populations such as those with HBV infection or chronic liver diseases. Standard screening protocols often include semi-annual ultrasound

examinations and the measurement of alpha-fetoprotein (AFP), although AFP's sensitivity and specificity can vary significantly (Debes et al., 2021). Ultrasound (US) technique, which uses US waves to reveal anomalies in the liver, is effective in detecting liver lesions that are suspected to be HCC. AFP, a glycoprotein generated by the liver and detectable in the blood, is considered a tumor marker as high levels can be related with HCC. These two tests (US and AFP) are performed alone or together to rule out the existence of HCC in person who are at high risk of developing HCC (URL_1).

Recent advances have expanded screening tools to include dynamic contrast-enhanced imaging techniques such as computed tomography (CT) scans and magnetic resonance imaging (MRI), which can detect smaller and early-stage tumors more effectively compared to US alone. However, the accessibility and cost of these advanced imaging methods can limit their routine use in some settings (Giustini et al., 2023).

Contrast-enhanced US (CEUS) is another diagnostic modality that offers real-time imaging advantages, particularly in differentiating HCC from other hepatic lesions. It is effective in patients with atypical radiological features or those who cannot undergo CT/MRI due to contraindications (Kale, 2021).

Emerging biomarkers and liquid biopsy techniques are under investigation to improve the early detection of HCC. These include circulating tumor DNA and other molecular markers that could potentially supplement traditional imaging and biochemical tests, providing a more comprehensive and early diagnosis (Manea et al., 2023).

Despite advancements, significant challenges remain in the accurate early diagnosis of HCC due to the disease's complexity and the subtlety of early-stage manifestations. This necessitates ongoing research into more sensitive diagnostic tools and the refinement of existing modalities to improve the prognosis and management of HCC (Nadarević et al., 2021; Wang and Wei, 2020).

1.6. Staging

HCC staging is critical for guiding treatment decisions and predicting prognosis (Torimura & Iwamoto, 2021). The Barcelona Clinic Liver Cancer (BCLC) system is used for clinical staging of liver cancer. It was first proposed by Llovet JM in 1999 and updated by the American Association for the Study of Liver Disease in 2005 (Figure 3 and Figure 4).

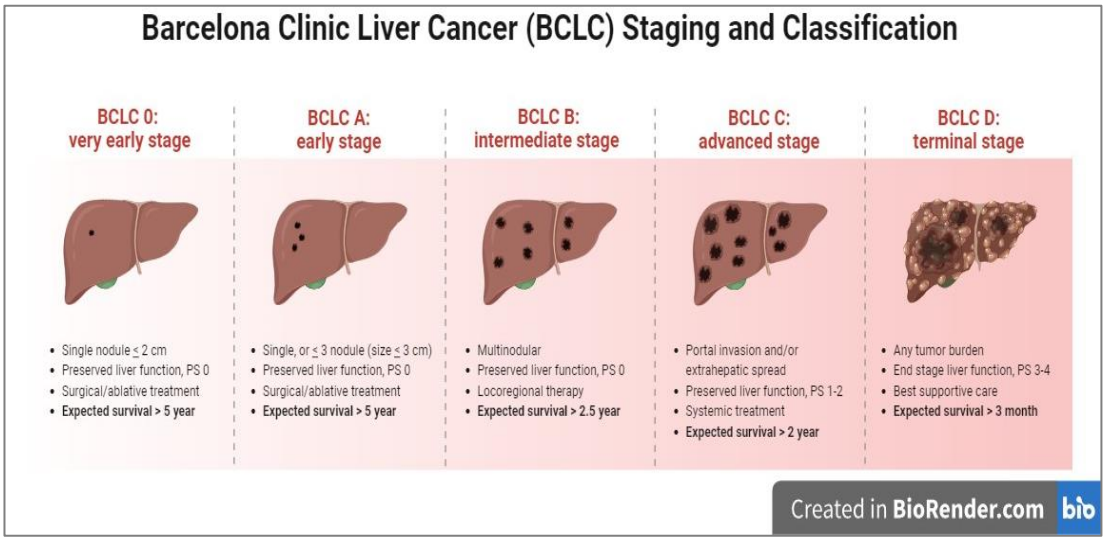


Figure 3: BCLC staging and classification

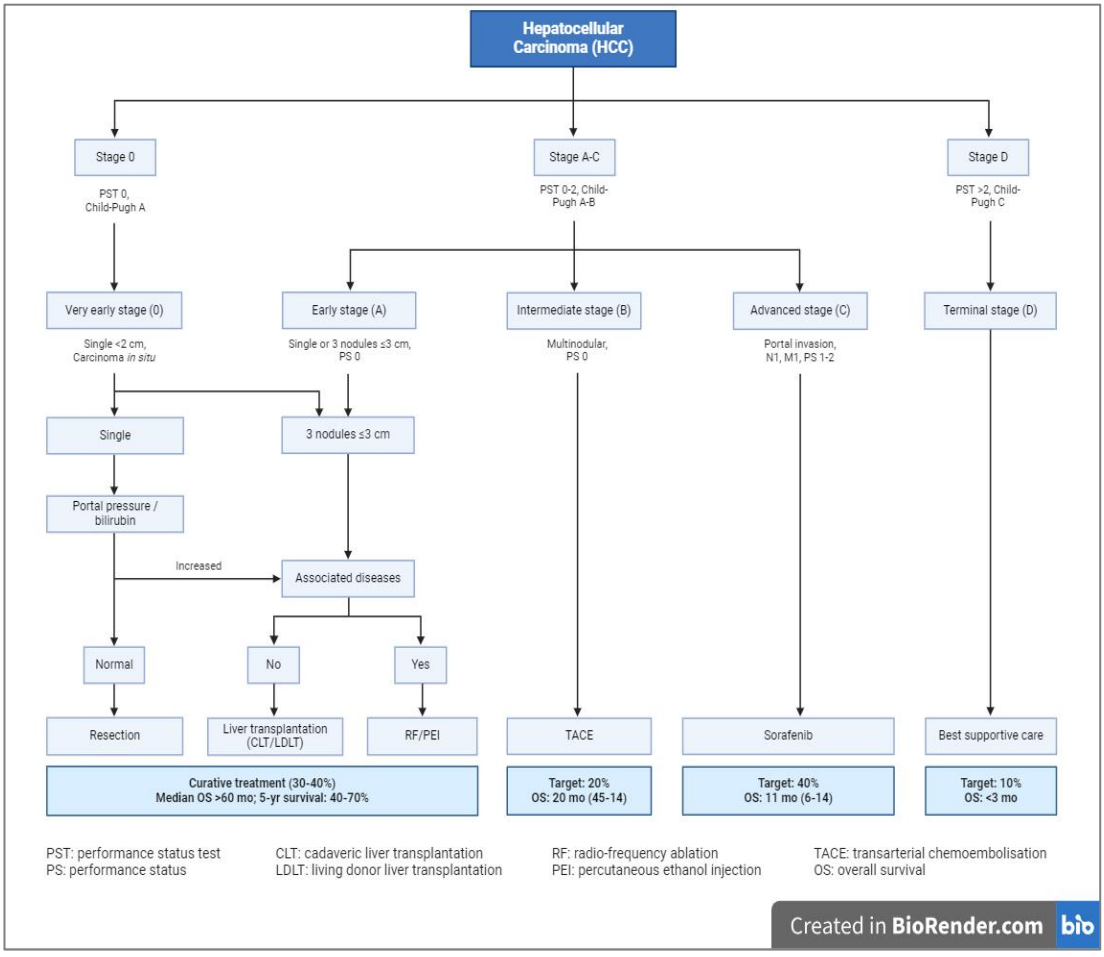


Figure 4: BCLC staging and classification

The implementation of this system will assist in assessing the patient's condition, providing accurate treatment alternatives, and predicting the patient's prognosis (Sun et al., 2024). The BCLC staging system is a system that classifies liver cancer patients and divides patients into five different stages: stage 0 (very early), stage A (early), stage B (intermediate), stage C (advanced), and stage D (terminal). This system also provides treatment recommendations. These recommendations are based on factors such as liver function, number of tumor, tumor size, vascular invasion, portal hypertension, extrahepatic metastasis, and physical condition (Sun et al., 2024; Wang et al., 2024).

The need for accurate staging is underscored by the heterogeneous nature of HCC, where patients with the same stage can have different outcomes based on sub-stage factors and response to treatment. This has led to continuous updates and refinements in staging systems to better match therapeutic options with patient-specific tumor characteristics and liver function status, aiming to optimize outcomes and enhance quality of life for patients with HCC (Golfieri et al., 2019).

1.7. Treatment

Various surgical and non-surgical approaches are used in the treatment of HCC. Surgical approaches include liver resection and liver transplantation. Liver resection provides long-term survival for anatomically limited tumors and remains the standard first-line therapy, while transplantation is preferred, especially for patients with chronic liver disease (Glantzounis et al., 2021; O'Leary et al., 2020).

Non-surgical approaches for HCC treatment include transarterial chemoembolization (TACE), transarterial radiation (TARE), percutaneous local ablation, and microwave ablation (MWA). TACE is considered the standard treatment options for intermediate-stage HCC, providing localized chemotherapy to tumors (Kotsifa et al., 2022). TARE is used to deliver high-dose radiation to tumors and can achieve complete response in some patients (Taylor et al., 2019). Percutaneous local ablation methods, including MWA and radiofrequency ablation (RFA), are particularly used for early-stage tumors (Chevallier et al., 2023). Systemic therapy options, such as tyrosine kinase inhibitors (TKIs) and immunotherapies, are widely used, with drugs like lenvatinib and sorafenib being prominent in this category (Marquardt et al., 2019; Meng et al., 2020).

In conclusion, the treatment of HCC involves a variety of surgical and non-surgical methods selected based on the patient's overall condition, tumor stage, and liver function, with various combinations expanding the treatment options for patients (Kawano et al., 2022).

1.7.1. Surgical approaches

Surgical approaches in the treatment of HCC primarily include resection and liver transplantation. Resection is the preferred method for patients with early-stage HCC

and well-preserved liver function. This approach is indicated for tumors confined to one lobe of the liver without vascular invasion, offering a potential cure for those without significant cirrhosis (Sugawara & Hibi, 2021) (Figure 5).

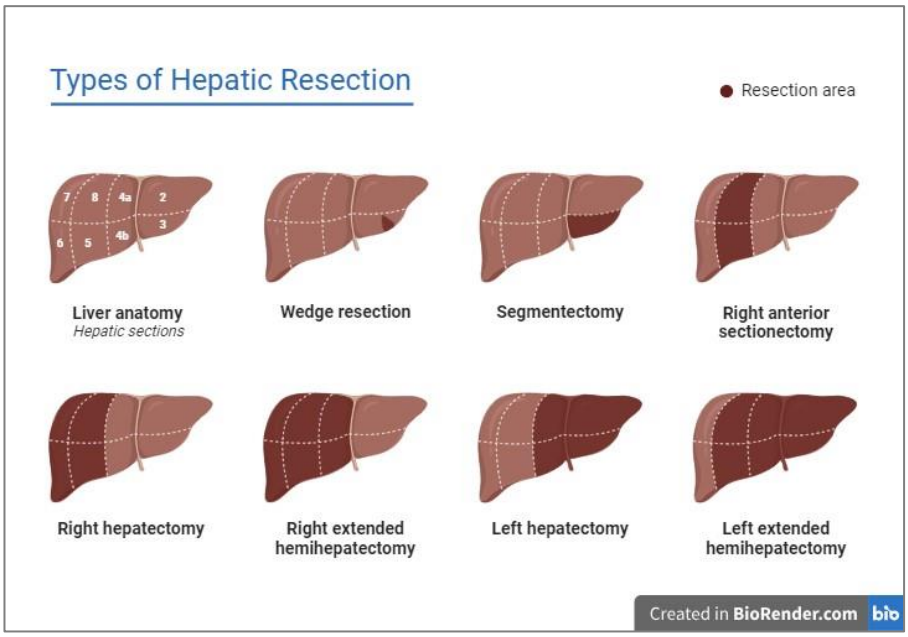


Figure 5: Types of hepatic resection

Liver transplantation, on the other hand, addresses both the tumor and the underlying liver disease, making it the optimal treatment for patients with decompensated cirrhosis or tumors meeting the Milan criteria (a solitary tumor $\leqslant$ 5 cm or up to three nodules each $\leqslant$ 3 cm). Transplantation not only removes the tumor but also eliminates cirrhosis, significantly improving long-term survival rates (Kow, 2019).

Recent advancements in surgical techniques and patient selection criteria have expanded the applicability of these surgical treatments, leading to improved outcomes. Laparoscopic liver resection has become increasingly popular due to its minimally invasive nature, which reduces blood loss and shortens hospital stays while providing similar long-term oncologic outcomes as open resection (Wilson & Geller, 2019). Despite the advances, the decision between resection and transplantation must consider tumor characteristics, liver function, and patient performance status to ensure the best possible outcomes (Makary et al., 2020).

In conclusion, surgical resection and liver transplantation are the cornerstone treatment options for HCC, with each approach tailored to the patient's specific clinical condition, offering the best chance for long-term survival (Kow, 2019; Lurje et al., 2019).

1.7.2. Non surgical approaches

Non-surgical approaches in the treatment of HCC include various methods tailored to the patient's overall condition and the stage of the tumor. TACE is considered the standard treatment option for intermediate-stage HCC, providing localized chemotherapy to tumors by blocking their blood supply, thus directly destroying tumor cells and enhancing the effect of chemotherapy (Llovet et al., 2021b). TARE delivers high doses of radiation to the tumor and can achieve complete responses in some patients, making it an effective alternative for those who do not respond to TACE (O'Leary et al., 2020).

Percutaneous local ablation methods, especially effective for early-stage HCC, contain RFA and MWA. RFA is effective in treating HCCs up to 3 cm in size, providing results equivalent to surgical resection. For larger tumors, a combination of TACE and RFA is recommended (Chevallier et al., 2023). RFA destroys tumor cells by heating them, while MWA can eliminate larger tumors more quickly and, when used with RFA, provides a broader ablation area (Luerken et al., 2022). MWA destroys tumor cells using high-frequency microwaves and typically operates faster than RFA (Glassberg et al., 2019). Systemic therapy options play a crucial role in managing advanced HCC. TKIs and immunotherapies are prominent in this category. Drugs like lenvatinib and sorafenib slow the progression of HCC and extend patient survival by inhibiting tumor growth and spread (Chami et al., 2023).

In conclusion, non-surgical approaches in HCC treatment are selected based on the tumor stage and the patient's overall condition. Methods such as TACE, TARE, RFA, and MWA, used in various combinations, expand treatment options and improve patient outcomes. These approaches are tailored to individual patient needs, allowing for the determination of the most suitable treatment strategy (Chami et al., 2023; Chevallier et al., 2023; Llovet et al., 2021; Luerken et al., 2022; Xu et al., 2019;).

1.7.3. Disadvantages of current treatments

The treatment of HCC has several surgical and non-surgical approaches, each with its own disadvantages. Surgical approaches like resection and liver transplantation are often limited by the patient's overall health and liver function. Resection is only suitable for a small proportion of patients with anatomically confined tumors and well-preserved liver function; however, it often results in high recurrence rates (Ince et al., 2019). Liver transplantation, while potentially curative, is hindered by the limited availability of donor organs and significant waiting times, which can lead to disease progression (Wege et al., 2019).

Non-surgical approaches also come with challenges. TACE is effective for intermediate-stage HCC but can cause significant side effects such as postembolization syndrome, which includes pain, fever, and nausea (Kotsifa et al., 2022). TARE can achieve high tumor control but poses risks such as radiation-induced liver disease (Andreana et al., 2012).

Percutaneous local ablation methods, like MWA and RFA, are effective for small tumors but are limited by the size and position of the tumors, often leading to incomplete ablation and tumor recurrence (Luerken et al., 2022). MWA, while faster than RFA, can cause complications such as bile duct injury and bleeding (Esmail et al., 2020).

Systemic therapy, including TKIs and immunotherapies, often has limited efficacy due to the underlying liver disease and can cause significant side effects such as hypertension, hand-foot syndrome, and immune-related adverse events (Ghavimi et al., 2020; Sadagopan et al., 2024).

In conclusion, while surgical and non-surgical treatments for HCC provide critical options, they each have limitations and potential adverse effects (D'Alessio & Fulgenzi, 2021).

1.8. Drug Discovery

Drug discovery is a complex, lengthy, and costly process that involves identifying, synthesizing, characterizing, and screening compounds for therapeutic efficacy. Typically, it takes about 10-15 years and costs between $1 billion and $2.6 billion to develop a new medicine, from the early discovery phase to regulatory approval and market entry (Annett, 2021; Rennane et al., 2021; Sun et al., 2022). The drug discovery process includes several stages: target identification, lead compound discovery, optimization, preclinical testing, and clinical trials (Figure 6) (Deore et al., 2019).

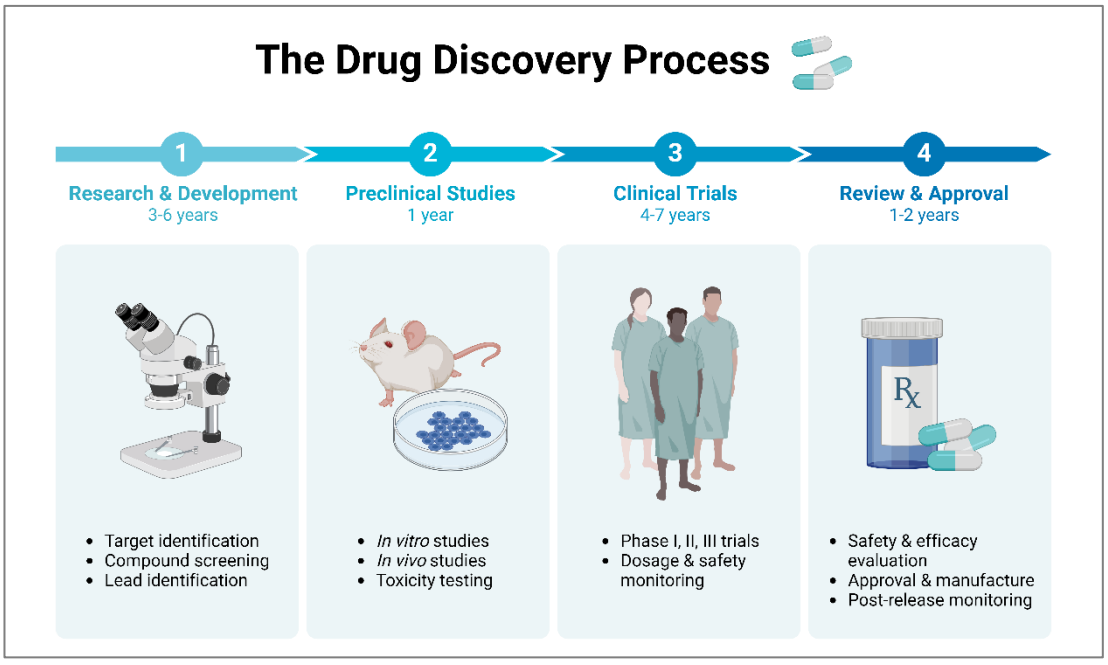


Figure 6. The drug discovery process (created in biorender.com)

Despite technological developments, such as high-throughput screening and artificial intelligence, which aim to reduce time and cost, drug discovery remains an expensive and time-consuming endeavor. The high failure rate of candidate compounds during clinical trials contributes significantly to the overall cost (Chan et al., 2019). One of the major disadvantages is that only a small fraction of discovered compounds eventually becomes a marketed drug (Figure 7), leading to substantial financial risk for pharmaceutical companies (Berdigaliyev & Aljofan, 2020; DiMasi et al., 2003).

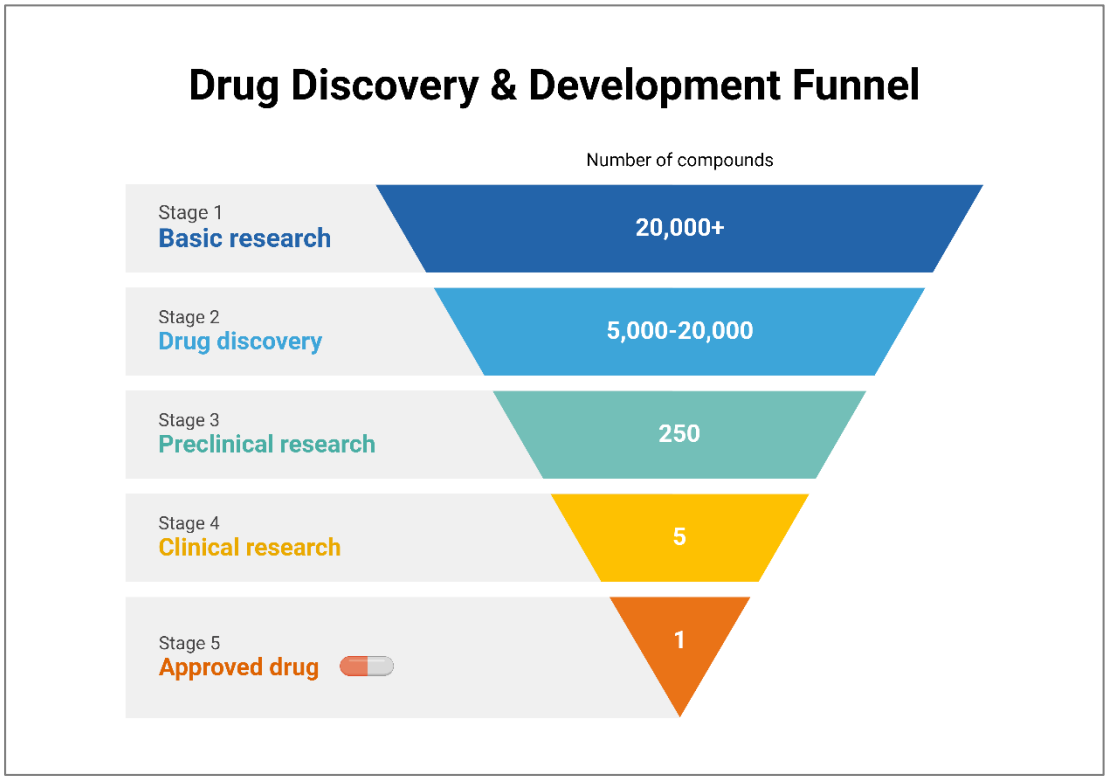


Figure 7. Drug discovery and development funnel (created in biorender.com)

Moreover, the increasing cost of research and development, along with regulatory requirements, further complicates the drug discovery process. Efforts to streamline and expedite this process include drug repurposing and the use of computational methods to predict therapeutic potential and side effects (Shaker et al., 2021; Vora et al., 2023). Despite these efforts, the inherent complexity and high attrition rate of drug development continue to pose significant challenges (Kumar et al., 2021; Waring et al., 2015).

In summary, drug discovery is a vital but challenging process with high costs, long timelines, and numerous obstacles, requiring innovative approaches to improve efficiency and success rates.

1.9. Drug Repurposing

Drug repurposing is a strategy that seeks to find new therapeutic uses for existing, shelved, or clinically tested drugs, offering a promising alternative to traditional drug discovery methods. This approach is advantageous as it significantly reduces the time and cost involved in drug development since the pharmacokinetics, pharmacodynamics, and safety profiles of these drugs are already well-established, allowing pre-clinical studies to be bypassed (Babu, 2020). The process of drug repurposing involves systematic screenings, often supported by bioinformatics and computational technologies, to identify potential new indications for existing drugs (Tanoli, 2021a).

One of the primary benefits of drug repurposing is that it allows for quicker clinical deployment, reducing both developmental risk and the need for extensive early-phase clinical trials (Ng et al., 2021). This approach has proven particularly useful in response to urgent medical needs where existing drugs were rapidly evaluated for new applications (Bhagat & Butle, 2021; Low et al., 2020). The potential for significant cost savings and reduced development timelines makes drug repurposing an attractive strategy for both common and rare diseases (Verma et al., 2022).

1.10. Machine Learning (ML)

ML has become a pivotal tool in drug repurposing, offering significant improvements in efficiency and accuracy. This approach leverages large datasets and advanced algorithms to identify novel therapeutic uses for existing drugs. The process involves systematic screenings supported by bioinformatics and computational technologies, such as deep learning, which help in predicting bioactivities and molecular properties with high accuracy (Ekins et al., 2019). ML methods can analyze gene expression profiles, chemical structures, and biological activities to uncover potential drug-disease associations, making the repurposing process more targeted and efficient (Figure 8) (Alharbi and Vakanski, 2023; Tanoli et al., 2021b; Walters and Vakanski, 2023; Wu and Wangi 2023).

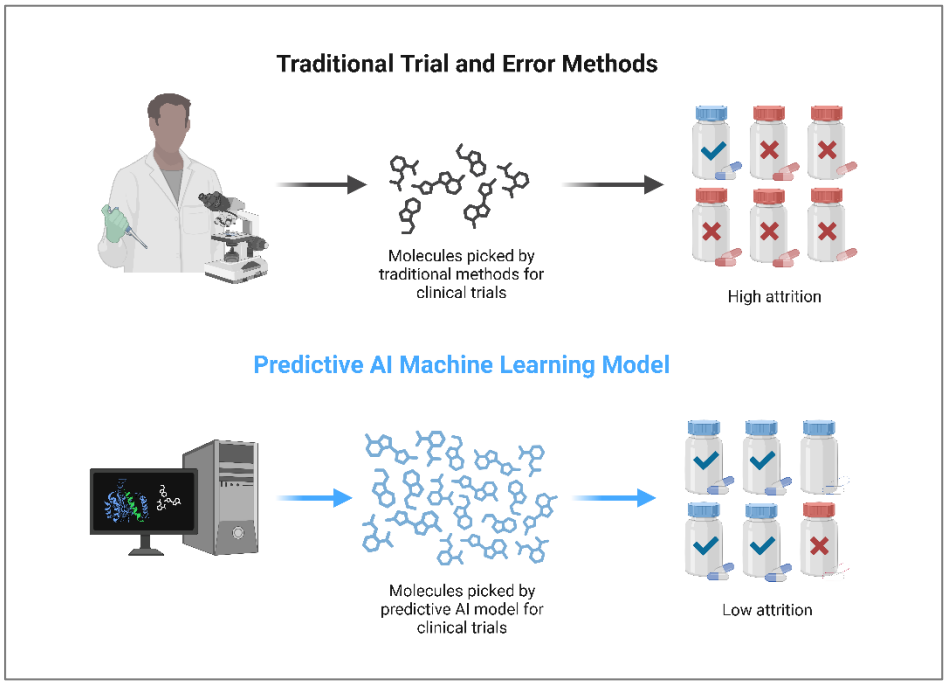


Figure 8. ML in drug discovery (created in biorender.com)

One of the primary advantages of using ML in drug repurposing is the reduction in time and cost associated with drug development. By utilizing existing safety and efficacy data, ML models can quickly identify promising candidates, thus accelerating the transition from discovery to clinical trials (Urbina et al., 2021). Moreover, ML approaches can handle vast amounts of heterogeneous data, integrating various types of biological, chemical, and clinical information to enhance prediction accuracy and uncover new therapeutic potentials (Issa et al., 2020). Ongoing advancements in ML techniques and the development of more sophisticated models continue to enhance the potential of drug repurposing, making it a vital strategy in modern drug discovery (Sherine Glory et al., 2022).

In summary, ML significantly enhances drug repurposing efforts by efficiently analyzing large datasets to identify new drug indications, thereby reducing development costs and timelines while overcoming traditional drug discovery challenges (Tanoli et al., 2021b).

1.11. MDeePred

MDeePred is an innovative deep learning-based approach for predicting binding affinity in drug discovery. This method leverages multi-channel protein featurization to identify interactions between bioactive small molecules and target proteins. MDeePred integrates various protein features, such as sequence, evolutionary,

structural, and physicochemical properties, into multiple 2D vectors, which are then fed into deep neural networks to predict compound-target protein interactions.

One of the primary advantages of MDeePred is its ability to handle large datasets and integrate various biological, chemical, and clinical information to uncover new therapeutic potentials. This significantly reduces the time and cost associated with drug development, enabling faster clinical applications. The use of MDeePred accelerates drug development processes and reduces costs due to its high prediction accuracy and extensive data integration capabilities.

In conclusion, MDeePred is a powerful and innovative tool for drug discovery and repurposing, providing more accurate and effective predictions through deep learning and multi-channel protein featurization. This method plays a crucial role in modern pharmaceutical research, accelerating drug development processes and reducing costs (Rifaioglu, 2020; Rifaioglu et al., 2021).

1.12. *In Silico* Studies

In silico studies aim to study biological processes using computer-aided simulations and modeling. These methods play an important role in the drug discovery process and save time and cost compared to traditional laboratory studies. Drug discovery is the process of finding and developing new therapeutic drugs and generally includes stages such as target identification, compound screening, optimization of candidate compounds, and preclinical and clinical testing. *In silico* approaches can be used at various stages of this process and are widely used in areas such as molecule screening, drug-target interaction prediction, and side effect analysis (Shaker et al., 2021).

Molecular screening and virtual screening methods help identify potential drug candidates by rapidly screening thousands of compounds in large databases. This process uses techniques such as molecular docking and molecular dynamics simulations to predict the potential of compounds to interact with biological targets. These methods accelerate the evaluation of potential candidates in the early stages of the drug discovery process and enable more targeted screening (Rognan, 2017).

Drug-target interactions are predicted through *in silico* modeling to evaluate critical parameters such as biological activity, binding affinity and specificity of candidate molecules. This evaluation increases the success rate of the drug discovery process and helps prevent possible failures in the preclinical stages (Terstappen & Reggiani, 2001).

In addition, *in silico* methods can reduce the risk of failure in clinical trials by predicting possible side effects and toxicity profiles. Pharmacokinetic and pharmacodynamic modeling allows the selection of safer and more effective compounds by simulating the absorption, distribution, metabolism, and excretion (ADME) processes of drug candidates (Butina et al., 2002).

As a result, *in silico* studies provide time and cost savings in the drug discovery process, allowing for rapid, efficient and targeted evaluation of potential drug candidates. These studies are used as a critical complementary tool alongside traditional methods, accelerating processes and increasing success rates.

1.12.1. Enrichment analyses

Enrichment analysis is an important statistical and bioinformatics method used to extract meaningful information from biological data. This analysis is used to determine whether certain categories are overrepresented in lists of biological entities such as gene sets, proteins, or metabolites. The main goal of enrichment analysis is to determine whether gene sets or biological pathways are overrepresented in a given condition over another, and this analysis is often used to discover disease-associated pathways, common gene interactions, or regulatory factors (Subramanian et al., 2005).

Gene set enrichment analysis (GSEA) is an important tool for interpreting gene expression data in terms of biological functions. GSEA allows genes to be grouped together by common biological functions, chromosomal locations, or regulatory mechanisms by considering gene sets. This method is very effective in revealing common biological pathways in various cancer-related data sets. For example, GSEA can reveal common biological pathways between independent studies in cases where single gene analyses show little similarity (Murohashi et al., 2010).

Because enrichment analysis often requires complex algorithms, it is common for these analyses to be performed using computer software. These software deliver results by running algorithms that are computationally demanding and often include statistical requirements such as multiple testing corrections (Tilford & Siemers, 2009). In addition, enrichment analyses used in the analysis of regulatory elements allow biological information to be extrapolated beyond gene sets and can also assess regulatory elements such as cytosine-phosphate-guanosine dinucleotide (CpG) sites, miRNAs, and transcription factors (García-Moreno et al., 2022).

In conclusion, enrichment analysis is a highly effective method for drawing meaningful conclusions from biological data. This analysis helps researchers better understand biological processes and develop hypotheses about diseases, treatment responses, or other biological conditions.

1.12.2. SwissADME analysis

SwissADME is a bioinformatics tool used in drug discovery and development processes, and plays an important role in evaluating the pharmacokinetic properties, drug-like properties and chemical suitability of drug candidates. This tool stands out as a critical tool in determining new therapeutic areas of existing drugs, especially in drug repurposing studies. SwissADME evaluates the bioavailability and efficacy of potential therapeutic molecules by predicting ADME properties from molecular structures (Daina et al., 2017).

One of the key features of SwissADME is the assessment of drug-like properties. SWISSADME evaluates whether a molecule is a drug candidate using drug-like rules such as "Lipinski's Rule of Five". This assessment plays a critical role in determining whether drugs are effective on biological targets (Azzam, 2023). In addition, SwissADME is an important tool for predicting pharmacokinetic properties. This tool predicts pharmacokinetic properties such as gastrointestinal absorption (GIA), blood brain barrier (BBB) permeability and P-gp substrate. Thus, it provides comprehensive information about the behavior of drugs in the human body (Daina et al., 2017). In addition, SwissADME also estimates the solubility and biological membrane permeability of molecules. These estimates are factors that directly affect the efficacy of drugs, and by estimating these parameters, SwissADME helps optimize bioavailability (Fan et al., 2022).

The role of SwissADME in drug repurposing involves the discovery of new uses for existing drugs for different diseases. In this process, the evaluation of ADME and drug-like properties provided by SwissADME provides a great advantage in identifying drugs that can be repurposed. Drug repurposing has the potential to provide therapeutic solutions with lower costs and shorter development times (Pushpakom et al., 2018). This tool allows drugs to be developed as more effective and safer therapeutic options.

1.12.3. Molecular docking

Molecular docking is a computer-aided method based on three-dimensional structural analysis of interactions between small molecules (ligands) and large biomolecules (e.g. proteins). This method predicts how ligands bind to a given target molecule and determines the strength and nature of these interactions by calculating their binding energies. Molecular docking is often used in drug discovery and development and is an effective tool for modeling interactions between biological targets and potential drug candidates (Pinzi & Rastelli, 2019).

Molecular docking plays an important role in drug discovery processes. This method helps to understand the structural basis of drug interactions by predicting the binding affinity of a drug to its biological target. Docking is used for many purposes, such as determining structural activity relationships, measuring binding affinity, and predicting potential toxicity. At the same time, docking supports the design of new drug candidates by allowing the identification of molecules suitable for the active sites of target proteins (Agarwal & Mehrotra, 2016).

Molecular docking is also of great importance in drug repurposing. This method is used to evaluate the potential activities of existing drugs on different biological targets and thus helps to identify new therapeutic indications. Docking accelerates the process of discovering the reusability of drugs for various diseases and reduces development costs. This process contributes to the rapid development of new treatment options, especially for rare diseases and emergency situations (Shaikh et al., 2022).

Molecular docking also offers potential new uses by examining the effectiveness of existing chemotherapeutic drugs on new targets for various diseases, such as cancer. This method accelerates the clinical validation process by testing the effectiveness of clinically approved drugs against new biological targets (Kumar & Kumar, 2019).

1.13. The Aim of the Thesis

This study aims to identify new candidate drug molecules for HCC therapeutics by repurposing compounds among existing compounds found in molecule databases. To achieve this goal, we used MDeePred, a recently developed deep learning-based method. Drug repurposing has the potential to rapidly expand available treatment options by shortening the development process and reducing costs. MDeePred shows promise in accelerating the drug discovery process using already-existing data and computational techniques. MDeePred's computational capabilities enable the identification of potential therapeutic targets and drug candidates with high accuracy rates. Our proposed approach holds promise in streamlining the drug discovery process by leveraging existing data and computational methods to identify potential therapeutic options not limited HCC but also for other cancers.

CHAPTER 2

MATERIALS & METHODS

This section consists of four parts in total. These are “collecting data from databases”, “preparation of data for MDeePred method and selection of the drug-target intereactions”, “*in silico* validation of predicted small molecules targets hepatocellular carcinoma (HCC) transferase by SwissADME and molecular docking”, and “literature based validation of novel drug-target interaction predictions towards drug repurposing”. A schematic summary of the methods and techniques used in each part is provided in Figure 9.

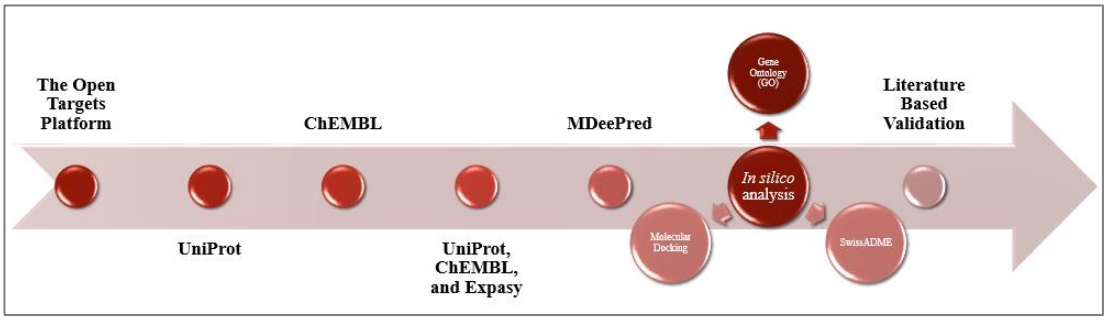


Figure 9. A schematic summary of the methods and techniques

2.1. Collecting Data from Databases

In this study, HCC-related genes were systematically identified using the Open Targets Platform database, a comprehensive resource that integrates various types of evidence to establish the associations between genes and diseases (Koscielny et al., 2017). The Open Targets Platform is a great tool for identifying disease-associated genes. By combining genetic data, clinical results, and biological databases, it aids in the identification of putative genetic targets contributing to disease etiology (Ochoa et al., 2021). The focus was placed on genetic associations and somatic mutations as the primary data types to ensure the relevance and specificity of the identified genes to HCC. These data types were chosen because genetic associations provide insights into inherited predispositions to HCC, while somatic mutations highlight the acquired genetic alterations commonly observed in HCC tumors (Müller et al., 2020; Yang et al., 2023).

To refine the selection of HCC-related genes, we employed a quantitative approach. Specifically, we calculated the arithmetic mean of the relevance scores for genetic associations and somatic mutations. This method allowed us to integrate and balance the contributions of both inherited and somatic genetic factors, thus identifying the most pertinent genes involved in HCC pathogenesis. The genes with the highest combined scores were selected as the actual HCC-related genes for further analysis in this research (Ochoa et al., 2021).

The comprehensive list of these HCC-related genes, along with their respective relevance scores, is provided in Table S1.

The protein products of the selected HCC-related genes were validated using the UniProt database. UniProt is a comprehensive protein database that includes structural and functional information about proteins, providing detailed biological and biochemical functions (UniProt Consortium, 2019). In this study, UniProt was utilized to verify and characterize the proteins encoded by HCC-related genes.

To compile and validate the biochemical activity data of the proteins, referred to as targets in this thesis, encoded by each HCC-related gene, the ChEMBL database was employed. ChEMBL is a comprehensive chemical biology database that includes biological activities, particularly interactions with drugs. In other words, ChEMBL offers extensive details on small compounds with recognized biological activity and their potential for use in drug research. A manually curated compound-protein activity dataset was created using the ChEMBL database (Smirnov et al., 2018; Zdrazil et al., 2024). Several filters were applied to this dataset:

Target Organism: The dataset was filtered to include only data related to *Homo sapiens* (humans).

Target Type: Only data pertaining to single protein targets were selected.

Assay Type: The focus was on binding assays.

Standard Unit: Data were standardized to the molar (M) unit.

Standard Type: Half-maximal inhibitory concentration (IC_{50}) measurements were used.

Standard Relation: Data were filtered based on standard relations such as = and < (Table 1).

Table 1: Filtration Criteria for Genes in ChEMBL Database

Criteria	**Filtration**
Target organism	*Homo sapiens*
Target type	Single protein
Assay type	Binding
Standard unit	Molar
Standard type	IC_{50}
Standard relation	=, <
pChEMBL value	Exist

It was noted that the dataset contained repeated measurements from different experiments. To address this, the median bioactivity of each pair was calculated and used this as a single bioactivity measurement. Additionally, bioactivity measurements lacking the pChEMBL value, which represents the half-maximal response on a negative logarithmic scale, were excluded. A data point with a pChEMBL value shows that the record has been curated and is therefore considered reliable (Rifaioglu et al., 2020).

After data filtering, the dataset was grouped using UniProt, ChEMBL, and Expasy databases. The filtered gene set is provided in Table S3. This grouping facilitated a better understanding of the functional and biochemical properties of the HCC-related genes.

Finally, for the deep learning-based binding affinity prediction using the MDeePred tool, we selected the enzyme class of "HCC associated transferases" as our final dataset. This class served as an appropriate model to identify potential drug targets related to HCC.

This methodological approach allowed for accurate identification of the biochemical activities of HCC-related genes and the identification of potential therapeutic targets.

Transferases play crucial roles in modulating protein function and activity, processes that are indispensable in the context of carcinogenesis. These enzymes facilitate various post-translational modifications and are fundamental in regulating cellular pathways that drive cancer development and progression.

For instance, phosphate transferases regulate signal transduction pathways by phosphorylating proteins. This phosphorylation can activate or deactivate proteins, thereby controlling various cellular processes such as growth of cell, differentiation of cell, and apoptosis. Dysregulation of these pathways can lead to uncontrolled cell proliferation, a hallmark of cancer (Krause and Van Etten, 2005).

Similarly, methyl transferases and hydroxyl transferases play significant roles in epigenetic modifications and post-translational modifications of proteins. Methyl transferases are involved in the methylation of DNA and histones, which can alter the patterns of gene expression without altering the underlying DNA sequence. These epigenetic changes are crucial in cancer, where aberrant methylation patterns can cause the silence of tumor suppressor genes or the activation of oncogenes (Robertson, 2001). Hydroxyl transferases, on the other hand, contribute to the hydroxylation of proteins, influencing their stability, localization, and interaction with other cellular molecules. These modifications can have profound effects on cellular functions such as DNA repair, metabolism, and response to hypoxia, all of which are relevant to cancer biology (Ploumakis and Coleman, 2015).

By targeting transferases involved in these fundamental molecular mechanisms, we aim to gain insights into their potential as therapeutic targets and elucidate their roles in driving oncogenic processes.

2.2. Preparation of Data for MDeePred Method and Selection of the Drug-Target Intereactions

The MDeePred method was employed as a deep learning tool to identify the final drug candidates for HCC. According to Rifaioğlu et al. (2021), MDeePred utilizes bioactivity drug target data specifically focused on "transferases," a class of enzymes integral to various biochemical processes, including those related to cancer. This method involves training and testing datasets derived from the bioactivity profiles of these enzymes, which play crucial roles in the pathogenesis of HCC.

MDeePred leverages the Simplified Molecular Input Line Entry System (SMILES) to generate representations of chemical compounds. SMILES is a standardized notation that encodes a molecule's structure using short ASCII strings (Drefahl, 2011; Rifaioğlu et al., 2021). In MDeePred, each compound is depicted as a 200x200 pixel two-dimensional image derived from its SMILES representation. This transformation into 2D images allows the deep learning model to analyze the spatial and structural characteristics of the molecules more effectively.

The SMILES notation provides comprehensive information necessary for creating these 2D images, capturing the molecule's atomic connectivity and stereochemistry. This detailed representation ensures that the generated images accurately reflect the molecular structure, which is essential for the model to predict bioactivity with high accuracy (Moret et al., 2021; Quiros et al., 2018).

Open-access bioactivity databases supply the standardized SMILES data, which are crucial for constructing the training and testing datasets. These databases compile extensive bioactivity profiles for a wide range of compounds, allowing MDeePred to learn from a diverse set of molecular interactions. The training process involves exposing the deep learning model to a variety of bioactive compounds, enabling it to

discern patterns and relationships between molecular structures and their biological activities.

Through this sophisticated approach, MDeePred can predict the bioactivity of new compounds, identifying potential drug candidates that may inhibit or modulate the activity of transferases involved in HCC. By converting chemical structures into image-based representations, MDeePred harnesses the power of convolutional neural networks (CNNs) to process and analyze molecular features with high precision.

This innovative method provides a robust platform for drug discovery, facilitating the identification of effective therapeutic agents for HCC. By integrating advanced deep learning techniques with detailed molecular data, MDeePred offers a powerful tool for accelerating the development of targeted cancer therapies.

Subsequently, to predict new drug-target interactions (DTI), we utilized the MDeePred tool, which was specifically trained on the "HCC-related transferases" dataset, to screen over one million small molecule drug entries from the ChEMBL database (v24). ChEMBL is a comprehensive database providing biological activity data crucial for drug discovery and drug development processes (Gaulton et al., 2012).

MDeePred employs deep learning algorithms to predict the potential effects of small molecule compounds on HCC-related transferases based on this dataset. During the screening process, each compound is transformed into a 200x200 pixel two-dimensional image derived from its SMILES notation, and these images are used as input for the deep learning model. This allows MDeePred to analyze the structural characteristics of the compounds and make predictions about their bioactivity.

Following these comprehensive DTI predictions, a statistical measurement was conducted to determine the bioactivities of the small molecules on their target proteins. In this phase, each compound's potential interaction with the target proteins is evaluated based on specific biochemical and biophysical criteria. Statistical analyses are performed to enhance the accuracy of the model's predictions and minimize potential false positives.

To distinguish the common features of the target proteins and better understand their biological functions, an ontology-based enrichment analysis was performed. In this analysis, Gene Ontology (GO) terms were used to examine the molecular functions and biological processes of the proteins. Annotations that were overrepresented based on GO Molecular Function and Biological Process ontology terms were prioritized according to their statistical relevance to the target proteins (Jin et al., 2015).

Enrichment analysis is used to determine whether proteins involved in specific biological functions or processes are represented at a higher rate. This method helps identify critical biological pathways and functions in which HCC-related transferases play a role. These tests are essential for categorizing proteins into functional groups

and determining their involvement in specific biological processes (Zhang et al., 2017).

In conclusion, this methodology provides a significant step in identifying potential therapeutic targets for HCC and the small molecule drug candidates that can interact with them. The use of the MDeePred tool, integrating deep learning and bioinformatics methods, contributes to the development of new and effective strategies for HCC treatment.

2.3. *In Silico* Validation of Predicted Small Molecules Targets HCC Transferase by SwissADME and Molecular Docking

Using the SwissADME online tool, we conducted a comprehensive analysis of small molecule drug candidate compounds targeting "HCC-related transferases." This analysis covered several critical aspects, including water solubility, physicochemical properties, lipophilicity, pharmacokinetics, drug-likeness, and medicinal chemistry friendliness. SwissADME provides a detailed assessment of these parameters, which are essential for understanding the potential efficacy and safety of the drug candidates.

The oral bioavailability and brain penetration potentials of the selected small molecules were evaluated using the oral bioavailability radar and the Brain Or IntestinaL EstimateD permeation method (BOILED Egg).

Oral bioavailability radar provides a visual representation of the drug-likeness of the molecules by plotting six key physicochemical properties: lipophilicity (LIPO), size (SIZE), polarity (POLAR), solubility (INSOLU), flexibility (FLEX), and saturation (INSATU). Each property is assessed to determine if the molecule falls within the optimal range for oral bioavailability. Compounds falling within the pink zone are considered to have favorable drug-likeness profiles, which indicate good oral bioavailability potential.

The BOILED Egg model was utilized to predict the passive GIA and brain penetration of the molecules. This model uses the lipophilicity (WLOGP) and the topological polar surface area (TPSA) of the compounds to classify them into three categories: high GIA, high brain penetration, or molecules that do not fall into either category. The resulting data is visualized on a 2D plot where molecules are placed within a 'yolk' (high probability of brain penetration) and/or 'white' (high probability of GIA) zone.

The physicochemical properties analyzed included molecular weight, hydrogen bond donors and acceptors, and TPSA. Lipophilicity was evaluated using multiple models such as XLogP3, WLogP, and MLogP, which predict the partition coefficient (log P) of compounds. Water solubility predictions were generated using the ESOL model, Ali model, and SILICOS-IT model, offering insights into the aqueous solubility of the compounds. These properties are crucial for predicting the absorption and distribution of the drug candidates in biological systems.

Pharmacokinetic properties were assessed to predict the behavior of the compounds in the body, including absorption, distribution, metabolism, excretion (ADME), and potential toxicity. Parameters such as blood-brain barrier permeability, GIA, and P-glycoprotein substrate status were considered. Additionally, the tool evaluated the potential for cytochrome P450 interactions, which are vital for understanding the metabolic stability and potential drug-drug interactions of the compounds.

Drug-likeness was evaluated using established rules such as Lipinski's Rule of Five, which considers factors like log P, molecular weight, hydrogen bond donors, and acceptors. Compounds were also assessed for compliance with other drug-likeness filters such as the Ghose filter, Veber rules, Egan rules, and Muegge rules. Medicinal chemistry friendliness was examined to identify potential structural alerts, pan-assay interference compounds (PAINS), and Brenk alerts, which can indicate undesirable chemical features that might interfere with biological assays (Daina et al., 2017).

Following the SwissADME analysis, molecular docking studies were performed using CB-Dock2, a web server that facilitates blind docking simulations. Blind docking was conducted by inputting the SDF file of each drug compound and the PDB files of five different target proteins into the server. CB-Dock2 identified potential binding sites on the target proteins and calculated the binding affinities of the drug candidates. In addition, molecular docking studies were performed for 6 small molecules and HCC drugs with kinase domain of all transferases.

The docking analysis focused on the docking poses with the highest Vina scores, indicating the strongest binding affinities. Vina scores provide an estimation of the binding free energy between the drug candidate and the target protein, with more negative scores indicating stronger predicted interactions. These docking poses were analyzed to identify the key interactions and binding conformations that could contribute to the efficacy of the drug candidates against HCC-related transferases.

This integrated approach, combining SwissADME and CB-Dock2, provided a thorough evaluation of the small molecule drug candidates, highlighting their potential as effective therapies for HCC by targeting specific transferases involved in the disease (Abhishek et al., 2023; Khachatryan et al., 2023).

In addition, we compared the values of our small compounds to lenvatinib, regorafenib, and sorafenib, which are presently used to treat advanced HCC.

2.4. Literature Based Validation of Novel Drug-Target Interaction Predictions towards Drug Repurposing

To support the DTI evidence presented in published scientific reports, we conducted a comprehensive literature review to validate the "HCC-related transferases" DTI pairs predicted by MDeePred. This step is crucial for corroborating the computational predictions with experimental data, ensuring the reliability and applicability of the

findings in a real-world context. The literature review focused on interactions between target proteins and drugs, gathering evidence that these interactions have been experimentally validated. We examined studies in the literature to collect empirical evidence supporting the drug-target interactions (DTIs) identified by MDeePred. By analyzing these studies, we aimed to confirm the predicted DTIs and enhance the confidence in our computational predictions.

CHAPTER 3

RESULTS

3.1. Data Sets

Hepatocellular carcinoma (HCC) is one of the most common and fatal liver cancers worldwide, and genetic factors play a crucial role in the pathogenesis of the disease. In this context, a comprehensive analysis was performed using the Open Target Platform database to identify genes associated with HCC. The Open Target Platform is a powerful tool for identifying disease-gene associations by integrating various genetic and biomedical data sources. Using this platform to discover genes responsible for HCC is critical for better understanding the genetic basis of the illness and identifying particular genes and mutations that contribute to its progression. The Open Targets Platform allows for genetic associations, somatic mutations, biological pathways, and their linkages to diseases such as. This makes it easier to identify the genes that are directly responsible for the pathophysiology of HCC. As a result of this analysis, 7853 genes thought to be associated with HCC were identified. However, additional selection criteria were applied to more clearly understand the direct relationship of these genes with HCC and to reach more specific results by taking into account somatic mutations. As a result of this selective analysis, which took into account genetic relationships and somatic mutations, 673 genes were identified that were determined to be more strongly associated with HCC. These genes allow a better understanding of genetic factors that may potentially play a crucial role in HCC pathogenesis. Table S1 provides a detailed list of these 673 genes, providing more in-depth information on each gene's association with HCC. The selection of "genetic associations and somatic mutations" was critical in identifying genes directly related with HCC. Genetic associations indicate disease-linked variations and their possible involvement in the development of HCC, whereas somatic mutations indicate genetic alterations that occur inside cancer cells and directly contribute to tumor growth. Combining these two criteria makes the identification of HCC-associated genes more selective and therapeutically relevant. Thus, combining the two approaches not only reveals HCC-associated genes, but also improves our knowledge of their role in HCC. This stage is critical for identifying prospective biomarkers and therapeutic targets, which improves the accuracy and usefulness of the research.

Identification of these genes may make a significant contribution to the in-depth understanding of the molecular mechanisms of HCC. In particular, these genes may elucidate biological pathways and processes that play critical roles in the pathogenesis and progression of HCC.

Furthermore, identification of these genes allows the identification of potential therapeutic targets. The development of novel drugs that can be used in the treatment of HCC may be possible through the inhibition or modulation of these genes and their products. Further studies on these genes may offer new opportunities for early diagnosis of HCC and development of personalized treatment strategies.

Additionally, more comprehensive functional analyzes on these genes may contribute to a better understanding of the biological basis of HCC. This provides new information about the pathophysiology of the disease, allowing the development of innovative strategies for the prevention, early diagnosis and effective treatment of HCC.

In this study, genetic associations and somatic mutation scores were calculated by taking the arithmetic mean in order to more specifically determine the genes associated with HCC. Genetic association scores determine the associations of genes with HCC, while somatic mutation scores reflect the mutation frequencies of these genes observed in HCC patients. The arithmetic mean of both scores was used to more comprehensively evaluate the genes' associations with the disease.

Genes with arithmetic mean values of 0.25 and above were selected as genes with a significant relationship with HCC. The 0.25 cut-off value was used to filter genes highly related with HCC, ensuring that only the most important genes are picked while limiting false positives. This threshold balances the gene pool, preventing excessive noise but not rejecting viable candidates. By integrating genetic association and somatic mutation data, this cut-off improves research reliability and provides biologically meaningful results. As a result of this selection process, 106 genes with strong genetic and somatic associations with HCC were identified. These genes are critical to better understand the biological and molecular basis of HCC.

