SYSTEMS BIOLOGY ANALYSIS OF KINASE INHIBITORS IN LIVER CANCER CELLS USING NEXT GENERATION SEQUENCING DATA

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

BY

KÜBRA NARCI

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

SEPTEMBER 2021

SYSTEMS BIOLOGY ANALYSIS OF KINASE INHIBITORS IN LIVER CANCER CELLS USING NEXT GENERATION SEQUENCING DATA

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

Date: 06.09.2021

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 : Kübra Narcı

Signature : _____

ABSTRACT

SYSTEMS BIOLOGY ANALYSIS OF KINASE INHIBITORS IN LIVER CANCER CELLS USING NEXT GENERATION SEQUENCING DATA

Narcı, Kübra
Ph.D, Department of Medical Informatics
Supervisor: Assoc. Prof. Yeşim Aydın Son
Co-Supervisor: Prof. Dr. Rengül Çetin Atalay

September 2021, 155 pages

The underlying mechanism for the development of Hepatocellular Carcinoma (HCC) is highly complex due to tissue heterogeneity. Although the traditional approaches mainly focus on a single gene or locus, understanding the variations in the signaling pathways of cancerogenic cells during hepatocarcinogenesis may help to develop novel strategies for treatment and drug development to prevent cancer progression in the patients.

This thesis study primarily focuses on unveiling the transcriptome sequencing of differentially expressed genes in HCC, which mainly concentrate on known disease signaling pathways. For this purpose, RNA-seq data of two HCC cell lines were targeted by three different kinase inhibitors and two of their combinations with Sorafenib. The functional pathways enriched with differentially expressed genes were identified by solving a graph problem called as Prize Collecting Steiner Tree (PCST) on human interactome generating inhibitor specific networks. As a result of this study, we found that combinatory treatment of Sorafenib with PIK-75 to HCC cell lines Huh7 and Mahlavu stimulates apoptosis, while TGX-221 with Sorafenib strikingly promotes cell growth antagonizing cellular death, especially for Mahlavu cell line. The states of transcriptomes for different kinase inhibitors were visualized using Cytoscape and molecular interactions were scanned deeply to understand synergistic or antagonistic effects of these kinase inhibitory treatments.

Hence, this study provides comprehensive pathways analysis for differential kinase inhibitor reactions of HCC. Using these data, novel HCC drug targets were identified which may lead to more cost-effective and diverse treatment options available for the treatment of liver cancer.

Keywords: Hepatocellular Carcinoma, Cellular Signaling Pathways, RNA Sequencing, Prize Collecting Steiner Algorithm, Network Modeling

ÖZ

YENİ NESİL DİZİLEME TEKNİĞİ KULLANILARAK KARACİĞER KANSERİNİN SİSTEM BİYOLOJİSİ ANALİZİ

Narcı, Kübra
Doktora, Tıp Bilişimi Bölümü
Tez Yöneticisi: Doç. Dr. Yeşim Aydın Son
Eş Danışman: Prof. Dr. Rengül Çetin Atalay

Eylül 2021, 155 sayfa

Hepatosellüler Kanser (HCC) gelişimi altında yatan mekanizma, kanser dokularının heterojenliği nedeniyle oldukça karmaşıktır. Geleneksel yaklaşımlar temelde tek gen veya lokusa odaklansa da hepatokarsinogenez sırasında kanserojen hücrelerin sinyal yollarındaki varyasyonları anlamak, hastalardaki kanserin ilerlemesini önlemek adına tedavi ve ilaç keşfi için yeni stratejiler geliştirmeye yardımcı olabilir.

Bu doktora tez çalışmasının amacı, HCC içerisinde değişken eksprese edilen genlerin başlıca bilinen hastalık sinyal yolaklarında yoğunlaşarak transkriptom dizilemesi yoluyla ortaya çıkarılmasıdır. Bu amaçla, iki çeşit HCC hücre hattı, üç farklı kinaz inhibitörünün tekli veya Sorafenib ile kombinasyonları olacak şekilde hedeflenerek RNA dizilemesi elde edildi. Price Collecting Steiner Tree (PCST) algoritmasının insan transkriptom ağı üzerinde, diferansiyel kontrol edilen genlerle çözümü bize bu genlerle yoğunlaştırılmış fonksiyonel yolaklar sunmuştur. Bu tez çalışması sonucunda, Sorafenib ve PIK-75 inhibitörlerinin birlikte kullanılması ile hem Huh7 hem de Mahlavu hücre hatlarında apoptozu uyardığını buna karşın TGX-221 inhibitörü ile Sorafenib'in birlikte kullanılmasının çarpıcı bir şekilde hücre büyümesini desteklediğini bulduk. Bu kombinasyonun özellikle Mahlavu hücre hattında Sorafenib ile oluşan hücresel ölümü antagonize ettiği gösterilmiştir. Farklı kinaz inhibitörleri ile elde edilen diferansiyel gen ekspresyon statüleri, Cytoscape aracı kullanılarak görselleştirilmiş ve bu kinaz inhibitörlerinin olası sinerjistik ve antagonistik etkilerini anlamak adına yolaklar

içerisindeki moleküler etkileşimler derinlemesine analiz edilerek karşılaştırmalar yapılmıştır.

Dolayısı ile, bu çalışma HCC hücrelerinin kinaz inhibitörleri karşısında oluşturduğu diferansiyel ekspresyonları incelemek adına kapsamlı bir yaklaşım sunmaktadır. Bu veriler ışığında, karaciğer kanser tedavisi için daha uygun maliyetli seçenekleri çeşitlendirebilecek yeni ilaç hedefleri belirlenmiştir.

Anahtar Sözcükler: Hepatoselüler Kanser, Hücresel Sinyal Yolakları, RNA dizileme, PCST, Ağ Modellemesi

To My Family

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitude to Dear Prof. Dr. Rengül Atalay who was the principal investigator of the project that I am in. She is the advisor of this thesis project. I am grateful to her for an original thesis topic and her continuous support, motivation, and immense knowledge. I cannot thank her enough for her continuous guidance and endless patience.

I also would like to thank Assoc. Prof. Yeşim Aydın Son for guiding me through the progress of this thesis and accepting to be there for me. Yet, I have known her for 12 years now, I was involved in many of the courses that she was instructing, I have learned a lot from her. I am grateful to her both for her professional guidance and friendly-looking face.

I express my gratitude to Assoc. Prof. Nurcan Tunçbağ and Assoc. Prof. Özlen Konu for being a part of my thesis committee. They contributed to this thesis study with their priceless commends. I want to thank them their positive attitude towards me, their help during all thesis meetings, and their participation.

Among examining committee members, Prof. Dr. Ayse Elif Erson Bensan and Assist. Prof. Aybar Can Acar are greatly acknowledged for their participation to the thesis defense.

I am thankful to all CANSIL members, being a part of this group was an honor for me. I want to acknowledge Deniz Cansen Kahraman and Altan Koyaş for editing the paper that originated from this study and leading and performing the additional experimental parts. I also acknowledge Dr. Tülin Erşahin, who is a former member of the laboratory, as performing RNA-seq experiments. I am also grateful to my dear friend Damla Gözen she and I were great companions during the whole doctorate process.

I am thankful to all my coworkers from Seven Bridges Genomics family; especially to Alper Döm and Duygu Kabakçı for always listening to my complaining when I was down, being there for me and supporting me in my difficult times. Furthermore, apart from being a side-to-side coworker to me, Elif Arslan is one of my best friends, she never denies her assistance to me, and she contributed this thesis study with her valuable comments and edits.

During my thesis preparation process, one of the people who was always with me was Burcu Yaldız. She and I overcome big challenges in the doctorate together. I always feel her support for me even if we are separated by kilometers now.

Sadly enough, my other physically distant friends are Yasin Tümtaş and Seda Koyuncu. Happily, their existence through many video conferences were our way of meditation during our full plate of workdays.

I have also never lost the connection with my friends Berke Gürkan and Alişan Kayabölen. These two multi-talented people are the spirit on my way to become a successful scientist.

I am also grateful to my dearest friends Fatma Akıncı, Ömer Aydın, Esra Bozdoğan, Ayşenur Erdoğan, Songül Taş and Erdem Aydın (cutest member of the crew) for their never-ending patience to all my shortcomings. They are my joy, traveling companions, shoulders to cry on and the most confidant fellows.

This study is foremost dedicated to my family, my father Hayati Narcı, my mother Fatma Narcı, my sister Ebru Narcı Bulut and my brother D. Eren Narcı. I cannot imagine being successful without their guidance. They always accepted me as I am, loved me and respected my wishes. I would like to thank them for supporting me spiritually throughout my life.

I acknowledge the Scientific and Technological Research Council of Turkey (TUBİTAK) for its thesis grant under 2211 BIDEB, ODTU BAP BAP-08-11-2015-013, and TUBITAK-110S388.

TABLE OF CONTENTS

ABSTRACT iv
ÖZ vi
DEDICATION viii
ACKNOWLEDGMENTS ix
TABLE OF CONTENTS xi
LIST OF TABLES xv
LIST OF FIGURES xvii
CHAPTERS
1. INTRODUCTION 1
1.1. Hepatocellular Carcinoma 1
1.1.1. Risk Factors 2
1.1.2. Initiation and Development 3
1.1.3. Acquired Capabilities Through Development of HCC 4
1.1.4. Genetic Heterogeneity 6
1.2. PI3K/AKT/mTOR Signaling Pathway 6
1.2.1. PI3K Kinase Classes 8
1.3. Current therapies for HCC 9
1.3.1. Sorafenib Inhibitor 9
1.3.2. PIK-75 and TGX-221 Inhibitors 11
1.3.3. Combinational Therapy 11
1.3.4. Drug Repurposing 12
1.4. A Survey on the Methods Used for Conventional Drug Target Studies 12
1.4.1. RNA Sequencing 13
1.4.2. Protein-Protein Interaction Databases 15
1.4.3. Protein-Protein Interaction Prediction 16
1.4.4. Network/Pathway Analysis 17
1.4.5. Visualization of the Biological Networks 18

1.4.6.The Systems Biology Approach 18
1.4.7.Omics Integrator 19
1.4.7.1.Prize Collecting Steiner Tree (PCST) algorithm 20
1.5. Preliminary Experimental Analysis at CANSYL 21
1.5.1.Molecular and Cellular Characterization of HCC in the Presence of Small Molecule Isoform Spesific PI3K Inhibitors 21
1.5.2.Migration Analysis of the Inhibitors 23
1.5.3.Synergistic cytotoxicity analysis 23
1.6. Motivation and Rationality 25
2. MATERIALS AND METHODS 29
2.1. Materials 29
2.1.1.HCC Cell Lines 29
2.1.2.Kinase Inhibitors 30
2.2. Methods 31
2.2.1.Cell Culture 32
2.2.2.RNA Sequencing Analysis 33
2.2.3.Preparation and Sequencing of Total RNAs 33
2.2.4.Quality Assessment of Raw Files 34
2.2.5.Alignment 34
2.2.6.Alignment Quality 34
2.2.7.Quantification 35
2.2.8.Differential Expression Analysis 35
2.2.9.Dendrogram Analysis 36
2.2.10.Correlation Analysis 36
2.2.11.Venn Analysis 36
2.2.12.Gene Ontology (GO) Analysis 36
2.2.13.Network Analysis 37
2.2.14.Optimization of Forest Parameters 38
2.2.15.Forest Module Runs 39
2.2.16.Randomization Tests 40
2.2.17.Centrality Measurements 40

2.2.18.Effective Visualization of the Optimal Networks 41
2.2.19.Clustering 42
2.2.20.Annotation 42
2.2.21.GO Analysis and Mapping 43
2.3. Prioritization of the Nodes in the Optimal Networks 43
2.4. Knock-out Experiment 43
3. RESULTS 45
3.1. RNA Sequencing Results 45
3.1.1.Quality Check Reports 45
3.1.2.Alignments and Gene Counts 46
3.1.3.Differential Expression Analysis 49
3.1.4.Correlation Analysis of Kinase Inhibitors 52
3.1.5.Top 50 Common DEGs 53
3.1.6.Differentially Expressed Untranslated Transcripts 56
3.1.7.Gene Enrichment Analysis of Differential Expression Patterns 58
3.2. Network-Based Interpretation of the Data 61
3.2.1.Optimal PCST Networks 61
3.2.2.Comparison of Optimal Network Nodes 63
3.2.3.Cluster Specific Gene Ontologies of Optimal PCST Networks 64
3.2.4.Comparison of Cluster Specific Gene Enrichments 76
3.2.5.Prioritizied Genes as Drug Targets 79
3.2.6.The Effect of Gene Removals from the Optimal Networks 81
4. DISCUSSION 83
5. CONCLUSION 95
REFERENCES 97
APPENDICES
APPENDIX A 115
APPENDIX B 121
APPENDIX C 123
APPENDIX D 127
APPENDIX E 131
APPENDIX F 147

APPENDIX F 147
CURRICULUM VITAE 151

LIST OF TABLES

Table 1: Kinase inhibitors used in RNA library construction. 31

Table 2: Experimental evidence codes used in the study. 37

Table 3: Forest parameters and their effect into the re-generated sub-interactomes. 39

Table 4: Total sequences, sequence length and GC% content of the FASTQ files of the HCC Cell lines. 46

Table 5: Number of total processed sequences, alignment rate in percentages, and the coverage of the alignments for HCC Cell lines. 47

Table 6. Housekeeping genes used for BCV calculation. 49

Table 7: The number of differentially expressed genes and untranslated transcripts for HCC cell lines under various inhibitor treatment conditions. 50

Table 8: Pearson correlation matrix between Huh7 and Mahlavu inhibitor treatments. . 53

Table 9. Ensembl ID, gene description, gene name and the regulation type in Huh7 cell line of the most common differentially expressed genes. 55

Table 10. Ensembl ID, gene description, gene name and the regulation type in Mahlavu cell line of the most common differentially expressed genes. 56

Table 11: Selected parameters for PSCT analysis using forest-tuner and numbers of nodes, terminals, and prizes of generated networks. 62

Table 12. Significant gene ontologies for inhibitor specific networks in Huh7 cell line. 78

Table 13. Significant gene ontologies for inhibitor specific networks in Mahlavu cell line. 79

Table 14: Prioritized genes as potential drug targets in Huh7 and Mahlavu HCC lines. 82

Table 15. Differentially expressed *DUSP* genes in Huh7 cells and their target kinases based on Huang & Tan, 2012. 91

Table 16: Cluster specific gene ontologies for PIK-75 treated Huh7 cells 131

Table 17: Cluster specific gene ontologies for TGX-221 treated Huh7 cells 133

Table 18: Cluster specific gene ontologies for PIK-75 and Sorafenib treated Huh7 cells 133

Table 19: Cluster specific gene ontologies for TGX-221 and Sorafenib treated Huh7 cells.....135

Table 20: Cluster specific gene ontologies for Sorafenib treated Huh7 cells 138

Table 21: Cluster specific gene ontologies for PIK-75 treated Mahlavu cells.....140

Table 22: Cluster specific gene ontologies for TGX-221 treated Mahlavu cells.....141

Table 23: Cluster specific gene ontologies for PIK-75 and Sorafenib treated Mahlavu cells.....141

Table 24: Cluster specific gene ontologies for PI3Ki-beta and Sorafenib treated Mahlavu cells.....145

Table 25: Cluster specific gene ontologies for Sorafenib treated Mahlavu cells.....146

LIST OF FIGURES

Figure 1:Percentages of new cancer cases and cancer deaths worldwide in 2018 (Bray et al., 2018).....2

Figure 2: Risk factors and development of HCC.. 3

Figure 3: Acquiring tumorigenic capabilities from normal liver to Hepatocellular carcinoma development 4

Figure 4: Overview of PI3K/Akt/mTOR signaling pathway. 7

Figure 5: Chemical structures of multi-kinase inhibitor Sorafenib, and of p110α and p110β inhibitors PIK-75 and TGX-221. 11

Figure 6: RNA-seq analysis workflow.....15

Figure 7: An example representation to show how the PCST algorithm performs its calculation 21

Figure 8. Characterization of HCC cells in the presence of small molecules inhibitors .22

Figure 9. Real-time cell growth analysis.....24

Figure 10: The aim of the study 27

Figure 11: Expression and activity of critical components of PI3K/AKT and RAF/MEK/ERK signaling pathways in Huh7 and Mahlavu. 30

Figure 12: Schematic representation of the kinase inhibitors used in the study, their target proteins, and the downstream proteins of their targets. 31

Figure 13: Pipeline summarizing the methods used in this thesis study.....32

Figure 14: Summary of RNA-seq analysis pipeline 33

Figure 15: Summarization of the methods used for effective visualization of optimal networks. 42

Figure 16: A. Library sizes per samples, B. gene distributions per samples resulted by HTSeq-count 48

Figure 17: Venn diagrams representing common and unique Huh7 and Mahlavu DEG numbers. 51

Figure 18: Pearson correlations of kinase inhibitor treatments for Huh7 and Mahlavu ..52

Figure 19: Heatmaply dendrograms representing top 50 most common DEGs of Huh7 and Mahlavu cell lines 54

Figure 20: Heatmaply dendrograms representing untranslated transcripts of Huh7 and Mahlavu cell lines 57

Figure 21: Heatmap of gene expressions illustrated as dendrograms for Huh7 cell line.. 59

Figure 22: Heatmap of gene expressions illustrated as dendrograms for Mahlavu cell line. 60

Figure 23: Venn diagrams representing the common and unique number of optimal network nodes of Huh7 and Mahlavu cells. 63

Figure 24: A schematic representation of an optimal network of DEGs upon PI3K-α inhibitor (PIK-75) treatment of Huh7 cell line.....65

Figure 25: A schematic representation of an optimal network of DEGs upon PI3K-β inhibitor (PIK-75) treatment of Huh7 cell line.....66

Figure 26: A schematic representation of an optimal network of DEGs upon PI3K-α inhibitor (PIK-75) with Sorafenib treatment of Huh7 cell line.....68

Figure 27: A schematic representation of an optimal network of DEGs upon PI3K-β inhibitor (PIK-75) with Sorafenib treatment of Huh7 cell line.....69

Figure 28: A schematic representation of an optimal network of DEGs upon Sorafenib treatment of Huh7 cell line.....70

Figure 29: A schematic representation of an optimal network of DEGs upon PI3K-α inhibitor (PIK-75) treatment of Mahlavu cell line. 71

Figure 30: A schematic representation of an optimal network of DEGs upon PI3K-β inhibitor (PIK-75) treatment of Mahlavu cell line. 72

Figure 31: A schematic representation of an optimal network of DEGs upon PI3K-α inhibitor (PIK-75) with Sorafenib treatment of Mahlavu cell line.....74

Figure 32: A schematic representation of an optimal network of DEGs upon PI3K-β inhibitor (PIK-75) with Sorafenib treatment of Mahlavu cell line 75

Figure 33. A schematic representation of an optimal network of DEGs upon Sorafenib treatment of Mahlavu cell line 76

Figure 34: Heatmap of gene enrichments illustrated as dendrograms for Huh7 and Mahlavu cell lines. 77

Figure 35. Prioritized nodes for Huh7 and Mahlavu were ranked by betweenness centrality values of randomized networks for each inhibitor treatments. 80

Figure 36: Up (red) and down (blue) regulations of the prioritized genes in Huh7 and Mahlavu cell lines, from which treatment they are selected was pointed out.....81

Figure 37. The sub-networks representing different DNA-repair mechanisms in different Huh7 treatments 89

Figure 38. The sub-networks illustrating difference in EGR1 activity in Mahlavu cells.92

Figure 39. Realtime cell growth analysis of Huh7 and Mahlavu cells with increasing concentrations. 115

Figure 40. Wound healing assay for 24 and 48 hours for cell migration.....116

Figure 41. Cell cycle analysis with flow cytometry.....117

Figure 43. Real-time cell growth analysis with various concantrations and combinations. 118

Figure 44. SynergyFinder view for Huh7 treated with Sorafenib and PI3Ki-α. 119

Figure 45. SynergyFinder view for Huh7 treated with Sorafenib and PI3Ki-β. 119

Figure 46. SynergyFinder view for Mahlavu treated with Sorafenib and PI3Ki-α.....120

Figure 47. SynergyFinder view for Mahlavu treated with Sorafenib and PI3Ki-β.....120

Figure 48: Quality score plot for DMSO treated Huh7 and Mahlavu cells 127

Figure 49: Quality score plots for PIK-75 treated Huh7 and Mahlavu cells 127

Figure 50: Quality score plots for TGX-221 treated Huh7 and Mahlavu cells.....128

Figure 51: Quality score plots for Sorafenib treated Huh7 and Mahlavu cells.....128

Figure 52: Quality score plots for combinational treatment of PIK-75 and Sorafenib to Huh7 and Mahlavu cells.....128

Figure 53: Quality scores for combinational treatment of TGX-221 and Sorafenib to Huh7 and Mahlavu cells.....129

Figure 54. Copyright permission of the image used in the Figure 1.....147

Figure 55. The copyright permission of the image used in Figure 11. 148

Figure 56. Copyright permission of the figure used in Figure 14. 149

LIST OF ABBREVIATIONS

HCC	Hepatocellular Carcinoma
FDA	Food and Drug Administration
RNA-seq	RNA Sequencing
ROS	Reactive Oxygen Species
PCST	Prize Collecting Steiner Tree Algorithm
PPI	Protein-Protein Interaction
KEGG	Kyoto Encyclopedia of Genes and Genomes
DMSO	Dimethyl Sulphoxide
DEG	Differentially Expressed Genes
GSEA	Gene Set Enrichment Analysis
GO	Gene Ontology
PI3K	Phosphatidylinositol 3-kinase
PI3K-α	Phosphatidylinositol 3-kinase alpha inhibitor (PIK-75)
PI3K-β	Phosphatidylinositol 3-kinase beta inhibitor (TGX-221)
MAPK	Mitogen activated protein kinase
VEGF	Vascular endothelial growth factor
NF-κB	Nuclear factor κB
PTEN	Phosphatase and tensin homologue
PDGF	Platelet-derived growth factor
mTOR	Mammalian target of rapamycin
EGF	Epithelial growth factor
FGF	Fibroblast growth factor
HGF	Hepatocyte growth factor
TGF	Transforming growth factor
TAM	Tumor-associated macrophage
ALPHA	PI3K-α inhibitor (PIK-75) treatment
BETA	PI3K-β inhibitor (TGX-221) treatment

SOR	Sorafenib treatment
SALPHA	PI3K-α inhibitor (PIK-75) and Sorafenib combinatory treatment
SBETA	PI3K-β inhibitor (TGX-221) and Sorafenib combinatory treatment

CHAPTER I

INTRODUCTION

In this chapter, initially, the concepts on Hepatocellular Carcinoma (HCC), major HCC signaling pathway PI3K/Akt/mTOR, and kinase inhibitors which constitute the main biomedical subject area of this thesis are represented to better express the motivation. Then, a survey on the bioinformatics techniques and methods like next-generation techniques used for gene expression analysis, conventional RNA-seq analysis, network construction to analyze expression data, and systematic approaches for gene expression analysis are reviewed. The initial experimental analysis results, which constitute the main source of motivation that forms the basis of this doctoral study are explained. Finally, the general motivation and rationality of the study are detailed.

1.1. Hepatocellular Carcinoma

According to 2020 WHO-Global Cancer Observatory (GCO) reports, every one of five men and every one of the 6 women will be diagnosed with one of type of cancer in their lifetime; and unfortunately, one of eight men and one of eleven women will lose their life just before 75 years old because of the disease. Considering the numbers, it is not shocking that cancer is one of the most dangerous diseases in the world (Bray et al., 2018). The estimated number of cases for mortality and morbidity of cancer in the World are 9.556.027 and 18.078.957 in 2018 in all ages and sex types. According to the Global Cancer Observatory in Turkey, cancer incidences are above the world average. Turkey is facing a drastic increase in the absolute number of cancer cases, by 2040 the number is expected to be raised by 91.6% for mortality and 75.1% for morbidity in Turkey (World Health Organization, 2015).

Primary liver cancer originates from hepatocytes and cholangiocytes. It is the 6th most common cancer in the world by mortality and it is 3rth to colorectum cancer by mortality (World Health Organization, 2015). Hepatocellular carcinoma (HCC), also called hepatoma or HCC, is the most common type of primary liver cancer (Perzet et al, 2006). HCC, which constitutes 75% of primary liver cancers, is the 5th most common and the 3rd most lethal cancer in the world as shown in Figure 1 (Bray et al., 2018; Perz et al, 2006). While the death rates from other cancers are decreasing due to advances in diagnosis and therapeutics, the incidence and the mortality of HCC follow

an increasing trend due to the high rate of obesity-associated with liver diseases (Aleksandrova et al., 2016; B. Sun & Karin, 2012).

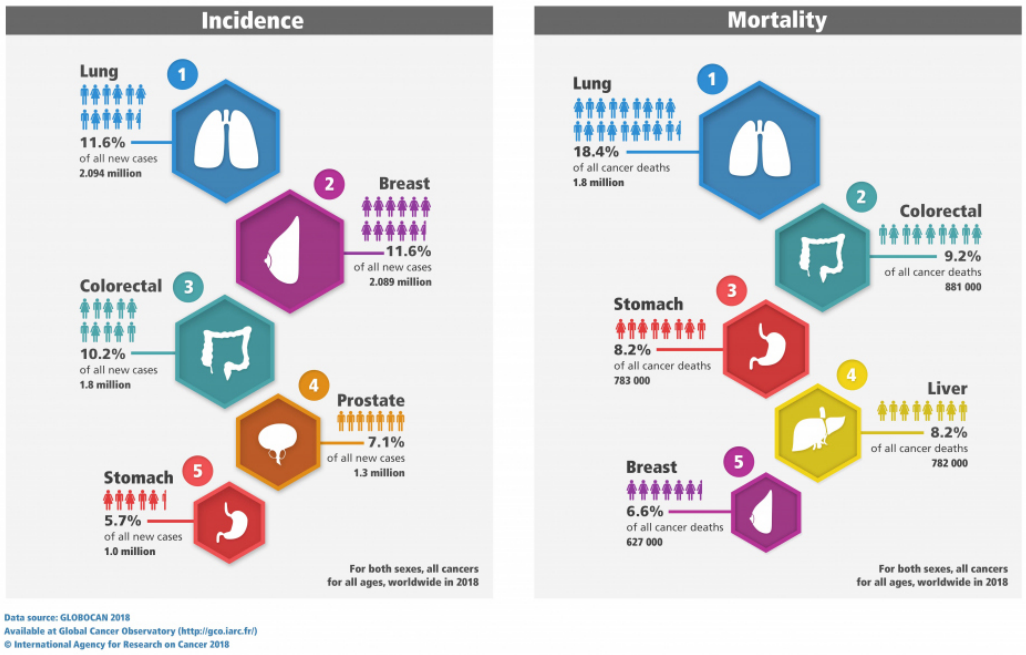


Figure 1:Percentages of new cancer cases and cancer deaths worldwide in 2018 (Bray et al., 2018).

1.1.1. Risk Factors

Hepatocellular carcinoma (HCC) develops essentially from hepatocytes and people with liver scarring (cirrhosis) are more likely to develop HCC. Liver scars result from chronic liver diseases. Furthermore, Hepatitis B (HBV) or C (HCV) virus infections are known to be directly linked to HCC (Jung et al., 2000; Ozcelik et al., 2003; Sun & Karin, 2012). In South Asia, high rates of HCC correlates to endemic HBV infections and in western countries the main risk of HCC is related to HCV infection (Bosch et al., 2004).

Among the environmental risk factors mentioned above, alcohol consumption is also linked to HCC. In a large cohort study considering multiple factors like family, drinking history and diabetes, they show that current or heavy drinkers were exposed to a high risk of HCC (Ogimoto et al., 2004). Moreover, dietary exposure to aflatoxin B1 significantly increases HCC risk, which mainly depends on geographical conditions. In Asia and Africa, high incidence of aflatoxin B1 increases susceptibility to HCC (Hamid et al., 2013). Cigarette smoking is causally associated with HCC, and

multiple effects of heavy drinking and smoking have been studied previously (Kuper et al., 2000).

Type 2 diabetes is also associated with the risk of HCC. Type 2 diabetes resulting from the constitutive stimulation of the insulin-like growth factor (IGF) pathway is considered as promoting HCC related pathways (El-Serag et al., 2006). In a large cohort HCC-control genome-wide association study, more than 20 loci are correlated as modulating type 2 diabetes (Salanti et al., 2009). High overall body mass rate (obesity) and diabetes, together, are believed to be a factor that may play role in %16 of HCC cases (Schlesinger et al., 2013). Apart from environmental or life-style related risk factors (listed in Figure 2), there are genetic bases of HCC. Several genetic abnormalities can give rise to liver injury, fibrosis and cirrhosis development. Human monogenic disorders of AAT deficiency, porphyria, hemochromatosis and tyrosinemia type I are the other syndromes associated with high risk of HCC (Dragani, 2010).

1.1.2. Initiation and Development

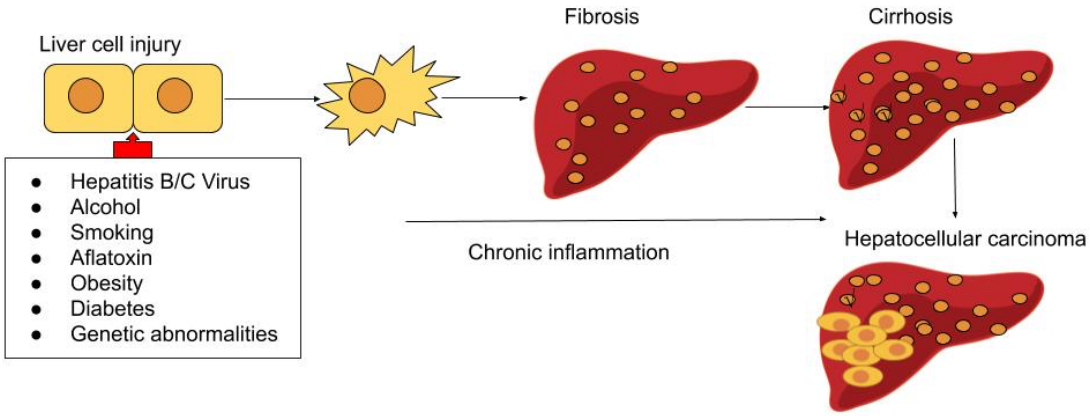


Figure 2: Risk factors and development of HCC. Risks factors including both environmental hepatitis B or C virus, excess alcohol consumption, cigarette smoking, aflatoxin B explosion, and genetic predispositions like obesity, diabetes and genomic abnormalities lead to liver injury, and through continuous chronic inflammations, fibrosis and cirrhosis induces hepatocellular carcinoma.

Development of HCC is described as a multicomplex biological process, where chronic liver disease is initiated due to a chronic inflammation or tissue damage. Injured hepatocytes promote hepatocyte death which is followed by liver regeneration through activation of chemokines and remodeling of cellular signaling events and finally uncontrolled cellular growth and proliferation. Increased genomic instability based on increased amount of reactive oxygen species (ROS) causes oxidative stress

which leads to accumulation of somatic mutations and finally leading to carcinogenesis. The increased proportion of proliferating cells leads to induction of several cell signaling pathways involved in liver regeneration, such as growth factor signaling, cell differentiation, angiogenesis and cell survival (Ersahin et al., 2015; Farazi & DePinho, 2006).

1.1.3. Acquired Capabilities Through Development of HCC

Transformation of a normal liver cell to hepatocellular carcinoma malignancy occurs through acquiring cancer features which in turn attenuate cellular growth and proliferation. During tumor invasion and progression, proliferation and cell survival signals are mainly activated through continuous activation of growth factors. Cellular differentiation is mainly due to multiple factors affected from dysregulation of signaling pathways. Supporting proliferative signaling, avoiding growth suppressors and cell death signals, enabling continuous replication, initiation of angiogenesis, invasion and metastasis, dysregulation energy mechanisms and resisting to the immune destruction are major hepatocellular carcinoma (HCC) hallmarks to gain malignancy (Ersahin et al., 2015; Hanahan & Weinberg, 2011). Acquired tumorigenic capabilities through the normal liver to HCC listed in Figure 3.

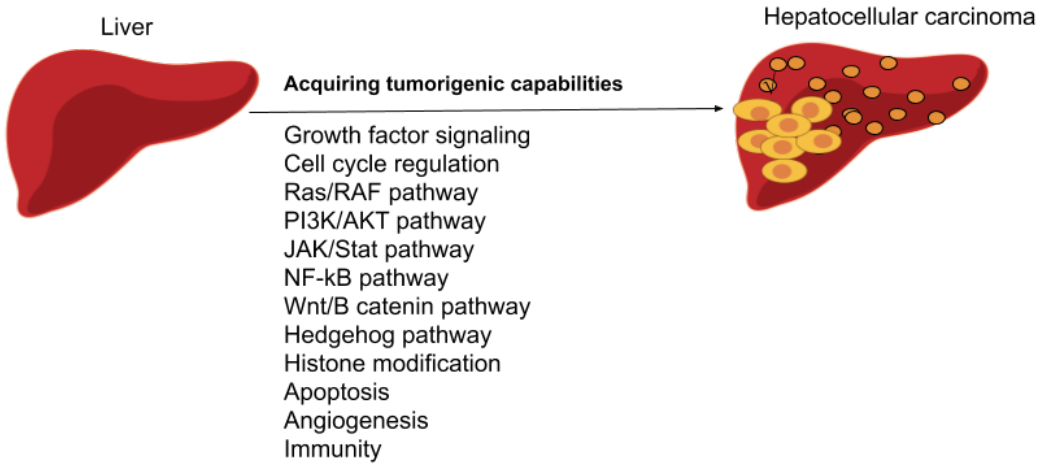


Figure 3: Acquiring tumorigenic capabilities from normal liver to Hepatocellular carcinoma development. Several signaling pathways are activated to provoke cell growth, proliferation, invasion, angiogenesis and metastasis, and control cell cycle, cell death and immunity destruction (Ersahin et al., 2015).

Downregulation of tumor suppressors like p53, Rb and p16 proteins, upregulation of c-myc and cyclin D1 and overexpression of E2F members stimulate cellular growth and proliferation that constitute activation of survival pathways. Endothelium growth factor (EGF), Insulin growth factor (IGF) and Hepatocyte growth factor (HGF-MET)

transmit the proliferation signal through PI3K/AKT/mTOR and RAS/RAF/MEK/ERK pathways (Moeini et al., 2012).

In HCC, growth arrest and DNA damage 45G (GADD45G) is frequently inactivated, and *GADD45G* gene in JAK/STATA3 pathway provokes senescence (Li Zhang et al., 2014b). Gained resistance to transforming growth factor beta (TGF-β) inhibition is commonly found in early stages of hepatocyte tumorigenesis. In advanced stages, TGF-β is downregulated to ensure resistance to cell death (Thomson et al., 2011). Moreover, increased levels of growth factors and promotion of anti-apoptotic pathways like nuclear factor kappa-light-chain-enhancer of activated B cells (NF-kB) signaling, diminish of death receptors like DR5 and Fas, and mutations in the tumor suppressor genes like p53 provide escape from apoptosis in HCC cells (Okamoto et al., 2007).

Telomere dysfunctions, oncogene activations, DNA damages and ROS maintained oxidative stress leads to permanent cell cycle arrest. In order to evade this stage, HCC gains replicative immortality mainly through overexpression of telomerases to maintain telomere length stable (Oh et al., 2003). Recently it was found that there is a correlation between telomere length and the aggressive behavior of HCC (Ozturk et al., 2009). Vascularization is another feature of HCC. Angiogenesis and vascularization evolve by interactions of tumor cells and vascular endothelial cells. The balance between them is unsettled due to increased amount of pro-angiogenic factors like vascular endothelial growth factor (VEGF), angiopoietins (Ang), fibroblast growth factor (FGF) and platelet-derived growth factor (PDGF) (Papetti & Herman, 2002). Neuropilin receptors (NRP) is a membrane bound factor in vascularization of HCC. Pro-angiogenic signals are majorly released in response to the hypoxic condition and nutrient depletion during tumorigenesis. *VEGF* and *FGF* upregulation are found in malignant types of HCC (Yoshiji et al., 2002).

For invasion and metastasis of HCC, cell detachment is onset. Promotion of twist, snail, slug, vimentin and zeb1 and zeb2 and decrease in E-cadherin and HNF-4Kα appear on epithelial-mesenchymal transition (EMT) enabling poor prognosis (Thomson et al., 2011). Glucose is the major suppliant of the tumor cells in order to provide energy for continuous growth and proliferation. Tumor cells reprogram the cellular energy mechanism to achieve this feature. The mechanistic target of rapamycin (mTOR), found to be downstream of PI3K/AKT pathway, monitors cellular energy level. When there is sufficient energy need, mTOR promotes biosynthesis through suppressing autophagy which supplies recycled metabolites to the cell, while PI3K/AKT pathway attenuates glucose uptake (Qiao et al., 2016b).

Based on the tumor microenvironment, in HCC development, growth factors, cytokines, chemokines and ROHs productions are increased leading to tumor initiation and contributing progression. Pro-inflammatory cytokines like Interleukin 1 beta (IL-β), Tumor necrosis factor alpha (TNF-α) and Platelet-derived growth factor (PDGF) together with Kuppfer cells and Hematopoietic stem cells (HSCs) stimulates cellular growth, inflammation, metastasis and invasion mainly through NF-kB and AKT

signaling pathways (Manning & Cantley, 2007). HCC cells evade immune protection mainly through immunosuppressors like Programmed death-ligand 1 (PD-L1), Indoleamine-pyrrole 2,3-dioxygenase (IDO), cytokines (IL-6, IL-10, VEGF and TGF-β). Macrophages are important factors in immune destruction system since they infiltrate leukocytes. The status of tumor-associated-microphases and other immune cells are important in tumor microenvironment for required response (Cavallo et al., 2011).

1.1.4. Genetic Heterogeneity

Genome-wide molecular profiling of hepatocellular carcinoma (HCC) tissues using whole genome or whole exome sequencing techniques contributed to a growing understanding of the genetic background of cancer progression (Dragani, 2010; Schulze et al., 2016). Several molecular abbreviations across the tumors are identified which form subtypes of HCC tumors. HCC tumors are classified into two groups according to their aggressiveness in biological features. The first group is called as aggressive type of HCC with increased genetic instability, proliferation, activation of survival pathways, damage in tumor suppressor with larger tumor size, poor prognosis, and high rate of recurrence. The aggressive tumors are also sub-grouped according to TGF-beta pathway activation, cholangioma-like gene signature, vascular invasion and stemness markers. Besides these HCC major types, different DNA mutations and environmental exposures like smoking, drinking or toxics contribute as other levels of heterogeneity (Goossens et al., 2015).

Several aberrant genes are acknowledged in HCC pathogenesis including *P53* (Tumor protein P53), *PTEN*, Breast cancer type 2 (*BRCA2*), *SMAD2* genes, c-myc and cyclin D1 proteins (Bae et al., 2007; Brito et al., 2012). Besides to these genetic abbreviations, there are several overexpressed signaling pathways observed in HCC. Stimulation of these pathways is associated with mostly tyrosine kinases which are commonly a part of phosphatidylinositol-3-kinase (PI3K)/AKT/mTOR pathway (Ersahin, Ozturk, & Cetin-Atalay, 2015). Another significant mechanism underlie in HCC development is MAPK pathway mostly activated though Ras protein and vascular endothelial growth factor (VEGF) pathway. Studies show that heterogenic nature of HCC is mostly caused by the variations of mutations and alterations in expression levels of these key proteins involved in these mentioned signaling pathways (Moeini et al., 2012).

1.2. PI3K/AKT/mTOR Signaling Pathway

PI3K/AKT/mTOR is the major signaling pathway in cell cycle. It directly regulates the expression of proteins involved in protein synthesis, cellular proliferation, survival, metabolism, and differentiation which all these aspects may contribute to cancer cells ability to survive and progress. The constitutive activation of PI3K/AKT/mTOR signaling pathway is frequently observed in liver cancers. Phosphatidylinositol 3-kinase (PI3K), a serine/threonine protein kinase (Akt) and mammalian target of

rapamycin (mTOR) are the major proteins in the pathway. The stimulation of these proteins are characteristically through tyrosine kinases, hormones and mitogenic factors (Ruggero & Sonenberg, 2005). The PI3K/AKT/mTOR pathway is illustrated in Figure 4.

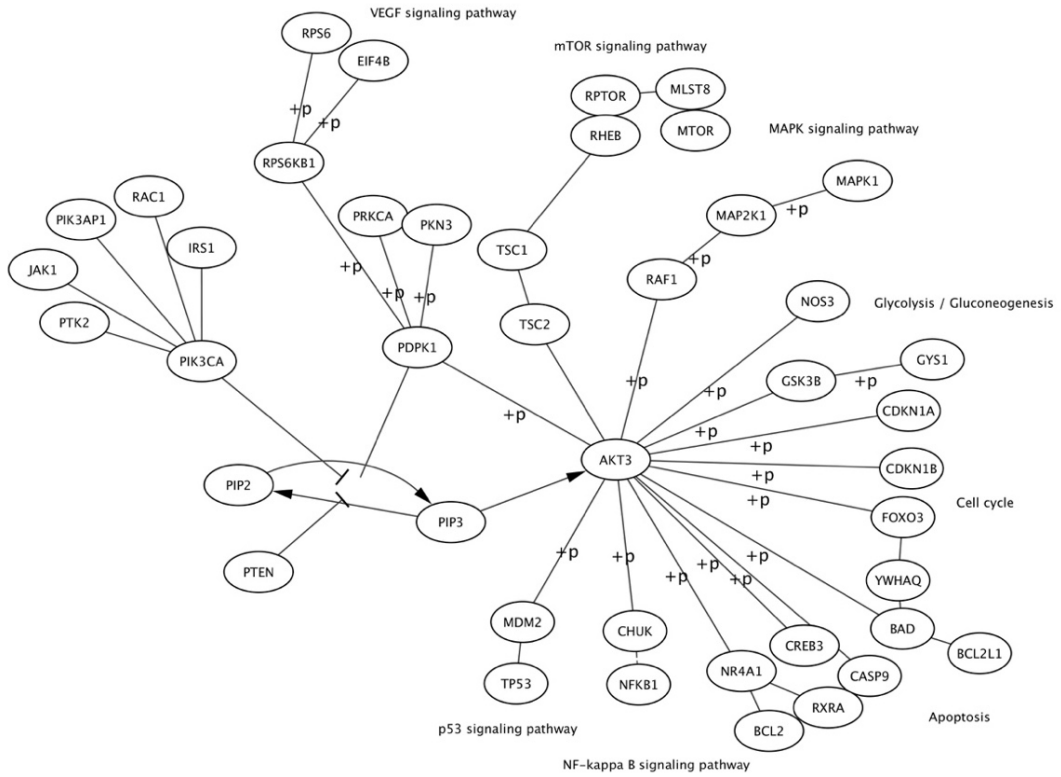


Figure 4: Overview of PI3K/Akt/mTOR signaling pathway. The activation of PI3K (PIK3CA protein) stimulates conversion of PIP2 to PIP3. PTEN negatively regulates the pathway by dephosphorylation of PIP3. Phosphorylation and activation of Akt (AKT3 protein) effects. The network is extracted from the KEGG PI3K/AKT signaling pathway.

Incessant Akt phosphorylation is the key factor for abnormal activation of PI3K/AKT/mTOR signaling pathway which is frequently due to inactivating mutations or loss of heterozygosity in a tumor suppressor protein, Phosphatase and tensin homolog (PTEN) antagonizes for Akt activation by dephosphorylating of PIP3, or mutations activating *PIK3CA* gene, or damage in the negative-feedback loop from mTOR signaling pathway (Bae et al., 2007; Buontempo et al., 2011; Engelman, 2009; Kawamura et al., 1999). When Receptor Tyrosine Kinases (RTKs) or G protein-coupled receptors (GPCRs) are induced by growth factors, phosphatidylinositol 4,5-biphosphate (PIP2) to phosphatidylinositol 3,4,5-triphosphate (PIP3) conversion occurs at the cell membrane by class IA and class IB PI3Ks. This conversion presents docking sites for phosphoinositide-dependent kinase 1 (PDK1) and phosphoinositide-dependent kinase 2 (PDK2) which one by one will activate Akt (Fujimoto et al., 2012).

mTOR kinase can connect to both PI3K/AKT and Raf/MEK/ERK signaling pathways through RAS. mTOR is found in two forms: as mTORC1 or mTORC2, but PI3K and MAPK signaling pathways control mTORC1 through phosphorylation of TSC2. Which in turn targets p70S6K leading to decrease of insulin signaling for PI3K, forming a negative feedback loop (Villanueva et al., 2008). Akt protein has several downstream effects like activating CREB and mTOR, inhibiting p27, localizing forkhead box O (FOXO) to cytoplasm. mTOR is a key protein in PI3K/Akt/mTOR signaling since it acts on both upstream and downstream of Akt. mTOR regulates the protein synthesis of the key molecules necessary for cell growth, proliferation, and angiogenesis. Akt induces cell survival through positive regulation of IkB kinase which is the master regulator of NF-kB cells (Porta et al., 2014). Akt promotes cell survival by avoiding proapoptotic signals from BAD and BAX. Akt expression also phosphorylates MDM2 which upsets p53-mediated apoptosis and GSK-α having role in gluconeogenesis controlling cellular energy. (A Villanueva & Llovet, 2013)

1.2.1. PI3K Kinase Classes

Phosphatidylinositol 3-kinases (PI3Ks) are from a family of lipid kinases. They phosphorylate PtdIns lipids to PIP3 second messengers on the cell membrane. They are the main catalyzers of the PI3K/AKT/mTOR signaling pathway. In the family there are three classes which are differentiated through their coding genes, structures, and substrate preferences structures (Vanhaesebroeck et al., 2010). Among them, class I PI3Ks are the most studied type because of their fundamental functions. It is a unique ability of class I PI3Ks to catalyze the phosphorylation of PIP2 to PIP3. Class Ia Phosphatidylinositol 3-kinase (PI3K) are heterodimeric lipid kinases. The studies found that increased levels of PIP3 is also related to carcinogenesis and hence rather than class II and class III PI3Ks, specifically class I PI3Ks are related to cancer development as only class I generates a substrate for PTEN (Fruman & Rommel, 2014; Jia et al., 2008). PTEN is frequently lost in many cancer types leading constitutive activation of PI3K (Suzuki et al., 1998).

Two of Class I members of PI3K have heterodimeric subunits of p110 and p85 regulatory subunits respectively, Class IA p110-α and Class IB p110-β are well studied enzymes in cancer. *PIK3CA* (phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha) gene encoded PI3K isoform p110α is activated through receptor tyrosine kinases (RTKs) and *RAS* oncogene, mutations in these proteins mostly regulates growth, metabolism, and angiogenesis (Zhao & Vogt, 2010).

The other PI3K isoform p110β (encoded from *PIK3CB* gene) is regulated exclusively by G protein-coupled receptors (GPCRs) and has critical functions in inflammatory cells (He et al., 2015). They show that PTEN-negative cancers, p110-β is very critical. PTEN-lost cells depend on p110-β activity for proliferation. In an animal model of prostate tumor epithelium, PTEN-negative cells showed a diminished tumorigenesis through decreased Akt phosphorylation with inhibition of p110-β but not by p110-α (Berenjeno et al., 2012).

On the other hand, in another study, they show that the main PI3K activity is on to p110-β inhibitor when PTEN is mutated. p110-β inhibitory treatment results with an increased activity of PI3K signaling in PTEN-deficient models of prostate cancer, probably through a feedback inhibition through IGF1R causing activation of androgen receptor kinases and so of isoform p110-α. Thus, they suggested a combined inhibition of p110-α and p110-β isoforms for an efficient tumor regression (Schwartz et al., 2016). In another study, they also proved unequal steatosis actions of PI3K α and β isoforms. They observed significantly more liver lipid accumulation in knockout of p110-α than p110-β in high-fat-diet fed (HFD-fed) mice. They conclude that, PI3K p110-α incite steatosis possibly through one of three ways; atypical protein kinase C activation, inducing lipogenesis, promoting fatty acid uptake from the blood (Chattopadhyay et al., 2011). Therefore, those multiple studies shows that the actions of the p110-α and p110-β isoforms are variable by the cell content and cellular mechanism.