Table S2 details these 106 selected genes and their genetic associations and somatic mutation scores. In addition, Table 2 shows top 20 selected genes and their genetic associations and somatic mutation scores. This table presents each gene's association with HCC from both a genetic and somatic mutation perspective, providing researchers with more in-depth information about the pathogenesis of the disease.

Table 2: Top 20 Genes and Their Genetic Associations and Somatic Mutations Scores

Target gene	Target ID	Score.overall	Genetic association score	Somatic mutation score	Arithmetic mean
TP53	ENSG00000141510	1.0	1.0	1.0	1
CTNNB1	ENSG00000168036	1.0	1.0	1.0	1
TERT	ENSG00000164362	1.0	0.84	0.71	0.78
PIK3CA	ENSG00000121879	1.0	0.5	1.0	0.75
AXIN1	ENSG00000103126	1.0	0.0	1.0	0.50
MET	ENSG00000105976	1.0	0.0	0.99	0.50
NFE2L2	ENSG00000116044	0.99	0.0	0.97	0.49
IDH1	ENSG00000138413	0.98	0.0	0.96	0.48
HRAS	ENSG00000174775	0.97	0.0	0.95	0.47
CDKN2A	ENSG00000147889	0.96	0.0	0.94	0.47
CREBBP	ENSG00000005339	0.94	0.0	0.93	0.46
NRAS	ENSG00000213281	0.93	0.0	0.92	0.46
IDH2	ENSG00000182054	0.93	0.0	0.91	0.46
ARID2	ENSG00000189079	0.90	0.0	0.87	0.43
GNAS	ENSG00000087460	0.86	0.0	0.86	0.43
KRAS	ENSG00000133703	0.83	0.0	0.82	0.41
FGFR1	ENSG00000077782	1.0	0.0	0.80	0.40
TSC2	ENSG00000103197	0.82	0.0	0.78	0.39
SF3B1	ENSG00000115524	0.79	0.0	0.76	0.38
APC	ENSG00000134982	0.86	0.0	0.75	0.37

Identification of these genes may help to better understand the molecular mechanisms of HCC and elucidate critical biological pathways associated with the disease. In

particular, it may be possible to utilize these genes as biomarkers for early HCC diagnosis and may also provide valuable information for monitoring disease progression and treatment response.

Additionally, further studies on these genes may contribute to the identification of new therapeutic targets. Inhibitors or modulators of these genes may allow the development of new strategies in the treatment of HCC. Therefore, detailed studies on the relationships of these genes with HCC may provide significant advances in the management of the disease.

In this research, the ChEMBL database was used to create a compound-protein activity array dataset for target proteins. The ChEMBL database is a comprehensive data source containing biological activities, pharmacological information and chemical structure information. Half-maximal inhibitory concentration (IC_{50}) and pChEMBL values were taken into account in the selection of compounds for target proteins. While the IC_{50} value expresses the half-maximal inhibitory concentration of a compound, the pChEMBL value is the logarithmic form of this IC_{50} value and allows a more sensitive assessment of the compound's potency. In addition to these criteria, some filtering processes were applied and the details are reported in Table 1.

After the compound-protein activity array was generated, data were analyzed using ChEMBL, UniProt, and Expasy databases. UniProt provides comprehensive information on protein sequences and functions, while the Expasy database provides bioinformatics tools and resources. 46,400 data for 106 genes were grouped through these databases.

In the first stage of the analysis, it was checked whether the target proteins were enzymes or not. Then, the enzyme classes of the selected enzyme targets were determined. This step is critical to understanding the biological activities and roles of target proteins.

As a result, 22 target proteins (38,794 data in total) were identified as enzymes belonging to the transferase class. Transferases are enzymes that transfer functional groups from one molecule to another, and this class of enzymes may play a particularly important role in the biochemical pathways of HCC. Table 3 provides a detailed list of these targets identified as transferases and their associated data. Identification of these targets can make significant contributions to the identification of potential therapeutic targets in HCC treatment and new drug development processes.

Table 3: Identification of 22 Targets as Transferases

Genes	ChEMBL ID	Genes Name
TERT	CHEMBL2916	Telomerase reverse transcriptase
PIK3CA	CHEMBL4005	PI3-kinase p110-alpha subunit
MET	CHEMBL3717	Hepatocyte growth factor receptor
CREBBP	CHEMBL5747	CREB-binding protein
FGFR1	CHEMBL3650	Fibroblast growth factor receptor 1
ACVR2A	CHEMBL5616	Activin receptor type-2A
PRKACA	CHEMBL4101	cAMP-dependent protein kinase alpha-catalytic subunit
ATM	CHEMBL3797	Serine-protein kinase ATM
ROS1	CHEMBL5568	Proto-oncogene tyrosine-protein kinase ROS
AKT1	CHEMBL4282	Serine/threonine-protein kinase AKT
SETD2	CHEMBL3108647	Histone-lysine N-methyltransferase SETD2
FLT3	CHEMBL1974	Tyrosine-protein kinase receptor FLT3
KMT2A	CHEMBL1293299	Histone-lysine N-methyltransferase MLL
FLT4	CHEMBL1955	Vascular endothelial growth factor receptor 3
KIT	CHEMBL1936	Stem cell growth factor receptor
ERBB3	CHEMBL5838	Receptor tyrosine-protein kinase erbB-3
MAP3K1	CHEMBL3956	Mitogen-activated protein kinase kinase kinase 1
KDR	CHEMBL279	Vascular endothelial growth factor receptor 2
ALK	CHEMBL4247	ALK tyrosine kinase receptor
JAK3	CHEMBL2148	Tyrosine-protein kinase JAK3
RET	CHEMBL2041	Tyrosine-protein kinase receptor RET
NTRK1	CHEMBL2815	Nerve growth factor receptor Trk-A

The targets listed in Table 3 are essential to the pathophysiology of HCC because they influence the survival, proliferation, and metastasis of tumor cells. TERT and PIK3CA are associated with telomerase activity and the phosphatidylinositol 3-kinase (PI3K)/AKT signaling pathway, which promotes cell immortalization and survival. Tyrosine kinase receptors, such as MET, ALK, RET, and ROS1, promote HCC cell invasion and metastasis by controlling proliferation, migration, and differentiation. Epigenetic regulators such as CREBBP and SETD2 regulate gene expression via histone modifications, and their dysregulation leads to tumor growth. Signaling pathways including AKT1 and MAP3K1 affect cell survival and apoptosis. When they are active, they enhance tumor development. VEGF receptors (KDR, FLT4, FGFR1, and FLT3) and FGF receptors promote tumor angiogenesis by supplying vital nutrients for development. KIT and ERBB3 receptors control cellular development and differentiation, which promotes tumor proliferation. JAK3 and NTRK1 influence

signaling pathways and immunological responses, whereas ATM and KMT2A affect DNA repair and gene expression. Protein kinases like ACVR2A and PRKACA have roles in signaling pathways that control cell proliferation and differentiation, and their dysregulation contributes to tumor development. Given their critical roles in HCC, these genes are promising therapeutic targets, and inhibitors or modulators have the ability to stop or delay disease development (Cetin Atalay, 2015).

3.2. MDeePred Results

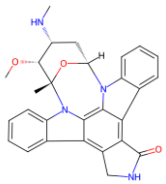
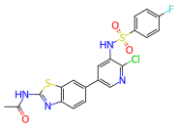
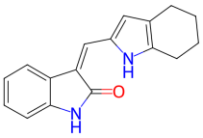
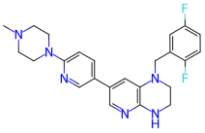
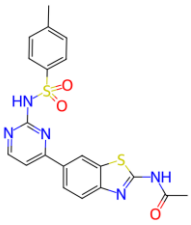
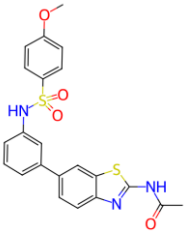
In this research, MDeePred technique was chosen to identify potential drug candidates for HCC. MDeePred is a method that predicts drug target interactions (DTI) using machine learning (ML) approaches. During the preparation of the datasets, data containing multiple data obtained from different experiments of the same gene for the same molecule were deduplicated. During the deduplication process, the median bioactivity value was used to handle duplicate data. This approach reduced noise in the data sets, resulting in more consistent and reliable results.

As a result of the deduplication process, the 38,794 data points initially available for 22 transferases were reduced to 30,821 data. This deduplicated data set was used in the training and testing stages of the model.

Following these preparations, the MDeePred process was carried out. MDeePred trained its model using the training dataset and then identified potential drug-target interactions by making predictions on the test dataset. This process resulted in 380 drug-target interactions (DTIs) associated with HCC (Table S3). These DTIs provide an important resource for identifying new drug candidates and target proteins that could potentially be used in the treatment of HCC.

Among 380 DTIs whose target-compound relationship has been studied in the literature, it was decided to select 6 DTIs to be used in further studies (Table 4). This selection was made considering that these DTIs have high potential in the treatment of HCC and are amenable to further experimental studies. These selected DTIs may contribute to the development of innovative therapeutic approaches in the treatment of HCC and allow significant advances in disease management.

Table 4: Selected Drug-Target Interactions List

Ligand (Drug/Compound)	Molecular Formula of Ligands	2D Structure of Ligands	Target Protein
CHEMBL388978 (Staurosporine)	C28H26N4O3		Tyrosine-protein kinase receptor FLT3 (FLT3)
CHEMBL1615189	C20H14ClFN4O3S2		PI3-kinase p110-alpha subunit (PIK3CA)
CHEMBL328029	C17H16N2O		Fibroblast growth factor receptor 1 (FGFR1)
CHEMBL1165499	C24H26F2N6		ALK tyrosine kinase receptor (ALK)
CHEMBL1773581	C20H17N5O3S2		Serine/threonine-protein kinase AKT (AKT1)
CHEMBL1773601	C22H19N3O4S2		Serine/threonine-protein kinase AKT (AKT1)

This study demonstrates that ML methods can be used effectively in identifying potential drug candidates for HCC and reveals that the MDeePred technique is a valuable tool in such research. Going forward, experimental studies on these DTIs may contribute to the development of new and effective therapeutic strategies in the treatment of HCC.

3.3. Enrichment Analyses of MDeePred Results

Molecular function enrichment analysis for HCC (Hepatocellular carcinoma) is of great importance in understanding the molecular mechanisms of the disease, therapeutic targets, diagnosis and prognosis of the disease, and evaluating the response to treatment. These analyzes serve as a critical tool in elucidating the biological basis of HCC and important pathways involved in the pathogenesis of the disease. Determining the molecular functions of HCC-associated genes and proteins is essential to identify potential therapeutic targets and develop innovative strategies in disease management.

In this study, enrichment analyzes of the results obtained with the MDeePred technique were performed. Enrichment analyzes enable assessing whether sets of genes and proteins are associated with specific biological functions or pathways. As a result of these analyses, HCC-associated molecular functions were grouped into two main categories: transmembrane receptor protein tyrosine kinase activity and adenosine triphosphate (ATP) binding.

Transmembrane receptor protein tyrosine kinase activity (Figure 10) refers to the activity of enzymes located on the cell surface that detect extracellular signals and initiate intracellular signal transduction pathways. These kinases play critical roles in many biological processes such as cell growth, differentiation, metabolism and apoptosis. In HCC, abnormalities in tyrosine kinase activity are important factors that promote the proliferation and metastasis of cancer cells. Therefore, tyrosine kinases attract attention as potential therapeutic molecules that can be targeted in the treatment of HCC.

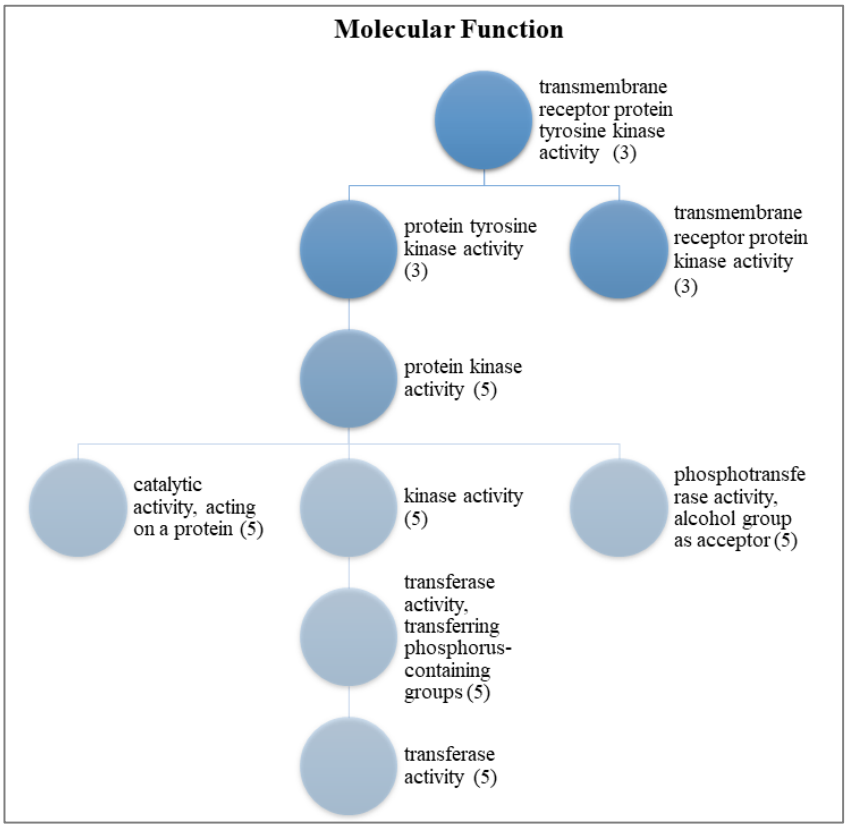


Figure 10: Transmembrane receptor protein tyrosine kinase activity of protein set. (3): FGFR1, ALK and FLT3; (5): FGFR1, ALK, AKT1, FLT3 and PIK3CA

ATP binding (Figure 11) involves the binding of the ATP molecule to proteins and energy transfer processes. ATP is a central component of the cellular energy cycle and plays a vital role in regulating cellular activities and meeting energy needs. In HCC, disruption of ATP binding mechanisms can lead to significant changes in cell metabolism and energy balance. Therefore, proteins and pathways involved in ATP binding may be considered as potential targets in HCC treatment.

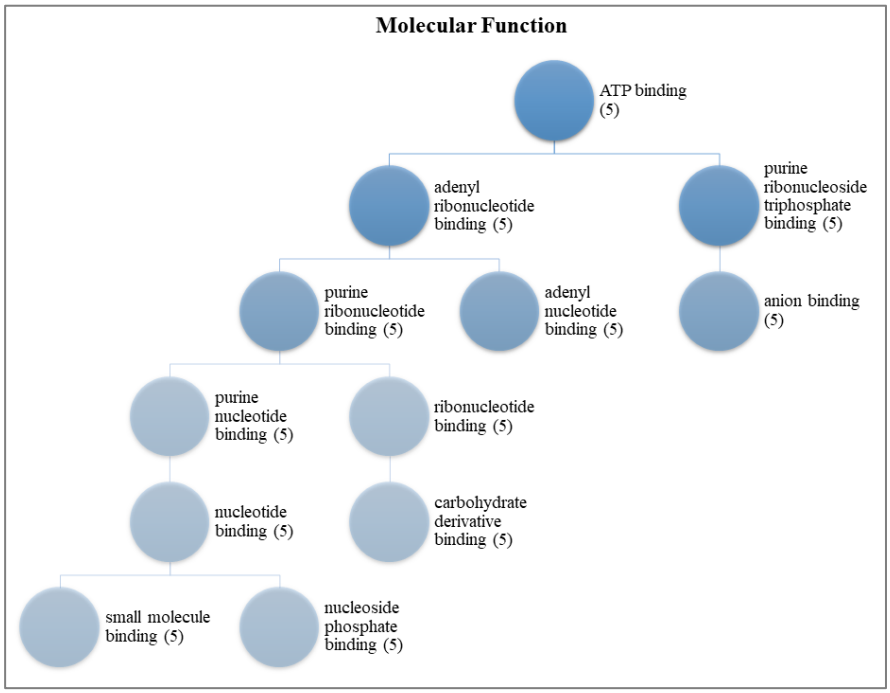


Figure 11: ATP binding of protein set. (5): FGFR1, ALK, AKT1, FLT3 and PIK3CA

Biological process analyzes have been used to understand the biological functions and pathways involved in HCC-related genes and proteins in a broader perspective. As a result of these analyses, 27 main biological process categories were identified (Figure 12). These categories cover a variety of biological processes such as cell cycle regulation, signal transduction, metabolic processes, cell death mechanisms, immune responses, and DNA repair. Understanding these processes provides in-depth insight into the pathogenesis and disease progression of HCC.

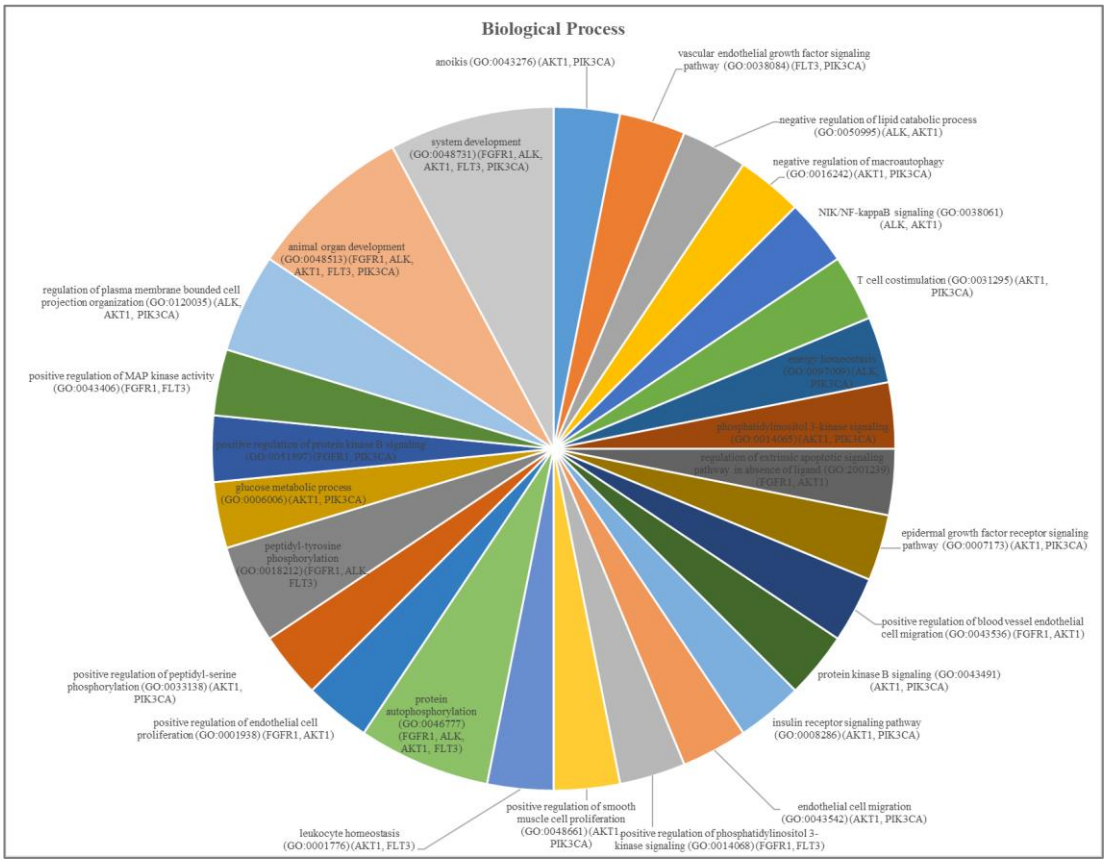


Figure 12: Biological process analyses of protein set

Molecular function enrichment analysis of HCC is a critical tool in understanding the biological basis and molecular mechanisms of this type of cancer. These analyzes allow the development of new approaches for early diagnosis of the disease, prognosis and evaluation of response to treatment. Consequently, molecular function enrichment analysis plays an important role in HCC research and treatment strategies.

3.4. SwissADME and Molecular Docking Results

SwissADME and molecular docking tools are critical in the process of evaluating, optimizing and selecting potential drug candidates for HCC treatment. These tools can contribute to the development of more effective and safer treatments by accelerating the drug development process. SwissADME provides important information in the drug design process by evaluating the pharmacokinetic and pharmacodynamic properties of drug candidates. Molecular docking simulates the interactions of drug candidates with target proteins and predicts the strength and specificity of these interactions.

The schematic diagram of oral bioavailability is a visual tool used to quickly evaluate the pharmacokinetic properties of a drug candidate. This diagram analyzes properties of potential drug candidates such as lipophilicity, size, polarity, insolubility, insaturation, and flexibility. These parameters affect the absorption in the gastrointestinal tract, passage into the systemic circulation and bioavailability of the drug candidate. Oral bioavailability predictions are particularly important in drug design and development because these properties can directly affect the effectiveness and safety of the drug. A schematic diagram showing the oral bioavailability of drug candidate compounds, lenvatinib, regorafenib, and sorafenib is presented in Figure 13.

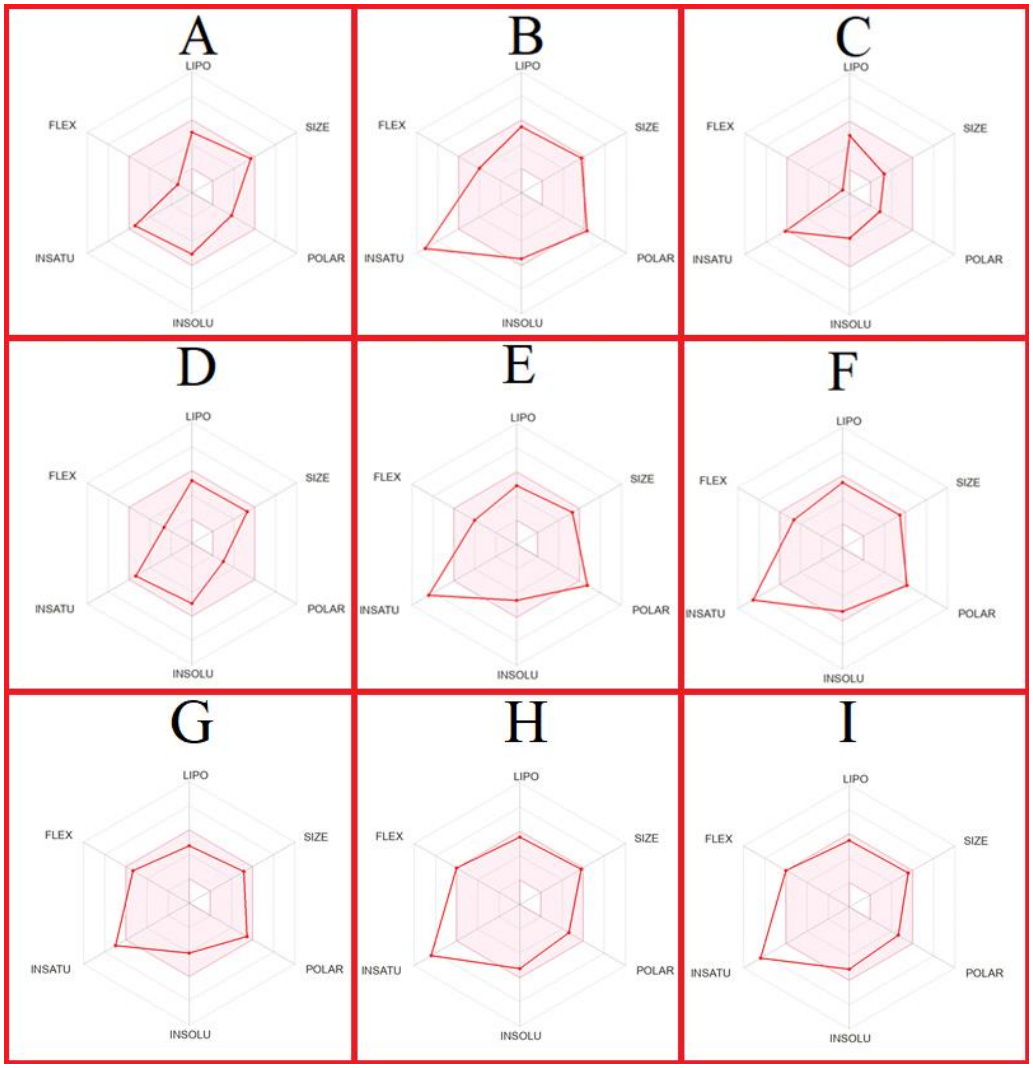


Figure 13: Schematic diagram of oral bioavailability of the drug candidate compounds and drugs. (A) CHEMBL388978. (B) CHEMBL1615189. (C) CHEMBL328029. (D) CHEMBL1165499. (E) CHEMBL1773581. (F) CHEMBL1773601 (G) Lenvatinib. (H) Regorafenib. (I) Sorafenib.

The BOILED Egg diagram is a graphical tool used to estimate the overall absorption, distribution, metabolism, and excretion (ADME) properties of a molecule. This diagram visually represents the gastrointestinal absorption (GIA) ability and likelihood of crossing the BBB of potential drug candidates. The BOILED Egg diagram is an important guide when evaluating the GIA and access to the central nervous system (CNS) of drug molecules. The BOILED Egg diagram of drug candidate compounds and approved HCC drugs is shown in Figure 14.

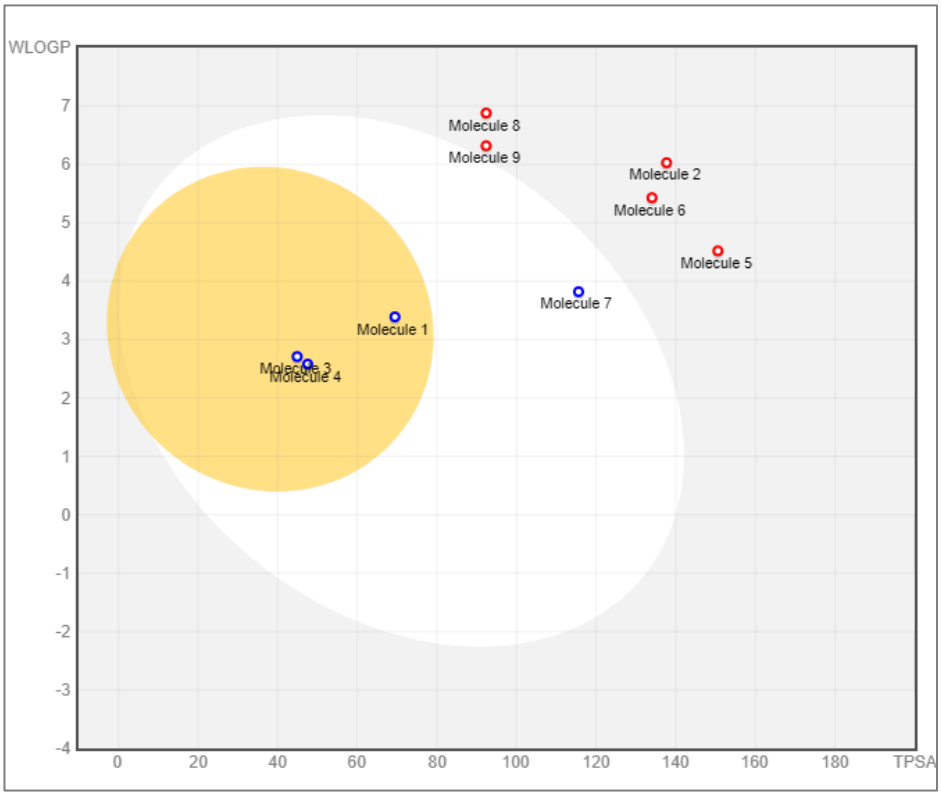


Figure 14: BOILED Egg diagram of the drug candidate compounds and drugs. (1) CHEMBL388978. (2) CHEMBL1615189. (3) CHEMBL328029. (4) CHEMBL1165499. (5) CHEMBL1773581. (6) CHEMBL1773601 (7) Lenvatinib. (8) Regorafenib. (9) Sorafenib.

Predictive findings regarding the physicochemical properties, lipophilicity, water solubility, drug-likeness and medicinal chemistry of drug candidate compounds and lenvatinib, regorafenib, and sorafenib are presented in detail in Table 5-9. These tables comprehensively evaluate the molecular properties of drug candidates and predict their potential therapeutic effects using this information.

Physicochemical properties: Includes properties such as molecular weight, polar surface area, hydrophobic and hydrophilic balance of molecules. These properties affect the drug's ability to cross cell membranes and reach the target (Table 5).

Lipophilicity: The degree of lipophilicity of molecules indicates their affinity for hydrophobic environments. High lipophilicity makes it easier to cross cell membranes, whereas excessive lipophilicity may increase the risk of toxicity (Table 6).

Water Solubility: The solubility of molecules in water is a critical factor in terms of bioavailability and distribution in the body. Low solubility may limit the absorption and effectiveness of the drug (Table 7).

Drug Similarity: The drug similarity of molecules is evaluated based on certain criteria such as Lipinski's Rules. These rules are used to predict the oral bioavailability and pharmacological activity of molecules (Table 8).

Medicinal Chemistry: The chemical structure and functional groups of molecules determine their potential biological activity and therapeutic effects (Table 9).

These analyzes are part of a comprehensive evaluation to identify the most appropriate drug candidates for the treatment of HCC. This process enables important steps towards drug development and contributes to finding safer and more effective treatments for patients.

Additionally, molecular docking analyzes were applied to 6 drug-target interactions, containing 5 different protein targets and 6 drug candidates, selected after MDeePred analysis and classification. In addition, molecular docking analyzes were applied to lenvatinib, regorafenib, and sorafenib. Molecular docking analyzes are used to predict how drug candidates dock at the binding sites of target proteins and the strength of these interactions. These analyzes examine the interaction of drug candidates with target proteins at the molecular level, providing a better understanding of their potential therapeutic effects.

Table 5: Physicochemical Properties of the Drug Candidate Compounds and Drugs

Drug candidate compounds and drugs*	CHEMBL388978	CHEMBL1615189	CHEMBL328029	CHEMBL1165499	CHEMBL1773581	CHEMBL1773601	Lenvatinib	Regorafenib	Sorafenib
Formula	C28H26N4O3	C20H14ClFN4O3S2	C17H16N2O	C24H26F2N6	C20H17N5O3S2	C22H19N3O4S2	C21H19ClN4O4	C21H15ClF4N4O3	C21H16ClF3N4O3
Molecular weight	466.53 g/mol	476.93 g/mol	264.32 g/mol	436.50 g/mol	439.51 g/mol	453.53 g/mol	426.85 g/mol	482.82 g/mol	464.82 g/mol
Num. heavy atoms	35	31	20	32	30	31	30	33	32
Num. arom. heavy atoms	20	21	11	18	21	21	16	18	18
Fraction Csp3	0.32	0.05	0.24	0.33	0.10	0.09	0.19	0.10	0.10
Num. rotatable bonds	2	6	1	4	6	7	8	9	9
Num. H-bond acceptors	4	6	1	5	6	5	5	8	7
Num. H-bond donors	2	2	2	1	2	2	3	3	3
Molar Refractivity	139.39	119.25	83.92	135.40	117.04	122.97	112.86	112.44	112.48
TPSA	69.45 Å²	137.67 Å²	44.89 Å²	47.53 Å²	150.56 Å²	134.01 Å²	115.57 Å²	92.35 Å²	92.35 Å²

*: SMILES of CHEMBL388978: CN[C@@H]1C[C@H]2O[C@]([C@@H]1OC)(C)n1c3ccccc3c3c1c1n2c2ccccc2c1c1c3CNC1=O
SMILES of CHEMBL1615189: CC(=O)Nc1nc2c(s1)cc(cc2)c1cnc(c(c1)NS(=O)(=O)c1ccc(cc1)F)Cl
SMILES of CHEMBL328029: O=C1Nc2c(/C/1=C/c1cc3c([nH]1)CCCC3)cccc2
SMILES of CHEMBL1165499: CN1CCN(CC1)c1ccc(cn1)c1cnc2c(c1)N(CCN2)Cc1cc(F)ccc1F
SMILES of CHEMBL1773581: CC(=O)Nc1nc2c(s1)cc(cc2)c1ccnc(n1)NS(=O)(=O)c1ccc(cc1)C
SMILES of CHEMBL1773601: COc1ccc(cc1)S(=O)(=O)Nc1cccc(c1)c1ccc2c(c1)sc(n2)NC(=O)C
SMILES of Lenvatinib: COc1cc2nccc(c2cc1C(=O)N)Oc1ccc(c(c1)Cl)NC(=O)NC1CC1
SMILES of Regorafenib: CNC(=O)c1nccc(c1)Oc1ccc(c(c1)F)NC(=O)Nc1ccc(c(c1)C(F)(F)F)Cl
SMILES of Sorafenib: CNC(=O)c1nccc(c1)Oc1ccc(cc1)NC(=O)Nc1ccc(c(c1)C(F)(F)F)Cl

Table 6: Lipophilicity of the Drug Candidate Compounds and Drugs

Drug candidate compounds and drugs	CHEMBL388978	CHEMBL1615189	CHEMBL328029	CHEMBL1165499	CHEMBL1773581	CHEMBL1773601	Lenvatinib	Regorafenib	Sorafenib
Log Po/w (iLOGP)	3.19	2.50	2.51	3.29	1.60	2.47	2.68	3.51	3.42
Log Po/w (XLOGP3)	3.24	4.03	2.95	3.59	3.05	4.01	2.75	4.17	4.07
Log Po/w (WLOGP)	3.39	6.03	2.71	2.58	4.52	5.43	3.82	6.88	6.32
Log Po/w (MLOGP)	2.60	2.54	2.44	3.09	1.29	2.38	2.10	3.28	2.91
Log Po/w (SILICOS-IT)	3.02	4.34	4.34	3.51	3.23	3.90	2.70	4.21	3.78
Consensus Log Po/w	3.09	3.89	2.99	3.21	2.74	3.64	2.81	4.41	4.10

Table 7: Water Solubility of the Drug Candidate Compounds and Drugs

Drug candidate compounds and drugs	CHEMBL388978	CHEMBL1615189	CHEMBL328029	CHEMBL1165499	CHEMBL1773581	CHEMBL1773601	Lenvatinib	Regorafenib	Sorafenib
Log *S* (ESOL)	-5.06	-5.44	-3.68	-4.96	-4.61	-5.22	-4.09	-5.27	-5.11
Solubility	4.02e-03 mg/ml; 8.62e-06 mol/l	1.73e-03 mg/ml; 3.62e-06 mol/l	5.54e-02 mg/ml; 2.10e-04 mol/l	4.78e-03 mg/ml; 1.10e-05 mol/	1.08e-02 mg/ml; 2.46e-05 mol/l	2.75e-03 mg/ml; 6.06e-06 mol/l	3.50e-02 mg/ml; 8.21e-05 mol/l	2.59e-03 mg/ml; 5.37e-06 mol/l	3.62e-03 mg/ml; 7.79e-06 mol/l
Class	Moderately soluble	Moderately soluble	Soluble	Moderately soluble	Moderately soluble	Moderately soluble	Moderately soluble	Moderately soluble	Moderately soluble
Log *S* (Ali)	-4.37	-6.62	-3.56	-4.27	-5.88	-6.53	-4.83	-5.82	-5.71
Solubility	1.98e-02 mg/ml; 4.25e-05 mol/l	1.13e-04 mg/ml; 2.38e-07 mol/l	7.36e-02 mg/ml; 2.79e-04 mol/l	2.32e-02 mg/ml; 5.31e-05 mol/l	5.82e-04 mg/ml; 1.32e-06 mol/l	1.35e-04 mg/ml; 2.97e-07 mol/l	6.29e-03 mg/ml; 1.47e-05 mol/l	7.35e-04 mg/ml; 1.52e-06 mol/l	8.98e-04 mg/ml; 1.93e-06 mol/l
Class	Moderately soluble	Poorly soluble	Soluble	Moderately soluble	Moderately soluble	Poorly soluble	Moderately soluble	Moderately soluble	Moderately soluble
Log *S* (SILICOS-IT)	-7.59	-8.74	-5.76	-7.65	-7.90	-8.37	-6.84	-8.86	-8.60
Solubility	1.19e-05 mg/ml; 2.56e-08 mol/l	8.70e-07 mg/ml; 1.82e-09 mol/l	4.63e-04 mg/ml; 1.75e-06 mol/l	9.86e-06 mg/ml; 2.26e-08 mol/l	5.57e-06 mg/ml; 1.27e-08 mol/l	1.94e-06 mg/ml; 4.29e-09 mol/l	6.16e-05 mg/ml; 1.44e-07 mol/l	6.63e-07 mg/ml; 1.37e-09 mol/l	1.16e-06 mg/ml; 2.50e-09 mol/l
Class	Poorly soluble	Poorly soluble	Moderately soluble	Poorly soluble	Poorly soluble	Poorly soluble	Poorly soluble	Poorly soluble	Poorly soluble

Table 8: Druglikeness of the Drug Candidate Compounds and Drugs

Drug candidate compounds and drugs	Lipinski	Ghose	Veber	Egan	Muegge	Bioavailability Score
CHEMBL388978	Yes; 0 violation	No; 1 violation: MR>130	Yes	Yes	No; 1 violation: #rings>7	0.55
CHEMBL1615189	Yes; 0 violation	No; 1 violation: WLOGP>5.6	Yes	No; 2 violations: WLOGP>5.88, TPSA>131.6	Yes	0.55
CHEMBL328029	Yes; 0 violation	Yes	Yes	Yes	Yes	0.55
CHEMBL1165499	Yes; 0 violation	No; 1 violation: MR>130	Yes	Yes	Yes	0.55
CHEMBL1773581	Yes; 0 violation	Yes	No; 1 violation: TPSA>140	No; 1 violation: TPSA>131.6	No; 1 violation: TPSA>150	0.55
CHEMBL1773601	Yes; 0 violation	Yes	Yes	No; 1 violation: TPSA>131.6	Yes	0.55
Lenvatinib	Yes; 0 violation	Yes	Yes	Yes	Yes	0.55
Regorafenib	Yes; 0 violation	No; 2 violations: MW>480, WLOGP>5.6	Yes	No; 1 violation: WLOGP>5.88	Yes	0.55
Sorafenib	Yes; 0 violation	No; 1 violation: WLOGP>5.6	Yes	No; 1 violation: WLOGP>5.88	Yes	0.55

Table 9: Medicinal Chemistry of the Drug Candidate Compounds and Drugs

Drug candidate compounds and drugs	CHEMBL388978	CHEMBL1615189	CHEMBL328029	CHEMBL1165499	CHEMBL1773581	CHEMBL1773601	Lenvatinib	Regorafenib	Sorafenib
PAINS	0 alert	0 alert	0 alert	0 alert	0 alert	0 alert	0 alert	0 alert	0 alert
Brenk	0 alert	1 alert: 2-halo_pyridine	1 alert: michael_acceptor_1	0 alert	0 alert	0 alert	0 alert	0 alert	0 alert
Leadlikeness	No; 1 violation: MW>350	No; 2 violations: MW>350, XLOGP3>3.5	Yes	No; 2 violations: MW>350, XLOGP3>3.5	No; 1 violation: MW>350	No; 2 violations: MW>350, XLOGP3>3.5	No; 2 violations: MW>350, Rotors>7	No; 3 violations: MW>350, Rotors>7, XLOGP3>3.5	No; 3 violations: MW>350, Rotors>7, XLOGP3>3.5
Synthetic accessibility	4.93	3.37	2.82	3.62	3.31	3.30	2.88	3.04	2.87

Figure 15 shows the optimal positions of drug candidates for human HCC in docking into the binding sites of their target proteins. This visual details how each drug candidate interacts with the target protein and the key contact points in the binding site. Molecular docking analyzes produce a detailed profile of drug-target interactions by determining the binding energies of drug molecules (vina score), binding cavity volume (Å^3) and contact residues located in binding sites.

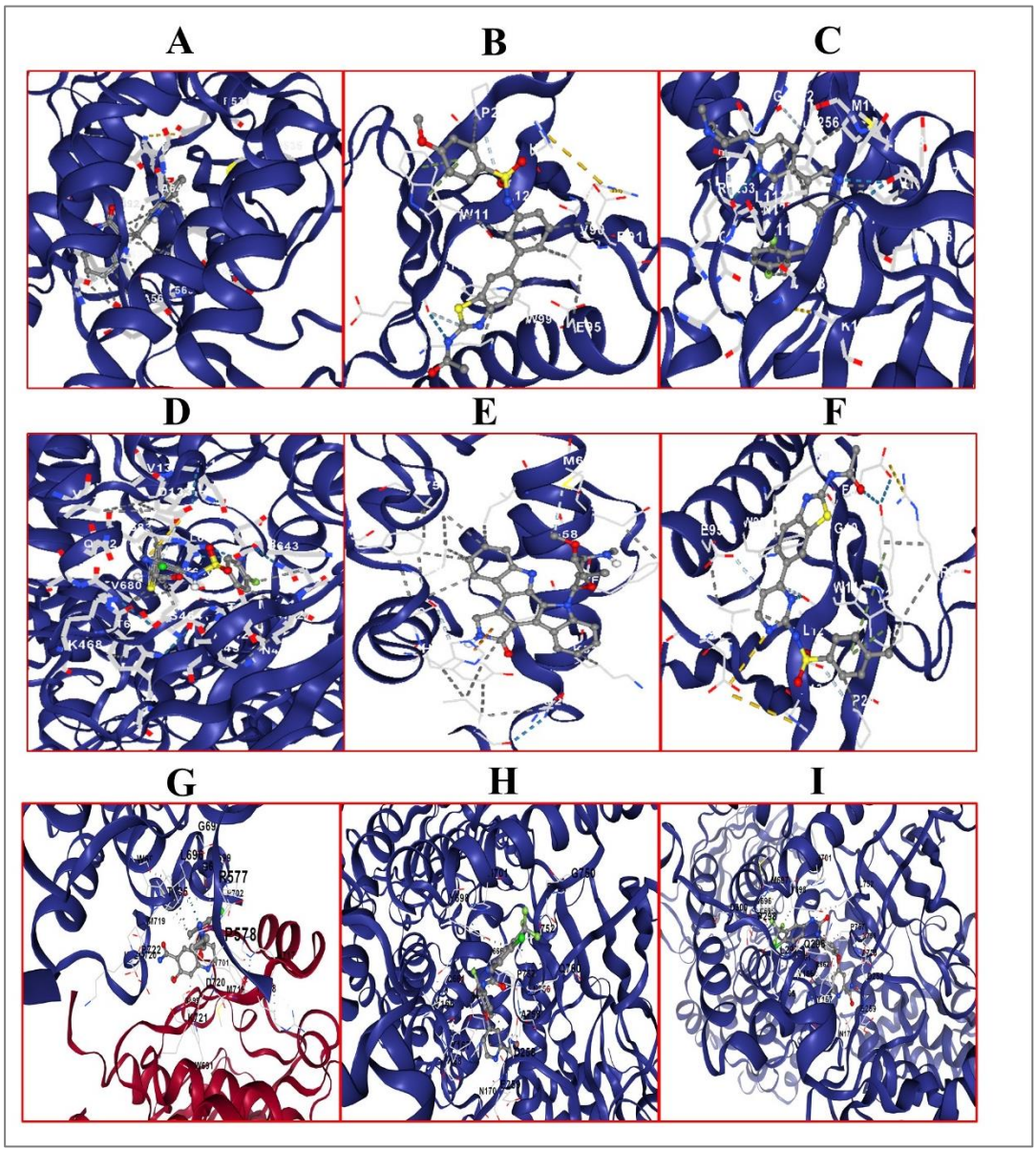


Figure 15: The best poses in the molecular docking of DTIs and drugs. (A) FGFR1 and CHEMBL328029. (B) ALK and CHEMBL1165499. (C) AKT1 and CHEMBL1773601. (D) AKT1 and CHEMBL1773581. (E) FLT3 and CHEMBL388978. (F) PIK3CA and CHEMBL1615189. (G) FGRF1 and levatinib. (H) PIK3CA and regorafenib. (I) PIK3CA and sorafenib.

Blue dotted lines for hydrogen bonds, yellow for electrostatic interactions, and grey for hydrophobic interactions, allows to observe how our drug compounds interact with target proteins.

In Table 10 and Table S4, the results of molecular docking analyzes of drug candidates and lenvatinib, regorafenib, and sorafenib are summarized. This table includes important parameters such as vina scores, binding cavity volume (Å³) and contacting amino acid residues (Table S4) obtained for each drug candidate. The vina score indicates the binding affinity of the drug molecule with the target protein, with lower scores indicating stronger binding. Cavity volume (Å³) refers to the physical size of the binding site, and these values are used to evaluate how well the drug molecule fits into the binding site. Contact residues indicate through which amino acid residues the drug molecule interacts with the target protein.

As a result of these molecular docking analyses, enzymes belonging to the transferase class of 6 DTI were identified as target proteins. These transferases are enzymes that play a critical role in the biochemical pathways of HCC, and drug candidates that interact with these enzymes can be evaluated as potential therapeutic agents. Table 10 presents details of these drug-target interactions, providing important information about the efficacy and safety of each drug candidate.

Table 10: Molecular Docking Results of Small Molecules-Transferases and Drugs-Transferases

Ligand (Drug/Compound)	Target Protein	Vina Score	Cavity Volume (Å^3)
CHEMBL328029	FGFR1	-8.7	543
CHEMBL1165499	ALK	-9.6	6263
CHEMBL1773601	AKT1	-7.8	110
CHEMBL1773581	AKT1	-7.4	110
CHEMBL388978	FLT3	-9.6	1932
CHEMBL1615189	PIK3CA	-9.6	904
Lenvatinib	FGRF1	-9.1	8731
Lenvatinib	ALK	-9.0	1932
Lenvatinib	AKT1	-6.7	110
Lenvatinib	FLT3	-8.5	832
Lenvatinib	PIK3CA	-8.9	6263
Regorafenib	FGRF1	-9.9	8731
Regorafenib	ALK	-9.6	1932
Regorafenib	AKT1	-7.2	110
Regorafenib	FLT3	-8.6	832
Regorafenib	PIK3CA	-10.0	2313
Sorafenib	FGRF1	-9.9	348
Sorafenib	ALK	-8.9	1932
Sorafenib	AKT1	-7.0	152
Sorafenib	FLT3	-10.3	832
Sorafenib	PIK3CA	-10.3	2313

In addition, in this study, molecular docking was performed to determine the binding affinities of six small molecules, identified as potential drug candidates for HCC, to the kinase domains of the targets of six DTIs. For three drug molecules (lenvatinib, regorafenib and sorafenib) currently used in the treatment of advanced HCC, binding

affinities to the kinase domains of the targets were also determined via molecular docking (Table 11 and Table S5).

Table 11: Molecular Docking Results of Small Molecules/Drugs-Kinase Domains of Transferases

Ligand (Drug/compound)	**Target Protein**	**Vina score**	**Cavity volume (Å^3)**
CHEMBL328029	FGFR1	-10.5	1933
Lenvatinib	FGFR1	-9.3	3591
Regorafenib	FGFR1	-11.0	1933
Sorafenib	FGFR1	-11.2	3591
CHEMBL1165499	ALK	-11.3	2409
Lenvatinib	ALK	-8.7	2409
Regorafenib	ALK	-9.6	2409
Sorafenib	ALK	-9.8	2409
CHEMBL1773601	AKT1	-10.8	2670
CHEMBL1773581	AKT1	-10.4	2670
Lenvatinib	AKT1	-9.2	2670
Regorafenib	AKT1	-10.1	2202
Sorafenib	AKT1	-10.3	2670
CHEMBL388978	FLT3	-9.2	805
Lenvatinib	FLT3	-8.3	2684
Regorafenib	FLT3	-10.0	2684
Sorafenib	FLT3	-9.8	805
CHEMBL1615189	PIK3CA	-9.5	3260
Lenvatinib	PIK3CA	-8.7	3260
Regorafenib	PIK3CA	-9.6	3260
Sorafenib	PIK3CA	-9.8	1188

Beyond the six DTIs, the binding affinities of the six small molecules, identified as potential drug candidates for HCC, were also determined for the kinase domains of all the targeted proteins within the scope of this study using molecular docking (Table S5).

The results of this study represent an important step towards identifying and optimizing new drug candidates that can be used in the treatment of HCC. Molecular docking analyzes can play a critical role in the drug development process, contributing to the development of more effective and safe treatments. Going forward, experimental studies on these results may enable the development of innovative therapeutic strategies in the treatment of HCC.

3.5. Literature Based Validation of Novel Drug-Target Interaction Predictions towards Drug Repurposing

A literature review of 380 DTIs found publications supporting the target-compound relationship of only 6 DTIs. These literature-supported interactions provide important information to evaluate the use of potential drug candidates in the treatment of HCC. Table 12 lists these literature-supported DTI estimates along with their references. This table contains literature validation of each interaction and IC_{50} values obtained from the ChEMBL database.

Table 12: Literature Verified Selected DTI Prediction of MDeePred

Ligand (Drug/Compound)	**Target Protein**	**Experimental Bioactivity IC_{50} (nM)**	**Reference**
CHEMBL328029	Fibroblast growth factor receptor 1 (FGFR1)	10500	Tinivella et al. (2022)
CHEMBL1165499	ALK tyrosine kinase receptor (ALK)	33	Das et al., (2021)
CHEMBL1773601	Serine/threonine-protein kinase AKT (AKT1)	1160	Jovanović et al. (2018)
CHEMBL1773581	Serine/threonine-protein kinase AKT (AKT1)	1260	Jovanović et al. (2018)
CHEMBL388978 (Staurosporine)	Tyrosine-protein kinase receptor FLT3 (FLT3)	1	Janssen et al. (2018)
CHEMBL1615189	PI3-kinase p110-alpha subunit (PIK3CA)	6.3	Jovanović et al. (2018)

IC_{50} values indicate the concentration at which a compound inhibits the activity of a particular biological target by 50%. These values play a critical role in evaluating the effectiveness of a drug candidate. Lower IC_{50} values indicate that the compound has a higher inhibitory potency on the target and therefore may be a more effective drug candidate.

Detailed examination of these 6 DTIs supported in the literature constitutes an important step to identify potential drug candidates that can be used in the treatment of HCC. Confirmation of these interactions would enable these drugs to be evaluated prospectively in clinical studies and experimental studies. Additionally, analysis of IC_{50} values helps determine optimal dosages of these compounds and conduct toxicity assessments.

This research plays an important role in the process of identifying and optimizing potential drug candidates for the treatment of HCC. Identification of DTIs supported by the literature allows for a more comprehensive examination of the efficacy and safety profiles of these compounds. This contributes to the development of innovative and effective therapeutic strategies in the treatment of HCC.

CHAPTER 4

DISCUSSION

Hepatocellular carcinoma (HCC) is a prevalent malignant tumor in the digestive system. It holds the fifth position in terms of incidence and third in terms of fatality rate among global malignant tumors, posing a significant risk to health and survival. Liver cancer often develops unnoticed, and the majority of cases are identified at an intermediate or advanced stage, leading to a grim outlook (Tian et al., 2023). HCC is unquestionably a resistant kind of cancer, making therapy more difficult. Although systemic drug therapy has improved survival rates in HCC patients, the progress in treatment outcomes remains gradual and insufficient, particularly when compared to other cancers (Attia et al., 2020). Additionally, the process of developing new drugs is both a lengthy and expensive process. As stated by the Eastern Research Group (ERG), it typically takes 10-15 years, and the average success rate for the development of a new molecular entity is a mere 2.01% (Pan et al., 2022).

Drug repurposing, also known as “new uses for old drugs”, involves finding new applications for approved or experimental drugs that are not within the range of the original medical indication. The key benefit of this method is that the pharmacokinetic, pharmacodynamic, and toxicity profiles of the drugs have already been determined in the initial preclinical and phase I studies. As a result, these drugs can quickly advance to phase II and phase III clinical trials, significantly reducing the associated development costs (Zhang et al., 2020).

The goal of this thesis is to identify possible therapeutic compounds for HCC by repurposing current medication molecules using a machine learning (ML) approach named MDeePred. MDeePred is able to identify potential drug candidates for the targets, we conducted a drug-target interactions (DTIs) study on genes that responsible for HCC. We obtained 380 DTIs using the MDeePred method (Table S3). As a result of the literature review for 380 DTIs, 6 DTIs were selected further studies (Table 4). There are 5 different transferases related to HCC within 6 DTIs as mentioned below. These transferases are FGFR1, ALK, AKT1, FLT3, and PIK3CA. All of these transferases have important roles in a variety metabolic processes and any defect in them is effective in the formation and progression of HCC.

FGFR1 belongs to the type 4 receptor tyrosine kinase family (FGFR1–4), which binds to fibroblast growth factors (FGFs) (Krook et al., 2021). The interaction between the ligand and the FGF receptor is vital in governing various developmental processes during embryonic growth and is crucial for tissue repair and upholding tissue balance

in adults. When FGFs bind to any of these receptors, it triggers signal transduction through the induction of receptor dimerization, leading to subsequent downstream signaling events (Schultheis et al., 2014). Overexpression of FGFR1 has been found to have important roles in the HCC (Zhao et al., 2021).

Anaplastic lymphoma kinase (ALK) is a significant molecular target in the receptor tyrosine kinase family, holding vast relevance in drug discovery, particularly for cancer treatments. ALK is a member of the insulin receptor superfamily and has an important role in multiple malignancies, HCC being one of them. Research has indicated ALK's role in advancing cancer and metastasis, highlighting its importance as a top target for new anticancer drug development (Gowhari Shabgah et al., 2021; Jia et al., 2014).

AKT1, one of the well-preserved subtypes of protein kinase B (AKT), is a crucial component of the phosphatidylinositol 3-kinase (PI3K)-Akt signaling cascade. PI3K/Akt/mammalian target of the rapamycin (mTOR) signaling pathway promotes cell growth, invasion, angiogenesis, and prevents cell apoptosis in various cancers (Guo et al., 2019). Its malfunction is central to the emergence of a wide range of human cancers (Revathidevi and Munirajan, 2019). Initial research identified mutations and imbalances in the expression of the PI3K/Akt/mTOR signaling pathway in HCC (Sun et al., 2021). Loss of phosphatase and tensin homolog (PTEN) tumorsupressor protein leads to hyperactivity in the PI3K/Akt pathway, which promotes cell survival and resistance to treatments in several malignancies, including liver cancer (Durmaz et al., 2016; Aydin et al., 2016).

FLT3 is a receptor tyrosine kinase expressed by immature hematopoietic cells. It is crucial for stem cell development and immune system functioning. Its ligand, produced by marrow stromal and other cells, collaborates with various growth factors to boost the growth of stem cells, dendritic cells, progenitor cells, and natural killer cells (Gilliland and Griffin, 2002). Given the decrease in tumor formation capacity in HCC cells after inhibiting FLT3 activity, it presents as a promising therapeutic target for HCC treatment (Aydin et al., 2016).

Receptor tyrosine kinase dependent PI3K are a family of lipid kinases. They are classified into three classes (Class I-A&B, II, and III) according to their substrate specificities and structural characteristics. In mammals, class I PI3Ks split further into subclasses IA and IB, differentiated by their regulatory methods. Class IA PI3Ks combine a p110 catalytic subunit with a p85 regulatory one. Three closely related class IA catalytic isoforms are encoded by the PIK3CA, PIK3CB, and PIK3CD genes: p110a, p110b, and p110d, as outlined by Verret et al. (2019). These PI3Ks are fundamental to starting the Akt/mTOR signaling pathway, governing crucial cellular functions and activities, such as cell growth, survival, movement, and metabolism. Therefore, PIK3CA is linked to the progression of various cancers, including HCC, positioning it as a potential therapeutic target (Rascio et al., 2021; Vique-Sanchez et al., 2022).

To conclude that since the aforementioned transferases (FGFR1, ALK, AKT1, FLT3 and PIK3CA) have an important role in the formation of HCC, it is important to find drug molecules that affect these transferases in the treatment of HCC.

Following the enrichment analyses, molecular functions were classified into two main categories: transmembrane receptor protein tyrosine kinase activity (Figure 10) and ATP binding (Figure 11). Meanwhile, analyses of biological processes were categorized into 27 primary groups (Figure 12). To conclude that, all transferases have kinase activity and ATP binding activity. As the 5 transferases of which we have explained above, kinases take part in important metabolic mechanisms in the cell, so any deterioration in their function causes HCC (Cicenes et al., 2018). As tyrosine kinase inhibitors (TKIs) are the backbone of systematic therapy for advanced HCC (Chen et al., 2022). It is important for HCC treatment that our drug compounds bind to and inhibit our targets with tyrosine kinase activity. As a result of the literature review, it is seen that these 6 drug compounds are completely new drug candidates for treatment of HCC (Table 12).

As far as we know, for HCC treatment, the predicted drug molecules have never been tested against these transferases using *in silico*, *in vitro*, or *in vivo* assays with respect to HCC. After that, SwissADME (Figure 13, Figure 14, and Table 5-9) and molecular docking (Figure 15, Table 10, and Table 11) were determined for 6 DTIs, which contains 5 different transferases and 6 drug candidates, and HCC approved drugs (lenvatinib, regorafenib, and sorafenib).

The oral bioavailability radar offers a swift evaluation of a compound's drug-likeness by evaluating six physicochemical properties: saturation, lipophilicity, polarity, size, solubility, and flexibility (Adbullahi and Adeniji, 2020). The pink area indicates the optimal range for each property. Lipophilicity (XLOGP3) ranged between -0.7 and +5.0. For size, molecular weights were between 150 and 500 g/mol. Polarity, defined by topological polar surface area (TPSA), ranged from 20 to 130 $Å^2$, while solubility (log S) did not exceed 6. In addition, saturation, indicated by the fraction of carbons in sp^3 hybridization, was not less than 0.25, and flexibility was defined by a maximum of 9 rotatable bonds (El-Nashar et al., 2022). Oral bioavailability radar of CHEMBL388978, CHEMBL1615189, CHEMBL328029, CHEMBL1165499, CHEMBL1773581, CHEMBL1773601, and HCC approved drugs (lenvatinib, regorafenib, and sorafenib) were shown in Figure 13. In addition, physicochemical properties of CHEMBL388978, CHEMBL1615189, CHEMBL328029, CHEMBL1165499, CHEMBL1773581, CHEMBL1773601, and HCC approved drugs (lenvatinib, regorafenib, and sorafenib) were shown in Table 5.

Molecular weight (size), XLOGP3 (lipophilicity), fraction csp^3 (saturation), TPSA (polarity), and number of rotatable bonds (flexibility) properties of CHEMBL388978, CHEMBL1615189, CHEMBL328029, CHEMBL1165499, CHEMBL1773581, and CHEMBL1773601 are used in the oral bioavailability radar for determination of the favourable zone. CHEMBL388978, CHEMBL328029, and CHEMBL1165499 are within favourable zone for lipophilicity, size, polarity, solubility, saturation, and

flexibility. CHEMBL1615189 and CHEMBL1773601 are within favourable zone for lipophilicity, size, polarity, solubility, and flexibility except saturation. CHEMBL1773581 is within favourable zone for lipophilicity, size, solubility, and flexibility except saturation and polarity. The analyzed drugs approved for HCC treatment Lenvatinib, regorafenib, and sorafenib meets all criteria except saturation.

In the BOILED Egg diagram, compounds situated within the white elliptical area signify a high likelihood of passive absorption in the gastrointestinal tract. Therefore, compounds fall in the white ellipse region can be absorbed by the human intestinal system. Meanwhile, compounds positioned in the yellow elliptical area, or the "yolk", indicate a strong chance of penetrating the BBB to reach the central nervous system (CNS). Hence, compounds fall in the yellow ellipse region can pass through the BBB easily (Şahin and Dege, 2022). BOILED Egg diagram of CHEMBL388978, CHEMBL1615189, CHEMBL328029, CHEMBL1165499, CHEMBL1773581, CHEMBL1773601, and HCC approved drugs (lenvatinib, regorafenib, and sorafenib) were shown in Figure 14. CHEMBL388978, CHEMBL328029, and CHEMBL1165499 showed the high BBB penetration and gastrointestinal absorption (GIA) ability according to the BOILED Egg diagram. In addition, when we look at approved drugs, only lenvatinib demonstrated high gastrointentinal absorption.

Five different approaches (iLOGP, XLOGP3, WLOGP, MLOGP, and SILICOS-IT) are used for estimation of lipophilicity (Log Po/w). These methods illustrate alternative approaches to determining how lipophilic (or hydrophobic) a molecule is. Log Po/w measures, such as iLOGP, were determined for n-octanol and water based on the free energies of solvation using the generalized-born and solvent accessible surface area (GB/SA) model by Daina et al. (2014), XLOGP3 is an atomic-based approach with correction factors and a knowledge library developed by Cheng et al. (2007), WLOGP uses a purely atomistic method derived from the fragment system of Wildman and Crippen (1999), MLOGP is a representative topological method that uses a linear relationship with 13 molecular descriptors, as outlined by Moriguchi et al. (1992,1994). Meanwhile, SILICOS-IT combines both approaches, leveraging 27 fragments and 7 topological descriptors. The consensus Log Po/w is the arithmetic mean of the values predicted by the five proposed methods. The averaging process minimizes the possible biases or errors of various approaches, resulting in a more reliable estimate (Daina et al., 2017). Log Po/w is ranged between −0.7 and +5.0 according to the oral bioavailability radar. Log Po/w of CHEMBL388978, CHEMBL1615189, CHEMBL328029, CHEMBL1165499, CHEMBL1773581, CHEMBL1773601, and HCC approved drugs (lenvatinib, regorafenib, and sorafenib) were shown in Table 6. CHEMBL388978, CHEMBL1615189, CHEMBL328029, CHEMBL1165499, CHEMBL1773581, and CHEMBL1773601 are within acceptable range for consensus Log Po/w, indicating favorable characteristics for oral absorption. Lenvatinib is falling within the acceptable range for consensus Log Po/w. Regorafenib and sorafenib are falling within the acceptable range except Log Po/w (WLOGP).

Water solubility (Log S) is ranged between insoluble < -10 < poorly < -6 < moderately < -4 < soluble < -2 very soluble < 0 highly (Henning et al., 2023; Yağlıoğlu et al.,

2022). Log S values of CHEMBL388978, CHEMBL1615189, CHEMBL328029, CHEMBL1165499, CHEMBL1773581, CHEMBL1773601, and HCC approved drugs (lenvatinib, regorafenib, and sorafenib) were shown in Table 7. CHEMBL388978, CHEMBL1165499, and CHEMBL1773581 are within acceptable range for moderately soluble class, while CHEMBL328029 is within acceptable range for soluble class. CHEMBL1615189 and CHEMBL1773601 are within acceptable range for poorly soluble class. Moreover, lenvatinib, regorefenib, and sorafenib are within acceptable range for moderately soluble class.

The druglikeness of the candidate compounds is determined using SwissADME, which uses rule-based filters and the Abbot bioavailability score to identify eligibility based on key pharmacokinetics characteristics. SwissADME uses filtering methods on chemical collections to remove molecules that don't meet the desired pharmacokinetics criteria. This is achieved using five specific rule-based filters created by prominent pharmaceutical corporations [Lipinski (Pfizer), Ghose (Amgen), Veber (GSK), Egan (Pharmacia), and Muegge (Bayer)]. These filters aim to improve the standard of specialized chemical libraries. Furthermore, the Abbott bioavailability score is designed to estimate the chance that a molecule will exhibit a minimum of 10% oral bioavailability in rats or detectable Caco-2 permeability. It forecasts the likelihood of a molecule achieving F>10%, taking into account its primary charge at biological pH in a rat model (Ranjith and Ravikumar, 2019). Druglikeness of the drug candidate compounds and approved drugs were shown in the Table 8. According to druglikeness scores, CHEMBL388978 is within acceptable range for 3 different rule-based filters (Lipinsky, Veber, and Egan) and Abbot bioavailability score. CHEMBL1615189 is within acceptable range for 3 different rule-based filters (Lipinsky, Veber, and Muegge) and Abbot bioavailability score. CHEMBL328029 is within acceptable range for all rule-based filters (Lipinsky, Ghose, Veber, Egan, and Muegge) and Abbot bioavailability score. CHEMBL1165499 is within acceptable range for all rule-based filters except Ghose (Lipinsky, Veber, Egan, and Muegge) and Abbot bioavailability score. CHEMBL1773581 is within acceptable range for 2 different rule-based filters (Lipinsky and Ghose) and Abbot bioavailability score. Lastly, CHEMBL1773601 is within acceptable range for all rule-based filters except Egan (Lipinsky, Ghose, Veber, and Muegge) and Abbot bioavailability score. These assessments show that the compounds have desirable properties for medication development, with various degrees of adherence to recognized pharmacokinetic criteria. When we look at approved drugs, lenvatinib is within acceptable range for all rule-based filters and Abbot bioavailability scor, while regorafenib and sorafenib are within acceptable range for all rule-based filters except Ghose and Egan.

Pan-assay interference compounds (PAINS) have a tendency to nonspecifically interact with multiple biological targets instead of specifically influencing a single desired target. Another tool used in the assessment is Brenk filters. Brenk filters out unwanted functionalities due to potential toxicity or unfavorable pharmacokinetics (Iwaloye et al., 2022). “Lead-likeness” is a strategic guideline for selecting initial points for chemical optimization that have the highest probability of yielding “drug-like” candidates at the conclusion of drug discovery programs. The synthetic

accessibility score of a compound, which measures the ease of synthesis, ranges from 1 (very easy) to 10 (very difficult) (Daina et al., 2017). Medicinal chemistry of the drug candidate compounds and approved drugs were shown in Table 9. According to PAINS filter, all drug candidate compounds and all approved drugs (lenvatinib, regorafenib, and sorafenib) had no alert. Four drug candidate compounds (CHEMBL388978, CHEMBL1165499, CHEMBL1773581 and CHEMBL1773601) and all approved drugs (lenvatinib, regorafenib, and sorafenib) had no alert in Brenk filter. However, two drug candidate compounds (CHEMBL1615189 and CHEMBL328029) had 1 alert in Brenk filter. Another key factor is leadlikeness, which evaluates a compound's overall potential as a starting point for therapeutic development. According to leadlikeness, only CHEMBL328029 had no violation. Lastly, the synthetic accessibility is another parameter, which measures how easily compounds can be produced using standard synthetic methods. Synthetic accessibility of all the drug candidate compounds are close to very easy. However, among the leadlike compounds, CHEMBL328029 is highly accessible, which makes it a viable synthetic option for more research and development. Despite the single alert in the Brenk filter, CHEMBL328029 stands out in the group because to this mix of advantageous features.

Molecular docking is an important tool used to predict the binding behaviors of small molecules to their target proteins, identifying potential sites and affinities crucial for drug development (Gaillard, 2018; Moharana et al., 2023). Here, the docking results provides insights into the molecular interactions specific to HCC for the MDeePred DTIs. “Vina score” mentioned in the Table 10 and Table 11 is an entirely empirical scoring system. It comprises Gaussian steric interaction components, a set repulsion element, piecewise linear hydrophobic and hydrogen-bond interaction components, and an entropy component based on the count of rotatable bonds (Gaillard, 2018). The greater the negative value of the vina score, the greater the Gibbs binding energy for drug-target complexes. This increases the binding potential of drug-target complexes. Contact residues and bonds showed contact amino acids and bond structures between ligands and target proteins. Teal dotted lines illustrate hydrogen bond interactions. Yellow dotted lines show electrostatic interactions, and the hydrophobic interactions are denoted by grey dotted lines (Moharana et al., 2023). As a result, we can clearly see from the molecular docking results (both blind docking and docking for kinase domain) that our drug compounds in the 6 DTIs bind to our transferases (Figure 15, Table 10, and Table 11). In addition, the vina scores for lenvatinib, regorafenib, and sorafenib, which are used to treat advanced HCC, were compared to the vina scores of six small molecules identified in our study. For ALK, CHEMBL1165499 was shown to have a negative vina score (also known as Gibbs binding energy) that was greater than sorafenib and equal to regorafenib. For AKT1, CHEMBL1773601 and CHEMB1773581 were found to have negative vina scores that were greater than those of any other medication. For FLT3, CHEMBL388978 was shown to have a greater negative vina score than regorafenib and lenvatinib. For PIK3CA, it was discovered that CHEMBL1615189 had a larger negative vina score than lenvatinib. The six small molecules included in our study are potential medication candidates to be employed in the treatment of HCC, according to comparisons with lenvatinib, regorafenib, and

sorafenib, which are utilized in the treatment of advanced HCC. In addition, the molecular docking results for the kinase domains of the targets indicate that the six small molecules identified in this study also exhibited effective binding affinities to the kinase domains of targets beyond those discussed within the scope of this study. This suggests that the six small molecules identified here have the potential to be used as kinase inhibitors in the treatment of HCC (Table S4).

Finally, we performed literature survey about the MDeePred predicted small molecules (Table 12) (Das et al., 2021; Janssen et al., 2018; Jovanović et al., 2018; Tinivella et al., 2022). Tinivella et al. (2022) determined potential biological targets of small molecules with *in silico* repositioning strategies and predicted their interactions using ligand-based similarity predictions and molecular docking analyses. Das et al. (2021) identified new kinases as therapeutic drug targets in their study. They used computational methods to predict the activity of small molecules. Jovanović et al. (2018) in their study, the 3D pharmacophore structure of selective kinase inhibitors was examined using molecular modeling and 3D-QSAR methods. Validated with experimental data, the model showed high reliability in predicting the effectiveness of inhibitors. Janssen et al. (2018) mapped the activity profile of compounds across the kinase family using a ML approach. This model predicted the activities of novel kinase inhibitors.

In the study conducted by Rifaioglu et al. (2021), a performance comparison was made between MDeePred and existing advanced techniques. In this context, it was stated that the MDeePred model exhibited superior performance in many evaluation metrics when compared to prominent methods such as CGKronRLS, DeepDTA, SimBoost and Grid-RF, Grid-DNN, ECFP4-RF and ECFP-RF on the MoleculeNet platform. It was stated that MDeePred outperformed all of these methods in terms of consistency and accuracy in benchmark data sets such as Davis dataset, Filtered Davis dataset and PDBBind Refined dataset. It was emphasized that MDeePred showed strong results, especially at changing biological activity threshold values, in metrics such as concordance index (CI), mean squared error (MSE) and Spearman rank correlation. In addition, it was stated that MDeePred's hybrid deep neural network architecture, which extracts complex representations from target proteins and compounds separately and then combines them, significantly contributes to improved prediction capabilities. This is particularly evident when compared to models based solely on molecular fingerprint-based features. MDeePred was found to be less sensitive to performance degradation at different biological activity thresholds. These results demonstrate that MDeePred’s innovative protein characterization approach and deep learning architecture provide a powerful alternative to current state-of-the-art methods for binding affinity prediction and hold significant potential in computational drug discovery.

In addition, in the study conducted by Rifaioglu et al. (2021), an *in vitro* comparative analysis of MDeePred was also performed. In this context, the effects of selected kinase inhibitors on Huh7 and Mahlavu human cancer cell lines were confirmed with the predictions obtained from the MDeePred analysis. In the study, differential gene

expression analyses were performed using well-characterized inhibitors targeting kinase pathways such as staurosporine, rapamycin, NVP-BEZ235 and alsterpaullone. RNA sequencing and subsequent gene expression profiling showed that the predicted binding affinities of MDeePred correlated with the expression changes observed in kinase-related pathways after treatment with these inhibitors. In particular, the study showed that pathways involving protein kinases were significantly changed in response to inhibitors, which supported the predicted mechanism of action of MDeePred. For example, rapamycin treatment targeted the PI3K/Akt/mTOR pathway, leading to upregulation of alternative protein kinases, highlighting possible compensatory mechanisms. Similarly, inhibitors targeting kinases in the cell cycle and MAPK signaling pathway, such as NVP-BEZ235 and alsterpaullone, resulted in downregulation of kinases with low predicted binding affinities. This demonstrated that MDeePred accurately identified the kinases most affected by specific inhibitors, consistent with the observed gene expression profile changes.

CHAPTER 5

CONCLUSION

Hepatocellular carcinoma (HCC) is the most common primary liver cancer with a high mortality rate due to limited treatment options. Although current systemic drug treatments slightly increase patients' survival rates, these treatments generally extend lifespan by only a few months and are insufficient to improve patients' quality of life. The development of new small molecule chemotherapeutics is quite time-consuming and costly, so alternative strategies are of great importance. In this context, the drug repurposing approach emerges as an important strategy in the rapid and effective identification and implementation of new treatment options for this deadly disease.

In this study, we aimed to identify potential drug candidates for HCC treatment by repurposing existing compounds using a machine learning (ML) tool called MDeePred. The open target platform, UniProt, ChEMBL, and Expasy databases were used to create a dataset for MDeePred to predict drug-target interactions (DTIs). The physicochemical properties, lipophilicity, water solubility, drug-likeness and medicinal chemistry properties of the drug candidates were meticulously determined and curated.

Enrichment analyzes of DTIs were performed, resulting in 6 out of 380 DTIs identified by MDeePred being selected for further analysis. The vast majority of these drug candidates fall within conventional ranges in terms of drug properties and demonstrate target docking capabilities. The findings reveal the binding efficiency of selected drug compounds to identified targets associated with HCC. These results suggest that the drug repurposing strategy is a potentially promising avenue for treating HCC.

The current therapy options for HCC differ based on the illness stage, tumor size, the patient's overall health, and liver function. However, the most commonly approved treatments include surgical resection, liver transplantation, local ablative therapies, TACE, molecular targeted medicines, and immunotherapy. Our research focuses on molecular targeted medicines, especially TKIs such as sorafenib and lenvatinib, which are authorized therapy for advanced HCC. These small molecule medicines decrease tumor development and angiogenesis. Our research identified six small compounds that interact with genes that have kinase activity, including FGFR1, ALK, and FLT3 proteins having tyrosine kinase activity. These results imply that the six small compounds might be used as kinase or TKIs in the treatment of HCC. After *in vitro* and *in vivo* testing, these potential drug candidates will be ready for clinical usage.

In conclusion, this study identified potential drug candidates that can be further evaluated experimentally for the treatment of HCC using the ML-based *in silico* MDeePred method. Additionally, this study highlights that the MDeePred deep learning tool can play an important role in drug repurposing studies in cancer treatment. In other words, we showed that ML techniques can be applied to drug repurposing in HCC to find new therapeutic agents with very drug-like qualities that are comparable to those of approved HCC medications. MDeePred's high accuracy rates and comprehensive data processing capacity accelerate the drug discovery process and increase the potential to develop more effective treatment options. This approach aims to contribute to the development of new therapeutic strategies for HCC and other types of cancer, providing better quality of life and long-term survival rates for patients.

REFERENCES

Abdullahi, M., & Adeniji, S. E. (2020). In-silico molecular docking and ADME/Pharmacokinetic prediction studies of some novel carboxamide derivatives as anti-tubercular agents. *Chemistry Africa*, *3*(4), 989-1000. https://doi.org/10.1007/s42250-020-00162-3

Abhishek, A. G., Saini, A., Nagarajan, K., Bhatt, P., Kapoor, G., Mehta, S., Mishra, S., Shrivastava, S., Ahlawat, A., & Shrivastava, A. (2023). Comparative in-silico screening of potent peptide leads using docking strategy & AI approaches for the treatment of liver cancer. *Journal of Pharmaceutical Negative Results*, 3144-3152.

Agarwal, S., & Mehrotra, R. (2016). An overview of molecular docking. *JSM chem*, *4*(2), 1024-1028.

Alameri, M. M., Kong, A. S., Aljaafari, M. N., Ali, H. A., Eid, K., Sallagi, M. A., Cheng, W. H., Abushelaibi, A., Lim, S. E., Loh, J. Y., & Lai, K. S. (2023). Aflatoxin contamination: An overview on health issues, detection and management strategies. *Toxins (Basel)*, *15*(4). https://doi.org/10.3390/toxins15040246

Alharbi, F., & Vakanski, A. (2023). Machine learning methods for cancer classification using gene expression data: A review. *Bioengineering (Basel)*, *10*(2). https://doi.org/10.3390/bioengineering10020173

Andreana, L., Isgro, G., Marelli, L., Davies, N., Yu, D., Navalkissoor, S., & Burroughs, A. K. (2012). Treatment of hepatocellular carcinoma (HCC) by intra-arterial infusion of radio-emitter compounds: trans-arterial radio-embolisation of HCC. *Cancer Treat Rev*, *38*(6), 641-649. https://doi.org/10.1016/j.ctrv.2011.11.004

Annett, S. (2021). Pharmaceutical drug development: high drug prices and the hidden role of public funding. *Biol Futur*, *72*(2), 129-138. https://doi.org/10.1007/s42977-020-00025-5

Antwi, S. O., Craver, E. C., Nartey, Y. A., Sartorius, K., & Patel, T. (2022). Metabolic risk factors for hepatocellular carcinoma in patients with nonalcoholic fatty liver disease: A prospective study. *Cancers (Basel)*, *14*(24). https://doi.org/10.3390/cancers14246234

Attia, Y. M., Ewida, H., & Ahmed, M. S. (2020). Successful stories of drug repurposing for cancer therapy in hepatocellular carcinoma. In *Drug Repurposing in Cancer Therapy* (pp. 213-229). https://doi.org/10.1016/b978-0-12-819668-7.00008-7

Awuchi, C. G., Amagwula, I. O., Priya, P., Kumar, R., Yezdani, U., & Khan, M. G. (2020). Aflatoxins in foods and feeds: A review on health implications, detection, and control. *Bull. Environ. Pharmacol. Life Sci*, *9*, 149-155.

Aydin, M. M., Bayin, N. S., Acun, T., Yakicier, M. C., & Akcali, K. C. (2016). Role of FLT3 in the proliferation and aggressiveness of hepatocellular carcinoma. *Turk J Med Sci*, *46*(2), 572-581. https://doi.org/10.3906/sag-1501-173

Azzam, K. (2022). SwissADME and pkCSM Webservers Predictors: An integrated Online Platform for Accurate and Comprehensive Predictions for In Silico ADME/T Properties of Artemisinin and Its Derivatives. *Kompleksnoe Ispolzovanie Mineralnogo Syra=Complex Use of Mineral Resources*, *325(2)*, 14-21. https://doi.org/10.31643/2023/6445.13

Babu, B. M. (2020). Drug repurposing and it's implications in therapy: An overview [Review]. *International Journal of Research in Pharmaceutical Sciences*, *11*(3), 4418-4423. https://doi.org/10.26452/ijrps.v11i3.2661

Berdigaliyev, N., & Aljofan, M. (2020). An overview of drug discovery and development. *Future medicinal chemistry*, *12*(10), 939-947.

Bhagat, R. T., & Butle, S. R. (2021). Drug repurposing: a review. *Journal of Pharmaceutical Research International*, *33*(31B), 161-169.

Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., & Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca Cancer J Clin*, *74*(3), 229-263. https://doi.org/10.3322/caac.21834

Brown, J. S., Amend, S. R., Austin, R. H., Gatenby, R. A., Hammarlund, E. U., & Pienta, K. J. (2023). Updating the Definition of Cancer. *Mol Cancer Res*, *21*(11), 1142-1147. https://doi.org/10.1158/1541-7786.MCR-23-0411

Buch, S., Innes, H., Lutz, P. L., Nischalke, H. D., Marquardt, J. U., Fischer, J., Weiss, K. H., Rosendahl, J., Marot, A., Krawczyk, M., Casper, M., Lammert, F., Eyer, F., Vogel, A., Marhenke, S., von Felden, J., Sharma, R., Atkinson, S. R., McQuillin, A., Stickel, F. (2023). Genetic variation in TERT modifies the risk of hepatocellular carcinoma in alcohol-related cirrhosis: results from a genome-wide case-control study. *Gut*, *72*(2), 381-391. https://doi.org/10.1136/gutjnl-2022-327196

Butina, D., Segall, M. D., & Frankcombe, K. (2002). Predicting ADME properties in silico: methods and models. *Drug discovery today*, *7*(11), P83-S88. https://doi.org/10.1016/S1359-6446(02)02288-2

Cai, P., Zheng, H., She, J., Feng, N., Zou, H., Gu, J., Yuan, Y., Liu, X., Liu, Z., & Bian, J. (2020). Molecular mechanism of aflatoxin-induced hepatocellular carcinoma derived from a bioinformatics analysis. *Toxins (Basel)*, *12*(3). https://doi.org/10.3390/toxins12030203

Campos-Valdez, M., Monroy-Ramirez, H. C., Armendariz-Borunda, J., & Sanchez-Orozco, L. V. (2021). Molecular mechanisms during hepatitis B infection and the effects of the virus variability. *Viruses*, *13*(6). https://doi.org/10.3390/v13061167

Cannella, R., Zins, M., & Brancatelli, G. (2024). ESR Essentials: diagnosis of hepatocellular carcinoma-practice recommendations by ESGAR. *Eur Radiol*, *34*(4), 2127-2139. https://doi.org/10.1007/s00330-024-10606-w

Chami, P., Jarnagin, W., Abou-Alfa, G. K., Harding, J., Kim, N., Lin, H., El Homsi, M., Crane, C., & Hajj, C. (2023). Non-surgical locoregional therapies alone or

in combination with systemic therapy in patients with hepatocellular carcinoma. *Cancers (Basel)*, *15*(6). https://doi.org/10.3390/cancers15061748

Chan, H. C. S., Shan, H., Dahoun, T., Vogel, H., & Yuan, S. (2019). Advancing drug discovery via artificial intelligence. *Trends Pharmacol Sci*, *40*(8), 592-604. https://doi.org/10.1016/j.tips.2019.06.004

Chen, R., Li, Q., Xu, S., Ye, C., Tian, T., Jiang, Q., Shan, J., & Ruan, J. (2022). Modulation of the tumour microenvironment in hepatocellular carcinoma by tyrosine kinase inhibitors: from modulation to combination therapy targeting the microenvironment. *Cancer Cell Int*, *22*(1), 73. https://doi.org/10.1186/s12935-021-02435-4

Cheng, T., Zhao, Y., Li, X., Lin, F., Xu, Y., Zhang, X., Li, Y., Wang, R., & Lai, L. (2007). Computation of octanol-water partition coefficients by guiding an additive model with knowledge [Article]. *Journal of Chemical Information and Modeling*, *47*(6), 2140-2148. https://doi.org/10.1021/ci700257y

Chevallier, O., Zhao, K., Marinelli, B., & Yarmohammadi, H. (2023). Image-guided percutaneous locoregional therapies for hepatocellular carcinoma. *Chin Clin Oncol*, *12*(2), 17. https://doi.org/10.21037/cco-22-119

Chidambaranathan-Reghupaty, S., Fisher, P. B., & Sarkar, D. (2021). Hepatocellular carcinoma (HCC): Epidemiology, etiology and molecular classification. *Adv Cancer Res*, *149*, 1-61. https://doi.org/10.1016/bs.acr.2020.10.001

Cho, E. J., Chung, G. E., Yoo, J. J., Cho, Y., Shin, D. W., Kim, Y. J., Yoon, J. H., Han, K., & Yu, S. J. (2023). The association between alcohol consumption and the risk of hepatocellular carcinoma according to glycemic status in Korea: A nationwide population-based study. *PLoS Med*, *20*(6), e1004244. https://doi.org/10.1371/journal.pmed.1004244

Chrysavgis, L., Giannakodimos, I., Diamantopoulou, P., & Cholongitas, E. (2022). Non-alcoholic fatty liver disease and hepatocellular carcinoma: Clinical challenges of an intriguing link. *World J Gastroenterol*, *28*(3), 310-331. https://doi.org/10.3748/wjg.v28.i3.310

Chuang, Y. C., Tsai, K. N., & Ou, J. J. (2022). Pathogenicity and virulence of Hepatitis B virus. *Virulence*, *13*(1), 258-296. https://doi.org/10.1080/21505594.2022.2028483

Cicenas, J., Zalyte, E., Bairoch, A., & Gaudet, P. (2018). Kinases and Cancer. *Cancers (Basel)*, *10*(3). https://doi.org/10.3390/cancers10030063

Çetin Atalay, R. (2015). Molecular biology of liver cancer. *Reviews in Cell Biology and Molecular Medicine*, 206-243.

D'Alessio, A., & Fulgenzi, C. A. M. (2021). Treating patients with advanced hepatocellular carcinoma and impaired liver function: Broadening the reach of anti-cancer therapy. *Liver Cancer International*, *2*(2), 31-32.

D'souza, S., Lau, K. C., Coffin, C. S., & Patel, T. R. (2020). Molecular mechanisms of viral hepatitis induced hepatocellular carcinoma. *World journal of gastroenterology*, *26*(38), 5759.

Daina, A., Michielin, O., & Zoete, V. (2014). iLOGP: a simple, robust, and efficient description of n-octanol/water partition coefficient for drug design using the GB/SA approach. *J Chem Inf Model*, *54*(12), 3284-3301. https://doi.org/10.1021/ci500467k

Daina, A., Michielin, O., & Zoete, V. (2017). SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci Rep*, *7*, 42717. https://doi.org/10.1038/srep42717

Das, S., Bhuyan, R., Goswami, A. M., & Saha, T. (2021). Kinome analyses of Candida albicans, C. parapsilosis and C. tropicalis enable novel kinases as therapeutic drug targets in candidiasis. *Gene*, *780*, 145530. https://doi.org/10.1016/j.gene.2021.145530

Dash, S., Aydin, Y., Widmer, K. E., & Nayak, L. (2020). Hepatocellular carcinoma mechanisms associated with chronic HCV infection and the impact of direct-acting antiviral treatment. *J Hepatocell Carcinoma*, *7*, 45-76. https://doi.org/10.2147/JHC.S221187

Debes, J. D., Romagnoli, P. A., Prieto, J., Arrese, M., Mattos, A. Z., Boonstra, A., & On Behalf Of The Escalon, C. (2021). Serum biomarkers for the prediction of hepatocellular carcinoma. *Cancers (Basel)*, *13*(7). https://doi.org/10.3390/cancers13071681

Deore, A. B., Dhumane, J. R., Wagh, R., & Sonawane, R. (2019). The stages of drug discovery and development process. *Asian Journal of Pharmaceutical Research and Development*, *7*(6), 62-67.

Dhandayuthapani, S., Chidamabaram, A., Thekkumalai, M., & Batra, J. (2022). a narrative review on etiology and mechanism of aflatoxin B1 in the induction of hepatocellular carcinoma. *Journal of Pharmaceutical Negative Results*, 1097-1103.

DiMasi, J. A., Hansen, R. W., & Grabowski, H. G. (2003). The price of innovation: new estimates of drug development costs. *J Health Econ*, *22*(2), 151-185. https://doi.org/10.1016/S0167-6296(02)00126-1

Donne, R., & Lujambio, A. (2023). The liver cancer immune microenvironment: Therapeutic implications for hepatocellular carcinoma. *Hepatology*, *77*(5), 1773-1796. https://doi.org/10.1002/hep.32740

Drefahl, A. (2011). CurlySMILES: A chemical language to customize and annotate encodings of molecular and nanodevice structures [Article]. *Journal of Cheminformatics*, *3*(1), Article 1. https://doi.org/10.1186/1758-2946-3-1

Durmaz, I., Guven, E. B., Ersahin, T., Ozturk, M., Calis, I., & Cetin-Atalay, R. (2016). Liver cancer cells are sensitive to Lanatoside C induced cell death independent of their PTEN status. *Phytomedicine*, *23*(1), 42-51. https://doi.org/10.1016/j.phymed.2015.11.012

Ekins, S., Puhl, A. C., Zorn, K. M., Lane, T. R., Russo, D. P., Klein, J. J., Hickey, A. J., & Clark, A. M. (2019). Exploiting machine learning for end-to-end drug discovery and development. *Nat Mater*, *18*(5), 435-441. https://doi.org/10.1038/s41563-019-0338-z

El-Nashar, H. A. S., El-Labbad, E. M., Al-Azzawi, M. A., & Ashmawy, N. S. (2022). A new xanthone glycoside from Mangifera indica L.: Physicochemical properties and in vitro anti-skin aging activities. *Molecules*, *27*(9). https://doi.org/10.3390/molecules27092609

Emaduldeen, H., Shamran, D. J., Mahdi, D. J., & TA, M. M. (2022). Risk factors of cancer: a review. *Science Archives*, *3 (4)*, 284-288.

Esmail, A., Shieha, O., Galal, A., & Bayoumi, D. (2020). Percutaneous microwave ablation of hepatocellular carcinoma: Proper lesion selection and possible complications. *The Medical Journal of Cairo University*, *88*(December), 2355-2366.

Fan, Z., Wang, S., Xie, Z., & Li, Z. (2022). ADME prediction for Breast Cancer Drugs in Computer-Aided Drug Design. Proceedings of the 11th International Conference on Informatics, Environment, Energy and Applications, P14-18. https://doi.org/10.1145/3533254.3533257

Gaillard, T. (2018). Evaluation of autoDock and autoDock vina on the CASF-2013 benchmark. *J Chem Inf Model*, *58*(8), 1697-1706. https://doi.org/10.1021/acs.jcim.8b00312

Gangopadhyay, A., Ibrahim, R., Theberge, K., May, M., & Houseknecht, K. L. (2022). Non-alcoholic fatty liver disease (NAFLD) and mental illness: Mechanisms linking mood, metabolism and medicines. *Front Neurosci*, *16*, 1042442. https://doi.org/10.3389/fnins.2022.1042442

Garcia-Moreno, A., López-Domínguez, R., Villatoro-García, J. A., Ramirez-Mena, A., Aparicio-Puerta, E., Hackenberg, M., Pascual-Montano, A., & Carmona-Saez, P. (2022). Functional enrichment analysis of regulatory elements. *Biomedicines*, *10*(3), 590. https://doi.org/10.3390/biomedicines10030590

Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., Light, Y., McGlinchey, S., Michalovich, D., Al-Lazikani, B., & Overington, J. P. (2012). ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res*, *40*(Database issue), D1100-1107. https://doi.org/10.1093/nar/gkr777

Ghavimi, S., Apfel, T., Azimi, H., Persaud, A., & Pyrsopoulos, N. T. (2020). Management and treatment of hepatocellular carcinoma with immunotherapy: a review of current and future options. *J Clin Transl Hepatol*, *8*(2), 168-176. https://doi.org/10.14218/JCTH.2020.00001

Gilliland, D. G., & Griffin, J. D. (2002). The roles of FLT3 in hematopoiesis and leukemia. *Blood*, *100*(5), 1532-1542. https://doi.org/10.1182/blood-2002-02-0492

Giustini, A. B., Ioannou, G. N., Sirlin, C., & Loomba, R. (2023). Available modalities for screening and imaging diagnosis of hepatocellular carcinoma—Current gaps and challenges. *Alimentary pharmacology & therapeutics*, *57*(10), 1056.

Glantzounis, G. K., Karampa, A., Peristeri, D. V., Pappas-Gogos, G., Tepelenis, K., Tzimas, P., & Cyrochristos, D. J. (2021). Recent advances in the surgical management of hepatocellular carcinoma. *Ann Gastroenterol*, *34*(4), 453-465. https://doi.org/10.20524/aog.2021.0632

Glassberg, M. B., Ghosh, S., Clymer, J. W., Qadeer, R. A., Ferko, N. C., Sadeghirad, B., Wright, G. W., & Amaral, J. F. (2019). Microwave ablation compared with radiofrequency ablation for treatment of hepatocellular carcinoma and liver metastases: a systematic review and meta-analysis. *Onco Targets Ther*, *12*, 6407-6438. https://doi.org/10.2147/OTT.S204340

Glory S.J., Devi, P. D., & Ezhumalai, P. (2022). *Impact of Machine and Deep Learning in Drug Repurposing* 2022 5th International Conference on Contemporary Computing and Informatics (IC3I),

Golfieri, R., Bargellini, I., Spreafico, C., & Trevisani, F. (2019). Patients with barcelona clinic liver cancer stages B and C hepatocellular carcinoma: Time for a Subclassification. *Liver Cancer*, *8*(2), 78-91. https://doi.org/10.1159/000489791

Gosnell, J. M., Golovko, G., Arroyave, E., Moghe, A., Kueht, M. L., Saldarriaga, O. A., McKinney, K. H., Stevenson, H. L., & Ferguson, M. R. (2024). Disparate outcomes in Hispanic patients with metabolic dysfunction-associated steatotic liver disease/steatohepatitis and type 2 diabetes: Large cohort study. *World J Diabetes*, *15*(5), 886-897. https://doi.org/10.4239/wjd.v15.i5.886

Gowhari Shabgah, A., Ezzatifar, F., Aravindhan, S., Olegovna Zekiy, A., Ahmadi, M., Gheibihayat, S. M., & Gholizadeh Navashenaq, J. (2021). Shedding more light on the role of Midkine in hepatocellular carcinoma: New perspectives on diagnosis and therapy. *IUBMB Life*, *73*(4), 659-669. https://doi.org/10.1002/iub.2458

Guo, K., Tang, W., Zhuo, H., & Zhao, G. (2019). Recent Advance of Akt Inhibitors in Clinical Trials. *ChemistrySelect*, *4*(31), 9040-9044. https://doi.org/10.1002/slct.201901293

Hamid, A. S., Tesfamariam, I. G., Zhang, Y., & Zhang, Z. G. (2013). Aflatoxin B1-induced hepatocellular carcinoma in developing countries: Geographical distribution, mechanism of action and prevention. *Oncol Lett*, *5*(4), 1087-1092. https://doi.org/10.3892/ol.2013.1169

Henning, N., Kannigadu, C., Aucamp, J., van Rensburg, H. D. J., & David, D. D. (2023). Probing benzothiadiazine-1, 1-dioxide ethylene glycol derivatives against Leishmania: synthesis and in vitro efficacy evaluation.

Huang, D. Q., Mathurin, P., Cortez-Pinto, H., & Loomba, R. (2023). Global epidemiology of alcohol-associated cirrhosis and HCC: trends, projections and risk factors. *Nat Rev Gastroenterol Hepatol*, *20*(1), 37-49. https://doi.org/10.1038/s41575-022-00688-6

Huang, J., Lok, V., Ngai, C. H., Chu, C., Patel, H. K., Thoguluva Chandraseka, V., Zhang, L., Chen, P., Wang, S., Lao, X. Q., Tse, L. A., Xu, W., Zheng, Z. J., & Wong, M. C. S. (2021). Disease burden, risk factors, and recent trends of liver cancer: A global country-level analysis. *Liver Cancer*, *10*(4), 330-345. https://doi.org/10.1159/000515304

Ince, V., Gundogan, E., Tolan, K., Kayaalp, C., & Yilmaz, S. (2020). Resection of adrenal metastasis invading left renal vein following living donor liver transplantation for hepatocellular carcinoma. *J Gastrointest Cancer*, *51*(1), 311-313. https://doi.org/10.1007/s12029-019-00252-7

Issa, N. T., Stathias, V., Schurer, S., & Dakshanamurthy, S. (2021). Machine and deep learning approaches for cancer drug repurposing. *Semin Cancer Biol*, *68*, 132-142. https://doi.org/10.1016/j.semcancer.2019.12.011

Ivanov, A. V., Valuev-Elliston, V. T., Tyurina, D. A., Ivanova, O. N., Kochetkov, S. N., Bartosch, B., & Isaguliants, M. G. (2017). Oxidative stress, a trigger of hepatitis C and B virus-induced liver carcinogenesis [Article]. *Oncotarget*, *8*(3), 3895-3932. https://doi.org/10.18632/oncotarget.13904

Iwaloye, O., Elekofehinti, O. O., Olawale, F., Chukwuemeka, P. O., Kikiowo, B., & Folorunso, I. M. (2023). Fragment-based drug design, 2D-QSAR and DFT calculation: Scaffolds of 1, 2, 4, triazolo [1, 5-a] pyrimidin-7-amines as potential inhibitors of Plasmodium falciparum dihydroorotate dehydrogenase. *Letters in Drug Design & Discovery*, *20*(3), 317-334. https://doi.org/10.2174/1570180819666220422120707

Jacob, R., Prince, D. S., Kench, C., & Liu, K. (2023). Alcohol and its associated liver carcinogenesis. *J Gastroenterol Hepatol*, *38*(8), 1211-1217. https://doi.org/10.1111/jgh.16248

Janssen, A. P. A., Grimm, S. H., Wijdeven, R. H. M., Lenselink, E. B., Neefjes, J., van Boeckel, C. A. A., van Westen, G. J. P., & van der Stelt, M. (2019). Drug discovery maps, a machine learning model that visualizes and predicts kinome-inhibitor interaction landscapes. *J Chem Inf Model*, *59*(3), 1221-1229. https://doi.org/10.1021/acs.jcim.8b00640

Jazieh, A.-R., Kim, A., Cong, Z., & Jones, W. (2023). Prevalence of cancer-related risk factors and associated cancer burden: A targeted literature review. In: American Society of Clinical Oncology.

Jia, S. W., Fu, S., Wang, F., Shao, Q., Huang, H. B., & Shao, J. Y. (2014). ALK gene copy number gain and its clinical significance in hepatocellular carcinoma. *World J Gastroenterol*, *20*(1), 183-192. https://doi.org/10.3748/wjg.v20.i1.183

Jiang, Y., Han, Q., Zhao, H., & Zhang, J. (2021). The mechanisms of HBV-induced hepatocellular carcinoma. *J Hepatocell Carcinoma*, *8*, 435-450. https://doi.org/10.2147/JHC.S307962

Jin, B., Wang, W., Du, G., Huang, G. Z., Han, L. T., Tang, Z. Y., Fan, D. G., Li, J., & Zhang, S. Z. (2015). Identifying hub genes and dysregulated pathways in hepatocellular carcinoma [Article]. *European Review for Medical and Pharmacological Sciences*, *19*(4), 592-601.

https://www.scopus.com/inward/record.uri?eid=2-s2.0-84942256203&partnerID=40&md5=59e03940cfd43be29504c292cafe8fb5

Jovanović, M., Nikolic, K., Gagic, Z., & Agbaba, D. (2018). Molecular modeling and analysis of the 3d pharmacophore structure of the selective pi3k-α inhibitors as antitumor agents [Article]. *Arhiv za Farmaciju*, *68*(4), 860-873. https://doi.org/10.5937/ARHFARM1804860J

Juanola, O., Martinez-Lopez, S., Frances, R., & Gomez-Hurtado, I. (2021). Non-alcoholic fatty liver disease: Metabolic, genetic, epigenetic and environmental risk factors. *Int J Environ Res Public Health*, *18*(10). https://doi.org/10.3390/ijerph18105227

Kale, A. (2020). Hepatocellular carcinoma: diagnosis and surveillance. *Hepatocellular Carcinoma-Challenges and Opportunities of a Multidisciplinary Approach*.

Kaur, S. P., Talat, A., Karimi-Sari, H., Grees, A., Chen, H. W., Lau, D. T. Y., & Catana, A. M. (2022). Hepatocellular carcinoma in hepatitis B virus-infected patients and the role of hepatitis B surface antigen (HBsAg). *J Clin Med*, *11*(4). https://doi.org/10.3390/jcm11041126

Kawano, Y., Kaneya, Y., Aoki, Y., Yoshioka, M., Matsushita, A., Shimizu, T., Ueda, J., Takata, H., Taniai, N., Kanda, T., Hirakata, A., Suzuki, H., & Yoshida, H. (2022). Medical treatment for hepatocellular carcinoma in Japan. *J Nippon Med Sch*, *89*(2), 154-160. https://doi.org/10.1272/jnms.JNMS.2022_89-224

Kew, M. C. (2013). Aflatoxins as a cause of hepatocellular carcinoma. *Journal of Gastrointestinal & Liver Diseases*, *22*(3).

Khachatryan, H., Olszowy, B., Barrero, C. A., Gordon, J., & Perez-Leal, O. (2023). Identification of inhibitors of tubulin polymerization using a crispr-edited cell line with endogenous fluorescent tagging of beta-tubulin and histone H1. *Biomolecules*, *13*(2). https://doi.org/10.3390/biom13020249

Kim, D. Y. (2024). Changing etiology and epidemiology of hepatocellular carcinoma: Asia and worldwide. *J Liver Cancer*, *24*(1), 62-70. https://doi.org/10.17998/jlc.2024.03.13

Koscielny, G., An, P., Carvalho-Silva, D., Cham, J. A., Fumis, L., Gasparyan, R., Hasan, S., Karamanis, N., Maguire, M., Papa, E., Pierleoni, A., Pignatelli, M., Platt, T., Rowland, F., Wankar, P., Bento, A. P., Burdett, T., Fabregat, A., Forbes, S., Dunham, I. (2017). Open Targets: a platform for therapeutic target identification and validation. *Nucleic Acids Res*, *45*(D1), D985-D994. https://doi.org/10.1093/nar/gkw1055

Kow, A. W. C. (2019). Transplantation versus liver resection in patients with hepatocellular carcinoma. *Transl Gastroenterol Hepatol*, *4*, 33. https://doi.org/10.21037/tgh.2019.05.06

Krause, D. S., & Van Etten, R. A. (2005). Tyrosine kinases as targets for cancer therapy. *New England Journal of Medicine*, *353*(2), 172-187.

Krook, M. A., Reeser, J. W., Ernst, G., Barker, H., Wilberding, M., Li, G., Chen, H. Z., & Roychowdhury, S. (2021). Fibroblast growth factor receptors in cancer: genetic alterations, diagnostics, therapeutic targets and mechanisms of resistance. *Br J Cancer*, *124*(5), 880-892. https://doi.org/10.1038/s41416-020-01157-0

Kumar, R., Harilal, S., Gupta, S. V., Jose, J., Thomas Parambi, D. G., Uddin, M. S., Shah, M. A., & Mathew, B. (2019). Exploring the new horizons of drug repurposing: A vital tool for turning hard work into smart work. *Eur J Med Chem*, *182*, 111602. https://doi.org/10.1016/j.ejmech.2019.111602

Kumar, S., & Kumar, S. (2019). Molecular docking: a structure-based approach for drug repurposing. In In Silico Drug Design (pp. 161-189). Elsevier. https://doi.org/10.1016/B978-0-12-816125-8.00006-7

Lampimukhi, M., Qassim, T., Venu, R., Pakhala, N., Mylavarapu, S., Perera, T., Sathar, B. S., & Nair, A. (2023). A review of incidence and related risk factors in the development of hepatocellular carcinoma. *Cureus*, *15*(11), e49429. https://doi.org/10.7759/cureus.49429

Li, Y. T., Wu, H. L., & Liu, C. J. (2021). Molecular mechanisms and animal models of hbv-related hepatocellular carcinoma: With emphasis on metastatic tumor antigen 1. *Int J Mol Sci*, *22*(17). https://doi.org/10.3390/ijms22179380

Liu, C. Y., Chen, K. F., & Chen, P. J. (2015). Treatment of liver cancer. *Cold Spring Harb Perspect Med*, *5*(9), a021535. https://doi.org/10.1101/cshperspect.a021535

Llovet, J. M., De Baere, T., Kulik, L., Haber, P. K., Greten, T. F., Meyer, T., & Lencioni, R. (2021). Locoregional therapies in the era of molecular and immune treatments for hepatocellular carcinoma. *Nat Rev Gastroenterol Hepatol*, *18*(5), 293-313. https://doi.org/10.1038/s41575-020-00395-0

Llovet, J. M., Kelley, R. K., Villanueva, A., Singal, A. G., Pikarsky, E., Roayaie, S., Lencioni, R., Koike, K., Zucman-Rossi, J., & Finn, R. S. (2021). Hepatocellular carcinoma. *Nat Rev Dis Primers*, *7*(1), 6. https://doi.org/10.1038/s41572-020-00240-3

Low, Z. Y., Farouk, I. A., & Lal, S. K. (2020). Drug repositioning: New approaches and future prospects for life-debilitating diseases and the COVID-19 pandemic outbreak. *Viruses*, *12*(9). https://doi.org/10.3390/v12091058

Luerken, L., Haimerl, M., Doppler, M., Uller, W., Beyer, L. P., Stroszczynski, C., & Einspieler, I. (2022). Update on percutaneous local ablative procedures for the treatment of hepatocellular carcinoma. *Rofo*, *194*(10), 1075-1086. https://doi.org/10.1055/a-1768-0954 (Aktueller Stand zu perkutanen lokalablativen Verfahren beim hepatozellulären Karzinom.)

Lurje, I., Czigany, Z., Bednarsch, J., Roderburg, C., Isfort, P., Neumann, U. P., & Lurje, G. (2019). Treatment strategies for hepatocellular carcinoma (-) a multidisciplinary approach. *Int J Mol Sci*, *20*(6). https://doi.org/10.3390/ijms20061465

Ma, J., Liu, Y., Guo, Y., Ma, Q., Ji, C., & Zhao, L. (2021). Transcriptional profiling of aflatoxin b1-induced oxidative stress and inflammatory response in macrophages. *Toxins (Basel)*, *13*(6). https://doi.org/10.3390/toxins13060401

Makary, M. S., Khandpur, U., Cloyd, J. M., Mumtaz, K., & Dowell, J. D. (2020). Locoregional therapy approaches for hepatocellular carcinoma: recent advances and management strategies. *Cancers (Basel)*, *12*(7). https://doi.org/10.3390/cancers12071914

Manea, I., Iacob, R., Iacob, S., Cerban, R., Dima, S., Oniscu, G., Popescu, I., & Gheorghe, L. (2023). Liquid biopsy for early detection of hepatocellular carcinoma. *Front Med (Lausanne)*, *10*, 1218705. https://doi.org/10.3389/fmed.2023.1218705

Manieri, E., Herrera-Melle, L., Mora, A., Tomas-Loba, A., Leiva-Vega, L., Fernandez, D. I., Rodriguez, E., Moran, L., Hernandez-Cosido, L., Torres, J. L., Seoane, L. M., Cubero, F. J., Marcos, M., & Sabio, G. (2019). Adiponectin accounts for gender differences in hepatocellular carcinoma incidence. *J Exp Med*, *216*(5), 1108-1119. https://doi.org/10.1084/jem.20181288

Marquardt, J. U., Saborowski, A., Czauderna, C., & Vogel, A. (2019). The changing landscape of systemic treatment of advanced hepatocellular carcinoma: New targeted agents and immunotherapies. *Target Oncol*, *14*(2), 115-123. https://doi.org/10.1007/s11523-019-00624-w

Matsushita, H., & Takaki, A. (2019). Alcohol and hepatocellular carcinoma. *BMJ Open Gastroenterol*, *6*(1), e000260. https://doi.org/10.1136/bmjgast-2018-000260

Meng, M., Li, W., Yang, X., Huang, G., Wei, Z., Ni, Y., Han, X., Wang, J., & Ye, X. (2020). Transarterial chemoembolization, ablation, tyrosine kinase inhibitors, and immunotherapy (TATI): A novel treatment for patients with advanced hepatocellular carcinoma. *J Cancer Res Ther*, *16*(2), 327-334. https://doi.org/10.4103/jcrt.JCRT_101_20

Michelotti, A., de Scordilli, M., Palmero, L., Guardascione, M., Masala, M., Roncato, R., Foltran, L., Ongaro, E., & Puglisi, F. (2021). NAFLD-related hepatocarcinoma: the malignant side of metabolic syndrome. *Cells*, *10*(8). https://doi.org/10.3390/cells10082034

Min, L., Fink-Gremmels, J., Li, D., Tong, X., Tang, J., Nan, X., Yu, Z., Chen, W., & Wang, G. (2021). An overview of aflatoxin B1 biotransformation and aflatoxin M1 secretion in lactating dairy cows. *Anim Nutr*, *7*(1), 42-48. https://doi.org/10.1016/j.aninu.2020.11.002

Moharana, M., Maharana, P. C., Pattanayak, S. K., & Khan, F. (2024). Effect of temperature on hepatitis a virus and exploration of binding mode mechanism

of phytochemicals from tinospora cordifolia: an insight into molecular docking, MM/GBSA, and molecular dynamics simulation study. *J Biomol Struct Dyn*, *42*(2), 598-614. https://doi.org/10.1080/07391102.2023.2194429

Montalto, G., Cervello, M., Giannitrapani, L., Dantona, F., Terranova, A., & Castagnetta, L. A. (2002). Epidemiology, risk factors, and natural history of hepatocellular carcinoma. *Ann N Y Acad Sci*, *963*, 13-20. https://doi.org/10.1111/j.1749-6632.2002.tb04090.x

Moret, M., Grisoni, F., Brunner, C., & Schneider, G. (2021). Leveraging molecular structure and bioactivity with chemical language models for drug design.