1.3. Current therapies for HCC

The options for Hepatocellular cancer (HCC) treatments are curative resection of the tumor cells, liver transplantation from a health donor, radiofrequency ablation (RFA), trans-arterial chemoembolization (TACE) and systemic targeted agents like sorafenib or Regorafenib (Raza & Sood, 2014).

Treatment alternatives for HCC heavily depend on the stage of the tumor, the reserve of the optimal liver function and in general patient performance. Depending on the patient's compatibility, liver transplantation or surgical resection of the tumor are the only options if the patient is in early stages of the cancer. However, even if the surgical operations are successful in first side, regeneration of the carcinoma cells is common, and transplantation is restricted by liver obtainability. Moreover, when the patient is not compatible to surgical resection due to late diagnosis or advanced stages of medical treatment, chemotherapy and radiotherapy are the other systematic treatment options, yet they are generally ineffective since the majority of HCC cases are noticed in advanced stages (Llovet et al., 2012; Omata et al., 2010).

Advanced molecular studies in HCC have found interpenetrating actions in various signaling pathways and some novel proteins representing key targets for new molecular therapeutic options. Yet, for now, Sorafenib and recently approved drug Regorafenib are the only targeted agents to cure HCC.

1.3.1. Sorafenib Inhibitor

For the patients whose major treatment options are not promising, Sorafenib (Nexavar, BAY43-9006) is accepted as the standard systemic treatment for staged HCC (Llovet et al., 2012). Unfortunately, Sorafenib treatment improves the patient survival rate only by 2.5 months (Cheng et al., 2009). The main reason behind the ineffectiveness of Sorafenib is the fact that RAS oncogene, which is the main target of this drug, is

frequently mutated establishing an ineffective docking site to the inhibitor (Llovet, Ricci, et al., 2008).

While surveying on the effect of new therapeutic agents like sunitinib, linifanib, erlotinib, linifanib, evorolimus and brivanib to HCC continue in phase II and phase III, they do not indicate any survival benefit. The reasons for phase II and phase III failure in HCC have been examined. The main reasons for unsuccessfulness of these drugs is the heterogenic nature of HCC and cross talks between the major signaling pathways having role on growth, angiogenesis, progression and apoptosis and lack of information on the major drivers for tumor progression (Llovet & Hernandez-Gea, 2014). Moreover, as the signaling pathways involved in major life processes are redundant, they compensate each other through some key molecular regulations, when drivers of a hyperactive pathway are inhibited. Which makes the signaling pathways with superfluous functions due to the potential cross-talks between them, which could be a reason for the ineffectiveness of these multi-kinase inhibitors (Moeini et al., 2012).

For patients who developed tolerance to Sorafenib or failed to get benefit of it, Regorafenib (Bayer, BAY73-4506) is replaced by Sorafenib. Regarding the fact that, like Sorafenib, Regorafenib is also a multi-kinase inhibitor (Personeni et al. , 2018) and proposing the same treatment strategy. According to the clinical trials, the median survival rate of Regorafenib is only 10.5 months. Thus, it also failed to benefit for overall survival (Bruix et al., 2017).

Hopefully, other than multi-kinase inhibitors, there are also several inhibitors targeting PI3K/AKT/mTOR pathway such as, PI3K inhibitors, dual PI3K/mTOR pathway inhibitors, mTOR inhibitors and AKT inhibitors. Now, these inhibitors are in clinical development with potential therapeutically effect (Rodon et al., 2013). The chemical structures of the drug agents; Sorafenib, PIK-75 and TGX-221 studied in this thesis are represented in Figure 5.

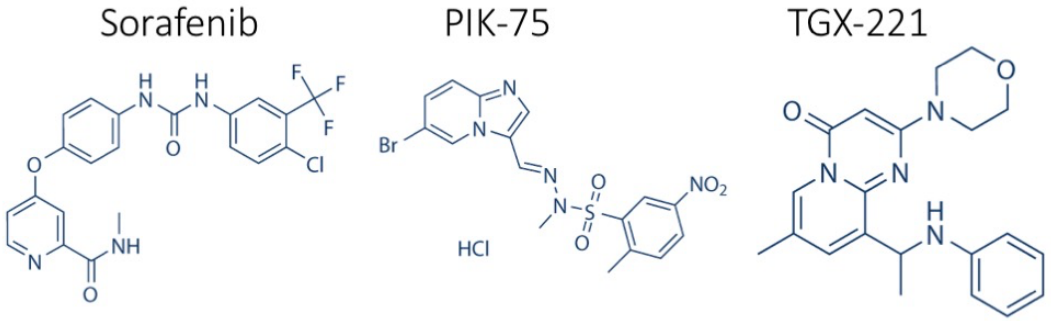


Figure 5: Chemical structures of multi-kinase inhibitor Sorafenib, and of p110α and p110β inhibitors PIK-75 and TGX-221. Sources: Sorafenib PubChem ID is 216239, PIK-75 PubChem ID is 10275789 and TGX-221 PubChem ID is 9907093.

1.3.2. PIK-75 and TGX-221 Inhibitors

One of main components of PI3K/AKT/mTOR pathway is PI3Ks. The PI3Ks family compose four class. Among the members, class I PI3Ks are able to activate AKT. They are heterodimeric lipid kinases composed of p110 catalytic subunits and p85 regulatory subunits (Fruman & Rommel, 2014). PIK-75 is a p110α inhibitor, reducing cell survival by decreasing mitochondrial activity shown previously in ASM cells and lung fibroblasts. TGX-221 is a p110β inhibitor, partially inhibiting platelet-ECC interaction, aggregation, and granule binding in the ECC model.

Previously, the effects of p110α and p110β inhibitors to Huh7 and Mahlavu cell lines were studied globally at CANSYL laboratory, METU. According to the experimental results, PIK-75 was very effective to inhibit cellular growth and reduce migration in both cell lines while TGX-221 was not effective at all. Furthermore, in this study, single Sorafenib and its combinational effects with p110α and p110β inhibitors were also considered. It was found that while combinational therapy of PIK-75 and Sorafenib significantly reduces cellular proliferation in Huh7 and Mahlavu, addition of TGX-221 to Sorafenib treatment attenuates cellular growth antagonizing Sorafenib action. Yet, the molecular action underlying this has not been revealed in that study (Ersahin, 2014).

1.3.3. Combinational Therapy

As discussed before, it is now clear that one drug targeting to one receptor is not an effective way for considering the redundancy in the signaling pathways (Maggiora, 2011). Heterogenic nature of HCC allows for cross-talks between the main signaling pathways and generally inhibition of just one super-active component not enough to retain normal functions of the cell (Llovet & Hernandez-Gea, 2014). Instead of using

single drug targets, as shown in the CANSYL laboratory, combinational targeted therapies can increase the efficiency of the existing drugs and a synergistic effect can be gained through combination of the multi-kinase inhibitor Sorafenib with other anti-angiogenic or anti-proliferative inhibitory molecules (Ersahin, 2014).

Recently, new therapeutic methods have been started to use combinations of inhibitory agents targeting PI3K/AKT/mTOR pathway like Sorafenib with other anti-angiogenic and anti-proliferative agents. For now, combinations of Sorafenib to Brivanib or Erlotinib did not exceed the survival level provided by single Sorafenib treatment (W. Sun & Cabrera, 2018). But in a more recent study, they show that the combinational therapy of bevacizumab targeting VEGF and atezolizumab targeting directly to an immune checkpoint protein managed to show superior survival rate than classical treatment of Sorafenib (Ray, 2020).

In HCC, gain of function mutations, oncogene activations or overexpression of growth factors activates Ras/MEK/ERK and PI3K/AKT/mTOR pathways. Components of these pathways are having an active role in gaining resistance to the targeted therapies. That is why these two pathways are the major targets for co-inhibition-based combinational therapies. Therefore, in the light of previously discussed studies, more combinations might be worth trying.

1.3.4. Drug Repurposing

Using the old known drugs for therapy of new diseases for unexplored medical uses called as drug repurposing (or drug repositioning and drug reprofiling) is a former method but it is gaining a great attraction in recent years again. Drug development process is complex, costly, lengthy, and often not successful because of various reasons (Talevi & Bellera, 2020). Frequently, animal models might not be coherent to the target disease and drug mechanism. Therefore, new clinical applications of the existing drugs instead of novel drug search approaches would be more practical. A great amount time is also be preserved since the long clinical trials would be skipped since it is already approved for use in humans. The other advantage of using a known drug would be the fact that there could be so much studies on that the molecular mechanism of it should be already known (Shim & Liu, 2014).

1.4. A Survey on the Methods Used for Conventional Drug Target Studies

Considering the failure of present drugs to eliminate cancer cells in Hepatocellular carcinoma (HCC), identification of novel drug-target connections could be an pivotal step for HCC therapy (Llovet et al., 2012). Currently, the discovery of drug targets against cancer mostly focuses on molecular agents with aberrant functions in regulatory signaling pathways. Experimental investigation of a single gene or locus is costly and time consuming while the only benefit of them is exploration of a new oncogene. Frequently, study of a single gene fails to satisfy the need of solving the complex interactions of its pathogenesis. Actually, the network of cancer invasion,

progress and growth is multifaceted due to interlaced regulatory signaling pathways and the connections of key or hub oncogenic proteins (Medina-Franco et al., 2013). New emerging computational methods can facilitate the work. Hence, rather than traditional experimental methods, new computational techniques should be used to find novel drug target relationships.

High throughput data like genomic, epigenetic, transcriptomic, or proteomic all together made it possible to study complex molecular mechanisms for drug discovery. Following the rise of high throughput genomic technologies, several studies using array or sequencing based methods enabled more comprehensive molecular profiling of HCC. Bioinformatic analysis of the high throughput data would not only accelerate immediate drug target identification, candidate searching or eliminating false positives but also it can provide a broad range search to understand the mechanism of the effects and even to understand side effects of the drugs or cancer recurrence mechanism (Paolini et al., 2006). Furthermore, the only benefit of high-capacity technologies is not only scanning a broad range of samples simultaneously, but the major advantage of them is to group and compare these gene profiles to find disease-causing elements. Recently, using RNA sequencing techniques to analyze the whole transcriptome of HCC cells has become very popular. As usage of RNA-seq became prevalent, downstream analysis methods started to range, and the demand for variety methods is needed.

1.4.1. RNA Sequencing

Currently, instead of concentrating on just a single molecule, it is common to use next generation sequencing techniques to profile the whole transcriptome of the cell comparing the normal versus the conditional (treatment) status. Concordantly, a set of genes affected though the treatments can be discovered for further study. RNA sequencing is one of the most used high capacity methods for rapid and reliable gene expression profiling (Chu & Corey, 2012).

RNA-seq application is very easy and straightforward. The steps of the RNA-seq analysis workflow are described in Figure 6: First, an appropriate experimental design including negative controls are conducted. Since the structure of the transcriptome studies, gene expression level is not countable itself, and so it requires comparative analysis. It is necessary to match every treatment sample to its control. The results of the RNA sequencing experiment are transferred into computational language by the FASTQ files. Besides the reads, FASTQ file also contains quality degrees of each base. After quality checks and refinements are performed, the next step is alignment of the reads to the reference genome and quantification of the reads aligned to the genes. One classical way of alignment and gene counting is done by TopHat (Trapnell et al., 2009) with Bowtie. The other methods are GSNAP (Wu and Nacu, 2010), MapSplice (Wang et al. 2010a), RUM (Grant et al. 2011), and STAR (Dobin et al. 2013). For quantification of the genes or transcripts, assembled reads to the reference

genome are counted. Cufflinks (Trapnell et al. 2010) and HTSeq-count (Anders et al., 2015) algorithms are the most commonly used quantification tools.

Generally, the next step for RNA-seq analysis is performing differential expression analysis. RNA-seq is frequently used to compare gene expression levels between two conditions like drug treated sample vs mock or untreated control and identify up and downregulated genes for each condition (Kukurba & Montgomery, 2015). Two of the most commonly used tools for this analysis are both from Bioconductor R using a model based on negative binomial distribution; DESeq (Robles et al., 2012) and edgeR (Robinson et al., 2009) packages. Generally, they are preferred to each other for their unique functions. For example, edgeR packages preserve some functions for non-replicated RNA-seq analysis and would be beneficial for the analysis in a condition if RNA-seq is high-quality.

In order to understand biological importance and functional relevancies between differential expressed genes, a variety of tertiary analysis can be performed. One way to associate these genes to their biological processes, functions or cellular components is to implement a gene set enrichment analysis (GSEA). There are a lot of methods for this analysis; the most common ones are Enrichr (Kuleshov et al., 2016), DAVID (Dennis Jr. et al., 2003) and FunRich (Pathan et al., 2015). As a result of GSEA analysis significant gene enrichments can be viewed and linked to altered cellular functions. Another approach to link differentially expressed genes could be using pathway analysis. The genes acting together on a cellular pathway can be observed though mapping them with expression levels. By this way, which signaling pathways are mostly affected by pathogenesis of a gene can be detected. In order to do that, KEGG (Kanehisa & Goto, 2000) and BioGRID (Stark et al., 2006) tools are the most popular online databases. Their wide range of sources includes whole pathways that provide a compressive way of gene signaling.

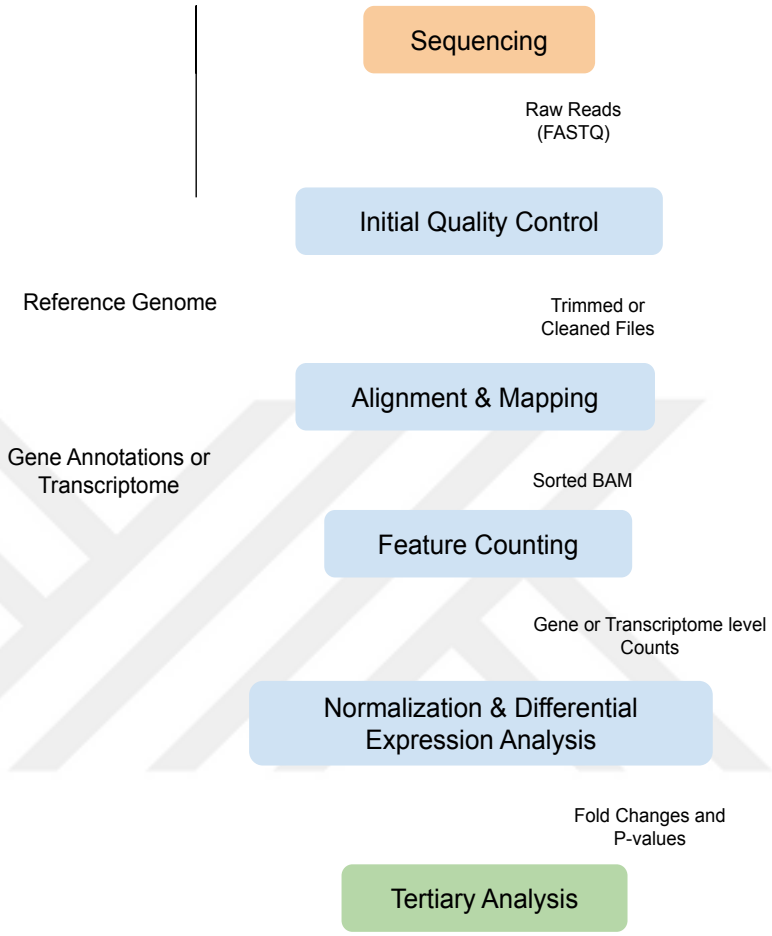


Figure 6: RNA-seq analysis workflow. The first step of RNA-seq analysis is to implement quality control. Then basing a reference genome alignment and mapping of the reads generates a BAM file. A sorted BAM file can be used to count genomics features using gene annotations or a transcriptome. Afterwards, comparing to a negative control, normalization of the counts and differential expression analysis is performed. Finally, resulting up or down regulated genes can be used for several tertiary analysis.

1.4.2. Protein-Protein Interaction Databases

The complex cellular metabolism regulation consist of a multifaceted network or pathway of proteins working together (Lander et al., 2001). Proteins form different types of interactions to make cellular processes. Therefore, to understand biological function of a cellular pathway, as well as acting proteins, their interactions also should be well characterized.

In order to detect protein-protein interactions (PPIs), experimental methods like yeast two hybrid (Y2H) (Ito et al., 2001), affinity purification (Rigaut et al., 1999), x-ray crystallography (Tong et al., 2001), NMR spectroscopy (Tong et al., 2001) or fluorescence resonance energy transfer (FRED) (Yan & Marriott, 2003) need to be

used. However, these assays are not high-throughput, time-dependent, and costly. Design of a single experiment can only cover a small number of interactions. Consequently, for a systemic analysis of transcriptome, these previous efforts should be well conserved and linked.

There is a wide source of publicly available PPI databases covering these experimental data to be used in network modeling. The Database of Interacting Proteins (DIP) (Xenarios et al., 2002), Molecular Interaction Database (MINT) (Chatr-aryamontri et al., 2007), The Biomolecular Interaction Network Database (BIND) (Bader et al., 2003), The Human Metabolome Database (HMDB) (Wishart et al., 2012), STITCH (Kuhn et al., 2013), Reactome (Fabregat et al., 2018), Kyoto Enclopedia of Genes and Genomes (KEGG) (Kanehisa & Goto, 2000), iRefWeb (Turner et al., 2010), STRING (Szklarczyk et al., 2015) are some of them.

DIP and MINT are two databases to catalogue experimentally determined PPIs. Besides the experimentally validated PPIs, different from these two, BIND curates interactions using PubMed records. The process is performed with scientists. Thus, these databases can be very useful to get the most reliable interaction data. HMDB is a collection of human metabolites and provides interactions of small metabolites like peptides, lipids, amino acids, nucleic acids, organics acids, vitamins, minerals, drugs, and toxins. STITCH is a database for protein to chemical interactions underlying many cellular signaling pathways. The database consists of experimental and manually curated verifications. Reactome and KEGG are two services providing a search base for a systematic analysis of gene-to-gene functions and the cellular pathways the genes have role on. For each gene, genomic information, annotations, and pathways are provided through the datasets. iRefWeb provides a method to generate “keys” to give protein interactions using publicly available sources. This method provides an alternative understanding and many groupings to the original database sources. iRefWeb includes many levels of evidence ready to be query. STRING provides protein-protein interaction (PPI) interaction proofs, and the interpretations of them are used to level the confidence of the protein interactions. Conserved neighborhood, co-occurrence, fusion information, co-expression, experiment, other databases, and text mining methods construct the baseline of the proofs. The evidence was collected for each protein interaction and used to calculate a confidence score. There are four confidence levels provided by the tool: low confidence (0.15 and above), medium confidence (0.4 and above), high confidence (0.7 and above) and the highest confidence (0.9 and above). Using high confidences will restrict the number of interactions to a more confident level since the provided experimental proofs become more accurate.

1.4.3. Protein-Protein Interaction Prediction

Since experimental methods for PPI are very expensive and time-consuming, there is a huge demand for computational methods for accurate prediction of PPIs. Combination of structural and evolutionary information on proteins and usage of

bioinformatic methods to predict their interactions makeup the basis of PPI prediction. Phylogenetic profiling, prediction of protein coevolution, conserved gene neighborhood, gene fusion and identification of structural protein binding patterns are the most used methods for PPI predictions.

Prediction Server for Protein-Protein Interactions (PSOPIA) is a tool for predicting PPI using sequence similarities, statistical propensities, and homologous protein distances. Protein Link Explorer (PLEX) (Date & Marcotte, 2005) uses phylogenetic profile to identify functionally linked proteins. Database of Ligand-Receptor Partners (DLRP) (Graeber & Eisenberg, 2001) is an effort of DIP in order to get protein ligand and receptor connections. Moreover, for computational modeling and analysis of protein structures as well as interaction predictions PrePPI algorithms are commonly used (Q. C. Zhang et al., 2012). PRISM is also a very useful prediction tool using protein structure and sequence conservation in protein binding sites together (Aytuna et al., 2005).

1.4.4. Network/Pathway Analysis

Especially following the emergence of next-generation sequencing technologies, generation of high-throughput measurements of molecular changes in the cell is wide. As these quantitative analyses provide comprehensive analysis, the functions, and processes active in the cells can be visualized globally. While transcriptomics has arisen as a prevailing approach to observe global changes, it has its own limitations. For example, a time dependent RNA-seq analysis can reveal expression patterns of genes yet cannot specify the exact pathways driving the gene. In order to link transcription factors to those gene patters, ChIP-seq analysis should be provided. ChIP-seq is a very well-known method to understand the mechanism between the transcription factors (and other regulatory factors) and gene expression.

Considering the complexity of the cellular mechanism, no single analysis set can cover the levels of functions the cell has role on. Hence, multiple levels of methods should be applied to solve sophisticated mechanisms of cellular systems. Besides, these different next-generation applications generate many levels of information, each dataset will bring its own analysis necessities. Therefore, joint analysis of these data requires new computational approaches of investigation too.

Recently, network modeling approaches are proposed to solve collective problems of biological data. Network modeling requires a set of interactome, which is available through public databases like STRING or iRefWeb. Previous experimental analysis generally stored in biological databases generates a huge source. The interactome data construct the backbone of network analysis. Unfortunately, since the biological databases used to construct interactome data generally composed of user generated data, and so the network includes millions of connections (Fabregat et al., 2018; Kanehisa & Goto, 2000; Kuhn et al., 2013; Razick et al., 2008; Wishart et al., 2012). It is a tremendous amount of data to be searched in. Moreover, generally as some of

the genes become hot topic of study, it attracts more people to work on them which eventually generates too many connections for this gene in the network. Dense networks ultimately become "hairballs" and too much studied nodes bring bias to the network. These two challenges form a basis for generated network analysis models.

1.4.5. Visualization of the Biological Networks

Biological networks are the basic representations of the molecules like proteins or genes. Visualizations of the biological networks or pathways can be achieved through various drawing tools like OmicsNet (Zhou & Xia, 2018), NAVIGaTOR (Brown et al., 2009), Osprey (Breitkreutz et al., 2003), Cytoscape (Shannon et al., 2003) and Arena3D (Pavlopoulos et al., 2008). These tools perform both mapping of the genes or proteins to the nodes, visualize the connections and provide statistical or topological analysis. The most obvious advantage using such a tool is to shape the interactome data as optimal to the needs of the analysis. For example, the researcher may change the color and size of the nodes according to the significance of the genes or group the nodes according to a specific function in a very fast and organized way.

From the visualization tools, Cystoscope is one the most popular ones for biological network analysis. It is an open-source tool and a platform for a set of other applications (also called as plugins) developed by world-wide users. Its collection includes applications providing scalable analysis for annotation, clustering, enrichment, and topological analysis.

1.4.6. The Systems Biology Approach

Collecting many levels of cellular data like transcriptome, proteome and metabolome from the same organism makes possible the use of network-oriented research. The aim of the systems biology approach is to understand these multi-layer biological networks through design and application of experiments and data analysis. Mathematical models are used to characterize biological systems and to predict cellular responses to the aberrant functions (Kitano, 2002). Quantitative whole-cell measurements form the major network components. The most attractive way of systems approach is to use these measurements to conduct networks and then use these networks for analysis of other cellular measurements or experiments. This approach is called top-down modeling, an application of reverse engineering (Kholodenko et al., 2002).

Considering the complex molecular networks responsible for cancer maturity and progressions, the research focused on the use of systems biology approaches to understand molecular networks altered by malignant transformation. Cancer systems biology studies incorporate cellular signaling pathways critical for cancer initiation, development, malignancy, and metastasis. Most importantly, targeted therapy opportunities can be widened using this approach and novel drugs can be found for clinical research (Werner et al., 2014).

Atlas of Cancer Signaling Network (ACSN) is a very powerful resource for usage of cancer systems biology analysis. Using Google Maps engine, it utilizes tools for analysis and visualizations of cancer signaling networks. The cell signaling mechanisms frequently disrupted in tumorigenesis, called as Hallmarks of Cancer, are formed as "geographic-like map" enabling zoom-in functions (Kuperstein et al., 2015). Besides, KEGG, and Reactome databases can be also useful for cancer signaling analysis pathways.

1.4.7. Omics Integrator

In order to perform an integrated analysis of proteomic, gene expression and epigenetic data and connect them through a protein-protein interaction (PPI) network, Omics Integrator software (Tuncbag et al., 2013) can be used. This package uses Prize Collecting Steiner Tree (PCST) solution to integrate high-throughput data to PPI network. There are two modules of Omics Integrator called Garnet and Forest performing a joint analysis of RNA-seq and ChIP-seq data.

Omics Integrator uses user defined high-throughput data like gene expressions or transcription factor affinities as prices. The interactome, like PPI network, is used to calculate edge costs as the prize nodes being traveled. As the edge cost gets higher it is less likely to be visited by the algorithm. By this way, all the prize nodes, the input data, will be collected. Ultimately, the least costly solution will be selected as the optimal network. The same group who developed the software was used this approach to model patient specific pathways in Glioblastome. They used tumor specific phosphoproteomic data and human interactome to construct disease modulated networks. They also developed a unique strategy to select targets for clinical research (Tuncbag, Milani, et al., 2016).

In 2017, Steiner prize collecting approach was used to generate a "humanized" network from known gene interactions found in yeast on the toxicity of synuclein (alpha-syn) protein. This way, the genes linked to Parkinson's disease mapped to the human network and thus pathogenetic genes in patient driven transcriptome are estimated and druggable network components (proteins) were elucidated (Khurana et al., 2017). In a multidisciplinary study examining the Huntington Disease metabolism in induced pluripotent stem cells (IPSCs) using Omics Integrator they found that ATP level is ablated in stem cells compared to differentiated cell while mitochondrial-related mRNA expression is balanced or upregulated in them (The HD iPSC Consortium, 2020).

In another study, Omics Integrator was used as a part of a toolkit for identification of therapeutic components with unknown or novel modes of actions (MoAs). In that interesting study, they examine the drugs found to be somehow beneficial in Huntington's Disease and group them with the same functions. Finally, autophagy activation is correlated with antihistamine effects of the drugs (Patel-Murray et al., 2020).

In conclusion, these previous studies state that in order to discover the dysregulation patterns of the diseases and to propose new drug targets, PCST algorithm through a simultaneous reconstruction of the molecular pathways is very useful.

1.4.7.1. Prize Collecting Steiner Tree (PCST) algorithm

The base of the Omics Integrator is the Prize Collecting Steiner Tree (PCST) algorithm. The function of the PCST algorithm is to find an optimal tree using a reference network. The function travels through the reference network, including the terminal nodes with prizes travelling through the interactome nodes which have costs of edges on condition if included. The terminal nodes are the ones given to the function to connect with and edges are network links containing the cost of travel. The task is to find shortest paths between the prize nodes avoiding the costs on the edges. The algorithm minimizes the cost of all edges by walking on prize nodes as much as possible. An example showing how the PCST working is represented in Figure 7.

Through the investigation stage for the terminal list, since the magnitude of the sets differs, various types of forests (in various complexity including one or more trees) can be generated. Hence, the parameters depend highly on the distribution of prizes and numbers of the nodes.

The whole function:

Equation 1: PCST algorithm

$$f'(F)= \sum_{v \in VF} p^1(v) + \sum_{e \in EF} c(e) + \omega \cdot \kappa$$

$$p^1(v)= \beta * p(v) - \mu * degree(v)$$

Here, p(v) is the prize of a node, c(e) is the cost of edge functions. β, ω, κ, and μ are normalizing factors. In order to construct meaningful trees from the input terminal lists, the normalizing factors must be fine-tuned.

ω is the cost of starting with a new tree controlling the number of trees in the forest. As ω gets smaller there will be more trees. β parameter controls the hubs in the network. Higher β value attenuates more hubs in the network and generates bigger networks. μ value also controls hubs in the network. To escape from a potential hub bias, reasoning if it is highly studied or has the greatest degree in the network, μ value competes against the degree of the nodes in the interactome. As μ increases, a smaller

number of hubs are dominated in the forest. When μ is zero, the control is cancelled. D is the last parameter which adjusts the depth of the tree.

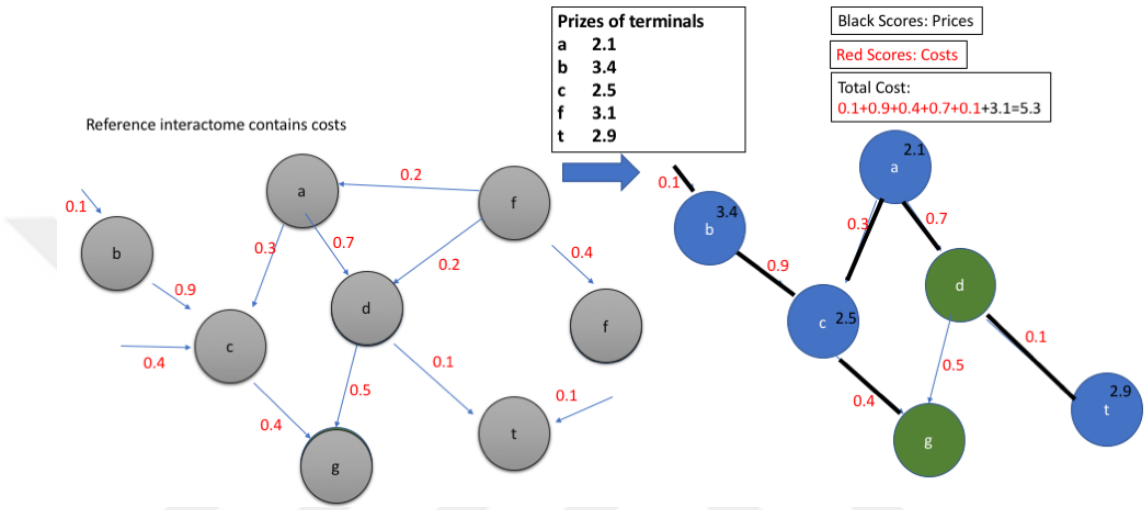


Figure 7: An example representation to show how the PCST algorithm performs its calculation. Grey interactome nodes with blue edges and red cost scores construct the backbone of the analysis. Given a set of terminal nodes, prizes are mapped into the interactome and connected through the shortest paths. Then the cost is calculated by summarizing the costs and prizes which if it is not able to be connected.

1.5. Preliminary Experimental Analysis at CANSYL

The experimental analysis investigating the PI3K isoform specific inhibition of the HCC cells on PTEN context was performed at CANSYL laboratory (METU) previously.

1.5.1. Molecular and Cellular Characterization of HCC in the Presence of Small Molecule Isoform Spesific PI3K Inhibitors

The expression levels and the phosphorylation status of key proteins in PI3K/Akt/mTOR and RAF/MEK/ERK signaling pathways were reported by our group, and in correlation with their PTEN status, Mahlavu cells display hyper-activated cell survival proteins (Durmaz et al., 2016).

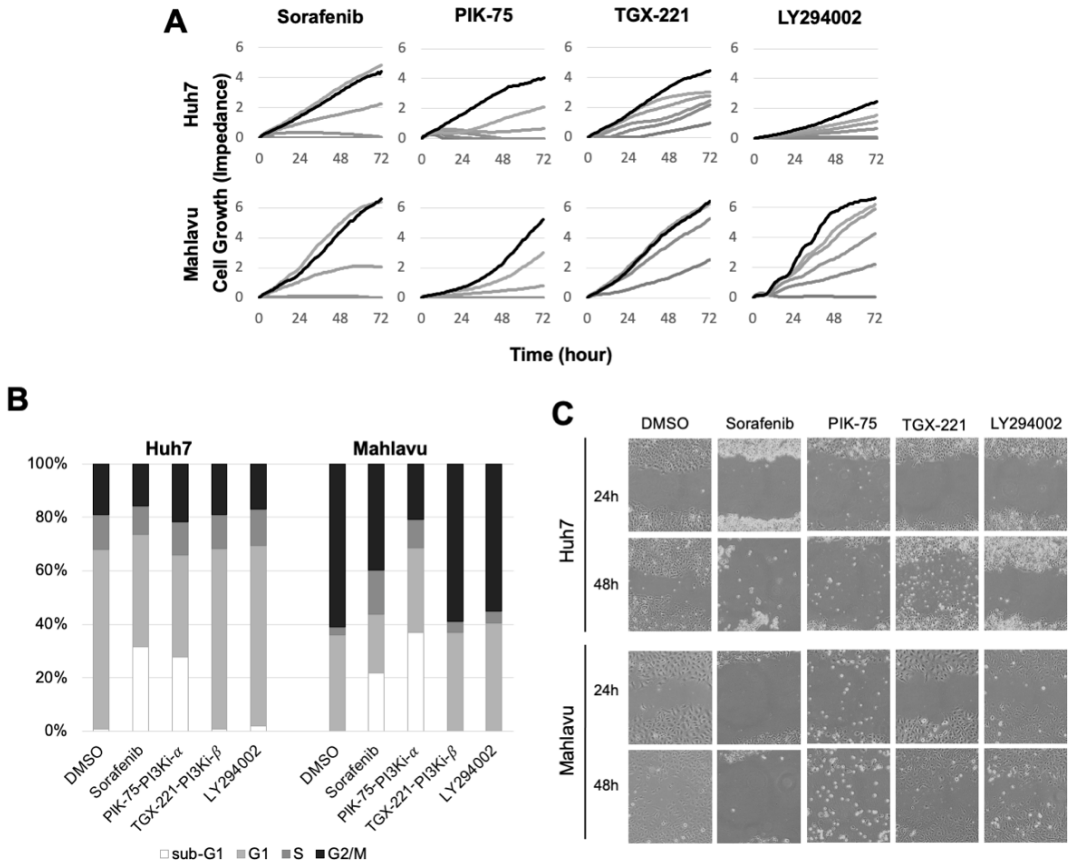


Figure 8. Characterization of HCC cells in the presence of small molecules inhibitors. Realtime cell growth analysis of Huh7 and Mahlavu cells with increasing concentrations (40µM, 20µM, 10µM, 5µM, 2.5µM) of Sorafenib, PI3K inhibitor LY294002, PI3Ki-β inhibitor (TGX-22) and PI3Ki-α (1µM, 0.5µM, 0.25µM, 0.125µM, 0.0625µM) PI3Ki-α (PIK-75) along with DMSO vehicle control (Control is black and increasing drug concentrations is given in grey level, highest concentration is being the darkest) A. Cell cycle analysis with flow cytometry. Sub-G1population represents apoptotic cells B). Wound healing assay for 24 and 48 hours for cell migration. C. 10µM of Sorafenib, LY294002 and PI3Ki-β (TGX-221) and 0.1µM of PI3Ki-α (PIK-75) were used for cell cycle and migration assays (Narci et al., 2021).

Initially, Sorafenib, LY294002, PI3K inhibitor p110α subunit specific (PIK-75) and PI3K inhibitor p110-β subunit specific (TGX-221) were assessed for their cytotoxic bioactivity and their effect on cell cycle progression with Huh7 and Mahlavu (Figure 8A or Figure 39). G1, S and G2/M cell cycle phases were analyzed separately to calculate viable cell distributions among them (Figure 8B or Figure 40). Sub-G1 percentage demonstrating apoptotic cells were also calculated. Cell cycle distribution remain stable for both cell lines and all inhibitor treatments. In both cell lines, Sorafenib and PIK-75 treatments showed stimulation of apoptosis through increase in sub-G1 population. In Huh7, Sorafenib was more active while PIK-75 functioned more

in Mahlavu cells which was more aggressive than Huh7 cell line by PTEN-loss based hyper-active Akt stimulation.

1.5.2. Migration Analysis of the Inhibitors

In order to analyze the effects of selected inhibitors on cell migration, wound-healing assay was performed. The percentages of wound closures after 48 hours of initial scratch were calculated for Huh7 and Mahlavu. It was observed that Sorafenib and PIK-75 reduced migration significantly ($p < 0.001$) in both Huh7 and Mahlavu (Figure 8C or Figure 41).

1.5.3. Synergistic cytotoxicity analysis

Since none of the treatments alone was fully effective to inhibit growth and stimulate apoptosis, the value of co-treatments of Sorafenib with PIK-75 and TGX-221 through real-time cell growth analysis (Figure 8C or Figure 41) were addressed. A synergistic effect of Sorafenib and PIK-75 treatments was observed on growth of both cell lines. TGX-221+ Sorafenib combinatory treatment also resulted in synergistic growth inhibition on Huh7 cell line.

On the other hand, TGX-221 displayed a growth inhibition of Mahlavu, TGX-221 co-treatment with Sorafenib resulted in an antagonistic effect and stimulated cellular growth. Furthermore, Sorafenib and PIK-75 treatment had more drastic effect on Mahlavu compared to Huh7. Therefore, these findings indicate that in PTEN deficient Mahlavu cells, constitutive activation of PI3K/Akt signaling mainly depend on p110-α (Narci et al., 2021) (Figure 9 or Figure 42, 43, 44, 45, and 46).

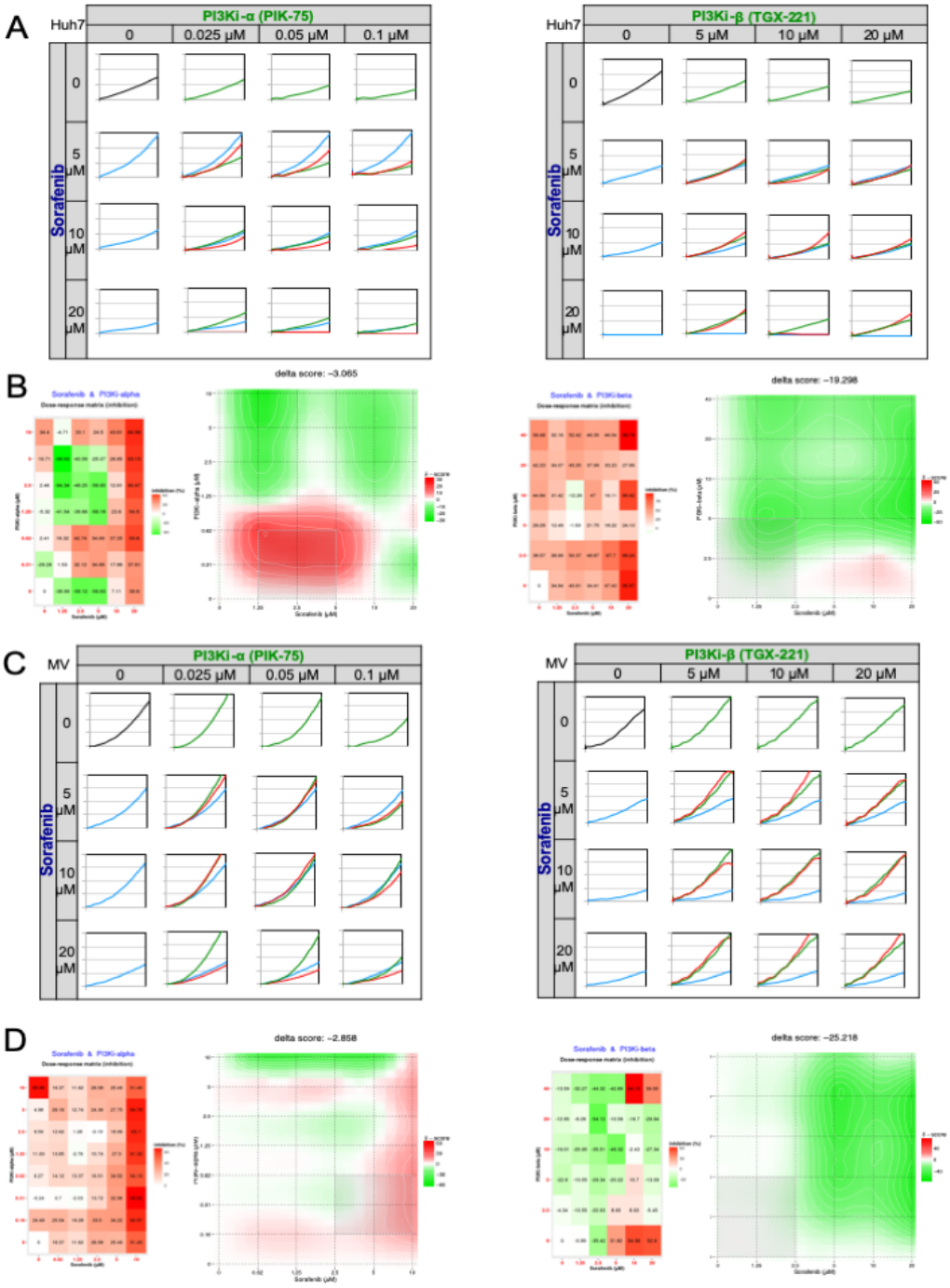


Figure 9. Real-time cell growth analysis. Human liver cancer cells Huh7 (A,B) and Mahlavu (MV) (C,D) were treated with the Sorafenib, PI3Ki-α and PI3Ki-β alone or in combination with increasing concentrations as indicated. Cell index measurements were obtained by RT-CES software. DMSO was used as negative control A. B. 72 hours of the percent growth inhibition values were used to calculate drug interactions with The SynergyFinder web application. Positive delta score reflects synergistic and negative score reflects antagonistic drug interactions. Experiments were performed in triplicate (Narci et al., 2021).

1.6. Motivation and Rationality

Hepatocellular carcinoma (HCC) is the most common subtype of liver cancer (Perz et al., 2006). The disease stages in a multi-step process where it originates from a liver injury and continues inflammation leads to developing cancerous cells (Farazi & DePinho, 2006). Often, for patients with HCC the only treatment option is chemotherapy as the disease can be diagnosed in the late stages where surgical options are not applicable (Llovet et al., 2012; Omata et al., 2010). Sorafenib (Nexavar, BAY43-9006) is the only targeted therapy for HCC yet the increase in the survival rate is not high for the patients and none of the other therapeutic agents can be approved by FDA (Liu et al., 2009; Llovet, Di Bisceglie, et al., 2008). Hence, there is an open demand for search of new targeted therapies for patients with HCC in advanced stages.

Acquired resistance and tumor recurrence are major drawbacks that Sorafenib suffers from. Sorafenib is a multi-kinase inhibitor targeting Raf, VEGFR and PDGFR which are the main controllers of cell proliferation and angiogenesis signaling pathways. Yet even these kinases were inhibited, carcinogenesis properties in the cell cannot completely diminish. Cellular pathways like proliferation, cell cycle, apoptosis, inflammation, angiogenesis, and migration are all integrated with redundant control of molecular pathways like Raf/MEK/ERK, PI3K/Akt, mTOR, NF-kB and p53. Therefore, using a single agent cannot remove the cancer cells completely. On the other hand, more than one type of kinase inhibitor should be used to avoid cross-talks and avoid compensation between the signaling pathways.

The discovery of effective agents against HCC was achieved mostly through identification of aberrant functions and dysregulated proteins from regulatory the signaling pathways (A Villanueva & Llovet, 2013). Traditionally, research laboratories focus their work into a single gene or a part of signaling pathway. With respect to the fact that these analyses are the most valuable evidence of protein interactions, they cannot solve the complex cancer mechanism as in-vivo techniques are costly and it takes a lot of time to do a series of analyses. Cancer development is a conclusion of several aberrant interactions and functions.

Earlier studies found that genomic variations in HCC is high and a wide-ranging investigation on molecular mechanisms of HCC should be done to fully understand the multi-stage development underlying behind (Ersahin et al., 2015). Current next generation analysis techniques assisted these studies to be implemented instantly. For example, a broad range scan of samples can be performed in hours using Illumina. Using these high throughput data allows to identify disease related gene profiles by enhancement of several bioinformatic methods.

In this thesis study, the benefit of high throughput techniques to get gene profiles of two Hepatocellular carcinoma cells were used. (PI3K/AKT/mTOR targets as it has been the driving signaling pathway in HCC are especially selected. Sorafenib, multi-kinase inhibitor, the only FDA approved agent, PI3K-α and PI3K-β inhibitors (PIK-75 and TGX-221) which targets two isoforms of PI3K were one by one treated to Huh7

and Mahlavu HCC lines. Furthermore, since single usage of these agents was proved to be ineffective to HCC, the combinational effect of, Sorafenib and PI3K isoform inhibitors (PIK-75 and TGX-221) were treated in combination too. Finally, the treatments with DMSO treated controls (12 in total) were scanned using RNA sequencing method Figure 10).

There are several bioinformatics tools for RNA-seq data analysis. The classical way of RNA-seq analysis pipeline includes basic preprocessing of the raw data followed with the identification of dysregulated genes. Even, the traditional workflow suggests the enrichment analysis after identification of up and downregulated genes, these methods only reveal the functional interactions inside the input gene set. A joint analysis using previous protein interactions may be advantageous for RNA-seq gene expression data. By this way, filtered or hidden connections between the genes may be revealed.

In this study, a set of protein-protein interactions from STRING dataset was used as backbone of gene relationships for Prize Collecting Steiner Tree (Ljubic et al., 2005) algorithm which is implemented in Omics Integrator software (Tuncbag et al., 2016) Up and downregulated genes were connected through the backbone network. This approach provided us to uncover some other gene profiles which either were not highly dysregulated or lost during RNA-seq analysis steps. By this way, functionally relevant pathways as trees for each condition were searched through not only with dysregulated genes, but also other connecting genes with the help of PCST.

Generation of kinase inhibitor treated HCC specific optimal networks allowed us to create a way to represent molecular events. Cytoscape, a java tool, was provided an effective visualization of the networks. Network representations were generated using several Cytoscape plug-ins displaying gene expression levels, interactions, regulations, processes, and functional and biological enrichments. Furthermore, a group of genes were identified as potential drug targets for HCC therapy using the benefit of network topology. Network centrality measures endorsed to select the most network specific targets while eliminating well studied ones. Overall, this systems level network approach provided an efficient network comparison to identify unique and shared gene profiles and functions of different kinds of kinase inhibitor treatments.

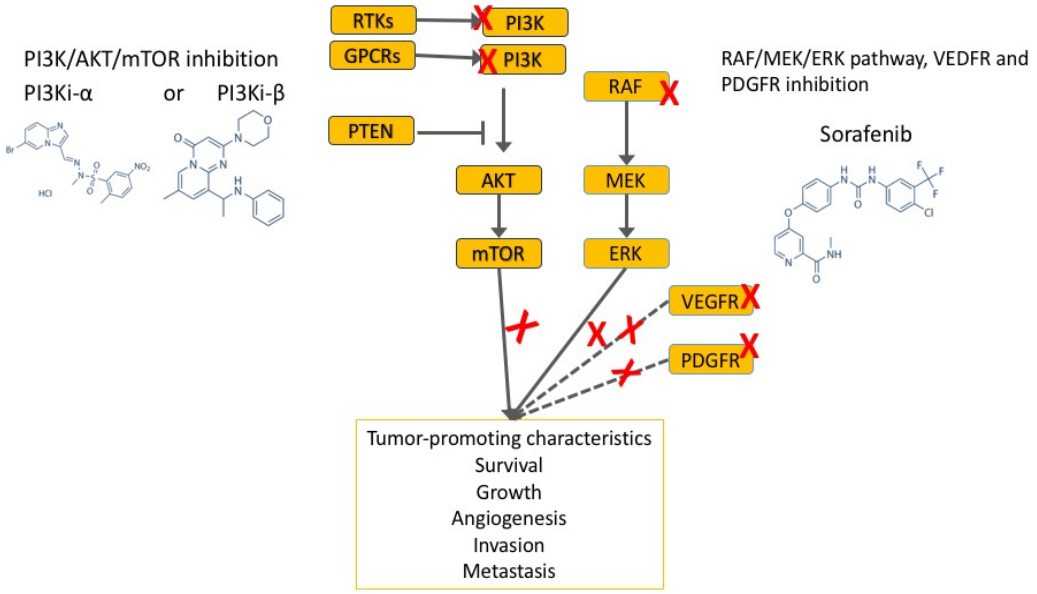


Figure 10: The aim of the study. The overall objective of this study is to find the most effective inhibitors targeting PI3K/AKT/mTOR signaling pathway to diminish tumor promoting characteristics like survival, growth, angiogenesis, invasion, and metastasis. Sorafenib and PI3K class IA and class IB inhibitors PI3K-α (PIK-75) and PI3K-β (TGX-221) are combinational treatments to analyze synergistic effects of inhibition of PI3K/AKT/mTOR and RAF signaling pathways.

CHAPTER II

MATERIALS AND METHODS

This chapter is divided into two sections as materials and methods. In the materials chapter, Hepatocellular Carcinoma (HCC) cell lines used in RNA-seq experiment, and the kinase inhibitors used to treat cell lines are explained. In the methods section, initially, the workflow followed to perform bioinformatics analysis is explained. The methods and tools used in RNA-sequencing and network analysis are detailed. The tools, tool versions, and parameters used in statistical, heatmap and visualization analysis are described in this chapter. Omics Integrator, network construction tool, optimization and run parameters are also part of this chapter.