Moriguchi, I., Hirono, S., Liu, Q., NAKAGOME, I., & MATSUSHITA, Y. (1992). Simple method of calculating octanol/water partition coefficient. *Chemical and pharmaceutical bulletin*, *40*(1), 127-130.

Moriguchi, I., Hirono, S., Nakagome, I., & Hirano, H. (1994). Comparison of reliability of log P values for drugs calculated by several methods. *Chemical and pharmaceutical bulletin*, *42*(4), 976-978.

Muller, M., Bird, T. G., & Nault, J. C. (2020). The landscape of gene mutations in cirrhosis and hepatocellular carcinoma. *J Hepatol*, *72*(5), 990-1002. https://doi.org/10.1016/j.jhep.2020.01.019

Mungamuri, S. K., & Mavuduru, V. A. (2020). Role of epigenetic alterations in aflatoxin-induced hepatocellular carcinoma. *Liver Cancer International*, *1*(2), 41-50. https://doi.org/10.1002/lci2.20

Murohashi, M., Hinohara, K., Kuroda, M., Isagawa, T., Tsuji, S., Kobayashi, S., Umezawa, K., Tojo, A., Aburatani, H., & Gotoh, N. (2010). Gene set enrichment analysis provides insight into novel signalling pathways in breast cancer stem cells. British journal of cancer, 102(1), 206-212. https:// doi:10.1038/sj.bjc.6605468

Nadarevic, T., Giljaca, V., Colli, A., Fraquelli, M., Casazza, G., Miletic, D., & Stimac, D. (2021). Computed tomography for the diagnosis of hepatocellular

carcinoma in adults with chronic liver disease. *Cochrane Database Syst Rev*, *10*(10), CD013362. https://doi.org/10.1002/14651858.CD013362.pub2

Ng, Y. L., Salim, C. K., & Chu, J. J. H. (2021). Drug repurposing for COVID-19: Approaches, challenges and promising candidates. *Pharmacol Ther*, *228*, 107930. https://doi.org/10.1016/j.pharmthera.2021.107930

O'Leary, C., Mahler, M., & Soulen, M. C. (2021). Liver-directed therapy for hepatocellular carcinoma. *Chin Clin Oncol*, *10*(1), 8. https://doi.org/10.21037/cco-20-51

Ochoa, D., Hercules, A., Carmona, M., Suveges, D., Gonzalez-Uriarte, A., Malangone, C., Miranda, A., Fumis, L., Carvalho-Silva, D., Spitzer, M., Baker, J., Ferrer, J., Raies, A., Razuvayevskaya, O., Faulconbridge, A., Petsalaki, E., Mutowo, P., Machlitt-Northen, S., Peat, G., McDonagh, E. M. (2021). Open Targets Platform: supporting systematic drug-target identification and prioritisation. *Nucleic Acids Res*, *49*(D1), D1302-D1310. https://doi.org/10.1093/nar/gkaa1027

Pan, X., Lin, X., Cao, D., Zeng, X., Yu, P. S., He, L., Nussinov, R., & Cheng, F. (2022). Deep learning for drug repurposing: Methods, databases, and applications. *WIREs Computational Molecular Science*, *12*(4). https://doi.org/10.1002/wcms.1597

Parra, N. S., Ross, H. M., Khan, A., Wu, M., Goldberg, R., Shah, L., Mukhtar, S., Beiriger, J., Gerber, A., & Halegoua-DeMarzio, D. (2023). Advancements in the diagnosis of hepatocellular carcinoma. *International Journal of Translational Medicine*, *3*(1), 51-65. https://doi.org/10.3390/ijtm3010005

Ploumakis, A., & Coleman, M. L. (2015). OH, the places you'll go! Hydroxylation, gene expression, and cancer. *Mol Cell*, *58*(5), 729-741. https://doi.org/10.1016/j.molcel.2015.05.026

Pinzi, L., & Rastelli, G. (2019). Molecular docking: shifting paradigms in drug discovery. *International journal of molecular sciences*, *20*(18), 4331. https://doi.org/10.3390/ijms20184331

Pouwels, S., Sakran, N., Graham, Y., Leal, A., Pintar, T., Yang, W., Kassir, R., Singhal, R., Mahawar, K., & Ramnarain, D. (2022). Non-alcoholic fatty liver disease (NAFLD): a review of pathophysiology, clinical management and effects of weight loss. *BMC Endocr Disord*, *22*(1), 63. https://doi.org/10.1186/s12902-022-00980-1

Powell, E. E., Wong, V. W., & Rinella, M. (2021). Non-alcoholic fatty liver disease. *Lancet*, *397*(10290), 2212-2224. https://doi.org/10.1016/S0140-6736(20)32511-3

Pushpakom, S., Iorio, F., Eyers, P. A., Escott, K. J., Hopper, S., Wells, A., Doig, A., Guilliams, T., Latimer, J., & McNamee, C. (2019). Drug repurposing: progress, challenges and recommendations. *Nature reviews Drug discovery*, *18*(1), 41-58. https:// doi:10.1038/nrd.2018.168

Quiros, M., Grazulis, S., Girdzijauskaite, S., Merkys, A., & Vaitkus, A. (2018). Using SMILES strings for the description of chemical connectivity in the Crystallography Open Database. *J Cheminform*, *10*(1), 23. https://doi.org/10.1186/s13321-018-0279-6

Ranjith, D., & Ravikumar, C. (2019). SwissADME predictions of pharmacokinetics and drug-likeness properties of small molecules present in Ipomoea mauritiana Jacq. *Journal of Pharmacognosy and Phytochemistry*, *8*(5), 2063-2073.

Rascio, F., Spadaccino, F., Rocchetti, M. T., Castellano, G., Stallone, G., Netti, G. S., & Ranieri, E. (2021). The pathogenic role of PI3K/AKT pathway in cancer onset and drug resistance: An updated review. *Cancers (Basel)*, *13*(16). https://doi.org/10.3390/cancers13163949

Rennane, S., Baker, L., & Mulcahy, A. (2021). Estimating the cost of industry investment in drug research and development: A review of methods and results. *Inquiry*, *58*, 469580211059731. https://doi.org/10.1177/00469580211059731

Revathidevi, S., & Munirajan, A. K. (2019). Akt in cancer: Mediator and more. *Semin Cancer Biol*, *59*, 80-91. https://doi.org/10.1016/j.semcancer.2019.06.002

Rifaioğlu, A. S. (2020). Deep learning for prediction of drug-target interaction space and protein functions. Doctoral dissertation, Middle East Technical University, Ankara.

Rifaioglu, A. S., Cetin Atalay, R., Cansen Kahraman, D., Dogan, T., Martin, M., & Atalay, V. (2021). MDeePred: novel multi-channel protein featurization for deep learning-based binding affinity prediction in drug discovery. *Bioinformatics*, *37*(5), 693-704. https://doi.org/10.1093/bioinformatics/btaa858

Robertson, K. D. (2001). DNA methylation, methyltransferases, and cancer. *Oncogene*, *20*(24), 3139-3155.

Rognan, D. (2017). The impact of in silico screening in the discovery of novel and safer drug candidates. *Pharmacology & therapeutics*, *175*, 47-66. https://doi.org/10.1016/j.pharmthera.2017.02.034

Roudot-Thoraval, F. (2021). Epidemiology of hepatitis C virus infection. *Clin Res Hepatol Gastroenterol*, *45*(3), 101596. https://doi.org/10.1016/j.clinre.2020.101596

Sadagopan, N., & He, A. R. (2024). Recent progress in systemic therapy for advanced hepatocellular carcinoma. *Int J Mol Sci*, *25*(2). https://doi.org/10.3390/ijms25021259

Samy, A. M., Kandeil, M. A., Sabry, D., Abdel-Ghany, A. A., & Mahmoud, M. O. (2024). From NAFLD to NASH: Understanding the spectrum of non-alcoholic liver diseases and their consequences. *Heliyon*, *10*(9), e30387. https://doi.org/10.1016/j.heliyon.2024.e30387

Sawicki, T., Ruszkowska, M., Danielewicz, A., Niedzwiedzka, E., Arlukowicz, T., & Przybylowicz, K. E. (2021). A Review of colorectal cancer in terms of epidemiology, risk factors, development, symptoms and diagnosis. *Cancers (Basel)*, *13*(9). https://doi.org/10.3390/cancers13092025

Schultheis, A. M., Bos, M., Schmitz, K., Wilsberg, L., Binot, E., Wolf, J., Buttner, R., & Schildhaus, H. U. (2014). Fibroblast growth factor receptor 1 (FGFR1)

amplification is a potential therapeutic target in small-cell lung cancer. *Mod Pathol*, *27*(2), 214-221. https://doi.org/10.1038/modpathol.2013.141

Shaker, B., Ahmad, S., Lee, J., Jung, C., & Na, D. (2021). In silico methods and tools for drug discovery. *Comput Biol Med*, *137*, 104851. https://doi.org/10.1016/j.compbiomed.2021.104851

Sherman, M. (2005). Hepatocellular carcinoma: epidemiology, risk factors, and screening. Seminars in liver disease,

Shaikh, N., Linthoi, R., Swamy, K., Karthikeyan, M., & Vyas, R. (2023). Comprehensive molecular docking and dynamic simulations for drug repurposing of clinical drugs against multiple cancer kinase targets. *Journal of Biomolecular Structure and Dynamics*, *41*(16), 7735-7743. https://doi.org/10.1080/07391102.2022.2124453

Sidharthan, S., & Kottilil, S. (2014). Mechanisms of alcohol-induced hepatocellular carcinoma. *Hepatol Int*, *8*(2), 452-457. https://doi.org/10.1007/s12072-013-9494-4

Smirnov, P., Kofia, V., Maru, A., Freeman, M., Ho, C., El-Hachem, N., Adam, G. A., Ba-Alawi, W., Safikhani, Z., & Haibe-Kains, B. (2018). PharmacoDB: an integrative database for mining in vitro anticancer drug screening studies. *Nucleic Acids Res*, *46*(D1), D994-D1002. https://doi.org/10.1093/nar/gkx911

Song, C., Yang, J., Wang, Y., Ding, G., Guo, L., & Qin, J. (2024). Mechanisms and transformed products of aflatoxin B1 degradation under multiple treatments: a review. *Crit Rev Food Sci Nutr*, *64*(8), 2263-2275. https://doi.org/10.1080/10408398.2022.2121910

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., & Lander, E. S. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, *102*(43), 15545-15550. https://doi.org/10.1073/pnas.0506580102

Subramaniyan, V., Chakravarthi, S., Jegasothy, R., Seng, W. Y., Fuloria, N. K., Fuloria, S., Hazarika, I., & Das, A. (2021). Alcohol-associated liver disease: A review on its pathophysiology, diagnosis and drug therapy. *Toxicol Rep*, *8*, 376-385. https://doi.org/10.1016/j.toxrep.2021.02.010

Sugawara, Y., & Hibi, T. (2021). Surgical treatment of hepatocellular carcinoma. *Biosci Trends*, *15*(3), 138-141. https://doi.org/10.5582/bst.2021.01094

Sun, D., Gao, W., Hu, H., & Zhou, S. (2022). Why 90% of clinical drug development fails and how to improve it? *Acta Pharm Sin B*, *12*(7), 3049-3062. https://doi.org/10.1016/j.apsb.2022.02.002

Sun, E. J., Wankell, M., Palamuthusingam, P., McFarlane, C., & Hebbard, L. (2021). Targeting the PI3K/Akt/mTOR pathway in hepatocellular carcinoma. *Biomedicines*, *9*(11). https://doi.org/10.3390/biomedicines9111639

Sun, J., Qi, C., Liu, Y., Gao, F., Fu, X., & Tian, Y. (2024). Evaluation of multiple liver cancer scoring systems. *Adv Biol (Weinh)*, *8*(2), e2300301. https://doi.org/10.1002/adbi.202300301

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca Cancer J Clin*, *71*(3), 209-249. https://doi.org/10.3322/caac.21660

Şahin, S., & Dege, N. (2022). (E)-N-(3-chlorophenyl)-1-(5-nitro-2-(piperidin-1-yl)phenyl)methanimine: X-Ray, DFT, ADMET, boiled-egg model, druggability, bioavailabilty, and human cyclophilin D (CypD) inhibitory activity. *Journal of Molecular Structure*, *1250*. https://doi.org/10.1016/j.molstruc.2021.131744

Taniai, M. (2020). Alcohol and hepatocarcinogenesis. *Clin Mol Hepatol*, *26*(4), 736-741. https://doi.org/10.3350/cmh.2020.0203

Tanoli, Z., Seemab, U., Scherer, A., Wennerberg, K., Tang, J., & Vaha-Koskela, M. (2021). Exploration of databases and methods supporting drug repurposing: a

comprehensive survey. *Brief Bioinform*, *22*(2), 1656-1678. https://doi.org/10.1093/bib/bbaa003

Tanoli, Z., Vaha-Koskela, M., & Aittokallio, T. (2021). Artificial intelligence, machine learning, and drug repurposing in cancer. *Expert Opin Drug Discov*, *16*(9), 977-989. https://doi.org/10.1080/17460441.2021.1883585

Taylor, A. C., Maddirela, D., & White, S. B. (2019). Role of radioembolization for biliary tract and primary liver cancer. *Surg Oncol Clin N Am*, *28*(4), 731-743. https://doi.org/10.1016/j.soc.2019.07.001

Terstappen, G. C., & Reggiani, A. (2001). In silico research in drug discovery. *Trends in pharmacological sciences*, *22*(1), 23-26. https://doi.org/10.1016/S0165-6147(00)01584-4

Tian, S., Chen, Y., Zhang, Y., & Xu, X. (2023). Clinical value of serum AFP and PIVKA-II for diagnosis, treatment and prognosis of hepatocellular carcinoma. *J Clin Lab Anal*, *37*(1), e24823. https://doi.org/10.1002/jcla.24823

Tilford, C. A., & Siemers, N. O. (2009). Gene set enrichment analysis. *Protein networks and pathway analysis*, 99-121.

Tinivella, A., Pinzi, L., Gambacorta, G., Baxendale, I., & Rastelli, G. (2022). Identification of potential biological targets of oxindole scaffolds via in silico repositioning strategies. *F1000Res*, *11*. https://doi.org/10.12688/f1000research.109017.2

Torimura, T., & Iwamoto, H. (2021). Optimizing the management of intermediate-stage hepatocellular carcinoma: Current trends and prospects. *Clin Mol Hepatol*, *27*(2), 236-245. https://doi.org/10.3350/cmh.2020.0204

UniProt, C. (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res*, *47*(D1), D506-D515. https://doi.org/10.1093/nar/gky1049

Urbina, F., Puhl, A. C., & Ekins, S. (2021). Recent advances in drug repurposing using machine learning. *Curr Opin Chem Biol*, *65*, 74-84. https://doi.org/10.1016/j.cbpa.2021.06.001

URL$_1$: Liver (Hepatocellular) Cancer Screening (PDQ®)–Health Professional Version. (2024). Access link https://www.cancer.gov/types/liver/hp/liver-screening-pdq#:~:text=The%20rationale%20for%20screening%20for,HCC%20have%20previously%20undiagnosed%20cirrhosis.

Verma, S., Vaghela, K., & Zaveri, M. (2022). Regulatory perspective of drug repurposing: Methods, regulatory pathways and hurdles. *International Journal of Drug Regulatory Affairs*, *10*(2), 39-45. https://doi.org/10.22270/ijdra.v10i2.520

Verret, B., Cortes, J., Bachelot, T., Andre, F., & Arnedos, M. (2019). Efficacy of PI3K inhibitors in advanced breast cancer. *Annals of Oncology*, *30*, x12-x20. https://doi.org/10.1093/annonc/mdz381

Vescovo, T., Refolo, G., Vitagliano, G., Fimia, G. M., & Piacentini, M. (2016). Molecular mechanisms of hepatitis C virus-induced hepatocellular carcinoma. *Clin Microbiol Infect*, *22*(10), 853-861. https://doi.org/10.1016/j.cmi.2016.07.019

Villar, S., Ortiz-Cuaran, S., Abedi-Ardekani, B., Gouas, D., Nogueira da Costa, A., Plymoth, A., Khuhaprema, T., Kalalak, A., Sangrajrang, S., Friesen, M. D., Groopman, J. D., & Hainaut, P. (2012). Aflatoxin-induced TP53 R249S mutation in hepatocellular carcinoma in Thailand: association with tumors developing in the absence of liver cirrhosis. *PLoS One*, *7*(6), e37707. https://doi.org/10.1371/journal.pone.0037707

Vique-Sánchez, J. L., & Benítez-Cardoza, C. G. (2022). A potential PIK3CA inhibitor to develop an anticancer drug. *ChemistrySelect*, *7*(27). https://doi.org/10.1002/slct.202202301

Vora, L. K., Gholap, A. D., Jetha, K., Thakur, R. R. S., Solanki, H. K., & Chavda, V. P. (2023). Artificial intelligence in pharmaceutical technology and drug

delivery design. *Pharmaceutics*, *15*(7). https://doi.org/10.3390/pharmaceutics15071916

Walters, W. P., & Barzilay, R. (2021). Applications of deep learning in molecule generation and molecular property prediction. *Acc Chem Res*, *54*(2), 263-270. https://doi.org/10.1021/acs.accounts.0c00699

Wang, H., Gao, C., Li, X., Chen, F., & Li, G. (2024). Camptothecin enhances the anti-tumor effect of low-dose apatinib combined with PD-1 inhibitor on hepatocellular carcinoma. *Sci Rep*, *14*(1), 7140. https://doi.org/10.1038/s41598-024-57874-6

Wang, W., & Wei, C. (2020). Advances in the early diagnosis of hepatocellular carcinoma. *Genes Dis*, *7*(3), 308-319. https://doi.org/10.1016/j.gendis.2020.01.014

Wang, Y., Sun, X., Chen, C., Ge, H., Sun, J., Li, E., Cai, Z., Fu, Q., Sun, X., Wu, J., Ye, M., Cao, W., Chen, Q., Wei, X., Han, X., Sun, K., Yan, Q., Huang, W., Wu, L., Liang, T. (2024). Optimizing hepatocellular carcinoma disease staging systems by incorporating tumor micronecrosis: A multi-institutional retrospective study. *Cancer Lett*, *585*, 216654. https://doi.org/10.1016/j.canlet.2024.216654

Waring, M. J., Arrowsmith, J., Leach, A. R., Leeson, P. D., Mandrell, S., Owen, R. M., Pairaudeau, G., Pennie, W. D., Pickett, S. D., Wang, J., Wallace, O., & Weir, A. (2015). An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nat Rev Drug Discov*, *14*(7), 475-486. https://doi.org/10.1038/nrd4609

Wege, H., Li, J., & Ittrich, H. (2019). Treatment lines in hepatocellular carcinoma. *Visc Med*, *35*(4), 266-272. https://doi.org/10.1159/000501749

Wei, L., & Ploss, A. (2021). Mechanism of hepatitis B virus cccDNA formation. *Viruses*, *13*(8). https://doi.org/10.3390/v13081463

Wildman, S. A., & Crippen, G. M. (1999). Prediction of physicochemical parameters by atomic contributions. *Journal of chemical information and computer sciences*, *39*(5), 868-873.

Wilson, G. C., & Geller, D. A. (2019). Evolving surgical options for hepatocellular carcinoma. *Surg Oncol Clin N Am*, *28*(4), 645-661. https://doi.org/10.1016/j.soc.2019.06.006

Wu, Y., & Wang, G. (2018). Machine learning based toxicity prediction: from chemical structural description to transcriptome analysis. *Int J Mol Sci*, *19*(8). https://doi.org/10.3390/ijms19082358

Xu, Z., Xie, H., Zhou, L., Chen, X., & Zheng, S. (2019). The combination strategy of transarterial chemoembolization and radiofrequency ablation or microwave ablation against hepatocellular carcinoma. *Anal Cell Pathol (Amst)*, *2019*, 8619096. https://doi.org/10.1155/2019/8619096

Yaglioglu, S. A., Gurbuz, D. G., Dolarslan, M., & Demirtas, I. (2022). First determination of anticancer, cytotoxic, and in silico ADME evaluation of secondary metabolites of endemic Astragalus leucothrix Freyn & Bornm. *Turk J Chem*, *46*(1), 169-183. https://doi.org/10.3906/kim-2104-23

Yang, T. H., Chan, C., Yang, P. J., Huang, Y. H., & Lee, M. H. (2023). Genetic susceptibility to hepatocellular carcinoma in patients with chronic hepatitis virus infection. *Viruses*, *15*(2). https://doi.org/10.3390/v15020559

Younossi, Z. M., & Henry, L. (2021). Epidemiology of non-alcoholic fatty liver disease and hepatocellular carcinoma. *JHEP Rep*, *3*(4), 100305. https://doi.org/10.1016/j.jhepr.2021.100305

Yu, W., Ma, Y., Shrivastava, S. K., Srivastava, R. K., & Shankar, S. (2022). Chronic alcohol exposure induces hepatocyte damage by inducing oxidative stress, SATB2 and stem cell-like characteristics, and activating lipogenesis. *J Cell Mol Med*, *26*(7), 2119-2131. https://doi.org/10.1111/jcmm.17235

Zdrazil, B., Felix, E., Hunter, F., Manners, E. J., Blackshaw, J., Corbett, S., de Veij, M., Ioannidis, H., Lopez, D. M., Mosquera, J. F., Magarinos, M. P., Bosc, N.,

Arcila, R., Kiziloren, T., Gaulton, A., Bento, A. P., Adasme, M. F., Monecke, P., Landrum, G. A., & Leach, A. R. (2024). The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Res*, *52*(D1), D1180-D1192. https://doi.org/10.1093/nar/gkad1004

Zhang, C., Peng, L., Zhang, Y., Liu, Z., Li, W., Chen, S., & Li, G. (2017). The identification of key genes and pathways in hepatocellular carcinoma by bioinformatics analysis of high-throughput data. *Med Oncol*, *34*(6), 101. https://doi.org/10.1007/s12032-017-0963-9

Zhang, X., Guan, L., Tian, H., Zeng, Z., Chen, J., Huang, D., Sun, J., Guo, J., Cui, H., & Li, Y. (2021). Risk factors and prevention of viral hepatitis-related hepatocellular carcinoma. *Front Oncol*, *11*, 686962. https://doi.org/10.3389/fonc.2021.686962

Zhang, Z., Zhou, L., Xie, N., Nice, E. C., Zhang, T., Cui, Y., & Huang, C. (2020). Overcoming cancer therapeutic bottleneck by drug repurposing. *Signal Transduct Target Ther*, *5*(1), 113. https://doi.org/10.1038/s41392-020-00213-8

Zhao, Z., Song, J., Zhang, D., Wu, F., Tu, J., & Ji, J. (2021). Oxysophocarpine suppresses FGFR1-overexpressed hepatocellular carcinoma growth and sensitizes the therapeutic effect of lenvatinib. *Life Sci*, *264*, 118642. https://doi.org/10.1016/j.lfs.2020.118642

APPENDIX

SUPPLEMENTARY TABLES

Table S1: All Targets Associated with HCC After Data Type Filtering

Target Name	Target Symbol	Overall Association Scores	Genetic Association Scores	Somatic Mutation Scores	Drug Scores	Pathways & Systems Biology Scores	RNA Expression Scores	Text Mining Scores	Animal Model Scores
tumor protein p53	TP53	1.0	1.0	1.0	0.0	0.7152195485738936	0.0	0.30033021847080205	0.25
catenin beta 1	CTNNB1	1.0	1.0	1.0	0.0	0.6581859646946654	0.0	0.23098058920399783	0.0
axin 1	AXIN1	1.0	0.0	1.0	0.0	1.0	0.0	0.12303580566311544	0.0
MET proto-oncogene, receptor tyrosine kinase	MET	1.0	0.0	0.9915123456790124	1.0	0.0	0.0	0.04230123609121459	0.0
fibroblast growth factor receptor 1	FGFR1	1.0	0.0	0.7986111111111111	1.0	0.0	0.0	0.21770072794453407	0.0
kinase insert domain receptor	KDR	1.0	0.0	0.5059156378600823	1.0	0.25	0.0	0.20383866416785226	0.0
platelet derived growth factor receptor alpha	PDGFRA	1.0	0.0	0.4805555555555557	0.9993859229924646	0.0	0.10823473706565383	0.2626506472001884	0.0
Raf-1 proto-oncogene, serine/threonine kinase	RAF1	1.0	0.0	0.4814814814814815	1.0	0.25	0.0	0.105406028785698	0.0
fms related tyrosine kinase 3	FLT3	1.0	0.0	0.5505952380952381	1.0	0.0	0.0	0.04895905328798186	0.0
platelet derived growth factor receptor beta	PDGFRB	1.0	0.0	0.4819444444444445	1.0	0.0	0.0	0.21368191472270465	0.0
fms related tyrosine kinase 4	FLT4	1.0	0.0	0.5326829805996472	1.0	0.0	0.0	0.09394877194955117	0.0
ret proto-oncogene	RET	1.0	0.0	0.5042438271604939	1.0	0.0	0.02796321498196705	0.11195600901584819	0.0
fibroblast growth factor receptor 2	FGFR2	1.0	0.0	0.4166666666666667	1.0	0.0	0.10296000000000001	0.2279852117105633	0.0
KIT proto-oncogene, receptor tyrosine kinase	KIT	1.0	0.0	0.5326829805996472	1.0	0.0	0.0	0.006	0.0
B-Raf proto-oncogene, serine/threonine kinase	BRAF	1.0	0.0	0.4166666666666667	1.0	0.0	0.009590970835463432	0.14071122072090142	0.0
phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha	PIK3CA	1.0	0.5	1.0	0.1	0.25	0.0	0.14530945936311657	0.0
CD274 molecule	CD274	1.0	0.0	0.2777777777777778	1.0	0.0	0.0	0.31453379905487555	0.0
fibroblast growth factor receptor 4	FGFR4	1.0	0.0	0.2777777777777778	1.0	0.0	0.0	0.2819590336600308	0.0
fibroblast growth factor receptor 3	FGFR3	1.0	0.0	0.2777777777777778	1.0	0.0	0.0	0.10701768123254814	0.0
telomerase reverse transcriptase	TERT	1.0	0.8419444444444443	0.711398898898899	0.0	0.0	0.06114595711989072	0.2998225453076396	0.0
DNA topoisomerase I	TOP1	1.0	0.0	0.1388888888888889	1.0	0.0	0.0	0.155899933106576	0.0
epidermal growth factor receptor	EGFR	0.9889033690370091	0.0	0.48042929292929293	0.8182777351054156	0.25	0.015633249479851088	0.30841182648142323	0.0
nuclear factor, erythroid 2 like 2	NFE2L2	0.9861488324613179	0.0	0.9710778486820153	0.0	0.0	0.0	0.07905584448256438	0.190540000000000001

Table S1: (cont.)

isocitrate dehydrogenase (NADP(+)) 1, cytosolic	IDH1	0.9757839263791365	0.0	0.9635802469135802	0.0	0.0	0.016559116501942018	0.18466103688765695	0.0
HRas proto-oncogene, GTPase	HRAS	0.9655228070697964	0.0	0.947787628425856	0.0	0.0	0.0	0.15961660779546386	0.0
cyclin dependent kinase inhibitor 2A	CDKN2A	0.9643727643583659	0.0	0.9375694656423823	0.0	0.0	0.10561075816007524	0.18182363697 88107	0.0
estrogen receptor 1	ESR1	0.9519781878015823	0.0	0.5033436213 99177	0.75	0.5	0.11194458618510383	0.25774309518 12522	0.0
DNA polymerase epsilon, catalytic subunit	POLE	0.9355827938466809	0.0	0.4802350427 3504275	0.8155240331629202	0.0	0.0	0.0	0.0
CREB binding protein	CREBBP	0.9353246276167473	0.0	0.9284357387278583	0.0	0.0	0.0	0.062	0.0
isocitrate dehydrogenase (NADP(+)) 2, mitochondrial	IDH2	0.9276793075494287	0.0	0.9110339506 17284	0.0	0.0	0.0	0.14980821238930225	0.0
NRAS proto-oncogene, GTPase	NRAS	0.9276289828184364	0.0	0.9172839506 17284	0.0	0.0	0.00809438597841138	0.08855219769751527	0.0
DNA polymerase delta 1, catalytic subunit	POLD1	0.9254042800765004	0.0	0.4166666666666667	0.8155240331629202	0.0	0.0	0.05142222222222223	0.0
AT-rich interaction domain 2	ARID2	0.8975792476262491	0.0	0.8680265222849968	0.0	0.0	0.0	0.26597452807 12706	0.0
colony stimulating factor 1 receptor	CSF1R	0.8958985846956575	0.0	0.3101851851851852	0.8123008832 39736	0.0	0.00521449173 6277451	0.05152949483497103	0.0
GNAS complex locus	GNAS	0.8643253914380926	0.0	0.8603395061728395	0.0	0.0	0.00706305313 2937941	0.03190000000 0000005	0.0
APC regulator of WNT signaling pathway	APC	0.8617515044760831	0.0	0.7446247329 05983	0.0	0.6341215558174885	0.0	0.0	0.0
discoidin domain receptor tyrosine kinase 2	DDR2	0.8548531243600074	0.0	0.4166666666666667	0.725	0.0	0.01534845374 4889225	0.22254461400856643	0.0
ABL proto-oncogene 1, non-receptor tyrosine kinase	ABL1	0.8530055555555555	0.0	0.2777777777777778	0.7673611111 11111	0.0	0.0	0.14579999999999999	0.0
KRAS proto-oncogene, GTPase	KRAS	0.8270290899955751	0.0	0.8159722222222222	0.0	0.0	0.0	0.09951180996017617	0.0
TSC complex subunit 2	TSC2	0.8223728912477103	0.0	0.7777111760883691	0.0	0.0	0.0	0.17864686063736468	0.0
splicing factor 3b subunit 1	SF3B1	0.7874111111111111	0.0	0.7638888888888888	0.0	0.0	0.0	0.2117	0.0
interleukin 6 signal transducer	IL6ST	0.7595346974403909	0.0	0.7105475040257648	0.0	0.0	0.0	0.19594877365 85046	0.0
AT-rich interaction domain 1A	ARID1A	0.7282025652730801	0.0	0.7029912522 61052	0.0	0.0	0.0	0.22690181710825336	0.0
major histocompatibility complex, class II, DQ alpha 1	HLA-DQA1	0.7225721750469736	0.7225721750469736	0.0	0.0	0.0	0.0	0.0	0.0
insulin like growth factor 2 receptor	IGF2R	0.7190700746893237	0.0	0.6944444444444444	0.0	0.0	0.0	0.09850252097951709	0.0
phosphatase and tensin homolog	PTEN	0.7142386544095812	0.0	0.6610289725 20908	0.0	0.2904947189 5720985	0.0	0.31548385758 4629	0.0
retinoid X receptor alpha	RXRA	0.7071105414537384	0.0	0.6944444444444444	0.0	0.0	0.0	0.05066438803717578	0.0
transcription factor 7 like 2	TCF7L2	0.7060830726587126	0.0	0.5305426954 73251	0.0	0.5604923708790421	0.0	0.11659525120221949	0.0
kinesin family member 1B	KIF1B	0.6995011950776868	0.6514500281 20844	0.0	0.0	0.0	0.0	0.19220466782 73713	0.0
CCR4-NOT transcription complex subunit 9	CNOT9	0.6957521220793593	0.0	0.6944444444444444	0.0	0.0	0.00523071053 9659667	0.0	0.0
caspase 8	CASP8	0.6951975105906835	0.0	0.6756365740740741	0.0	0.0	0.0	0.17604842864948486	0.0
splicing factor 3b subunit 2	SF3B2	0.6793944444444445	0.0	0.6770833333333334	0.0	0.0	0.0	0.02080000000 0000003	0.0
neurofibromin 1	NF1	0.6730377888636936	0.0	0.6680555555555555	0.0	0.0	0.0	0.04484009977324263	0.0
activin A receptor type 2A	ACVR2A	0.6696192495126705	0.0	0.6684636939571149	0.0	0.0	0.0	0.01040000000 0000001	0.0

Table S1: (cont.)

lysine methyltransferase 2C	KMT2C	0.6657109658678286	0.0	0.6634331880900508	0.0	0.0	0.0	0.0205	0.0
polybromo 1	PBRM1	0.6632685273368606	0.0	0.6490575396825397	0.0	0.0	0.0	0.12789888888888887	0.0
ATM serine/threonine kinase	ATM	0.6623314873129 14	0.0	0.5592206790123457	0.0	0.3732758936 3974 8	0.0	0.08812651401568204	0.0
ATRX chromatin remodeler	ATRX	0.6556311728395061	0.0	0.6556311728395061	0.0	0.0	0.0	0.0	0.0
HNF1 homeobox A	HNF1A	0.6552025662516333	0.0	0.5939663973414934	0.0	0.0	0.0	0.24494467564055972	0.0
androgen receptor	AR	0.6505240334339668	0.0	0.4805555555 5555557	0.225	0.5	0.0	0.08616231272124636	0.0
AKT serine/threonine kinase 1	AKT1	0.6486487590891844	0.0	0.5555555555555556	0.2	0.0	0.0	0.2834839252456263	0.0
SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4	SMARCA4	0.6472277056260018	0.0	0.5491469478737998	0.0	0.2620701025512506	0.0	0.27584742987491906	0.0
cyclin dependent kinase inhibitor 1A	CDKN1A	0.6356357270112526	0.0	0.3624614197530864	0.0	0.5125359370641707	0.0	0.2923599150792937	0.0
catenin delta 1	CTNND1	0.6337585517736111	0.0	0.3107638888888889	0.0	0.5474765175710412	0.0	0.07731955782312923	0.0
protein kinase cAMP-activated catalytic subunit alpha	PRKACA	0.625	0.0	0.625	0.0	0.0	0.0	0.0	0.0
SET domain containing 2, histone lysine methyltransferase	SETD2	0.6184285314829113	0.0	0.5552083333333334	0.0	0.0	0.0	0.2528807925983116	0.0
DnaJ heat shock protein family (Hsp40) member B1	DNAJB1	0.6139797938061801	0.0	0.5902777777777778	0.0	0.0	0.0061897875384249405	0.09205704742986459	0.0
erb-b2 receptor tyrosine kinase 3	ERBB3	0.6022607315794631	0.0	0.5147707231040565	0.25	0.0	0.0	0.22491007627865964	0.0
TSC complex subunit 1	TSC1	0.5994610138888888	0.0	0.5599392361111111	0.0	0.0	0.0	0.1580871111111111	0.0
hypoxia inducible factor 1 subunit alpha	HIF1A	0.5988293637550938	0.0	0.4814814814814815	0.0	0.3270091424 4957735	0.0	0.3203603699 5096107	0.0
MDM2 proto-oncogene	MDM2	0.5971261493852453	0.0	0.4826388888888889	0.0	0.3213271367 2665114	0.0160308882 14530907	0.2983819122115699	0.0
tumor protein p63	TP63	0.5965456588544794	0.0	0.4166666666666667	0.0	0.4869080045 3349176	0.0	0.04923888888888889	0.0
RB transcriptional corepressor 1	RB1	0.5953011170696666	0.0	0.5560875896057348	0.0	0.0	0.0	0.1568541098557252	0.0
transducin beta like 1 X-linked receptor 1	TBL1XR1	0.5945910493827161	0.0	0.5861111111111111	0.0	0.0	0.0	0.07631944444444445	0.0
BRCA2 DNA repair associated	BRCA2	0.5944784103249043	0.0	0.5642010381593715	0.0	0.0	0.0	0.12110948866213152	0.0
exportin 1	XPO1	0.5811090256000897	0.0	0.5015432098765432	0.0	0.2787798824737049	0.0	0.08883760594608214	0.0
AT-rich interaction domain 1B	ARID1B	0.5795966392318244	0.0	0.5686299725651577	0.0	0.0	0.0	0.0987000000000001	0.0
transformation/transcription domain associated protein	TRRAP	0.5783380384583254	0.0	0.5128205128205128	0.0	0.2620701025512506	0.0	0.0	0.0
major histocompatibility complex, class II, DQ beta 1	HLA-DQB1	0.5749026631525533	0.558725	0.0	0.0	0.0	0.0	0.06471065261021341	0.0
BLM RecQ like helicase	BLM	0.5740002410698014	0.0	0.5503472222222222	0.0	0.0	0.0050672181649 46112	0.2100268594104308	0.0
mechanistic target of rapamycin kinase	MTOR	0.5727270781670275	0.0	0.4802350427 3504275	0.1463611111111111	0.0	0.0	0.3049187590 1188986	0.0
notch receptor 1	NOTCH1	0.5726441803112589	0.0	0.5268454218106996	0.0	0.0	0.0	0.1831950340022372	0.0
Janus kinase 2	JAK2	0.5723777039170864	0.0	0.4809027777777778	0.1	0.25	0.0	0.2045243352537779	0.0

Table S1: (cont.)

MHC class I polypeptide-related sequence A	MICA	0.5723063475180985	0.4959176003468815	0.0	0.0	0.0	0.0	0.3055549886848679	0.0
large tumor suppressor kinase 1	LATS1	0.5703243854436789	0.0	0.5183917548500881	0.0	0.0	0.0	0.2077305223743 6304	0.0
axin 2	AXIN2	0.5695687718501588	0.0	0.5362225651577504	0.0	0.0	0.1009800000000001	0.0729108602316 7641	0.0
membrane associated guanylate kinase, WW and PDZ domain containing 1	MAGI1	0.5688180827886711	0.0	0.5688180827886711	0.0	0.0	0.0	0.0	0.0
SMAD family member 4	SMAD4	0.5669502894919922	0.0	0.5045438957475995	0.0	0.0	0.0	0.2496255749775 7104	0.0
WT1 transcription factor	WT1	0.5628203537469423	0.0	0.5038580246913581	0.0	0.0	0.0436003826000 1573	0.2164713684001 0772	0.0
LDL receptor related protein 1B	LRP1B	0.5624908315177922	0.0	0.5609908315177923	0.0	0.0	0.0	0.006	0.0
signal transducer and activator of transcription 3	STAT3	0.5616830795949018	0.0	0.4819444444444445	0.0	0.0	0.0038231089785598714	0.3172553810557 8577	0.0
ROS proto-oncogene 1, receptor tyrosine kinase	ROS1	0.5609159069586652	0.0	0.5562114197530865	0.0	0.0	0.0106153848502 0833	0.0141000000000 0001	0.0
ALK receptor tyrosine kinase	ALK	0.5607288628065256	0.0	0.5055941358024691	0.0	0.0	0.0	0.2205389080162 2555	0.0
lysine methyltransferase 2A	KMT2A	0.5592055033383725	0.0	0.5415509259 25926	0.0	0.0	0.0	0.0706183096497 8585	0.0
lymphoid enhancer binding factor 1	LEF1	0.5588248547450064	0.0	0.4819444444 4444445	0.0	0.0	0.09504	0.2652816412022 4806	0.0
nucleoporin 98	NUP98	0.5586271367521367	0.0	0.5586271367521367	0.0	0.0	0.0	0.0	0.0
BCL2 apoptosis regulator	BCL2	0.5585410480596 47	0.0	0.4837962962962963	0.0	0.27229961708123734	0.0	0.0600286274373 7181	0.0
MYC proto-oncogene, bHLH transcription factor	MYC	0.5580777475466973	0.0	0.4837962962962963	0.0	0.0	0.0052287695684767035	0.2948019074156145	0.0
protein tyrosine phosphatase receptor type B	PTPRB	0.5577487863823485	0.0	0.5098379629 62963	0.0	0.0	0.0045974107744469299	0.1896	0.0
SUZ12 polycomb repressive complex 2 subunit	SUZ12	0.5569387304118836	0.0	0.5444315843621399	0.0	0.0	0.0	0.0500285841989 7464	0.0
ANTXR cell adhesion molecule 1	ANTXR1	0.5567958696 44154	0.0	0.5336555918663761	0.0	0.0	0.0	0.0925611111111112	0.0
lysine demethylase 6A	KDM6A	0.5562876823 79349	0.0	0.5525876823 79349	0.0	0.0	0.0	0.0148000000000000002	0.0
neurotrophic receptor tyrosine kinase 1	NTRK1	0.5559287824412941	0.0	0.5	0.2	0.0	0.0	0.0533590419716 4604	0.0
Janus kinase 1	JAK1	0.5550099773535351	0.0	0.4798280423280423	0.0	0.25	0.0	0.1141374152294 3456	0.0
nuclear receptor corepressor 1	NCOR1	0.5533784050785039	0.0	0.5459104938271605	0.0	0.0	0.0048862012620 90616	0.0277000000000000002	0.0
forkhead box A1	FOXA1	0.5517119813535 04	0.0	0.4826388888888889	0.0	0.0	0.0	0.2762923698584 6023	0.0
erb-b2 receptor tyrosine kinase 4	ERBB4	0.5508286756453423	0.0	0.4804292929 2929293	0.25	0.0	0.0	0.0710944444444445	0.0
MDS1 and EVI1 complex locus	MECOM	0.5502035881634597	0.0	0.5346628679962013	0.0	0.0	0.0057352315053256775	0.0596138888888889	0.0
SMAD family member 3	SMAD3	0.5476457226977884	0.0	0.4819444444 4444445	0.0	0.0	0.0	0.2628051130133759	0.0
heterogeneous nuclear ribonucleoprotein A2/B1	HNRNPA2B1	0.5449423981481482	0.0	0.4814814814814815	0.0	0.0	0.0	0.2538436666666667	0.0
APC membrane recruitment protein 1	AMER1	0.5446033950617284	0.0	0.5336033950617284	0.0	0.0	0.0	0.0440000000000000004	0.0
kelch like ECH associated protein 1	KEAP1	0.5413374302684604	0.0	0.5334809447128288	0.0	0.0	0.0	0.0314259422225 26315	0.0
large tumor suppressor kinase 2	LATS2	0.5409468257117662	0.0	0.4814814814814815	0.0	0.0	0.0	0.2378613769211 3885	0.0

Table S1: (cont.)

bromodomain containing 4	BRD4	0.5406746153335232	0.0	0.4819444444444445	0.0	0.0	0.0	0.23492068355631493	0.0
DEAD-box helicase 3 X-linked	DDX3X	0.5404599877262534	0.0	0.4837962962962963	0.0	0.0	0.0	0.2266547657198283	0.0
tripartite motif containing 24	TRIM24	0.5393013120265299	0.0	0.4826388888888889	0.0	0.0	0.007975558238768034	0.22310500000000003	0.0
F-box and WD repeat domain containing 7	FBXW7	0.5383821956185575	0.0	0.4861111111111116	0.0	0.0	0.0	0.20908433802985535	0.0
FAT atypical cadherin 4	FAT4	0.5379644207747739	0.0	0.5219650205761316	0.0	0.0	0.05466426746109662	0.021	0.0
CCCTC-binding factor	CTCF	0.5371270000336295	0.0	0.4807098765432099	0.0	0.0	0.0	0.22566849396167835	0.0
lysine methyltransferase 2D	KMT2D	0.5367155349794239	0.0	0.5367155349794239	0.0	0.0	0.0	0.0	0.0
MYCN proto-oncogene, bHLH transcription factor	MYCN	0.5365379629629963	0.0	0.4837962962962963	0.0	0.0	0.0	0.21096666666666666	0.0
SMAD family member 2	SMAD2	0.5345736926932448	0.0	0.5015432098765432	0.0	0.0	0.006318843720557774	0.12931355627989186	0.0
BCL6 corepressor like 1	BCORL1	0.5340701388888889	0.0	0.4805555555555557	0.0	0.0	0.0	0.21405833333333335	0.0
menin 1	MEN1	0.5328018838708323	0.0	0.5126028806584362	0.0	0.0	0.0	0.08079601284958428	0.0
proteasome 26S subunit, non-ATPase 2	PSMD2	0.5325445263227183	0.0	0.1827417695473251	0.1	0.4757479728224776	0.0	0.0	0.0
fizzy and cell division cycle 20 related 1	FZR1	0.5324556327160493	0.0	0.5324556327160493	0.0	0.0	0.0	0.0	0.0
novel transcript	AL031847.2	0.5316358024691358	0.0	0.5316358024691358	0.0	0.0	0.0	0.0	0.0
neurofibromin 2	NF2	0.5316305689247922	0.0	0.5086805555555556	0.0	0.0	0.0	0.09180005347694632	0.0
capicua transcriptional repressor	CIC	0.5314521604938272	0.0	0.5146604938271605	0.0	0.0	0.0	0.06716666666666667	0.0
DNA methyltransferase 3 alpha	DNMT3A	0.5304657236665058	0.0	0.4814814814814815	0.0	0.0	0.0	0.1959369687400971	0.0
perforin 1	PRF1	0.5295794444444445	0.0	0.4819444444444445	0.0	0.0	0.0	0.0	0.190540000000000001
E1A binding protein p300	EP300	0.529352295383308	0.0	0.4166666666666667	0.0	0.4173192436 06814	0.0	0.07079746598639455	0.0
zinc finger homeobox 3	ZFHX3	0.5287114027777777	0.0	0.509765625	0.0	0.0	0.0	0.07578311111111112	0.0
T-box 3	TBX3	0.5280305555555556	0.0	0.5086805555555556	0.0	0.0	0.0	0.07740000000000001	0.0
mitogen-activated protein kinase kinase 1	MAP2K1	0.5245195176514912	0.0	0.4166666666666667	0.2824722222222222	0.25	0.003150151484532672	0.14929618547375884	0.0
BCL6 corepressor	BCOR	0.5244957231040565	0.0	0.5147707231040565	0.0	0.0	0.0	0.0389	0.0
dicer 1, ribonuclease III	DICER1	0.5239339265285772	0.0	0.5095486111111111	0.0	0.0	0.012899088757195015	0.05180833333333345	0.0
tet methylcytosine dioxygenase 1	TET1	0.5225722769141387	0.0	0.4861111111111116	0.0	0.0	0.010892276675642798	0.14100365135626883	0.0
mitogen-activated protein kinase kinase kinase 1	MAP3K1	0.5213145833333334	0.0	0.5091145833333334	0.0	0.0	0.0	0.0488	0.0
MYB proto-oncogene, transcription factor	MYB	0.5201984567901234	0.0	0.5026234567901234	0.0	0.0	0.0	0.07030000000000002	0.0
lysine demethylase 5A	KDM5A	0.5199160276441358	0.0	0.5086805555555556	0.0	0.0	0.015775498797221673	0.03793055555555556	0.0
ASXL transcriptional regulator 1	ASXL1	0.5179694919278253	0.0	0.5179694919278253	0.0	0.0	0.0	0.0	0.0
mutS homolog 6	MSH6	0.5175740249625231	0.0	0.5057870370 37037	0.0	0.0	0.009557891329375588	0.042899999999999994	0.0
mitogen-activated protein kinase kinase 2	MAP2K2	0.5168103055555555	0.0	0.4166666666666667	0.2824722222222222	0.25	0.0	0.027964888888888893	0.0
erb-b2 receptor tyrosine kinase 2	ERBB2	0.5167211187548029	0.0	0.4166666666666667	0.2915	0.0	0.004958456731293913	0.24182593688187395	0.0
calmodulin binding transcription activator 1	CAMTA1	0.5165257936507937	0.0	0.4811507936 5079366	0.0	0.0	0.0	0.1415	0.0

Table S1: (cont.)

clathrin heavy chain	CLTC	0.5151932379349046	0.0	0.5087682379349046	0.0	0.0	0.0	0.025700000000000004	0.0
transforming growth factor beta receptor 2	TGFBR2	0.5145411498826575	0.0	0.4166666666666667	0.0	0.3649107375740053	0.0	0.059821189402240548	0.0
staphylococcal nuclease and tudor domain containing 1	SND1	0.5145059631341882	0.0	0.4805555555555557	0.0	0.0	0.0	0.1358016303145306	0.0
myosin heavy chain 9	MYH9	0.5126028806584362	0.0	0.512602880658362	0.0	0.0	0.0	0.0	0.0
collagen type II alpha 1 chain	COL2A1	0.5122056353070112	0.0	0.509375	0.0	0.0	0.011322541228044951	0.0	0.0
RAN binding protein 2	RANBP2	0.5120876543209877	0.0	0.5084876543209876	0.0	0.0	0.0	0.014400000000000003	0.0
phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit beta	PIK3CB	0.5107222222222222	0.0	0.4819444444444445	0.1	0.0	0.0	0.034	0.0
nuclear receptor corepressor 2	NCOR2	0.5097222222222222	0.0	0.5097222222222222	0.0	0.0	0.0	0.0	0.0
protein tyrosine phosphatase receptor type K	PTPRK	0.5096188271604939	0.0	0.5042438271604939	0.0	0.0	0.0	0.0215	0.0
glypican 3	GPC3	0.5096175403798529	0.0	0.4166666666666667	0.0	0.0	0.11874097684591407	0.3190297273656718 7	0.0
Janus kinase 3	JAK3	0.5094803591470257	0.0	0.5049803591470258	0.0	0.0	0.0	0.018000000000000002	0.0
calcium voltage-gated channel subunit alpha1 D	CACNA1D	0.5091145833333334	0.0	0.5091145833333334	0.0	0.0	0.0	0.0	0.0
spermine oxidase	SMOX	0.5050592116280685	0.0	0.4807098765432099	0.0	0.0	0.013269015763727505	0.0915	0.0
protein tyrosine phosphatase non-receptor type 13	PTPN13	0.5050426116235932	0.0	0.4800925925925926	0.0	0.0	0.034967454442270 47	0.0842589852607709 8	0.0
folliculin	FLCN	0.5045432098765432	0.0	0.5015432098765432	0.0	0.0	0.0	0.012	0.0
mutS homolog 2	MSH2	0.5045118905828747	0.0	0.4819444444444445	0.0	0.0	0.011729374590504852	0.0850567291801632 6	0.0
Werner syndrome RecQ like helicase	WRN	0.5016212962962963	0.0	0.4837962962962963	0.0	0.0	0.0	0.0713	0.0
secreted frizzled related protein 4	SFRP4	0.5009076574074074	0.0	0.4837962962962963	0.0	0.0	0.0	0.068445444444444445	0.0
ERCC excision repair 3, TFIIH core complex helicase subunit	ERCC3	0.4987868240740 7404	0.0	0.4837962962962963	0.0	0.0	0.0	0.059962111111111111	0.0
BRCA1 DNA repair associated	BRCA1	0.4987662652464 1393	0.0	0.4819444444444445	0.0	0.0	0.0	0.067287283207877 93	0.0
calreticulin	CALR	0.4985246435333 6484	0.0	0.4837962962962963	0.0	0.0	0.013533906293537094	0.0528983194844800 8	0.0
DNA polymerase theta	POLQ	0.4980527777777778	0.0	0.4809027777777778	0.0	0.0	0.0686	0.0	0.0
dynamin 2	DNM2	0.4978462962962963	0.0	0.4837962962962963	0.0	0.0	0.0	0.056200000000000001	0.0
activin A receptor type 1	ACVR1	0.4964737962962963	0.0	0.4837962962962963	0.0	0.0	0.0	0.050710000000000005	0.0
AKT serine/threonine kinase 2	AKT2	0.4961176939 1160027	0.0	0.4166666666666667	0.2	0.0	0.0	0.2289152200908456	0.0
tet methylcytosine dioxygenase 2	TET2	0.4961064514898059	0.0	0.4166666666666667	0.0	0.0	0.0	0.1790643134 0825282	0.2381750000 0000003
C-X-C motif chemokine receptor 4	CXCR4	0.4946831754 2109247	0.0	0.4166666666666667	0.0	0.0	0.0	0.3120660350 1770314	0.0
spalt like transcription factor 4	SALL4	0.4932813293571418	0.0	0.4166666666666667	0.0	0.0	0.0	0.3064586507619005	0.0
MALT1 paracaspase	MALT1	0.492425	0.0	0.4819444444444445	0.0	0.0	0.0	0.041922222222222222	0.0
nuclear receptor coactivator 2	NCOA2	0.4913231481481482	0.0	0.4814814814814815	0.0	0.0	0.0	0.039366666666666667	0.0
ABL proto-oncogene 2, non-receptor tyrosine kinase	ABL2	0.4911111111111116	0.0	0.4861111111111116	0.0	0.0	0.0	0.020000000000000004	0.0
stromal antigen 2	STAG2	0.4908962962962963	0.0	0.4837962962962963	0.0	0.0	0.0	0.0284	0.0
Cbl proto-oncogene	CBL	0.4908121631219 3915	0.0	0.4166666666666667	0.0	0.2965819858 2108986	0.0	0.0	0.0

Table S1: (cont.)

atypical chemokine receptor 3	ACKR3	0.4904848443977607	0.0	0.4166666666666667	0.0	0.0	0.006644202271681783	0.29231973213696205	0.0
lysine demethylase 5C	KDM5C	0.4903292929292929295	0.0	0.48042929292929293	0.0	0.0	0.0	0.0396	0.0
IKAROS family zinc finger 1	IKZF1	0.4898157088325562	0.0	0.4809027777777778	0.0	0.0	0.016316379493006 01	0.028399999999999998	0.0
interleukin 7 receptor	IL7R	0.4897674625239779	0.0	0.4826388888888889	0.0	0.0	0.006075912715801404	0.025813888888888893	0.0
Kruppel like factor 6	KLF6	0.4896506742959838	0.0	0.4166666666666667	0.0	0.0	0.0	0.2919360305172685	0.0
protein phosphatase 6 catalytic subunit	PPP6C	0.4892962962962963	0.0	0.4837962962962963	0.0	0.0	0.0	0.022000000000000002	0.0
Kruppel like factor 4	KLF4	0.4892917721325513	0.0	0.4166666666666667	0.0	0.0	0.011050201957 06419	0.2855892209937323	0.0
SH2B adaptor protein 3	SH2B3	0.4884234241560314	0.0	0.4819444444444445	0.0	0.0	0.013310817404282369	0.020000000000000004	0.0
ERCC excision repair 4, endonuclease catalytic subunit	ERCC4	0.4883992610503722	0.0	0.48115079365079366	0.0	0.0	0.018436206596 20696	0.020800000000000003	0.0
succinate dehydrogenase complex flavoprotein subunit A	SDHA	0.4877814814814815	0.0	0.4814814814814815	0.0	0.0	0.0	0.0252	0.0
T cell leukemia homeobox 3	TLX3	0.4876388888888889	0.0	0.4826388888888889	0.0	0.0	0.0	0.020000000000000004	0.0
ethanolamine kinase 1	ETNK1	0.4873492581130266	0.0	0.4826388888888889	0.0	0.0	0.018841476896550798	0.0	0.0
ubiquitin protein ligase E3 component n-recognin 5	UBR5	0.4870277777777778	0.0	0.4809027777777778	0.0	0.0	0.0	0.024500000000000004	0.0
microtubule associated scaffold protein 1	MTUS1	0.4868888888888887	0.4763888888888886	0.0	0.0	0.0	0.0	0.042	0.0
lysine acetyltransferase 6A	KAT6A	0.4858555555555556	0.0	0.4805555555555557	0.0	0.0	0.0	0.021200000000000004	0.0
epidermal growth factor receptor pathway substrate 15	EPS15	0.4856444444444443	0.0	0.4819444444444445	0.0	0.0	0.0	0.014800000000000002	0.0
mediator complex subunit 12	MED12	0.4847507936507936 5	0.0	0.48115079365079366	0.0	0.0	0.0	0.014400000000000003	0.0
zinc finger protein 331	ZNF331	0.4844959760178878	0.0	0.4814814814814815	0.0	0.0	0.012057978145 58599	0.0	0.0
afadin, adherens junction formation factor	AFDN	0.4841388888888889	0.0	0.4826388888888889	0.0	0.0	0.0	0.006	0.0
zinc finger CCCH-type, RNA binding motif and serine/arginine rich 2	ZRSR2	0.4837962962962963	0.0	0.4837962962962963	0.0	0.0	0.0	0.0	0.0
TAL bHLH transcription factor 1, erythroid differentiation factor	TAL1	0.4837962962962963	0.0	0.4837962962962963	0.0	0.0	0.0	0.0	0.0
interferon regulatory factor 4	IRF4	0.4837962962962963	0.0	0.4837962962962963	0.0	0.0	0.0	0.0	0.0
MLLT10 histone lysine methyltransferase DOT1L cofactor	MLLT10	0.4837962962962963	0.0	0.4837962962962963	0.0	0.0	0.0	0.0	0.0
homeobox A11	HOXA11	0.4837350885164541	0.0	0.4166666666666667	0.0	0.0	0.017242796648086516	0.26061022222222224	0.0
inhibitor of nuclear factor kappa B kinase subunit beta	IKBKB	0.4836152235054762	0.0	0.4166666666666667	0.0	0.0	0.0	0.26779422735523806	0.0
promyelocytic leukemia	PML	0.4832134898397811	0.0	0.4166666666666667	0.0	0.0	0.0	0.2661872926924575	0.0
notch receptor 2	NOTCH2	0.4830284410424 3325	0.0	0.4166666666666667	0.0	0.0	0.0	0.2654470975 0306617	0.0
BAF chromatin remodeling complex subunit BCL11B	BCL11B	0.4826388888888889	0.0	0.4826388888888889	0.0	0.0	0.0	0.0	0.0

Table S1: (cont.)

eukaryotic translation initiation factor 3 subunit E	EIF3E	0.4826388888888889	0.0	0.4826388888888889	0.0	0.0	0.0	0.0	0.0
N-myc downstream regulated 1	NDRG1	0.4820581214881947	0.0	0.4166666666666667	0.0	0.0	0.0	0.2615658192861121	0.0
ETS transcription factor ERG	ERG	0.4819444444444445	0.0	0.4819444444444445	0.0	0.0	0.0	0.0	0.0
PR/SET domain 1	PRDM1	0.4816924816621056 5	0.0	0.4166666666666667	0.0	0.2537032599 8175586	0.0	0.0144000000000000 03	0.0
BCL9 transcription coactivator	BCL9	0.4815277026667710 3	0.0	0.4166666666666667	0.0	0.0	0.0050734770621 63406	0.2571892653061 2246	0.0
BUB1 mitotic checkpoint serine/threonine kinase B	BUB1B	0.4815134294368143 6	0.0	0.4166666666666667	0.0	0.0	0.1049304177740 8385	0.2127513098476 6458	0.0
mitochondrial ribosomal protein L36	MRPL36	0.4814814814814815	0.0	0.4814814814814815	0.0	0.0	0.0	0.0	0.0
AF4/FMR2 family member 3	AFF3	0.4814814814814815	0.0	0.4814814814814815	0.0	0.0	0.0	0.0	0.0
ubiquitin specific peptidase 6	USP6	0.4814814814814815	0.0	0.4814814814814815	0.0	0.0	0.0	0.0	0.0
ATPase plasma membrane Ca2+ transporting 3	ATP2B3	0.4814814814814815	0.0	0.4814814814814815	0.0	0.0	0.0	0.0	0.0
ERCC excision repair 2, TFIIH core complex helicase subunit	ERCC2	0.4814439823628123	0.0	0.4166666666666667	0.0	0.0	0.0	0.2591092627845 8246	0.0
CCAAT enhancer binding protein alpha	CEBPA	0.4814352989259465	0.0	0.4166666666666667	0.0	0.0	0.0	0.1542026903335187	0.1905400000000001
glutamate ionotropic receptor NMDA type subunit 2A	GRIN2A	0.4811507936507 9366	0.0	0.4811507936507 9366	0.0	0.0	0.0	0.0	0.0
clathrin heavy chain like 1	CLTCL1	0.4811507936507 9366	0.0	0.4811507936507 9366	0.0	0.0	0.0	0.0	0.0
FA complementation group A	FANCA	0.4809027777777778	0.0	0.4809027777777778	0.0	0.0	0.0	0.0	0.0
protein tyrosine phosphatase non-receptor type 11	PTPN11	0.4807721850516147	0.0	0.4166666666666667	0.0	0.0	0.0	0.2564220735397921	0.0
cyclin dependent kinase 12	CDK12	0.4807098765432099	0.0	0.4807098765432099	0.0	0.0	0.0	0.0	0.0
BAF chromatin remodeling complex subunit BCL11A	BCL11A	0.4807098765432099	0.0	0.4807098765432099	0.0	0.0	0.0	0.0	0.0
endothelial PAS domain protein 1	EPAS1	0.4806284779608973 6	0.0	0.4166666666666667	0.0	0.0	0.0	0.2558472451769226	0.0
signal transducer and activator of transcription 4	STAT4	0.4805076475706178	0.4202887001878941 6	0.0	0.0	0.0	0.0	0.2408757895308946	0.0
caspase recruitment domain family member 11	CARD11	0.4803240740740740 7	0.0	0.4803240740740740 7	0.0	0.0	0.0	0.0	0.0
spen family transcriptional repressor	SPEN	0.4800925925925926	0.0	0.4800925925925926	0.0	0.0	0.0	0.0	0.0
zinc finger protein 521	ZNF521	0.4800347222222222	0.0	0.4800347222222222	0.0	0.0	0.0	0.0	0.0
BCL6 transcription repressor	BCL6	0.4785326388888889	0.0	0.4166666666666667	0.0	0.0	0.0	0.0209000000000000 05	0.2381750000000003
tropomyosin 3	TPM3	0.4782542015435738 4	0.0	0.4166666666666667	0.0	0.0	0.0106828138921 64526	0.2416022222222225	0.0
LCK proto-oncogene, Src family tyrosine kinase	LCK	0.4779500000000000 4	0.0	0.4166666666666667	0.2361111111111111	0.0	0.0	0.0203000000000000 2	0.0
protein tyrosine phosphatase receptor type C	PTPRC	0.4772346951332031	0.0	0.4166666666666667	0.0	0.0	0.0077875408075 71794	0.1120167644945 6875	0.1905400000000001
CYLD lysine 63 deubiquitinase	CYLD	0.4767167816577 3884	0.0	0.4166666666666667	0.0	0.0	0.0096401839003 60177	0.2359159337863 5073	0.0
forkhead box P1	FOXP1	0.4743629283037 2895	0.0	0.4166666666666667	0.0	0.0	0.0	0.2307850465482 4894	0.0

Table S1: (cont.)

ATPase Na+/K+ transporting subunit alpha 1	ATP1A1	0.4715894529478458	0.0	0.4166666666666667	0.0	0.0	0.0	0.21969114512471655	0.0
protein phosphatase, Mg2+/Mn2+ dependent 1D	PPM1D	0.47124533333333335	0.0	0.4166666666666667	0.0	0.0	0.0	0.21831466666666668	0.0
exostosin glycosyltransferase 1	EXT1	0.4696368055555556	0.0	0.4166666666666667	0.0	0.0	0.0	0.21188055555555557	0.0
cyclin dependent kinase inhibitor 3	CDKN3	0.46949039181922636	0.4375	0.0	0.0	0.0	0.11009682653616469	0.040195666666666664	0.0
succinate dehydrogenase complex iron sulfur subunit B	SDHB	0.4684250000000004	0.0	0.4166666666666667	0.0	0.0	0.0	0.20703333333333335	0.0
TNF alpha induced protein 3	TNFAIP3	0.46839872839506175	0.0	0.4166666666666667	0.0	0.0	0.0	0.036873555555555555	0.190540000000000001
RecQ like helicase 4	RECQL4	0.468275	0.0	0.4166666666666667	0.0	0.0	0.0	0.20643333333333336	0.0
BRCA1 associated RING domain 1	BARD1	0.4682158201960016	0.0	0.4166666666666667	0.0	0.0	0.01394238176401465	0.2	0.0
myocardin related transcription factor A	MRTFA	0.4641208333333333337	0.0	0.4166666666666667	0.0	0.0	0.0	0.18981666666666666	0.0
STIL centriolar assembly protein	STIL	0.4637682512741777 6	0.0	0.4166666666666667	0.0	0.0	0.013064261467599642	0.1826	0.0
baculoviral IAP repeat containing 3	BIRC3	0.4584916026077097 7	0.0	0.4166666666666667	0.0	0.0	0.0	0.16729974376417228	0.0
mutL homolog 1	MLH1	0.4557709274376417 5	0.0	0.4166666666666667	0.0	0.0	0.0	0.15641704308390023	0.0
von Hippel-Lindau tumor suppressor	VHL	0.4546254560727872 6	0.0	0.4166666666666667	0.0	0.0	0.0	0.15183515762448235	0.0
SRY-box 2	SOX2	0.4545158407071012	0.0	0.4166666666666667	0.0	0.0	0.0	0.1513966961617381	0.0
phosphoinositide-3-kinase regulatory subunit 1	PIK3R1	0.4542739086356765	0.0	0.4166666666666667	0.1	0.0	0.0	0.10598452343159487	0.0
programmed cell death 1 ligand 2	PDCD1LG2	0.4524387940759637	0.0	0.4166666666666667	0.0	0.0	0.0	0.1430885096371882	0.0
XPC complex subunit, DNA damage recognition and repair factor	XPC	0.4521452966269841 6	0.0	0.4166666666666667	0.0	0.0	0.0	0.1419145198412698 2	0.0
SSX family member 1	SSX1	0.4518190544144424	0.0	0.4166666666666667	0.0	0.0	0.1308762176576958	0.021900000000000003	0.0
PLAG1 zinc finger	PLAG1	0.4508166666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.1366	0.0
GATA binding protein 2	GATA2	0.44998125	0.0	0.4166666666666667	0.0	0.0	0.0	0.1332583333333333	0.0
catenin alpha 2	CTNNA2	0.4467091070466468	0.0	0.3100198412698413	0.0	0.3566691467290076	0.1049400000000002	0.014000000000000002	0.0
DEAD-box helicase 5	DDX5	0.4466041666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.1197500000000001	0.0
WNT inhibitory factor 1	WIF1	0.4448606018285497 5	0.0	0.4166666666666667	0.0	0.0	0.0	0.1127757406475322	0.0
phospholipase C gamma 1	PLCG1	0.4437436638321995 4	0.0	0.4166666666666667	0.0	0.0	0.0	0.1083079886621315 2	0.0
F-box protein 11	FBXO11	0.4415666666666666	0.0	0.4166666666666667	0.0	0.0	0.0	0.0996000000000000 1	0.0
protein phosphatase 2 scaffold subunit Aalpha	PPP2R1A	0.4410666666666666	0.0	0.4166666666666667	0.0	0.0	0.0	0.0976	0.0
FAT atypical cadherin 1	FAT1	0.4409659722222222	0.0	0.4166666666666667	0.0	0.0	0.0	0.0971972222222222 4	0.0
spleen associated tyrosine kinase	SYK	0.4408361143121173	0.0	0.4166666666666667	0.0	0.0	0.0	0.0966777905820252 9	0.0
signal transducer and activator of transcription 6	STAT6	0.4389903825113379	0.0	0.4166666666666667	0.0	0.0	0.0	0.0892948633786848 1	0.0
phosphatidylinositol-3,4,5-trisphosphate dependent Rac exchange factor 2	PREX2	0.4388004166666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.088535	0.0
neurotrophic receptor tyrosine kinase 3	NTRK3	0.4387444444444445	0.0	0.4166666666666667	0.0	0.0	0.0	0.0883111111111113	0.0

Table S1: (cont.)

solute carrier family 34 member 2	SLC34A2	0.43868244846000704	0.0	0.4166666666666667	0.0	0.0	0.05819203614006318	0.062200000000000005	0.0
serine and arginine rich splicing factor 2	SRSF2	0.43838818594 10431	0.0	0.4166666666666667	0.0	0.0	0.0	0.08688607709750566	0.0
FUS RNA binding protein	FUS	0.43819700222 7344	0.0	0.4166666666666667	0.0	0.0	0.00706522157 67075905	0.08298124376417236	0.0
speckle type BTB/POZ protein	SPOP	0.43801666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0854	0.0
retinoic acid receptor alpha	RARA	0.43749927166 48472	0.0	0.4166666666666667	0.0	0.0	0.0	0.08333041999272202	0.0
tripartite motif containing 33	TRIM33	0.43735625	0.0	0.4166666666666667	0.0	0.0	0.0	0.08275833333333334	0.0
FA complementation group D2	FANCD2	0.43576861712 63715	0.0	0.4166666666666667	0.0	0.0	0.01617380413 7342988	0.06921944444444444	0.0
zinc finger and BTB domain containing 16	ZBTB16	0.43560347222222223	0.0	0.4166666666666667	0.0	0.0	0.0	0.07574722222222222	0.0
BCL3 transcription coactivator	BCL3	0.43556666666666666	0.0	0.4166666666666667	0.0	0.0	0.0	0.0756	0.0
patched 1	PTCH1	0.43525435709886956	0.0	0.4166666666666667	0.0	0.0	0.0	0.07435076172881151	0.0
checkpoint kinase 2	CHEK2	0.43521942651 03698	0.0	0.4166666666666667	0.0	0.0	0.01334660786 7704492	0.06827921365583273	0.0
FA complementation group G	FANCG	0.43431935855530024	0.0	0.4166666666666667	0.0	0.0	0.01824922699770185	0.0625	0.0
neuregulin 1	NRG1	0.43396308152409213	0.0	0.4166666666666667	0.0	0.0	0.01353023371 6829067	0.06317222222222223	0.0
mutY DNA glycosylase	MUTYH	0.43350864002267575	0.0	0.4166666666666667	0.0	0.0	0.0	0.06736789342403629	0.0
paired box 5	PAX5	0.43350493367346943	0.0	0.4166666666666667	0.0	0.0	0.0	0.06735306802721089	0.0
nuclear factor of activated T cells 2	NFATC2	0.43284230555 55556	0.0	0.4166666666666667	0.0	0.0	0.0	0.06470255555555557	0.0
ETS variant 4	ETV4	0.43251319444444447	0.0	0.4166666666666667	0.0	0.0	0.0	0.06338611111111112	0.0
BCL2 associated X, apoptosis regulator	BAX	0.43235171387 78767	0.0	0.4166666666666667	0.0	0.0	0.0	0.06274018884484003	0.0
DCC netrin 1 receptor	DCC	0.43222436968 82377	0.0	0.32187273057371096	0.0	0.3422933889124869	0.0	0.08516518319090724	0.0
chromodomain helicase DNA binding protein 4	CHD4	0.43221666666 66667	0.0	0.4166666666666667	0.0	0.0	0.0	0.062200000000000005	0.0
nuclear receptor binding SET domain protein 2	NSD2	0.43206666666 66667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0616	0.0
DNA damage inducible transcript 3	DDIT3	0.43205996768 19175	0.0	0.4166666666666667	0.0	0.0	0.0	0.06157320406100331	0.0
ERCC excision repair 5, endonuclease	ERCC5	0.43168761224 4898	0.0	0.4166666666666667	0.0	0.0	0.0	0.06008378231292517	0.0
CBFA2/RUNX1 translocation partner 3	CBFA2T3	0.43116448221260667	0.0	0.4166666666666667	0.0	0.0	0.05159126218376003	0.014400000000000003	0.0
protein tyrosine phosphatase receptor type T	PTPRT	0.42974722222 22222	0.0	0.4166666666666667	0.0	0.0	0.0	0.05232222222222223	0.0
drosha ribonuclease III	DROSHA	0.42969767308 50898	0.0	0.4166666666666667	0.0	0.0	0.00387905776 5808491	0.0504	0.0
Bruton tyrosine kinase	BTK	0.42872675000000005	0.0	0.4166666666666667	0.0	0.0	0.0	0.04824033333333333336	0.0
PBX homeobox 1	PBX1	0.42871192057478624	0.0	0.4166666666666667	0.0	0.0	0.04818101563 2478126	0.0	0.0
serine and arginine rich splicing factor 3	SRSF3	0.42864166666 66667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0479	0.0
forkhead box O4	FOXO4	0.42861833333 33334	0.0	0.4166666666666667	0.0	0.0	0.0	0.04780666666666667	0.0
NK2 homeobox 1	NKX2-1	0.42769722222222223	0.0	0.4166666666666667	0.0	0.0	0.0	0.04412222222222222	0.0
SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily b, member 1	SMARCB1	0.42765000000000003	0.0	0.4166666666666667	0.0	0.0	0.0	0.043933333333333345	0.0

Table S1: (cont.)

PMS1 homolog 2, mismatch repair system component	PMS2	0.4274916666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0433000000000000005	0.0
RUNX1 translocation partner 1	RUNX1T1	0.4264666666666666	0.0	0.4166666666666667	0.0	0.0	0.0	0.0392000000000000006	0.0
ring finger protein 43	RNF43	0.4264205833333333	0.0	0.4166666666666667	0.0	0.0	0.0	0.0390156666666666667	0.0
RUNX family transcription factor 1	RUNX1	0.4261666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.038	0.0
cAMP responsive element binding protein 3 like 1	CREB3L1	0.4250666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0336	0.0
cadherin 11	CDH11	0.4249666666666666	0.0	0.4166666666666667	0.0	0.0	0.0	0.0332	0.0
ETS transcription factor ELK4	ELK4	0.4245666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0316	0.0
transcription factor 7 like 1	TCF7L1	0.4242666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0304	0.0
protection of telomeres 1	POT1	0.4237802777777777777	0.0	0.4166666666666667	0.0	0.0	0.0	0.028454444444444445	0.0
WASP actin nucleation promoting factor	WAS	0.4237666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.028399999999999998	0.0
paired box 3	PAX3	0.4234197747977777054	0.0	0.4166666666666667	0.0	0.0	0.023990210302193272	0.006800000000000000005	0.0
bone morphogenetic protein receptor type 1A	BMPR1A	0.4230583333333333337	0.0	0.4166666666666667	0.0	0.0	0.0	0.025566666666666666668	0.0
RAD21 cohesin complex component	RAD21	0.4230085491339976	0.0	0.4166666666666667	0.0	0.0	0.003976942205977979	0.023600000000000000003	0.0
RNA binding motif protein 10	RBM10	0.4228525	0.0	0.4166666666666667	0.0	0.0	0.0	0.024743333333333333336	0.0
nibrin	NBN	0.422783627362230	0.0	0.4166666666666667	0.0	0.0	0.004527646260070069	0.022455555555555556	0.0
leucine rich repeats and immunoglobulin like domains 3	LRIG3	0.4227666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.024400000000000000005	0.0
DEAD-box helicase 6	DDX6	0.4226916666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.024100000000000000003	0.0
SET binding protein 1	SETBP1	0.4224166666666666	0.0	0.4166666666666667	0.0	0.0	0.0	0.023000000000000000003	0.0
EWS RNA binding protein 1	EWSR1	0.4222777777777778	0.0	0.4166666666666667	0.0	0.0	0.0	0.022444444444444444445	0.0
EBF transcription factor 1	EBF1	0.4222043798010 3	0.0	0.4166666666666667	0.0	0.0	0.004839418209269565	0.020000000000000000004	0.0
histone cluster 1 H3 family member b	HIST1H3B	0.4216145523348752 7	0.0	0.4166666666666667	0.0	0.0	0.019791542672834 37	0.0	0.0
MYCL proto-oncogene, bHLH transcription factor	MYCL	0.4213459165829017 7	0.0	0.4166666666666667	0.0	0.0	0.008813249246116114	0.014800000000000000002	0.0
ras homolog family member H	RHOH	0.4211666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.018000000000000000002	0.0
RAD51 paralog B	RAD51B	0.4210666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0176	0.0
nuclear receptor binding SET domain protein 3	NSD3	0.4209666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0172	0.0
nuclear receptor subfamily 4 group A member 3	NR4A3	0.4208994168258737	0.0	0.4166666666666667	0.0	0.0	0.005694751432862773	0.014400000000000000003	0.0
LIM domain only 2	LMO2	0.4208666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.016800000000000000002	0.0
LIM domain containing preferred translocation partner in lipoma	LPP	0.4206666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.016	0.0
SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 1	SMARCD1	0.4206666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.016	0.0
insulin receptor substrate 4	IRS4	0.4205666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.015600000000000000003	0.0

Table S1: (cont.)

metabolism of cobalamin associated B	MMAB	0.42049914302850694	0.0	0.4166666666666667	0.0	0.0	0.015329905447361042	0.0	0.0
mitogen-activated protein kinase kinase kinase 13	MAP3K13	0.4203666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.014800000000000002	0.0
REL proto-oncogene, NF-kB subunit	REL	0.4202666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.014400000000000003	0.0
myogenic differentiation 1	MYOD1	0.4202666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.014400000000000003	0.0
SUFU negative regulator of hedgehog signaling	SUFU	0.4202666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.014400000000000003	0.0
cysteinyl-tRNA synthetase	CARS	0.41987475465658874	0.0	0.4166666666666667	0.0	0.0	0.012832351959688152	0.0	0.0
ETS variant 6	ETV6	0.4194166666666666	0.0	0.4166666666666667	0.0	0.0	0.0	0.011000000000000003	0.0
MN1 proto-oncogene, transcriptional regulator	MN1	0.41938157635073525	0.0	0.4166666666666667	0.0	0.0	0.010859638736274260	0.0	0.0
E74 like ETS transcription factor 4	ELF4	0.4186666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.008	0.0
kinetochore scaffold 1	KNL1	0.41808427412237875	0.0	0.4166666666666667	0.0	0.0	0.005670429822848301	0.0	0.0
cut like homeobox 1	CUX1	0.4166666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0	0.0
novel transcript	AP003108.2	0.4166666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0	0.0
Rho guanine nucleotide exchange factor 12	ARHGEF12	0.4166666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0	0.0
melanocyte inducing transcription factor	MITF	0.4166666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0	0.0
BCL9 like	BCL9L	0.4166666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0	0.0
mastermind like transcriptional coactivator 2	MAML2	0.4166666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0	0.0
DEAD-box helicase 10	DDX10	0.4166666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0	0.0
myeloid leukemia factor 1	MLF1	0.4166666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0	0.0
bromodomain containing 3	BRD3	0.4166666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0	0.0
FEV transcription factor, ETS family member	FEV	0.4166666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0	0.0
post-GPI attachment to proteins 3	PGAP3	0.4166666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0	0.0
lysine acetyltransferase 6B	KAT6B	0.4166666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0	0.0
Fli-1 proto-oncogene, ETS transcription factor	FLI1	0.4166666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0	0.0
exostosin glycosyltransferase 2	EXT2	0.4166666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0	0.0
PR/SET domain 16	PRDM16	0.4166666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0	0.0
Cbl proto-oncogene C	CBLC	0.4166666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0	0.0
SH3 domain containing GRB2 like 1, endophilin A2	SH3GL1	0.4166666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0	0.0
Rap1 GTPase-GDP dissociation stimulator 1	RAP1GDS1	0.4166666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0	0.0
ubiquitin specific peptidase 8	USP8	0.4166666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0	0.0
cell division cycle 73	CDC73	0.4166666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0	0.0

Table S1: (cont.)

TNF receptor associated factor 7	TRAF7	0.4166666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0	0.0
huntingtin interacting protein 1	HIP1	0.4166666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0	0.0
ribosomal protein L5	RPL5	0.4166666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0	0.0
potassium voltage-gated channel subfamily J member 5	KCNJ5	0.4166666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0	0.0
FA complementation group E	FANCE	0.4166666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0	0.0
coiled-coil domain containing 6	CCDC6	0.4166666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0	0.0
elongation factor for RNA polymerase II	ELL	0.4166666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0	0.0
leucine zipper like transcription regulator 1	LZTR1	0.4166666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0	0.0
fumarate hydratase	FH	0.4166666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0	0.0
CCR4-NOT transcription complex subunit 3	CNOT3	0.4166666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0	0.0
partner and localizer of BRCA2	PALB2	0.4166666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0	0.0
transcription factor binding to IGHM enhancer 3	TFE3	0.4166666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0	0.0
CD79b molecule	CD79B	0.4166666666666667	0.0	0.4166666666666667	0.0	0.0	0.0	0.0	0.0
cyclin dependent kinase 14	CDK14	0.4138900470595945	0.39198282459096256	0.0	0.0	0.0	0.0	0.08762888987452773	0.0
DEK proto-oncogene	DEK	0.41305078008334406	0.0	0.3472222222222222	0.0	0.0	0.0	0.2633142314444875	0.0
MET transcriptional regulator MACC1	MACC1	0.41246662207668994	0.0	0.33984375	0.0	0.0	0.0028969807138033137	0.28920394132284705	0.0
serine/threonine kinase 11	STK11	0.4119060998119336	0.0	0.3472222222222222	0.0	0.0	0.0	0.2587355103588455	0.0
glutamate ionotropic receptor kainate type subunit 1	GRIK1	0.4089696331454392	0.4030446331454392	0.0	0.0	0.0	0.0	0.023700000000000006	0.0
SIX homeobox 1	SIX1	0.4084286210867638 7	0.0	0.3472222222222222	0.0	0.0	0.0247510540665891	0.2338251269844127	0.0
cytochrome P450 family 2 subfamily C member 8	CYP2C8	0.4023248831432459 4	0.0	0.3391203703703 7035	0.0	0.0	0.1148536098538 3936	0.2017720022675 7372	0.0
mucin 16, cell surface associated	MUC16	0.4018199882435 4903	0.0	0.3816054059109 6154	0.0	0.0	0.0	0.0808583293303 5005	0.0
patatin like phospholipase domain containing 3	PNPLA3	0.4012590619221 5466	0.3336220335212275	0.0	0.0	0.0	0.0158581239107 98407	0.2635000585322426	0.0
SET domain bifurcated histone lysine methyltransferase 1	SETDB1	0.3987958209561 6023	0.0	0.3532021604938 2713	0.0	0.0	0.0	0.2595351356764 9283	0.0
LIF receptor alpha	LIFR	0.3903583270933104	0.0	0.3394097222222227	0.0	0.0	0.0626924494844 7719	0.1759311086023626	0.0
peroxisome proliferator activated receptor gamma	PPARG	0.3808817161456 1983	0.0	0.2777777777777778	0.0	0.0	0.0	0.3114372717011754	0.0
Rho GTPase activating protein 5	ARHGAP5	0.3804311728395 0617	0.0	0.3395061728395062	0.0	0.0	0.0	0.1637	0.0
cyclin dependent kinase inhibitor 1B	CDKN1B	0.3799043092702 6553	0.0	0.3472222222222222	0.0	0.0	0.0	0.1307283481921732	0.0
homeobox A13	HOXA13	0.3782149747387 6103	0.0	0.3472222222222222	0.0	0.0	0.0458228361790 7244	0.1036053050976 7855	0.0

Table S1: (cont.)

far upstream element binding protein 1	FUBP1	0.3781801528985051	0.0	0.36548353909465 02	0.0	0.0	0.0	0.0507864552154195	0.0
enhancer of zeste 2 polycomb repressive complex 2 subunit	EZH2	0.3768659604906673	0.0	0.27777777777777778	0.0	0.0	0.009991746236453524	0.3063113220199 5026	0.0
CUB and Sushi multiple domains 3	CSMD3	0.3716770061728 3955	0.0	0.3662770061728 3954	0.0	0.0	0.0	0.0216	0.0
AF4/FMR2 family member 1	AFF1	0.3669114197530 8633	0.0	0.3562114197530 8634	0.0	0.0	0.0	0.0428000000000 0005	0.0
EPH receptor A3	EPHA3	0.3647559115903 0224	0.0	0.3095619658119658	0.2	0.0	0.0050148528005555992	0.0439246573047 1516	0.0
death domain associated protein	DAXX	0.3632305555555556	0.0	0.3472222222222222	0.0	0.0	0.0	0.0640333333333333	0.0
EPH receptor A7	EPHA7	0.3596064814814815	0.0	0.3096064814814815	0.2	0.0	0.0	0.0	0.0
PHD finger protein 6	PHF6	0.3578222222222222	0.0	0.3472222222222222	0.0	0.0	0.0	0.0424	0.0
nuclear receptor binding SET domain protein 1	NSD1	0.3563331433998101	0.0	0.3517331433998101	0.0	0.0	0.0	0.0184	0.0
CCHC-type zinc finger nucleic acid binding protein	CNBP	0.3551222222222223	0.0	0.3472222222222222	0.0	0.0	0.0	0.0316	0.0
nucleophosmin 1	NPM1	0.3549764309072029	0.0	0.27777777777777778	0.0	0.0	0.0	0.2855319864627584	0.0
SRC proto-oncogene, non-receptor tyrosine kinase	SRC	0.3545552372118658	0.0	0.27777777777777778	0.2611111111111111	0.0	0.0	0.1034971349067 9226	0.0
neurobeachin	NBEA	0.3544108807889 7653	0.0	0.3533599887766554	0.0	0.0	0.0042035680492 84553	0.0	0.0
kinectin 1	KTN1	0.3543095238095238	0.0	0.3363095238095238	0.0	0.0	0.0	0.0720000000000 0001	0.0
cyclin dependent kinase inhibitor 2C	CDKN2C	0.3534707566388802	0.0	0.3472222222222222	0.0	0.0	0.0249941376663 19298	0.0	0.0
SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily e, member 1	SMARCE1	0.3534222222222222	0.0	0.3472222222222222	0.0	0.0	0.0	0.0248000000000 0003	0.0
signal transducer and activator of transcription 5B	STAT5B	0.3524222222222222	0.0	0.3472222222222222	0.0	0.0	0.0	0.0208000000000 0003	0.0
thyroid hormone receptor associated protein 3	THRAP3	0.3505940329218106	0.0	0.3488940329218106 5	0.0	0.0	0.0	0.0068000000000 00005	0.0
tenascin C	TNC	0.3483800095559295	0.0	0.3241705246913580 4	0.0	0.0	0.0	0.0968379394582 8595	0.0
integrin subunit alpha V	ITGAV	0.3477904299546446	0.0	0.3098958333333337	0.0	0.0	0.0	0.1515783864852 4494	0.0
MYD88 innate immune signal transduction adaptor	MYD88	0.3472222222222222	0.0	0.3472222222222222	0.0	0.0	0.0	0.0	0.0
NGFI-A binding protein 2	NAB2	0.3472222222222222	0.0	0.3472222222222222	0.0	0.0	0.0	0.0	0.0
G protein subunit alpha 11	GNA11	0.3472222222222222	0.0	0.3472222222222222	0.0	0.0	0.0	0.0	0.0
protein kinase C beta	PRKCB	0.3463817292344362	0.0	0.3101851851851852	0.1	0.0	0.0055548393295534596	0.0976442993197279	0.0
aldehyde dehydrogenase 2 family member	ALDH2	0.3460753710141 1895	0.0	0.27777777777777778	0.0	0.0	0.06596	0.2438748173898 0898	0.0
collagen type III alpha 1 chain	COL3A1	0.3460277777777777777	0.0	0.3402777777777778	0.0	0.0	0.0	0.0230000000000 0003	0.0
forkhead box O1	FOXO1	0.3448953643915 8874	0.0	0.27777777777777778	0.0	0.0	0.0065245558003 83977	0.2655705438772953	0.0
protein tyrosine phosphatase receptor type D	PTPRD	0.3443796759259 2594	0.0	0.3096064814814815	0.0	0.0	0.0	0.1390927777777 7776	0.0
CAP-Gly domain containing linker protein 1	CLIP1	0.3442926143941289	0.0	0.3392857142857143	0.0	0.0	0.0200276004336 58324	0.0	0.0
collagen type I alpha 1 chain	COL1A1	0.3441660943877551	0.0	0.3211419753086 4197	0.0	0.0	0.0	0.0920964763164525	0.0

Table S1: (cont.)

zinc finger E-box binding homeobox 1	ZEB1	0.34395507153780047	0.0	0.27777777777777778	0.0	0.0	0.0	0.2647091750400908	0.0
A-kinase anchoring protein 9	AKAP9	0.34296296296296297	0.0	0.33796296296296297	0.0	0.0	0.0	0.020000000000000004	0.0
phosphodiesterase 4D interacting protein	PDE4DIP	0.34187690286 96957	0.0	0.339506172839 5062	0.0	0.0	0.00948292012 0758064	0.0	0.0
FAT atypical cadherin 3	FAT3	0.34009971509 97151	0.0	0.340099715099 97151	0.0	0.0	0.0	0.0	0.0
gephyrin	GPHN	0.33958333333 333335	0.0	0.339583333333333335	0.0	0.0	0.0	0.0	0.0
sperm antigen with calponin homology and coiled-coil domains 1	SPECC1	0.33958333333 333335	0.0	0.339583333333333335	0.0	0.0	0.0	0.0	0.0
epithelial cell transforming 2 like	ECT2L	0.33950617283 95062	0.0	0.339506172839 5062	0.0	0.0	0.0	0.0	0.0
kinesin family member 5B	KIF5B	0.33940972222 222227	0.0	0.339409722222222227	0.0	0.0	0.0	0.0	0.0
PPFIA binding protein 1	PPFIBP1	0.33888888888 888885	0.0	0.338888888888888885	0.0	0.0	0.0	0.0	0.0
colony stimulating factor 3 receptor	CSF3R	0.33759313592 92531	0.0	0.27777777777777778	0.225	0.0	0.00341926128 7802502	0.030164888888888894	0.0
WNK lysine deficient protein kinase 2	WNK2	0.33671305778 50998	0.0	0.321437757201646 13	0.0	0.0	0.05025675788 937017	0.024400000000000005	0.0
baculoviral IAP repeat containing 6	BIRC6	0.33656932686 65491	0.0	0.323394326866 65491	0.0	0.0	0.0	0.052700000000000004	0.0
Fc fragment of IgG receptor IIb	FCGR2B	0.33535480887 38885	0.0	0.27777777777777778	0.0	0.0	0.08947827986 499665	0.0	0.190540000000000001
protein tyrosine kinase 6	PTK6	0.33521111111 11111	0.0	0.27777777777777778	0.2	0.0	0.0	0.066900000000000002	0.0
zinc and ring finger 3	ZNRF3	0.33465599678 701147	0.0	0.320698302469135 83	0.0	0.0	0.05280855504 928017	0.006800000000000005	0.0
translocated promoter region, nuclear basket protein	TPR	0.33454838664 421993	0.0	0.326748386644 21996	0.0	0.0	0.0	0.031200000000000006	0.0
E74 like ETS transcription factor 3	ELF3	0.33408958333 333333	0.0	0.27777777777777778	0.0	0.0	0.0	0.225247222222222224	0.0
musashi RNA binding protein 2	MSI2	0.33392435710 824997	0.0	0.27777777777777778	0.0	0.0	0.00823096397 4249373	0.220928111111111113	0.0
insulin like growth factor 2 mRNA binding protein 2	IGF2BP2	0.33353938458 2392	0.0	0.311342592592 5926	0.0	0.0	0.0	0.088787167959 1976	0.0
PR/SET domain 2	PRDM2	0.33147687037 03704	0.0	0.309606481481 4815	0.0	0.0	0.0	0.087481555555555556	0.0
terminal nucleotidyltransferase 5C	TENT5C	0.33115277777 77778	0.0	0.27777777777777778	0.0	0.0	0.0	0.213500000000000002	0.0
heat shock protein 90 alpha family class A member 1	HSP90AA1	0.33106611224 4898	0.0	0.27777777777777778	0.1	0.0	0.0	0.168708893424 0363	0.0
ral guanine nucleotide dissociation stimulator	RALGDS	0.33006388888 88889	0.0	0.310763888888888889	0.0	0.0	0.0	0.0772	0.0
heat shock protein 90 alpha family class B member 1	HSP90AB1	0.32779023068 907887	0.0	0.138888888888888889	0.1	0.0	0.00413047058 1479387	0.281698742944 40306	0.0
SET domain containing 1B, histone lysine methyltransferase	SETD1B	0.32708518518 518515	0.0	0.310185185185 1852	0.0	0.0	0.0	0.0676	0.0
mucin 4, cell surface associated	MUC4	0.32647872218 075774	0.0	0.309275793650 79366	0.0	0.0	0.0	0.068811714119 85633	0.0
chromodomain helicase DNA binding protein 2	CHD2	0.32541277777 77778	0.0	0.27777777777777778	0.0	0.0	0.0	0.0	0.190540000000000001
ring finger protein 213	RNF213	0.32475323387 153404	0.0	0.323473936899 8628	0.0	0.0	0.00511718788 6684939	0.0	0.0
motor neuron and pancreas homeobox 1	MNX1	0.32454259259 25926	0.0	0.311342592592 5926	0.0	0.0	0.0	0.052800000000000001	0.0
catenin delta 2	CTNND2	0.32402044836 275345	0.0	0.309606481481 4815	0.0	0.0	0.05765586752 508776	0.0	0.0

Table S1: (cont.)

MLLT1 super elongation complex subunit	MLLT1	0.3239638888888889	0.0	0.3107638888888889	0.0	0.0	0.0	0.052800000000000001	0.0
zinc finger MYM-type containing 3	ZMYM3	0.3234310699588477	0.0	0.3234310699588477	0.0	0.0	0.0	0.0	0.0
zinc finger MYM-type containing 2	ZMYM2	0.3224261294128984	0.0	0.3107638888888889	0.0	0.0	0.007735164716008403	0.043211111111111112	0.0
splicing factor proline and glutamine rich	SFPQ	0.3218777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.1764	0.0
TNF receptor superfamily member 14	TNFRSF14	0.3216319166666667	0.0	0.2777777777777778	0.0	0.0	0.0	0.175416555555555555	0.0
zinc finger protein 429	ZNF429	0.3209447873799725 6	0.0	0.3209447873799725 6	0.0	0.0	0.0	0.0	0.0
Rho guanine nucleotide exchange factor 10	ARHGEF10	0.3206983024691358 3	0.0	0.3206983024691358 3	0.0	0.0	0.0	0.0	0.0
PMS1 homolog 1, mismatch repair system component	PMS1	0.3206983024691358 3	0.0	0.3206983024691358 3	0.0	0.0	0.0	0.0	0.0
ezrin	EZR	0.3200820277813613 6	0.0	0.2777777777777778	0.0	0.0	0.0034565459502562 91	0.1676807573697758	0.0
tripartite motif containing 31	TRIM31	0.3199821227258289 3	0.3025713146564457 7	0.0	0.0	0.0	0.0048222726244483 43	0.0675	0.0
family with sequence similarity 135 member B	FAM135B	0.3199588477366255	0.0	0.3199588477366255	0.0	0.0	0.0	0.0	0.0
APOBEC1 complementation factor	A1CF	0.3199141944444447	0.0	0.3107638888888889	0.0	0.0	0.0	0.036601222222222222	0.0
septin 9	SEPTIN9	0.3178138888888889	0.0	0.3107638888888889	0.0	0.0	0.0	0.028200000000000003	0.0
non-POU domain containing octamer binding	NONO	0.3175175925925926	0.0	0.3113425925925926	0.0	0.0	0.0	0.0247	0.0
mitogen-activated protein kinase 1	MAPK1	0.3159232570820043	0.0	0.2777777777777778	0.0	0.0	0.0	0.1525819172 1690614	0.0
Rho guanine nucleotide exchange factor 10 like	ARHGEF10L	0.3158198412698413	0.0	0.3100198412698413	0.0	0.0	0.0	0.0232	0.0
GLI family zinc finger 1	GLI1	0.3150847447412313	0.0	0.2777777777777778	0.0	0.0	0.0	0.1492278678 5381423	0.0
bone morphogenetic protein 5	BMP5	0.3150477180213157	0.0	0.2777777777777778	0.0	0.0	0.1490797609 7415157	0.0	0.0
elastin	ELN	0.3147002811979554	0.0	0.2777777777777778	0.0	0.0	0.0197895849 22263306	0.1388946426 0414908	0.0
fibulin 2	FBLN2	0.3146479929066945 4	0.0	0.3104166666666667	0.0	0.0	0.0169253049 60111465	0.0	0.0
kallikrein related peptidase 2	KLK2	0.3134425925925926	0.0	0.3113425925925926	0.0	0.0	0.0	0.0084000000000000001	0.0
thyroid hormone receptor interactor 11	TRIP11	0.3131537313059805	0.0	0.3101851851851852	0.0	0.0	0.0118741844831 81251	0.0	0.0
hes related family bHLH transcription factor with YRPW motif 1	HEY1	0.3126044166666666 6	0.0	0.2777777777777778	0.0	0.0	0.0	0.1393065555555555 57	0.0
eukaryotic translation initiation factor 4A2	EIF4A2	0.3125	0.0	0.3125	0.0	0.0	0.0	0.0	0.0
calcium activated nucleotidase 1	CANT1	0.3115305273662208 3	0.0	0.3101851851851852	0.0	0.0	0.0053813687241426175	0.0	0.0
Mab-21 domain containing 2	MB21D2	0.3113425925925926	0.0	0.3113425925925926	0.0	0.0	0.0	0.0	0.0
nuclear mitotic apparatus protein 1	NUMA1	0.3113425925925926	0.0	0.3113425925925926	0.0	0.0	0.0	0.0	0.0
ubiquitin specific peptidase 44	USP44	0.3113425925925926	0.0	0.3113425925925926	0.0	0.0	0.0	0.0	0.0
EMAP like 4	EML4	0.3107638888888889	0.0	0.3107638888888889	0.0	0.0	0.0	0.0	0.0
transcription factor 12	TCF12	0.3107638888888889	0.0	0.3107638888888889	0.0	0.0	0.0	0.0	0.0

Table S1: (cont.)

golgin A5	GOLGA5	0.3107638888888889	0.0	0.31076388888888889	0.0	0.0	0.0	0.0	0.0
forkhead box O3	FOXO3	0.3104980510991434	0.0	0.27777777777777778	0.0	0.0	0.0	0.13088109328546238	0.0
dynactin subunit 1	DCTN1	0.3104166666666667	0.0	0.31041666666666667	0.0	0.0	0.0	0.0	0.0
DDB1 and CUL4 associated factor 12 like 2	DCAF12L2	0.3104166666666667	0.0	0.31041666666666667	0.0	0.0	0.0	0.0	0.0
A-Raf proto-oncogene, serine/threonine kinase	ARAF	0.3104166666666667	0.0	0.31041666666666667	0.0	0.0	0.0	0.0	0.0
growth arrest specific 7	GAS7	0.3104166666666667	0.0	0.31041666666666667	0.0	0.0	0.0	0.0	0.0
FGFR1 oncogene partner	FGFR1OP	0.3101851851851852	0.0	0.3101851851851852	0.0	0.0	0.0	0.0	0.0
regulator of G protein signaling 7	RGS7	0.3101851851851852	0.0	0.3101851851851852	0.0	0.0	0.0	0.0	0.0
ninein	NIN	0.3100198412698413	0.0	0.31001984126698413	0.0	0.0	0.0	0.0	0.0
nucleoporin 214	NUP214	0.3097993827160494	0.0	0.30979938271760494	0.0	0.0	0.0	0.0	0.0
glutamate metabotropic receptor 3	GRM3	0.3096064814814815	0.0	0.30960648148814815	0.0	0.0	0.0	0.0	0.0
contactin associated protein like 2	CNTNAP2	0.3096064814814815	0.0	0.30960648148814815	0.0	0.0	0.0	0.0	0.0
SET nuclear proto-oncogene	SET	0.2777777777777778	0.0	0.30847777777777778	0.0	0.0	0.0	0.1228	0.0
CCR4-NOT transcription complex subunit 1	CNOT1	0.1388888888888889	0.0	0.3068736384145513	0.0	0.2721514161923291	0.0	0.0	0.0
IL2 inducible T cell kinase	ITK	0.2777777777777778	0.1	0.30528888888888894	0.0	0.0	0.0	0.022600000000000006	0.0
moesin	MSN	0.2777777777777778	0.0	0.30362152777777778	0.0	0.0	0.0	0.1033750000000001	0.0
BTG anti-proliferation factor 1	BTG1	0.2777777777777778	0.0	0.3010273888888889	0.0	0.0	0.0	0.09299844444444444	0.0
cyclin E1	CCNE1	0.2777777777777778	0.0	0.3009519296825389	0.0	0.0	0.07252	0.045397367142850023	0.0
Jun proto-oncogene, AP-1 transcription factor subunit	JUN	0.2777777777777778	0.0	0.3001662411790969 6	0.0	0.0	0.04656251592927088	0.06885940208115644	0.0
homeobox A9	HOXA9	0.2777777777777778	0.0	0.3001441666666666	0.0	0.0	0.0	0.08946555555555558	0.0
C-C motif chemokine receptor 7	CCR7	0.2777777777777778	0.0	0.3000963810426 6984	0.0	0.0	0.009956824866057742	0.08484915756354251	0.0
serum/glucocorticoid regulated kinase 1	SGK1	0.2777777777777778	0.0	0.2993577991303498	0.0	0.0	0.0076491921731484365	0.08292044444444445	0.0
caudal type homeobox 2	CDX2	0.2777777777777778	0.0	0.2990950366591081	0.0	0.0	0.0	0.08526903552532122	0.0
paired related homeobox 1	PRRX1	0.2777777777777778	0.0	0.2988570277777778	0.0	0.0	0.0	0.08431699999999999	0.0
WW domain containing transcription regulator 1	WWTR1	0.2777777777777778	0.0	0.2985555555555556	0.0	0.0	0.0	0.0831111111111111	0.0
beta-2-microglobulin	B2M	0.2777777777777778	0.0	0.2971180962503971 5	0.0	0.0	0.0	0.07736127389047734	0.0
thyroid stimulating hormone receptor	TSHR	0.2777777777777778	0.0	0.2967521801050032	0.0	0.0	0.009669620945028422	0.07160000000000001	0.0
MDM4 regulator of p53	MDM4	0.2777777777777778	0.0	0.2961560179988662	0.0	0.0	0.0	0.07351296088435375	0.0
platelet derived growth factor subunit B	PDGFB	0.2777777777777778	0.0	0.2958027777777778	0.0	0.0	0.0	0.0721	0.0
filamin A	FLNA	0.2777777777777778	0.0	0.2953801111111111	0.0	0.0	0.0	0.07040933333333335	0.0
poly(A) binding protein cytoplasmic 1	PABPC1	0.2777777777777778	0.0	0.2938126303854875 5	0.0	0.0	0.0	0.06413941043083901	0.0
BCL10 immune signaling adaptor	BCL10	0.2777777777777778	0.0	0.2931444444444443	0.0	0.0	0.0	0.06146666666666666	0.0
interleukin 21 receptor	IL21R	0.2777777777777778	0.0	0.2930777777777777	0.0	0.0	0.0	0.061200000000000004	0.0

Table S1: (cont.)

homeobox D13	HOXD13	0.292862427512434	0.0	0.27777777777777778	0.0	0.0	0.06033859893862489	0.0	0.0
QKI, KH domain containing RNA binding	QKI	0.2924725421120653	0.0	0.27777777777777778	0.0	0.0	0.0038028790085876314	0.0570888888888889	0.0
guanine monophosphate synthase	GMPS	0.2924422983563337	0.0	0.27777777777777778	0.0	0.0	0.010393103574349988	0.05403892517006803	0.0
protein kinase cAMP-dependent type I regulatory subunit alpha	PRKAR1A	0.29204539909297056	0.0	0.27777777777777778	0.0	0.0	0.0	0.05707048526077098	0.0
poly(rC) binding protein 1	PCBP1	0.29165	0.0	0.27777777777777778	0.0	0.0	0.0	0.055488888888888886	0.0
R-spondin 2	RSPO2	0.2908777777777778	0.0	0.27777777777777778	0.0	0.0	0.0	0.0524	0.0
cyclin D2	CCND2	0.2901925316111336	0.0	0.27777777777777778	0.0	0.0	0.0	0.04965901533423286	0.0
paired box 8	PAX8	0.2900917500000000004	0.0	0.27777777777777778	0.0	0.0	0.0	0.0492558888888889	0.0
DNA polymerase gamma, catalytic subunit	POLG	0.2899777777777778	0.0	0.27777777777777778	0.0	0.0	0.0	0.0488	0.0
glypican 5	GPC5	0.2898777777777778	0.0	0.27777777777777778	0.0	0.0	0.0	0.0484	0.0
ribonuclease A family member 1, pancreatic	RNASE1	0.2897160555555556	0.0	0.27777777777777778	0.0	0.0	0.0	0.047753111111111114	0.0
prostate transmembrane protein, androgen induced 1	PMEPA1	0.28966181669903085	0.0	0.27777777777777778	0.0	0.0	0.0034563502912772024	0.0460000000000000006	0.0
GATA binding protein 1	GATA1	0.2890027777777778	0.0	0.27777777777777778	0.0	0.0	0.0	0.0448999999999999995	0.0
class II major histocompatibility complex transactivator	CIITA	0.2887944444444446	0.0	0.27777777777777778	0.0	0.0	0.0	0.0440666666666666667	0.0
ETS variant 5	ETV5	0.2887277777777778	0.0	0.27777777777777778	0.0	0.0	0.0	0.0438000000000000006	0.0
nuclear factor I B	NFIB	0.2882777777777778	0.0	0.27777777777777778	0.0	0.0	0.0	0.042	0.0
lymphocyte cytosolic protein 1	LCP1	0.2879027777777778	0.0	0.27777777777777778	0.0	0.0	0.0	0.0405	0.0
FA complementation group F	FANCF	0.28782916285152715	0.0	0.27777777777777778	0.0	0.0	0.012162465663744419	0.0348	0.0
roundabout guidance receptor 2	ROBO2	0.2874256990374347 6	0.0	0.27777777777777778	0.0	0.0	0.0213562913369127 54	0.0291000000000000004	0.0
transcription elongation factor A1	TCEA1	0.2869777777777778	0.0	0.27777777777777778	0.0	0.0	0.0	0.0368	0.0
MPL proto-oncogene, thrombopoietin receptor	MPL	0.2864523333333333336	0.0	0.27777777777777778	0.0	0.0	0.0	0.0346982222222222222	0.0
embryonic ectoderm development	EED	0.2862444444444444447	0.0	0.27777777777777778	0.0	0.0	0.0	0.0338666666666666667	0.0
ras homolog family member A	RHOA	0.2858305555555556	0.0	0.27777777777777778	0.0	0.0	0.0	0.0322111111111111114	0.0
GATA binding protein 3	GATA3	0.2857222222222222	0.0	0.27777777777777778	0.0	0.0	0.0	0.0317777777777777778	0.0
cyclin B1 interacting protein 1	CCNB1IP1	0.2838638888888889	0.0	0.27777777777777778	0.0	0.0	0.0	0.0243444444444444445	0.0
transcription factor EB	TFEB	0.2834527777777778	0.0	0.27777777777777778	0.0	0.0	0.0	0.0227	0.0
cAMP responsive element binding protein 3 like 2	CREB3L2	0.2827777777777778	0.0	0.27777777777777778	0.0	0.0	0.0	0.0200000000000000004	0.0
KIAA1549	KIAA1549	0.2827777777777778	0.0	0.27777777777777778	0.0	0.0	0.0	0.0200000000000000004	0.0
forkhead box L2	FOXL2	0.2823777777777778	0.0	0.27777777777777778	0.0	0.0	0.0	0.0184	0.0
DGCR8 microprocessor complex subunit	DGCR8	0.2820138888888889	0.0	0.27777777777777778	0.0	0.0	0.0	0.0169444444444444445	0.0
ASPSCR1 tether for SLC2A4, UBX domain containing	ASPSCR1	0.2818777777777778	0.0	0.27777777777777778	0.0	0.0	0.0	0.0163999999999999998	0.0

Table S1: (cont.)