2.1. Materials

HCC cell lines and kinase inhibitors used in this study are explained.

2.1.1. HCC Cell Lines

In this study two types of HCC cell lines were used: Huh7 and Mahlavu. Mahlavu is a mesenchymal-like, undifferentiated, very aggressive type of HCC where no tumor suppressor protein, PTEN expression is seen. PTEN protein antagonizes for Akt activation through ensuring dephosphorylation of PIP3. When there is no PTEN expression, Akt protein is constitutively expressed as like its downstream proteins (Schwartz et al., 2016). A previous study (Buontempo et al., 2011) proved that besides overexpression of proteins in the AKT pathway, the RAF/ERK signaling pathway is also hyperactive in Mahlavu cells (Figure 11).

Unlike poorly differentiated Mahlavu cells, epithelial-like, tumorigenic Huh-7 cells is a well differentiated HCC type. As shown in Figure 11, there is PTEN expression blocking Akt in the usual mechanism and so the downstream proteins of AKT are not overexpressed.

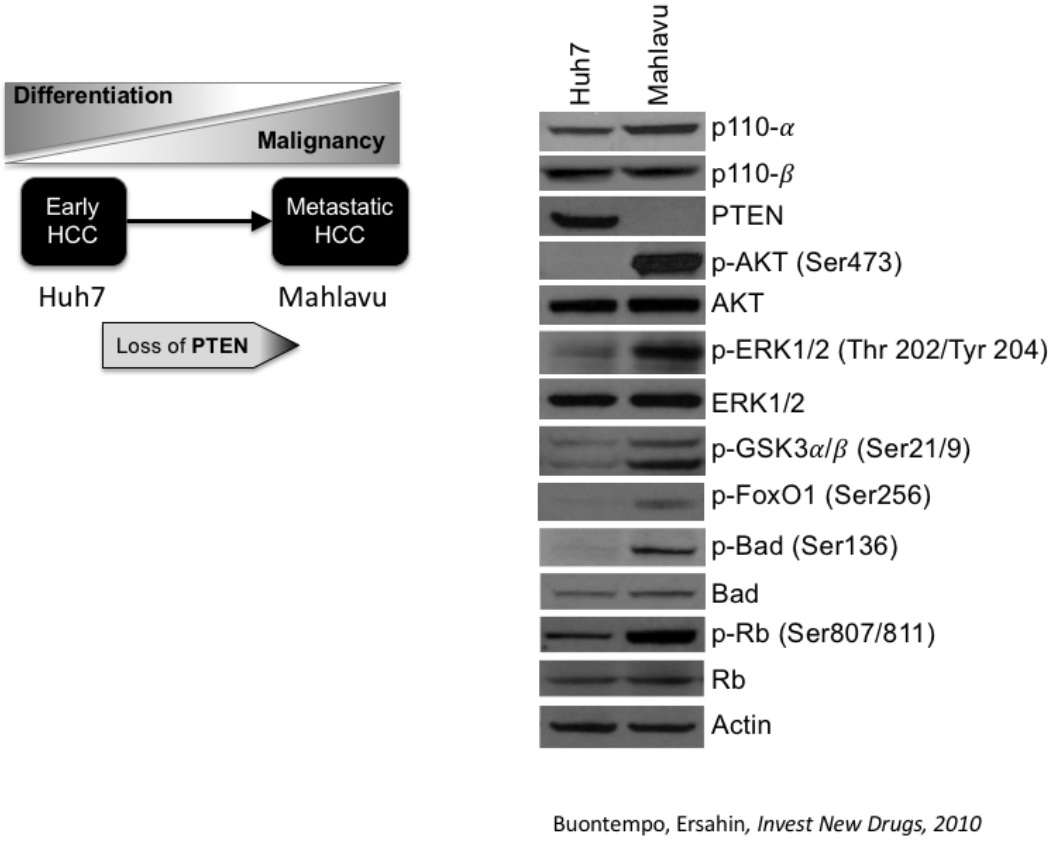


Figure 11: Expression and activity of critical components of PI3K/AKT and RAF/MEK/ERK signaling pathways in Huh7 and Mahlavu (Buontempo et al., 2011) (Copyright 2021).

2.1.2. Kinase Inhibitors

Mahlavu and Huh7 cell lines were treated with the inhibitors which were listed in Table 1 with their targets. Sorafenib as a multi-kinase inhibitor of RAF and VEGFR and two isoforms of PI3K class I inhibitors PIK-75 and TGX-221 targeting PI3K were used. The target proteins of the inhibitors used in the study can be viewed from Figure 12, and AKT is the main downstream protein for all treatments. Sorafenib (Nexavar) was purchased from Bayer Health-care Pharmaceuticals, Inc., NJ USA, Inhibitors PIK-75 (cat#528116) and TGX-221(cat#528113) were purchased from Calbiochem.

Towards understanding of the alterations in the signaling pathways of diseased cells during hepatocarcinogenesis, a global study model was constructed comprising two HCC cell lines treated with kinase inhibitors alone or in combinations; PIK-75, TGX-221 and Sorafenib alone treatments and PIK-75 + Sorafenib and TGX-221 + Sorafenib combinatory treatments. DMSO treatment was used as negative control for both cell lines. Therefore, for the RNA-seq experiment, cDNA libraries prepared from 2 cell lines per 5 treatments and 1 control for each in total 12 samples.

Table 1: Kinase inhibitors used in RNA library construction.

Inhibitor	**Target**
Sorafenib (BAY 43-9006)	B-Raf and VEGFR
PI3Ki-α (PI3K alpha inhibitor VIII) (PIK-75)	P110α isoform of PI3K (class IA)
PI3Ki-β (PI3K beta inhibitor VI) (TGX-221)	P110β isoform of PI3K (class IB)

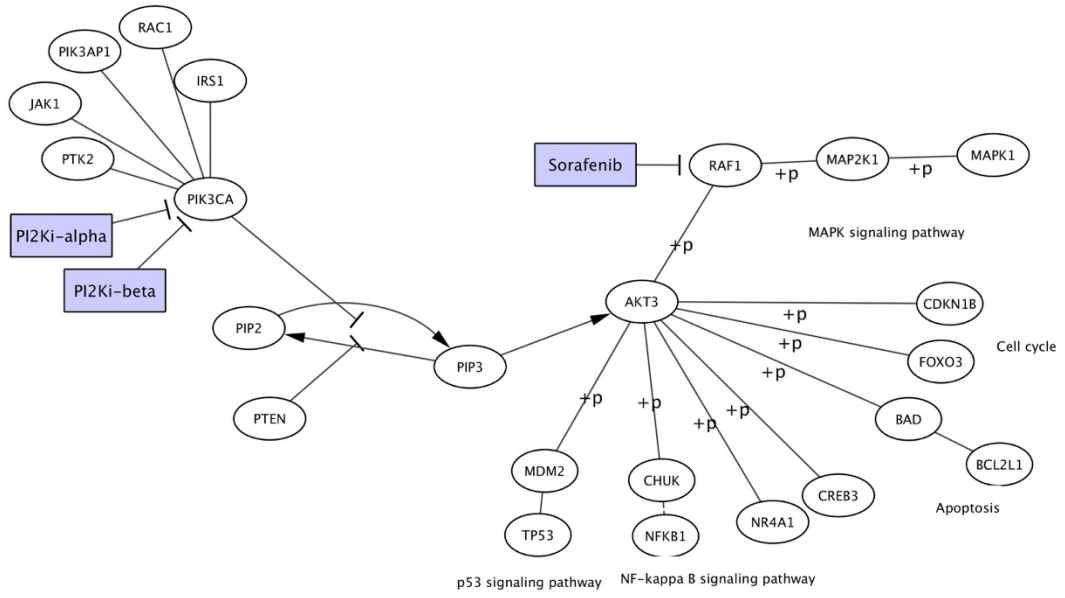


Figure 12: Schematic representation of the kinase inhibitors used in the study, their target proteins, and the downstream proteins of their targets.

2.2. Methods

Figure 13 displays the workflow of the methods used in this study. The experimental design included two HCC cell lines; Huh7 and Mahlavu. These cell lines were treated with three types of kinase inhibitors, two of their combinations, and DMSO as negative control as described in the materials section. RNA-seq libraries were prepared from 12 treatment sets and they were sent for sequencing using Illumina Genome Analyzer. Further RNA-seq analysis as follows; raw data processing using FASTQC, alignment and mapping using Bowtie with Tophat tools, and differential expression analysis

using EdgeR. Omics Integrator with the integration of STRING human PPI provided kinase inhibition modeling of the networks. Topology analysis of the networks performed with NetworkX. Dendrogram, clustering and enrichment analysis were executed using several R packages. BINGO tool offered gene enrichment analysis. Glay was used to cluster network, and ClusterMaker was used to box the clusters. Cytoscape was the main tool where visualization, clustering and enrichment analysis performed in. In the following section each process will be explained in further.

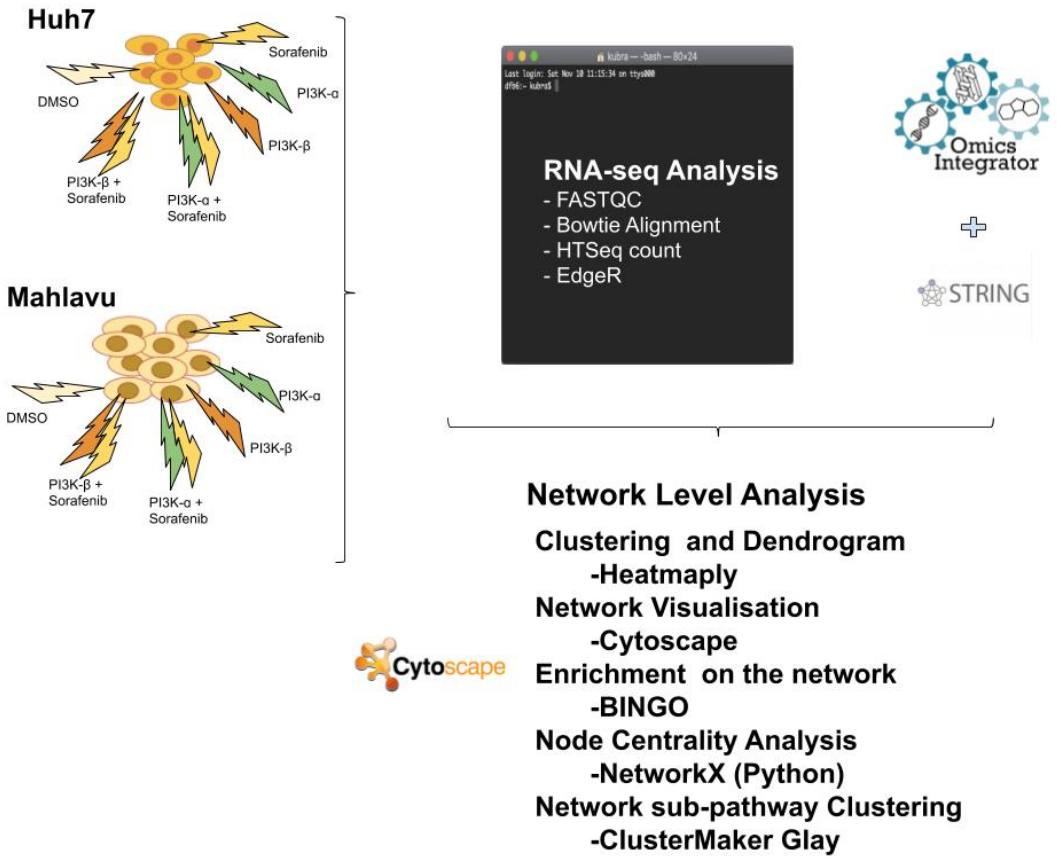


Figure 13: Pipeline summarizing the methods used in this thesis study. RNA-seq analysis was performed on kinase inhibitor treated Huh7 and Mahlavu cell lines followed by a network level analysis.

2.2.1. Cell Culture

Mahlavu and Huh7, HCC cell lines were cultured in DMEM medium, supplemented with 10% fetal bovine serum (FBS), 1% penicillin/streptomycin (P/S) and 1% non-essential amino acids (NEA) and incubated in humidified 37◦C incubator with 5% CO2. The cell maintenance was performed by one of the members of the CANSYL laboratory.

2.2.2. RNA Sequencing Analysis

RNA-seq analysis was done through these steps summarized in Figure 14: total RNA preparation, sequencing, raw file quality control, alignment of the reads to the reference genome, quality assessment of alignment files, read counting per feature and differential expression analysis. Commend line of the tools was provided in Appendix C.

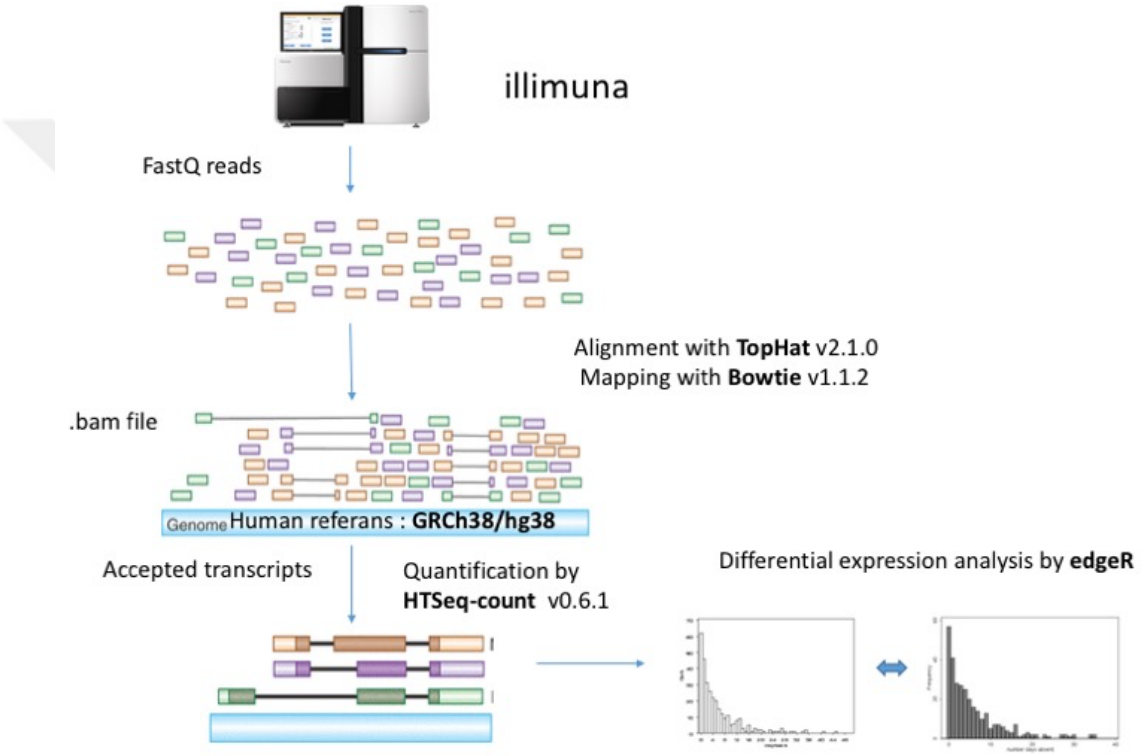


Figure 14: Summary of RNA-seq analysis pipeline. Raw FASTQ reads were aligned and mapped to the human reference genome, accepted transcripts were quantified, and differential expression analysis was applied sequentially. A part of this figure taken from edgeR publication (Robinson et al., 2009) and illumina distributed the figures publicly.

2.2.3. Preparation and Sequencing of Total RNAs

Total RNA was isolated with NucleoSpin RNA II Kit (Macherey-Nagel) according to the manufacturer's protocol (MN, Duren, Germany) in CANSYL previously. The protocol was applied with small modifications such as 30 min of DNA digestion instead of 15min and 2-step elution with 20µl water instead of one elution with 60µl. RNA concentration was measured with NanoDrop and A260/A280, A260/A230 ratios were checked for RNA quality and purity. Total RNAs were provided to BGI Tech (https://en.genomics.cn/) for sequencing. RIN values were acquired from the Agilent

Bioanalyzer system, and they were above 0.8 for all samples. 12 total RNA samples were sequenced by Illumina Genome Analyzer in a sequencing center.

2.2.4. Quality Assessment of Raw Files

RNA reads were processed by Illumina Hiseq 2000 (SE50). RNA sequencing by Illumina Genome Analyzer generates output sequencing files in FASTQ format. Besides the raw sequences, this format also stores quality scores per base. For a proper and significant alignment of the sequences, technical errors should be eliminated. It is necessary to remove low-quality bases and duplicates and clean the remaining adaptors.

In this study, 12 FASTQ files were first assessed through the FASTQC (Andrews, 2010) tool for quality control. FASTQC is a well-known quality assessment tool developed by Babraham Bioinformatics for FASTQ formatted data. FASTQC tool generates quality reports evaluating per base and per tile sequence quality, per base and sequence content, GC content, sequence length distribution, duplication level, overrepresented sequences, and adapter content. In order to clean FASTQC data, second party applications like Cutadapt (Compeau et al., 2013) must be used, after a careful analysis of FASTQC reports.

2.2.5. Alignment

Since the quality of RNA-seq of all FASTQ were outperforming by the per base and sequence quality, in this analysis there was no need for a trimming step. Single-end reads were aligned to the reference human genome (GRCh38/hg38) using a split read aligner algorithm TopHat V2.1.0 (Trapnell et al., 2009). TopHat is a frequently used splice junction mapper for RNA-seq reads. Given a reference genome, it can read and align high throughput short reads. TopHat itself features an ultrafast mapper Bowtie (v2.2.6) algorithm for alignment and analyses the mapped reads to detect exome-intron junctions.

TopHat generates a list of files containing mapped and unmapped reads separately in BAM and SAM formats with indexes and insertion, deletion, and junction files in txt formatted BED files.

2.2.6. Alignment Quality

The quality of alignment files assessed using a JAVA based tool CollectAlignmentSummaryMetrics from Picard (Broad Institute, 2018). The tool processes a SAM or BAM file input and as output produces a number of read quality metrics like the number of high-quality reads, read lengths, number of noise reads, bad cycles, strand balance and mismatches.

2.2.7. Quantification

The number of mapped reads to the genome was counted using HTSeq-count v0.6.1 (Anders et al., 2015) tool. Given aligned sequencing reads with a file with genomic feature list, HTSeq-count quantifies the reads map per features. For quantification analysis, human gene split regions (GRCh38 v84) in a GTF file format was used to count how many transcripts map to each gene, generating a gene level count matrix.

2.2.8. Differential Expression Analysis

EdgeR (Robinson et al., 2009), a Bioconductor package, is a widely used method for differential expression analysis. EdgeR constructs a negative binomial model based on empirical Bayes estimates, exact test or linear models using the RNA count data.

Normally, edgeR requires biological replicates in order to estimate biological variation inside the samples. In this experimental design, there were no biological replicates of the samples to inherit the in-sample variation. EdgeR solves the no-replication problem by suggesting a different dispersion calculation method to estimate variation within each sample compared to housekeeping genes. A set of housekeeping genes in Hepatocellular Carcinoma was well characterized in (Ersahin et al., 2014), this set was used to estimate biological coefficient of variation (BCV) value manually. BCV estimation was performed for both Huh7 and Mahlavu cells separately in this analysis and the values were very close to each other, and usage of the same value did not affect the differential expression gene status. Hence, BCV was used as 0.045, originally calculated for Huh7, for both cell lines.

Gene level count matrices of 12 RNA-seq treatment sets were used for differential expression analysis. DMSO treated Huh7 and Mahlavu HCC samples were used as negative control. Before edgeR analysis, genes with less than 5 readings were filtered out using counts per million constraint (cpm < 5). A biological model was constructed by taking BCV as 0.045 and differential analysis performed using exactTest function of EdgeR package. Finally, I selected the top differentially expressed genes according to following filters; p-value<0.01, FDR<0.01, and logFC (log2 of fold changes) ranges less than -2 & over +2 for Huh7 cell lines and -1.5 & over +1.5 for Mahlavu cell where the FDR is defined as False Discovery Rate and the logFC is logarithmic Fold Change.

Before filtration, the genes were annotated using the ensembl gene id to gene symbol, entrez id and unigene names. Gene annotations were obtained using org.Hs.eg.db R package (Marc Carlson, 2016) from Bioconductor. R scripts including the functions to calculate biological variance using housekeeping genes and to perform differential expression analysis were included in the GitHub repository mentioned in Appendix B.

2.2.9. Dendrogram Analysis

Heatmap representation is one of the most popular graphical methods for visualization of big-data supplying color encoding cells that represent numbers. Heatmaps are frequently used to segment expression data, find similar or different expression patterns, and perform closeness analysis between the samples.

Heatmaply (Galili et al., 2012) is a very powerful way of investigating clusters in a high dimensional data since the final heatmap result is visualized as an interactive graph offering inspection over the cells making zoom-in functions available. In this study, all heatmaps were visualized through heatmaply using default *hclust* clustering by using Euclidean as distance measure. The codes to generate heatmaps in this study was provided in the GitHub repository mentioned in Appendix B with the produced html files.

2.2.10. Correlation Analysis

In order to assess the similarity co-efficiencies between the different treatments in the same cell line, Pearson correlations were calculated. A count matrix was generated containing gene logFC values (without any filtration) of all treatments per HCC line. R "cor" function was used to calculate Pearson correlations and the "corrplot" function was used to graph the results.

Pearson correlation is a number between -1 and +1, as the value approaches to -1 negative correlation gets stronger, whereas as it approaches to +1 that indicates a significant positive correlation.

2.2.11. Venn Analysis

In order to calculate and draw Venn schemes, the online tool Venn Diagram (http://bioinformatics.psb.ugent.be/webtools/Venn/) was used. It is a very handy tool both to generate Venn diagrams and to get intersections between the files. Venn diagrams generated in this study were used to compare top DEG genes and network nodes.

2.2.12. Gene Ontology (GO) Analysis

The Gene Ontology (GO) project aims to construct a model of biological systems computationally. A structured description of known biological information, which is designed as a tree of vocabularies, consists of multiple layers. The GO terms or vocabularies associated with biological processes, molecular functions or cellular components (Ashburner et al., 2000; Carbon et al., 2021).

Given a set of genes on the network, an open source tool BiNGO (Maere et al., 2005) maps functional terms to enriched genes to output Gene Ontology (GO) terms and their statistical features. BiNGO is a java-based tool integrated into Cytoscape. Tool finds the predominant or overrepresented gene ontology terms and links these terms in a map proving a good way of GO hierarch representation.

Statistically overrepresented GO terms were characterized using BiNGO in our analysis to have a better understanding of the processes that the selected genes have roles. A hypergeometric test was applied using to the gene sets or network nodes with the selected annotation file (H_sapiens_default) of Homo Sapiens.The default GO annotation file (GO_full) is used as ontology file. Input parameters were adapted to be used as the same for all analysis. A stringent Benjamini & Hochberg False Discovery Rate (FDR <0.005) was used to filter out non-significant GO terms. BiNGO results were kept in .bgo files and exported into R for series of downstream analysis.

Table 2: Experimental evidence codes used in the study.

Code	**Description**
EXP	Inferred from Experiment
IDA	Inferred from Direct Assay
IMP	Inferred from Mutant Phenotype
IGI	Inferred from Genetic Interaction
IEP	Inferred from Expression Pattern
IPI	Inferred from Physical Interaction

GO sets containing redundant and electronically annotated terms generate a huge noise for functional comparisons. These terms either originated electronically or have ambiguity in them. The GOs other than experimentally validated codes were filtered out (accepted codes shown in Table 2) to avoid suspicious GO terms and set a level to the analysis. The codes were matched through publicly available Gene Ontology Annotation (GOA) database (http://current.geneontology.org/annotations/).

2.2.13. Network Analysis

Hidden expression patterns in the differentially expression sets which could be lost in RNA-seq analysis can be explored through construction of protein-protein interaction networks. The classical downstream analysis of RNA-seq is performed using DEG sets, and the analysis is restricted to gene set enrichment analysis, clustering analysis or gene-based database searches. Using these types of analysis are limited by the input

gene lists. Furthermore, the rules for DEG filters or statistical limitation parameters are not specific and general usage of the parameters are experiment based. Hence, construction of an interaction pathway from DEG sets can uncover some expression which may be lost during up-stream analysis.

Omics Integrator is a package that aims to analyze proteomic, gene expression or epigenetic data starting from omic data with the aim of network construction by PCST formulation. It consists of two modules, Garner and Forest works one by one to construct a network connecting proteomic and gene expression or epigenetic data. For downstream RNA-seq analysis, Omics Integrator software Forest module determined multiple sub-pathways in the human interactome given top DEGs as input terminals.

PCST (Prize Collecting Steiner Tree) algorithm first used for biological network by (Tuncbag et al., 2016) aims to identify sub-networks from an interaction network given a set of weighted genes. By using PCST, the biologically meaningful interactions between the DEGs from human protein-protein interaction data were extracted. The PCST algorithm finds an optimal tree, including the terminal with prizes travelling through the interactome nodes which have costs of edges only if they are included. The task is to find the shortest paths between the prize nodes avoiding the costs on the edges. The algorithm minimizes the cost of all edges by passing through many prize nodes as possible.

STRING human protein-protein interaction database v.10 was used as backbone of PCST database in this study. V10 STRING database had 19,247 unique proteins with 8,548,002 interactions. PII in STRING is weighted by confidence scores using numerous proofs. The confidence score was restricted by a high confidence value (>0.7) with stronger experimental proof needed. In STRING edges are scored according to a confidence score determined through an algorithm. If an interaction is proved through experimentally the confidence score gets higher basically in a range of 0-1. After the filtration, there were 12,910 unique proteins with 333,324 edges. The cost of the edges in OmicsIntegrator is determined by getting negative logarithm of these scores, so the cost and the interaction confidence are negatively correlated. The price list given to the algorithm was a DEG list with ENSEMBL IDs, in order to match those to STRING first, DEG ENSEMBL IDs converted into Gene Names using Ensembl BioMart. Then, STRING PPI was converted from protein names to Gene Names using the same strategy. By this way, while using the same Gene Names was managed, untranslated transcripts from the DEG lists were lost.

2.2.14. Optimization of Forest Parameters

In order to construct meaningful trees using differentially expressed genes, input Forest parameters must be fine-tuned. The size and degree of the forests are expected to vary as the number of genes in the input files changes. Forest parameters depend highly on the distribution of prizes and numbers of the nodes and there could be hundreds of possible solutions for numbers of input terminals. The best combinations

of parameters for each DEG set were explored using forest-tuner script (https://github.com/gungorbudak/forest-tuner) which is PCST algorithm parameter tuner for ω, β and μ parameters. This script was used to find the best arrangements of the parameters to be used in the Forest module for each treatment.

Table 3: Forest parameters and their effect into the re-generated sub-interactomes.

Parameter	Description	Effect
ω	Cost of starting to a new tree	Less trees
β	Controls hubs	Larger networks
μ	Controls hubs	Smaller networks
D	Adjust depth of the tree	Longer trees

The parameters were tuned in the following ranges: ω (1-10.0 or 5-15), β (1-15.0), μ (0.01-0.05). Here, ω parameter tunes the number of trees in the network, β parameter increases the number of prices entering the tree and μ is another parameter that arranges the dominance of hub proteins in the network (Table 3). D parameter, which is not tuned and taken as 10, also controls the depth of the tree resulting with longer trees as it gets larger. From all possible solutions, the combination which generates a minimum degree network with the greatest number of prize nodes was selected. Config files including selected parameters per kinase inhibitor treatment were kept for forest analysis.

Commend line parameters to run Omics Integrator is provided in Appendix C and the corresponding codes and required files can be found in GitHub repository mentioned in Appendix B.

2.2.15. Forest Module Runs

Omics Integrator "forest.py" was run with the config files generated by forest-tuner.py. As a result of the runs, forest outputs a list of files whose content are explained below:

- Sample_info.txt summarizes the information on the run for debugging.
- Sample_optimalForest.sif includes the final optimal network in sample interaction file format (SIF). The file is ready to open with Cytoscape.
- Sample_augmentedForest.sif look likes optimal forest file yet it also includes all the edge interactions inherited directly from reference interactome.

Betweenness centrality in the node attribute file is calculated through this network.

- Sample_dummyForest.sif look likes optimal forest file yet it also includes the dummy nodes and interacting edges that the algorithm eliminated.

- Sample_edgeattributes.tsv includes information about the edges in the network; its weight in the interactome, the fraction of optimal networks containing the edges, separated by tabs.

- Sample_nodeattributes.tsv includes information about the nodes in the network, its prize and betweenness centralities, separated by tab.

2.2.16. Randomization Tests

In order to test the significance of the nodes appearing in the optimal nodes, each PCST network was subjected into 100 randomization tests using the Omics Integrator tool (specified by *–randomTerminals*=100). The tests were performed using a random set of terminals with respect to keeping node numbers, and original interactome set with same edge weights and optimization penalties the same.

Here, the probability that a node randomly be present in the network was expressed by its frequency of randomness in the network producing the 'specificity index. Therefore, less frequent nodes would be the most specific ones to the network (Tuncbag et al., 2013). Through the analysis, the nodes that appeared only once in the random networks was considered.

2.2.17. Centrality Measurements

Centrality measures are the indicators of the most valuable vertices in the graph for network analysis. Centrality is often used to identify influential nodes of the network and by this way provides a ranking which identifies the important nodes in the network. Degree, eigenvector and betweenness centralities calculated to estimate network topology. NetworkX python library (Hagberg et al., 2008) was used to calculate centrality measures. Here is the list of centrality measures used in this study:

- Degree Centrality: Proportion calculated by the number of connections a node has.

- Eigenvector Centrality: Computes the centrality for a node based on the centrality of its neighbors.

- Betweenness Centrality: Calculates how many times a node bridges along two node's shortest path.

2.2.18. Effective Visualization of the Optimal Networks

Omics Integrator outputs the optimal networks in SIF format. SIF files are ready to import Cytoscape as default, yet for a practical analysis the optimal network should be shaped and arranged efficiently.

Cytoscape includes many add-ons for biological network analysis including annotation, pathway enrichment and clustering. Cytoscape was used for network analysis in this study to use all of these functionalities together. The methods used for network representation are illustrated in Figure 15.

The classical way of a biological network analysis is to perform an enrichment analysis to understand the functions of these genes/nodes having a role on. Using all the inputs together for the enrichment analysis results in either too general GO terms to analyze or losing too definite terms as the statistical significance is vanished for big size networks, for example. The optimal network generated through forest includes sub-trees. Biologically, these branches are representing a group of proteins closely interacting each other. So, these sub-trees can be representation of functional groupings. In this study, a topology-based clustering method was performed to detect those groups. Then, enrichment analysis was performed using these clusters and after most important GO terms were selected, they were back mapped to the network generating a well-designed network representation summarizing the functional groups and which role they were engaged in.

The network nodes were sized along with the node labels with betweenness centrality values to emphasize their centrality. The hub nodes are represented with a bigger size. Afterwards, to understand the expression patterns and transcriptional flow direction, upregulated genes were colored to red while downregulated ones colored to blue. As the value gets stronger the darkness of the value increases. The final visualizations represented all gene relationships, up and downregulated genes and internal Steiner nodes. Also, highly connected groups and GO annotations were provided for an easy and efficient way to compare networks with each other. For network representations, these following files were used.

1. A SIF file containing the optimal network, forest module run results.

2. A TSV file containing the betweenness centralities, forest module run results.

3. A CSV file containing the genes, logFC values and other annotations, edgeR results.

4. A SIF file including the selected GO terms and their gene links, generated from merged .bgo files.

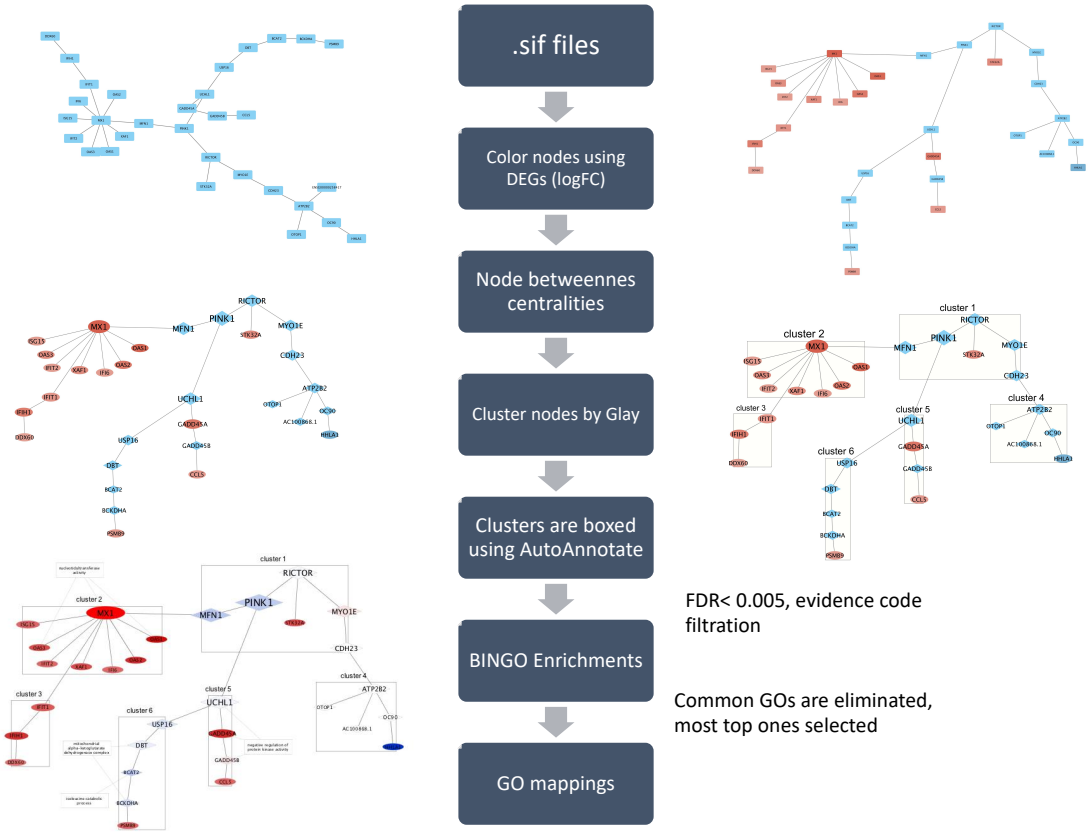


Figure 15: Summarization of the methods used for effective visualization of optimal networks.

2.2.19. Clustering

After SIF files were imported into the Cytoscape, yFiles layout algorithms were implemented and hierarchical layout was selected for visualizations. In order to cluster the networks, Glay algorithm (Su et al., 2010) using edge betweenness centrality was implemented. In this regard, the genes which are topologically close to each other in the network were better detected. Application of this strategy resulted in the most connected patterns in the networks.

2.2.20. Annotation

The Cytoscape plug-in AutoAnnotate tool (Kucera et al., 2016) was used to annotate the clusters. It is a very easy to use method since it can automatically detect the groups or work directly on the already generated ones. The autogenerated labels are also editable further. the clusters were annotated based on the selected GO terms to better represent the functions of the connected proteins inside the networks.

Therefore, this method provided a nice way to properly interpret the network and easy compare of the different networks.

2.2.21. GO Analysis and Mapping

In order to better understand what processes these clusters have role on, statistically overrepresented Gene Ontology (GO) terms were characterized using BiNGO for each cluster in networks. The lists were filtered out using a stringent FDR cut-off (0.005). Selected GO terms were imported into the network using their gene maps and mapped into the clusters using AutoAnnotate tool.

2.3. Prioritization of the Nodes in the Optimal Networks

Drug target genes were proposed by following a series of network filtrations based on network topology. The hub nodes needed to be avoided because they are commonly the most studied ones. Target node needed to be network specific and significant for the network. Thus, the strategy was following this idea: a node would be candidate to be target only if it occurs in the branches on random networks while presenting in more central areas on the optimal networks. In order to achieve these nodes, 100 randomization tests were performed to eliminate frequent nodes in random networks. the least frequent nodes (specificity index smaller than 0.01) for construction of optimal networks were used.

Then, three stage selection were performed for target selection:

1. The hub nodes of optimal networks were selected though using degree, eigenvector and betweenness centralities greater than 0.001.

2. From the remaining nodes, the ones that were predominant in the random network were eliminated using degree centrality of random networks bigger than 0.001.

3. Finally, top 20 target for each cell line were selected upon sorting by their betweenness centralities.

2.4. Knock-out Experiment

The prioritized nodes were silenced in the optimal networks using *–knockout* parameter provided by Omics Integrator tool. The nodes (genes) given to the tool one by one with the corresponding network, and the effects of the remove was analyzed afterwards. If exists, the number of effected node and the effected gene ontology were counted and summarized. The node would be important if either effecting more than two nodes or at least one gene ontology.

CHAPTER III

RESULTS

In this chapter, the results of the bioinformatics analysis performed as a part of the study are discussed. The chapter is separated into two sections: As a part of RNA sequencing results, the number, the quality, and the alignments of reads are explained first and, then the results of the differential expressions of the genes and similarity-based gene correlations are described. In the second section, network-based interpretation of the data, the results of optimal PCST network constructions, comparisons of the networks, network clustering, network gene enrichment, gene prioritization and the effect of node removals from the network are explained.

3.1. RNA Sequencing Results

The following section explains RNA sequencing results. The quality reports for raw FASTQ reads and alignments of the reads, differentially expressed genes, Pearson correlation or similarity analysis performed with dendrogram and analysis of expression patterns in HCC cell lines are discussed.

3.1.1. Quality Check Reports

12 Single-end FASTQ reads, were processed by Illumina Genome Analyzer, were analyzed using the FASTQC tool. The base quality scores of raw reads for all treatments were plotted and added into the Appendix D. The bases quality of all reads was above average and in good quality. According to FASTQC generated reports, single end reads 49bp in length were well sequenced and no end bias was seen for the reads. Total sequences, sequence lengths and GC% contents were represented in Table 4. Mean value of base quality scores was above the average. An average of 33,465,904 clean reads in 47-50% GC were generated.

Table 4: Total sequences, sequence length and GC% content of the FASTQ files of the HCC Cell lines.

HCC Cell Lines	**Quality Measures**		
Huh7	**Total sequences**	**Sequence Length**	**GC% Content**
PI3K-α inhibitor (PIK-75)	33192992	49	48
PI3K-β inhibitor (TGX-221)	32730853	49	50
PIK-75 + Sorafenib	34017394	49	49
TGX-221 + Sorafenib	28579376	49	49
Sorafenib	36506606	49	49
DMSO	37827459	49	48
Mahlavu	**Total sequences**	**Sequence Length**	**GC% Content**
PI3K-α inhibitor (PIK-75)	34302640	49	47
PI3K-β inhibitor (TGX-221)	29776671	49	49
PIK-75 + Sorafenib	31897951	49	49
TGX-221 + Sorafenib	37221329	49	49
Sorafenib	32116151	49	50
DMSO	33421422	49	49

3.1.2. Alignments and Gene Counts

Table 5 represents the alignment qualities per library. The aligned number of reads was on average 27-36 million reads, and the alignment rate per line was about 97%. The coverage of the alignments was at most 47X and at least 35X. Therefore, all the alignments were of high-quality and coverage.

Table 5: Number of total processed sequences, alignment rate in percentages, and the coverage of the alignments for HCC Cell lines.

HCC Cell Lines	**Library Quality Measures**		
Huh7	**Total reads**	**Alignment Rate (%)**	**Coverage**
PI3K-α inhibitor (PIK-75)	32184033	97.0	41
PI3K-β inhibitor (TGX-221)	31841448	97.3	40
PIK-75 + Sorafenib	33134724	97.4	42
TGX-221 + Sorafenib	27899790	97.6	35
Sorafenib	35665666	97.7	45
DMSO	36867755	97.5	47
Mahlavu	**Total reads**	**Alignment Rate (%)**	**Coverage**
PI3K-α inhibitor (PIK-75)	33219644	96.0	42
PI3K-β inhibitor (TGX-221)	29018356	97.5	37
PIK-75 + Sorafenib	30868036	96.8	39
TGX-221 + Sorafenib	36355199	97.7	46
Sorafenib	31292278	97.4	39
DMSO	32591562	97.5	41

In Table 5, the coverage of the alignment files calculated using the following equation, taking human transcriptome length as 39,841,315 bp.

Coverage = (Sequence Length x Total Sequences) / Human Transcriptome Length

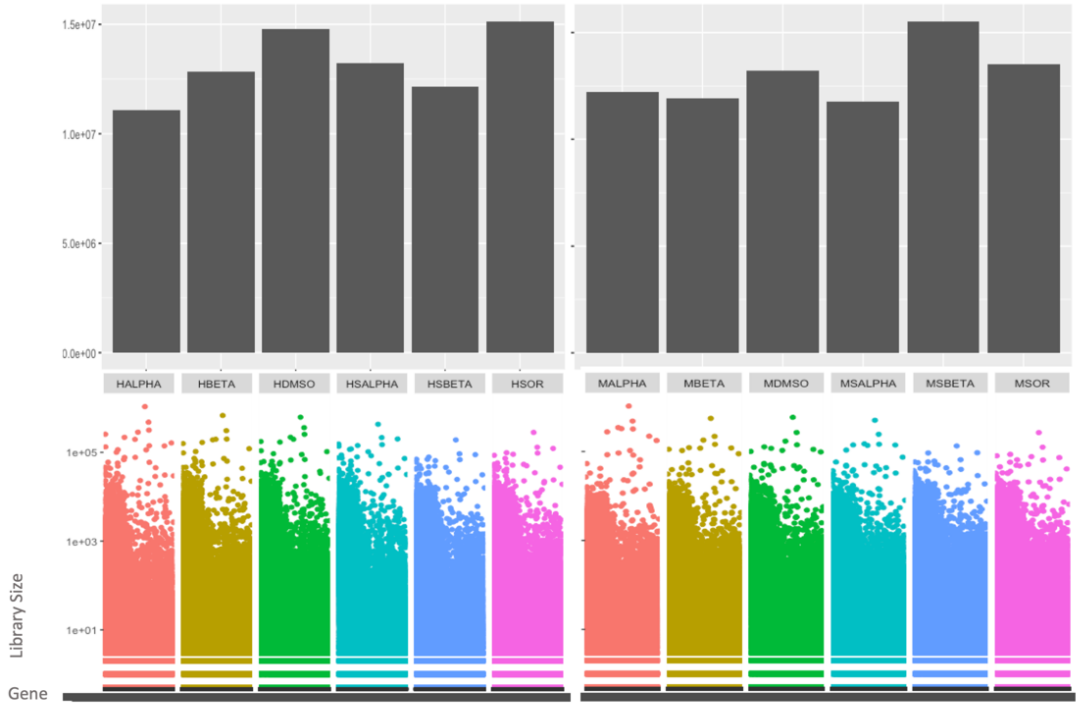


Figure 16: A. Library sizes per samples, B. gene distributions per samples resulted by HTSeq-count. There is no normalization in both plots. HALPHA; PI3Ki-α inhibitor (PIK-75), HSALPHA; PIK-75 and Sorafenib, HBETA; PI3Ki-β inhibitor (TGX-221), HSBETA; TGX-221 and Sorafenib, HSOR; Sorafenib treatments of Huh7 cells and MALPHA; PI3Ki-α inhibitor (PIK-75), MSALPHA; PIK-75 and Sorafenib, MBETA; PI3Ki-β inhibitor (TGX-221), MSBETA; TGX-221 and Sorafenib, H-MSOR; Sorafenib treatments of Mahlavu cells.

Mapped reads to the human transcriptome were counted by the HTSeq-count tool given a set of genomic features. Without any transformation, library sizes of genes were between 28000000-38000000 (Figure 16.A). The count of genes where well mapped were similarly distributed. Which indicates the fact that the sets were comparable to each other even though there was no replication. In Figure 16.B, genes are well distributed, and no distinguishable bias was observed. At the quantification step, no normalization was applied; the necessary normalization for differential gene expression analysis was provided inside the edgeR package.

3.1.3. Differential Expression Analysis

The systematic analysis of differential response of the HCC cell lines to the specified inhibitory elements were provided with RNA-seq analysis. RNA-seq experiment was performed for 5 Huh7 and 5 Mahlavu cell lines. EdgeR tool was used by taking DMSO treated cells as control to determine differentially expressed genes (DEGs). In order to estimate the dispersion among the cell lines, a set of housekeeping genes based on the assumption that their expression was not affected by the treatment with the inhibitors was used. Especially for Hepatocellular carcinoma, there is a comprehensive study exploring expression levels of tissue specific housekeeping genes (Ersahin et al., 2014). Using this set of genes listed in Table 6, the biological coefficient of variation (BCV) was estimated efficiently.

Table 6. Housekeeping genes used for BCV calculation.

Ensembl Gene ID	**Gene description**	**Gene name**
ENSG00000196470	siah E3 ubiquitin protein ligase 1	*SIAH1*
ENSG00000253729	protein kinase, DNA-activated, catalytic subunit	*PRKDC*
ENSG00000164924	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein zeta	*YWHAZ*
ENSG00000012048	BRCA1 DNA repair associated	*BRCA1*
ENSG00000128513	protection of telomeres 1	*POT1*
ENSG00000149269	p21 (RAC1) activated kinase 1	*PAK1*
ENSG00000135446	cyclin dependent kinase 4	*CDK4*
ENSG00000104290	frizzled class receptor 3	*FZD3*
ENSG00000161960	eukaryotic translation initiation factor 4A1	*EIF4A1*
ENSG00000120738	early growth response 1	*EGR1*
ENSG00000168539	cholinergic receptor muscarinic 1	*CHRM1*
ENSG00000149554	checkpoint kinase 1	*CHEK1*
ENSG00000207730	microRNA 200b	*MIR200B*

Raw results of edgeR resulted in between 900 to 11000 differentially expressed genes. When p value (0.01) filtration was applied 15 to 1500 of them remained to be significant. logFC using -/+2 criteria was applied for both cell lines to get expressive changes. However, the number of remaining genes for Mahlavu was significantly low using logFC 2. Considering many steps need to be performed for down-stream analysis A less stringent limitation for log was performed in Mahlavu cell line. On the other hand, Huh7 resulted in greater number discounting for single PI3Ki-β (TGX-221) treatment. Therefore, different logFC values on Huh7 and Mahlavu were applied. For Mahlavu cell lines, a less stringent logFC value (-1.5/+1.5) was used for further analysis.

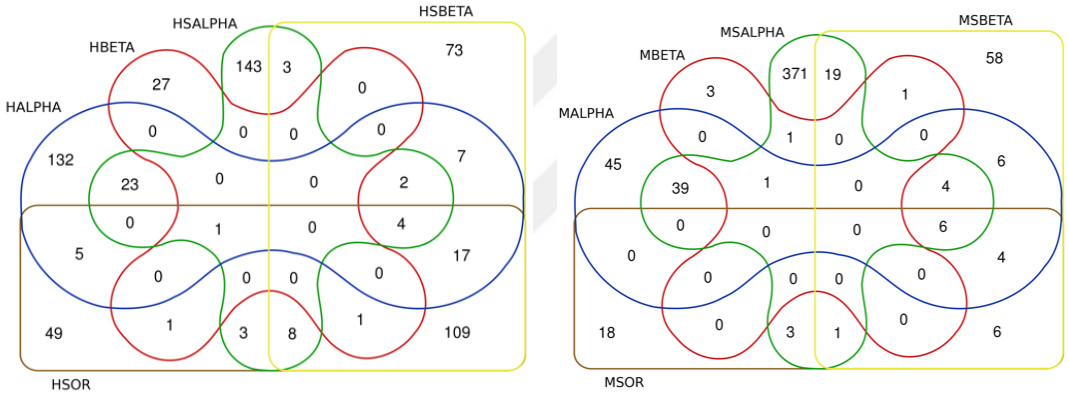
Table 7: The number of differentially expressed genes and untranslated transcripts for HCC cell lines under various inhibitor treatment conditions.