H3 histone family member 3B	H3F3B	0.2816501640261795	0.0	0.2777777777777778	0.0	0.0	0.015489544993606847	0.0	0.0
SIX homeobox 2	SIX2	0.2814777777777777	0.0	0.2777777777777778	0.0	0.0	0.0	0.014800000000000002	0.0
RAD17 checkpoint clamp loader component	RAD17	0.2813777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.014400000000000003	0.0
lon peptidase 1, mitochondrial	LONP1	0.2809519275406444 6	0.0	0.2777777777777778	0.0	0.0	0.012696599051466707	0.0	0.0
cytokine receptor like factor 2	CRLF2	0.2807548419001660	0.0	0.2777777777777778	0.0	0.0	0.011908256489552 71	0.0	0.0
nuclear receptor coactivator 1	NCOA1	0.2803971798783511	0.0	0.1736111111111111	0.0	0.0	0.010434001558221405	0.2358350685941043	0.0
damage specific DNA binding protein 2	DDB2	0.2798527777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0083	0.0
tripartite motif containing 27	TRIM27	0.2797777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.008	0.0
PC4 and SFRS1 interacting protein 1	PSIP1	0.2794967118947417 3	0.0	0.2777777777777778	0.0	0.0	0.006875736467855812	0.0	0.0
CREB regulated transcription coactivator 3	CRTC3	0.2794777777777777	0.0	0.2777777777777778	0.0	0.0	0.0	0.006800000000000005	0.0
FES proto-oncogene, tyrosine kinase	FES	0.2788649710405803 3	0.0	0.2777777777777778	0.0	0.0	0.004348773051210141	0.0	0.0
LSM14A mRNA processing body assembly factor	LSM14A	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
bromodomain adjacent to zinc finger domain 1A	BAZ1A	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
family with sequence similarity 47 member C	FAM47C	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
myosin VA	MYO5A	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
nascent polypeptide associated complex subunit alpha	NACA	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
SLIT-ROBO Rho GTPase activating protein 3	SRGAP3	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
formin binding protein 1	FNBP1	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
TAL bHLH transcription factor 2	TAL2	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
zinc finger protein 479	ZNF479	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
cyclic nucleotide binding domain containing 1	CNBD1	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
forkhead box R1	FOXR1	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
cleavage and polyadenylation factor I subunit 1	CLP1	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
MLLT3 super elongation complex subunit	MLLT3	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
PWWP domain containing 2A	PWWP2A	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
tropomyosin 4	TPM4	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
acyl-CoA synthetase long chain family member 6	ACSL6	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
Fc receptor like 4	FCRL4	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
RNA binding motif protein 15	RBM15	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
U2 small nuclear RNA auxiliary factor 1	U2AF1	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
family with sequence similarity 131 member B	FAM131B	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0

Table S1: (cont.)

solute carrier family 45 member 3	SLC45A3	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
Rho GTPase activating protein 26	ARHGAP26	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
factor interacting with PAPOLA and CPSF1	FIP1L1	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
ASXL transcriptional regulator 2	ASXL2	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
vav guanine nucleotide exchange factor 1	VAV1	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
methyl-CpG binding domain protein 1	MBD1	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
lysine acetyltransferase 7	KAT7	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
transmembrane protein 127	TMEM127	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
tec protein tyrosine kinase	TEC	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
myosin heavy chain 11	MYH11	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
homeobox D11	HOXD11	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
SBDS ribosome maturation factor	SBDS	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
septin 6	SEPTIN6	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
homeobox C11	HOXC11	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
centrosomal protein 89	CEP89	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
3-ketodihydrosphingosine reductase	KDSR	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
centriolin	CNTRL	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
striatin	STRN	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
CCR4-NOT transcription complex subunit 6	CNOT6	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
POU class 2 homeobox associating factor 1	POU2AF1	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein epsilon	YWHAE	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
La ribonucleoprotein domain family member 4B	LARP4B	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
LYL1 basic helix-loop-helix family member	LYL1	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
POZ/BTB and AT hook containing zinc finger 1	PATZ1	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
ELKS/RAB6-interacting/CAST family member 1	ERC1	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
NEDD4 binding protein 2	N4BP2	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
5'-nucleotidase, cytosolic II	NT5C2	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
phosphatidylinositol binding clathrin assembly protein	PICALM	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
cadherin 10	CDH10	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0
cullin 3	CUL3	0.2777777777777778	0.0	0.2777777777777778	0.0	0.0	0.0	0.0	0.0

Table S1: (cont.)

ankyrin 1	ANK1	0.27777777777777778	0.0	0.27777777777777778	0.0	0.0	0.0	0.0	0.0
BCL2 associated transcription factor 1	BCLAF1	0.27777777777777778	0.0	0.27777777777777778	0.0	0.0	0.0	0.0	0.0
paired box 7	PAX7	0.27777777777777778	0.0	0.27777777777777778	0.0	0.0	0.0	0.0	0.0
ETS variant 1	ETV1	0.27777777777777778	0.0	0.27777777777777778	0.0	0.0	0.0	0.0	0.0
LIM and SH3 protein 1	LASP1	0.27462496846853274	0.0	0.13888888888888889	0.0	0.0	0.0	0.23990274624631053	0.0
major histocompatibility complex, class II, DQ alpha 2	HLA-DQA2	0.27406162498721187	0.27406162498721187	0.0	0.0	0.0	0.0	0.0	0.0
intestine specific homeobox	ISX	0.27081555555555553	0.0	0.13888888888888889	0.0	0.0	0.12078	0.22267333333333333	0.0
SAMM50 sorting and assembly machinery component	SAMM50	0.26198282459096256	0.26198282459096256	0.0	0.0	0.0	0.0	0.0	0.0
AKT serine/threonine kinase 3	AKT3	0.24503114073654153	0.0	0.13888888888888889	0.2	0.0	0.0	0.09278026662887376	0.0
interleukin 2	IL2	0.2396224876216162	0.0	0.13888888888888889	0.0	0.0	0.0	0.12924238859454548	0.190540000000000001
C-C motif chemokine receptor 4	CCR4	0.23575221838122412	0.047919965431016995	0.0	0.1	0.0	0.0	0.2054277777777778	0.0
5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase	ATIC	0.23472222222222222	0.0	0.13888888888888889	0.0	0.0	0.0	0.2	0.0
structural maintenance of chromosomes 1A	SMC1A	0.23472222222222222	0.0	0.13888888888888889	0.0	0.0	0.0	0.2	0.0
cytochrome P450 family 26 subfamily B member 1	CYP26B1	0.21003433318879938	0.21003433318879938	0.0	0.0	0.0	0.0	0.0	0.0
cytoplasmic polyadenylation element binding protein 3	CPEB3	0.20064196994721709	0.0	0.13888888888888889	0.0	0.0	0.010677729524953 69	0.16473333333333334	0.0
transferrin receptor	TFRC	0.19168514882581333	0.0	0.13888888888888889	0.0	0.0	0.0	0.15696292660359112	0.0
cadherin 17	CDH17	0.18638899173579 61	0.0	0.13888888888888889	0.0	0.0	0.0	0.15166676951357388	0.0
carbohydrate sulfotransferase 11	CHST11	0.18591111111111111	0.0	0.17361111111111111	0.0	0.0	0.0	0.04920000000000001	0.0
protein tyrosine phosphatase non-receptor type 6	PTPN6	0.18427116666666665	0.0	0.17361111111111111	0.0	0.0	0.0	0.04264022222222222 5	0.0
crooked neck pre-mRNA splicing factor 1	CRNKL1	0.17729300276108187	0.0	0.17361111111111111	0.0	0.0	0.014727566599883066	0.0	0.0
transmembrane serine protease 2	TMPRSS2	0.17361111111111111	0.0	0.17361111111111111	0.0	0.0	0.0	0.0	0.0
bromodomain PHD finger transcription factor	BPTF	0.16164652777777778	0.0	0.13888888888888889	0.0	0.0	0.0	0.09103055555555557	0.0
caspase 9	CASP9	0.15989100590150712	0.0	0.13888888888888889	0.0	0.0	0.0	0.08400846805047293	0.0
BCR activator of RhoGEF and GTPase	BCR	0.15634947019905568	0.0	0.13888888888888889	0.0	0.0	0.0032452317915009628	0.0684	0.0
SRY-box 21	SOX21	0.15278888888888889	0.0	0.13888888888888889	0.0	0.0	0.0	0.05560000000000001	0.0
BCL2 like 12	BCL2L12	0.15228888888888889	0.0	0.13888888888888889	0.0	0.0	0.0	0.05360000000000001	0.0
flap structure-specific endonuclease 1	FEN1	0.15110350042 86339	0.0	0.13888888888888889	0.0	0.0	0.025556503857705064	0.0375	0.0
reticulon 4	RTN4	0.15094244444444443	0.0	0.13888888888888889	0.0	0.0	0.0	0.04821422222222214	0.0
JAZF zinc finger 1	JAZF1	0.15030555555555555	0.0	0.13888888888888889	0.0	0.0	0.0	0.045666666666666675	0.0

Table S1: (cont.)

cyclin C	CCNC	0.1501888888888889	0.0	0.13888888888888889	0.0	0.0	0.0	0.0452000000000000004	0.0
CD209 molecule	CD209	0.1488888888888889	0.0	0.13888888888888889	0.0	0.0	0.0	0.0400000000000000001	0.0
proline rich mitotic checkpoint control factor	PRCC	0.1463888888888889	0.0	0.13888888888888889	0.0	0.0	0.0	0.03	0.0
nuclear factor of activated T cells 4	NFATC4	0.1463888888888889	0.0	0.13888888888888889	0.0	0.0	0.0	0.03	0.0
inhibitor of DNA binding 3, HLH protein	ID3	0.1458888888888889	0.0	0.13888888888888889	0.0	0.0	0.0	0.0280000000000000004	0.0
Rho GTPase activating protein 35	ARHGAP35	0.1453388888888889	0.0	0.13888888888888889	0.0	0.0	0.0	0.0258000000000000003	0.0
nuclear receptor coactivator 4	NCOA4	0.1425888888888889	0.0	0.13888888888888889	0.0	0.0	0.0	0.0148000000000000002	0.0
SS18L1 subunit of BAF chromatin remodeling complex	SS18L1	0.1418888888888889	0.0	0.13888888888888889	0.0	0.0	0.0	0.012	0.0
kinetochore localized astrin (SPAG5) binding protein	KNSTRN	0.1400249749053657 8	0.0	0.13888888888888889	0.0	0.0	0.0045443440665907486	0.0	0.0
shootin 1	SHTN1	0.1397694425753442 7	0.0	0.13888888888888889	0.0	0.0	0.0035222147458215305	0.0	0.0
novel transcript	AC000093.1	0.1388888888888889	0.0	0.13888888888888889	0.0	0.0	0.0	0.0	0.0
histone cluster 1 H4 family member i	HIST1H4I	0.1388888888888889	0.0	0.13888888888888889	0.0	0.0	0.0	0.0	0.0
metastasis associated lung adenocarcinoma transcript 1	MALAT1	0.1388888888888889	0.0	0.13888888888888889	0.0	0.0	0.0	0.0	0.0
intersectin 1	ITSN1	0.1388888888888889	0.0	0.13888888888888889	0.0	0.0	0.0	0.0	0.0
centrosomal protein 290	CEP290	0.1388888888888889	0.0	0.13888888888888889	0.0	0.0	0.0	0.0	0.0
myelodysplastic syndrome 2 translocation associated	MDS2	0.1388888888888889	0.0	0.13888888888888889	0.0	0.0	0.0	0.0	0.0
NUT family member 2A	NUTM2A	0.1388888888888889	0.0	0.13888888888888889	0.0	0.0	0.0	0.0	0.0
hook microtubule tethering protein 3	HOOK3	0.1388888888888889	0.0	0.13888888888888889	0.0	0.0	0.0	0.0	0.0
cytochrome c oxidase subunit 6C	COX6C	0.1388888888888889	0.0	0.13888888888888889	0.0	0.0	0.0	0.0	0.0
WD repeat and coiled coil containing	WDCP	0.1388888888888889	0.0	0.13888888888888889	0.0	0.0	0.0	0.0	0.0
lamin A/C	LMNA	0.1388888888888889	0.0	0.13888888888888889	0.0	0.0	0.0	0.0	0.0
cysteinyl leukotriene receptor 2	CYSLTR2	0.1388888888888889	0.0	0.13888888888888889	0.0	0.0	0.0	0.0	0.0
S100 calcium binding protein A7	S100A7	0.1388888888888889	0.0	0.13888888888888889	0.0	0.0	0.0	0.0	0.0
SS18 subunit of BAF chromatin remodeling complex	SS18	0.1388888888888889	0.0	0.13888888888888889	0.0	0.0	0.0	0.0	0.0
myosin heavy chain 10	MYH10	0.1388888888888889	0.0	0.13888888888888889	0.0	0.0	0.0	0.0	0.0
microtubule actin crosslinking factor 1	MACF1	0.1388888888888889	0.0	0.13888888888888889	0.0	0.0	0.0	0.0	0.0
FKBP prolyl isomerase 1B	FKBP1B	0.1388888888888889	0.0	0.13888888888888889	0.0	0.0	0.0	0.0	0.0
ASH1 like histone lysine methyltransferase	ASH1L	0.1388888888888889	0.0	0.13888888888888889	0.0	0.0	0.0	0.0	0.0

Table S1: (cont.)

BAF chromatin remodeling complex subunit BCL7A	BCL7A	0.13888888888889	0.0	0.13888888888889	0.0	0.0	0.0	0.0	0.0
TCF3 fusion partner	TFPT	0.13888888888889	0.0	0.13888888888889	0.0	0.0	0.0	0.0	0.0
leptin receptor overlapping transcript like 1	LEPROTL1	0.13888888888889	0.0	0.13888888888889	0.0	0.0	0.0	0.0	0.0
hedgehog acyltransferase	HHAT	0.13888888888889	0.0	0.13888888888889	0.0	0.0	0.0	0.0	0.0
homocysteine inducible ER protein with ubiquitin like domain 1	HERPUD1	0.13888888888889	0.0	0.13888888888889	0.0	0.0	0.0	0.0	0.0
rabaptin, RAB GTPase binding effector protein 1	RABEP1	0.13888888888889	0.0	0.13888888888889	0.0	0.0	0.0	0.0	0.0
tolloid like 1	TLL1	0.08057964666886383	0.05196303111 9899715	0.0	0.0	0.0	0.0	0.06758888888889	0.0
EF-hand calcium binding domain 11	EFCAB11	0.05686289581001867	0.05019724345 3977504	0.0	0.0	0.0	0.01139087120 4370464	0.02160000000 0000005	0.0
complement C2	C2	0.05299662172 6988755	0.05299662172 6988755	0.0	0.0	0.0	0.0	0.0	0.0
ventricular zone expressed PH domain containing 1	VEPH1	0.03259810842480203	0.03259810842480203	0.0	0.0	0.0	0.0	0.0	0.0
DLC1 Rho GTPase activating protein	DLC1	0.02192894729757386	0.01571578919 0295426	0.0	0.0	0.0	0.0	0.01800000000 0000002	0.0
novel transcript	AC013489.1	0.01241987712 2258983	0.01241987712 2258983	0.0	0.0	0.0	0.0	0.0	0.0
heparan sulfate-glucosamine 3-sulfotransferase 1	HS3ST1	0.01040845741 0619371	0.00955375163 250691	0.0	0.0	0.0	0.00341882311 24498434	0.0	0.0
aspartylglucosaminidase	AGA	0.00967782634 9288978	0.00967782634 9288978	0.0	0.0	0.0	0.0	0.0	0.0

Table S2: All Genes and Their Genetic Associations and Somatic Mutations Scores

Target gene	Target ID	Score.overall	Genetic association score	Somatic mutation score	Arithmetic mean
TP53	ENSG00000141510	1.0	1.0	1.0	1
CTNNB1	ENSG00000168036	1.0	1.0	1.0	1
TERT	ENSG00000164362	1.0	0.84	0.71	0.78
PIK3CA	ENSG00000121879	1.0	0.5	1.0	0.75
AXIN1	ENSG00000103126	1.0	0.0	1.0	0.5
MET	ENSG00000105976	1.0	0.0	0.99	0.50
NFE2L2	ENSG00000116044	0.99	0.0	0.97	0.49
IDH1	ENSG00000138413	0.98	0.0	0.96	0.48
HRAS	ENSG00000174775	0.97	0.0	0.95	0.47
CDKN2A	ENSG00000147889	0.96	0.0	0.94	0.47
CREBBP	ENSG00000005339	0.94	0.0	0.93	0.46
NRAS	ENSG00000213281	0.93	0.0	0.92	0.46
IDH2	ENSG00000182054	0.93	0.0	0.91	0.46
ARID2	ENSG00000189079	0.90	0.0	0.87	0.43
GNAS	ENSG00000087460	0.86	0.0	0.86	0.43
KRAS	ENSG00000133703	0.83	0.0	0.82	0.41
FGFR1	ENSG00000077782	1.0	0.0	0.80	0.40
TSC2	ENSG00000103197	0.82	0.0	0.78	0.39
SF3B1	ENSG00000115524	0.79	0.0	0.76	0.38
APC	ENSG00000134982	0.86	0.0	0.75	0.37
HLA-DQA1	ENSG00000196735	0.72	0.72	0.0	0.36
IL6ST	ENSG00000134352	0.76	0.0	0.71	0.36
ARID1A	ENSG00000117713	0.73	0.0	0.70	0.35
IGF2R	ENSG00000197081	0.72	0.0	0.69	0.35
RXRA	ENSG00000186350	0.71	0.0	0.69	0.35
CNOT9	ENSG00000144580	0.70	0.0	0.69	0.35
SF3B2	ENSG00000087365	0.68	0.0	0.68	0.34
CASP8	ENSG00000064012	0.70	0.0	0.68	0.34
ACVR2A	ENSG00000121989	0.67	0.0	0.67	0.33
NF1	ENSG00000196712	0.67	0.0	0.67	0.33
KMT2C	ENSG00000055609	0.67	0.0	0.66	0.33
PTEN	ENSG00000171862	0.71	0.0	0.66	0.33
ATRX	ENSG00000085224	0.66	0.0	0.66	0,33
KIF1B	ENSG00000054523	0.70	0.65	0.0	0.33
PBRM1	ENSG00000163939	0.66	0.0	0.65	0.33

Table S2: (cont.)

PRKACA	ENSG00000072062	0.63	0.0	0.63	0.31
HNF1A	ENSG00000135100	0.66	0.0	0.59	0.30
DNAJB1	ENSG00000132002	0.61	0.0	0.59	0.30
TBL1XR1	ENSG00000177565	0.60	0.0	0.59	0.29
MAGI1	ENSG00000151276	0.57	0.0	0.57	0.28
ARID1B	ENSG00000049618	0.58	0.0	0.57	0.28
BRCA2	ENSG00000139618	0.59	0.0	0.56	0.28
LRP1B	ENSG00000168702	0.56	0.0	0.56	0.28
TSC1	ENSG00000165699	0.60	0.0	0.56	0.28
ATM	ENSG00000149311	0.66	0.0	0.56	0.28
HLA-DQB1	ENSG00000179344	0.58	0.56	0.0	0.28
NUP98	ENSG00000110713	0.56	0.0	0.56	0.28
ROS1	ENSG00000047936	0.56	0.0	0.56	0.28
RB1	ENSG00000139687	0.60	0.0	0.56	0.28
AKT1	ENSG00000142208	0.65	0.0	0.56	0.28
SETD2	ENSG00000181555	0.62	0.0	0.56	0.28
KDM6A	ENSG00000147050	0.56	0.0	0.55	0.28
FLT3	ENSG00000122025	1.0	0.0	0.55	0.28
BLM	ENSG00000197299	0.57	0.0	0.55	0.28
SMARCA4	ENSG00000127616	0.65	0.0	0.55	0.28
NCOR1	ENSG00000141027	0.55	0.0	0.55	0.27
SUZ12	ENSG00000178691	0.56	0.0	0.54	0.27
KMT2A	ENSG00000118058	0.56	0.0	0.54	0.27
KMT2D	ENSG00000167548	0.54	0.0	0.54	0.27
AXIN2	ENSG00000168646	0.57	0.0	0.54	0.27
MECOM	ENSG00000085276	0.55	0.0	0.53	0.27
ANTXR1	ENSG00000169604	0.56	0.0	0.53	0.27
AMER1	ENSG00000184675	0.55	0.0	0.53	0.27
KEAP1	ENSG00000079999	0.54	0.0	0.53	0.27
FLT4	ENSG00000037280	1.0	0.0	0.53	0.27
KIT	ENSG00000157404	1.0	0.0	0.53	0.27
FZR1	ENSG00000105325	0.53	0.0	0.53	0.27
AL031847.2	ENSG00000285629	0.53	0.0	0.53	0.27
TCF7L2	ENSG00000148737	0.71	0.0	0.53	0.27
NOTCH1	ENSG00000148400	0.57	0.0	0.53	0.26
FAT4	ENSG00000196159	0.54	0.0	0.52	0.26
LATS1	ENSG00000131023	0.57	0.0	0.52	0.26

Table S2: (cont.)

ASXL1	ENSG00000171456	0.52	0.0	0.52	0.26
ERBB3	ENSG00000065361	0.60	0.0	0.52	0.26
BCOR	ENSG00000183337	0.53	0.0	0.52	0.26
CIC	ENSG00000079432	0.53	0.0	0.52	0.26
TRRAP	ENSG00000196367	0.58	0.0	0.51	0.26
MEN1	ENSG00000133895	0.53	0.0	0.51	0.26
MYH9	ENSG00000100345	0.51	0.0	0.51	0.26
PTPRB	ENSG00000127329	0.56	0.0	0.51	0.26
ZFHX3	ENSG00000140836	0.53	0.0	0.51	0.26
NCOR2	ENSG00000196498	0.51	0.0	0.51	0.26
DICER1	ENSG00000100697	0.52	0.0	0.51	0.26
COL2A1	ENSG00000139219	0.51	0.0	0.51	0.26
MAP3K1	ENSG00000095015	0.52	0.0	0.51	0.26
CACNA1D	ENSG00000157388	0.51	0.0	0.51	0.26
CLTC	ENSG00000141367	0.52	0.0	0.51	0.25
NF2	ENSG00000186575	0.53	0.0	0.51	0.25
TBX3	ENSG00000135111	0.53	0.0	0.51	0.25
KDM5A	ENSG00000073614	0.52	0.0	0.51	0.25
RANBP2	ENSG00000153201	0.51	0.0	0.51	0.25
KDR	ENSG00000128052	1.0	0.0	0.51	0.25
MSH6	ENSG00000116062	0.52	0.0	0.51	0.25
ALK	ENSG00000171094	0.56	0.0	0.51	0.25
JAK3	ENSG00000105639	0.51	0.0	0.51	0.25
SMAD4	ENSG00000141646	0.57	0.0	0.51	0.25
RET	ENSG00000165731	1.0	0.0	0.50	0.25
PTPRK	ENSG00000152894	0.51	0.0	0.50	0.25
WT1	ENSG00000184937	0.56	0.0	0.50	0.25
ESR1	ENSG00000091831	0.95	0.0	0.50	0.25
MYB	ENSG00000118513	0.52	0.0	0.50	0.25
XPO1	ENSG00000082898	0.58	0.0	0.50	0.25
SMAD2	ENSG00000175387	0.54	0.0	0.50	0.25
FLCN	ENSG00000154803	0.51	0.0	0.50	0.25
NTRK1	ENSG00000198400	0.56	0.0	0.50	0.25

Table S3: Drug Target Interactions After MDeePred

COMPOUND ID	TARGET ID
CHEMBL1241270	CHEMBL4005
CHEMBL3763833	CHEMBL2148
CHEMBL1241490	CHEMBL4005
CHEMBL3895671	CHEMBL279
CHEMBL4215501	CHEMBL3717
CHEMBL4112408	CHEMBL4247
CHEMBL2418345	CHEMBL4005
CHEMBL3678511	CHEMBL3717
CHEMBL3671818	CHEMBL2148
CHEMBL4291436	CHEMBL4005
CHEMBL452812	CHEMBL1955
CHEMBL494921	CHEMBL4005
CHEMBL596583	CHEMBL4282
CHEMBL206198	CHEMBL4282
CHEMBL3642255	CHEMBL2148
CHEMBL3648908	CHEMBL4005
CHEMBL3976418	CHEMBL4247
CHEMBL3940191	CHEMBL4247
CHEMBL3915565	CHEMBL2815
CHEMBL570573	CHEMBL279
CHEMBL3319466	CHEMBL1936
CHEMBL2152127	CHEMBL4005
CHEMBL2088106	CHEMBL3650
CHEMBL509435	CHEMBL1974
CHEMBL1760157	CHEMBL4005
CHEMBL3964812	CHEMBL279
CHEMBL2018882	CHEMBL1955
CHEMBL3661583	CHEMBL2815
CHEMBL3948623	CHEMBL279
CHEMBL3134298	CHEMBL4005
CHEMBL1922224	CHEMBL4247
CHEMBL3398170	CHEMBL4247
CHEMBL3286814	CHEMBL4247
CHEMBL230686	CHEMBL279
CHEMBL3753512	CHEMBL2041
CHEMBL4168420	CHEMBL4005

Table S3: (cont.)

CHEMBL3687221	CHEMBL2041
CHEMBL493083	CHEMBL4005
CHEMBL2402840	CHEMBL3717
CHEMBL1614778	CHEMBL4005
CHEMBL3670213	CHEMBL4005
CHEMBL222539	CHEMBL2916
CHEMBL3753903	CHEMBL2041
CHEMBL1170550	CHEMBL2041
CHEMBL4166436	CHEMBL3717
CHEMBL1098311	CHEMBL4005
CHEMBL2431839	CHEMBL3717
CHEMBL77242	CHEMBL3650
CHEMBL412081	CHEMBL279
CHEMBL3919045	CHEMBL279
CHEMBL352992	CHEMBL279
CHEMBL3967141	CHEMBL279
CHEMBL1940105	CHEMBL1974
CHEMBL196609	CHEMBL279
CHEMBL460989	CHEMBL2815
CHEMBL3754722	CHEMBL279
CHEMBL3401762	CHEMBL3650
CHEMBL246355	CHEMBL279
CHEMBL74143	CHEMBL279
CHEMBL3286808	CHEMBL4247
CHEMBL558306	CHEMBL3717
CHEMBL170984	CHEMBL279
CHEMBL3786830	CHEMBL4247
CHEMBL1242032	CHEMBL279
CHEMBL3644994	CHEMBL2148
CHEMBL493315	CHEMBL279
CHEMBL3286829	CHEMBL4247
CHEMBL2322671	CHEMBL4005
CHEMBL3644956	CHEMBL2148
CHEMBL3667474	CHEMBL2148
CHEMBL191594	CHEMBL279
CHEMBL516429	CHEMBL4282
CHEMBL3775190	CHEMBL279

Table S3: (cont.)

CHEMBL3695646	CHEMBL4005
CHEMBL185140	CHEMBL2148
CHEMBL3889789	CHEMBL2148
CHEMBL3358977	CHEMBL279
CHEMBL209535	CHEMBL2815
CHEMBL3798482	CHEMBL2148
CHEMBL3612512	CHEMBL4005
CHEMBL572956	CHEMBL4005
CHEMBL3113146	CHEMBL4282
CHEMBL2030439	CHEMBL4005
CHEMBL176705	CHEMBL279
CHEMBL1928728	CHEMBL4005
CHEMBL101797	CHEMBL279
CHEMBL3745943	CHEMBL279
CHEMBL3666286	CHEMBL2815
CHEMBL3951451	CHEMBL279
CHEMBL3343366	CHEMBL279
CHEMBL2030438	CHEMBL4005
CHEMBL328029	CHEMBL3650
CHEMBL2386972	CHEMBL4005
CHEMBL2323141	CHEMBL2916
CHEMBL1915448	CHEMBL1974
CHEMBL377771	CHEMBL4282
CHEMBL1795619	CHEMBL5747
CHEMBL3962203	CHEMBL279
CHEMBL576982	CHEMBL279
CHEMBL4164843	CHEMBL3717
CHEMBL3596879	CHEMBL279
CHEMBL3915857	CHEMBL2148
CHEMBL1946262	CHEMBL4005
CHEMBL101464	CHEMBL279
CHEMBL233454	CHEMBL279
CHEMBL3775086	CHEMBL279
CHEMBL4078141	CHEMBL1974
CHEMBL112914	CHEMBL3650
CHEMBL3577116	CHEMBL279
CHEMBL3114771	CHEMBL2148

Table S3: (cont.)

CHEMBL3687223	CHEMBL5568
CHEMBL4281394	CHEMBL5568
CHEMBL4166436	CHEMBL4247
CHEMBL4078075	CHEMBL1974
CHEMBL3814951	CHEMBL2148
CHEMBL3904441	CHEMBL1974
CHEMBL3703988	CHEMBL3717
CHEMBL4161643	CHEMBL5568
CHEMBL4211826	CHEMBL2815
CHEMBL2312304	CHEMBL3717
CHEMBL2332844	CHEMBL1974
CHEMBL564940	CHEMBL2148
CHEMBL205514	CHEMBL4282
CHEMBL1681958	CHEMBL279
CHEMBL1762781	CHEMBL4005
CHEMBL598852	CHEMBL3650
CHEMBL380944	CHEMBL279
CHEMBL4244708	CHEMBL1974
CHEMBL245966	CHEMBL279
CHEMBL2151932	CHEMBL4005
CHEMBL1090598	CHEMBL4005
CHEMBL1822974	CHEMBL3717
CHEMBL1796242	CHEMBL4247
CHEMBL1940275	CHEMBL279
CHEMBL1823254	CHEMBL3717
CHEMBL3747077	CHEMBL2148
CHEMBL4112318	CHEMBL4005
CHEMBL403661	CHEMBL1974
CHEMBL3814934	CHEMBL2148
CHEMBL3941123	CHEMBL4247
CHEMBL3329651	CHEMBL1974
CHEMBL3931691	CHEMBL279
CHEMBL516640	CHEMBL3956
CHEMBL147761	CHEMBL279
CHEMBL1644558	CHEMBL4005
CHEMBL3958497	CHEMBL1936
CHEMBL3416150	CHEMBL1974

Table S3: (cont.)

CHEMBL4094670	CHEMBL2041
CHEMBL3586401	CHEMBL4005
CHEMBL4171749	CHEMBL5747
CHEMBL3937128	CHEMBL3717
CHEMBL3398175	CHEMBL4247
CHEMBL598434	CHEMBL279
CHEMBL1241679	CHEMBL279
CHEMBL3642425	CHEMBL2148
CHEMBL55594	CHEMBL3650
CHEMBL3673444	CHEMBL2815
CHEMBL206410	CHEMBL3650
CHEMBL4063976	CHEMBL3650
CHEMBL106571	CHEMBL2148
CHEMBL4209753	CHEMBL1974
CHEMBL2322769	CHEMBL4005
CHEMBL2322773	CHEMBL4005
CHEMBL1202478	CHEMBL3650
CHEMBL469025	CHEMBL279
CHEMBL3642334	CHEMBL2148
CHEMBL2057368	CHEMBL4005
CHEMBL2151323	CHEMBL2815
CHEMBL3900059	CHEMBL279
CHEMBL3654300	CHEMBL5747
CHEMBL1242033	CHEMBL4005
CHEMBL3673470	CHEMBL2815
CHEMBL1165499	CHEMBL4247
CHEMBL4114735	CHEMBL3717
CHEMBL2035028	CHEMBL4282
CHEMBL3335371	CHEMBL1936
CHEMBL155151	CHEMBL2916
CHEMBL3290305	CHEMBL4005
CHEMBL1077842	CHEMBL2148
CHEMBL1773601	CHEMBL4282
CHEMBL1258331	CHEMBL4282
CHEMBL240336	CHEMBL4005
CHEMBL1240566	CHEMBL279
CHEMBL3693558	CHEMBL4282

Table S3: (cont.)

CHEMBL3890788	CHEMBL2148
CHEMBL3753157	CHEMBL279
CHEMBL524708	CHEMBL279
CHEMBL2012699	CHEMBL4282
CHEMBL1773581	CHEMBL4282
CHEMBL4163608	CHEMBL4282
CHEMBL1091714	CHEMBL4005
CHEMBL2158227	CHEMBL1936
CHEMBL252856	CHEMBL4282
CHEMBL1689442	CHEMBL2916
CHEMBL1290744	CHEMBL4005
CHEMBL400057	CHEMBL1936
CHEMBL2442139	CHEMBL3717
CHEMBL209774	CHEMBL1974
CHEMBL3813725	CHEMBL4005
CHEMBL3353355	CHEMBL279
CHEMBL3218849	CHEMBL4247
CHEMBL3770633	CHEMBL4005
CHEMBL4249784	CHEMBL2148
CHEMBL3896249	CHEMBL279
CHEMBL92516	CHEMBL279
CHEMBL199723	CHEMBL1974
CHEMBL1277620	CHEMBL2041
CHEMBL527066	CHEMBL3717
CHEMBL264382	CHEMBL279
CHEMBL3672558	CHEMBL4282
CHEMBL3977135	CHEMBL4005
CHEMBL255902	CHEMBL279
CHEMBL219886	CHEMBL279
CHEMBL4284396	CHEMBL2148
CHEMBL3938428	CHEMBL2815
CHEMBL2418949	CHEMBL4005
CHEMBL2017973	CHEMBL4005
CHEMBL3941879	CHEMBL2148
CHEMBL1241950	CHEMBL4005
CHEMBL4175060	CHEMBL5747
CHEMBL1204020	CHEMBL2916

Table S3: (cont.)

CHEMBL594217	CHEMBL4005
CHEMBL2064508	CHEMBL4005
CHEMBL4114743	CHEMBL3717
CHEMBL2152254	CHEMBL4005
CHEMBL3904401	CHEMBL4005
CHEMBL226331	CHEMBL4282
CHEMBL3668018	CHEMBL279
CHEMBL1940275	CHEMBL1974
CHEMBL3667416	CHEMBL4005
CHEMBL1762782	CHEMBL4005
CHEMBL2323145	CHEMBL2916
CHEMBL4167072	CHEMBL3717
CHEMBL2086748	CHEMBL279
CHEMBL365585	CHEMBL279
CHEMBL2312652	CHEMBL3650
CHEMBL4074115	CHEMBL4247
CHEMBL3897417	CHEMBL2148
CHEMBL4241638	CHEMBL279
CHEMBL3648080	CHEMBL279
CHEMBL271131	CHEMBL279
CHEMBL1242844	CHEMBL4005
CHEMBL3891894	CHEMBL279
CHEMBL3952048	CHEMBL279
CHEMBL3347489	CHEMBL2916
CHEMBL1770494	CHEMBL2916
CHEMBL1957870	CHEMBL4005
CHEMBL3221125	CHEMBL3717
CHEMBL154216	CHEMBL2916
CHEMBL1170702	CHEMBL2916
CHEMBL2071451	CHEMBL3717
CHEMBL4067036	CHEMBL3717
CHEMBL3980022	CHEMBL279
CHEMBL3972478	CHEMBL3797
CHEMBL3589738	CHEMBL279
CHEMBL216703	CHEMBL279
CHEMBL3402763	CHEMBL3717
CHEMBL188978	CHEMBL279

Table S3: (cont.)

CHEMBL218439	CHEMBL2148
CHEMBL589864	CHEMBL4005
CHEMBL3754301	CHEMBL279
CHEMBL3952637	CHEMBL4247
CHEMBL595373	CHEMBL279
CHEMBL516639	CHEMBL279
CHEMBL3736182	CHEMBL4247
CHEMBL2387079	CHEMBL4005
CHEMBL388978	CHEMBL1974
CHEMBL2064492	CHEMBL4005
CHEMBL1822798	CHEMBL3717
CHEMBL2381381	CHEMBL4005
CHEMBL3426917	CHEMBL1955
CHEMBL1173697	CHEMBL4282
CHEMBL1644571	CHEMBL4005
CHEMBL566010	CHEMBL3717
CHEMBL1823651	CHEMBL4282
CHEMBL3582356	CHEMBL4005
CHEMBL3655869	CHEMBL4005
CHEMBL3398166	CHEMBL4247
CHEMBL3423384	CHEMBL2916
CHEMBL1615189	CHEMBL4005
CHEMBL3656350	CHEMBL1974
CHEMBL3593764	CHEMBL2148
CHEMBL1242661	CHEMBL4005
CHEMBL597592	CHEMBL2041
CHEMBL3695624	CHEMBL4005
CHEMBL3932688	CHEMBL4282
CHEMBL1088857	CHEMBL2148
CHEMBL3600770	CHEMBL4005
CHEMBL3394477	CHEMBL3717
CHEMBL1645108	CHEMBL4005
CHEMBL4218103	CHEMBL2815
CHEMBL3674504	CHEMBL3717
CHEMBL4164803	CHEMBL5747
CHEMBL1807717	CHEMBL2041
CHEMBL3334568	CHEMBL3717

Table S3: (cont.)

CHEMBL13354	CHEMBL3650
CHEMBL1630542	CHEMBL4005
CHEMBL3115500	CHEMBL4247
CHEMBL4214476	CHEMBL279
CHEMBL2088256	CHEMBL3650
CHEMBL3653155	CHEMBL4005
CHEMBL3354188	CHEMBL279
CHEMBL3746341	CHEMBL2148
CHEMBL156248	CHEMBL2916
CHEMBL462707	CHEMBL3717
CHEMBL3883989	CHEMBL4005
CHEMBL2032035	CHEMBL3717
CHEMBL3334553	CHEMBL3717
CHEMBL3950207	CHEMBL3717
CHEMBL1081852	CHEMBL4282
CHEMBL2312649	CHEMBL4282
CHEMBL1083305	CHEMBL4005
CHEMBL3690510	CHEMBL279
CHEMBL2103868	CHEMBL3717
CHEMBL1822801	CHEMBL3717
CHEMBL3925815	CHEMBL2148
CHEMBL1241389	CHEMBL4005
CHEMBL3290148	CHEMBL2148
CHEMBL429410	CHEMBL279
CHEMBL3671823	CHEMBL2148
CHEMBL3577117	CHEMBL279
CHEMBL3218861	CHEMBL4247
CHEMBL2171124	CHEMBL1936
CHEMBL529663	CHEMBL1974
CHEMBL493911	CHEMBL4282
CHEMBL2037224	CHEMBL2815
CHEMBL3645547	CHEMBL2148
CHEMBL176078	CHEMBL279
CHEMBL3687195	CHEMBL4247
CHEMBL3098318	CHEMBL279
CHEMBL498769	CHEMBL3717
CHEMBL2087482	CHEMBL4005

Table S3: (cont.)

CHEMBL4204150	CHEMBL3717
CHEMBL4284541	CHEMBL1974
CHEMBL4165354	CHEMBL3717
CHEMBL1915640	CHEMBL1974
CHEMBL3948901	CHEMBL4005
CHEMBL383030	CHEMBL279
CHEMBL524057	CHEMBL279
CHEMBL3667472	CHEMBL2148
CHEMBL3134602	CHEMBL4005
CHEMBL3899361	CHEMBL4005
CHEMBL485215	CHEMBL4282
CHEMBL3600686	CHEMBL4005
CHEMBL2071276	CHEMBL279
CHEMBL462338	CHEMBL279
CHEMBL4082186	CHEMBL279
CHEMBL3128062	CHEMBL4247
CHEMBL4248386	CHEMBL2148
CHEMBL98350	CHEMBL4005
CHEMBL3309962	CHEMBL3717
CHEMBL3673459	CHEMBL2815
CHEMBL3793268	CHEMBL4005
CHEMBL3959170	CHEMBL4247
CHEMBL4103824	CHEMBL5747
CHEMBL470192	CHEMBL4282
CHEMBL3813701	CHEMBL5747
CHEMBL2442129	CHEMBL3717
CHEMBL3355074	CHEMBL3797
CHEMBL3221129	CHEMBL3717
CHEMBL379372	CHEMBL279
CHEMBL3601923	CHEMBL4005
CHEMBL2448066	CHEMBL279
CHEMBL3113144	CHEMBL4282
CHEMBL46785	CHEMBL279
CHEMBL2409780	CHEMBL279
CHEMBL398989	CHEMBL279
CHEMBL270847	CHEMBL4282
CHEMBL2323783	CHEMBL3717

Table S3: (cont.)

CHEMBL3645022	CHEMBL2148
CHEMBL3814770	CHEMBL4005
CHEMBL4097807	CHEMBL2916
CHEMBL3943168	CHEMBL3650
CHEMBL3687205	CHEMBL2041
CHEMBL3981577	CHEMBL279
CHEMBL3967044	CHEMBL4247
CHEMBL245948	CHEMBL279
CHEMBL3600784	CHEMBL4005
CHEMBL3644972	CHEMBL2148
CHEMBL3639478	CHEMBL4005

Table S4: Molecular Docking Results of Small Molecules-Transferases and Drugs-Transferases

Ligand (Drug/Compound)	Target Protein	Vina Score	Cavity Volume (Å^3)	Contact Residues
CHEMBL328029	FGFR1	-8.7	543	ILE19 GLN24 LYS51 LEU54 PHE55 GLY58 GLN59 ILE61 MET62 VAL75 PHE91 VAL93 HIS96 ILE99 TYR100
CHEMBL1165499	ALK	-9.6	6263	VAL131 ASP133 GLU135 VAL136 ASN426 ILE427 ASN428 MET441 ALA442 LEU443 TRP446 VAL461 THR462 GLY463 SER464 LYS468 LEU639 LYS640 GLU642 GLN643 LEU645 THR679 VAL680 SER681 GLN682 ARG683
CHEMBL1773601	AKT1	-7.8	110	LYS8 GLU9 GLY10 TRP11 LEU12 HIS13 PRO24 TYR26 ARG41 VAL90 GLU91 GLU95 TRP99
CHEMBL1773581	AKT1	-7.4	110	LYS8 GLU9 GLY10 TRP11 LEU12 HIS13 PRO24 HIS89 VAL90 GLU91 GLU95 TRP99
CHEMBL388978	FLT3	-9.6	1932	GLY1121 LEU1122 GLY1123 HIS1124 GLY1125 VAL1130 ALA1148 LYS1150 VAL1180 LEU1196 GLU1197 LEU1198 MET1199 GLY1202 ASP1203 ARG1253 ASN1254 LEU1256 GLY1269 ASP1270
CHEMBL1615189	PIK3CA	-9.6	904	LEU484 GLY485 VAL492 ALA512 LYS514 GLU531 MET535 ILE545 VAL561 GLU562 TYR563 ALA564 GLY567 LEU630 ALA640 ASP641
Lenvatinib	FGRF1	-9.1	8731	ARG576 ARG577 PRO578 LEU595 SER596 SER597 LEU600 TRP691 PHE694 THR695 LEU696 GLY698 SER699 TYR701 PRO702 HIS717 ARG718 MET719 ASP720 LYS721 PRO722 SER723 ASN724 TYR730 ARG577 PRO578 LEU595 LEU600 TRP691 PHE694 THR695 LEU696 GLY697 GLY698 SER699 PRO700 TYR701 PRO702 HIS717 ARG718 MET719 ASP720 LYS721 PRO722 SER723 ASN724 ASN727
Lenvatinib	ALK	-9.0	1932	ARG1120 LEU1122 GLY1123 HIS1124 GLY1125 ALA1126 VAL1130 GLU1132 ALA1148 VAL1149 LYS1150 VAL1180 LEU1196 GLU1197 LEU1198 MET1199 ALA1200 GLY1201 GLY1202 ASP1203 LYS1205 SER1206 ASP1249 ARG1253 ASN1254 CYS1255 LEU1256 GLY1269 ASP1270 GLY1272 MET1290

Table S4: (cont.)

Ligand (Drug/Compound)	Target Protein	Vina Score	Cavity Volume (Å^3)	Contact Residues
Lenvatinib	AKT1	-6.7	110	VAL7 LYS8 GLU9 GLY10 TRP11 LEU12 HIS13 PRO24 ARG25 TYR26 ARG41 HIS89 VAL90 GLU91 GLU94 GLU95 GLU98 TRP99
Lenvatinib	FLT3	-8.5	832	TYR572 GLU573 SER574 GLN575 TYR589 TYR591 PHE621 ALA657 ARG810 ASP811 ASN816 ASP829 PHE830 GLY831 LEU832 ARG834 ILE836 TYR842 ARG845 GLY846 ASN847 ALA848 ARG849 LEU850 PRO851 MET855 SER859 LEU860 PHE861 GLU862 GLY863 ILE864 TYR865
Lenvatinib	PIK3CA	-8.9	6263	GLU127 MET130 VAL131 LYS132 ASP133 PRO134 GLU135 VAL136 ASN426 ILE427 ASN428 PHE430 ASP431 TYR432 THR435 LEU436 VAL437 SER438 MET441 ALA442 LEU443 TRP446 VAL461 THR462 GLY463 SER464 ASN465 PRO466 LYS468 LYS640 GLU642 GLN643 TYR644 LEU645 THR679 VAL680 GLN682 ARG683
Regorafenib	FGRF1	-9.9	8731	GLN574 ARG577 PRO578 TRP691 PHE694 THR695 LEU696 GLY698 SER699 PRO700 TYR701 PRO702 VAL704 HIS717 ARG718 MET719 ASP720 LYS721 PRO722 SER723 ASN724 TYR730 ARG577 PRO578 LEU595 SER597 LEU600 TRP691 PHE694 THR695 LEU696 GLY697 GLY698 SER699 PRO700 TYR701 PRO702 VAL704 LEU712 GLU715 GLY716 HIS717 ARG718 MET719 ASP720 LYS721 PRO722 SER723 ASN724 ARG734
Regorafenib	ALK	-9.6	1932	ARG1120 GLY1121 LEU1122 GLY1123 HIS1124 GLY1125 ALA1126 VAL1130 GLU1132 GLN1146 ALA1148 LYS1150 VAL1180 LEU1196 GLU1197 LEU1198 MET1199 ALA1200 GLY1201 GLY1202 ASP1203 LYS1205 SER1206 ASP1249 ARG1253 ASN1254 CYS1255 LEU1256 GLY1269 ASP1270 GLY1272 MET1273 MET1290
Regorafenib	AKT1	-7.2	110	ILE6 VAL7 LYS8 GLU9 GLY10 TRP11 LEU12 HIS13 PRO24 ARG25 TYR26 LYS39 GLU40 ARG41 HIS89 VAL90 GLU91 GLU95 GLU98 TRP99 THR101 ALA102 THR105

Table S4: (cont.)

Ligand (Drug/Compound)	Target Protein	Vina Score	Cavity Volume (Å^3)	Contact Residues
Regorafenib	FLT3	-8.6	832	TYR572 GLU573 SER574 GLN575 LEU576 GLN577 MET578 TYR589 TYR591 VAL592 ASP593 PHE594 ARG595 PHE621 LEU646 ARG655 GLU656 ALA657 SER660 GLU661 MET664 ARG810 ASP811 ASN816 ASP829 PHE830 GLY831 LEU832 ARG834 ILE836 TYR842 ASN847 ALA848 ARG849 LEU850 PRO851 MET855 SER859 LEU860 GLU862 GLY863 TYR865
Regorafenib	PIK3CA	-10.0	2313	TYR165 VAL166 TYR167 PRO168 ASN170 VAL196 ILE197 TYR250 LYS253 VAL254 CYS257 ASP258 GLU259 TYR260 LYS271 TYR272 SER275 MET286 LEU287 MET288 ALA289 SER292 SER295 GLN296 LEU297 PRO298 GLN661 ARG662 HIS665 PHE666 MET697 TYR698 HIS701 GLY750 PHE751 LEU752 ASN756 PRO757 ALA758 HIS759 GLN760 LEU761 GLY762 PRO786 ASP787 ILE788 LEU793 PHE794
Sorafenib	FGRF1	-9.9	348	LEU484 GLY485 GLU486 GLY487 ALA488 PHE489 GLY490 GLN491 VAL492 ALA512 VAL513 LYS514 MET515 LEU516 ASP524 ASP527 LEU528 GLU531 MET535 ILE545 VAL559 VAL561 GLU562 TYR563 ALA564 GLY567 ASN568 ARG570 GLU571 ARG627 ASN628 LEU630 ILE639 ALA640 ASP641 PHE642 LEU644 ALA645 THR657 THR658 ASN659
Sorafenib	ALK	-8.9	1932	ARG1120 LEU1122 GLY1123 HIS1124 GLY1125 ALA1126 VAL1130 GLU1132 ALA1148 VAL1149 LYS1150 VAL1180 LEU1196 GLU1197 LEU1198 MET1199 ALA1200 GLY1202 ASP1203 LYS1205 SER1206 GLU1210 ASP1249 ARG1253 ASN1254 LEU1256 GLY1269 ASP1270 GLY1272 MET1273
Sorafenib	AKT1	-7.0	152	LYS14 ARG15 GLY16 GLU17 TYR18 ILE19 LYS20 ARG23 LEU52 ASN53 ASN54 PHE55 THR65 GLU66 ARG67 PRO68 THR72 ILE74 ARG76 GLN79 THR82 VAL83 ILE84 GLU85 ARG86 THR87

Table S4: (cont.)