HCC treatments	**Differentially Expressed Genes**			
Huh7 vs DMSO	**Up**	**Down**	**Total**	**Untranslated Transcripts**
PI3K-α inhibitor (PIK-75)	139	52	191	20
PI3K-β inhibitor (TGX-221)	4	1	5	0
PIK-75 + Sorafenib	16	171	187	9
TGX-221 + Sorafenib	162	62	224	11
Sorafenib	127	71	198	11
Mahlavu vs DMSO	**Up**	**Down**	**Total**	**Untranslated Transcripts**
PI3K-α inhibitor (PIK-75)	66	39	105	22
PI3K-β inhibitor (TGX-221)	3	3	6	0
PIK-75 + Sorafenib	240	205	445	37
TGX-221 + Sorafenib	2	103	105	31
Sorafenib	17	21	38	12

Table 7 lists the numbers of DEGs. In both cell lines, as I expected, TGX-221 treatment had minor effects, and the DEG numbers were very low. In Huh7, the number of DEGs were greater for PI3Ki-α (PIK-75), whilst in PIK-75 + Sorafenib, DEG numbers were lower. The number of DEGs were also low for Sorafenib and TGX-221 alone

treatments, in their combinational treatment an increased number of DEGs, especially downregulated genes, were observed in Mahlavu cells.

The number of DEGs seen for PIK-75 treated Huh7 cells had more upregulated genes than PIK-75 + Sorafenib treatment and the number of downregulated genes for the combinatory treatment were greater than the single PIK-75 treatment. Therefore, a reversed reaction was observed for PIK-75 single and combinatory treatment with Sorafenib in the Huh7 cell line. Moreover, the response of Mahlavu to the same combination treatment was significantly different with a highly increased number of up and downregulated genes when compared to Sorafenib or PIK-75 inhibitor alone.



Huh7 Mahlavu

Figure 17: Venn diagrams representing common and unique Huh7 and Mahlavu DEG numbers. HALPHA; PI3Ki-α inhibitor (PIK-75), HSALPHA; PIK-75 and Sorafenib, HBETA; PI3Ki-β inhibitor (TGX-221), HSBETA; TGX-221 and Sorafenib, HSOR; Sorafenib treatments of Huh7 cells and MALPHA; PI3Ki-α inhibitor (PIK-75), MSALPHA; PIK-75 and Sorafenib, MBETA; PI3Ki-β inhibitor (TGX-221), MSBETA; TGX-221 and Sorafenib, H-MSOR; Sorafenib treatments of Mahlavu cells.

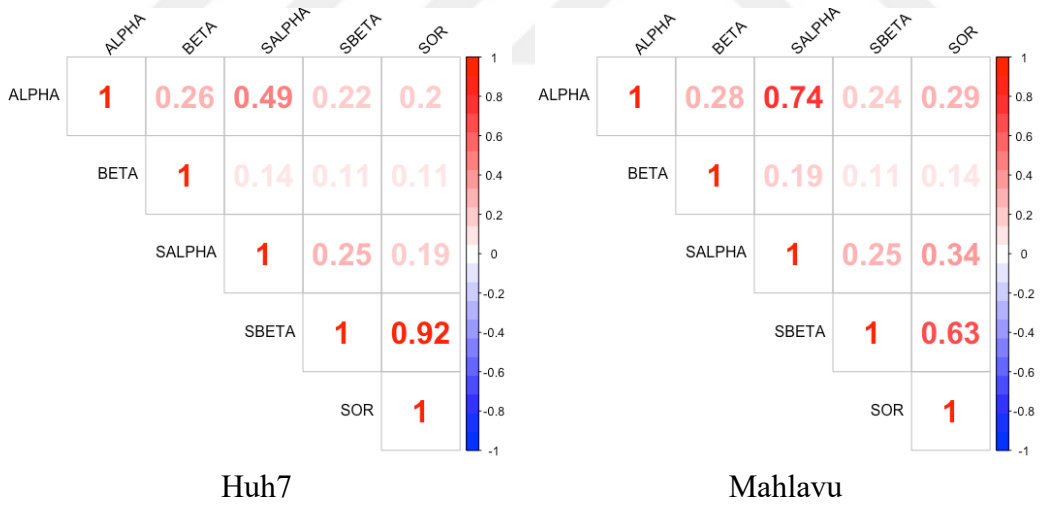
About 5% to 10% of the DEGs were found to be untranslated transcripts in Huh7 cells. PIK-75 treatment had the most untranslated transcripts in the DEG set. On the contrary, for PTEN-deficient HCC, Mahlavu cells, the number of untranslated transcripts were above average (30%). Especially, Sorafenib alone and TGX-221 + Sorafenib treatments had many untranslated transcripts. The number of untranslated transcripts is shown in Table 7.

Common and unique DEGs were identified using Venn diagrams and they were plotted separately for Huh7 and Mahlavu cell lines shown in Figure 17. Venn diagrams revealed that Sorafenib treatment and its combinational treatment with TGX-221 inhibitor shared a significant number of common genes in both cell lines. 70% (141 genes) in Huh7 and 45% (17 genes) in Mahlavu were shared between Sorafenib alone and TGX-221 + Sorafenib treatments. Furthermore, PIK-75 and Sorafenib

combination also had common genes in the Mahlavu cell line. 50% of the genes (50 of them) were common for those two treatments.

3.1.4. Correlation Analysis of Kinase Inhibitors

Pearson correlation analysis was conducted to see how comparable the differential expressions overall. logFC values of all significant (FDR and p-value > 0.01) differentially expressed genes (without logFC limitation) united for correlation analysis. As a result, a significant correlation was observed between Sorafenib treatment and its combinatorial treatment with TGX-221 (0.92) in Huh7 and (0.63) in Mahlavu, and between PIK-75 and its combinatorial treatment with Sorafenib in Mahlavu (0.74) (Figure 18). Negative correlation was not observed in any of the comparisons. Furthermore, a correlation analysis comparing Huh7 and Mahlavu cell lines together was conducted. As represented in Table 8, there was no correlation either in between two cell lines or in between the same drug treatments. This shows that PTEN-deficiency alters the molecular mechanism in HCC cell lines significantly and the correlation in DEGs are not notable.



Huh7 Mahlavu

Figure 18: Pearson correlations of kinase inhibitor treatments for Huh7 and Mahlavu. ALPHA; PI3Ki-α inhibitor (PIK-75), SALPHA; PIK-75 and Sorafenib, BETA; PI3Ki-β inhibitor (TGX-221), SBETA; TGX-221 and Sorafenib, SOR; Sorafenib treatments.

Table 8: Pearson correlation matrix between Huh7 and Mahlavu inhibitor treatments.

	Mahlavu				
Huh7	ALPHA	BETA	SOR	SALPHA	SBETA
ALPHA	0.02	0.04	0.02	0.27	0.07
BETA	0.14	0.08	0.02	0.03	0.04
SOR	0.26	0.15	0.18	0.22	0.18
SALPHA	0.02	0.10	0.14	0.14	0.12
SBETA	0.22	0.10	0.24	0.17	0.16

Abbreviations: ALPHA; PI3Ki-α inhibitor (PIK-75), SALPHA; PIK-75 and Sorafenib, BETA; PI3Ki-β inhibitor (TGX-221), SBETA; TGX-221 and Sorafenib, SOR; Sorafenib treatments.

3.1.5. Top 50 Common DEGs

The top 50 most differentially expressed genes are represented through a dendrogram in Figure 19. Top 50 genes were selected for each HCC cell line by ranking their sum of absolute logFC values per cell type.

Up and downregulated genes were clustered well in both cell limes. The separation was more absolute in Huh7. The most common downregulated genes in Huh7 were *DUSP5, PCNA, SPRY2*, and *WNK4* and upregulated genes were *VCAN, GADD45B, TBX4, HOXA2* and *DUSP8*. The details of the genes are listed in Table 9.

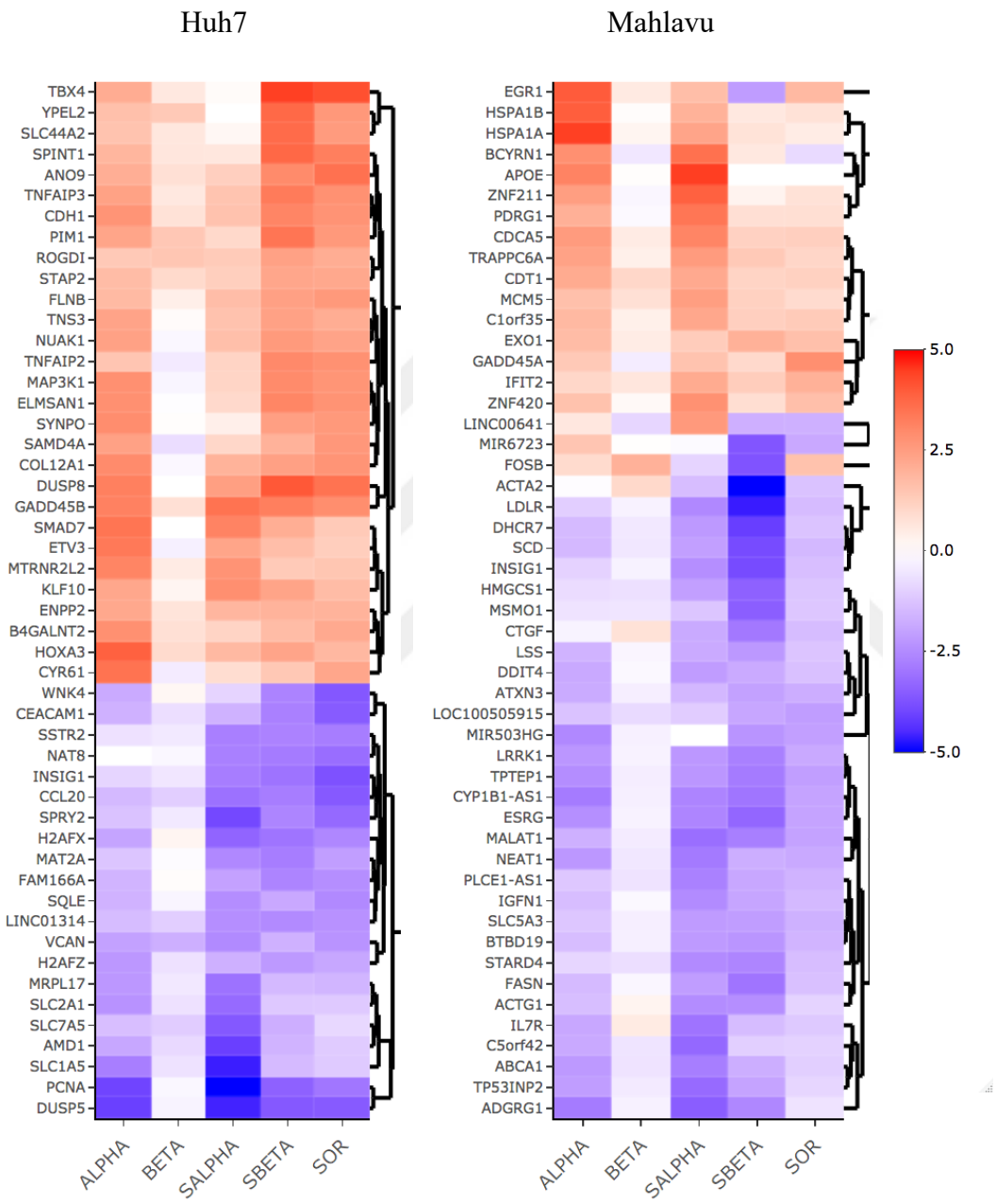


Figure 19: Heatmaply dendrograms representing top 50 most common DEGs of Huh7 and Mahlavu cell lines. Red signifies the upregulation while blue represents down. ALPHA; PI3Ki-α inhibitor (PIK-75), SALPHA; PIK-75 and Sorafenib, BETA; PI3Ki-β inhibitor (TGX-221), SBETA; TGX-221 and Sorafenib, SOR; Sorafenib treatments.

Table 9. Ensembl ID, gene description, gene name and the regulation type in Huh7 cell line of the most common differentially expressed genes.

Ensembl ID	**Description**	**Name**	**Regulation**
ENSG00000136158	sprouty RTK signaling antagonist 2	*SPRY2*	Down
ENSG00000132646	proliferating cell nuclear antigen	*PCNA*	Down
ENSG00000126562	WNK lysine deficient protein kinase 4	*WNK4*	Down
ENSG00000138166	dual specificity phosphatase 5	*DUSP5*	Down
ENSG00000121075	T-box transcription factor 4	*TBX4*	Up
ENSG00000099860	growth arrest and DNA damage inducible beta	*GADD45B*	Up
ENSG00000273793	dual specificity phosphatase 8	*DUSP8*	Up
ENSG00000105996	homeobox A2	*HOXA2*	Up
ENSG00000038427	versican	*VCAN*	Up

With respect to Huh7, common differentially expressed genes were not well separated in Mahlavu. Some of the genes like *EGR1, LINC00641, MIR6723, FOSB* and *ACTA2* genes were found to vary in different treatments. *HSPA1B, HSPA1A, APOE, ESGR, CYP1B1-AS1* and *TPTEP1* genes were the most upregulated genes and *LDLR, DHCR7* and *ADGRG1* were the most commonly downregulated genes, listed in Table 10.

When common differentially expressed genes in Huh7 and Mahlavu cells were compared, *INSIG1* gene was common to both cell lines and it was downregulated in most of the treatments. Moreover, while overexpression of *GADD45B* was mediated in Huh7, *GADD45A* was upregulated in Mahlavu cells.

Table 10. Ensembl ID, gene description, gene name and the regulation type in Mahlavu cell line of the most common differentially expressed genes.

Ensembl ID	**Description**	**Name**	**Regulation**
ENSG00000172893	7-dehydrocholesterol reductase	*DHCR7*	Down
ENSG00000205336	adhesion G protein-coupled receptor G1	*ADGRG1*	Down
ENSG00000130164	low density lipoprotein receptor	*LDLR*	Down
ENSG00000204389	heat shock protein family A (Hsp70) member 1A	*HSPA1A*	Up
ENSG00000204388	heat shock protein family A (Hsp70) member 1B	*HSPA1B*	Up
ENSG00000130203	apolipoprotein E	*APOE*	Up
ENSG00000100181	TPTE pseudogene 1	*TPTEP1*	Up
ENSG00000232973	CYP1B1 antisense RNA 1	*CYP1B1-AS1*	Up

3.1.6. Differentially Expressed Untranslated Transcripts

Following the identification of DEG lists, most of the down-stream analysis methods of RNA-seq analysis methods depend on the translated transcripts since they have protein annotations. After filtrations, 44 and 80 unique untranslated transcripts for Huh7 and Mahlavu, respectively, were identified. Apart from unidentified untranslated transcripts, most of them were either antisense RNAs or mitochondrial pseudogenes. In Figure 20, the dendrogram analysis of Huh7 untranslated transcripts is represented. Among the antisense RNAs, in Huh7, *TUBA1C* is found to be commonly downregulated gene. Also, *SLC1AS, DDIT4* and *HMCS1* were downregulated while *NCBP2* and *HOXA2* were upregulated in PIK-75 treatment group, *GABPB1* was downregulated but *RUSC1* was upregulated in PIK-75 and Sorafenib treatment, while *LYZ* was downregulated in Sorafenib treatments. Furthermore, mitochondrial pseudogenes were upregulated in PIK-75 treatment, while they were downregulated in Sorafenib and its PIK-75 combination.

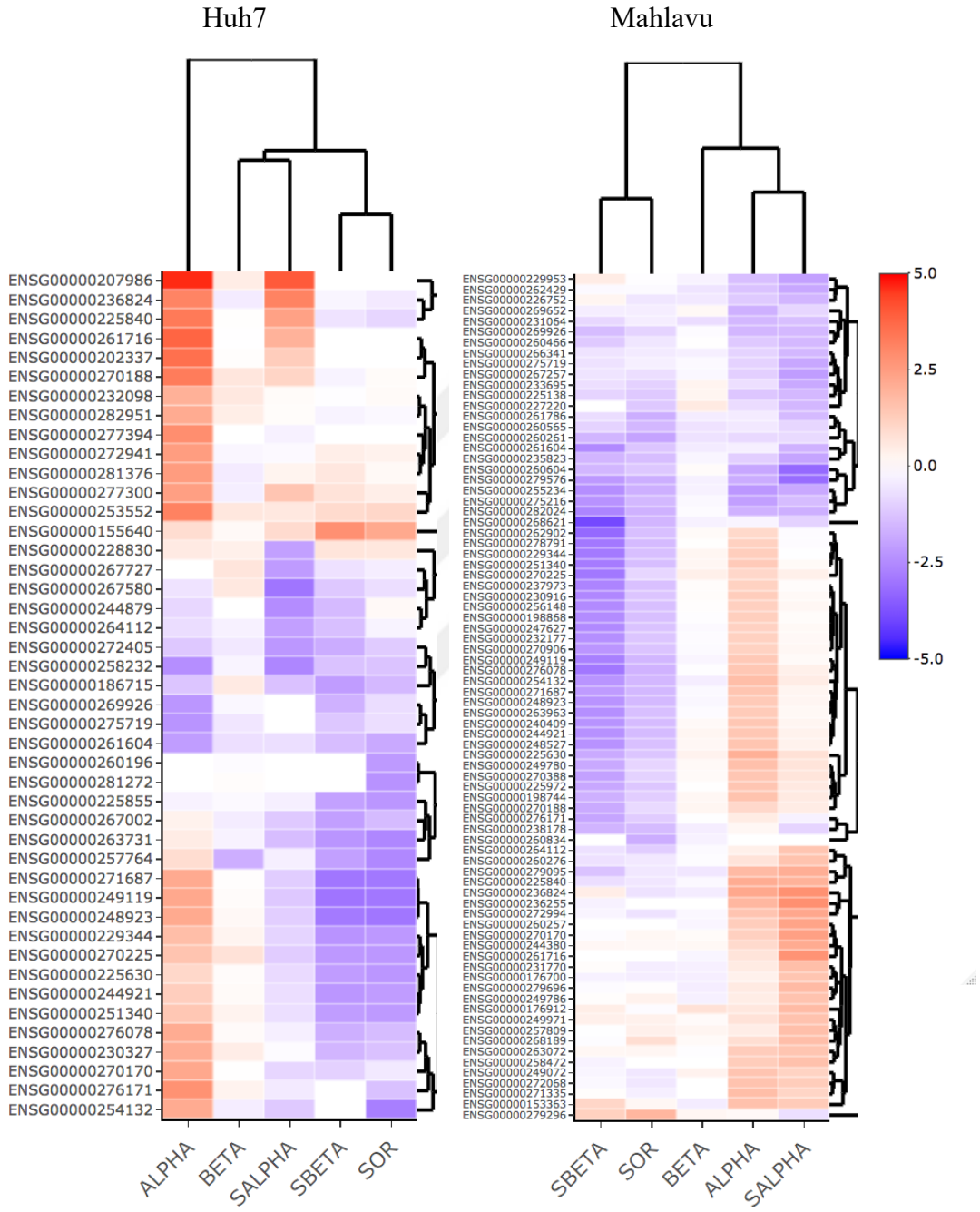


Figure 20: Heatmaply dendrograms representing untranslated transcripts of Huh7 and Mahlavu cell lines. Red signifies the upregulation while blue represents down. ALPHA; PI3Ki-α inhibitor (PIK-75), SALPHA; PIK-75 and Sorafenib, BETA; PI3Ki-β inhibitor (TGX-221), SBETA; TGX-221 and Sorafenib, SOR; Sorafenib treatments. Filtration criteria: p-value <0.01, logFC >2.0, <-2.0 for Huh7 and logFC >1.5, <-1.5 for Mahlavu cells.

Untranslated transcripts' dendrogram analysis for Mahlavu cells is shown in Figure 20. Among the antisense RNAs downregulation of *AXL, GMPR, DDIT4* and *THBS3* found in PIK-75 inhibitor treatment, *GAS6, HMGCS1, NFE2L1*, and *SLC9A3* downregulated and *ZNF213, EAF1, TMEM44, TYMS*, and *NCBP2* were upregulated in PIK-75 + Sorafenib. Moreover, mitochondrial pseudogenes were upregulated in PIK-75 treatment, conversely, they were downregulated in Sorafenib and TGX-221 combination. Furthermore, p53 regulation associated with the lncRNA was upregulated in this combination. programmed cell death 6 (*PDCD6*) was downregulated in Sorafenib treatment.

3.1.7. Gene Enrichment Analysis of Differential Expression Patterns

Since significant correlations between some of the specific treatments were observed, the shared patterns between sets were also explored in more detail. Huh7 and Mahlavu cell treatments were clustered using their logFC values to investigate expression patterns. Single PI3K-β inhibitor (TGX-221) analysis was excluded from both cell line analysis considering low number of DEGs. Clustering analysis was performed separately for the cell lines. Treatment-specific DEG sets joined for clustering analysis. United sets included 11033 and 11615 genes for Huh7 and Mahlavu respectively in total without any filtration. the genes that satisfied the specified limits kept: p-value <0.01, logFC >2.0, <-2.0 for Huh7 and logFC >1.5, <-1.5 for Mahlavu cells. So, dendrogram analysis was performed on 581 genes for Huh7 and 583 genes for Mahlavu cells.

In both HCC cell lines, the dendrogram was separated into two in the same way. Single PI3K-α inhibitor (PIK-75) treatment and Sorafenib with PIK-75 treatment were clustered together and single Sorafenib treatment and Sorafenib with TGX-221 were clustered together. In correlation analysis, the similarity between single Sorafenib and its combinatorial treatment in TGX-221 was greater in Huh7 and conversely single PIK-75 and its combinatorial treatment with Sorafenib was greater in Mahlavu.

Clustering was performed on the dendrograms to define common or unique expression patterns between the treatments. Huh7 and Mahlavu DEGs were divided into 8 and 6 clusters respectively. Gene ontologies belonging to those clusters were identified to characterize functional processes responding to different kinase inhibitors. Huh7 and Mahlavu dendrogram analysis are represented in Figure 21 and Figure 22 respectively.

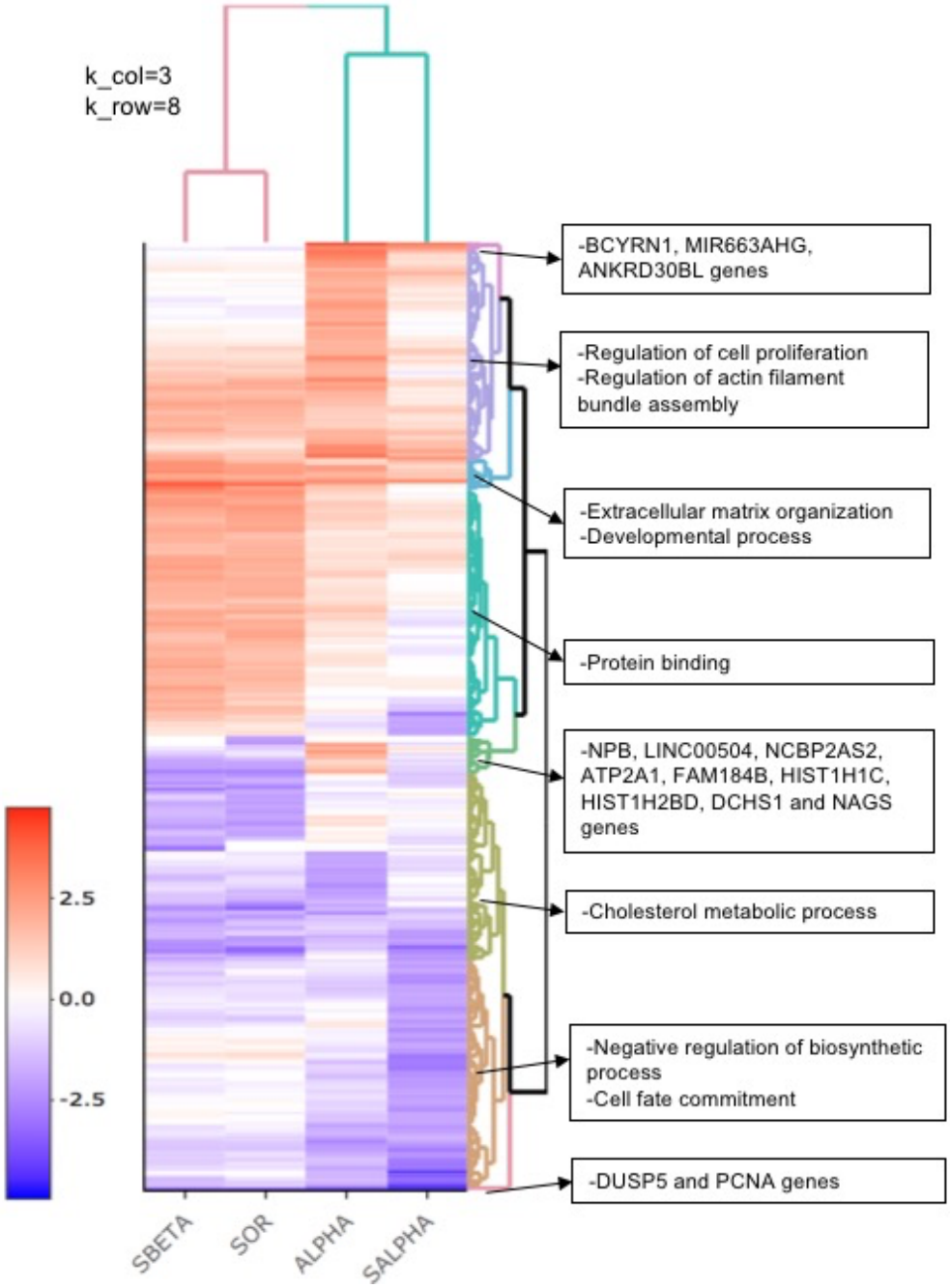


Figure 21: Heatmap of gene expressions illustrated as dendrograms for Huh7 cell line. Up and downregulated gene levels are colored as red and blue respectively, the intensity of the color indicates how strong the logFC value is. For more detailed analysis and to view interactive dendrograms please see html file in the GitHub repository. 8 clusters were generated by heatmaply and colored. Gene enrichment analysis was performed using BiNGO (FDR<0.05) and significant GOs selected according to the context. Clusters that do not show any significant enrichment were excluded. ALPHA; PI3Ki-α inhibitor (PIK-75), SALPHA; PIK-75 and Sorafenib, BETA; PI3Ki-β inhibitor (TGX-221), SBETA; TGX-221 and Sorafenib, SOR; Sorafenib treatments.

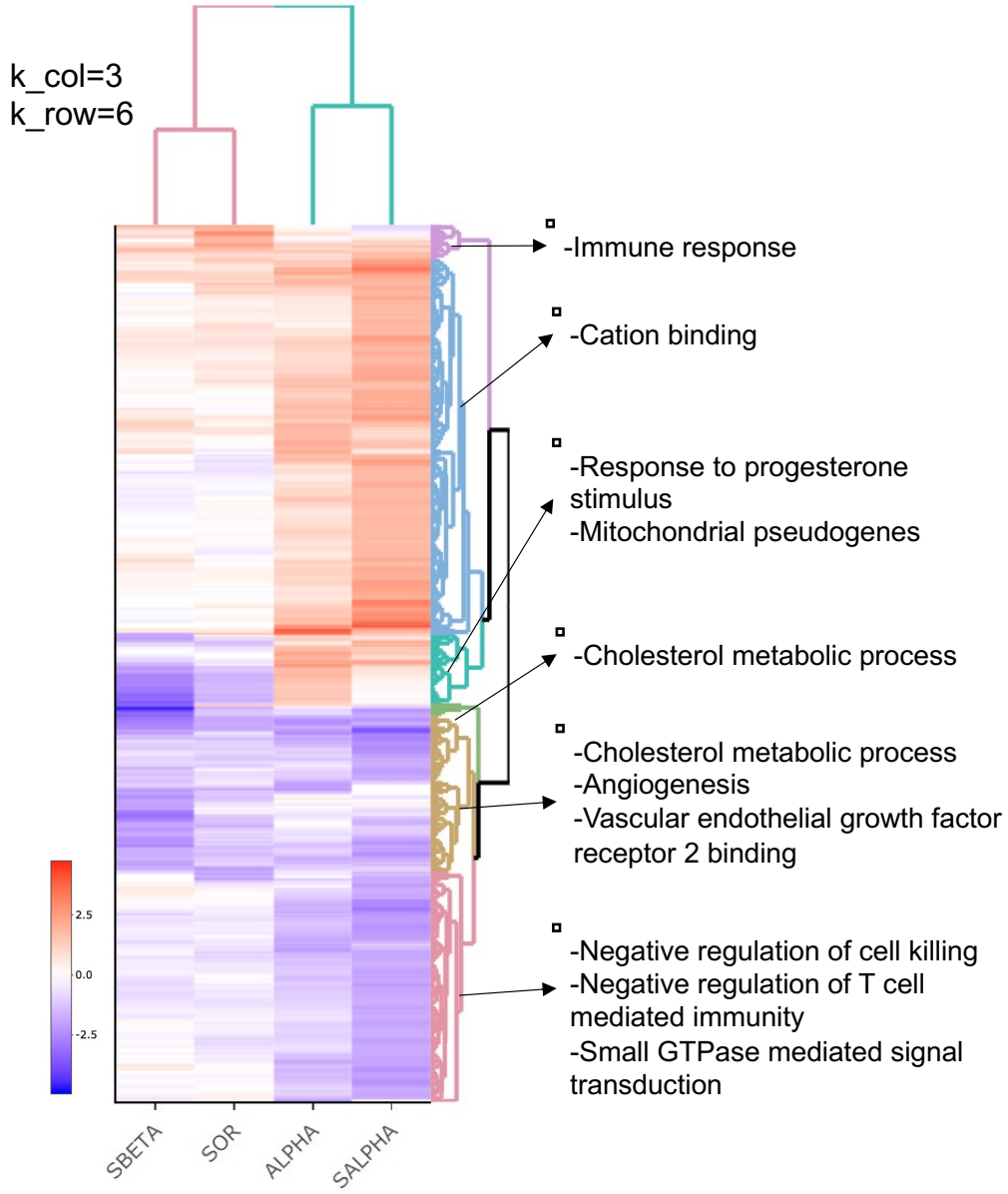


Figure 22: Heatmap of gene expressions illustrated as dendrograms for Mahlavu cell line. Up and downregulated gene levels are colored as red and blue respectively, the intensity of the color indicates how strong the logFC value is. For more detailed analysis and to view interactive dendrograms please see html file in the GitHub repository. 6 clusters were generated by heatmaply and colored. Gene enrichment analysis was performed using BiNGO (FDR<0.05) and significant GOs selected according to the context. Clusters that do not show any significant enrichment were excluded. ALPHA; PI3Ki-α inhibitor (PIK-75), SALPHA; PIK-75 and Sorafenib, BETA; PI3Ki-β inhibitor (TGX-221), SBETA; TGX-221 and Sorafenib, SOR; Sorafenib treatments.

Heatmap analysis of differentially expressed genes revealed functional expression patterns in HCC cells. For all treatments in HCC, positive regulation of extracellular matrix organization and developmental processes were observed while regulation of cell proliferation and actin filamentation bundle assembly ontologies were more active in PIK-75 treatment. *DUSP5* and *PCNA* genes were downregulated for all treatments. PIK-75 and Sorafenib combined treatment resulted in downregulation of genes enriched in negative regulation of biosynthetic processes and cell fate commitment ontologies. Likewise, cholesterol metabolic processes gene ontology was downregulated for TGX-221 + Sorafenib treatment. I also identified a group of genes including 2 histone family proteins, 1 long intergenic non-translating RNA, uncharacterized proteins *FAM184B* and *NCBP2AS2*, NBP, and *NAG5* and Ca+2 carrier receptor gene *ATP2A1* being downregulated in the treatment of PI3Ki-α alone while they were upregulated all the other Huh7.

Immune response was upregulated more significantly for Sorafenib treatment in Mahlavu treatments. Cation binding was upregulated for PIK-75 and its Sorafenib combination. Cholesterol metabolic processes, angiogenesis and vascular endothelial growth factor receptor 2 binding were downregulated for all treatments. A group of genes were upregulated in single Sorafenib and TGX-221 + Sorafenib treatment while downregulated in single PIK-75, PIK-75 + Sorafenib treatments. In this group, most of the genes were mitochondrial pseudogenes as they were also identified in dendrogram analysis of untranslated transcripts.

3.2. Network-Based Interpretation of the Data

A traditional way of RNA-seq analysis is to use only DEG sets for gene enrichment analysis which generally restricts the detection of some cellular events. The application of a conventional method, like Omics Integrator, to create a network from differentially expressed genes by connecting them through their known or physical protein-protein interactions can show hidden patterns. Omics Integrator adapts the Prize Collecting Steiner Tree (PCST) algorithm to connect differentially expressed genes by adding intermediate genes (or Steiner nodes) aiming the construction of the most optimal gene to gene network. Protein nodes in STRING human PPI was converted into Gene Names creating a reference gene network.

3.2.1. Optimal PCST Networks

Forest-tuner was run for the DEG list to find the best arrangements in these ranges; ω (1-10.0 or 5-15), β (1-15.0), μ (0.01-0.05). Forest-tuner lists the arrangements and the outcomes of the parameters running Forest. From the possible solutions, the parameters giving the maximum number of nodes with minimal network mean degree were selected. The selected and thus optimal ω, β and, μ parameters and their consequences on generated networks were listed in Table 11.

The number of prize nodes were the number of given proteins as input to the algorithm. Ensembl gene ids of DEGs were converted into Gene Names in this step matching to backbone reference network. Since some of the Ensembl gene ids did not correspond to any protein, as discussed before they are untranslated transcripts, not all input DEGs was used in Omics Integrator. The number of prize nodes and total number of nodes in the optimal networks contains both the input terminal proteins and the ones found to connect them (Steiner).

Table 11: Selected parameters for PSCT analysis using forest-tuner and numbers of nodes, terminals, and prizes of generated networks.

HCC treatments	**ω**	**β**	**μ**	**(Terminal+Steiner) Total Node #**	**Prize Node #**	**Mean degrees**
Huh7						
PI3K-α inhibitor (PIK-75)	7.75	5.50	0.04	(138+124) 262	171	24.83
PI3K-β inhibitor (TGX-221)	5.50	3.25	0.03	(5+12) 17	5	23.08
PIK-75 + Sorafenib	10.0	10.0	0.02	(145+101) 246	178	28.58
TGX-221 + Sorafenib	10.0	3.25	0.03	(178+147) 325	213	23.86
Sorafenib	10.0	7.75	0.05	(157+124) 281	187	27.19
Mahlavu						
PI3K-α inhibitor (PIK-75)	10.0	7.75	0.03	(52+63) 115	84	29.14
PI3K-β inhibitor (TGX-221)	7.75	5.50	0.03	(6+20) 26	6	16.7
PIK-75 + Sorafenib	10.0	3.25	0.01	(321+236) 547	409	30.31
TGX-221 + Sorafenib	10.0	5.50	0.03	(53+40) 93	75	34.65
Sorafenib	5.0	7.00	0.04	(16+15) 31	27	19.67

Since input DEG numbers for PI3K-β inhibitor (TGX-221) treated Huh7 and Mahlavu cells and Sorafenib treated Mahlavu cells were low, their networks were smaller than the others. The networks were optimized by limiting the number of trees to one and keeping their overall degrees as minimal as possible to avoid hairballs. That enabled

us to have more than one central hub node in the network and generated multi branches for the analysis.

3.2.2. Comparison of Optimal Network Nodes

Venn diagram comparing the nodes in the optimal networks is represented in Figure 23. 4 nodes (*DUSP5, PCNA, GADD45B* and *DUSP8* genes) were found to be shared by all networks excluding single PI3K-β inhibitor (TGX-221) treatment in Huh7. There were 10 nodes in Huh7 found to be common for Sorafenib including treatments (SALPHA, SBETA, and SOR). Gene enrichment of these proteins was associated with response to stimulus, damaged DNA binding and sterol metabolic process. *VCAN* gene was common for PI3Ki-α inhibitor (PIK-75) treatments (ALPHA and SALPHA) and single Sorafenib treatment (SOR). 25 common nodes were found for single PIK-75, TGX-221 + Sorafenib, and Sorafenib treatments (ALPHA, SBETA, and SOR). Gene enrichment analysis of these proteins showed that they were mainly acting on the regulation of cellular organization, migration, and cell cycle. Furthermore, treatment of PIK-75 + Sorafenib and TGX-211 + Sorafenib (SALPHA and SBETA) sharing 8 nodes enriched in N-acetyltransferase activity and Bcl3-Bcl10 complex. As shown previously through differentially expressed gene similarity analysis, combinational TGX-221 and single Sorafenib treatment shared 118 nodes enriched in cell junction control, signal transmission, regulation of MAPKKK, and interleukin-8 production.

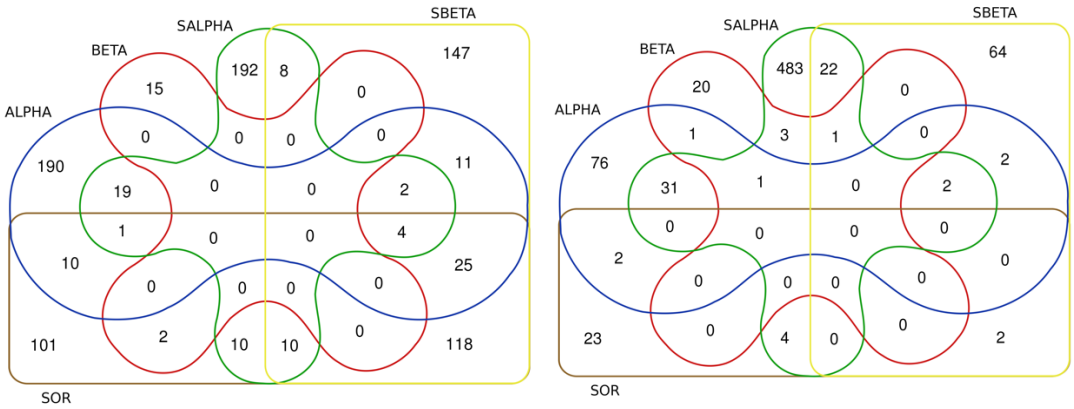


Figure 23: Venn diagrams representing the common and unique number of optimal network nodes of Huh7 and Mahlavu cells. ALPHA; PI3Ki-α inhibitor (PIK-75), SALPHA; PIK-75 and Sorafenib, BETA; PI3Ki-β inhibitor (TGX-221), SBETA; TGX-221 and Sorafenib, SOR; Sorafenib treatments.

The most common nodes were *INHBE*, *LRRK1, TP53INP2*, and *FOSB* genes in PTEN-deficient HCC Mahlavu treatments. Single PIK-75 and its combination with Sorafenib treatments (ALPHA and SALPHA) resulted in 31 common nodes enriched in GABA-

B receptor activity, Cdc42 protein signal transduction, regulation of inclusion body assembling and cholesterol efflux regulation. Moreover, PIK75 + Sorafenib and TGX-21 + Sorafenib (SALPHA and SBETA) shared 22 common proteins enriched in cholesterol metabolic process, oxidoreductase activity, and Bcl2-Bcl10 complex. *BCL3* gene was common for all combinatory treatment in HCC cell lines.

3.2.3. Cluster Specific Gene Ontologies of Optimal PCST Networks

A deeper understanding of the interacting genes and a better comparison of the networks were provided through a functional encoloring, sizing of the nodes, and a systematic usage of the network centrality which measures for clustering using Cytoscape tool. Optimal gene-to-gene networks predicted by PCST were imported into Cytoscape, gene's logFC values were attached and used to color the nodes to represent up and downregulated branches. The sizes of the nodes correlate with the betweenness centrality.

Systematic usage of networks was provided through the creation of a map for a practical comparison of different inhibitor treatments. For some of the networks, the greater number of nodes prevented effective comparison of the networks. Furthermore, using all nodes for gene ontology analysis would not be statistically significant because of large input sizes. Therefore, the big networks needed to be divided enabling a strong comparison strategy. To be able to investigate the networks more deeply, GO analysis applied into the network clusters. Finally, some significant gene ontologies for clusters were selected, and connected in the network using their gene associations. In Appendix E, whole list of Gene Ontologies found for each cluster were listed.

The optimal network generated using DEGs from PIK-75 treated Huh7 cell line is represented in Figure 24. Here, the nodes were colored by logFC values resulted from EdgeR analysis and color intensifies as the logFC value gets higher. Red and blue represented up and downregulation respectively. Steiner nodes were shaped as diamond while input transcriptome nodes were shown as ellipse. Node size was directly correlated with betweenness centrality of nodes. Clusters were generated using betweenness centralities of the nodes using a community cluster Glay algorithm and boxed for better representation. The clusters were separately explored by BINGO and only selected significant Gene Ontology was added to the network through associated genes.

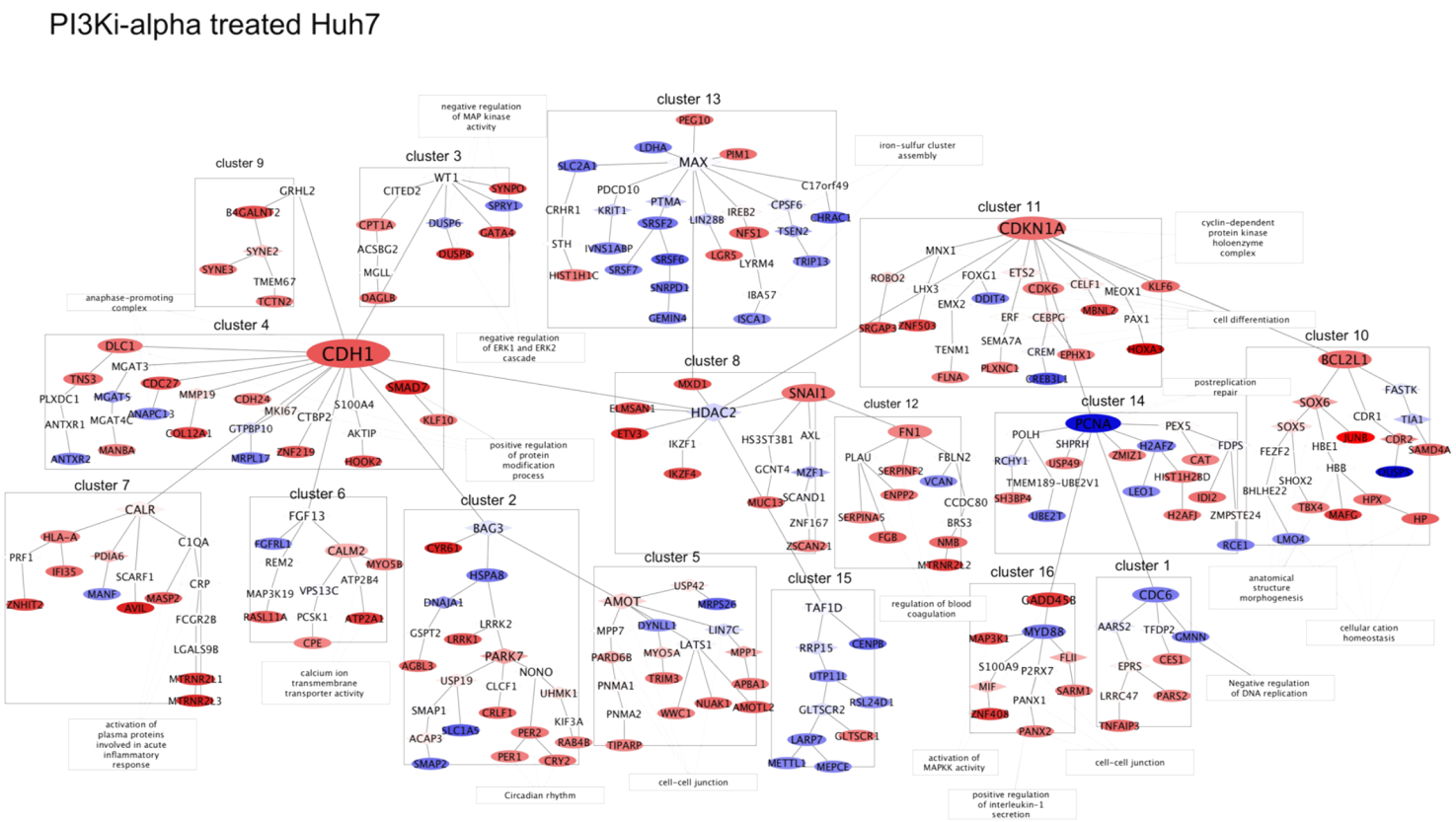


Figure 24: A schematic representation of an optimal network of DEGs upon PI3K-α inhibitor (PIK-75) treatment of Huh7 cell line. The network can be visualized through Cytoscape by using the .cys file in the GitHub repository.

CDH1 and *CDKN1A* genes in anaphase-promoting complex and cyclin dependent protein kinase holoenzyme were two central nodes of the PIK-75 network being upregulated. The other nodes connected to those have complex regulations. Importantly, negative regulation of Erk1 and Erk2, MAP kinase activity, DNA replication, cell differentiation and activation of plasma proteins involved in acute inflammatory response were found to be enriched in this network.

The network generated using DEGs from the TGX-221 treated Huh7 cell line is represented in Figure 25. Upregulation of cell redox homeostasis, cellular carbohydrate catabolic processes and oxidation-reduction activities were observed in that network.

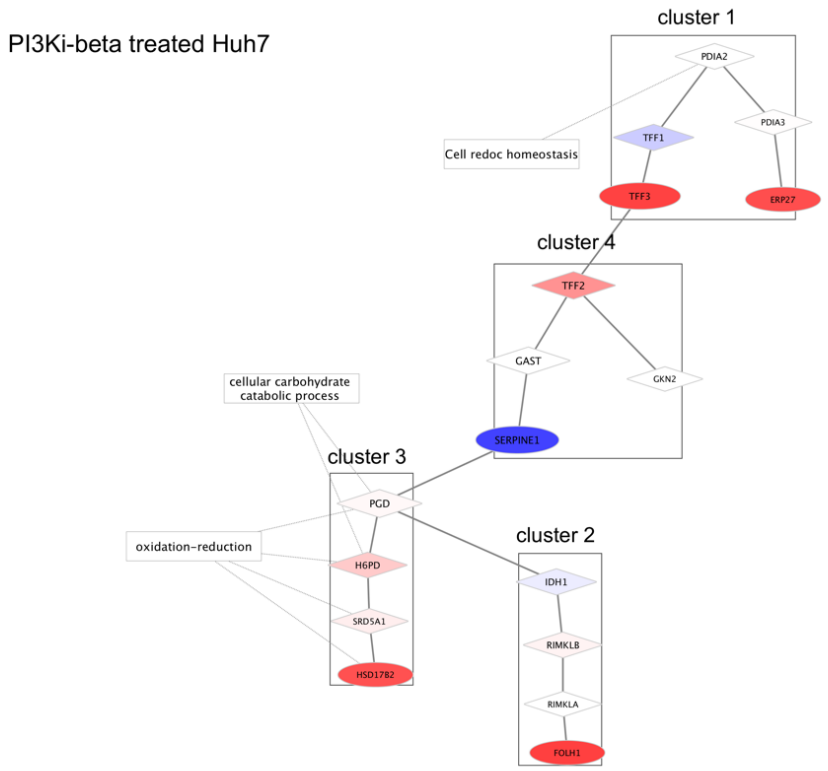


Figure 25: A schematic representation of an optimal network of DEGs upon PI3K-β inhibitor (PIK-75) treatment of Huh7 cell line. The network can be visualized through Cytoscape by using the .cys file in the GitHub repository.

The optimal network created from PI3K-α (PIK-75) + Sorafenib treated Huh7 cell line differentially expressed genes is represented in Figure 26. The network consists mainly of downregulated genes and the central gene in this network was downregulated JUN. *JUN* gene is a very prominent proto-oncogene encoding for a transcription factor. The

transcription factor is the main activator of ERK pathway which directly affects PI3K/AKT/mTOR pathway. The other two important downregulated genes were *MDM2* and *PCNA*, these two are related to regulation of microtubule depolymerization and cell proliferation respectively. Upregulation of *MTOR* gene would lead downregulation of several proteins related to the cellular stress response. Furthermore, negative regulation of MAP kinase activity and transmembrane receptor protein serine/threonine kinase signaling protein were observed which are mainly upregulated in HCC cancer.

TGX-221 + Sorafenib network is represented in Figure 27. The network was centered by the upregulated *CHD1* connected to upregulated *GAB2* and *SMAD7.* The other central node was slightly downregulated Steiner HDAC2 node. The network show some important gene enrichments in the clusters like positive regulation of cell proliferation, positive regulation of MAPKKK cascade, negative regulation of ERK1 and ERK2 cascade, positive regulation of JUN kinase activity, regulation of programmed cell death, apoptotic mitochondrial changes, inactivation of MAPK activity, negative regulation of fibroblast growth factor signaling pathway, negative regulation of nerve growth factor receptor signaling pathway, and Bcl3/NF-kappaB2 complex .

The optimal network generated through differentially expressed genes from Sorafenib treated Huh7 cells is shown in Figure 28. The network shares the same downregulated central gene, CDH1, with combinatory treated Sorafenib with TGX-221 represented in Figure 27. As opposed to its combinatory treatment with other isoform PIK-75, the network was mainly enriched with upregulated proteins. The main functional enrichments observed in the networks were positive regulation of cell-matrix adhesion, homeostatic process, negative regulation of ERK1 and ERK2 cascade, negative regulation of fibroblast growth factor receptor signaling pathway, cell differentiation, phosphate metabolic process, regulation of stress-activated MAPK cascade, GTPase activity, and oxidation reduction. Downregulated *PCNA* gene was hub node as like in combinatory treatment of Sorafenib with PIK-75. Single Sorafenib treatment shares some important patterns with its both treatments.

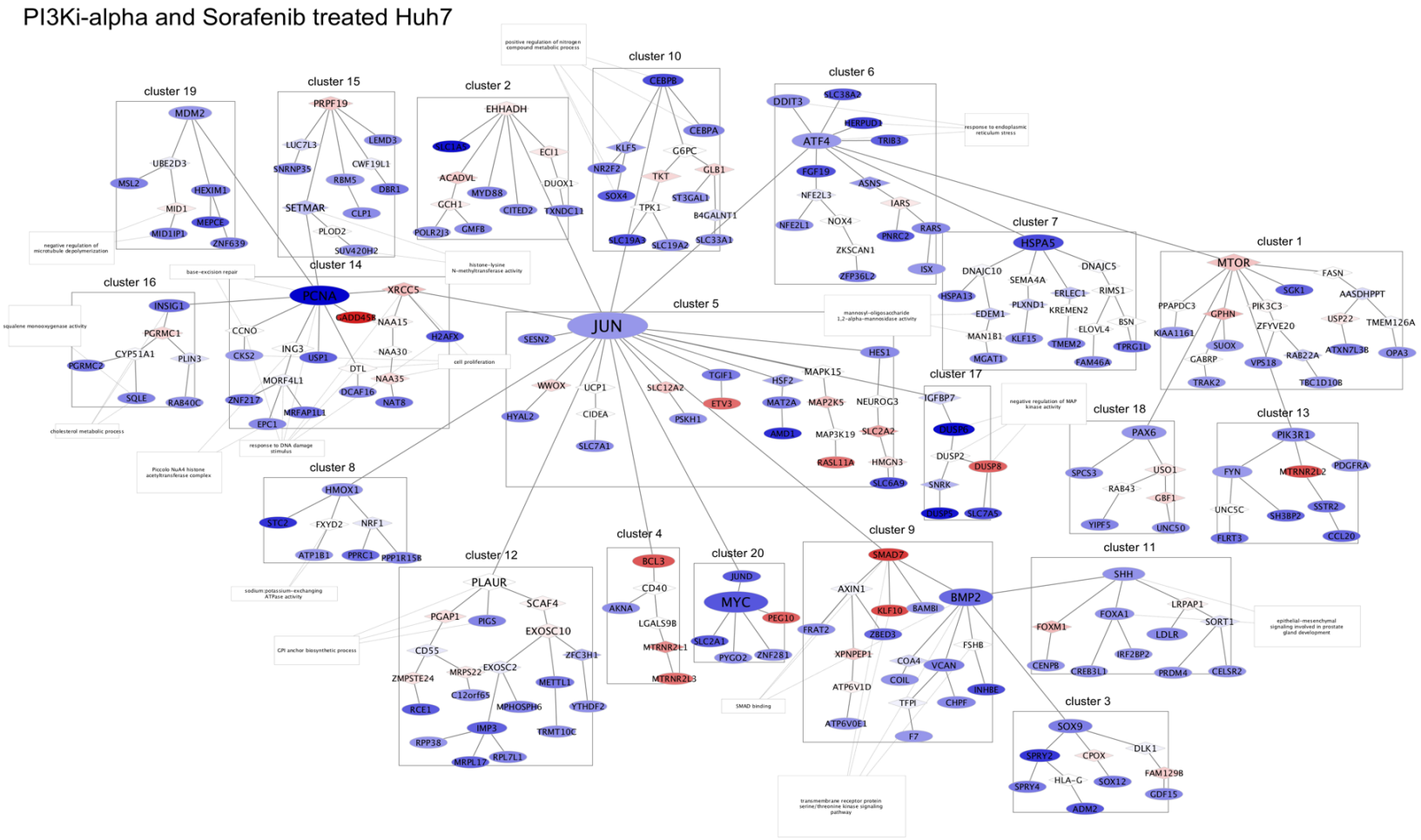


Figure 26: A schematic representation of an optimal network of DEGs upon PI3K-α inhibitor (PIK-75) with Sorafenib treatment of Huh7

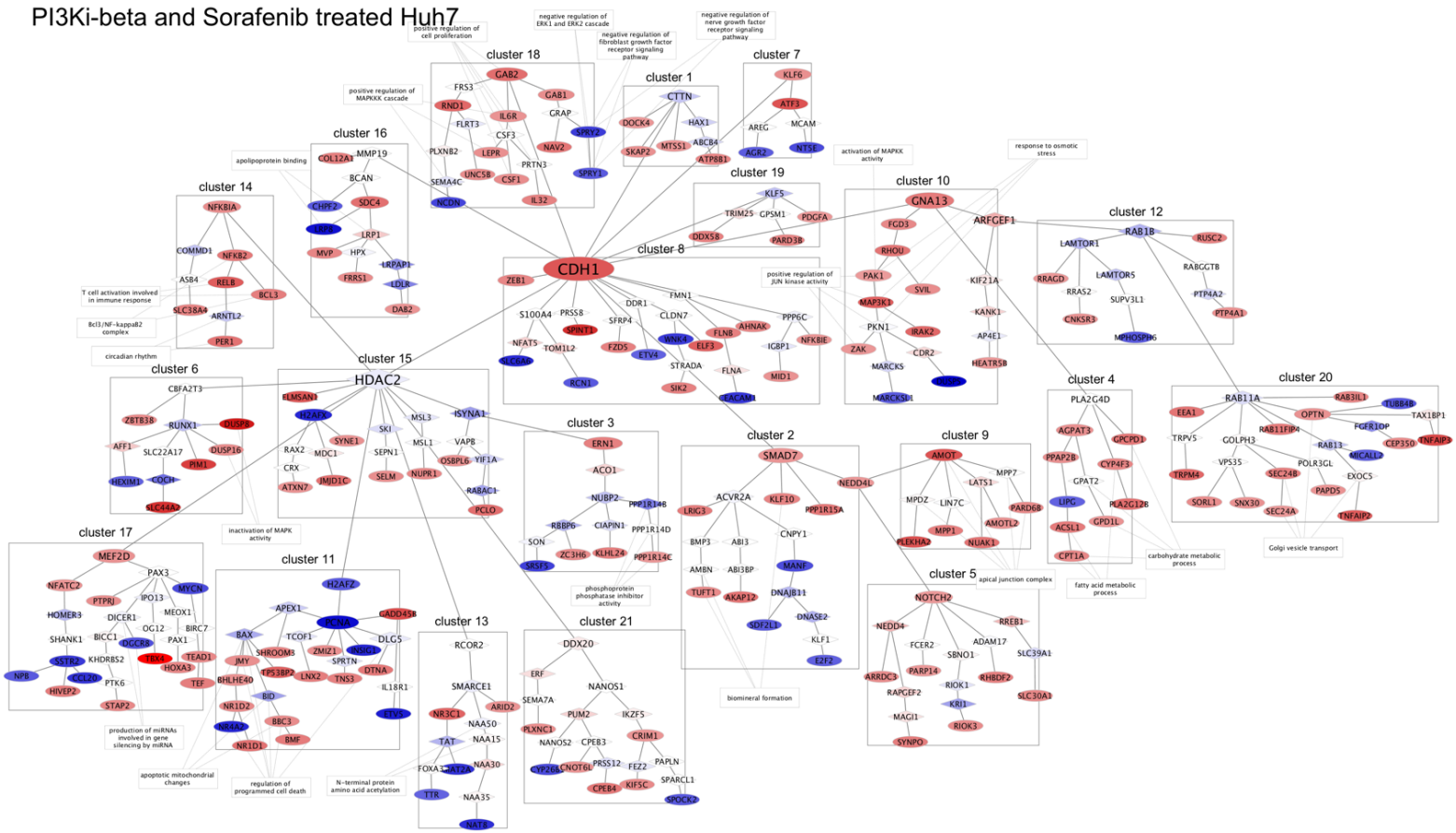


Figure 27: A schematic representation of an optimal network of DEGs upon PI3K-β inhibitor (PIK-75) with Sorafenib treatment

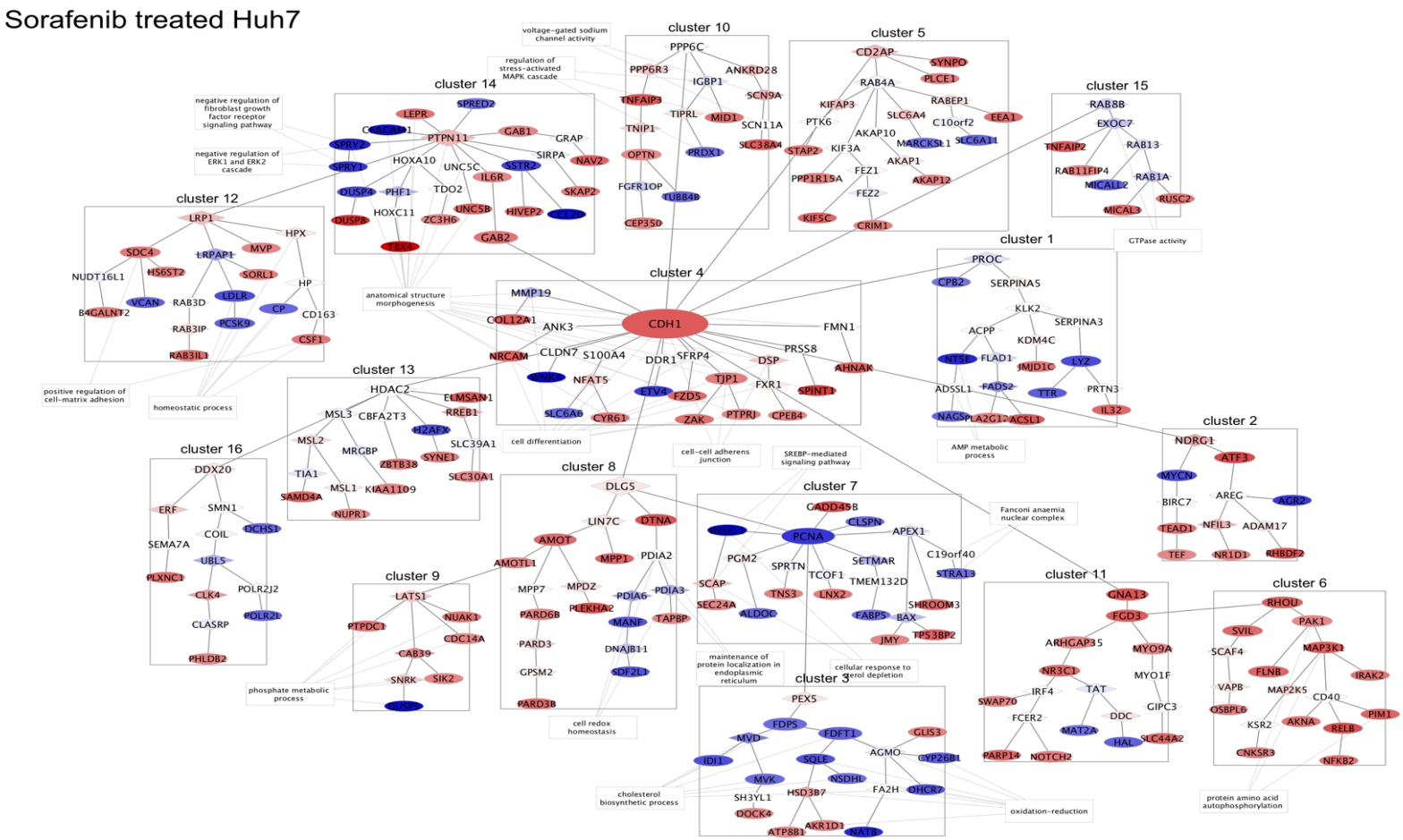


Figure 28: A schematic representation of an optimal network of DEGs upon Sorafenib treatment of Huh7 cell line. The network

The optimal network created from the DEGs extracted from PI3K-α inhibitor (PIK-75) treated Mahlavu cell line is represented in Figure 29. The central gene in the network was *CCND1*. *CCND1* gene expresses cyclin D1 protein which forms regulatory subunit of CDK4 or CDK6 whose activity was required for G1/S transition in the cell cycle. Upregulation of this protein also increases genes involved in growth arrest, such as *GADD45A*. The other important enriched functions in that network were negative regulation of cell proliferation, regulation of cell growth, negative regulation of body assembly, and positive regulation of cholesterol transport.

The optimal network generated from DEGSs from single TGX-221 treatment to Mahlavu cell is presented in Figure 30. As in Huh7 cells, the input DEG size lead to a smaller size network. All hub genes in this network were Steiner nodes. Some of the enriched pathways in this network were responses to molecules of bacterial origin, positive regulation of interleukin-6 production, pancreas development, and metallo-carboxypeptidase activity.

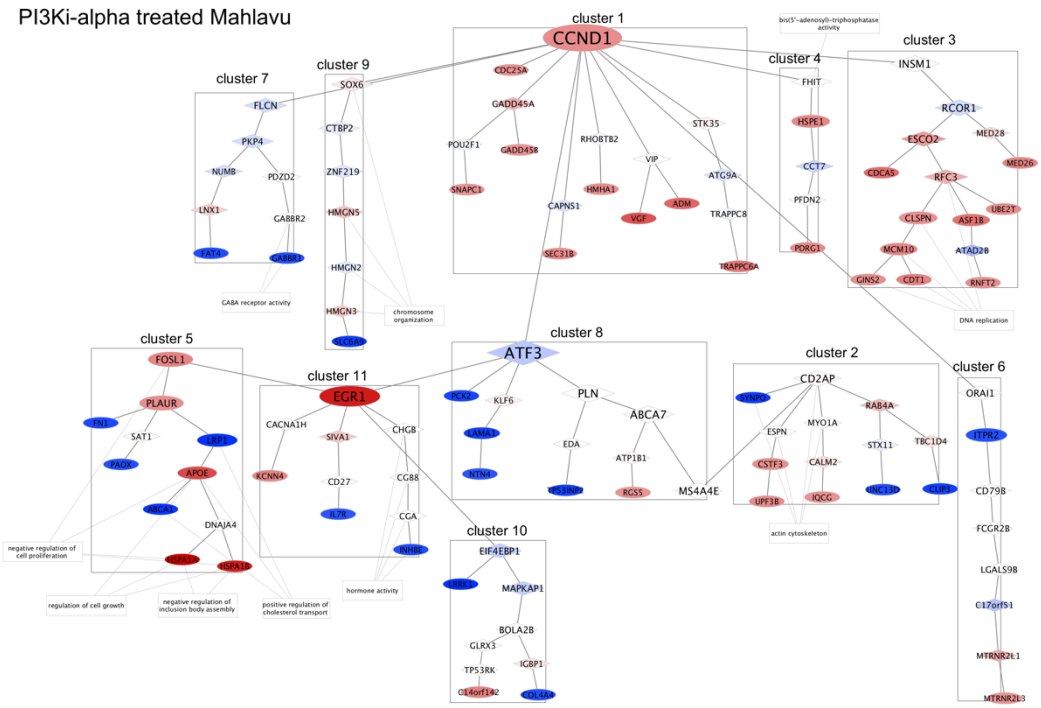


Figure 29: A schematic representation of an optimal network of DEGs upon PI3K-α inhibitor (PIK-75) treatment of Mahlavu cell line. The network can be visualized through Cytoscape by using the .cys file in the GitHub repository.

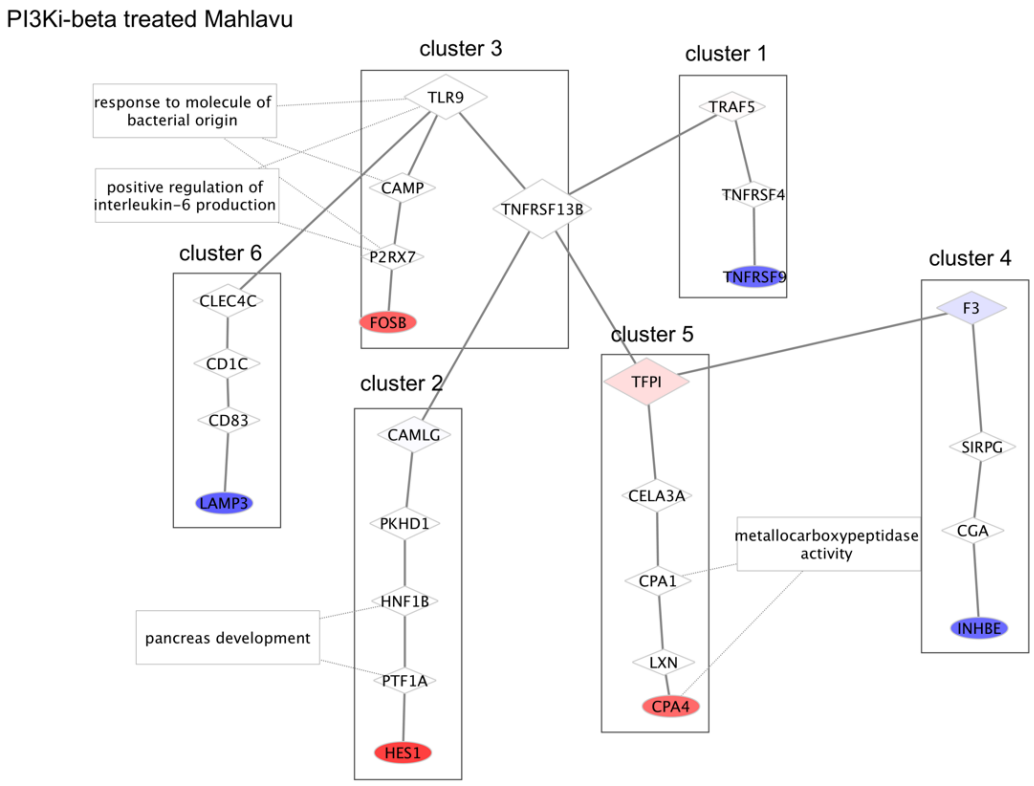


Figure 30: A schematic representation of an optimal network of DEGs upon PI3K-β inhibitor (PIK-75) treatment of Mahlavu cell line. The network can be visualized through Cytoscape by using the .cys file in the GitHub repository.

The network from the combinational treatment of PIK-75 + Sorafenib is shown in Figure 31. Similar to the Huh7 response shown in Figure 26, the main genes in this network were also downregulated. Downregulation of the *EGFR* gene was one of the important factors in this network. It was encoding a well-known protein associated with cancer. *TP53* gene was another central node in this network recognized as a Steiner in the network. Downregulation of this gene has many functions including DNA damage response, signal transduction resulting in induction of apoptosis and positive regulation of apoptosis. The other enriched functions in this network were positive regulation of transitional initiation, negative regulation of protein complex assembly, positive regulation of interferon-alpha production, NAD(P)H oxidase activity, generation of precursor metabolites and energy, regulation of GTPase mediated signal transduction, histone biotinylation, negative regulation of cholesterol storage, negative regulation of biosynthesis process, negative regulation of inclusion body assembly, cell-to-cell junction, mitotic cell cycle, and positive regulation of system process.

In Figure 32, the network generated from Mahlavu cells treated with TGX-221 + Sorafenib is shown. The hub gene in this network was upregulated MAPK1 gene

(Steiner node) but the many other genes in the network were downregulated. Especially, downregulation of the Bcl3/NF-kappaB2 complex was associated with cancer development. The other downregulation functions in this network were long-chain fatty acid transporter activity, SREBP-mediated signaling pathway, cholesterol biosynthetic process, nucleoside transport and oxidoreductase activity acting on the CH-CH group of donors, NAD or NADP as acceptor.

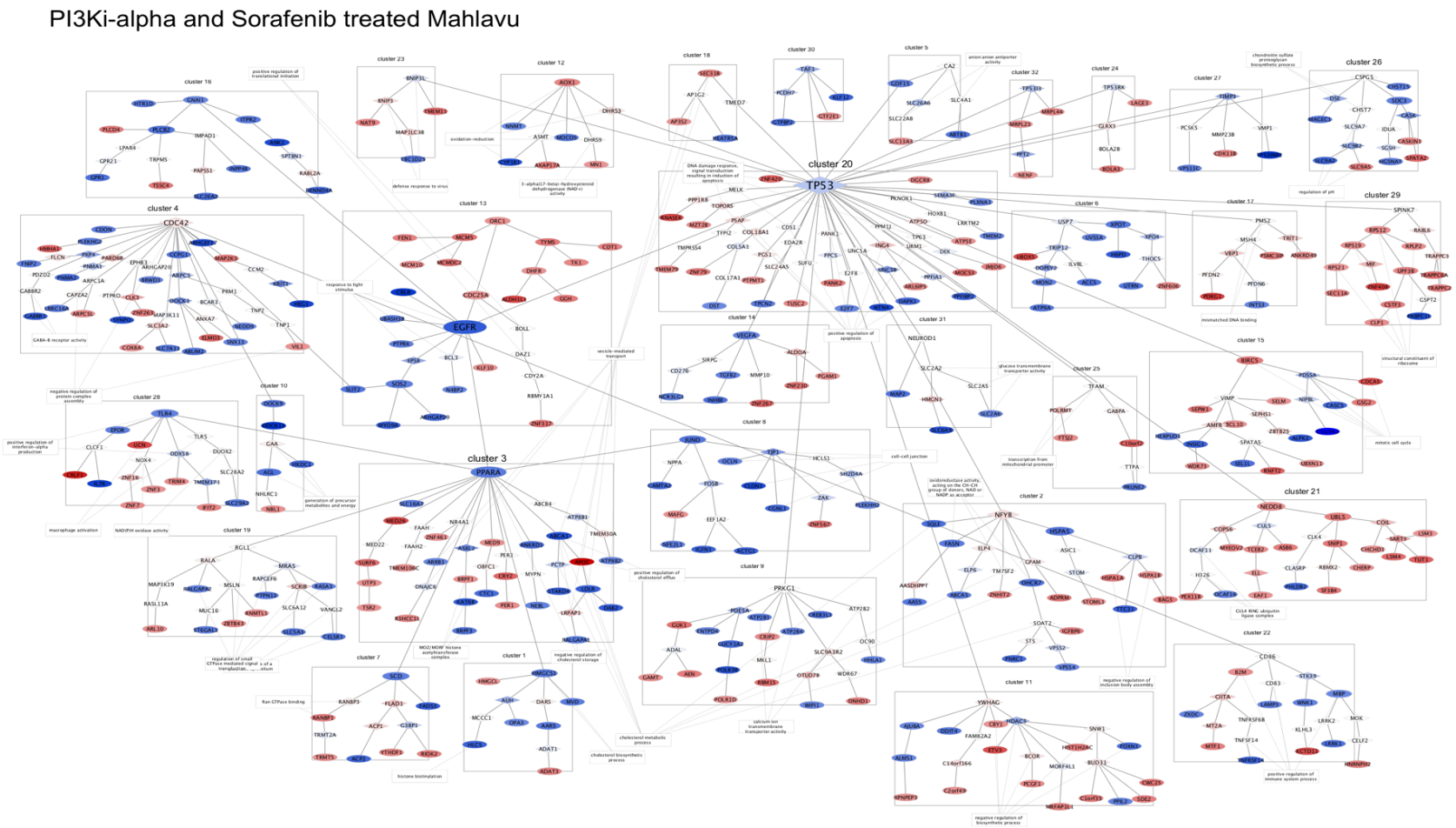


Figure 31: A schematic representation of an optimal network of DEGs upon PI3K-α inhibitor (PIK-75) with Sorafenib treatment of Mahlavu

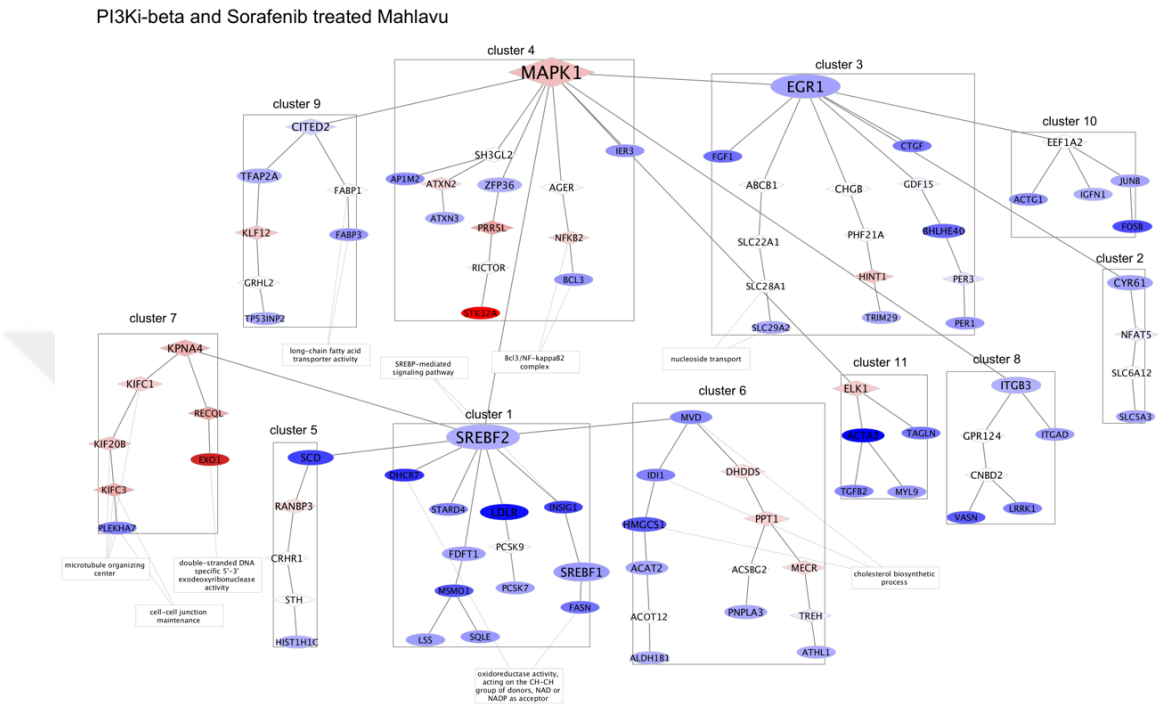


Figure 32: A schematic representation of an optimal network of DEGs upon PI3K-β inhibitor (PIK-75) with Sorafenib treatment of Mahlavu cell line. The network can be visualized though Cytoscape by using the .cys file in the GitHub repository.

The network constructed from the DEGs of Sorafenib treated Mahlavu cells is represented in Figure 33. As opposed to the Huh7, the same treatment affected minatory genes in the Mahlavu, and a smaller network was generated. Upregulation of the *MX1* gene with 10 other genes was observed in the network. The two clusters associated with that group were mainly related to interferon-induced dynamin-like GTPase with antiviral activity. Also, mitochondrial alpha-ketoglutarate dehydrogenase complex, isoleucine catabolic process, and negative regulation of protein kinase activity were among the enriched pathways in this network.

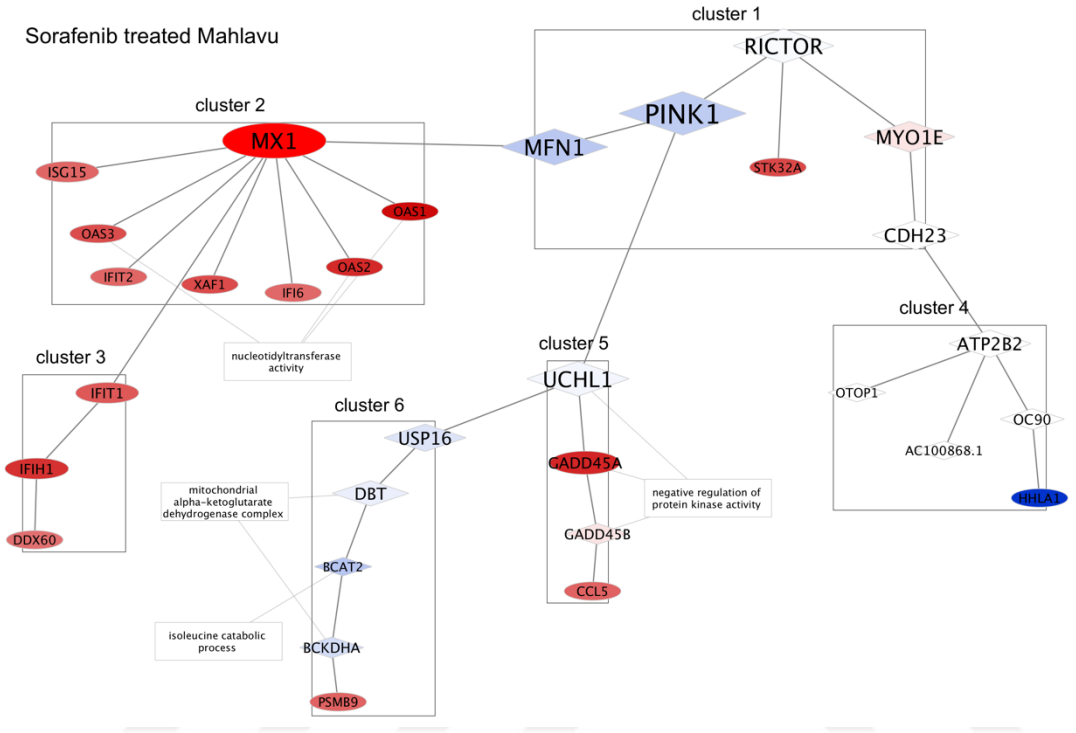


Figure 33. A schematic representation of an optimal network of DEGs upon Sorafenib treatment of Mahlavu cell line. The network can be visualized through Cytoscape by using the .cys file in the GitHub repository.

3.2.4. Comparison of Cluster Specific Gene Enrichments

The cluster-specific gene enrichments for each HCC cell line through a dendrogram analysis were compared in Figure 34. With respect to the similarities in DEG expressions, the functional processes in the optimal networks were not related pointedly. In both cell lines, single PI3Ki-β (TGX-221) and Sorafenib treatments were the closest in the dendrogram even though they do not share a significant number of processes. That was possibly due to the ineffectiveness of single TGX-221 treatment on the cells. Then, the closest treatments in terms of network functionalities were Sorafenib with PI3Ki-α (PIK-75) in Huh7 and TGX-221 in Mahlavu, yet in both the number of shared elements were very low.

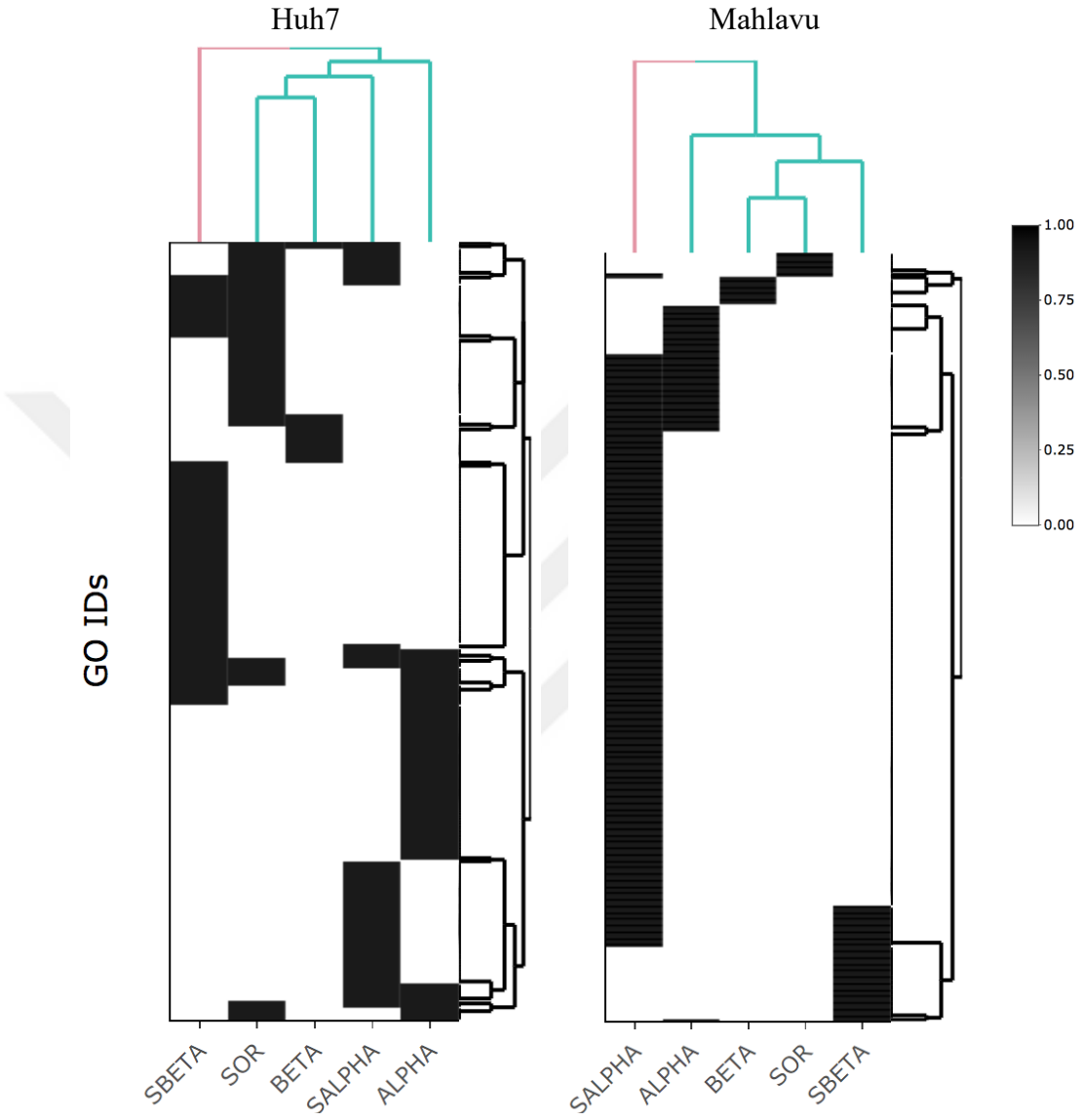


Figure 34: Heatmap of gene enrichments illustrated as dendrograms for Huh7 and Mahlavu cell lines. The dendrogram was plotted from a binary matrix; 1 for seen GO ids 0 for exclusion. For more detailed analysis and to view interactive dendrogram please see html files from the GitHub repository. ALPHA; PI3Ki-α inhibitor (PIK-75), SALPHA; PIK-75 and Sorafenib, BETA; PI3Ki-β inhibitor (TGX-221), SBETA; TGX-221 and Sorafenib, SOR; Sorafenib treatments.

All treatments except TGX-221 was enriched in MAP kinase activity regulation in Huh7. Lipid and steroid metabolic processes was observed for all treatments except PIK-75. Oxidation reduction was enriched for single TGX-221 and single Sorafenib treatments. In Sorafenib, SREB-mediated signaling pathway was also active. Combinational PIK-75 treatment represented regulation of DNA repair and SMAD binding gene ontologies with regulation of cell proliferation while combinational

TGX-221 resulted with Bcl3/NF-kappaB2 complex, regulations of apoptotic processes, cell morphogenesis and programmed cell death. Single PIK-75 and its Sorafenib combination represented an enrichment in development ontology (Table 12).

Table 12. Significant gene ontologies for inhibitor specific networks in Huh7 cell line.

Gene Ontologies	ALPHA	BETA	SALPHA	SBETA	SOR
Bcl3/NF-kappaB2 complex				■	
regulation of apoptotic process				■	
regulation of cell morphogenesis				■	
regulation of programmed cell death				■	
regulation of DNA repair			■		
SMAD binding			■		
anaphase-promoting complex	■				
SREBP-mediated signaling pathway					■
regulation of FGFR and NGFR signaling pathways				■	■
regulation of cell proliferation			■	■	
oxidation-reduction process		■			■
innate immune response	■			■	
cell differentiation	■			■	■
developmental process	■			■	■
negative regulation of ERK1 and ERK2 cascade	■			■	■
lipid metabolic process			■	■	■
steroid metabolic process		■	■		■
MAPK cascade	■		■	■	■

*ALPHA; PI3Ki-α inhibitor (PIK-75), SALPHA; PIK-75 and Sorafenib, BETA; PI3Ki-β inhibitor (TGX-221), SBETA; TGX-221 and Sorafenib, SOR; Sorafenib treatments.

Sorafenib represented enrichments in mitochondrial alpha-ketoglutarate dehydrogenase, PIK-75 inhibitor enriched in defense response, developmental processes and interleukin-6 production, and PIK-75 inhibitor enriched in negative regulation of cell proliferation and cell growth regulation in Mahlavu cell lines. Combinatory treatment of PIK-75 and its single treatment commonly were enriched in cholesterol transport and inclusion body assembly. The network for combinatory PIK-75 treatment was very large thus there were many GO hits mostly in immune processes and cell to cell junction organization. Only combinational PIK-75 enriched in defense response, generation of precursor metabolites and energy, mitotic cell cycle, oxidation-reduction process, regulations of apoptotic processes, cell death, and small GTPase mediated signal transduction (Table 13).

Table 13. Significant gene ontologies for inhibitor specific networks in Mahlavu cell line.

Gene Ontologies	ALPHA	BETA	SALPHA	SBETA	SOR
negative regulation of cell proliferation	■				
regulation of cell growth	■				
response to bacterium		■			
defense response			■		
generation of precursor metabolites and energy			■		
mitotic cell cycle			■		
oxidation-reduction process			■		
positive regulation of apoptotic process			■		
positive regulation of cell death			■		
regulation of small GTPase mediated signal transduction			■		
Bcl3/NF-kappaB2 complex				■	
SREBP-mediated signaling pathway				■	
mitochondrial alpha-ketoglutarate dehydrogenase complex					■
Cdc42 protein signal transduction	■		■		
positive regulation of interleukin-6 production		■	■		
apical junction complex			■	■	
cholesterol biosynthetic process	■		■	■	

*ALPHA; PI3Ki-α inhibitor (PIK-75), SALPHA; PIK-75 and Sorafenib, BETA; PI3Ki-β inhibitor (TGX-221), SBETA; TGX-221 and Sorafenib, SOR; Sorafenib treatments.

3.2.5. Prioritizied Genes as Drug Targets

The selection of drug targets using the randomization tests and network topology was well explained in Methods chapter. 20 gene were selected for further analysis in each cell line. Figure 35 represents the selected genes sorted with betweenness centrality of optimal networks in HCC lines. For Huh7 cells; *CDC27, CCDC80, AARS2, ACSBG2, CITED27* and *CDR2* genes in PI3Ki-α (PIK-75) inhibitor treatment, *RIMKLA* gene in PI3Ki-β (TGX-221) treatment, *CEBPB*, *DNAJC10, DLK1, ATP6V1D, EDEM1* and *DUSP8* genes in PIK-75 + Sorafenib treatment, *LIN7C* gene in TGX-221 + Sorafenib treatment, *EXOC7, FEZ1, GAB2, BIRC7, HOXA10* and *ANKRD28* genes Sorafenib inhibitor treatment. For Mahlavu cells; *ATP1B1, CACNA1H, CAPNS1, CCT7, ATG9A* and *BOLA2B* genes in PIK-75 inhibitor treatments, *CGA* and *TNFRSF4* genes in TGX-221 inhibitor treatment, *ALMS1, AOX1, BCL3, ANKRD1, CD276* and *ASIC1* genes in PIK-75 + Sorafenib treatment, *GDF15, AGER, FABP1, ACOT12, HMGCS1* and *CRHR1* genes TGX-221 + Sorafenib treatment were prioritized for further investigations.

Figure 35. Prioritized nodes for Huh7 and Mahlavu were ranked by betweenness centrality values of randomized networks for each inhibitor treatment. ALPHA; PI3Ki-α inhibitor (PIK-75), SALPHA; PIK-75 and Sorafenib, BETA; PI3Ki-β inhibitor (TGX-221), SBETA; TGX-221 and Sorafenib, SOR; Sorafenib treatments.

In Figure 36, the selected nodes are represented with their expression pattern for each treatment in Huh7 and Mahlavu cell lines. Both the intensity and the regulation pattern (up or down regulation) was depending on the kind of the treatment. *LIK7C* gene, for example, was selected through TGX-221 + Sorafenib treatment (Huh7), was upregulated, but the same gene was downregulated in PIK-75, TGX-221, and PIK-75 + Sorafenib treatments. *CCT7* and *CAPNS1* were two downregulated genes selected from Mahlavu cell line. They were selected through single PIK-75 treatment, but an

opposite action acquired with combination of PIK-75 with Sorafenib. Therefore, the prioritization of those genes should be more detailly observed.

Furthermore, nearly half of the prioritized genes were Steiner nodes like *FEZ1, ACSBG2, RIMKLA, BIRC7, HOXA10* and *CCDC80* in Huh7 and *FABP1, CRHR1, AGER, ACOT12, ASIC1, BOLA2B, CACNA1H, CGA,* and *TNFRSF4* in Mahlavu in the optimal networks. They were lost or hidden in the RNA-seq experiment, but they were prioritized only through Omics Integrator based network optimization and network-topology based significancy analysis.

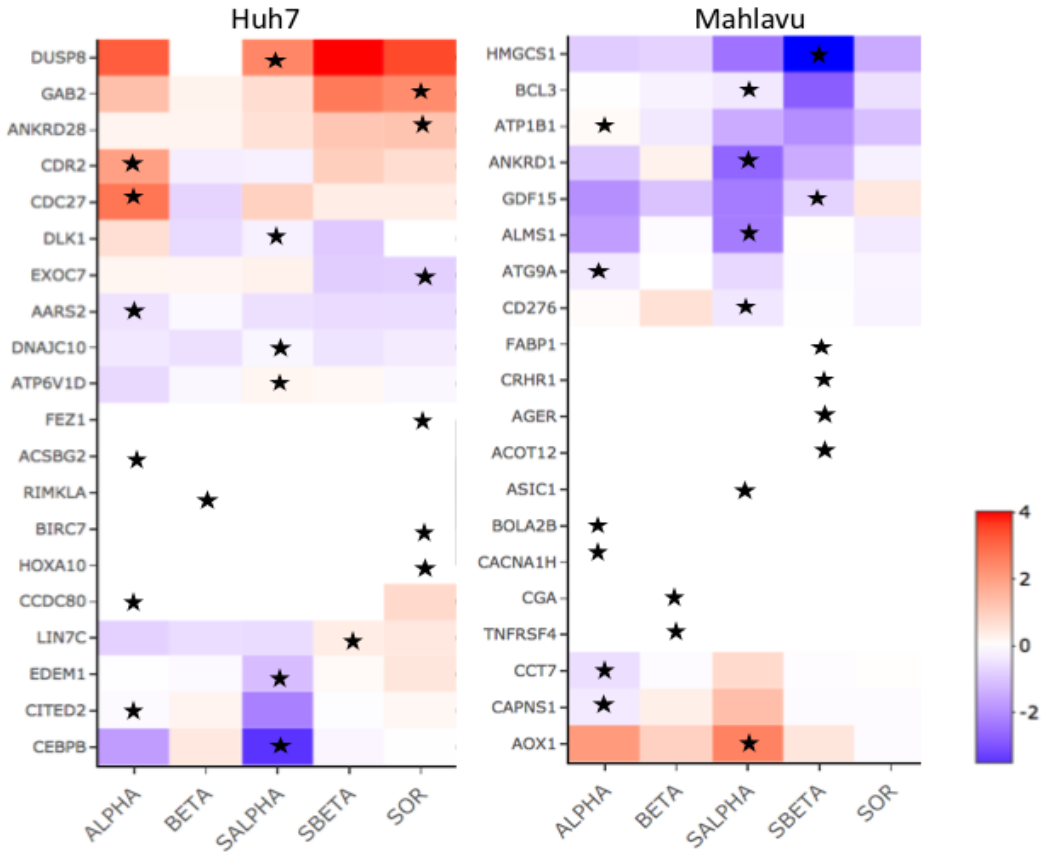


Figure 36: Up (red) and down (blue) regulations of the prioritized genes in Huh7 and Mahlavu cell lines, from which treatment they are selected was pointed out. ALPHA; PI3Ki-α inhibitor (PIK-75), SALPHA; PIK-75 and Sorafenib, BETA; PI3Ki-β inhibitor (TGX-221), SBETA; TGX-221 and Sorafenib, SOR; Sorafenib treatments.

3.2.6. The Effect of Gene Removals from the Optimal Networks

Simulating an *in-silico* knock-out experiment, the effects of node removals from the optimal networks were analyzed. In theory, the most important targets would be affecting maximum number of PPIs.

Each prioritized node was analyzed in the associated pathway. Furthermore, if the node exists in the other treatment with the same cell line it was also considered. Since some of the selected nodes were in the close to the branches, they have affected one or two nodes. Those were considered un-significant unless they have not associated into an important gene ontology. The deletion of the target protein and it's all connections, affecting nodes and functions of the deleted functions by the removals were studied and summarized in Table 14.

Table 14: Prioritized genes as potential drug targets in Huh7 and Mahlavu HCC lines.

Gene	**Treatment**	**Cell line**	**Function**	**Effect on the network**
AARS2	PIK-75	Huh7	Alanyl-TRNA Synthetase 2, Mitochondrial	5 nodes
CITED2	PIK-75	Huh7	Cbp/P300 Interacting Transactivator with Glu/Asp Rich Carboxy-Terminal Domain 2	5 nodes
DLK1	PIK-75 + Sorafenib	Huh7	Delta Like Non-Canonical Notch Ligand 1	3 nodes
DNAJC10	PIK-75+ Sorafenib	Huh7	DnaJ Heat Shock Protein Family (Hsp40) Member C10	5 nodes
GAB2	TGX-221 + Sorafenib	Huh7	GRB2 Associated Binding Protein 2	Positive regulation of cell proliferation
GAB2	Sorafenib	Huh7	GRB2 Associated Binding Protein 2	Erk1/Erk2 kinase pathway
BOLA2B	PIK-75	Mahlavu	BolA Family Member 2B	6 nodes
AOX1	PIK-75 + Sorafenib	Mahlavu	Aldehyde Oxidase 1	Positive regulation of oxidation reduction
AGER	TGX-221 + Sorafenib	Mahlavu	Advanced Glycosylation End-Product Specific Receptor	Apoptosis

Ultimately, in Huh7 cells removal of *AARS2* and *CITED2* genes in PIK-75 inhibitor treatment, *DLK1* and *DNAJC10* genes in PIK-75 + Sorafenib treatment, *HOXA10* gene in Sorafenib treatment and *GAB2* gene in both single Sorafenib and TGX-221 + Sorafenib treatments were found to be affecting at least 5 nodes in the networks and destructing key pathways. In Mahlavu, *BOLA2B* gene in PIK-75 inhibitor single treatment, *AOX1* in PIK-75 + Sorafenib treatment and *AGER* gene in TGX-221 + Sorafenib treatment in Mahlavu were found to be affecting important pathways, or at least 6 nodes.