Ligand (Drug/Compound)	Target Protein	Vina Score	Cavity Volume (Å^3)	Contact Residues
Sorafenib	FLT3	-10.3	832	TYR572 GLU573 SER574 GLN575 LEU576 GLN577 MET578 TYR591 VAL592 ASP593 PHE594 ARG595 PHE621 GLU656 ALA657 SER660 GLU661 MET664 ARG810 ASP811 ASN816 ASP829 PHE830 GLY831 LEU832 ARG834 ILE836 TYR842 ARG845 GLY846 ASN847 ALA848 ARG849 LEU850 PRO851 MET855 SER859 LEU860 GLU862 GLY863 ILE864 TYR865
Sorafenib	PIK3CA	-10.3	2313	TYR165 VAL166 TYR167 PRO168 PRO169 ASN170 ASP258 GLU259 TYR260 MET288 SER292 LEU293 GLN296 LEU297 PRO298 ASP300 GLN661 ARG662 HIS665 CYS695 GLY696 MET697 TYR698 LYS700 HIS701 GLY750 PHE751 LEU752 ASN756 PRO757 ALA758 GLN760

Table S5: Molecular Docking Results of Small Molecules and Kinase Domains of All Transferases

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
CHEMBL328029	FGFR1	-10.5	1933	Chain B: LEU484 GLY485 ALA488 PHE489 VAL492 ALA512 LYS514 GLU531 MET535 ILE545 VAL559 VAL561 GLU562 TYR563 ALA564 SER565 LYS566 GLY567 ASN568 ARG570 GLU571 ASP623 ARG627 ASN628 LEU630 ILE639 ALA640 ASP641 PHE642 ILE651 ASP652 LYS655 LYS656 THR657 THR658 PRO663

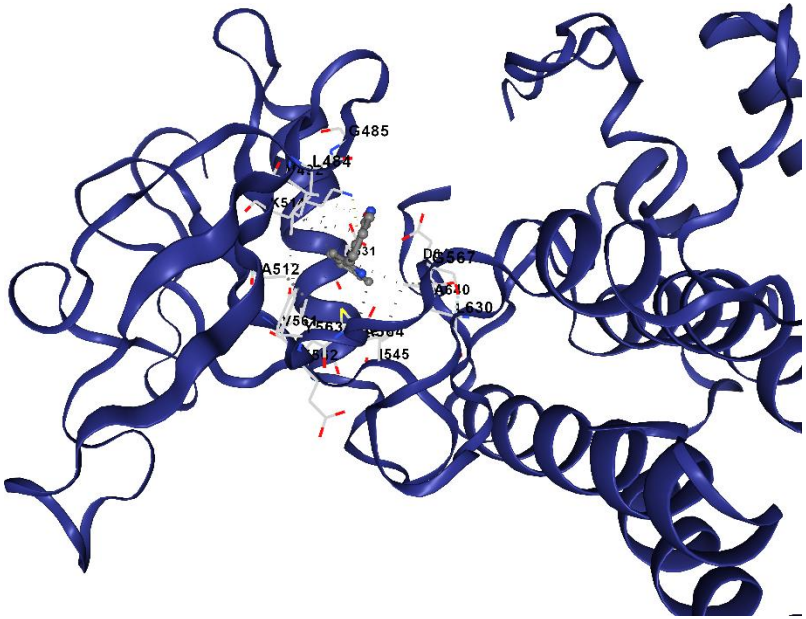


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
CHEMBL1165499	FGFR1	-9.7	3591	Chain A: GLY485 GLU486 GLY487 ALA488 PHE489 GLY490 GLN491 LYS514 MET515 LEU516 LYS517 ALA520 LYS523 ASP524 ASP527 LEU528 GLU531 ASN568 ARG570 GLU571 HIS621 ARG622 ASP623 ARG627 ASN628 ASP641 LEU644 ARG646 ASP647 ILE648 HIS649 HIS650 ILE651 ASP652 LYS655 LYS656 THR657 THR658 PRO663 MET667 TYR677

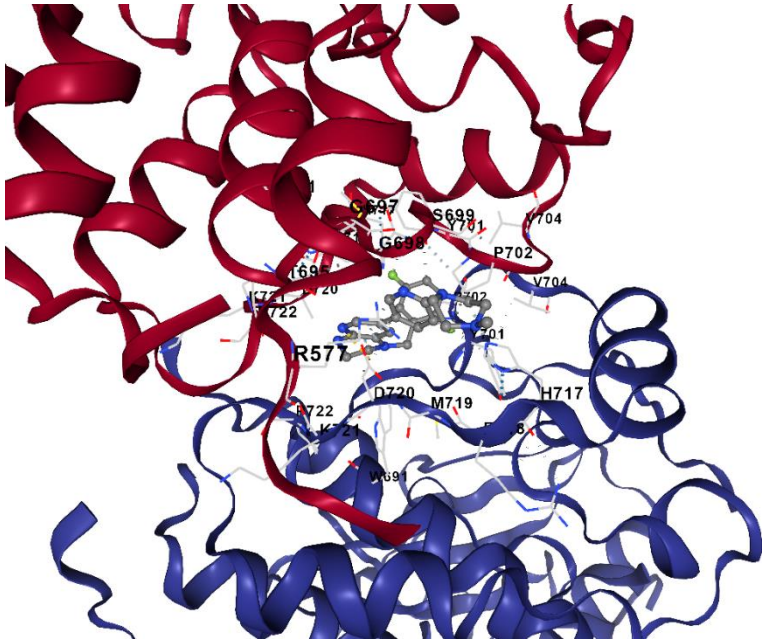


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
CHEMBL1773601	FGFR1	-10.9	3591	Chain A: LYS482 PRO483 LEU484 GLY485 GLU486 ALA488 PHE489 GLY490 GLN491 VAL492 ALA512 VAL513 LYS514 MET515 LEU516 ASP524 ASP527 LEU528 GLU531 MET532 MET535 ILE545 LEU547 VAL559 VAL561 GLU562 TYR563 ALA564 SER565 LYS566 GLY567 ASN568 ARG570 GLU571 ARG622 ASP623 ARG627 ASN628 LEU630 ALA640 ASP641 PHE642 GLY643 LEU644 ARG646 ASP647 ILE648 ILE651 ASP652 LYS655 LYS656 THR657 THR658 ASN659 PRO663 MET667

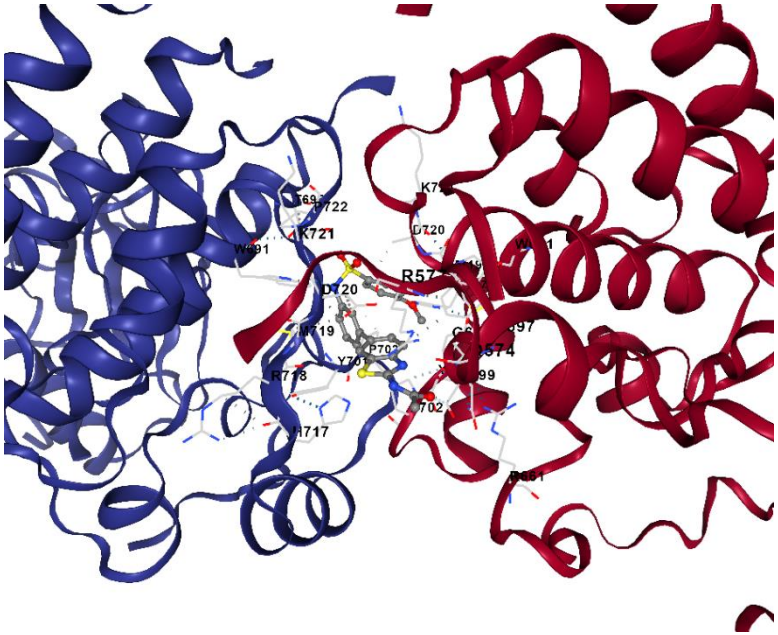


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
CHEMBL1773581	FGFR1	-10.8	3591	Chain A: LYS482 PRO483 LEU484 GLY485 GLU486 ALA488 PHE489 GLY490 GLN491 VAL492 ALA512 LYS514 MET515 LEU516 ASP524 ASP527 LEU528 GLU531 MET535 ILE545 VAL559 VAL561 GLU562 TYR563 ALA564 SER565 LYS566 GLY567 ASN568 ARG570 GLU571 ARG622 ASP623 ARG627 ASN628 LEU630 ALA640 ASP641 LEU644 ARG646 ASP647 ILE648 ILE651 ASP652 TYR653 LYS655 LYS656 THR657 THR658 ASN659 PRO663 VAL664

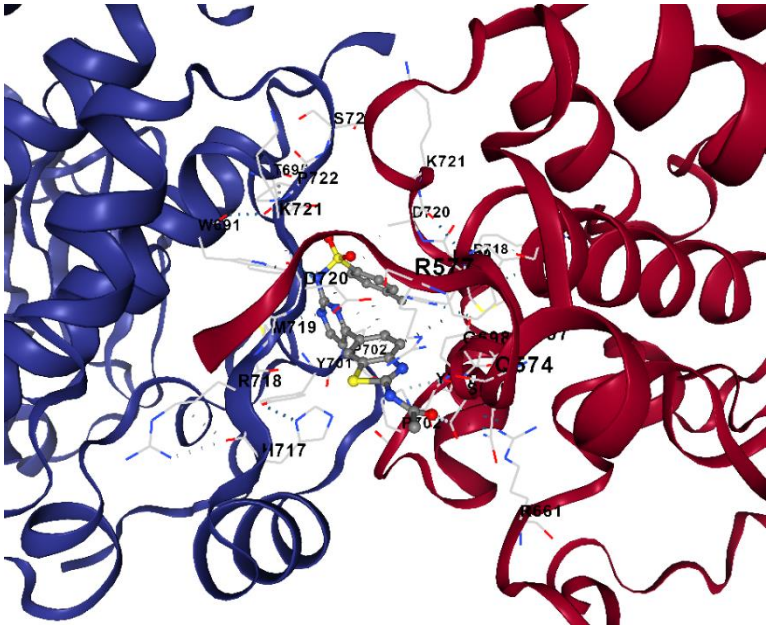


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
CHEMBL388978	FGFR1	-9.8	1933	Chain B: GLY487 ALA488 PHE489 GLY490 LYS514 HIS621 ARG622 ASP623 ARG627 ASN628 ASP641 LEU644 ILE651 ASP652 TYR653 TYR654 LYS655 LYS656 THR657 THR658 PRO663 VAL664 MET667 TYR677

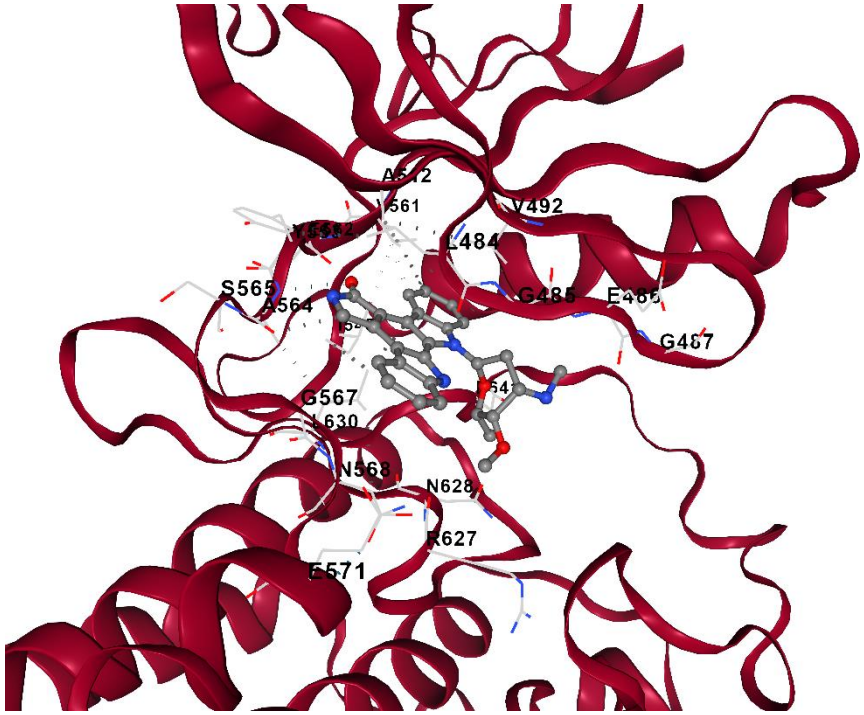


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
CHEMBL1615189	FGFR1	-10.9	1933	Chain B: PRO483 LEU484 GLY485 GLU486 ALA488 PHE489 GLY490 GLN491 VAL492 ALA512 VAL513 LYS514 MET515 LEU516 LYS517 ASP524 ASP527 LEU528 GLU531 MET535 ILE545 VAL559 VAL561 GLU562 TYR563 ALA564 SER565 LYS566 GLY567 ASN568 ARG570 GLU571 HIS621 ARG622 ASP623 ARG627 ASN628 LEU630 ALA640 ASP641 PHE642 GLY643 LEU644 ILE651 ASP652 TYR653 LYS655 THR657 THR658 PRO663 VAL664 MET667

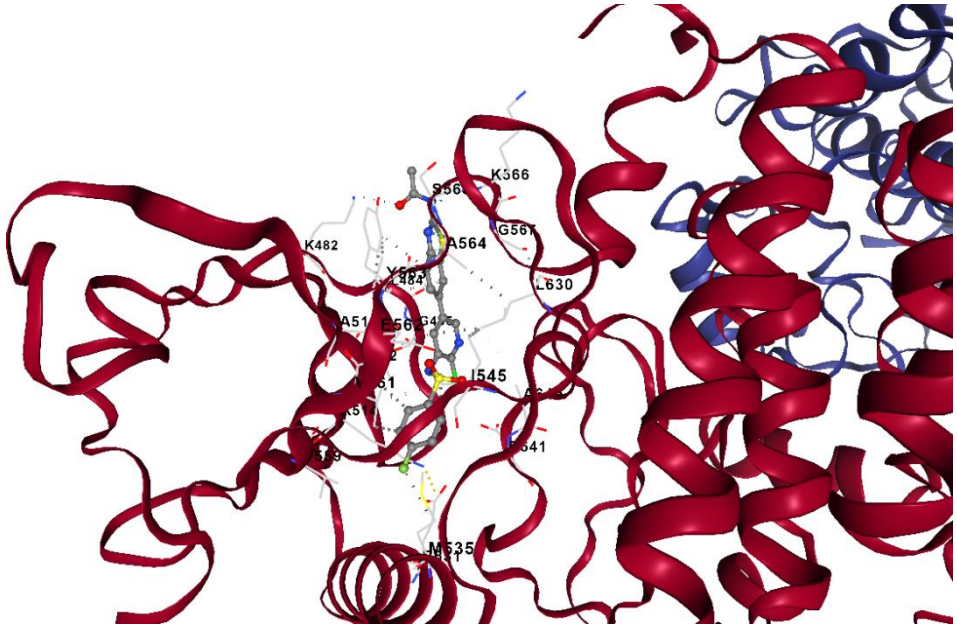


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
Lenvatinib	FGFR1	-9.3	3591	Chain A: LYS482 PRO483 LEU484 GLY485 GLU486 ALA488 PHE489 VAL492 ALA512 VAL513 LYS514 LEU516 ASP527 LEU528 GLU531 MET535 ILE545 VAL559 VAL561 GLU562 TYR563 ALA564 SER565 LYS566 GLY567 ASN568 ARG570 GLU571 HIS621 ARG622 ASP623 ARG627 ASN628 LEU630 ALA640 ASP641 LEU644 ALA645 ARG646 LYS655 LYS656 THR657 THR658 ASN659 PRO663 VAL664 TRP666

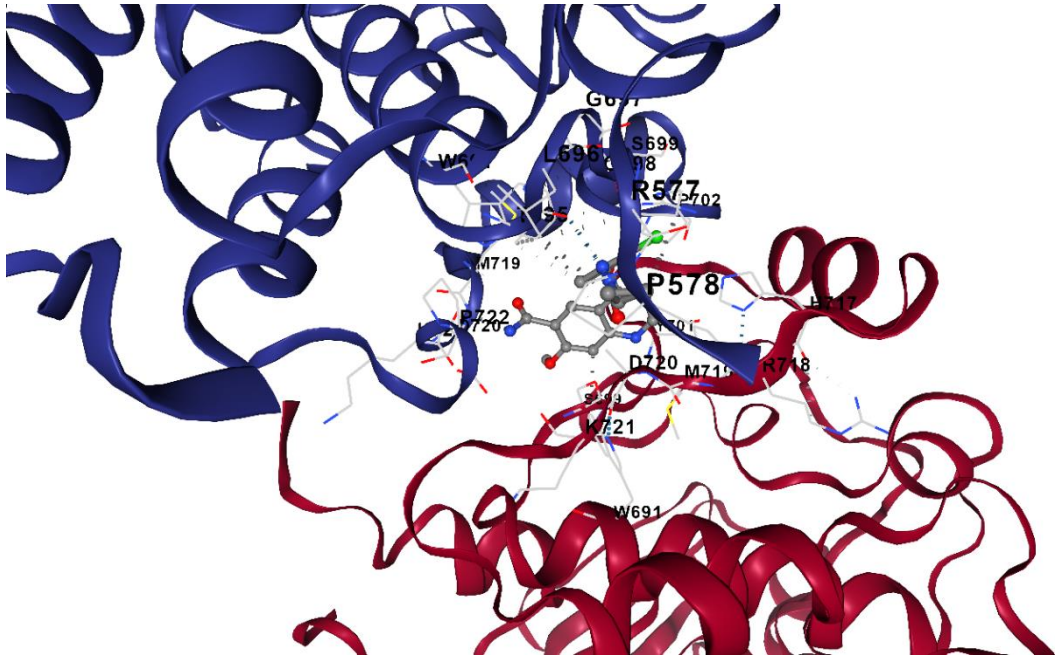


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
Regorafenib	FGFR1	-11.0	1933	Chain B: LEU484 GLY485 GLU486 GLY487 ALA488 PHE489 GLY490 GLN491 VAL492 ALA512 VAL513 LYS514 MET515 LEU516 LYS517 ALA520 ASP524 LEU525 ASP527 LEU528 GLU531 MET535 ILE545 LEU547 VAL559 ILE560 VAL561 GLU562 TYR563 ALA564 SER565 LYS566 GLY567 ASN568 ARG570 GLU571 HIS621 ARG622 ASP623 ARG627 ASN628 LEU630 ILE639 ALA640 ASP641 PHE642 GLY643 LEU644 ILE651 LYS656 THR657 THR658 ASN659 PRO663 MET667

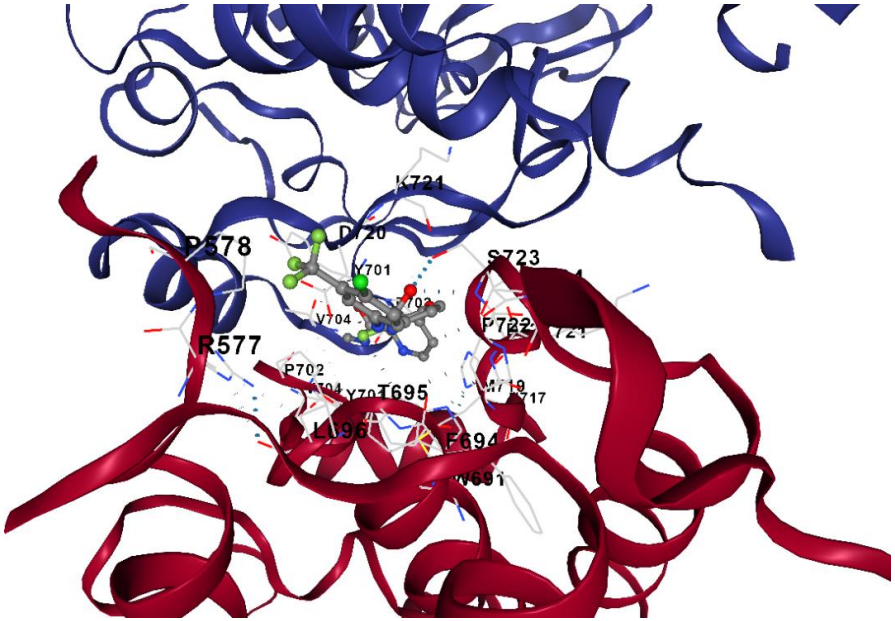


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
Sorafenib	FGFR1	-11.2	3591	Chain A: LYS482 PRO483 LEU484 GLY485 GLU486 GLY487 ALA488 PHE489 GLY490 GLN491 VAL492 ALA512 VAL513 LYS514 MET515 LEU516 LYS517 ALA520 LYS523 ASP524 ASP527 LEU528 GLU531 MET532 MET535 ILE545 VAL559 VAL561 GLU562 TYR563 ALA564 SER565 LYS566 GLY567 ASN568 ARG570 GLU571 ARG622 ASP623 ARG627 ASN628 LEU630 ALA640 ASP641 PHE642 LEU644 ARG646 ASP647 ILE651 ASP652 TYR653 LYS655 LYS656 THR657 THR658 ASN659 PRO663 VAL664 MET667

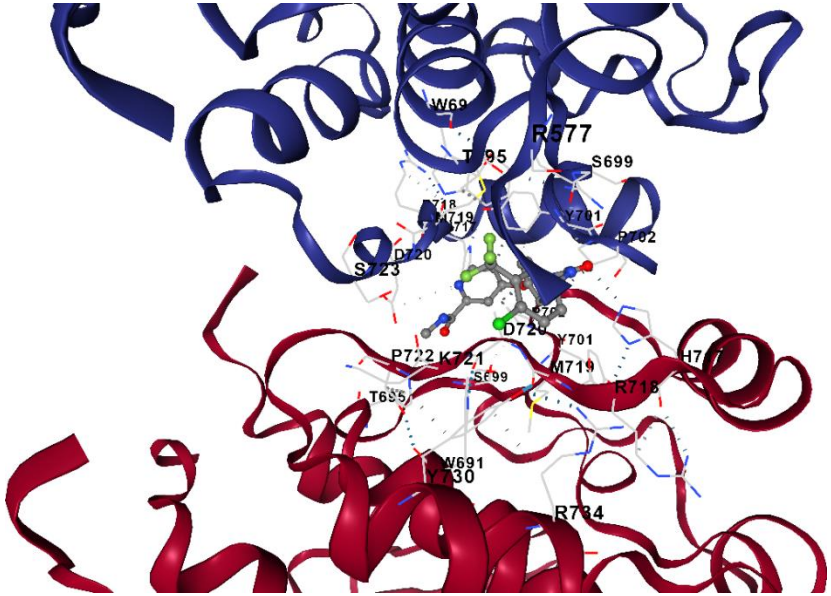


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
CHEMBL328029	ALK	-9.2	2409	Chain A: LEU1122 GLY1123 HIS1124 GLY1125 ALA1126 PHE1127 GLY1128 GLU1129 VAL1130 ALA1148 LYS1150 THR1151 LEU1152 PRO1153 ASP1160 ASP1163 PHE1164 GLU1167 LEU1196 GLU1197 LEU1198 MET1199 ALA1200 GLY1201 GLY1202 ASP1203 SER1206 ARG1253 LEU1256 GLY1269 ASP1270 ARG1275

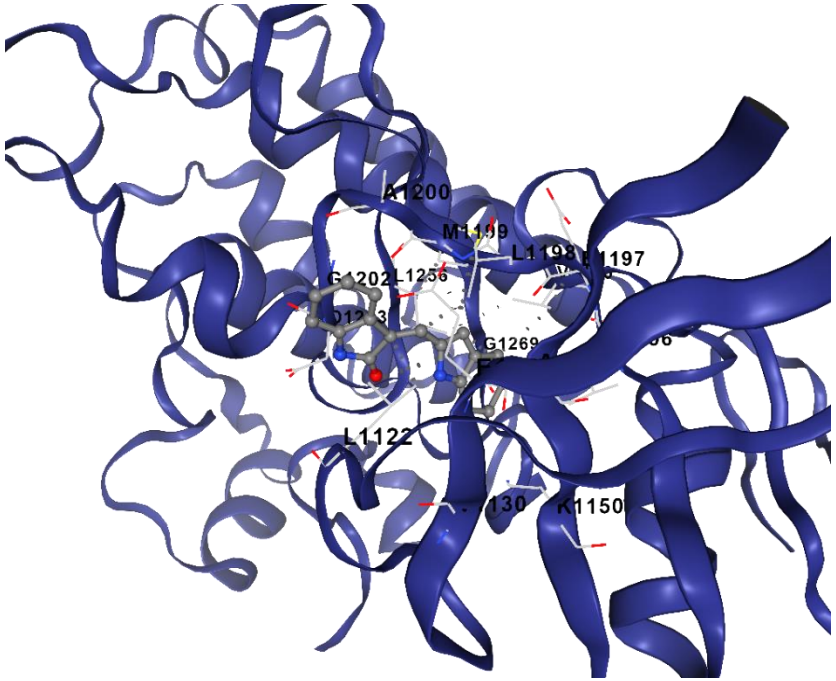


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
CHEMBL1165499	ALK	-11.3	2409	Chain A: LEU1122 GLY1123 HIS1124 GLY1125 ALA1126 PHE1127 VAL1130 ALA1148 LYS1150 VAL1180 ARG1181 LEU1196 GLU1197 LEU1198 MET1199 ALA1200 GLY1201 GLY1202 ASP1203 SER1206 GLU1210 ARG1253 ASN1254 CYS1255 LEU1256 GLY1269 ASP1270

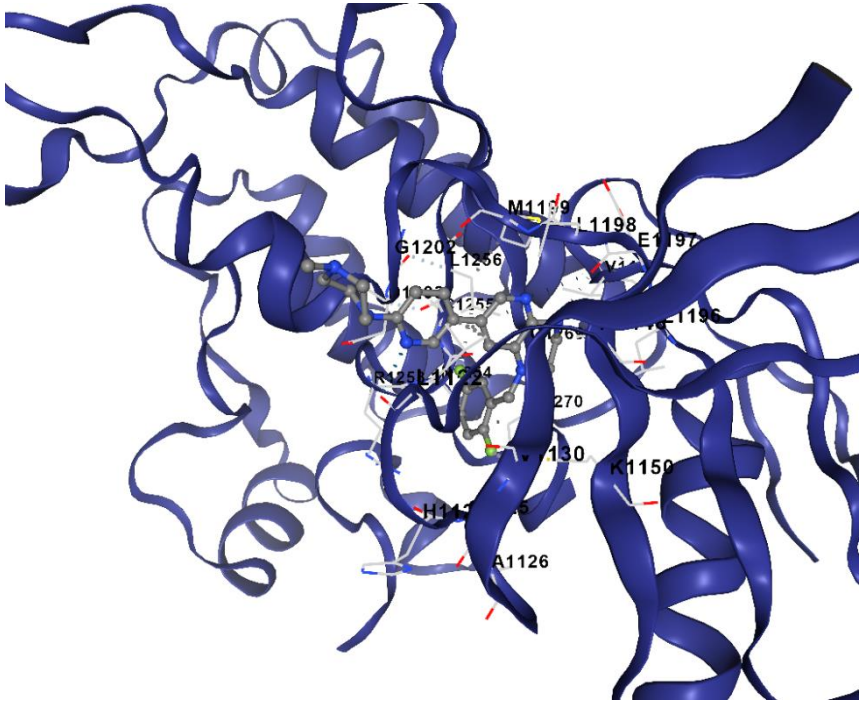


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
CHEMBL1773601	ALK	-10.3	2409	Chain A: GLY1121 LEU1122 GLY1123 HIS1124 GLY1125 ALA1126 PHE1127 VAL1130 TYR1131 GLU1132 ALA1148 VAL1149 LYS1150 GLU1167 VAL1180 LEU1196 GLU1197 LEU1198 MET1199 ALA1200 GLY1201 GLY1202 ASP1203 LYS1205 SER1206 ARG1253 ASN1254 CYS1255 LEU1256 GLY1269 ASP1270 TRP1295 GLU1321 MET1328

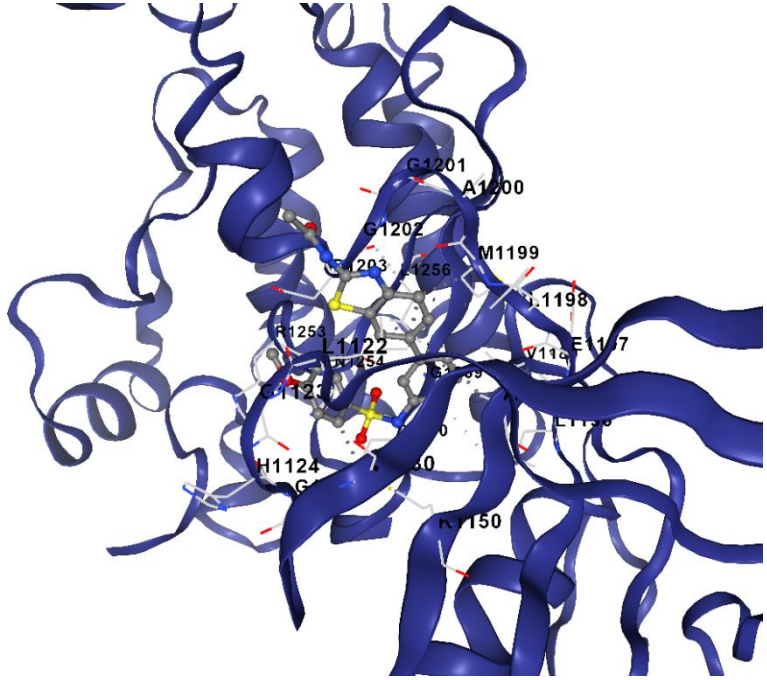


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
CHEMBL1773581	ALK	-9.2	2409	Chain A: LEU1122 GLY1123 HIS1124 GLY1125 ALA1126 PHE1127 VAL1130 TYR1131 GLU1132 ALA1148 LYS1150 LEU1196 GLU1197 LEU1198 MET1199 ALA1200 GLY1201 GLY1202 ASP1203 LYS1205 SER1206 GLU1210 ARG1253 ASN1254 LEU1256 PRO1260 GLY1269 ASP1270

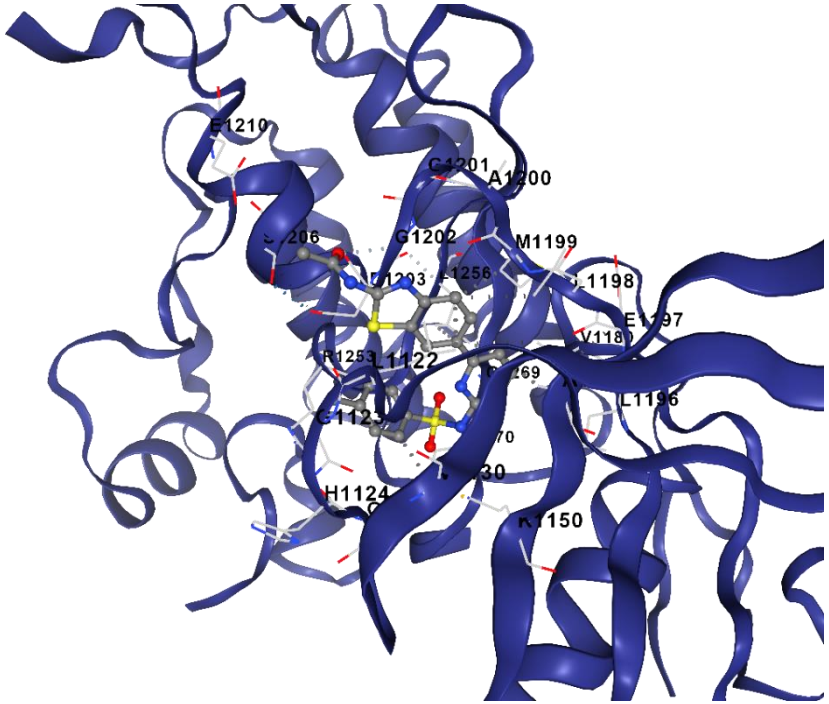


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
CHEMBL388978	ALK	-9.6	173	Chain A: TYR1096 CYS1097 PHE1098 ALA1099 GLY1100 LYS1101 ASN1243 HIS1244 PHE1245 ILE1246 HIS1247 ARG1248 ALA1274 ILE1277 TYR1278 ARG1279 PHE1306

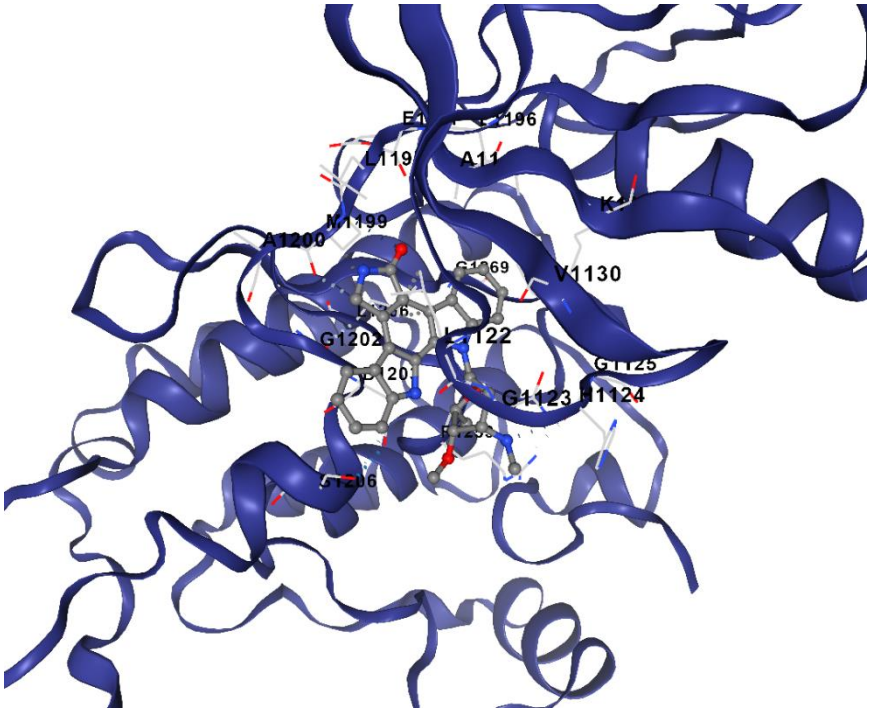


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
CHEMBL1615189	ALK	-9.5	2409	Chain A: ARG1120 LEU1122 GLY1123 HIS1124 GLY1125 ALA1126 PHE1127 VAL1130 GLU1132 ALA1148 LYS1150 GLU1167 VAL1180 LEU1196 GLU1197 LEU1198 MET1199 ALA1200 GLY1201 GLY1202 ASP1203 LYS1205 SER1206 GLU1210 ARG1253 ASN1254 LEU1256 PRO1260 GLY1269 ASP1270

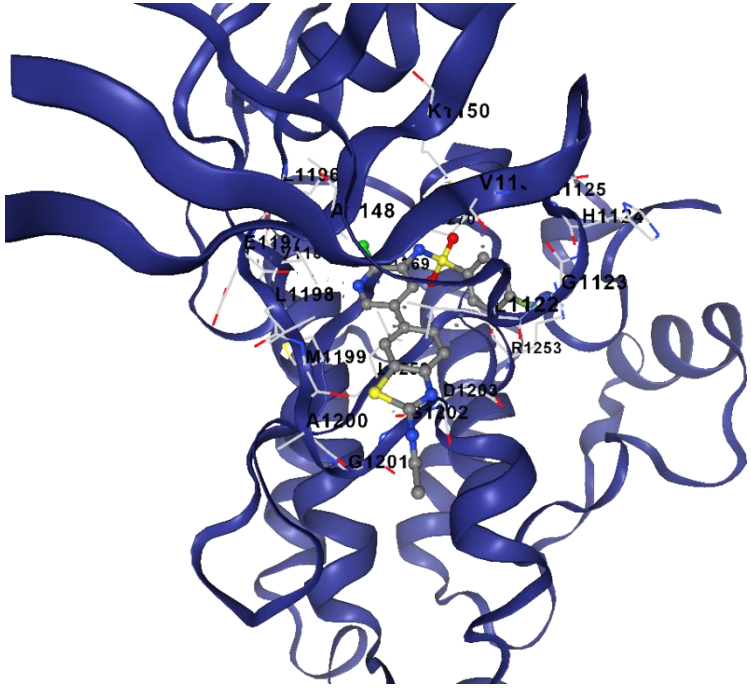


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
Lenvatinib	ALK	-8.7	2409	Chain A: LEU1122 GLY1123 HIS1124 GLY1125 ALA1126 PHE1127 VAL1130 GLU1132 ALA1148 LYS1150 LEU1196 GLU1197 LEU1198 MET1199 ALA1200 GLY1201 GLY1202 ASP1203 LYS1205 SER1206 ALA1252 ARG1253 ASN1254 CYS1255 LEU1256 PRO1260 GLY1269 ASP1270

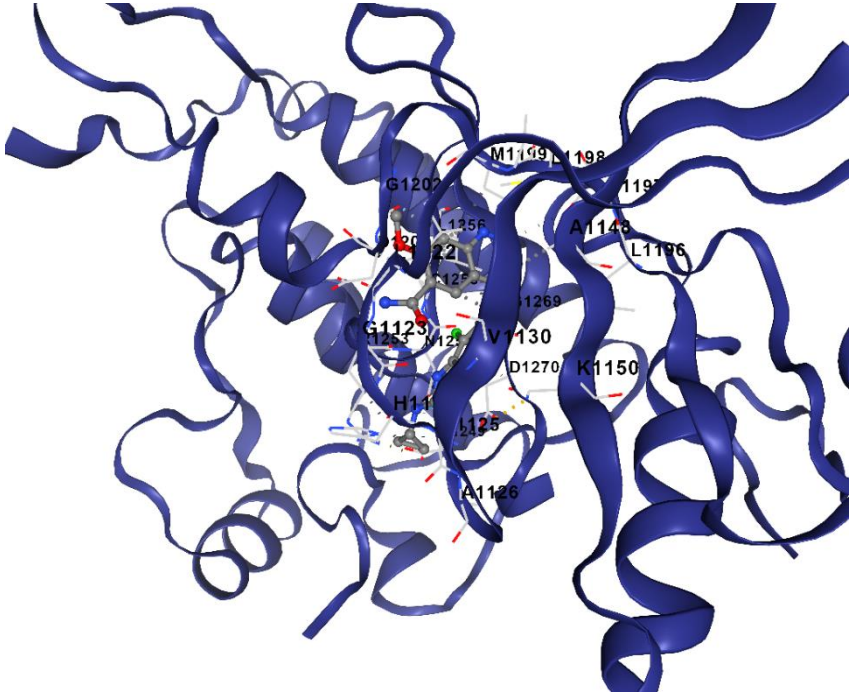


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
Regorafenib	ALK	-9.6	2409	Chain A: ARG1120 GLY1121 LEU1122 GLY1123 HIS1124 GLY1125 ALA1126 PHE1127 VAL1130 GLU1132 GLN1146 ALA1148 LYS1150 LEU1196 GLU1197 LEU1198 MET1199 ALA1200 GLY1201 GLY1202 ASP1203 LYS1205 SER1206 GLU1210 ALA1252 ARG1253 ASN1254 LEU1256 PRO1260 GLY1269 ASP1270

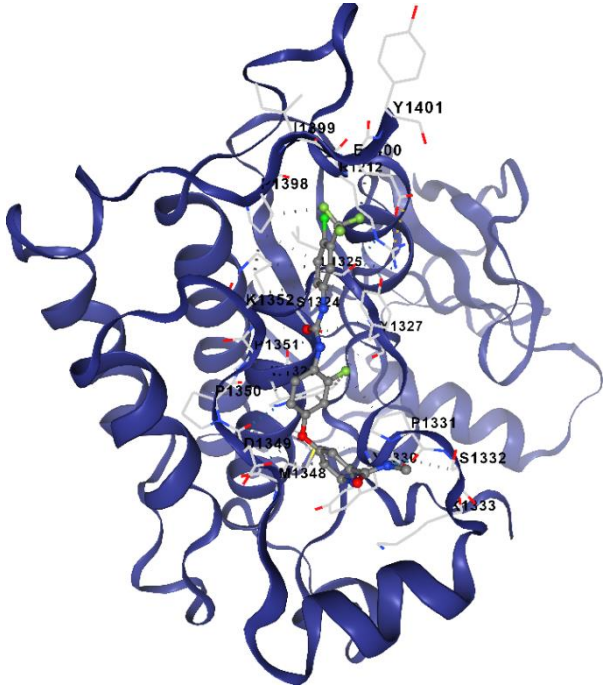


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
Sorafenib	ALK	-9.8	2409	Chain A: ARG1120 LEU1122 GLY1123 HIS1124 GLY1125 ALA1126 PHE1127 VAL1130 GLU1132 ALA1148 LYS1150 VAL1180 LEU1196 GLU1197 LEU1198 MET1199 ALA1200 GLY1201 GLY1202 ASP1203 LYS1205 SER1206 ARG1209 ASP1249 ARG1253 ASN1254 CYS1255 LEU1256 GLY1269 ASP1270 MET1273 MET1290 LEU1291 PRO1292

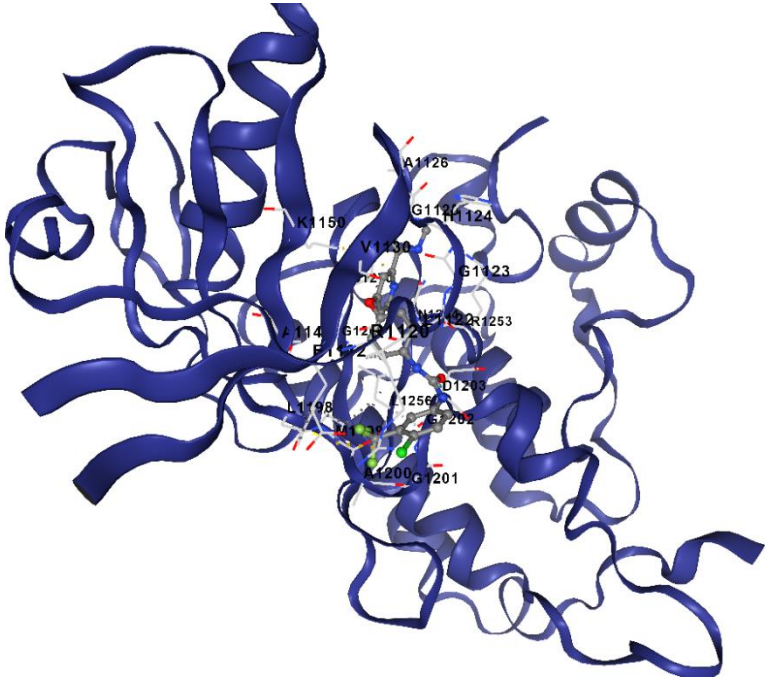


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
CHEMBL328029	AKT1	-8.9	2670	Chain A: LEU156 GLY157 LYS158 GLY159 THR160 PHE161 GLY162 VAL164 ALA177 LYS179 LEU181 ILE186 GLU191 HIS194 THR195 GLU198 THR211 MET227 GLU228 TYR229 ALA230 GLY233 GLU234 PHE236 PHE237 SER240 ARG241 ASP274 LYS276 GLU278 ASN279 MET281 THR291 ASP292 PHE293 GLY294 LEU295 TYR437 PHE438 ASP439 PHE442 Chain C: ARG2 ARG4 THR5 THR6 SER7 PHE8

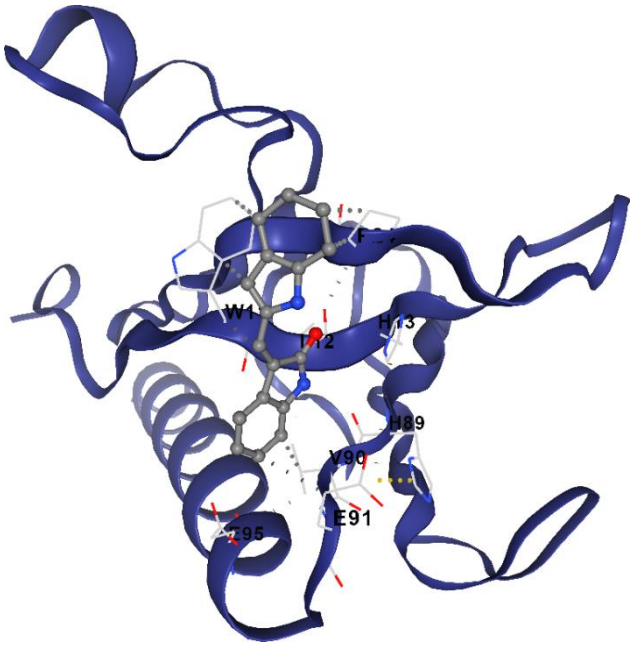


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
CHEMBL1165499	AKT1	-10.2	2202	Chain B: LEU156 GLY157 LYS158 GLY159 THR160 PHE161 GLY162 LYS163 VAL164 ALA177 LYS179 LEU181 ILE186 GLU191 HIS194 THR195 GLU198 THR211 PHE225 MET227 GLU228 TYR229 ALA230 GLU234 TYR272 ASP274 LYS276 GLU278 ASN279 MET281 THR291 ASP292 PHE293 GLY294 LEU295 PHE438 PHE442 Chain D: ARG4 THR5 THR6 SER7 PHE8 ALA9

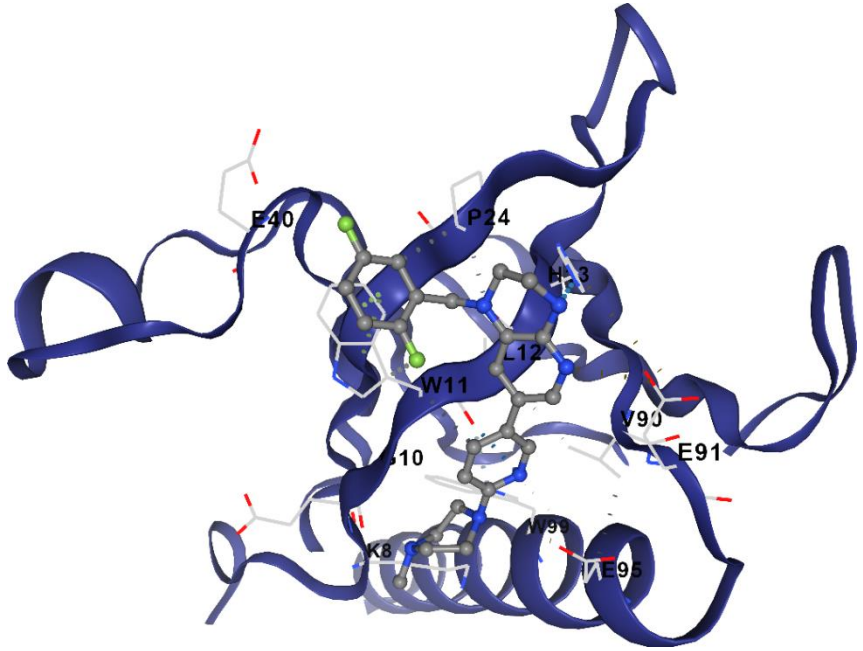


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
CHEMBL1773601	AKT1	-10.8	2670	Chain A: LEU156 GLY157 LYS158 GLY159 THR160 PHE161 GLY162 LYS163 VAL164 ALA177 LYS179 LEU181 ILE186 GLU191 HIS194 THR195 GLU198 THR211 PHE225 MET227 GLU228 TYR229 ALA230 GLU234 ASP274 LYS276 GLU278 ASN279 MET281 THR291 ASP292 PHE293 GLY294 LEU295 TYR437 PHE438 ASP439 PHE442 Chain C: ARG4 THR5 THR6 SER7 PHE8 ALA9

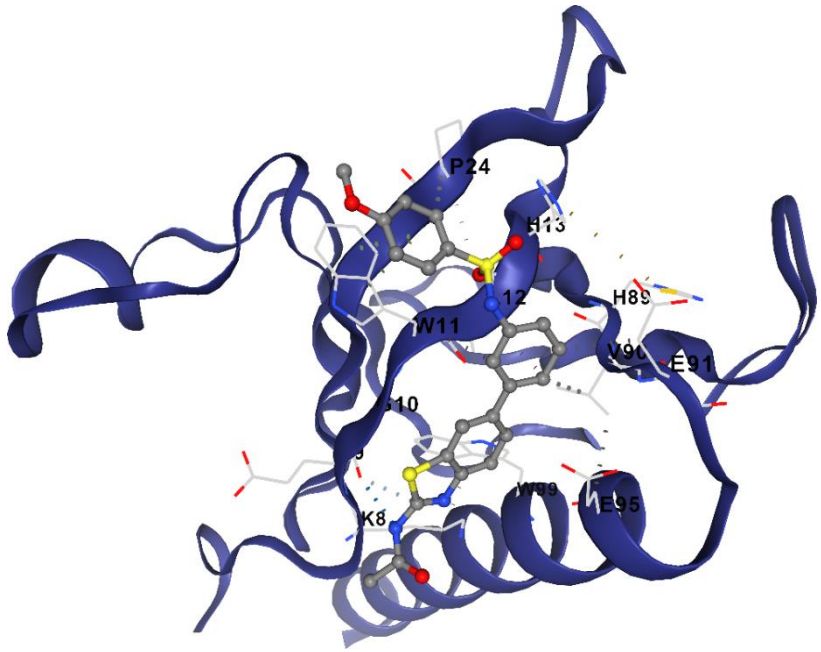


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
CHEMBL1773581	AKT1	-10.4	2670	Chain A: LEU156 GLY157 LYS158 GLY159 PHE161 GLY162 LYS163 VAL164 ALA177 LYS179 LEU181 ILE186 GLU191 HIS194 THR195 GLU198 THR211 MET227 GLU228 TYR229 ALA230 GLU234 ASP274 LYS276 GLU278 ASN279 MET281 THR291 ASP292 PHE293 GLY294 LEU295 PHE438 PHE442 Chain C: ARG4 THR5 THR6 SER7 PHE8 ALA9

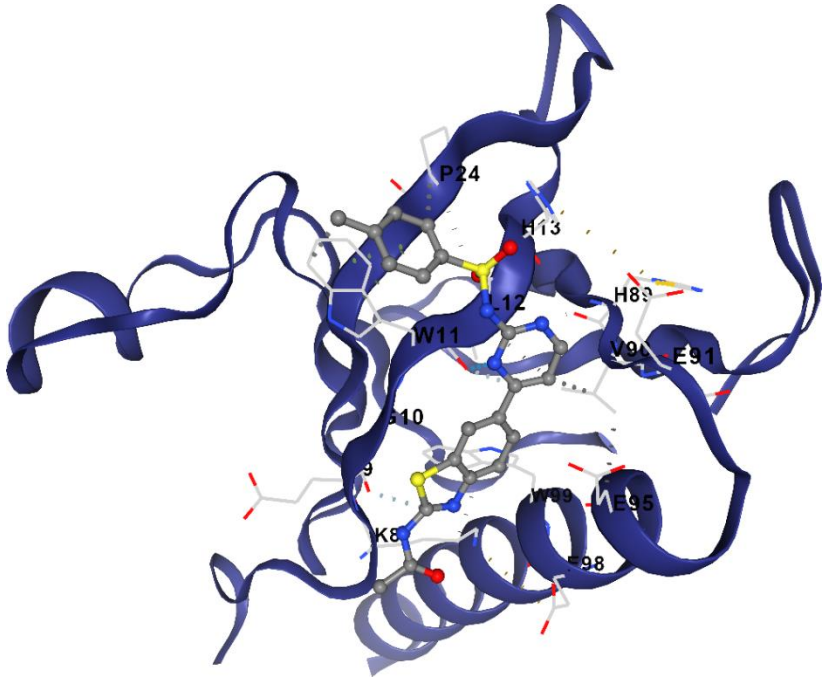


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
CHEMBL388978	AKT1	-10.0	2670	Chain A: LEU156 GLY157 LYS158 GLY159 THR160 PHE161 GLY162 LYS163 VAL164 LYS179 LEU181 ILE186 GLU191 HIS194 THR195 GLU198 PHE225 MET227 GLU234 ASP274 LYS276 GLU278 ASN279 MET281 THR291 ASP292 GLY294 LEU295 PHE438 ASP439 GLU441 PHE442 Chain C: ARG4 THR5 THR6 SER7 PHE8

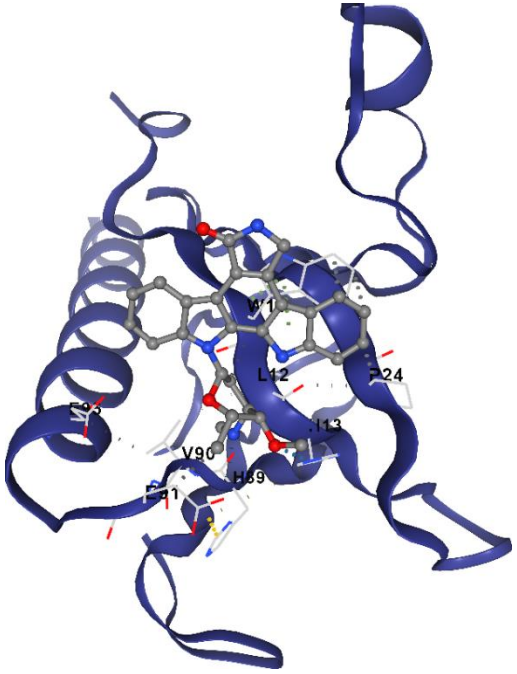


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
CHEMBL1615189	AKT1	-9.8	2202	Chain B: LEU156 GLY157 LYS158 GLY159 THR160 PHE161 GLY162 LYS163 VAL164 ALA177 MET178 LYS179 LEU181 ILE186 GLU191 HIS194 THR195 GLU198 THR211 MET227 GLU228 TYR229 ALA230 GLU234 PHE237 ASP274 LYS276 GLU278 ASN279 MET281 THR291 ASP292 PHE293 GLY294 LEU295 TYR437 PHE438 ASP439 GLU441 PHE442 Chain D: ARG4 THR5 THR6 SER7 PHE8 ALA9

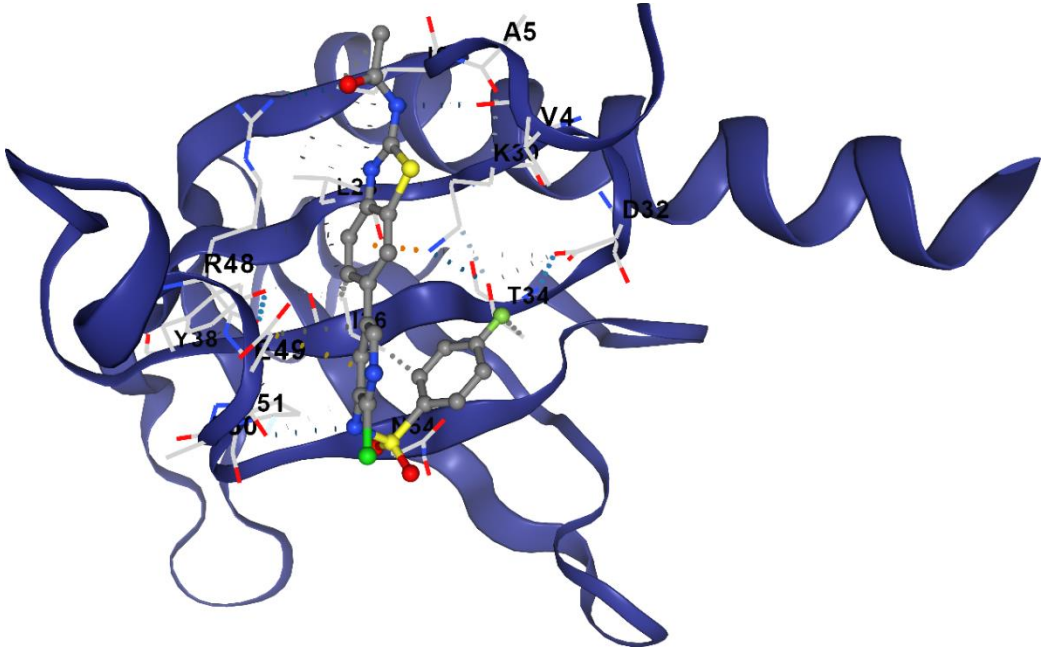


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
Lenvatinib	AKT1	-9.2	2670	Chain A: LEU156 GLY157 LYS158 GLY159 THR160 PHE161 GLY162 LYS163 VAL164 ALA177 LYS179 ILE180 LEU181 ILE186 GLU191 HIS194 THR195 GLU198 THR211 MET227 GLU228 TYR229 ALA230 GLU234 ASP274 LYS276 GLU278 ASN279 LEU280 MET281 THR291 ASP292 PHE293 GLY294 LEU295 TYR437 PHE438 ASP439 PHE442 Chain C: ARG4 THR5 THR6 SER7