CHAPTER VI

DISCUSSION

Hepatocellular carcinoma (HCC) has limited targeted treatment options such as multi-kinase inhibitor Sorafenib and recently approved drug Regorafenib (Perz et al., 2006, Liu et al., 2009; Llovet, Di Bisceglie, et al., 2008). Yet, none of the drugs can increase the survival of the patients by more than 10 months. Up to date, there are many studies in search of novel targets to cure HCC better. One of the reasons why those treatment strategies were not operative is the redundant functions of signaling pathways controlling proliferation, cell cycle, migration, angiogenesis, or apoptosis in precancerous cells during the chronic liver disease stage. Even multi-kinase inhibitors, e.g. Sorafenib, targeting Raf, VEGFR, and PDGFR proteins, cannot effectively prevent tumorigenic cross-talks between the signaling pathways. Therefore, the need for combinational therapies to evade multi-functioning pathways for HCC is massive and urgent.

The classical way to discover new therapeutic agents is to analyze differential expression patterns comparing cancerous versus normal-like or healthy cells. Following this research, abnormal functions or signaling pathways and deviant gene expressions are evaluated experimentally. Wet laboratory experiments focusing on singular a gene or a pathway is long, labor-intensive, and expensive. Yet, it is a well-known fact that cancer develops in a multi-stage process by enrolling many different protein interactions. It would also be costly and time consuming if enough comparisons were made in order to solve this multidimensional mechanism of each cancer type. Hence, new computational methods are immediately required encompassing many *in-vitro* analysis. Currently, using the sequencing platforms, and advanced bioinformatic tools and good practice pipelines to analyze those results would provide novel therapeutic targets. Nevertheless, it is pivotal to construct right workflows (sequence of analysis) enabling reproducible, scalable, and optimal analysis.

In this doctoral thesis, two Hepatocellular carcinoma (HCC) cell lines, Huh and Mahlavu, were treated with 3 kinds of drugs; Sorafenib as multi-kinase inhibitor, PIK-75 as PI3K-α isoform inhibitor, and TGX-221 as PI3K-β isoform inhibitor. As it was known that none of the drug treatments was able to eliminate cancerous cells alone, the cells treated with combinations of the drugs; PIK-75 with Sorafenib and TGX-221 with Sorafenib were also analyzed. A classical way of comparison of those treatments

would be acquiring differentially expressed gene lists and making functional enrichment analysis. Yet, through this analysis, it is possible to lose single genes or dwindle some functions since the RNA-seq experiment cannot cover all the gene expressions. In this study, instead of performing traditional analysis, biological network analysis using Omics Integrator was utilized. Omics Integrator connects the given nodes, namely differentially expressed genes, by adding other nodes from human protein-protein interaction network if needed. Thereby, the hidden patterns of the expression which may be absent in RNA-Seq analysis was unrevealed. There are other pathway or network analysis tools which could be used for DEG analysis but none of them could perform a simultaneous contraction of the treatment specific network. For example, using KEGG database, the DEG set could be mapped to PI3K/Akt/mTOR pathway yet the redundancy through other signaling pathways will be lost. STRING database can also construct DEG specific networks, but it won't consider the connecting paths or nodes between the given nodes.

Next step of the multi-level target analysis is to make powerfully comparisons of the networks. One challenge of associating biological networks is their complexity. Large DEG input sizes for Omics Integrator drives many protein-protein interactions leading huge size of output networks. Hence, direct node-to-node comparison of the networks becomes impossible. In order to compare networks, performing biological enrichment to the networks to relate the functions would be a better approach, but for the large networks it is also problematic. Biological enrichment analysis of large networks will result in either too general or too specific Gene Ontology (GO) terms with disrupted significance. In this study, the problem was solved by topologically clustering the networks which divides the nodes (genes) according to their connections in the network. Each cluster size kept the same for large networks and enrichment analysis was performed to the separate clusters. The clusters were identified using betweenness centrality measure, namely keeping the closest in interaction proteins together.

The real comparison of the different treatments to the cell lines was provided through Cytoscape visualizations. The optimal networks were imported into the Cytoscape tool, differential expression patterns of the genes were used to color the nodes, centrality measures were used to resize the nodes, node clusters were boxed, and functional enrichments were mapped back to the nodes. Herein, easy-to-observe treatment network were created not only to visualize all elements of the networks and but also compare single nodes with their expression pattern and biological function.

Differential expression analysis of the treatments was made using EdgeR. To perform a statistically strong analysis, EdgeR requires an estimation of dispersion for each cell line. Yet, in this analysis set, there were no biological replicates to calculate it. That was the major challenge of this study since all the downstream analysis basing DEG sets. In EdgeR documentation, the default BVC is 0,16, 0,4 was a precalculated value for biological replicates in human samples and 0.1 was calculated for technical replicates of a model organism. In our analysis, we decided to use housekeeping genes in HCC based on the assumption that the expression level of housekeeping genes does not vary by drug treatments. BVC was 0,045 in our analysis yet that value could be

too conservative. Alternatively, a number of randomly selected set of genes together with housekeeping genes could be used for BVC calculation for EdgeR DEG analysis. The resulting DEG sets could be compared with the current DEG set, hence the validity of our initial approach could be supported by this method.

According to our DEG results, PTEN deficient Mahlavu cell lines responded to inhibitors less severe than Huh7 since Mahlavu is a more aggressive type of HCC than Huh7. PI3K-β (TGX-221) inhibitor treatment resulted in a smaller number of DEGs for both cell lines. That was correlated with previous results performed in CANSYL (Figure 8 and 9). Singular TGX-221 treatment had minor effect in HCC treatment. All single drug treatments exhibited a lower number of DEGs in Mahlavu, representing the need for combinational treatment for PTEN-deficient cell lines. The ineffectiveness of TGX-221 in both cell lines and Sorafenib in Mahlavu were also observed through Pearson correlations and dendrogram analysis. The correlations were vital in combinational TGX-221 with Sorafenib and combinational PIK-75 with singular PIK-75 treatments.

A significant number of untranslated transcripts in DEG sets especially for Mahlavu treated with single Sorafenib or single TGX-221 were observed. Many of those untranslated transcripts were antisense RNA or mitochondrial pseudogenes. Mitochondrial dysfunctions were also observed in GO analysis using DEGs and network analysis in Sorafenib treatments. There are many researches showing the correlation of long untranslated RNAs to the diseases in the literature (Weilin et al., 2013; Lijuan Zhang et al., 2019b), yet deeper research is needed to uncover the relations of them in HCC or with drug treatments. Interestingly, mitochondrial pseudogenes were representing different patterns in PIK-75 single and combinational treatments in Huh7. Sorafenib must be acting into the pathways having a role in the cellular energy regulations. Another exciting finding was a lncRNA in p53 regulation was upregulated in Sorafenib with TGX-221 and untranslated transcript programmed cell death 6 (PDCD6) was downregulated in Sorafenib treatment.

The most common DEGs like *PCNA, DUSP5, VCAN, SPRY2, EGR1, FOSB, ACTA2, HSPA1A, APOE, LDLR,* and *GADD45B* were the most significant and most studied genes in cancer. A simple PubMed search of these genes (correlated with cancer in the last 10 years) resulted in 100 full text research articles. Also, those common genes show the similar expression regulation patterns in the other cancer cell lines. Hence the expression of the characteristics of the cell lines and conserved patterns with respect to the inhibitor treatments could be considered as a part of cancer hallmark.

In order to identify shared functional enrichments from the different treatments, heatmap analysis were used. Significant cell or treatment specific patterns were observed through the enrichment analysis with common DEG patterns. For example, positive regulation of extracellular matrix organization and development processes were observed for Huh7 cells while cell proliferation regulations were more actively regulated though PIK-75 treatments. Furthermore, negative regulation of cell fate

commitment was specific only to PIK-75 with Sorafenib treatment. Hence, PIK-75 + Sorafenib activated death related processes in Huh7.

Metabolic processes, angiogenesis and VEGFR2 binding were downregulated for all treatments indicating a decline in cancer development in Mahlavu cells. The immune response against the inhibitor treatments was stronger with single Sorafenib treatment. A group of genes was upregulated in single Sorafenib and combinatory TGX-221 and Sorafenib treatment while downregulated in single PIK-75, combinatory PIK-75 and Sorafenib treatments. Since this group was especially downregulated in TGX-221 + Sorafenib treatment, they could be related to their antagonistic effect on cell death.

Until network analysis part, all the findings were discovered using the traditional way of RNA-seq analysis. In order to compare differentially regulated signaling pathways inhibited through different kinds of drug targets, and to find novel targets for new treatment strategies, a systems level of understanding of differentially expressed genes were examined through optimal networks.

The network node comparisons indicated that most of the common genes were retained in Huh7 treatments. An increase in the number shared nodes for Sorafenib treatments was observed, and unsurprisingly most of those genes were active in lipid metabolism. The similarity between the combinational PIK-75 and PIK-75 treatments were retained while some of shared genes of combinational TGX-221 and Sorafenib treatments were lost in Mahlavu. The comparison of network specific gene ontologies also contributed the difference of combinational TGX-221 and Sorafenib treatments. Sorafenib treatment only enriched in mitochondrial alpha-ketoglutarate dehydrogenase complex in Mahlavu, same as Sorafenib treatment in Huh7 which also shows enrichment in oxidation-reduction processes. It was previously found that Sorafenib treatment supports tumor hypoxia in favor of tumor proliferation, angiogenesis and metastasis (Balkwill, 2004; Gerber, Hippe, Buhren, Müller, & Homey, 2009; Sethi & Kang, 2011; Zlotnik, Burkhardt, & Homey, 2011). Therefore, using Sorafenib as a single agent targeting the PI3K/Akt pathway can even increase the effect of modulating cell redox homeostasis.

The combinational treatment of TGX-221 to Mahlavu resulted with a small network showing cholesterol biosynthetic processes, apical junction complex, SREBP-mediated signaling pathway and Bcl3/NF-kB2 complex. Bcl3/NF-kB2 complex was also enriched in combinational treatment of TGX-221 of Huh7 cells. Yet, NF-kB2 was only upregulated in Mahlavu and Bcl3 downregulation is more apparent. Previous studies highlights the significance of Bcl3 in HCC as being a tumor marker and Bcl3 knock-down in HCC induced cell apoptosis in a study (Poveda et al., 2017; Tu et al., 2016). But none of those analysis was performed on a PTEN-deficient cell line, thus, the differential control of Bcl3/NF-kB2 complex need to be studied more in PTEN context.

As anticipated, Sorafenib and PIK-75 combination was more successful on Mahlavu cells. Mitotic cell cycle was activated and TP53 mediated apoptosis was induced, and

immune response processes were downregulated. Even single treatment of PIK-75 resulted with negative regulation of cell proliferation and regulation of cell growth. GTPase mediated signal transduction and Cdc42 protein signal transduction were observed for both PIK-75 treatments, probably, PIK-75 treatment caused GTPase activity to be transmitted through TGX-221 rather than PIK-75. At this point, it seems to that cell survival is mainly dependent on p110α rather than p110β isoform in PTEN-deficient HCC cell lines. A previous study using PTEN-deficient LNCaP cell line found that p110β is more significant for efficient inhibition of PI3K hyperactivation (Schwartz et al., 2016) and inhibition of p110β is more successful to inhibit cellular growth. Therefore, selective PI3K isoform activation is highly dependent on cellular type and context.

While single PIK-75 treatment in Huh7 activated some important cellular inflammatory related gene ontologies like; interleukin 1 alpha and beta secretion, and interleukin 6 production, no inflammatory response was seen for Mahlavu treatments. Single treatment of PIK-75 activated the negative regulation of cell proliferation and cell growth which were associated with *APOE, FOSL1, HSPA1B* and *HSPA1A* gene regulations. PIK-75 + Sorafenib treatment represented a reduction in cell proliferation, migration, vasculogenesis but mitochondrial apoptotic changes, and increased cell death mainly through negative regulation of transforming growth factor beta receptor signaling pathway were rising. PIK-75 + Sorafenib combination, interestingly, did not provoke an immune response involving a cytokine production. In PIK-75 combinational treatment to Huh7, GTPase activity was ablated by downregulation of *RAB1A, RAB13* and *RAB88* genes and fibroblast growth factor signaling pathway though *SPRY1* and *SPRY2* genes which were negatively regulated in Sorafenib treated Huh7 cells. The pathways including PI3K, MAPK and Ras cascades are redundant in function, PI3K is activated through small GTPases, in turn, acts as an inhibitor of PI3K by stimulation of PTEN (Z. Li et al., 2005; Yang et al., 2012). It is known that interactions of GTPases to PI3K are isoform specific, while Ras cannot bind p110β, RAC1 and CDC42 proteins can activate it. Hence, while single TGX-221 inhibitor treatment on Huh7 cells only affected oxidoreductase activities, its combinatorial treatment with Sorafenib caused more drastic changes than alpha isoform combination. It activated apoptotic changes and positive regulation of programmed cell death, upregulation of T-cell involved immune responses, activation of Bcl3/NF-kB2 complex and JNK cascade, regulation of ion transport and cellular migration. Previous studies on isoform specificity on PI3K activation indicated the main role of p110α on cell growth (Liu et al., 2009), yet combinational treatment of TGX-221 with multi-kinase inhibitor Sorafenib might be also effective for suppression of PI3K hyperactivation in Huh7 cell lines.

The growth arrest and DNA damage 45, Gadd45 protein (*GADD45* gene), is one of the proteins in the cell which is downregulated in response to oncogenic stress in order to regulate stimulation of cell cycle and cellular growth through interacting to cdk1, p21, MEKK4, MKK7 and p38 proteins (Li Zhang et al., 2014a). There is evidence that it's direct interaction with the DNA repair mechanism using proliferation cell nuclear antigen, PCNA, mediates epigenetic gene activation by repair-mediated DNA

demethylation (I. T. Chen et al., 1995). The clue on a tissue's proliferative activity can be deduced from its DNA repair mechanism. Proliferating cell nuclear antigen (PCNA) protein found in the nucleus is involved in RAD6-dependent repair pathway. Previously, it was found that in HCC cells, PCNA protein is positively regulated activating repair mechanisms with response to inactivating mutations in Gadd45; which is even more aggressive in well differentiated HCC (D.-D. Li et al., 2021; Venturi et al., 2008; Z.-L. Zhang et al., 2018). In this analysis, increased Gadd45 levels lead to downregulation of *PCNA* gene in single PIK-75, PIK-75 + Sorafenib, TGX-221 and Sorafenib combination, and single Sorafenib treatments in the Huh7 cells (Figure 37). In previous studies, it was found that Sorafenib treatment on Huh7 cells reduced expression of cyclin D1and PCNA proteins (Lijuan Zhang et al., 2019a).

Furthermore, it has been suggested that Gadd45 interaction to Cyclin dependent kinase inhibitor 1A, p21 protein (*CDKN1A* gene) enhances cyclin dependent kinase 6 (*CDK6* gene) activity to annulate G1 cell cycle arrest (Abbas & Dutta, 2009). P21 protein together with Gadd45 found to be upregulated only in PIK-75 treated Huh7 cells shows that cell cycle activity altered through PIK-75 treatment (Figure 37). With respect to that, Bcl-2-like1 (*BC2L1* gene), known as anti-apoptotic marker, also regulates cell death acting as a regulator of G2 checkpoint and inhibits activation caspases (Bruey et al., 2007) and it is upregulated in PIK-75 treatment opposing action of CDK6 gene. Yet, PIK-75 and Sorafenib combination downregulates AP-1 family proteins JUN kinase, activated transcription factor 4 (ATF4), and proto-oncogene MYC. ATF4 often was found to be overexpressed in HCC cell lines and its induction correlates with chemotherapeutic resistance (Z. Zhang et al., 2012). Activated p21 blocks activation of c-JUN N-terminal kinase (JNK) mediated apoptosis. Insulin induced gene 1 (*INSIG1* gene) is one of the liver specific proteins identified as a biomarker for HCC (G. Yu et al., 2007). Besides its control in cholesterol mechanism, this protein potentially includes in G0/G1 transition of cell growth. In this study, intriguingly, only for Sorafenib treated Huh7 cells, this protein maintained its downregulation together with PCNA.

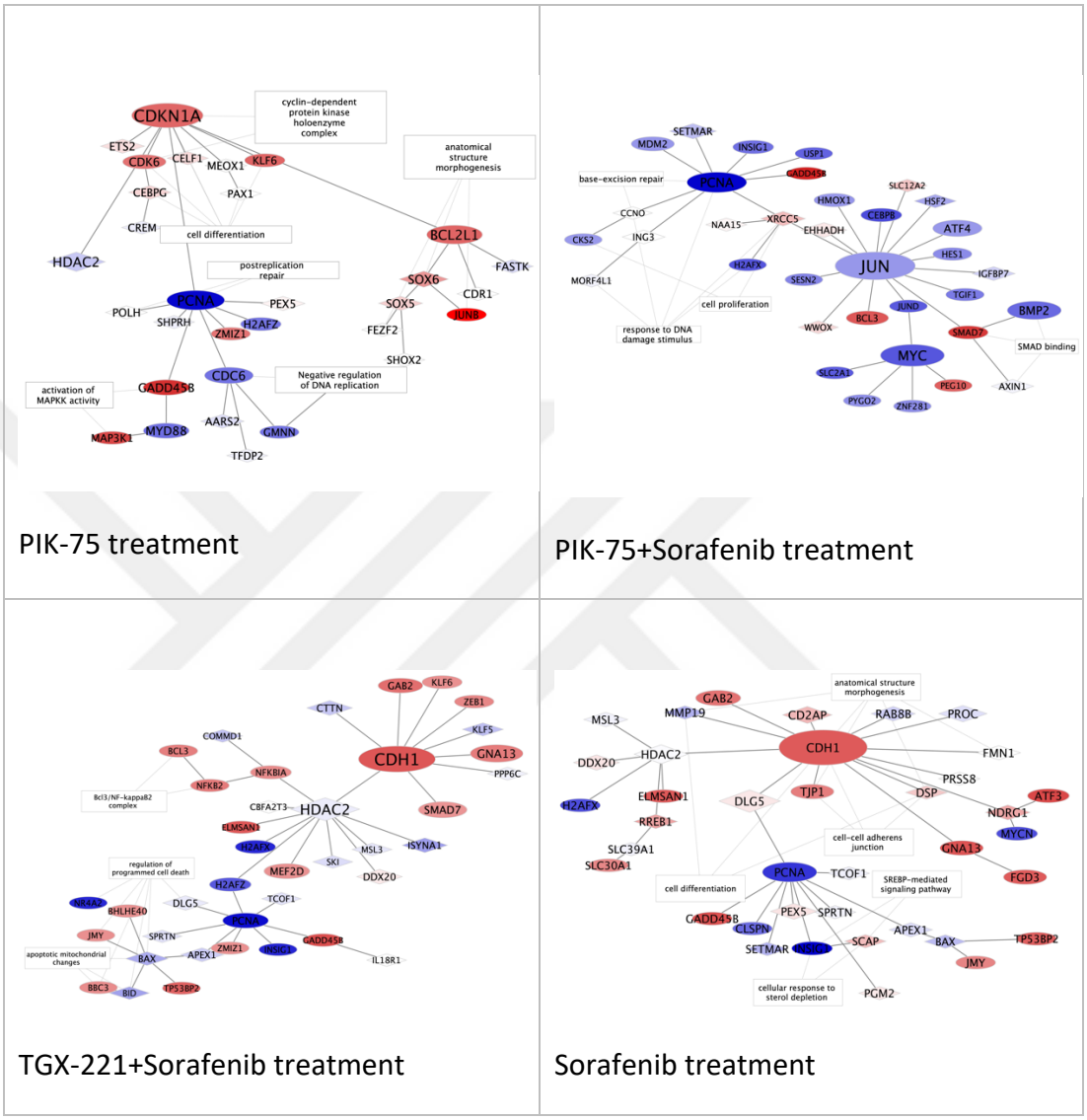


Figure 37. The sub-networks representing different DNA-repair mechanisms in different Huh7 treatments. The color of the nodes was arranged according to the logFC values from differential expression analysis, up regulated genes with red and down regulated genes with blue. Betweenness centralities were used to enlarge the hub genes (nodes). DEG genes were shaped as ellipses and Steiner nodes connecting them were diamonds.

MAP3K1, mitogen-activated protein 3 kinase 1, is a serine/threonine kinase and its autophosphorylation activates ERK, JNK and NF-kB signaling pathways. *MAP3K1* gene was found to be upregulated in PIK-75 alone, Sorafenib alone and Sorafenib and TGX-221 inhibitor treated Huh7 cells, which was an indication of no suppression of MAPK activations upon these treatments. Mammalian myeloid differentiation factor 88 (MyD88 protein) is Toll/interleukin (IL)-1 (TIR)-domain containing adapter protein involved in TLR signaling and an elevated level of MyD88 is found mostly on

metastatic HCC cells increasing overall cancer proliferation (Liang et al., 2013). PIK-75 treated Huh7 cells (alone and together with Sorafenib) responded with an ablated level of *MYD88* expression. Even MyD88 protein was downregulated, MAP3K1 still constitutively activated and its Sorafenib combinatory treatment MAP3K1 upregulation was not mediated in PIK-75 alone treatment.

While downregulation of MyD88 was not monitored, p21 activated kinase (*PAK1* gene), interleukin 1 receptor associated kinase 2 (*IRAK2* gene), RhoGEF and PH domain-containing protein 3 (*FGD3* gene), Supevillin (*SVIL* gene) and Ras homolog family member U protein (*RHOU* gene) were upregulated in Sorafenib alone and combinatory treatment with TGX-221 inhibitor Huh7 cells together with MAP3K1. PAK1, IRAK2 and SVIL were proteins found to be activated in many cancer types including HCC. PAK1 activates proliferation through p53/p21 pathway (Z.-L. Zhang et al., 2018), SVIL proteins are known to activate p38/Erk pathway for metastasis (X. Chen et al., 2018), IRAK2 is found to be involved in the IL1-induced upregulation of NF-kB signaling pathway (Flannery et al., 2011). Furthermore, it was found that PAK1 activation can be through Rho proteins. Therefore, constitutive activation of PAK1 mediated by Rho proteins also may an indication of cellular survival in Sorafenib and its combinatory treatment with TGX-221 in Huh7 cells.

The dual specificity phosphatases (DUSPs) enzyme family provides feedback inhibition on MAPKs towards a stress induction. They play a critical role in the regulation of oncogenic signaling, especially for ERK, p38 and JNK kinases. In this study, downregulation of *DUSP4, DUSP5* and *DUSP6* genes in Huh7 cell treatments İs described in Table 15. Downregulation of *DUSP* genes in cancer genes is part of constitutive activation of the MAPK pathway (C. Huang & Tan, 2012). Yet, *DUSP8* which is the dual inhibitor of p38 and JNK kinases is found to be upregulated in all Sorafenib treated Huh7 cells, which can be ablation of kinase activity. Moreover, interestingly, in Huh7 cells treated with TGX-221 and Sorafenib, besides to *DUSP8, DUSP16* was also upregulated while *DUSP4, DUSP5* and *DUSP6* were mainly stable. Moreover, it was previously found that in PTEN deficient HCC cell lines, DUSP proteins were significantly downregulated to keep mitogen activated protein kinases active apart from AKT activation (Khalid, Hussain, Manzoor, Saalim& Khaliq, 2017). In Mahlavu cells, none of the DUSP were found to be downregulated.

As a part of activated mTORC1 pathway in HCC cells increasing cellular triacylglyceride (TG), apolipoprotein E, ApoE protein production increases playing a role in lipid and lipoprotein metabolism. It has been known that ApoE protein level in serum with liver disease increases especially in Huh7 (Roberts et al., 2016) to meet a high level of cellular energy needs. In this thesis study, the level of ApoE protein in all Huh7 treatments remained stable while in Mahlavu PIK-75 inhibitor treated cells, upregulation of ApoE was identified. An increase in ApoE expression could be the result of downregulation of *LRP1* gene (LDL receptor-related protein 1) which enables secretion of ApoE from endosomes (Laatsch et al., 2012). Moreover, *ABCA1* gene has a role in cholesterol efflux by generation of ApoE containing high density-sized lipoprotein particles which were downregulated PIK-75 treated Mahlavu cells. Yet,

with respect to *LRP1*, in PIK-75 combined with Sorafenib treatment downregulated *LDLR* gene, low-density lipoprotein receptor. LDLR direct interaction to ApoE maintained cholesterol efflux.

Table 15. Differentially expressed *DUSP* genes in Huh7 cells and their target kinases based on Huang & Tan, 2012.

Gene Name	**Gene Targets**	**HCC Treatment Type**
DUSP4	Nuclear; Erk1/2, p38, JNK	Downregulation in Sorafenib
DUSP5	Nuclear; Erk1/2	Downregulation in PIK-75 alone, PIK-75 and Sorafenib combination, Sorafenib only and TGX-221 and Sorafenib combination
DUSP6	Cytoplasmic; Erk1/2	PIK-75 and Sorafenib combination
DUSP8	Cytoplasmic/Nuclear; p38, JNK	Upregulation in PIK-75 alone, PIK-75 and Sorafenib combination, Sorafenib only and TGX-221 and Sorafenib combination
DUSP16	Cytoplasmic/Nuclear; p38, JNK	Upregulation of TGX-221 and Sorafenib combination

PIK-75 + Sorafenib treated Mahlavu cells, death mechanism control was not in the same way as Huh7 did, in the network centered by *TP53* gene, more than 20 gene were differentially expressed: including *APOE, TP63, TP53I3, BCL3, TNFSF14, TP53INP2, DAPK1* and *SOS2* (Figure 38). Death Associated Protein Kinase1 (DAPK1) is a putative tumor suppressor and Tumor Protein p53 Inducible P53 nuclear protein 2 (TP53INP2) is dual regulator of autophagy. These two are downregulated in PIK-75 + Sorafenib treated Mahlavu which in turn may be downregulating cell death. Transcription factor nuclear factor kappa B (NF-kB) together with Bcl-3 protein (*BCL3* gene) as a complex has a central role, increasing activity in response to immune stress and inflammatory injuries, acting as an inhibitor of apoptosis. Bcl-3 over expression known to be increasing sensitivity to apoptosis (Poveda et al., 2017). In Mahlavu, TGX-221 combined with Sorafenib treatment revealed ablation of Bcl-3 may be the basis for its antagonistic action.

Early growth response 1 (EGR1), is one of the factors that targets PTEN, TP53 and JUN. In most cancers EGR1 loss correlates with its tumor suppressor position (Gregg & Fraizer, 2011; J. Yu et al., 2009). In this study, increase in EGR1 expression in single PIK-75 treatment and decrease in expression with TGX-221 + Sorafenib treatment in PTEN deficient Mahlavu cells were observed. EGR1 downregulation

together with JUNB and FOSB correlated with TGX-221 + Sorafenib treated Mahlavu cells antagonistic action to Sorafenib. Interestingly, even EGR1 downregulation was not seen in PIK-75 and Sorafenib combined treatment, EGR1 target TP53 was downregulated stimulating apoptosis (Figure 38).

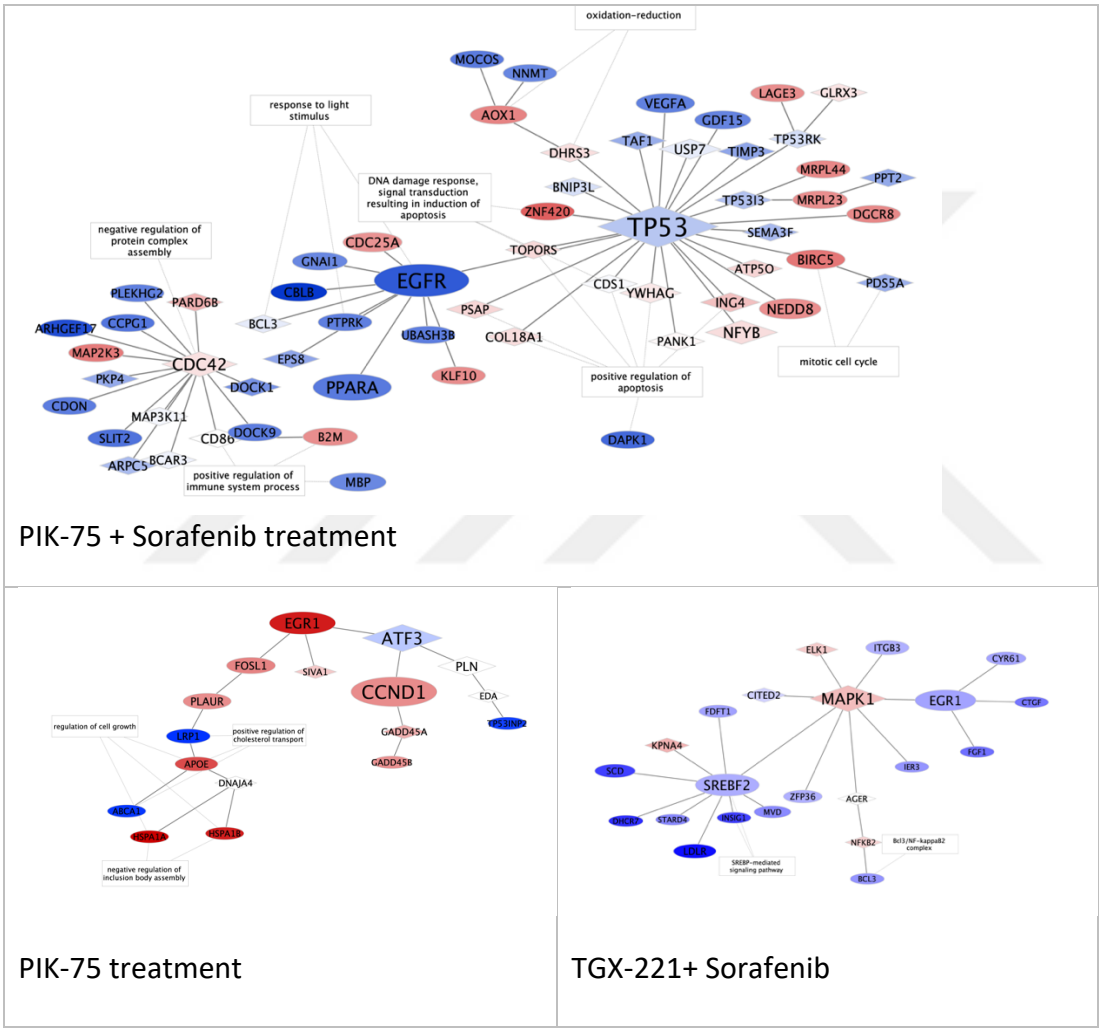


Figure 38. The sub-networks illustrating difference in EGR1 activity in Mahlavu cells. The color of the nodes was arranged according to the logFC values from differential expression analysis; up regulated genes with red and down regulated genes with blue.

The network analysis applied primarily using differentially expressed genes together with the human protein-protein interaction network lead us to identify several signaling pathways. Besides those network comparisons as discussed above, novel drug targets were proposed based on random network constructions and network topology features. Since the hub nodes are mainly part of hot topics or generally well-studied proteins, to detect more novel proteins, a series of filtration was applied as explained in the Methods chapter.

Within the proposed drug targets, besides to differentially expressed genes, there were genes mainly identified as Steiner nodes in the networks. For example, all the prioritized genes in TGX-221 with Sorafenib treatment network for Mahlavu cell line were Steiner. Growth differentiation factor 15 (*GDF15* gene) was one of them. This protein induces tumor angiogenesis as a response to chemotherapy (Dong et al., 2018) and so it could be targeted as potential co-treatment for HCC. Delta-like 1 homolog, *DLK1* gene was also a Steiner node of PIK-75 and Sorafenib treatment network in Huh7. DLK1 is a hepatic stem cell marker, and its overexpression is associated with the cell progression in HCC cells (J. Huang et al., 2007). In a previous study, they show that DLK1 knockdown suppresses cell proliferation and colony formation (Cai et al., 2016; Xu et al., 2012). Thus, DLK1 could be a potential target to inhibit progression gained through PI3K/AKT/mTOR pathway which is already targeted by PIK-75 and Sorafenib in Huh7. DLK may mediate the reduction of progenitor cell development after the malignancy is prohibited using the known kinase inhibitors. Thus, DLK1 should be more studied as therapeutical agent in HCC.

BCL3 and *ASIC* are the two genes were highlighted in Mahlavu PIK-75 + Sorafenib combinational treatment network. *BCL3*, is a protooncogene, regulating cell proliferation (Tu et al., 2016) through cell cycle in HCC. PIK-75 + Sorafenib treatment to highlighted cellular death and apoptosis, yet combining their action with BCL3, may inhibit excessive growth of Mahlavu cells. Acid sensing ion channel 1a (ASIC1a) is a proton gated cation channel regulating tumor migration and invasion recently identified as one of drug resistance genes in HCC (Y. Zhang et al., 2017).

BolA family member 2 (*BOLA2B* gene) stimulates cell proliferation and is associated with poor prognosis in HCC. In a recent study, it was revealed that knockout of *BOLA2* from a HCC type Hep3B cells demonstrated reduction in cell proliferation and tumor growth, hence BOLA2 would be a potential therapeutic target for the treatment of HCC metastasis (Luo et al., 2019). *BOLA2B* gene was a Steiner node and prioritized in PIK-75 treated Mahlavu cells. In PTEN-deficient Mahlavu cell, combinational treatment of PIK-75 + more successful than single treatment of PIK-75, hence a combinational treatment targeting to PI3K-α, Sorafenib and *BOLA2B* gene would be a better strategy. Hence, BOLA2 can be a potential drug target potentially combined with other for clinical studies.

There are many studies establishing diabetes as a the stimulant factor in HCC (Aleksandrova et al., 2016; El-Serag et al., 2004, 2006) tough the main mechanism is still unknown. Previously, it was shown that a high glucose in HCC, a diabetes characteristic, can fasten tumorigenesis process. Advanced glycosylation end product-specific receptor (AGER) was one of the proteins upregulated in diabetes related to Glucose metabolism. AGER was also found to be involved in liver carcinogenesis (Qiao et al., 2016a). *AGER* gene was a Steiner node in TGX-221 + Sorafenib treated Mahlavu network. The combination of TGX-221 + Sorafenib treatment resulted with growth promotion in PTEN deficient Mahlavu cells, so that inhibition of PI3K-β antagonizing Sorafenib function. Here, *AGER* inhibition together with Sorafenib could be a clinical strategy for aggressive Mahlavu cells. By this way, one of the energy

mechanisms would be blocked and a multi-kinase inhibitor, Sorafenib may potentially work better in inhibition of cancer cells growth.

GAB2, Grb2-associated binding protein 2 is a key protein in PI3K and ERK signaling pathways which is closely related to cell proliferation and tumor progression (Y. Chen et al., 2016). The overexpression of *GAB2* gene was identified in HCC tissues in a study. In that study, they also identified microRNA-663b targeting GAB2 to ablate cancerous cell proliferation and invasion (Guo et al., 2019). In this study, *AGER* was also upregulated in TGX-221 + Sorafenib and Sorafenib alone treated Huh7 cells. Maybe using microRNA-663b together with TGX-221 + Sorafenib combination or Sorafenib treatment may suppress cell proliferation which was still in function in Huh7 treatments.

AOX1, Aldehyde oxidase 1, is a highly expressed protein in the liver and associates with the generation of reactive oxygen species. Studies on AOX1 revealed that it is essential for energy generation and drug metabolism (Weigert et al., 2008). *AOX* gene was upregulated in PIK-75 and Sorafenib combinatory treatment in Mahlavu. Since the best working combination to eliminate the cancerous cells was PIK-75+Sorafenib in PTEN deficient Mahlavu, *AOX1* knockdown together with the same combination of treatment in Mahlavu cell may destroy the energy metabolism of the cell diminishing cellular progression.

Finally, *AOX1* and *AGER* genes were selected to be validated by mRNA expression through qPCR experiments. Mahlavu and Huh7 cells were treated with Sorafenib or its combinations with PIK-75 and TGX-221 inhibitors. The gene AGER was selected because it is a pure Steiner node and not found in any DEG lists. Whereas AOX1 is in both Steiner node and part of the DEG list. The experimental analysis was performed at the CANSYL laboratory by one of my collogues. Initial qPCR result correlated and validated the network analysis results. Knockdown experiments for AOX1 and AGER genes were also performed on HCC cells to investigate the effects of these genes on cell toxicity and proliferation. siRNA treatments resulted in significant knockdown of both genes at 25 nM concentrations in Mahlavu and Huh7 cells after 48 h of treatment. Moreover, real-time cell analysis has shown that, silencing AGER and AOX1 significantly inhibited growth of these cells with respect to negative control siRNA treatments. Overall, results from in vitro experiments have supported and validated the systems level network analysis results. AOX1 is considered one of the key biomarkers in HCC and abnormal expression of AOX1 is correlated with the poor prognosis (Jovel et al., 2018). AGER, also is shown as one of the main responsible factors in tumorigenesis of HCC cells in the presence of high glucose for diabetes (Qiao et al., 2016a).

CHAPTER V

CONCLUSION

Hepatocellular carcinoma (HCC), one of the major cancer types, is a leading cause of morbidity and mortality in patients with advanced liver disease. PI3K/Akt/mTOR is the key pathway in HCC since it is a hub to essential functions consisting of regulating cellular growth, glucose metabolism, apoptosis, cell proliferation, cell migration, cellular modification, and cell cycle progression. Consequently, mutations in this signaling pathway leads to constitutive expressions of Akt kinases. These kinases host many downstream effectors like Bcl-2, NF-kB, TFEB and MDM2 controlling cellular survival and metabolism. Sorafenib is the only FDA approved therapeutic agent to HCC serving as a multi-kinase inhibitor in this pathway, yet its single treatment cannot effectively remove cancer cells. One of the reasons for Sorafenib's ineffectiveness is redundancy in the major signaling pathways in HCC.

An emerging strategy to avoid the redundant signaling pathways' functioning is to combine kinase inhibitors to repress multiple compensatory pathways instantaneously. For an efficient screening of more than one combination of kinase inhibitors, next generation sequencing technologies can be used. Analysis of whole transcriptome statues of a targeted cell comparing to a negative control is very an effective approach to capture whole differential regulations of the signaling pathways. Moreover, since combinatory treatments affect more than one kinase, omics data should be interpreted through top to bottom strategy. As human transcriptome hints thousands of protein-protein interaction possibilities, a servant network-level understanding, and systems-level analysis are needed.

This compressive network-level analysis of RNA-seq data show that combinatory treatment of multi-kinase inhibitor Sorafenib and dual inhibitor of PI3Ki-α (PIK-75) stimulates apoptosis both in Huh7 and Mahlavu HCC cell lines correlated with the previous studies. While Sorafenib treatment is more effective in Huh7, PIK-75 treatment in Mahlavu is more successful than Sorafenib to stimulate cell death. The combinational therapy of PI3Ki-β (TGX-221) with Sorafenib may be more efficient in Huh7 cells than PIK-75 combination considering enriched apoptotic pathways in the network. Yet, PI3Ki-β (TGX-221) + Sorafenib treatment antagonizes apoptosis and stimulates growth in Mahlavu, whereas single treatment with TGX-221 has a limited action on both cell lines. This finding was correlated with the previous studies which also shown that PI3K signaling is mainly through p110-β isoform when PTEN is mutated and its inhibition leads p110-α isoform to be activated (Schwartz et al., 2016). Detailed networks assisting for more deep understanding of molecular level actions of

kinase inhibitions are marks of this study. The initial findings supporting these relations were exploited in CANSYL laboratory and details of the study were discussed in the introduction section.

Combination of targeted drugs to inhibit alternative compensatory pathways holds great promise for effective treatment of cancer including HCC. As it has been clearly shown in this study, system-level analysis of cellular networks in response to combination treatments and the investigation of the regulation signaling pathways are of necessity, because such treatments may result in an opposite action. The importance of context-dependent (PTEN status) PI3K/Akt/mTOR signaling inhibition must be taken into consideration during the use of isoform specific or pan-PI3K inhibitors in combination therapies with Sorafenib with respect to a resistance in HCC cells. In this study, many specific or general effects of kinase inhibitors were observed and represented through easy-to-observe visualizations of the gene-to-gene interaction networks.

Furthermore, through network-topology level prioritizations, this thesis proposes drug targets that potentially could be studied more in the future. Expression levels two of the predicted drug targets in this study, *AUX1* and *AGER* genes, were shown to be correlated with the network analysis in CANSYL laboratory. Their qRT-PCR expression verifications were complete. The preliminary results of silencing experiments were also indicated an efficient knock-down of those genes resulted with a decline in growth of these cells.

Finally, the work is presented with this thesis is a united study that born and grow in CANSYL laboratory. My contribution to prove the original hypothesis of the thesis brings the systems level analysis. My insight is that this study will set an example for other cancer studies based on the methodology developed in this thesis work.

REFERENCES

Abbas, T., & Dutta, A. (2009). p21 in cancer: intricate networks and multiple activities. *Nature Reviews. Cancer*, *9*(6), 400–414. https://doi.org/10.1038/nrc2657

Aleksandrova, K., Stelmach-Mardas, M., & Schlesinger, S. (2016). Obesity and Liver Cancer. *Recent Results in Cancer Research. Fortschritte Der Krebsforschung. Progres Dans Les Recherches Sur Le Cancer*, *208*, 177–198. https://doi.org/10.1007/978-3-319-42542-9_10

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166–169. https://doi.org/10.1093/bioinformatics/btu638

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., … Consortium, G. O. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics*, *25*(1), 25–29. https://doi.org/10.1038/75556

Aytuna, A. S., Gursoy, A., & Keskin, O. (2005). Prediction of protein–protein interactions by combining structure and sequence conservation in protein interfaces. *Bioinformatics*, *21*(12), 2850–2855. https://doi.org/10.1093/bioinformatics/bti443

Bader, G. D., Betel, D., & Hogue, C. W. V. (2003). BIND: the Biomolecular Interaction Network Database. *Nucleic Acids Research*, *31*(1), 248–250. https://doi.org/10.1093/nar/gkg056

Bae, J. J., Rho, J. W., Lee, T. J., Yun, S. S., Kim, H. J., Choi, J. H., Jeong, D., Jang, B. C., & Lee, T. Y. (2007). Loss of heterozygosity on chromosome 10q23 and mutation of the phosphatase and tensin homolog deleted from chromosome 10 tumor suppressor gene in Korean hepatocellular carcinoma patients. *Oncology Reports*, *18*(4), 1007–1013.

Berenjeno, I. M., Guillermet-Guibert, J., Pearce, W., Gray, A., Fleming, S., & Vanhaesebroeck, B. (2012). Both p110α and p110β isoforms of PI3K can modulate the impact of loss-of-function of the PTEN tumour suppressor. *The Biochemical Journal*, *442*(1), 151–159. https://doi.org/10.1042/BJ20111741

Bosch, F. X., Ribes, J., Díaz, M., & Cléries, R. (2004). Primary liver cancer: Worldwide

incidence and trends. *Gastroenterology*, *127*(5), S5–S16. https://doi.org/10.1053/j.gastro.2004.09.011

Bray, F., Ferlay, J., & Soerjomataram, I. (2018). Global Cancer Statistics 2018 : GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA CANCER J CLIN*, *68*, 394–424. https://doi.org/10.3322/caac.21492

Breitkreutz, B.-J., Stark, C., & Tyers, M. (2003). Osprey: a network visualization system. *Genome Biology*, *4*(3), R22. https://doi.org/10.1186/gb-2003-4-3-r22

Brito, A. F., Abrantes, A. M., Pinto-Costa, C., Gomes, A. R., Mamede, A. C., Casalta-Lopes, J., Gonçalves, A. C., Sarmento-Ribeiro, A. B., Tralhão, J. G., & Botelho, M. F. (2012). Hepatocellular Carcinoma and Chemotherapy: The Role of p53. *Chemotherapy*, *58*(5), 381–386. https://doi.org/10.1159/000343656

Broad Institute. (2018). Picard Tools - By Broad Institute. *Broad Institute, GitHub Repository*. http://broadinstitute.github.io/picard/

Brown, K. R., Otasek, D., Ali, M., McGuffin, M. J., Xie, W., Devani, B., Toch, I. L. van, & Jurisica, I. (2009). NAViGaTOR: Network Analysis, Visualization and Graphing Toronto. *Bioinformatics (Oxford, England)*, *25*(24), 3327–3329. https://doi.org/10.1093/bioinformatics/btp595

Bruey, J., Bruey-sedano, N., Luciano, F., Zhai, D., Balpai, R., Xu, C., Kress, C. L., Bailly-maitre, B., Li, X., Osterman, A., Matsuzawa, S., Terskikh, A. V, Faustin, B., & Reed, J. C. (2007). Bcl-2 and Bcl-X L Regulate Proinflammatory Caspase-1 Activation by Interaction with NALP1. *Cell*, *1*(045), 45–56. https://doi.org/10.1016/j.cell.2007.01.045

Bruix, J., Qin, S., Merle, P., Granito, A., Huang, Y.-H., Bodoky, G., Pracht, M., Yokosuka, O., Rosmorduc, O., Breder, V., Gerolami, R., Masi, G., Ross, P. J., Song, T., Bronowicki, J.-P., Ollivier-Hourmand, I., Kudo, M., Cheng, A.-L., Llovet, J. M., … Han, G. (2017). Regorafenib for patients with hepatocellular carcinoma who progressed on sorafenib treatment (RESORCE): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet (London, England)*, *389*(10064), 56–66. https://doi.org/10.1016/S0140-6736(16)32453-9

Buontempo, F., Ersahin, T., Missiroli, S., Senturk, S., Etro, D., Ozturk, M., Capitani, S., Cetin-Atalay, R., & Neri, M. L. (2011). Inhibition of Akt signaling in hepatoma cells induces apoptotic cell death independent of Akt activation status. *Investigational New Drugs*, *29*(6), 1303–1313. https://doi.org/10.1007/s10637-010-9486-3

Cai, C.-M., Xiao, X., Wu, B.-H., Wei, B.-F., & Han, Z.-G. (2016). Targeting endogenous DLK1 exerts antitumor effect on hepatocellular carcinoma through initiating cell differentiation. *Oncotarget*, *7*(44), 71466–71476.

https://doi.org/10.18632/oncotarget.12214

Carbon, S., Douglass, E., Good, B. M., Unni, D. R., Harris, N. L., Mungall, C. J., Basu, S., Chisholm, R. L., Dodson, R. J., Hartline, E., Fey, P., Thomas, P. D., Albou, L. P., Ebert, D., Kesling, M. J., Mi, H., Muruganujan, A., Huang, X., Mushayahama, T., … Elser, J. (2021). The Gene Ontology resource: Enriching a GOld mine. *Nucleic Acids Research*, *49*(D1), D325–D334. https://doi.org/10.1093/nar/gkaa1113

Cavallo, F., De Giovanni, C., Nanni, P., Forni, G., & Lollini, P. L. (2011). 2011: The immune hallmarks of cancer. *Cancer Immunology, Immunotherapy*, *60*(3), 319–326.

Chatr-aryamontri, A., Ceol, A., Palazzi, L. M., Nardelli, G., Schneider, M. V., Castagnoli, L., & Cesareni, G. (2007). MINT: the Molecular INTeraction database. *Nucleic Acids Research*, *35*(Database issue), D572-4. https://doi.org/10.1093/nar/gkl950

Chattopadhyay, M., Selinger, E. S., Ballou, L. M., & Lin, R. Z. (2011). Ablation of PI3K p110-α prevents high-fat diet-induced liver steatosis. *Diabetes*, *60*(5), 1483–1492. https://doi.org/10.2337/db10-0869

Chen, I. T., Smith, M. L., O'Connor, P. M., & Fornace, A. J. J. (1995). Direct interaction of Gadd45 with PCNA and evidence for competitive interaction of Gadd45 and p21Waf1/Cip1 with PCNA. *Oncogene*, *11*(10), 1931–1937.

Chen, X., Zhang, S., Wang, Z., Wang, F., Cao, X., Wu, Q., Zhao, C., Ma, H., Ye, F., Wang, H., & Fang, Z. (2018). Supervillin promotes epithelial- mesenchymal transition and metastasis of hepatocellular carcinoma in hypoxia via activation of the RhoA / ROCK-ERK / p38 pathway. *Journal of Experimental & Clinical Cancer Research*, *37*(128), 1–16.