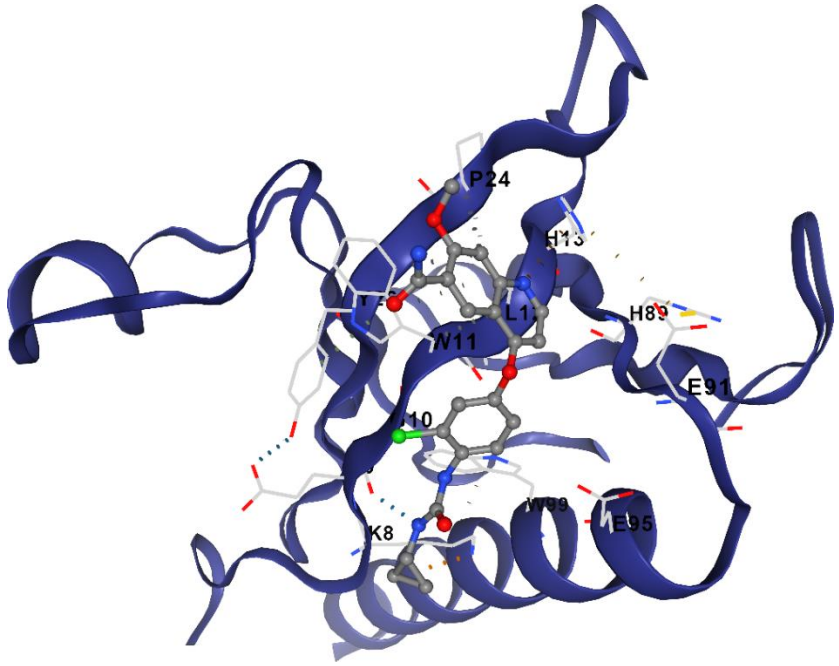


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
Regorafenib	AKT1	-10.1	2202	Chain B: LEU156 GLY157 LYS158 GLY159 THR160 PHE161 GLY162 VAL164 ALA177 LYS179 LEU181 GLU191 HIS194 THR195 GLU198 THR211 MET227 GLU228 TYR229 ALA230 GLU234 TYR272 ASP274 LYS276 GLU278 ASN279 MET281 THR291 ASP292 PHE293 GLY294 LEU295 CYS296 PHE438 GLU441 PHE442 Chain D: ARG4 THR5 THR6 SER7 PHE8 ALA9

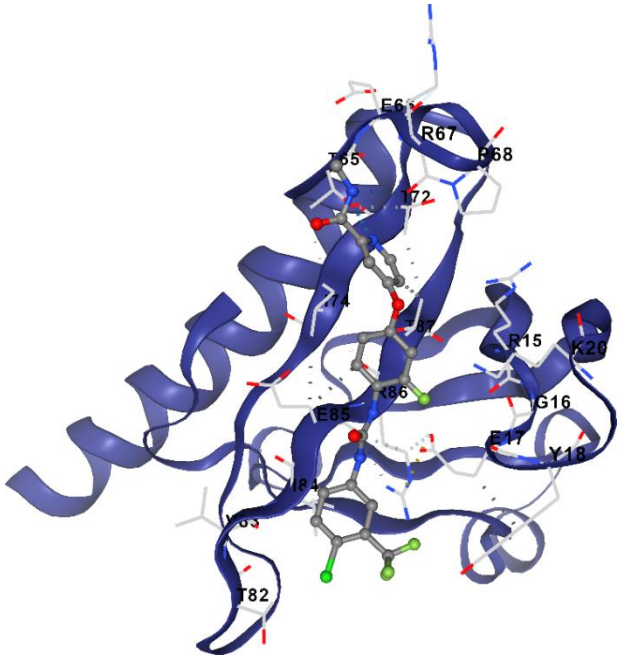


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
Sorafenib	AKT1	-10.3	2670	Chain A: LEU156 GLY157 LYS158 GLY159 THR160 PHE161 GLY162 VAL164 ALA177 LYS179 LEU181 ILE186 GLU191 HIS194 THR195 GLU198 THR211 MET227 GLU228 TYR229 ALA230 ASN231 GLU234 TYR272 ASP274 LYS276 GLU278 ASN279 MET281 THR291 ASP292 PHE293 GLY294 LEU295 PHE438 PHE442 Chain C: ARG4 THR5 THR6 SER7 PHE8 ALA9

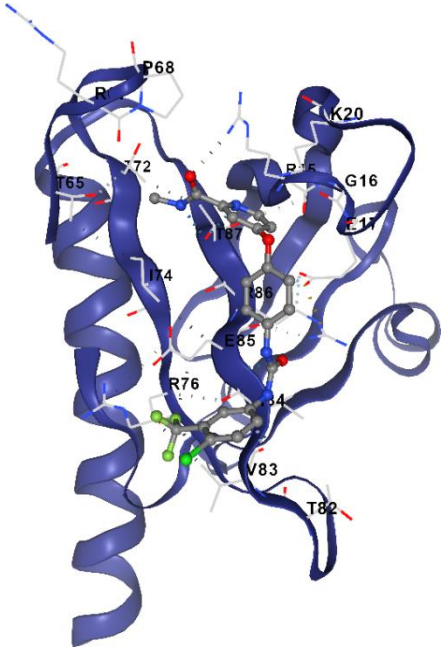


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
CHEMBL328029	FLT3	-9.9	2684	Chain B: LYS614 LEU616 GLY617 VAL624 ALA642 LYS644 ALA657 SER660 GLU661 MET664 MET665 LEU668 ILE674 VAL675 PHE691 GLU692 TYR693 CYS694 CYS695 TYR696 GLY697 ASP698 ASN701 LEU802 VAL808 HIS809 ARG810 ASP811 ARG815 LEU818 ILE827 CYS828 ASP829 PHE830 GLY831 LEU832 ALA833 ARG834 ILE836 ASP839 TYR842 ARG849 LEU850 PRO851

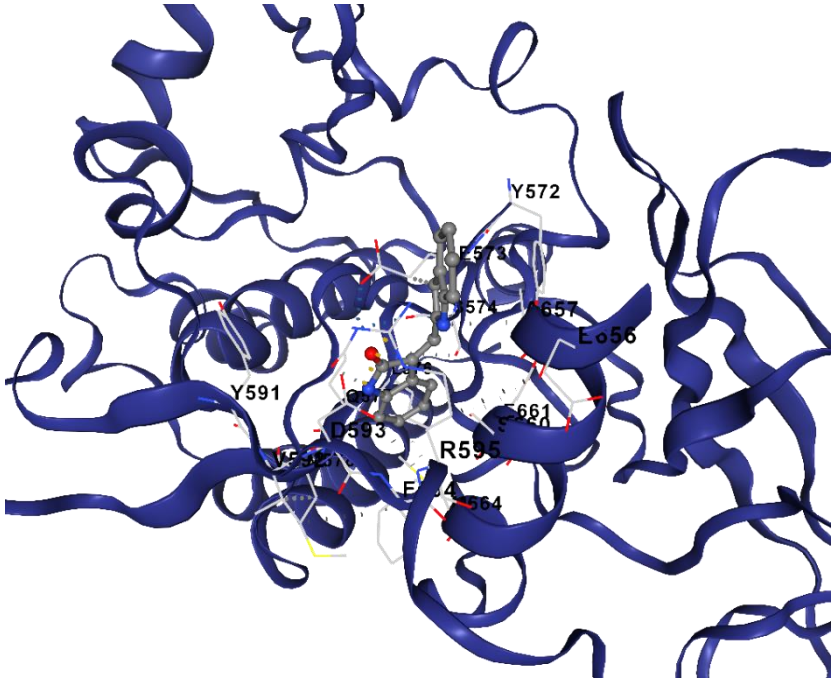


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
CHEMBL1165499	FLT3	-10.0	3007	Chain A: LYS614 VAL615 LEU616 GLY617 SER618 VAL624 ALA642 VAL643 LYS644 GLU661 MET665 VAL675 LEU689 PHE691 GLU692 TYR693 CYS694 CYS695 TYR696 GLY697 ASP698 LEU700 ASN701 ARG704 ARG815 LEU818 CYS828 ASP829 PHE830 ALA833 ARG834 ASP835 ILE836 SER838 ASP839 SER840 ASN841

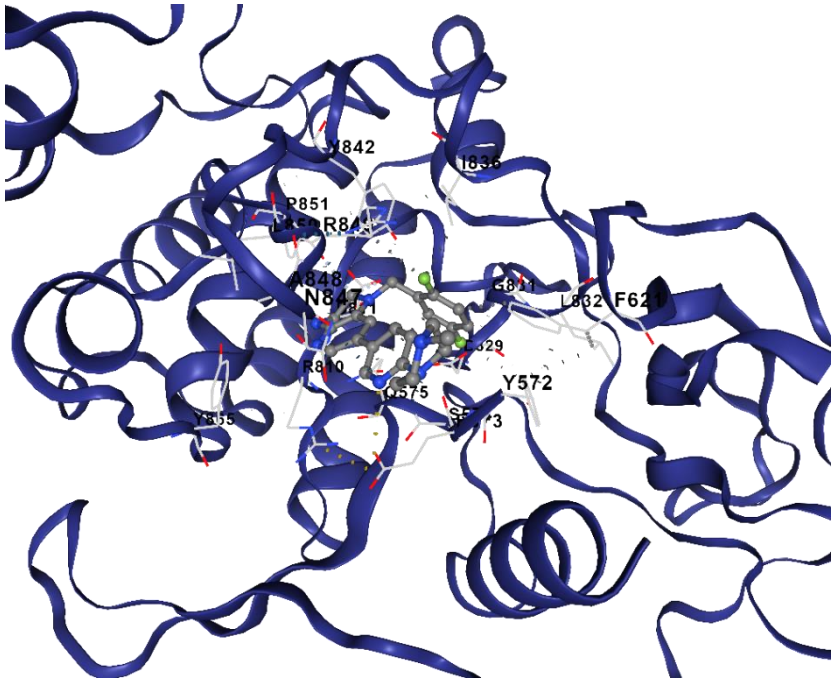


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
CHEMBL1773601	FLT3	-10.7	3007	Chain A: LEU616 GLY617 SER618 GLY619 ALA620 PHE621 VAL624 ALA642 LYS644 ALA657 SER660 GLU661 MET664 MET665 LEU668 ILE674 VAL675 PHE691 GLU692 TYR693 CYS694 CYS695 TYR696 GLY697 ASP698 LEU700 ASN701 TYR702 SER705 LEU802 VAL808 HIS809 ARG810 ARG815 LEU818 ILE827 CYS828 ASP829 PHE830 GLY831 LEU832 ALA833 ARG834 ASP835 ILE836 SER838 ASP839

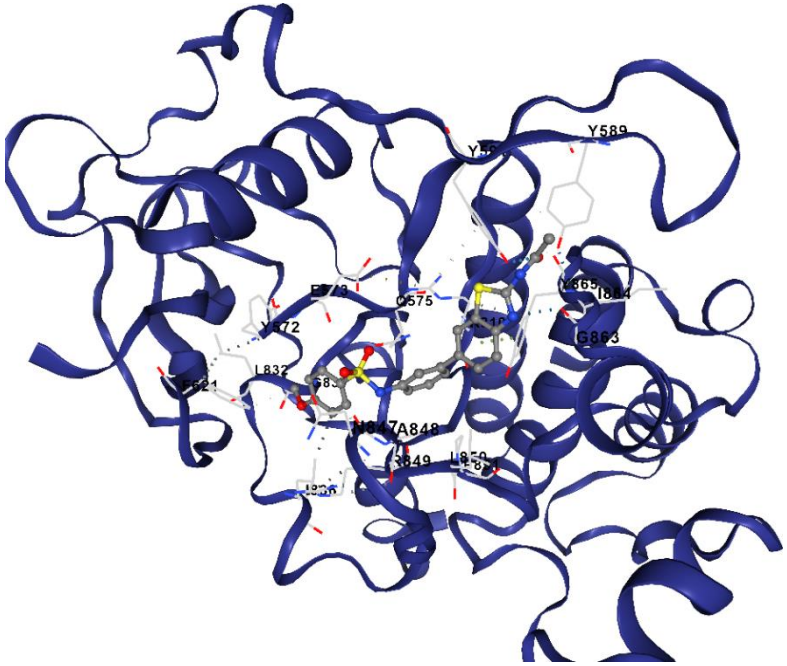


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
CHEMBL1773581	FLT3	-10.5	3007	Chain A: LEU616 GLY617 SER618 ALA620 PHE621 VAL624 ALA642 LYS644 ALA657 LEU658 SER660 GLU661 MET664 MET665 LEU668 ILE674 VAL675 PHE691 GLU692 TYR693 CYS694 CYS695 TYR696 GLY697 ASP698 LEU700 ASN701 SER705 LEU802 HIS809 ARG810 ASP811 ARG815 LEU818 ILE827 CYS828 ASP829 PHE830 GLY831 LEU832 ALA833 ARG834 ASP835 ILE836 MET837 SER838 ASP839 TYR842 ARG849 LEU850 PRO851

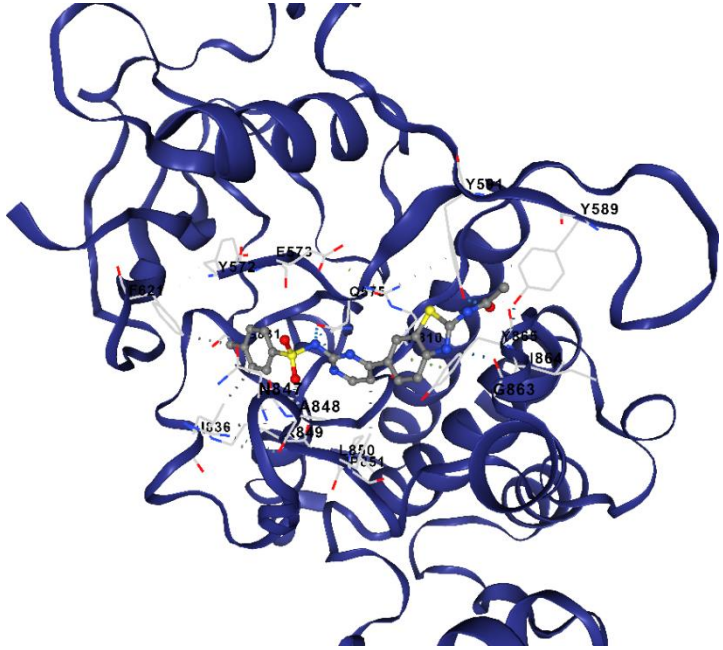


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
CHEMBL388978	FLT3	-9.2	805	Chain A: GLN903 ASN904 GLY905 PHE906 LYS907 MET908 ASP909 GLN910 GLU916 TYR919 ILE920 GLN923 SER924 ALA927 PHE928 ASP929 LYS932 Chain B: GLN903 ASN904 GLY905 PHE906 LYS907 ASP909 GLN910 GLU916 TYR919 ILE920 GLN923 SER924 TRP926 ALA927 PHE928 ASP929 LYS932

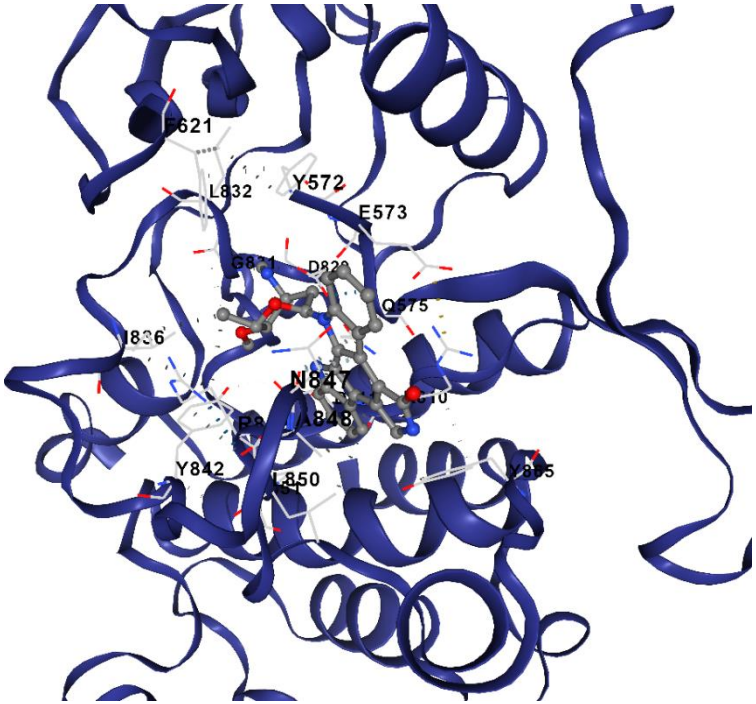


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
CHEMBL1615189	FLT3	-10.6	3007	Chain A: LEU616 GLY617 SER618 VAL624 ALA642 LYS644 GLU661 MET665 VAL675 PHE691 GLU692 TYR693 CYS694 CYS695 TYR696 GLY697 ASP698 ASN701 ARG815 LEU818 ILE827 CYS828 ASP829 PHE830 ALA833 ARG834 ASP835 SER838 ASP839

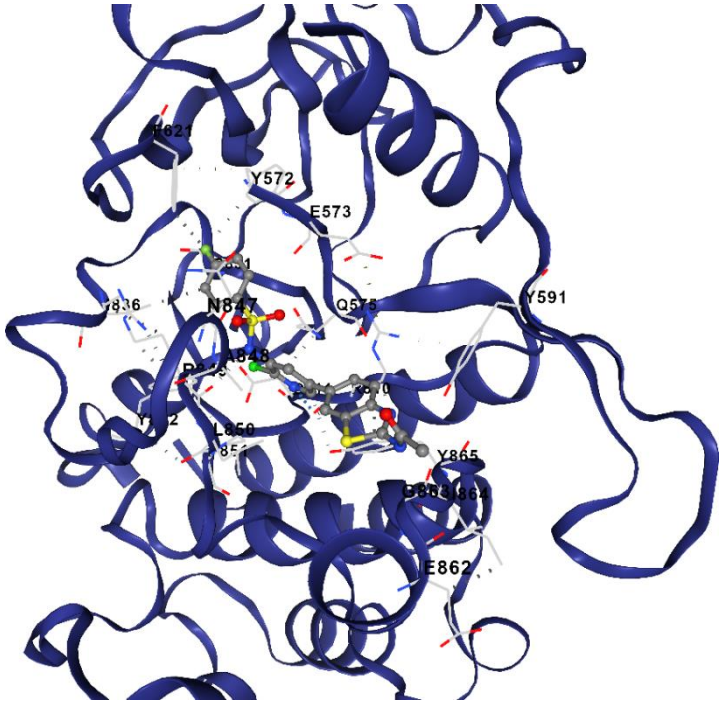


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
Lenvatinib	FLT3	-8.3	2684	Chain B: LYS614 VAL615 LEU616 GLY617 SER618 GLY619 VAL624 ALA642 LYS644 ALA657 GLU661 MET665 ILE674 VAL675 PHE691 GLU692 TYR693 CYS694 CYS695 TYR696 GLY697 ASP698 LEU699 LEU700 ASN701 TYR702 LEU703 ARG704 SER705 VAL808 HIS809 ARG810 ARG815 LEU818 ILE827 CYS828 ASP829 PHE830 ALA833 ARG834 ASP835 ILE836 SER838 ASP839 SER840 ASN841

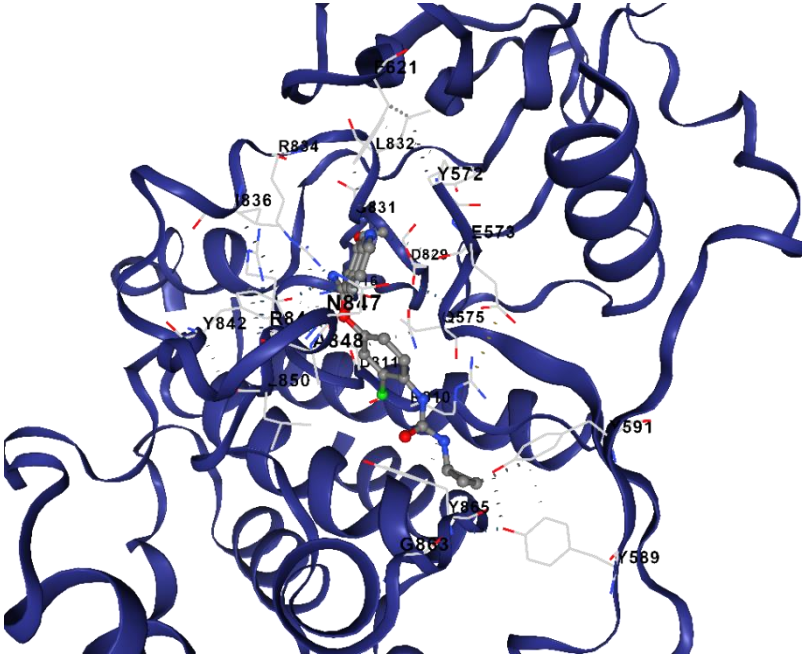


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
Regorafenib	FLT3	-10.0	2684	Chain B: LYS614 VAL615 LEU616 GLY617 SER618 VAL624 ALA642 LYS644 GLU661 MET665 VAL675 PHE691 GLU692 TYR693 CYS694 CYS695 TYR696 GLY697 ASP698 LEU700 ASN701 TYR702 ARG704 SER705 ARG815 LEU818 CYS828 ASP829 PHE830 GLY831 ALA833 ARG834 ASP835 SER838 ASP839 ASN841

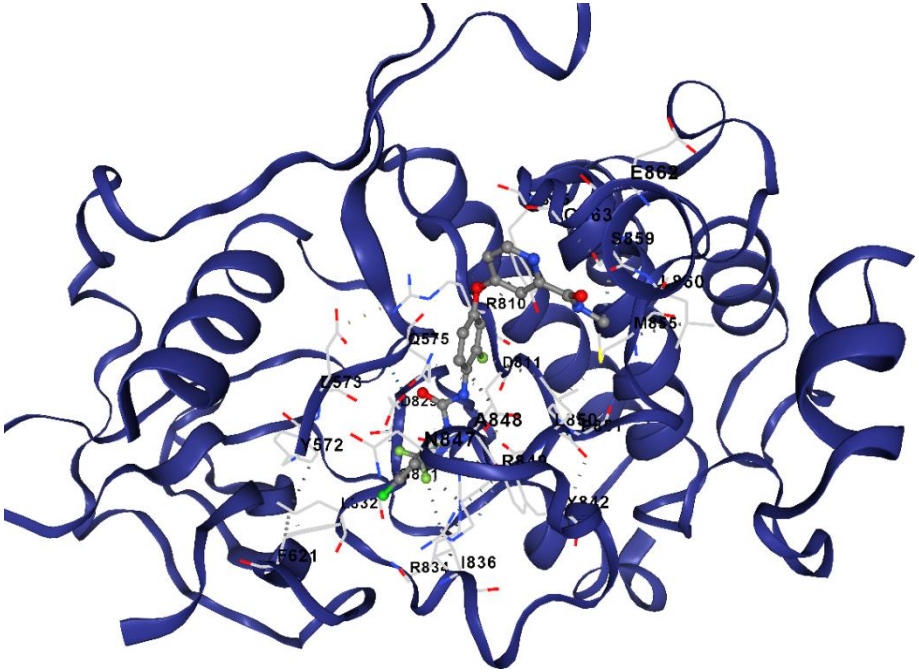


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
Sorafenib	FLT3	-9.8	805	Chain A: ASN904 GLY905 PHE906 LYS907 MET908 ASP909 GLN910 PRO911 PHE912 GLU916 TYR919 GLN923 SER924 ALA927 PHE928 ASP929 LYS932 Chain B: GLN903 ASN904 GLY905 PHE906 LYS907 MET908 ASP909 GLN910 GLU916 TYR919 ILE920 GLN923 SER924 TRP926 ALA927 PHE928 ASP929 LYS932

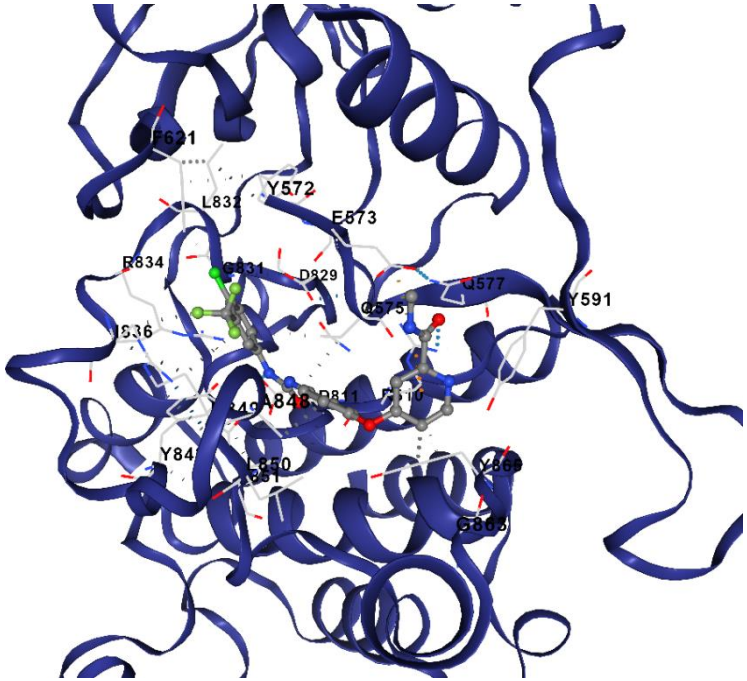


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
CHEMBL328029	PIK3CA	-8.9	1256	Chain A: VAL166 TYR167 PRO168 ASN170 LYS253 CYS257 ASP258 GLU259 TYR260 TYR272 SER275 MET286 MET288 LEU293 GLN296 LEU297 PRO298 GLN661 ARG662 HIS665 PHE666 TYR698 HIS701 GLY750 PHE751 LEU752 LEU755 ASN756 PRO757 ALA758 HIS759 GLN760 ASP787 ILE788 LEU793

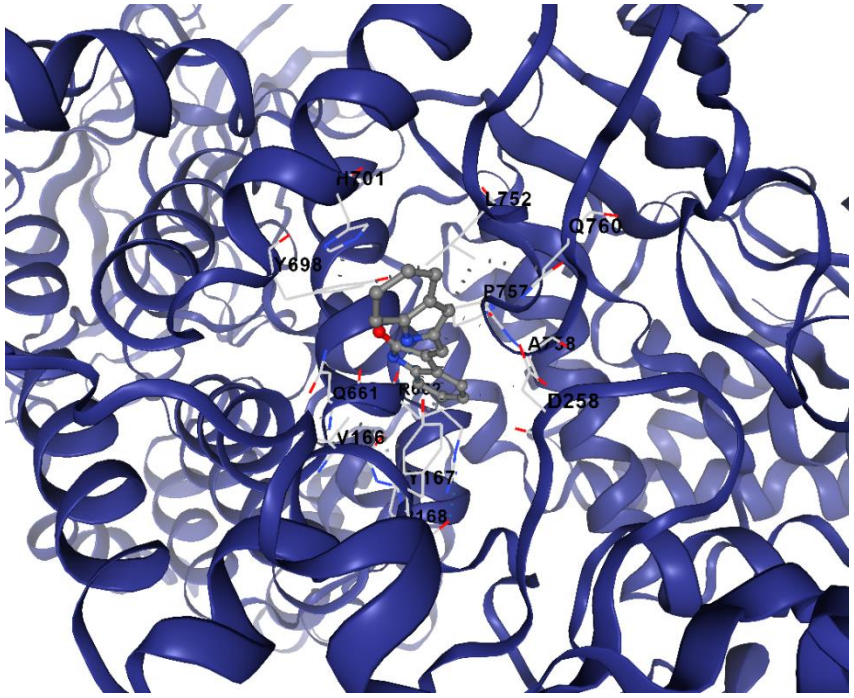


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
CHEMBL1165499	PIK3CA	-9.7	3260	Chain A: MET130 VAL131 LYS132 ASP133 PRO134 GLU135 VAL136 ASN426 ILE427 ASN428 LEU429 PHE430 ASP431 TYR432 THR433 THR435 LEU436 VAL437 SER438 MET441 ALA442 LEU443 TRP446 VAL461 THR462 GLY463 SER464 PRO466 LYS468 GLN643 TYR644 LEU645 LYS678 THR679 VAL680 SER681 GLN682 ARG683

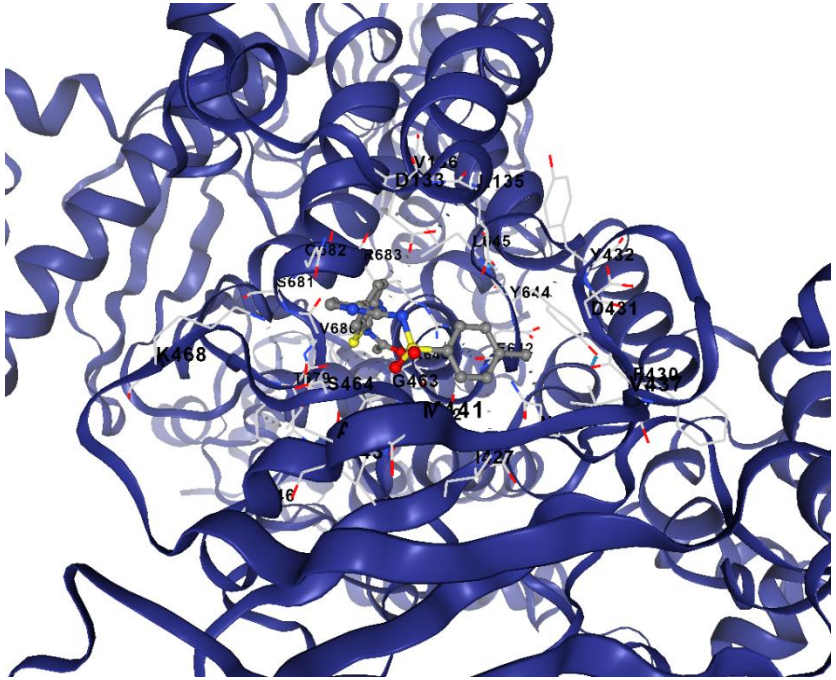


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
CHEMBL1773601	PIK3CA	-9.2	3409	Chain A: ASN170 GLU172 PRO178 GLU259 LYS271 TYR272 ARG274 SER275 MET278 LEU279 ASP626 SER629 GLN630 LEU632 ILE633 ARG662 PHE666 HIS670 LEU755 ASN756 ALA758 HIS759 LEU793 PHE794 ASN796 ASN797 GLU798 MET811 ARG818 GLU821 ASN822 GLN825 ASN826 ARG832 MET833 LEU834 PRO835 TYR836 GLY837 CYS838 LEU839 GLU849 VAL850 VAL851 ARG852 LYS924 GLN928

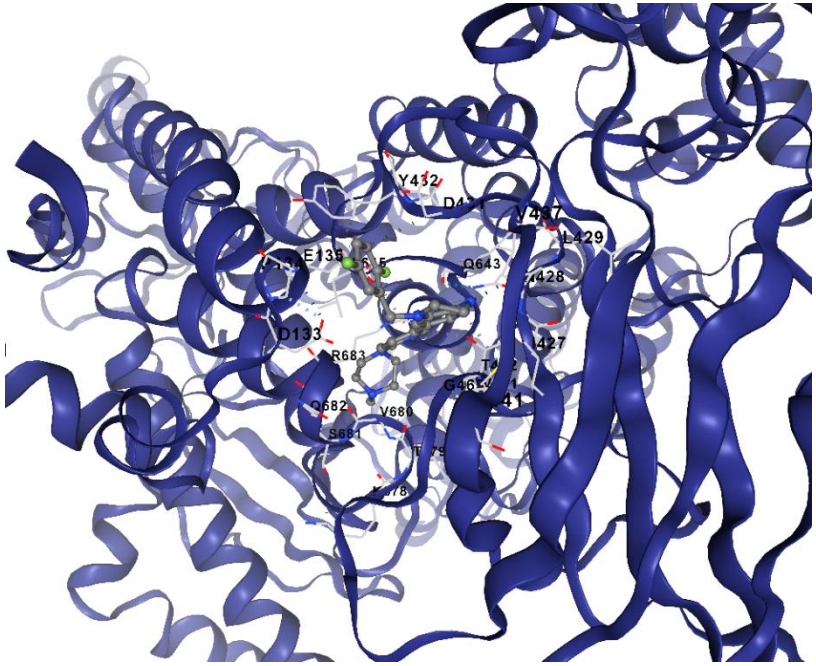


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
CHEMBL1773581	PIK3CA	-10.0	989	Chain A: PRO169 ASN170 VAL171 GLU172 GLU259 LYS271 TYR272 ARG274 SER275 MET278 LEU279 LYS594 GLU596 GLN597 GLU600 ASP625 ASP626 LYS627 SER629 GLN630 TYR631 LEU632 ILE633 ARG662 ILE663 PHE666 HIS670 LEU755 ASN756 HIS759 MET811 LEU814 GLN815 ARG818 GLU821 ASN822 ILE823 GLN825 ASN826 ARG832 MET833 LEU834 PRO835 TYR836 GLY837 CYS838 LEU839 GLU849 ARG992 GLN993 HIS994 ALA995 ASN996 LEU997

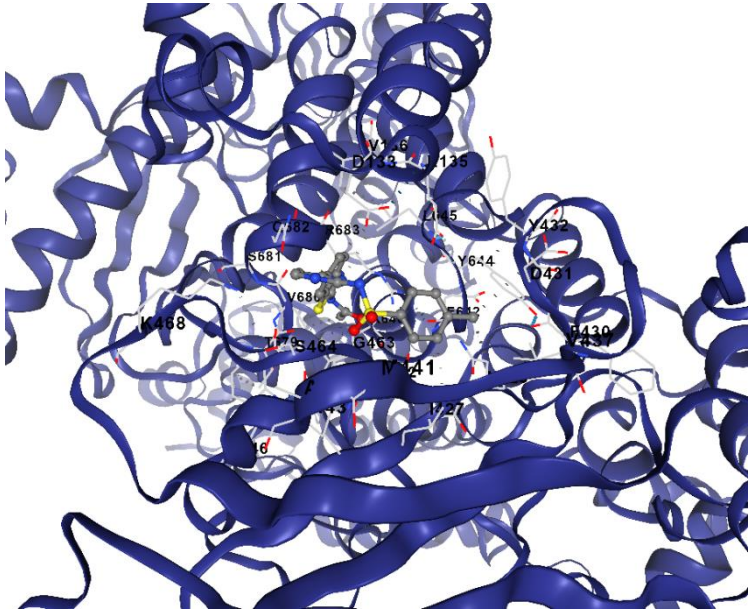


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
CHEMBL388978	PIK3CA	-9.1	989	Chain A: ASN170 GLU172 GLU259 LYS271 TYR272 ARG274 SER275 MET278 LEU279 ASP626 SER629 GLN630 ARG662 PHE666 LEU755 ASN756 HIS759 MET811 ARG818 GLU821 ASN822 GLN825 ARG832 MET833 LEU834 PRO835 TYR836 GLY837 CYS838 LEU839 GLU849 VAL851 LYS924 GLN928

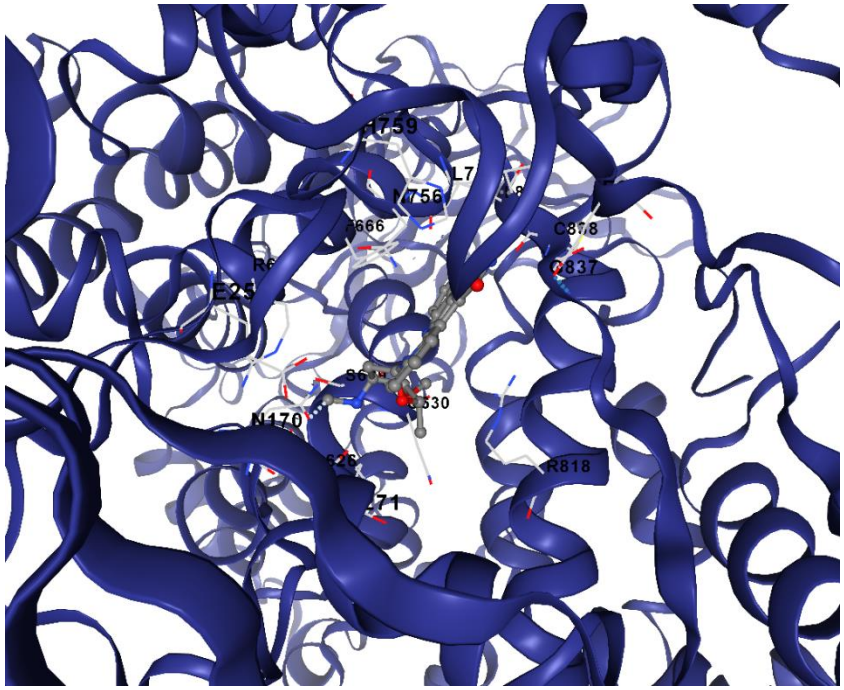


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
CHEMBL1615189	PIK3CA	-9.5	3260	Chain A: VAL131 LYS132 ASP133 PRO134 GLU135 VAL136 ASN426 ILE427 ASN428 PHE430 ASP431 TYR432 THR435 LEU436 VAL437 SER438 MET441 ALA442 LEU443 TRP446 VAL461 THR462 GLY463 SER464 ASN465 PRO466 LYS468 GLN643 TYR644 LEU645 LYS678 THR679 VAL680 SER681 GLN682 ARG683 LEU686

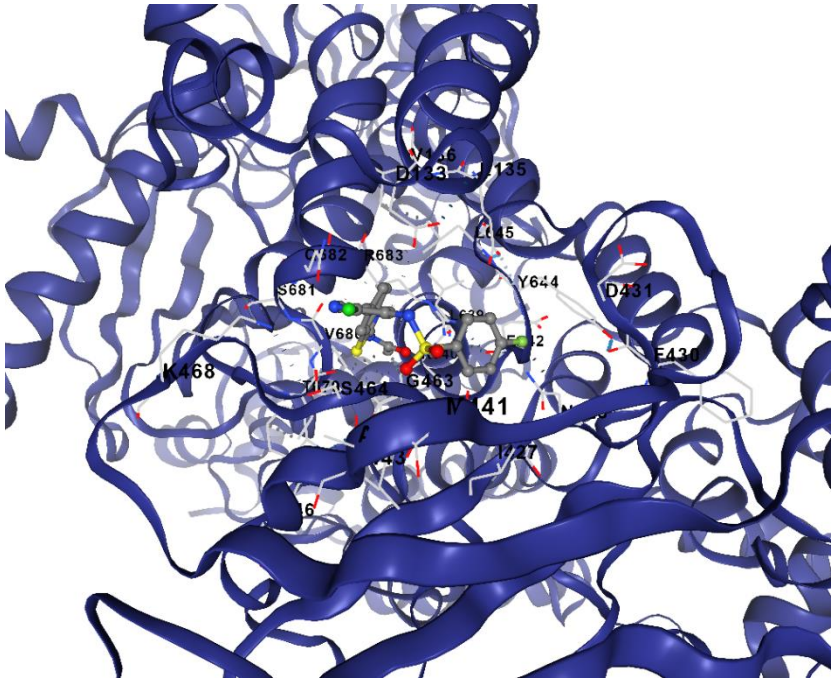


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
Lenvatinib	PIK3CA	-8.7	3260	Chain A: MET123 GLU127 MET130 VAL131 LYS132 ASP133 PRO134 GLU135 VAL136 ILE427 ASN428 PHE430 ASP431 TYR432 THR433 THR435 LEU436 VAL437 SER438 MET441 ALA442 TRP446 VAL461 THR462 GLY463 SER464 ASN465 PRO466 ASN467 LYS468 GLU642 GLN643 TYR644 LEU645 LYS678 THR679 VAL680 SER681 GLN682 ARG683

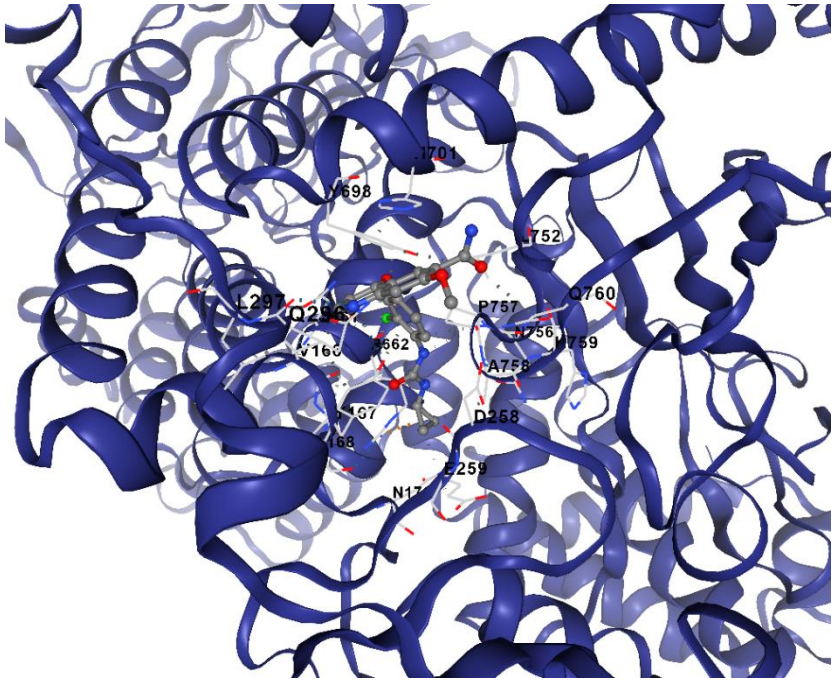


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
Regorafenib	PIK3CA	-9.6	3260	Chain A: VAL131 ASP133 PRO134 GLU135 VAL136 ASP138 ILE427 ASN428 PHE430 ASP431 TYR432 THR433 THR435 LEU436 VAL437 SER438 MET441 ALA442 TRP446 VAL461 THR462 GLY463 SER464 ASN465 PRO466 LYS468 TRP479 VAL483 GLU642 GLN643 TYR644 LEU645 LYS678 THR679 VAL680 GLN682 ARG683

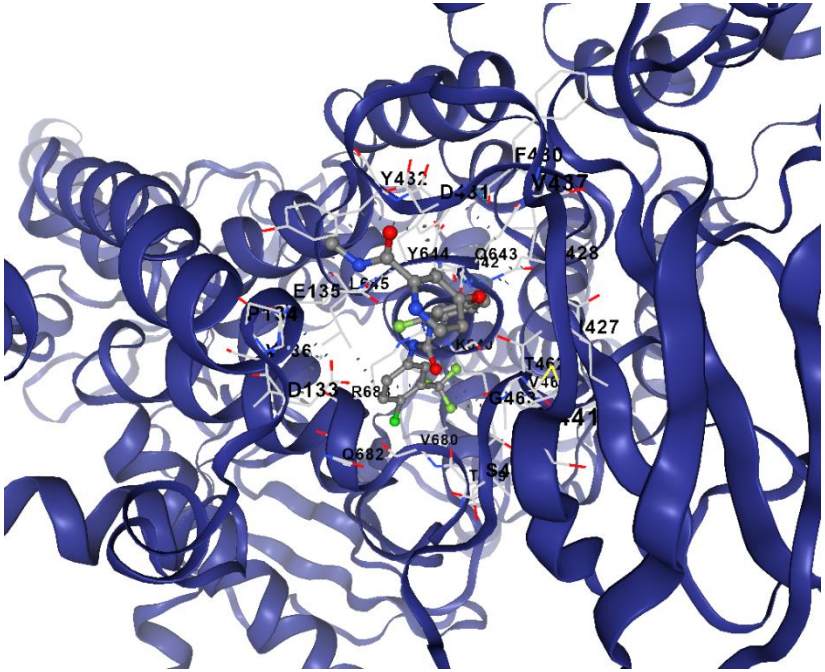
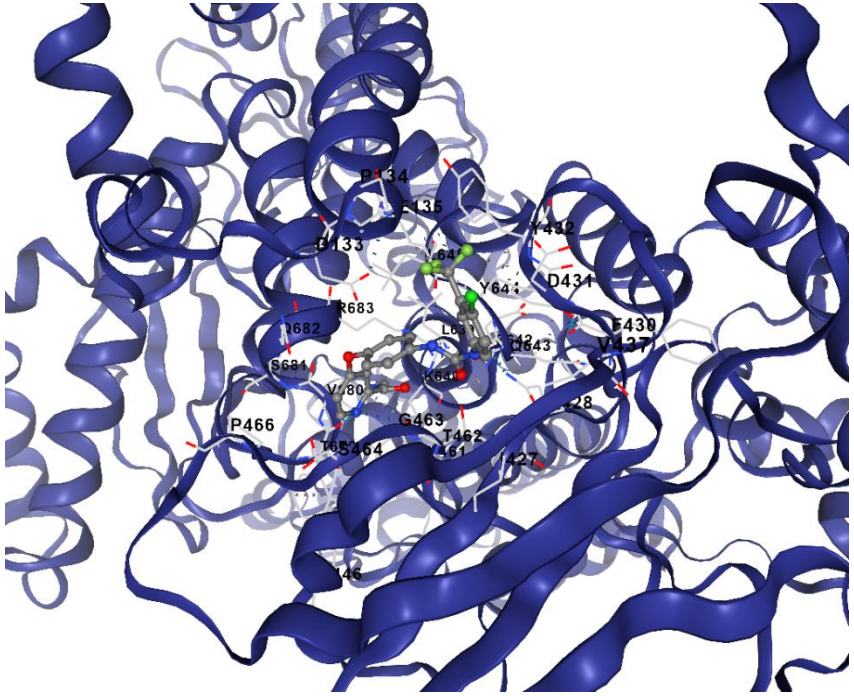


Table S5: (cont.)

Ligand (Drug/compound)	Target Protein	Vina score	Cavity volume (Å^3)	Contact residues
Sorafenib	PIK3CA	-9.8	1188	Chain A: GLU109 GLU110 LYS111 ILE112 LEU113 ASN114 ARG115 GLU116 GLY118 PHE119 GLY122 MET123 PRO124 VAL125 CYS126 ARG154 ASP300 CYS301 PHE302 THR303 CYS692 ARG693 ALA694 CYS695 GLY696 MET697 LEU699 LYS700 ASN703 ARG704 GLU707



CURRICULUM VITAE

Surname, Name	: Başer, Tuğçe
Nationality	: T.C.
Marital Status	: Single

EDUCATIONAL BACKGROUND:

Ph.D. **:** Middle East Technical University, Department of Medical Informatics (ANKARA) (2016-) 3.88/4.00
Thesis Title **:** "Application of Machine Learning Methods in Drug Development for Liver Cancer"
Advisor **:** Dr. Öğr. Üyesi Burçak OTLU SARITAŞ
Co-advisor **:** Prof. Dr. Rengül ÇETİN ATALAY

Master's Degree **:** Middle East Technical University, Graduate School of Natural and Applied Sciences, Department of Biology (2010-2013) 3.64/4.00

Thesis Title **:** "Determination of Lipid Profiles in Genetic Obese Mice Adipose Tissues Using High-Performance Liquid Chromatography"

Advisor **:** Prof. Dr. Feride SEVERCAN

Undergraduate **:** Middle East Technical University, Faculty of Science, Department of Biology (ANKARA) (2005-2010) 2.85/4.00

High School **:** M.E.V. Private Ankara Science High School (ANKARA) (2002-2005) 5.00/5.00

PROFESSIONAL EXPERIENCE:

October 2022 - Present: TÜBİTAK ARGES Senior Expert in Scientific Programs

September 2018 - October 2022: TÜBİTAK ARGES Expert in Scientific Programs

February 2016 - September 2018: TÜBİTAK ARGES Assistant Expert in Scientific Programs

December 2014 - February 2016: TÜBİTAK KBAG Assistant Expert in Scientific Programs

LANGUAGE PROFICIENCY:

Turkish (Native), English (Intermediate)

COMPUTER SKILLS:

Microsoft Word, Excel, PowerPoint, Windows 11

OTHER INFORMATION

INTERNSHIPS:

September 2010 - September 2013: Middle East Technical University
Molecular Biophysics Laboratory
TÜBİTAK Project
3 years, Master's Student
Supervisor: Prof. Dr. Feride SEVERCAN

July 2009 - August 2009: Hacettepe Hospital
Department of Basic Oncology
1.5 months, Summer Internship (mandatory)

2008-2010: Cancer, Molecular Biology, and Drug Resistance
Cancer Laboratory at Middle East Technical University
TÜBİTAK Project
Volunteer Project Assistant
Supervisor: Prof. Dr. Ufuk GÜNDÜZ

ACHIEVEMENTS:

Middle East Technical University High Honor Certificate for the Spring 2010 Semester

Three Middle East Technical University Honor Certificates between Fall 2008 and Spring 2009

First Place Certificate at M.E.V. Private Ankara Science High School (2005)

RESEARCH PROJECTS:

TÜBİTAK, Characterization of Structure, Function, and Content of Different Adipose Tissues in Inbred Obese Mouse Models, Project Code: 110S23, Project Supervisor: Prof. Dr. Feride SEVERCAN, Scholar: Tuğçe BAŞER (2011-2012)

Middle East Technical University Scientific Research Projects Unit, Targeting Cancer Cells with Doxorubicin-Loaded Nanoparticles, Project Supervisor: Prof. Dr. Ufuk GÜNDÜZ (2008-2010)

TÜBİTAK, Synthesis of Idarubicin-Loaded Magnetic Nanoparticles and Application to Mcf-7 Breast Cancer Cell Line, Project Code: 109T949, Project Supervisor: Prof. Dr. Ufuk GÜNDÜZ (2009)

TRAINING:

Ministry of Industry and Technology, Remote Education Platform Crisis Management Training, In-Service Training (Online) (16 December 2021)

Ministry of Industry and Technology, Remote Education Platform Effective Presentation Techniques, In-Service Training (Online) (20 October 2021)

BTK Academy, Introduction to Big Data Training (Online) (20 October 2021)

BTK Academy, Written Communication Techniques Training (Online) (19 October 2021)

Ministry of Industry and Technology, Remote Education Platform Importance of Motivation for Employee Productivity, In-Service Training (Online) (20 August 2021)

Ministry of Industry and Technology, Remote Education Platform Human Rights, In-Service Training (Online) (26 July 2021)

Ministry of Industry and Technology, Remote Education Platform Archival Legislation and Archival Awareness, In-Service Training (Online) (14 July 2021)

Ministry of Industry and Technology, Remote Education Platform Ethics in the Framework of Human Rights, In-Service Training (Online) (12 July 2021)

Ministry of Industry and Technology, Remote Education Platform Good Governance Principles in the Context of the Ombudsman Institution, In-Service Training (Online) (12 July 2021)

Ministry of Industry and Technology, Remote Education Platform Gender Equality in the Workplace Based on Human Rights, In-Service Training (Online) (12 July 2021)

Ministry of Industry and Technology, Remote Education Platform Risk Taking / Problem Solving and Decision Making / Result Orientation, In-Service Training (Online) (1 April 2021)

Ministry of Industry and Technology, Remote Education Platform Awareness Training on the Power of Communication / Zafer Kiraz, In-Service Training (Online) (1 March 2021)

Ministry of Industry and Technology, Remote Education Platform Official Correspondence Procedures and Principles, Mandatory Training (Online) (22 February 2021)

Ministry of Industry and Technology, Remote Education Platform Oratory and Effective Presentation Training, In-Service Training (Online) (16 February 2021)

Ministry of Industry and Technology, Remote Education Platform Body Language and Communication Training, In-Service Training (Online) (16 February 2021)

Ministry of Industry and Technology, Remote Education Platform Public Ethics and Ethical Behavior Principles for Public Officials, Mandatory Training (Online) (31 January 2021)

Ministry of Industry and Technology, Remote Education Platform Diction, Mandatory Training (Online) (30 January 2021)

Ministry of Industry and Technology, Remote Education Platform Importance of Content in Communication, Mandatory Training (Online) (29 January 2021)

Ministry of Industry and Technology, Remote Education Platform Public Speaking, Mandatory Training (Online) (29 January 2021)

Ministry of Industry and Technology, Remote Education Platform Common Postural Problems and Office Exercises for Office Workers, Mandatory Training (Online) (5 November 2020)

Ministry of Industry and Technology, Remote Education Platform Applied Excel Training, Mandatory Training (Online) (15 October 2020)

TÜBİTAK, Basic Photography Training (ANKARA) (3 October 2017 - 21 November 2017)

Middle East Technical University Basic First Aid Training (26.02.2010-28.05.2010)

CERTIFICATES:

Ministry of Health, First Aid Certificate (2021)

Yaşamartı, Participation Certificate at Yaşamartı Personal Development Summit: "Don’t Worry About Everything in Life" by Ozanser UĞURLU (ANKARA) (22 October 2017)

Yaşamartı, Participation Certificate at Yaşamartı Personal Development Summit: "The Power of Knowledge" by Prof. Dr. İlber ORTAYLI (ANKARA) (22 October 2017)

Yaşamartı, Participation Certificate at Yaşamartı Personal Development Summit: "Simplicity in Communication: Effective Listening & Correct Listening & Correct Interpretation" by Deniz BAYRAMOĞLU (ANKARA) (22 October 2017)

Yaşamartı, Participation Certificate at Yaşamartı Personal Development Summit: "The Future of Business Life" by Erdem AKSAKAL (ANKARA) (22 October 2017)

Gazi University, Faculty of Education, Department of Biology Teaching, Pedagogical Formation Certificate (ANKARA) (2012-2013)

Middle East Technical University Basic First Aid Certificate (2010)

EVENTS:

TEDxMETU, "Great Illusions" (ANKARA) (3 December 2017)

SOCIAL CLUBS:

METU Bio-Gen Society (2005-2010)

M.E.V. Private Ankara Science High School Volleyball Team (2002-2005)

M.E.V. Private Ankara Primary School Volleyball Team (1998-2002)