Chen, Y., Liu, Q., Wu, M., Li, M., Ding, H., Shan, X., Liu, J., Tao, T., Ni, R., & Chen, X. (2016). GAB2 promotes cell proliferation by activating the ERK signaling pathway in hepatocellular carcinoma. *Tumor Biology*, *37*(9), 11763–11773. https://doi.org/10.1007/s13277-016-5019-9

Cheng, A.-L., Kang, Y.-K., Chen, Z., Tsao, C.-J., Qin, S., Kim, J. S., Luo, R., Feng, J., Ye, S., Yang, T.-S., Xu, J., Sun, Y., Liang, H., Liu, J., Wang, J., Tak, W. Y., Pan, H., Burock, K., Zou, J., … Guan, Z. (2009). Efficacy and safety of sorafenib in patients in the Asia-Pacific region with advanced hepatocellular carcinoma: a phase III randomised, double-blind, placebo-controlled trial. *The Lancet Oncology*, *10*(1), 25–34. https://doi.org/10.1016/S1470-2045(08)70285-7

Chu, Y., & Corey, D. R. (2012). RNA Sequencing: Platform Selection, Experimental Design, and Data Interpretation. *Nucleic Acid Therapeutics*, *22*(4), 271–274. https://doi.org/10.1089/nat.2012.0367

Compeau, P. E. C., Pevzner, P. A., Tesler, G., Papoutsoglou, G., Roscito, J. G., Dahl, A.,

Myers, G., Winkler, S., Pippel, M., Sameith, K., Hiller, M., Francoijs, K.-J., Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., Nagarajan, N., Pop, M., Abbas, M. M., … Martin, M. (2013). Cutadapt removes adapter sequences from high-throughput sequencing reads kenkyuhi hojokin gan rinsho kenkyu jigyo. *EMBnet Journal*, *17*(1), 10–12. https://doi.org/10.14806/ej.17.1.200

Date, S. V, & Marcotte, E. M. (2005). Protein function prediction using the Protein Link EXplorer (PLEX). *Bioinformatics*, *21*(10), 2558–2559. https://doi.org/10.1093/bioinformatics/bti313

Dennis Jr, G., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, C. H., Lempicki, R. A., Dennis, G., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., & Lempicki, R. A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biology*, *4*(5), P3. https://doi.org/10.1186/gb-2003-4-9-r60

Dong, G., Zheng, Q. D., Ma, M., Wu, S. F., Zhang, R., Yao, R. R., Dong, Y. Y., Ma, H., Gao, D. M., Ye, S. L., Cui, J. F., Ren, Z. G., & Chen, R. X. (2018). Angiogenesis enhanced by treatment damage to hepatocellular carcinoma through the release of GDF15. *Cancer Medicine*, *7*(3), 820–830. https://doi.org/10.1002/cam4.1330

Dragani, T. A. (2010). Risk of HCC: Genetic heterogeneity and complex genetics. *Journal of Hepatology*, *52*(2), 252–257. https://doi.org/10.1016/j.jhep.2009.11.015

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/https://doi.org/10.1016/j.phymed.2015.11.012

El-Serag, H. B., Hampel, H., & Javadi, F. (2006). The association between diabetes and hepatocellular carcinoma: a systematic review of epidemiologic evidence. *Clinical Gastroenterology and Hepatology : The Official Clinical Practice Journal of the American Gastroenterological Association*, *4*(3), 369–380. https://doi.org/10.1016/j.cgh.2005.12.007

El-Serag, H. B., Tran, T., & Everhart, J. E. (2004). Diabetes increases the risk of chronic liver disease and hepatocellular carcinoma. *Gastroenterology*, *126*(2), 460–468. https://doi.org/10.1053/j.gastro.2003.10.065

Engelman, J. A. (2009). Targeting PI3K signalling in cancer: Opportunities, challenges and limitations. *Nature Reviews Cancer*, *9*(8), 550–562. https://doi.org/10.1038/nrc2664

Ersahin, T. (2014). *Systems biology approach for targeted therapy of liver cancer pi3k/akt/mtor pathway inhibitors: an ally or rival for sorafenib*. Bilkent University, Molecular Biology aand Genetics Department.

Ersahin, T., Carkacioglu, L., Can, T., Konu, O., Atalay, V., & Cetin-Atalay, R. (2014). Identification of novel reference genes based on MeSH categories. *PLoS ONE*, *9*(3). https://doi.org/10.1371/journal.pone.0093341

Ersahin, T., Ozturk, M., & Cetin-Atalay, R. (2015). Molecular biology of liver cancer. In *Rev. Cell Biol. Mol. Medicine* (Vol. 1, Issue 2, pp. 206–232). https://doi.org/10.1159/000343863

Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., Milacic, M., Roca, C. D., Rothfels, K., Sevilla, C., Shamovsky, V., Shorser, S., Varusai, T., Viteri, G., Weiser, J., … D'Eustachio, P. (2018). The Reactome Pathway Knowledgebase. *Nucleic Acids Research*, *46*(D1), D649–D655. https://doi.org/10.1093/nar/gkx1132

Farazi, P. A., & DePinho, R. A. (2006). Hepatocellular carcinoma pathogenesis: From genes to environment. *Nature Reviews Cancer*, *6*(9), 674–687. https://doi.org/10.1038/nrc1934

Flannery, S. M., Keating, S. E., Szymak, J., & Bowie, A. G. (2011). Human Interleukin-1 Receptor-associated Kinase-2 Is Essential for Toll-like Receptor-mediated Transcriptional and Post-transcriptional Regulation of Tumor Necrosis Factor ␣ *. *Journal of Biological Chemistry*, *286*(27), 23688–23697. https://doi.org/10.1074/jbc.M111.248351

Fruman, D. A., & Rommel, C. (2014). PI3K and cancer: lessons, challenges and opportunities. *Nature Reviews. Drug Discovery*, *13*(2), 140–156. https://doi.org/10.1038/nrd4204

Fujimoto, A., Totoki, Y., Abe, T., Boroevich, K. A., Hosoda, F., Nguyen, H. H., Aoki, M., Hosono, N., Kubo, M., Miya, F., Arai, Y., Takahashi, H., Shirakihara, T., Nagasaki, M., Shibuya, T., Nakano, K., Watanabe-Makino, K., Tanaka, H., Nakamura, H., … Nakagawa, H. (2012). Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nature Genetics*, *44*(7), 760–764. https://doi.org/10.1038/ng.2291

Galili, T., Callaghan, A. O., Sidi, J., & Sievert, C. (2012). heatmaply : an R package for creating interactive cluster heatmaps for online publishing. *Oxford Academic, Bioinformatics*, *December*, 1–2. https://doi.org/10.1093/bioinformatics/btx657/4562328

Goossens, N., Sun, X., & Hoshida, Y. (2015). Molecular classification of hepatocellular carcinoma: potential therapeutic implications. *Hepatic Oncology*, *2*(4), 371–379. https://doi.org/10.2217/hep.15.26

Graeber, T. G., & Eisenberg, D. (2001). Bioinformatic identification of potential autocrine

signaling loops in cancers from gene expression profiles. In *Nature genetics* (Vol. 29, Issue 3, pp. 295–300). https://doi.org/10.1038/ng755

Gregg, J., & Fraizer, G. (2011). Transcriptional Regulation of EGR1 by EGF and the ERK Signaling Pathway in Prostate Cancer Cells. *Genes & Cancer*, *2*(9), 900–909. https://doi.org/10.1177/1947601911431885

Guo, L., Li, B., Miao, M., Yang, J., & Ji, J. (2019). MicroRNA-663b targets GAB2 to restrict cell proliferation and invasion in hepatocellular carcinoma. *Mol Med Rep*, *19*(4), 2913–2920. https://doi.org/10.3892/mmr.2019.9934

Hagberg, A. A., Schult, D. A., & Swart, P. J. (2008). Exploring network structure, dynamics, and function using NetworkX. *Proceedings of the 7th Python in Science Conference (SciPy 2008)*, *SciPy*, 11–15.

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. *Oncology Letters*, *5*(4), 1087–1092. https://doi.org/10.3892/ol.2013.1169

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: The next generation. In *Cell* (Vol. 144, Issue 5, pp. 646–674).

He, Q., Johnston, J., Zeitlinger, J., City, K., & City, K. (2015). PI3K p110α and p110β have differential effects on Akt activation and protection against oxidative stress-induced apoptosis in myoblasts. *Cell Death Differ.*, *33*(4), 395–401. https://doi.org/10.1038/nbt.3121.ChIP-nexus

Huang, C., & Tan, T. (2012). DUSPs , to MAP kinases and beyond. *Cell & Bioscience*, *2*(1), 1. https://doi.org/10.1186/2045-3701-2-24

Huang, J., Zhang, X., Zhang, M., Zhu, J.-D., Zhang, Y.-L., Lin, Y., Wang, K.-S., Qi, X.-F., Zhang, Q., Liu, G.-Z., Yu, J., Cui, Y., Yang, P.-Y., Wang, Z.-Q., & Han, Z.-G. (2007). Up-regulation of DLK1 as an imprinted gene could contribute to human hepatocellular carcinoma. *Carcinogenesis*, *28*(5), 1094–1103. https://doi.org/10.1093/carcin/bgl215

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., & Sakaki, Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(8), 4569–4574. https://doi.org/10.1073/pnas.061034498

Jia, S., Liu, Z., Zhang, S., Liu, P., Zhang, L., Hyun, S., Zhang, J., Signoretti, S., Loda, M., Roberts, T. M., & Zhao, J. J. (2008). Kinase-dependent and -independent functions of the p110β phosphoinositide-3-kinase in cell growth, metabolic regulation and oncogenic transformation. *Nature*, *454*(7205), 776–779.

https://doi.org/10.1038/nature07091.

Jovel, J., Lin, Z., Okeefe, S., Willos, S., Wang, W., Zhang, G., & Patterson, J. (2018). A survey of Molecular heterogeneity in hepatocellular carcinoma. *Hepatic Oncology*, *5*(1), HEP10. https://doi.org/10.2217/hep-2018-0005

Jung, M., Drapier, J. C., Weidenbach, H., Renia, L., Oliveira, L., Wang, A., Beger, H. G., & Nussler, A. K. (2000). Effects of hepatocellular iron imbalance on nitric oxide and reactive oxygen intermediates production in a model of sepsis. *Journal of Hepatology*, *33*(3), 387–394.

Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, *28*(1), 27–30. https://doi.org/10.1093/nar/28.1.27

Kawamura, N., Nagai, H., Bando, K., Koyama, M., Matsumoto, S., Tajiri, T., Onda, M., Fujimoto, J., Ueki, T., Konishi, N., Shiba, T., & Emi, M. (1999). PTEN/MMAC1 mutations in hepatocellular carcinomas: Somatic inactivation of both alleles in tumors. *Japanese Journal of Cancer Research*, *90*(4), 413–418. https://doi.org/10.1111/j.1349-7006.1999.tb00763.x

Khalid, A., Hussain, T., Manzoor, S., Saalim, M., & Khaliq, S. (2017). PTEN : A potential prognostic marker in virus-induced hepatocellular carcinoma. *Tumor Biology*, 1–10. https://doi.org/10.1177/1010428317705754

Kholodenko, B. N., Kiyatkin, A., Bruggeman, F. J., Sontag, E., Westerhoff, H. V, & Hoek, J. B. (2002). Untangling the wires: a strategy to trace functional interactions in signaling and gene networks. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(20), 12841–12846. https://doi.org/10.1073/pnas.192442699

Khurana, V., Peng, J., Chung, C. Y., Auluck, P. K., Fanning, S., Tardiff, D. F., Bartels, T., Koeva, M., Eichhorn, S. W., Benyamini, H., Lou, Y., Nutter-Upham, A., Baru, V., Freyzon, Y., Tuncbag, N., Costanzo, M., San Luis, B. J., Schöndorf, D. C., Barrasa, M. I., … Lindquist, S. (2017). Genome-Scale Networks Link Neurodegenerative Disease Genes to α-Synuclein through Specific Molecular Pathways. *Cell Systems*, *4*(2), 157-170.e14. https://doi.org/10.1016/j.cels.2016.12.011

Kitano, H. (2002). Systems biology: a brief overview. *Science (New York, N.Y.)*, *295*(5560), 1662–1664. https://doi.org/10.1126/science.1069492

Kucera, M., Isserlin, R., Arkhangorodsky, A., & Bader, G. D. (2016). AutoAnnotate: A Cytoscape app for summarizing networks with semantic annotations. *F1000Research*, *5*, 1717. https://doi.org/10.12688/f1000research.9090.1

Kuhn, M., Szklarczyk, D., Pletscher-Frankild, S., Blicher, T. H., von Mering, C., Jensen,

L. J., & Bork, P. (2013). STITCH 4: integration of protein–chemical interactions with user data. *Nucleic Acids Research*, *42*(D1), D401–D407. https://doi.org/10.1093/nar/gkt1207

Kukurba, K. R., & Montgomery, S. B. (2015). RNA Sequencing and Analysis Kimberly. *Cold Spring Harbor Protocols*, *2015*(11), 951–969. https://doi.org/10.1101/pdb.top084970.RNA

Kuleshov, M. V, Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S. L., Jagodnik, K. M., Lachmann, A., McDermott, M. G., Monteiro, C. D., Gundersen, G. W., & Ma'ayan, A. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*, *44*(W1), W90-7. https://doi.org/10.1093/nar/gkw377

Kuper, H., Tzonou, A., Kaklamani, E., Hsieh, C. C., Lagiou, P., Adami, H. O., Trichopoulos, D., & Stuver, S. O. (2000). Tobacco smoking, alcohol consumption and their interaction in the causation of hepatocellular carcinoma. *International Journal of Cancer*, *85*(4), 498–502.

Kuperstein, I., Bonnet, E., Nguyen, H.-A., Cohen, D., Viara, E., Grieco, L., Fourquet, S., Calzone, L., Russo, C., Kondratova, M., Dutreix, M., Barillot, E., & Zinovyev, A. (2015). Atlas of Cancer Signalling Network: a systems biology resource for integrative analysis of cancer data with Google Maps. *Oncogenesis*, *4*, e160. https://doi.org/10.1038/oncsis.2015.19

Laatsch, A., Panteli, M., Sornsakrin, M., Hoffzimmer, B., Grewal, T., & Heeren, J. (2012). Low Density Lipoprotein Receptor-Related Protein 1 Dependent Endosomal Trapping and Recycling of Apolipoprotein E. *PLoS ONE*, *7*(1), 1–10. https://doi.org/10.1371/journal.pone.0029385

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., … Trust:, T. W. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860–921. https://doi.org/10.1038/35057062

Li, D.-D., Zhang, J.-W., Zhang, R., Xie, J.-H., Zhang, K., Lin, G.-G., Han, Y.-X., Peng, R.-X., Han, D.-S., Wang, J., Yang, J., & Li, J.-M. (2021). Proliferating cell nuclear antigen (PCNA) overexpression in hepatocellular carcinoma predicts poor prognosis as determined by bioinformatic analysis. *Chinese Medical Journal*, *134*(7).

Li, Z., Dong, X., Dong, X., Wang, Z., Liu, W., Deng, N., Ding, Y., Tang, L., Hla, T., Zeng, R., Li, L., & Wu, D. (2005). Regulation of PTEN by Rho small GTPases. *Nature Cell Biology*, *7*(4), 399–404. https://doi.org/10.1038/ncb1236

Liang, B., Chen, R., Wang, T., Cao, L., Liu, Y., Yin, F., Zhu, M., & Fan, X. (2013).

Myeloid Differentiation Factor 88 Promotes Growth and Metastasis of Human Hepatocellular Carcinoma. *Clinical Cancer Research*, *12*(1245), 1–13. https://doi.org/10.1158/1078-0432.CCR-12-1245

Liu, P., Cheng, H., Roberts, T. M., & Zhao, J. J. (2009). Targeting the phosphoinositide 3-kinase pathway in cancer. *Nature Reviews Drug Discovery*, *8*(8), 627–644. https://doi.org/10.1038/nrd2926

Llovet, J. M., Di Bisceglie, A. M., Bruix, J., Kramer, B. S., Lencioni, R., Zhu, A. X., Sherman, M., Schwartz, M., Lotze, M., Talwalkar, J., & Gores, G. J. (2008). Design and endpoints of clinical trials in hepatocellular carcinoma. *Journal of the National Cancer Institute*, *100*(10), 698–711. https://doi.org/10.1093/jnci/djn134

Llovet, J. M., Ducreux, M., Lencioni, R., Di Bisceglie, A. M., Galle, P. R., Dufour, J. F., Greten, T. F., Raymond, E., Roskams, T., De Baere, T., Mazzaferro, V., Bernardi, M., Bruix, J., Colombo, M., & Zhu, A. (2012). EASL-EORTC Clinical Practice Guidelines: Management of hepatocellular carcinoma. *Journal of Hepatology*, *48*(5), 599–641. https://doi.org/10.1016/j.ejca.2011.12.021

Llovet, J. M., & Hernandez-Gea, V. (2014). Hepatocellular carcinoma: Reasons for phase III failure and novel perspectives on trial design. *Clinical Cancer Research*, *20*(8), 2072–2079. https://doi.org/10.1158/1078-0432.CCR-13-0547

Llovet, J. M., Ricci, S., Mazzaferro, V., Hilgard, P., Gane, E., Blanc, J.-F., de Oliveira, A. C., Santoro, A., Raoul, J.-L., Forner, A., Schwartz, M., Porta, C., Zeuzem, S., Bolondi, L., Greten, T. F., Galle, P. R., Seitz, J.-F., Borbath, I., Häussinger, D., … Bruix, J. (2008). Sorafenib in Advanced Hepatocellular Carcinoma. *New England Journal of Medicine*, *359*(4), 378–390. https://doi.org/10.1056/NEJMoa0708857

Luo, J., Wang, D., Zhang, S., Kuan, H., Haijun, W., Li, J., Wang, Z., & Tao, Y. (2019). BolA family member 2 enhances cell proliferation and predicts a poor prognosis in hepatocellular carcinoma with tumor hemorrhage. *Journal of Cancer*, *4*(6), 1–32. https://doi.org/doi:10.7150/jca.31829

Maere, S., Heymans, K., & Kuiper, M. (2005). BiNGO: A Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks. *Bioinformatics*, *21*(16), 3448–3449. https://doi.org/10.1093/bioinformatics/bti551

Maggiora, G. M. (2011). The reductionist paradox: are the laws of chemistry and physics sufficient for the discovery of new drugs? *Journal of Computer-Aided Molecular Design*, *25*(8), 699–708. https://doi.org/10.1007/s10822-011-9447-8

Manning, B. D., & Cantley, L. C. (2007). AKT/PKB signaling: navigating downstream. *Cell*, *129*(7), 1261–1274. https://doi.org/10.1016/j.cell.2007.06.009

Marc Carlson. (2016). *org.Hs.eg.db: Genome wide annotation for Human. R package*

version 3.4.0.

Medina-Franco, J. L., Giulianotti, M. A., Welmaker, G. S., & Houghten, R. A. (2013). Shifting from the single to the multitarget paradigm in drug discovery. *Drug Discovery Today*, *18*(9–10), 495–501. https://doi.org/10.1016/j.drudis.2013.01.008

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

Narci, K., Kahraman, D. C., Koyas, A., Ersahin, T., Tuncbag, N., & Atalay, R. C. (2021). Context dependent isoform specific PI3K inhibition confers drug resistance in Hepatocellular carcinoma cells. *BioRxiv*, 2021.03.17.435839. https://doi.org/10.1101/2021.03.17.435839

Ogimoto, I., Shibata, A., Kurozawa, Y., Nose, T., Yoshimura, T., Suzuki, H., Iwai, N., Sakata, R., Fujita, Y., Ichikawa, S., Fukuda, K., & Tamakoshi, A. (2004). Risk of death due to hepatocellular carcinoma among drinkers and ex-drinkers. Univariate analysis of JACC study data. *The Kurume Medical Journal*, *51*(1), 59–70.

Oh, B.-K., Jo Chae, K., Park, C., Kim, K., Jung Lee, W., Han, K., & Nyun Park, Y. (2003). Telomere shortening and telomerase reactivation in dysplastic nodules of human hepatocarcinogenesis. *Journal of Hepatology*, *39*(5), 786–792.

Okamoto, T., Sanda, T., & Asamitsu, K. (2007). NF-kappa B signaling and carcinogenesis. *Current Pharmaceutical Design*, *13*(5), 447–462. http://www.ncbi.nlm.nih.gov/pubmed/17348842

Omata, M., Lesmana, L. A., Tateishi, R., Chen, P. J., Lin, S. M., Yoshida, H., Kudo, M., Lee, J. M., Choi, B. I., Poon, R. T. P., Shiina, S., Cheng, A. L., Jia, J. D., Obi, S., Han, K. H., Jafri, W., Chow, P., Lim, S. G., Chawla, Y. K., … Sarin, S. K. (2010). Asian pacific association for the study of the liver consensus recommendations on hepatocellular carcinoma. *Hepatology International*, *4*(2), 439–474. https://doi.org/10.1007/s12072-010-9165-7

Ozcelik, D., Ozaras, R., Gurel, Z., Uzun, H., & Aydin, S. (2003). Copper-mediated oxidative stress in rat liver. *Biological Trace Element Research*, *96*(1–3), 209–215. https://doi.org/10.1385/BTER:96:1-3:209

Ozturk, M., Arslan-Ergul, A., Bagislar, S., Senturk, S., & Yuzugullu, H. (2009). Senescence and immortality in hepatocellular carcinoma. *Cancer Letters*, *286*(1), 103–113. https://doi.org/10.1016/j.canlet.2008.10.048

Paolini, G. V, Shapland, R. H. B., van Hoorn, W. P., Mason, J. S., & Hopkins, A. L. (2006). Global mapping of pharmacological space. *Nature Biotechnology*, *24*(7), 805–815. https://doi.org/10.1038/nbt1228

Papetti, M., & Herman, I. M. (2002). Mechanisms of normal and tumor-derived angiogenesis. *American Journal of Physiology. Cell Physiology*, *282*(5), C947-70. https://doi.org/10.1152/ajpcell.00389.2001

Patel-Murray, N. L., Adam, M., Huynh, N., Wassie, B. T., Milani, P., & Fraenkel, E. (2020). A Multi-Omics Interpretable Machine Learning Model Reveals Modes of Action of Small Molecules. *Scientific Reports*, *10*(1), 954. https://doi.org/10.1038/s41598-020-57691-7

Pathan, M., Keerthikumar, S., Ang, C. S., Gangoda, L., Quek, C. Y. J., Williamson, N. A., Mouradov, D., Sieber, O. M., Simpson, R. J., Salim, A., Bacic, A., Hill, A. F., Stroud, D. A., Ryan, M. T., Agbinya, J. I., Mariadason, J. M., Burgess, A. W., & Mathivanan, S. (2015). FunRich: An open access standalone functional enrichment and interaction network analysis tool. *Proteomics*, *15*(15), 2597–2601. https://doi.org/10.1002/pmic.201400515

Pavlopoulos, G. A., O'Donoghue, S. I., Satagopam, V. P., Soldatos, T. G., Pafilis, E., & Schneider, R. (2008). Arena3D: visualization of biological networks in 3D. *BMC Systems Biology*, *2*, 104. https://doi.org/10.1186/1752-0509-2-104

Personeni, N., Pressiani, T., Santoro, A., & Rimassa, L. (2018). Regorafenib in hepatocellular carcinoma: latest evidence and clinical implications. *Drugs in Context*, *7*, 212533. https://doi.org/10.7573/dic.212533

Perz, J. F., Armstrong, G. L., Farrington, L. A., Hutin, Y. J. F., & Bell, B. P. (2006). The contributions of hepatitis B virus and hepatitis C virus infections to cirrhosis and primary liver cancer worldwide. *Journal of Hepatology*, *45*(4), 529–538. https://doi.org/10.1016/j.jhep.2006.05.013

Porta, C., Paglino, C., & Mosca, A. (2014). Targeting PI3K/Akt/mTOR Signaling in Cancer. *Frontiers in Oncology*, *4*, 64. https://doi.org/10.3389/fonc.2014.00064

Poveda, J., Sanz, A. B., Carrasco, S., Ruiz-Ortega, M., Cannata-Ortiz, P., Sanchez-Niño, M. D., & Ortiz, A. (2017). Bcl3: a regulator of NF-κB inducible by TWEAK in acute kidney injury with anti-inflammatory and antiapoptotic properties in tubular cells. *Experimental &Amp; Molecular Medicine*, *49*, e352. https://doi.org/10.1038/emm.2017.89

Qiao, Y., Zhang, X., Zhang, Y., Wang, Y., Xu, Y., Liu, X., Sun, F., & Wang, J. (2016a). High glucose stimulates tumorigenesis in hepatocellular carcinoma cells through ager-dependent O-glcnacylation of C-jun. *Diabetes*, *65*(3), 619–632. https://doi.org/10.2337/db15-1057

Qiao, Y., Zhang, X., Zhang, Y., Wang, Y., Xu, Y., Liu, X., Sun, F., & Wang, J. (2016b). High Glucose Stimulates Tumorigenesis in Hepatocellular Carcinoma Cells Through AGER-Dependent O-GlcNAcylation of c-Jun. *Diabetes*, *65*(3), 619–632.

https://doi.org/10.2337/db15-1057

Ray, K. (2020). A combination therapy to keep hepatocellular carcinoma in check. *Nature Reviews Gastroenterology & Hepatology*, *17*(8), 452. https://doi.org/10.1038/s41575-020-0337-1

Raza, A., & Sood, G. K. (2014). Hepatocellular carcinoma review: current treatment, and evidence-based medicine. *World Journal of Gastroenterology*, *20*(15), 4115–4127. https://doi.org/10.3748/wjg.v20.i15.4115

Razick, S., Magklaras, G., & Donaldson, I. M. (2008). iRefIndex: A consolidated protein interaction database with provenance. *BMC Bioinformatics*, *9*(1), 405. https://doi.org/10.1186/1471-2105-9-405

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., & Seraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. *Nature Biotechnology*, *17*(10), 1030–1032. https://doi.org/10.1038/13732

Roberts, J. L., He, B., Erickson, A., & Moreau, R. (2016). Improvement of mTORC1-driven overproduction of apoB-containing triacylglyceride-rich lipoproteins by short-chain fatty acids , 4-phenylbutyric acid and (R) - α -lipoic acid , in human hepatocellular carcinoma cells. *Biochimica et Biophysica Acta*, *1861*, 166–176. https://doi.org/10.1016/j.bbalip.2015.12.001

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2009). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Robles, J. a, Qureshi, S. E., Stephen, S. J., Wilson, S. R., Burden, C. J., & Taylor, J. M. (2012). Efficient experimental design and analysis strategies for the detection of differential expression using RNA-Sequencing. *BMC Genomics*, *13*(1), 484. https://doi.org/10.1186/1471-2164-13-484

Rodon, J., Dienstmann, R., Serra, V., & Tabernero, J. (2013). Development of PI3K inhibitors: lessons learned from early clinical trials. *Nature Reviews. Clinical Oncology*, *10*(3), 143–153. https://doi.org/10.1038/nrclinonc.2013.10

Ruggero, D., & Sonenberg, N. (2005). The Akt of translational control. *Oncogene*, *24*(50), 7426–7434. https://doi.org/10.1038/sj.onc.1209098

Salanti, G., Southam, L., Altshuler, D., Ardlie, K., Barroso, I., Boehnke, M., Cornelis, M. C., Frayling, T. M., Grallert, H., Grarup, N., Groop, L., Hansen, T., Hattersley, A. T., Hu, F. B., Hveem, K., Illig, T., Kuusisto, J., Laakso, M., Langenberg, C., … Ioannidis, J. P. A. (2009). Underlying genetic models of inheritance in established type 2 diabetes associations. *American Journal of Epidemiology*, *170*(5), 537–545.

https://doi.org/10.1093/aje/kwp145

Schlesinger, S., Aleksandrova, K., Pischon, T., Jenab, M., Fedirko, V., Trepo, E., Overvad, K., Roswall, N., Tjønneland, A., Boutron-Ruault, M. C., Fagherazzi, G., Racine, A., Kaaks, R., Grote, V. A., Boeing, H., Trichopoulou, A., Pantzalis, M., Kritikou, M., Mattiello, A., … Nöthlings, U. (2013). Diabetes mellitus, insulin treatment, diabetes duration, and risk of biliary tract cancer and hepatocellular carcinoma in a European Cohort. *Annals of Oncology*, *24*(9), 2449–2455. https://doi.org/10.1093/annonc/mdt204

Schulze, K., Nault, J. C., & Villanueva, A. (2016). Genetic profiling of hepatocellular carcinoma using next-generation sequencing. *Journal of Hepatology*, *65*(5), 1031–1042. https://doi.org/10.1016/j.jhep.2016.05.035

Schwartz, S., Wongvipat, J., Trigwell, C. B., Hancox, U., Carver, B. S., Rodrik-outmezguine, V., Will, M., Yellen, P., Stanchina, E. De, Scher, H. I., Barry, S. T., & Sawyers, C. L. (2016). *Feedback supression of PI3Kα signaling in PTEN deficient tumors is relieved by selective inhibition of PI3K β*. *27*(1), 109–122. https://doi.org/10.1016/j.ccell.2014.11.008.Feedback

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape : A Software Environment for Integrated Models of Biomolecular Interaction Networks Cytoscape : A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, *Karp 2001*, 2498–2504. https://doi.org/10.1101/gr.1239303

Shim, J. S., & Liu, J. O. (2014). Recent advances in drug repositioning for the discovery of new anticancer drugs. *International Journal of Biological Sciences*, *10*(7), 654–663. https://doi.org/10.7150/ijbs.9224

Stark, C., Breitkreutz, B.-J., Reguly, T., Boucher, L., Breitkreutz, A., & Tyers, M. (2006). BioGRID: a general repository for interaction datasets. *Nucleic Acids Research*, *34*(Database issue), D535–D539. https://doi.org/10.1093/nar/gkj109

Su, G., Kuchinsky, A., Morris, J. H., States, D. J., & Meng, F. (2010). GLay: Community structure analysis of biological networks. *Bioinformatics*, *26*(24), 3135–3137. https://doi.org/10.1093/bioinformatics/btq596

Sun, B., & Karin, M. (2012). Review Obesity , inflammation , and liver cancer. *Journal of Hepatology*, *56*(3), 704–713. https://doi.org/10.1016/j.jhep.2011.09.020

Sun, W., & Cabrera, R. (2018). Systemic Treatment of Patients with Advanced, Unresectable Hepatocellular Carcinoma: Emergence of Therapies. *Journal of Gastrointestinal Cancer*, *49*(2), 107–115. https://doi.org/10.1007/s12029-018-0065-8

Suzuki, A., de la Pompa, J. L., Stambolic, V., Elia, A. J., Sasaki, T., del Barco Barrantes, I., Ho, A., Wakeham, A., Itie, A., Khoo, W., Fukumoto, M., & Mak, T. W. (1998). High cancer susceptibility and embryonic lethality associated with mutation of the PTEN tumor suppressor gene in mice. *Current Biology : CB*, *8*(21), 1169–1178. https://doi.org/10.1016/s0960-9822(07)00488-5

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K. P., Kuhn, M., Bork, P., Jensen, L. J., & Von Mering, C. (2015). STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research*, *43*(D1), D447–D452. https://doi.org/10.1093/nar/gku1003

Talevi, A., & Bellera, C. L. (2020). Challenges and opportunities with drug repurposing: finding strategies to find alternative uses of therapeutics. *Expert Opinion on Drug Discovery*, *15*(4), 397–401. https://doi.org/10.1080/17460441.2020.1704729

The HD iPSC Consortium. (2020). Bioenergetic deficits in Huntington's disease iPSC-derived neural cells and rescue with glycolytic metabolites. *Human Molecular Genetics*, *29*(11), 1757–1771. https://doi.org/10.1093/hmg/ddy430

Thomson, S., Petti, F., Sujka-Kwok, I., Mercado, P., Bean, J., Monaghan, M., Seymour, S. L., Argast, G. M., Epstein, D. M., & Haley, J. D. (2011). A systems view of epithelial-mesenchymal transition signaling states. *Clinical and Experimental Metastasis*, *28*(2), 137–155. https://doi.org/10.1007/s10585-010-9367-3

Tong, A. H., Evangelista, M., Parsons, A. B., Xu, H., Bader, G. D., Page, N., Robinson, M., Raghibizadeh, S., Hogue, C. W., Bussey, H., Andrews, B., Tyers, M., & Boone, C. (2001). Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science (New York, N.Y.)*, *294*(5550), 2364–2368. https://doi.org/10.1126/science.1065810

Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics*, *25*(9), 1105–1111. https://doi.org/10.1093/bioinformatics/btp120

Tu, K., Liu, Z., Yao, B., Xue, Y., Xu, M., Dou, C., Yin, G., & Wang, J. (2016). BCL-3 promotes the tumor growth of hepatocellular carcinoma by regulating cell proliferation and the cell cycle through cyclin D1. *Oncology Reports*, *35*(4), 2382–2390. https://doi.org/10.3892/or.2016.4616

Tuncbag, N., Braunstein, A., Pagnani, A., Huang, S. C., Chayes, J., Borgs, C., Zecchina, R., & Fraenkel, E. (2013). Simultaneous Reconstruction of Multiple Signaling Pathways via the Prize-Collecting Steiner Forest Problem. *Journal of Computational Biology*, *20*(2), 124–136. https://doi.org/10.1089/cmb.2012.0092

Tuncbag, N., Gosline, S. J. C., Kedaigle, A., Soltis, A. R., Gitter, A., & Fraenkel, E.

(2016). Network-Based Interpretation of Diverse High-Throughput Datasets through the Omics Integrator Software Package. *PLoS Computational Biology*, *12*(4). https://doi.org/10.1371/journal.pcbi.1004879

Tuncbag, N., Milani, P., Pokorny, J. L., Johnson, H., Sio, T. T., Dalin, S., Iyekegbe, D. O., White, F. M., Sarkaria, J. N., & Fraenkel, E. (2016). Network Modeling Identifies Patient-specific Pathways in Glioblastoma. *Scientific Reports*, *6*, 1–12. https://doi.org/10.1038/srep28668

Turner, B., Razick, S., Turinsky, A. L., Vlasblom, J., Crowdy, E. K., Cho, E., Morrison, K., Donaldson, I. M., & Wodak, S. J. (2010). iRefWeb: interactive analysis of consolidated protein interaction data and their supporting evidence. *Database : The Journal of Biological Databases and Curation*, *2010*, baq023. https://doi.org/10.1093/database/baq023

Vanhaesebroeck, B., Guillermet-Guibert, J., Graupera, M., & Bilanges, B. (2010). The emerging mechanisms of isoform-specific PI3K signalling. *Nature Reviews Molecular Cell Biology*, *11*(5), 329–341. https://doi.org/10.1038/nrm2882

Venturi, A., Piaz, F. D., Giovannini, C., Gramantieri, L., Chieco, P., & Bolondi, L. (2008). Human hepatocellular carcinoma expresses specific PCNA isoforms: an in vivo and in vitro evaluation. *Laboratory Investigation; a Journal of Technical Methods and Pathology*, *88*(9), 995–1007. https://doi.org/10.1038/labinvest.2008.50

Villanueva, A, & Llovet, J. (2013). Targeted Therapies for Hepatocellular Carcinoma. *Gastroenterology*, *140*(5), 1410–1426. https://doi.org/10.1053/j.gastro.2011.03.006.Targeted

Villanueva, Augusto, Chiang, D. Y., Newell, P., Peix, J., Thung, S., Alsinet, C., Tovar, V., Roayaie, S., Minguez, B., Sole, M., Battiston, C., Van Laarhoven, S., Fiel, M. I., Di Feo, A., Hoshida, Y., Yea, S., Toffanin, S., Ramos, A., Martignetti, J. A., … Llovet, J. M. (2008). Pivotal role of mTOR signaling in hepatocellular carcinoma. *Gastroenterology*, *135*(6), 1972–1983, 1983.e1-11. https://doi.org/10.1053/j.gastro.2008.08.008

Weigert, J., Neumeier, M., Bauer, S., Mages, W., Schnitzbauer, A. A., Obed, A., Gröschl, B., Hartmann, A., Schäffler, A., Aslanidis, C., Schölmerich, J., & Buechler, C. (2008). Small-interference RNA-mediated knock-down of aldehyde oxidase 1 in 3T3-L1 cells impairs adipogenesis and adiponectin release. *FEBS Letters*, *582*(19), 2965–2972. https://doi.org/10.1016/j.febslet.2008.07.034

Weilin, W., Qiang, S., Zehui, W., Dongkai, Z., Jianfeng, W., Haiyang, X., & Shusen, Z. (2013). Mitochondrial dysfunction-related genes in hepatocellular carcinoma Weilin. *Frontiers in Bioscince*, *18*, 1141–1149.

Werner, H. M. J., Mills, G. B., & Ram, P. T. (2014). Cancer Systems Biology: a peek into

the future of patient care? *Nature Reviews. Clinical Oncology*, *11*(3), 167–176. https://doi.org/10.1038/nrclinonc.2014.6

Wishart, D. S., Jewison, T., Guo, A. C., Wilson, M., Knox, C., Liu, Y., Djoumbou, Y., Mandal, R., Aziat, F., Dong, E., Bouatra, S., Sinelnikov, I., Arndt, D., Xia, J., Liu, P., Yallou, F., Bjorndahl, T., Perez-Pineiro, R., Eisner, R., … Scalbert, A. (2012). HMDB 3.0—The Human Metabolome Database in 2013. *Nucleic Acids Research*, *41*(D1), D801–D807. https://doi.org/10.1093/nar/gks1065

World Health Organization. (2015). *World Cancer Report 2014* (p. Chapter 5.6). https://doi.org/9283204298

Xenarios, I., Salwinski, L., Duan, X. J., Higney, P., Kim, S.-M., & Eisenberg, D. (2002). DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Research*, *30*(1), 303–305. https://doi.org/10.1093/nar/30.1.303

Xu, X., Liu, R.-F., Zhang, X., Huang, L.-Y., Chen, F., Fei, Q.-L., & Han, Z.-G. (2012). DLK1 as a Potential Target against Cancer Stem/Progenitor Cells of Hepatocellular Carcinoma. *Molecular Cancer Therapeutics*, *11*(3), 629–638. https://doi.org/10.1158/1535-7163.mct-11-0531

Yan, Y., & Marriott, G. (2003). Analysis of protein interactions using fluorescence technologies. *Current Opinion in Chemical Biology*, *7*(5), 635–640.

Yang, H. W., Shin, M.-G., Lee, S., Kim, J.-R., Park, W. S., Cho, K.-H., Meyer, T., & Heo, W. Do. (2012). Cooperative activation of PI3K by Ras and Rho family small GTPases. *Molecular Cell*, *47*(2), 281–290. https://doi.org/10.1016/j.molcel.2012.05.007

Yoshiji, H., Kuriyama, S., Yoshii, J., Ikenaka, Y., Noguchi, R., Hicklin, D. J., Huber, J., Nakatani, T., Tsujinoue, H., Yanase, K., Imazu, H., & Fukui, H. (2002). Synergistic effect of basic fibroblast growth factor and vascular endothelial growth factor in murine hepatocellular carcinoma. *Hepatology (Baltimore, Md.)*, *35*(4), 834–842. https://doi.org/10.1053/jhep.2002.32541

Yu, G., Kim, S., Park, S., Cho, B., Yeom, Y., Kim, S., Kim, S., Chu, I., & Kim, D. (2007). Identification of molecular markers for the oncogenic differentiation of hepatocellular carcinoma. *Experimental and Molecular Medicine*, *39*(5), 641–652.

Yu, J., Zhang, S. S., Saito, K., Williams, S., Arimura, Y., Ma, Y., Ke, Y., Baron, V., Mercola, D., Feng, G.-S., Adamson, E., & Mustelin, T. (2009). PTEN regulation by Akt-EGR1-ARF-PTEN axis. *The EMBO Journal*, *28*(1), 21–33. https://doi.org/10.1038/emboj.2008.238

Zhang, Li, Yang, Z., & Liu, Y. (2014a). GADD45 proteins: Roles in cellular senescence

and tumor development. *Experimental Biology and Medicine*, *239*(7), 773–778. https://doi.org/10.1177/1535370214531879

Zhang, Li, Yang, Z., & Liu, Y. (2014b). GADD45 proteins: roles in cellular senescence and tumor development. *Experimental Biology and Medicine (Maywood, N.J.)*, *239*(7), 773–778. https://doi.org/10.1177/1535370214531879

Zhang, Lijuan, Li, S., Wang, R., Chen, C., Ma, W., & Cai, H. (2019a). Cytokine augments the sorafenib-induced apoptosis in Huh7 liver cancer cell by inducing mitochondrial fragmentation and activating MAPK-JNK signalling pathway. *Biomedicine & Pharmacotherapy*, *110*(October 2018), 213–223. https://doi.org/10.1016/j.biopha.2018.11.037

Zhang, Lijuan, Li, S., Wang, R., Chen, C., Ma, W., & Cai, H. (2019b). Cytokine augments the sorafenib-induced apoptosis in Huh7 liver cancer cell by inducing mitochondrial fragmentation and activating MAPK-JNK signalling pathway. *Biomedicine & Pharmacotherapy = Biomedecine & Pharmacotherapie*, *110*, 213–223. https://doi.org/10.1016/j.biopha.2018.11.037

Zhang, Q. C., Petrey, D., Deng, L., Qiang, L., Shi, Y., Thu, C. A., Bisikirska, B., Lefebvre, C., Accili, D., Hunter, T., Maniatis, T., Califano, A., & Honig, B. (2012). Structure-based prediction of protein-protein interactions on a genome-wide scale. *Nature*, *490*(7421), 556–560. https://doi.org/10.1038/nature11503

Zhang, Y., Zhang, T., Wu, C., Xia, Q., & Xu, D. (2017). ASIC1a mediates the drug resistance of human hepatocellular carcinoma via the Ca2+/PI3-kinase/AKT signaling pathway. *Laboratory Investigation*, *97*(1), 53–69. https://doi.org/10.1038/labinvest.2016.127

Zhang, Z.-L., Liu, G., Peng, L., Zhang, C., Jia, Y.-M., Yang, W.-H., & Mao, L. (2018). Effect of PAK1 gene silencing on proliferation and apoptosis in hepatocellular carcinoma cell lines MHCC97-H and HepG2 and cells in xenograft tumor. *Gene Therapy*, *25*(4), 284–296. https://doi.org/10.1038/s41434-018-0016-9

Zhang, Z., Yin, J., Zhang, C., Liang, N., Bai, N., Chang, A., Liu, Y., Li, Z., Tan, X., Li, N., Lv, D., Xiang, R., Tian, Y., & Liu, C. (2012). Activating transcription factor 4 increases chemotherapeutics resistance of human hepatocellular carcinoma. *Cancer Biology and Therapy*, *13*(6), 435–442. https://doi.org/10.4161/cbt.19295

Zhao, L., & Vogt, P. K. (2010). Hot-spot mutations in p110alpha of phosphatidylinositol 3-kinase (pI3K): differential interactions with the regulatory subunit p85 and with RAS. *Cell Cycle (Georgetown, Tex.)*, *9*(3), 596–600. https://doi.org/10.4161/cc.9.3.10599

Zhou, G., & Xia, J. (2018). OmicsNet: a web-based tool for creation and visual analysis of biological networks in 3D space. *Nucleic Acids Research*, *46*(W1), W514–W522.

https://doi.org/10.1093/nar/gky510

APPENDIX A

EXPERIMENTAL FIGURES

Characterization of HCC cells in the presence of small molecules inhibitors were explained in Figure 8. Realtime cell growth analysis of Huh7 and Mahlavu cells with increasing concentrations (40µM, 20µM, 10µM, 5µM, 2.5µM) of Sorafenib, PI3K inhibitor LY294002, PI3Ki-β inhibitor (TGX-22) and PI3Ki-α (1µM, 0.5µM, 0.25µM, 0.125µM, 0.0625µM) PI3Ki-α (PIK-75) along with DMSO vehicle control (Control is black and increasing drug concentrations is given in grey level, highest concentration is being the darkest). These figures are published at (Narci et al., 2021).

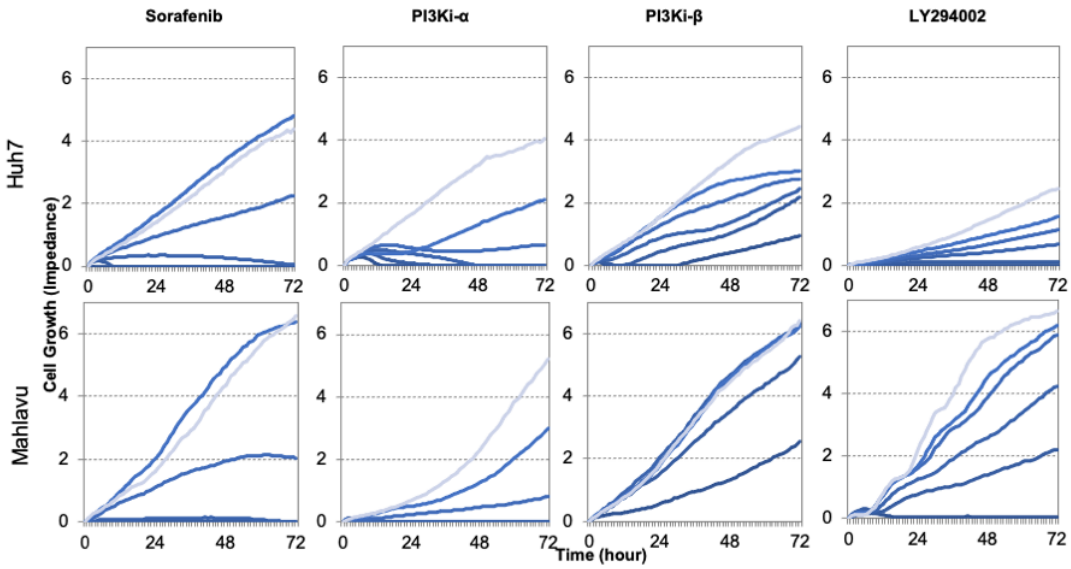


Figure 39. Realtime cell growth analysis of Huh7 and Mahlavu cells with increasing concentrations.

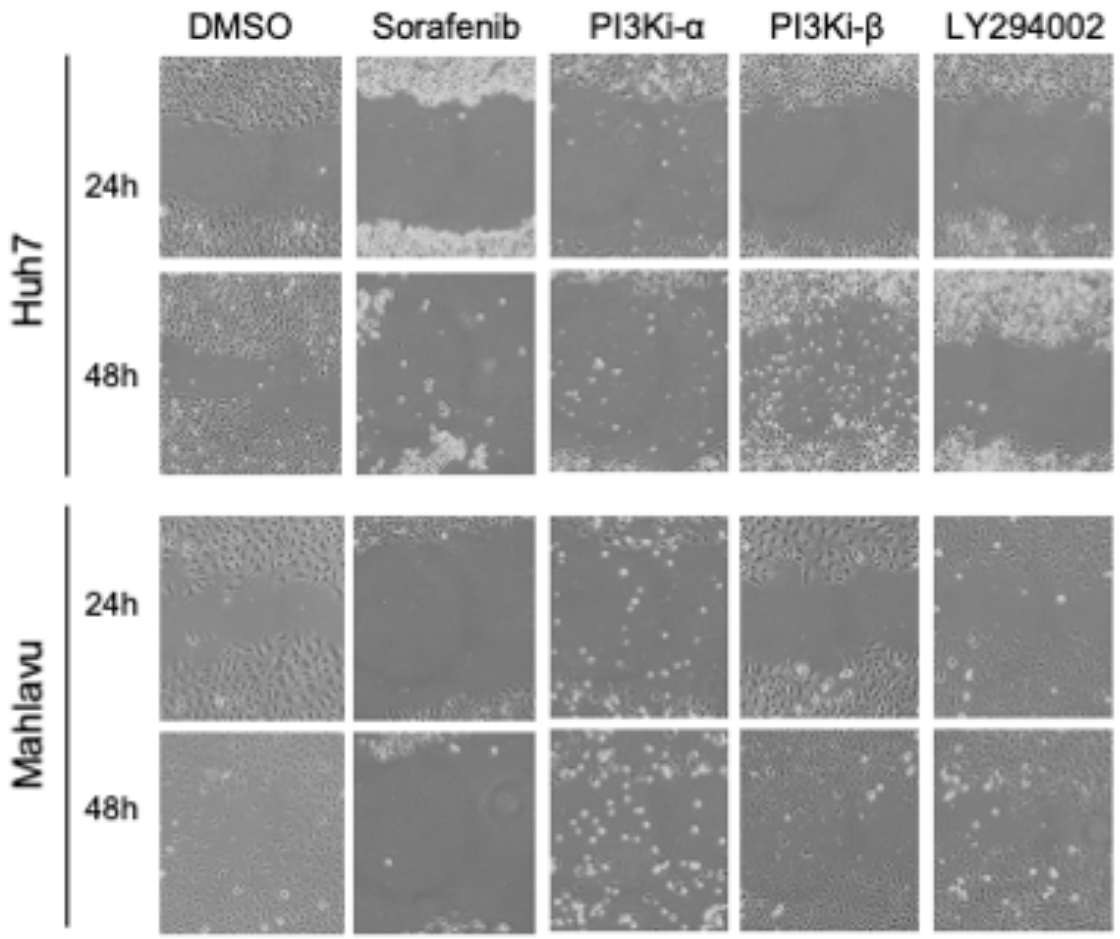


Figure 40. Wound healing assay for 24 and 48 hours for cell migration

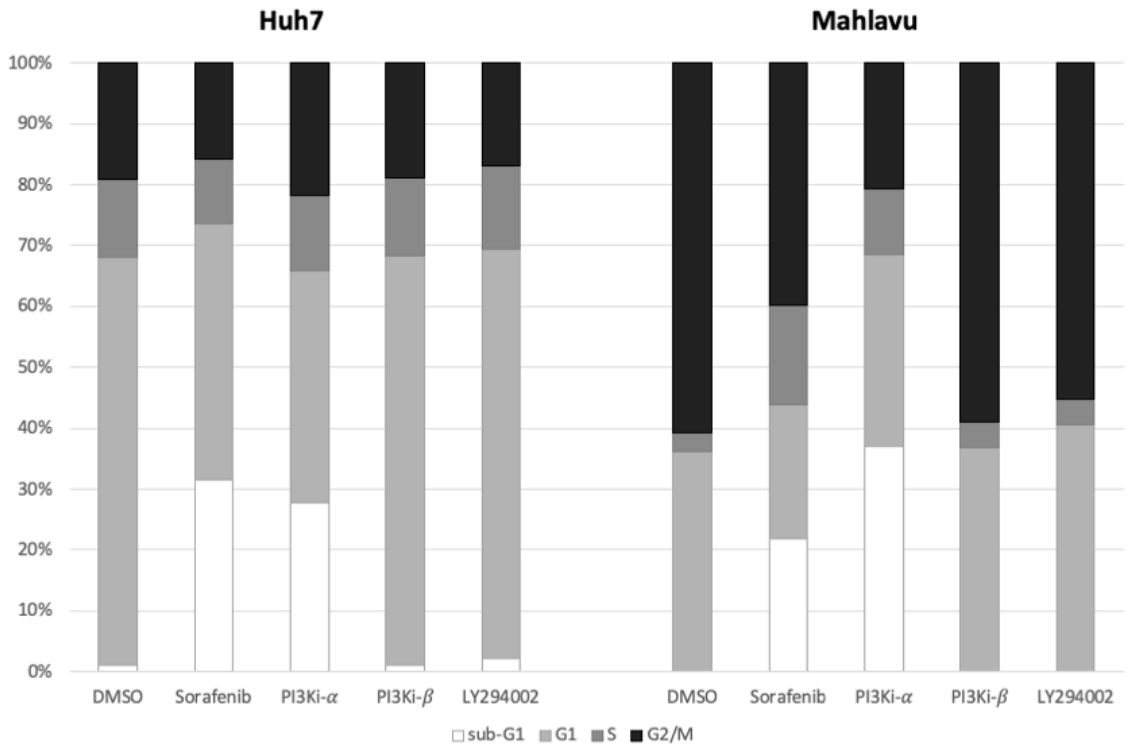


Figure 41. Cell cycle analysis with flow cytometry. Sub-G1population represents apoptotic cells. 10µM of Sorafenib, LY294002 and PI3Ki-β(TGX-221) and 0.1µM of PI3Ki-α(PIK-75) were used.

Real-time cell growth analysis was mentioned in Figure 10. Cell index measurements were obtained by RT-CES software. DMSO was used as negative control A. B. 72 hours of the percent growth inhibition values were used to calculate drug interactions with The SynergyFinder web application. Positive delta score reflects synergistic and negative score reflects antagonistic drug interactions. Experiments were performed in triplicate. Corresponding figures are published in (Narci et al., 2021).

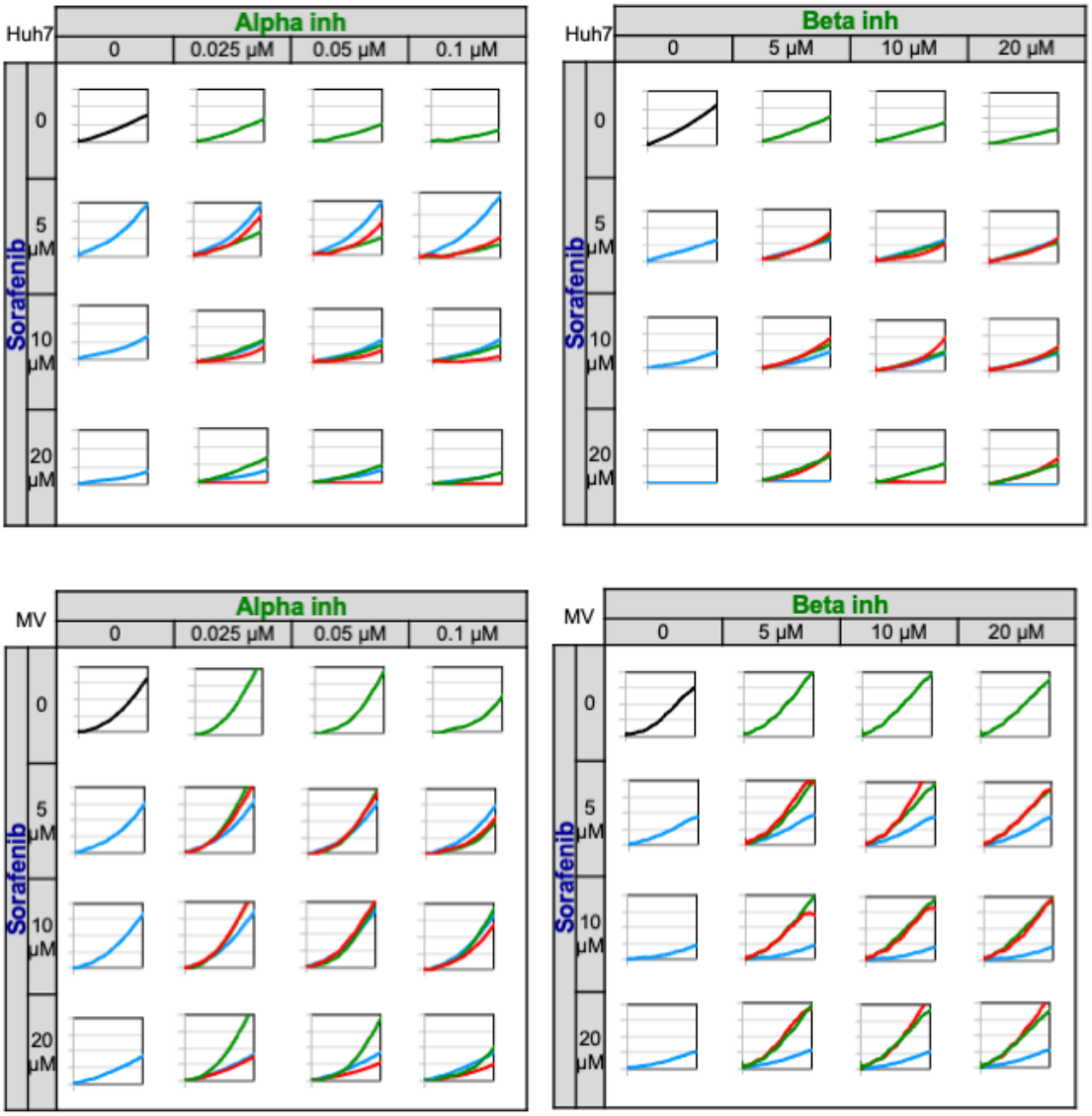


Figure 42. Real-time cell growth analysis. Human liver cancer cells Huh7 and Mahlavu (MV) were treated with the Sorafenib, PI3Ki-α and PI3Ki-β alone or in combination with increasing concentrations as indicated. Cell index measurements were obtained by RT-CES software. DMSO was used as negative control.

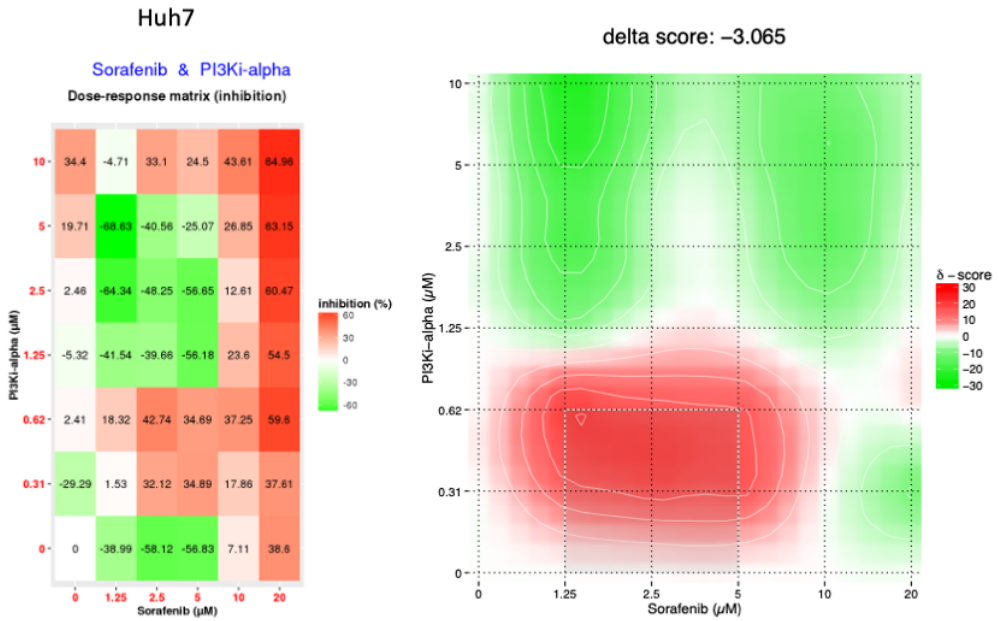


Figure 43. SynergyFinder view for Huh7 treated with Sorafenib and PI3Ki-α.

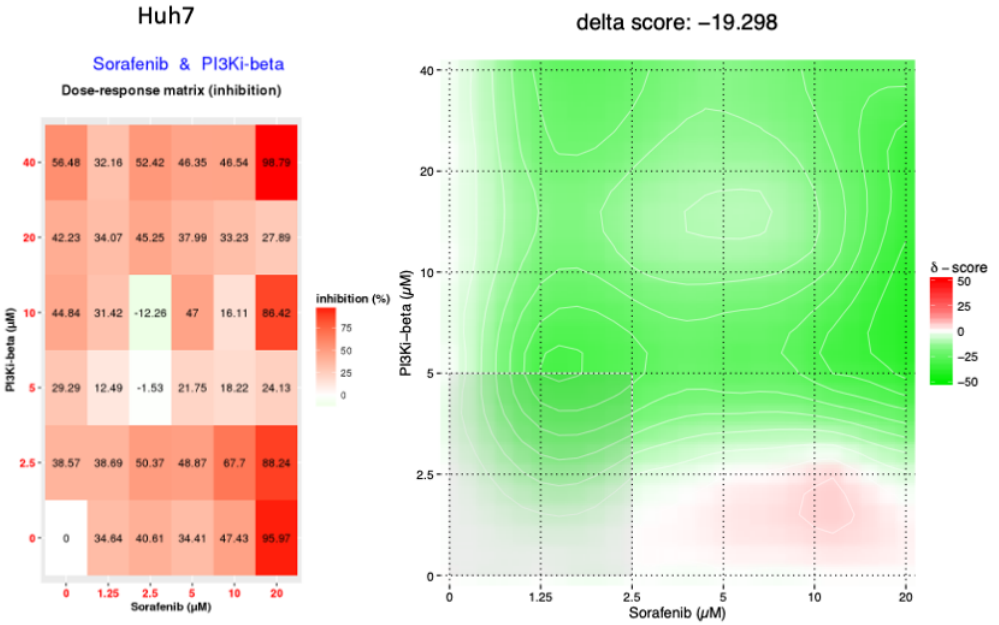


Figure 44. SynergyFinder view for Huh7 treated with Sorafenib and PI3Ki-β.

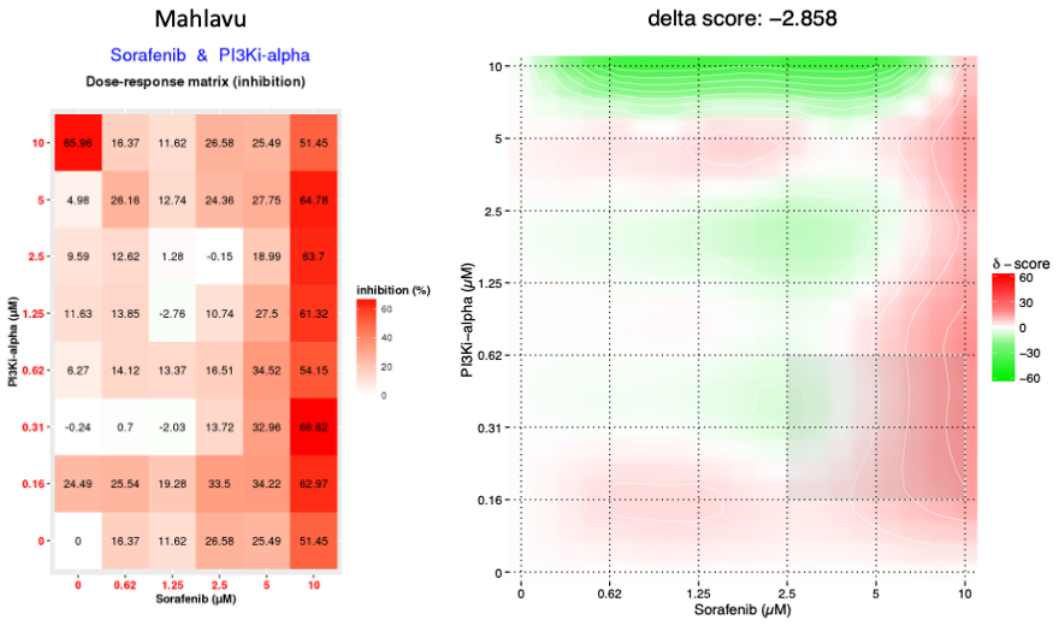


Figure 45. SynergyFinder view for Mahlavu treated with Sorafenib and PI3Ki-α.

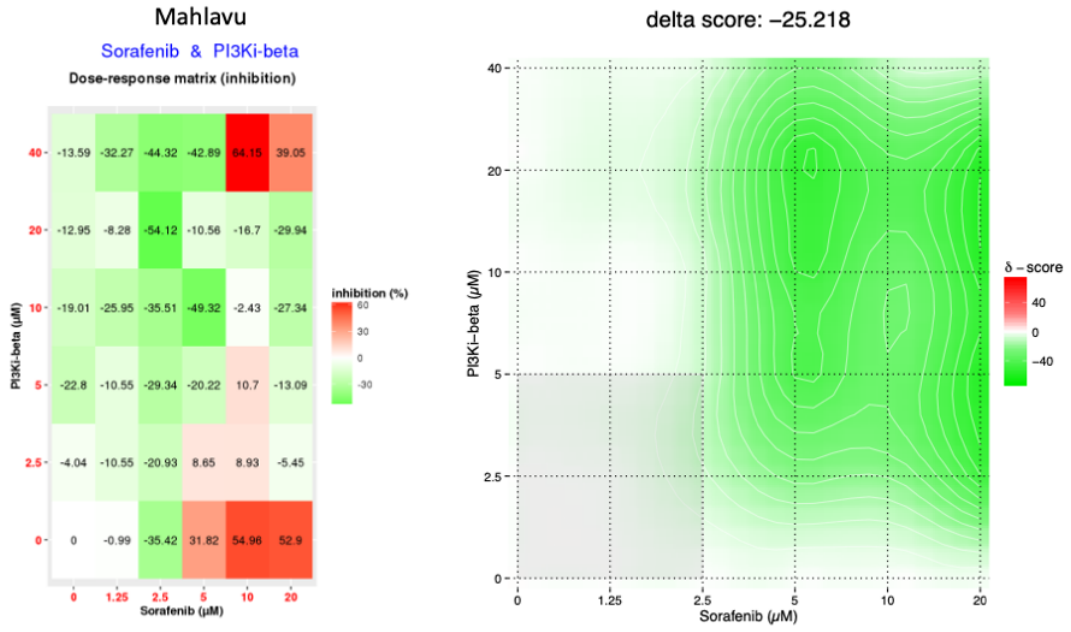


Figure 46. SynergyFinder view for Mahlavu treated with Sorafenib and PI3Ki-β.

APPENDIX B

GITHUB REPOSITORY

The raw RNA-seq files used in this study can be found in NCBI SRA with though PRJNA556552 id and the results of this study shared in CANSYL GitHub repository at; https://github.com/cansyl/Isoform-spesific-PI3K-inhibitor-analysis.

Codes: Includes the codes in this study.

- bingo_functions.r : reads a list of .bgo results from BINGO tool, merges, annotates and filters by evidence codes.
- common_dispersion.r: calculates common dispersion for Huh7 and Mahlavu RNA-seq differential expression analysis results using a list of housekeeping genes.
- correlation.r: Calculates and plots correlation between RNA-seq differential expression analysis results.
- edgeR_noRep.r: Given a count table including control vs sample expression values and filtration values for cpm, dispersion, FDR and P value Differentially Expressed Genes calculated and annotated and written into the output file.
- heatmaply.r: The heatmap function to draw heatmap dendrograms.
- network.py: Calculates centrality metrics for PCST networks using raw .sif files.
- prioritization.r: The code used to prioritize in the networks separately and visualize the nodes together using ggplot.

FASTQC reports: Includes FASTQC reports in zip files for each kinase inhibitors. HALPHA; PI3Ki-α inhibitor (PIK-75), HSALPHA; PIK-75 and Sorafenib, HBETA; PI3Ki-β inhibitor (TGX-221), HSBETA; TGX-221 and Sorafenib, HSOR; Sorafenib treatments of Huh7 cells and MALPHA; PI3Ki-α inhibitor (PIK-75), MSALPHA; PIK-75 and Sorafenib, MBETA; PI3Ki-β inhibitor (TGX-221), MSBETA; TGX-221 and Sorafenib, H-MSOR; Sorafenib treatments of Mahlavu cells.

Gene_Ontologies: Includes the excel file for gene ontologies found for each cluster for each treatment.

Heatmaply_images: The dendrograms created for this thesis study were made using heatmaply which offers interactive analysis though HTML generation. Produced HTML files were included in this directory.

Networks-SIF files: The network visualizations made though Cytoscape tool, the final network representations provided in the .cys formatted files provided in this directory. The files can be open using Cytoscape tool (File -> Open)

Required Files: Additional text formatted files required for running of the codes included.

APPENDIX C

COMMEND LINE AND PARAMETERS OF THE TOOLS

An example flow of the tools and parameters shown below.

1. RNA-Seq workflow:

```
fastqc  --noextract --outdir . --threads 1 input.fq.gz

bwa index
GRCh38_full_analysis_set_plus_decoy_hla.phiX174.fa ; tar
-cf
GRCh38_full_analysis_set_plus_decoy_hla.phiX174.fa.tar
GRCh38_full_analysis_set_plus_decoy_hla.phiX174.fa *.amb
*.ann *.bwt *.pac *.sa *.alt

tophat-2.1.0.Linux_x86_64/tophat2  --num-threads 8 --
output-dir ./tophat_out --no-coverage-search --keep-
fasta-order --GTF Homo_sapiens.GRCh38.84.gtf
./GRCh38_full_analysis_set_plus_decoy_hla.phiX174
input.fq.gz && mv tophat_out/align_summary.txt
tophat_out/input.fq.gz_align_summary.txt && mv
tophat_out/deletions.bed
tophat_out/input.fq.gz_deletions.bed && mv
tophat_out/insertions.bed
tophat_out/input.fq.gz_insertions.bed && mv
tophat_out/junctions.bed
tophat_out/input.fq.gz_junctions.bed && mv
tophat_out/accepted_hits.bam
tophat_out/input.fq.gz_accepted_hits.bam && mv
tophat_out/unmapped.bam
tophat_out/input.fq.gz_unmapped.bam

bamtools/bin/bamtools index -in
input.fq.gz_accepted_hits.bam

java -Xmx2048M -jar picard-tools-1.140/picard.jar
CollectAlignmentSummaryMetrics
INPUT=/TopHat2/tophat_out/input.fq.gz_accepted_hits.bam
```

```
OUTPUT=input.fq.gz_accepted_hits.summary_metrics.txt
REFERENCE_SEQUENCE=GRCh38_full_analysis_set_plus_decoy_hl
a.phiX174.fa VALIDATION_STRINGENCY=SILENT

htseq-count -f bam -r name -o
input.fq.gz_accepted_hits.annotated.sam
BamTools_Index/HDMSO.fq.gz_accepted_hits.bam
Homo_sapiens.GRCh38.84.gtf >
input.fq.gz_accepted_hits.table.txt

perl /edgeR-noReplicate/prepare_countMatrix.pl
control.fq.gz_accepted_hits.table.txt
input.fq.gz_accepted_hits.table.txt input_CountMatrix.txt

Rscript edgeR_predict_dispersion.r
boentempo_HK_genes_mart_export.txt count_matrix.txt >
analysis-output.txt

Rscript edgeR_noRep.r input_CountMatrix.txt 5 0.045 2 -2
0.05 input_CountMatrix_0.05_0.045_2.txt

Rscript heatmaply.r input_logFC_matrix.csv
```

2. Network construction workflow:

```
Rscript EdgeRfile2PCSTfile.r
input_CountMatrix_0.05_0.045_2.txt
input_045_2_PCSTfile.txt

python /forest-tuner.py --workingDir /home/ --forestPath
/OmicsIntegrator/scripts/forest.py --msgsteinerPath
/msgsteiner-1.3/msgsteiner --edgePath
/forest_interaction_filtbyzeropointseven_v2.txt --
prizePath input_CountMatrix_ALL_PCSTfile.txt --
outputsName input_CountMatrix_ALL_PCSTfile_output --
dataPath input_CountMatrix_ALL_PCSTfile_data.tsv --
logPath input_CountMatrix_ALL_PCSTfile_logs.log && zip -r
input_CountMatrix_ALL_PCSTfile_output.zip
/home/input_CountMatrix_ALL_PCSTfile_output/

python makeConf.py -w 7.75 -b 5.5 -D 10 -mu 0.04 -
garnetBeta undefined -noise undefined && python
```

```
/OmicsIntegrator/scripts/forest.py -p forest-
py_Prepare_PCSTinput/045_2_PCSTfile.txt -e
/forest_interaction_filtbyzeropointseven_v2.txt --msgpath
/msgsteiner-1.3/msgsteiner -c conf.txt --outlabel HALPHA
--shuffledPrizes 5

python makeConf.py -w 5 -b 7 -D 10 -mu 0.04 -garnetBeta
undefined -noise undefined && python
/OmicsIntegrator/scripts/forest.py -p /forest-
py_Prepare_PCSTinput/5_PCSTfile.txt -e
forest_interaction_filtbyzeropointseven_v2.txt --msgpath
/msgsteiner-1.3/msgsteiner -c conf.txt --outlabel
MSOR_1_5 --randomTerminals 100

python networkx.py -sif input_optimalForest.sif -o
input_networkx.txt

Rscript prioritization.r random_nodeattributes/
centrality_measures/
```

APPENDIX D

RNA-SEQ FASTQ QUALITY REPORTS

The quality scores across all bases of raw reads for Huh7 and Mahlavu cells; DMSO, PIK-75, TGX-221, Sorafenib, PIK-75 and Sorafenib combined and TGX-221 and Sorafenib combined treatments were represented in this section.

Huh7 Mahlavu

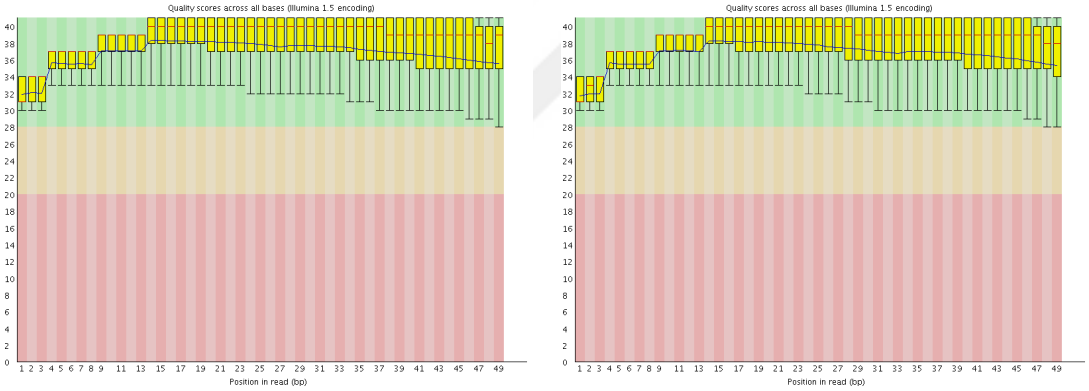


Figure 47: Quality score plot for DMSO treated Huh7 and Mahlavu cells

Huh7 Mahlavu

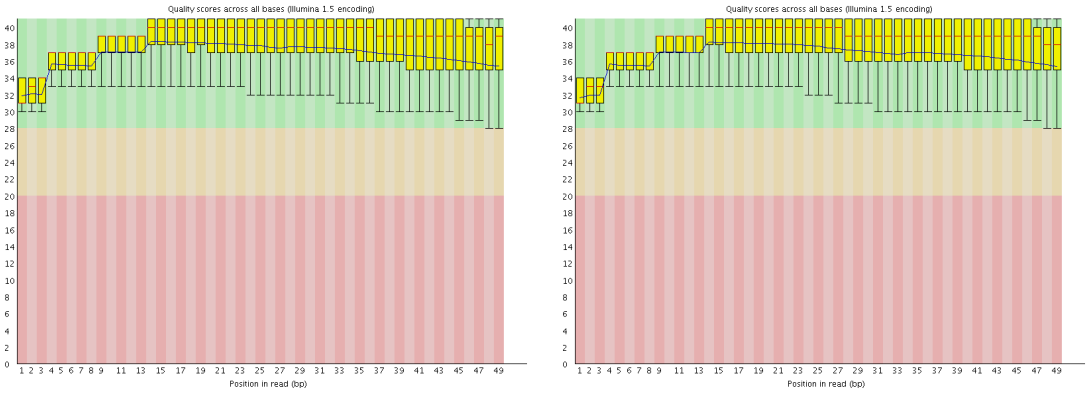


Figure 48: Quality score plots for PIK-75 treated Huh7 and Mahlavu cells

Huh7 Mahlavu

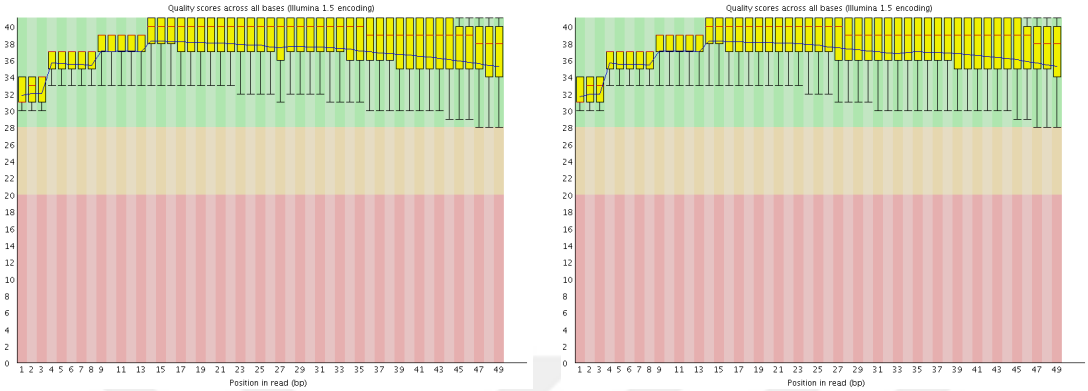


Figure 49: Quality score plots for TGX-221 treated Huh7 and Mahlavu cells

Huh7 Mahlavu

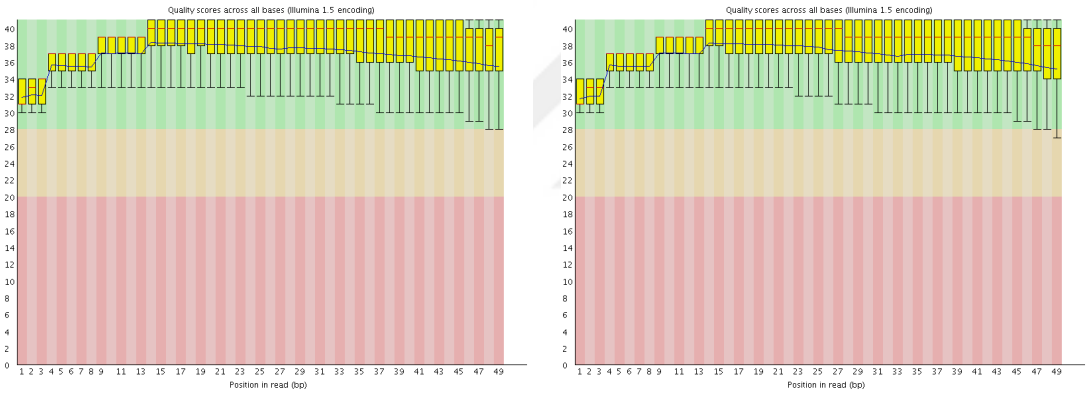


Figure 50: Quality score plots for Sorafenib treated Huh7 and Mahlavu cells

Huh7 Mahlavu

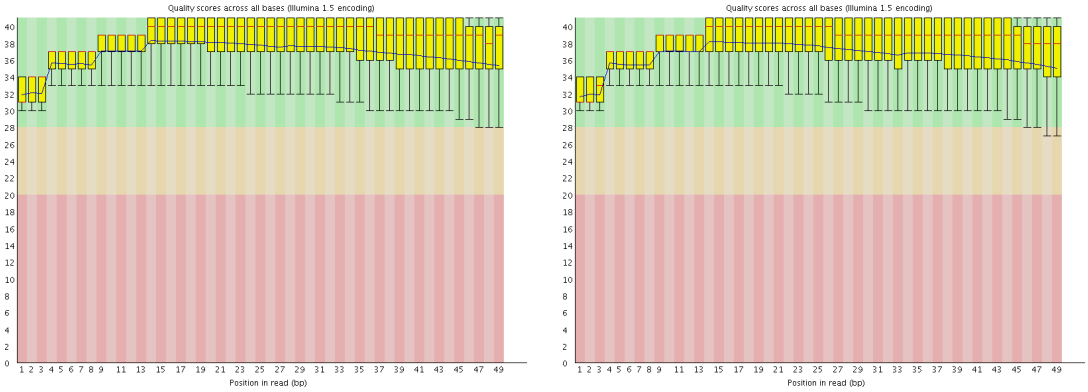


Figure 51: Quality score plots for combinational treatment of PIK-75 and Sorafenib to Huh7 and Mahlavu cells

Huh7 Mahlavu

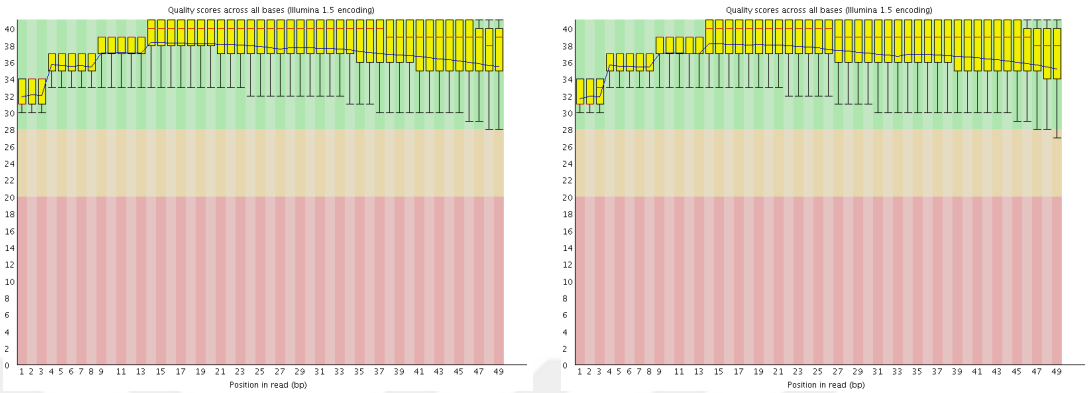


Figure 52: Quality scores for combinational treatment of TGX-221 and Sorafenib to Huh7 and Mahlavu cells

APPENDIX E

NETWORK CLUSTER GENE ENRICHMENTS

In this section, the cluster specific gene enrichments for kinase inhibitor treated Huh7 and Mahlavu cells were listed. The pathway gene set enrichment analysis was performed using BiNGO. Please see the methods section to find how the analyses were carried out. In the table which column shows the category of the enrichment: F: Cellular Function, P: Cellular Process, and C: Cellular Component.

Table 16: Cluster specific gene ontologies for PIK-75 treated Huh7 cells

GOID	Description	which	Evidence	Cluster
GO:0043565	sequence-specific DNA binding	F	IDA	11
GO:0030154	cell differentiation	P	IDA	11
GO:0000307	cyclin-dependent protein kinase holoenzyme complex	C	IDA	11
GO:0001701	in utero embryonic development	P	IEP	10
GO:0030492	hemoglobin binding	F	IDA	10
GO:0009653	anatomical structure morphogenesis	P	IMP	10
GO:0050801	ion homeostasis	P	IMP	10
GO:0007399	nervous system development	P	IMP	10
GO:0030003	cellular cation homeostasis	P	IDA	10
GO:0006333	chromatin assembly or disassembly	P	IMP	14
GO:0071103	DNA conformation change	P	IDA	14
GO:0006301	postreplication repair	P	IDA	14
GO:0008203	cholesterol metabolic process	P	IDA	14
GO:0005777	peroxisome	C	IDA	14
GO:0008299	isoprenoid biosynthetic process	P	IDA	14
GO:0034622	cellular macromolecular complex assembly	P	IDA	14
GO:0002455	humoral immune response mediated by circulating immunoglobulin	P	IMP	7
GO:0051604	protein maturation	P	IDA	7

Table 16: Cluster specific gene ontologies for PIK-75 treated Huh7 cells (continued)

GO:0002541	activation of plasma proteins involved in acute inflammatory response	P	IDA	7
GO:0002250	adaptive immune response	P	IMP	7
GO:0002253	activation of immune response	P	IDA	7
GO:0000152	nuclear ubiquitin ligase complex	C	IDA	4
GO:0005680	anaphase-promoting complex	C	IDA	4
GO:0031401	positive regulation of protein modification process	P	IDA	4
GO:0070695	FHF complex	C	IDA	4
GO:0030247	polysaccharide binding	F	IDA	12
GO:0005539	glycosaminoglycan binding	F	IDA	12
GO:0009611	response to wounding	P	IMP	12,16
GO:0006953	acute-phase response	P	IEP	12
GO:0030193	regulation of blood coagulation	P	IDA	12
GO:0061041	regulation of wound healing	P	IDA	12
GO:0006952	defense response	P	IDA	16
GO:0050716	positive regulation of interleukin-1 secretion	P	IDA	16
GO:0050718	positive regulation of interleukin-1 beta secretion	P	IDA	16
GO:0033198	response to ATP	P	IDA	16
GO:0045087	innate immune response	P	IMP	16
GO:0000186	activation of MAPKK activity	P	IMP	16
GO:0006139	nucleobase-containing compound metabolic process	P	IDA	13
GO:0016226	iron-sulfur cluster assembly	P	IDA	13
GO:0006418	tRNA aminoacylation for protein translation	P	IMP	1
GO:0034645	cellular macromolecule biosynthetic process	P	IGI	1
GO:0008156	negative regulation of DNA replication	P	IMP	1
GO:0006520	cellular amino acid metabolic process	P	IDA	1
GO:0005388	calcium-transporting ATPase activity	F	IDA	6
GO:0015085	calcium ion transmembrane transporter activity	F	IDA	6
GO:0016462	pyrophosphatase activity	F	IDA	6
GO:0016486	peptide hormone processing	P	IDA	6
GO:0017111	nucleoside-triphosphatase activity	F	IDA	6

Table 16: Cluster specific gene ontologies for PIK-75 treated Huh7 cells (continued)

GO:0005923	tight junction	C	IDA	5
GO:0070160	occluding junction	C	IDA	5
GO:0016327	apicolateral plasma membrane	C	IDA	5
GO:0043296	apical junction complex	C	IDA	5
GO:0001726	ruffle	C	IMP	5
GO:0043407	negative regulation of MAP kinase activity	P	IMP	3
GO:0070373	negative regulation of ERK1 and ERK2 cascade	P	IDA	3
GO:0000188	inactivation of MAPK activity	P	IMP	3
GO:0007623	circadian rhythm	P	IMP	2

Table 17: Cluster specific gene ontologies for TGX-221 treated Huh7 cells

GOID	**Description**	**which**	**Evidence**	**Cluster**
GO:0006621	protein retention in ER lumen	P	IMP	1
GO:0035437	maintenance of protein localization in endoplasmic reticulum	P	IMP	1
GO:0003756	protein disulfide isomerase activity	F	IDA	1
GO:0045454	cell redox homeostasis	P	IDA	1
GO:0003865	3-oxo-5-alpha-steroid 4-dehydrogenase activity	F	IDA	3
GO:0006740	NADPH regeneration	P	IMP	3
GO:0008202	steroid metabolic process	P	IDA	3
GO:0009055	electron carrier activity	F	IDA	3
GO:0016052	carbohydrate catabolic process	P	IDA	3
GO:0016229	steroid dehydrogenase activity	F	IDA	3
GO:0016491	oxidoreductase activity	F	IDA	3
GO:0017057	6-phosphogluconolactonase activity	F	IDA	3
GO:0019322	pentose biosynthetic process	P	IDA	3
GO:0044275	cellular carbohydrate catabolic process	P	IDA	3
GO:0055114	oxidation-reduction process	P	IDA	3

Table 18: Cluster specific gene ontologies for PIK-75 and Sorafenib treated Huh7 cells

GOID	**Description**	**Which**	**Evidence**	**Cluster**
GO:0007026	negative regulation of microtubule depolymerization	P	IMP	19

Table 18: Cluster specific gene ontologies for PIK-75 and Sorafenib treated Huh7 cells (continued)

GO:0031111	negative regulation of microtubule polymerization or depolymerization	P	IMP	19
GO:0000188	inactivation of MAPK activity	P	IDA	17
GO:0043407	negative regulation of MAP kinase activity	P	IDA	17
GO:0000165	MAPK cascade	P	IDA	17
GO:0016788	hydrolase activity, acting on ester bonds	F	IMP	17
GO:0008203	cholesterol metabolic process	P	IDA	16
GO:0005506	iron ion binding	F	IMP	16
GO:0008202	steroid metabolic process	P	IMP	16
GO:0004506	squalene monooxygenase activity	F	IDA	16
GO:0008398	sterol 14-demethylase activity	F	IDA	16
GO:0022618	ribonucleoprotein complex assembly	P	IDA	15
GO:0000375	RNA splicing, via transesterification reactions	P	IDA	15
GO:0018024	histone-lysine N-methyltransferase activity	F	IMP	15
GO:0000245	spliceosomal complex assembly	P	IMP	15
GO:0042054	histone methyltransferase activity	F	IDA	15
GO:0006974	response to DNA damage stimulus	P	IDA	14
GO:0032777	Piccolo NuA4 histone acetyltransferase complex	C	IDA	14
GO:0006301	postreplication repair	P	IDA	14
GO:0008283	cell proliferation	P	IMP	14
GO:0080008	CUL4 RING ubiquitin ligase complex	C	IMP	14
GO:0006282	regulation of DNA repair	P	IDA	14
GO:0006284	base-excision repair	P	IDA	14
GO:0007178	transmembrane receptor protein serine/threonine kinase signaling pathway	P	IDA	9
GO:0046332	SMAD binding	F	IPI	9
GO:0005391	sodium:potassium-exchanging ATPase activity	F	IDA	8
GO:0005890	sodium:potassium-exchanging ATPase complex	C	IDA	8
GO:0030433	ER-associated protein catabolic process	P	IMP	7
GO:0051787	misfolded protein binding	F	IPI	7
GO:0004571	mannosyl-oligosaccharide 1,2-alpha-mannosidase activity	F	IDA	7

Table 18: Cluster specific gene ontologies for PIK-75 and Sorafenib treated Huh7 cells (continued)

GO:0005788	endoplasmic reticulum lumen	C	IDA	7
GO:0003700	sequence-specific DNA binding transcription factor activity	F	IDA	6
GO:0034976	response to endoplasmic reticulum stress	P	IMP	6
GO:0006520	cellular amino acid metabolic process	P	IDA	6
GO:0071310	cellular response to organic substance	P	IDA	6
GO:0016255	attachment of GPI anchor to protein	P	IMP	12
GO:0006506	GPI anchor biosynthetic process	P	IDA	12
GO:0008654	phospholipid biosynthetic process	P	IDA	12
GO:0006650	glycerophospholipid metabolic process	P	IDA	12
GO:0008408	3'-5' exonuclease activity	F	IDA	12
GO:0042254	ribosome biogenesis	P	IMP	12
GO:0060738	epithelial-mesenchymal signaling involved in prostate gland development	P	IDA	11
GO:0043627	response to estrogen stimulus	P	IDA	11
GO:0014902	myotube differentiation	P	IMP	11
GO:0009725	response to hormone stimulus	P	IDA	11
GO:0008518	reduced folate carrier activity	F	IDA	10
GO:0046483	heterocycle metabolic process	P	IDA	10
GO:0051173	positive regulation of nitrogen compound metabolic process	P	IGI	10
GO:0031328	positive regulation of cellular biosynthetic process	P	IDA	10

Table 19: Cluster specific gene ontologies for TGX-221 and Sorafenib treated Huh7 cells

GOID	**Description**	**Which**	**Evidence**	**Cluster**
GO:0043296	apical junction complex	C	IDA	9
GO:0005911	cell-cell junction	C	IDA	9
GO:0030165	PDZ domain binding	F	IPI	9
GO:0000188	inactivation of MAPK activity	P	IMP	6
GO:0006469	negative regulation of protein kinase activity	P	IDA	6
GO:0017017	MAP kinase tyrosine/serine/threonine phosphatase activity	F	IDA	6
GO:0033549	MAP kinase phosphatase activity	F	IDA	6

Table 19: Cluster specific gene ontologies for TGX-221 and Sorafenib treated Huh7 cells (continued)

GO:0033673	negative regulation of kinase activity	P	IDA	6
GO:0005741	mitochondrial outer membrane	C	IDA	4
GO:0016298	lipase activity	F	IDA	4
GO:0005975	carbohydrate metabolic process	P	IDA	4
GO:0006631	fatty acid metabolic process	P	IDA	4
GO:0006641	triglyceride metabolic process	P	IMP	4
GO:0004864	protein phosphatase inhibitor activity	F	IDA	3
GO:0019208	phosphatase regulator activity	F	IDA	3
GO:0019888	protein phosphatase regulator activity	F	IDA	3
GO:0016192	vesicle-mediated transport	P	IMP	16, 20
GO:0048193	Golgi vesicle transport	P	IMP	20
GO:0030127	COPII vesicle coat	C	IDA	20
GO:0012507	ER to Golgi transport vesicle membrane	C	IDA	20
GO:0030282	bone mineralization	P	IDA	2
GO:0031214	biomineral tissue development	P	IMP	2
GO:0060348	bone development	P	IMP	2
GO:0001501	skeletal system development	P	IMP	2, 17
GO:0030154	cell differentiation	P	IDA	18
GO:0032502	developmental process	P	IMP	18
GO:0051387	negative regulation of nerve growth factor receptor signaling pathway	P	IMP	18
GO:0040037	negative regulation of fibroblast growth factor receptor signaling pathway	P	IMP	18
GO:0042127	regulation of cell proliferation	P	IDA	18
GO:0070373	negative regulation of ERK1 and ERK2 cascade	P	IMP	18
GO:0043410	positive regulation of MAPK cascade	P	IMP	18
GO:0008284	positive regulation of cell proliferation	P	IDA	18
GO:0046580	negative regulation of Ras protein signal transduction	P	IDA	18
GO:0051058	negative regulation of small GTPase mediated signal transduction	P	IMP	18
GO:0030099	myeloid cell differentiation	P	IDA	18
GO:0048705	skeletal system morphogenesis	P	IMP	17
GO:0060017	parathyroid gland development	P	IMP	17
GO:0035196	production of miRNAs involved in gene silencing by miRNA	P	IMP	17

Table 19: Cluster specific gene ontologies for TGX-221 and Sorafenib treated Huh7 cells (continued)

GO:0034189	very-low-density lipoprotein particle binding	F	IDA	16
GO:0010324	membrane invagination	P	IMP	16
GO:0034185	apolipoprotein binding	F	IPI	16
GO:0030662	coated vesicle membrane	C	IDA	16
GO:0021766	hippocampus development	P	IMP	16
GO:0033257	Bcl3/NF-kB2 complex	C	IDA	14
GO:0002250	adaptive immune response	P	IEP	14
GO:0002286	T cell activation involved in immune response	P	IDA	14
GO:0042752	regulation of circadian rhythm	P	IMP	14
GO:0002366	leukocyte activation involved in immune response	P	IDA	14
GO:0002520	immune system development	P	IMP	14
GO:0007623	circadian rhythm	P	IDA	14
GO:0008080	N-acetyltransferase activity	F	IDA	13
GO:0016407	acetyltransferase activity	F	IDA	13
GO:0016410	N-acyltransferase activity	F	IDA	13
GO:0006474	N-terminal protein amino acid acetylation	P	IDA	13
GO:0031365	N-terminal protein amino acid modification	P	IDA	13
GO:0016740	transferase activity	F	IDA	13
GO:0001836	release of cytochrome c from mitochondria	P	IMP	11
GO:0008637	apoptotic mitochondrial changes	P	IDA	11
GO:0010941	regulation of cell death	P	IMP	11
GO:0042981	regulation of apoptotic process	P	IGI	11
GO:0043067	regulation of programmed cell death	P	IDA	11
GO:0019104	DNA N-glycosylase activity	F	IDA	11
GO:0007257	activation of JUN kinase activity	P	IDA	10
GO:0043507	positive regulation of JUN kinase activity	P	IMP	10
GO:0043406	positive regulation of MAP kinase activity	P	IDA	10
GO:0006970	response to osmotic stress	P	IMP	10
GO:0008360	regulation of cell shape	P	IDA	10
GO:0000186	activation of MAPKK activity	P	IDA	10
GO:0022604	regulation of cell morphogenesis	P	IMP	10

Table 20: Cluster specific gene ontologies for Sorafenib treated Huh7 cells

GOID	Description	Which	Evidence	Cluster
GO:0043407	negative regulation of MAP kinase activity	P	IDA	14
GO:0034260	negative regulation of GTPase activity	P	IMP	14
GO:0051387	negative regulation of nerve growth factor receptor signaling pathway	P	IMP	14
GO:0040037	negative regulation of fibroblast growth factor receptor signaling pathway	P	IDA	14
GO:0051386	regulation of nerve growth factor receptor signaling pathway	P	IMP	14
GO:0009653	anatomical structure morphogenesis	P	IMP	4, 14
GO:0032502	developmental process	P	IMP	4, 14
GO:0070373	negative regulation of ERK1 and ERK2 cascade	P	IDA	14
GO:0051058	negative regulation of small GTPase mediated signal transduction	P	IMP	14
GO:0003824	catalytic activity	F	IDA	1, 3
GO:0046033	AMP metabolic process	P	IDA	1
GO:0006796	phosphate-containing compound metabolic process	P	IMP	9
GO:0043687	post-translational protein modification	P	IMP	9
GO:0008138	protein tyrosine/serine/threonine phosphatase activity	F	IDA	9
GO:0044267	cellular protein metabolic process	P	IDA	9
GO:0006470	protein dephosphorylation	P	IMP	9
GO:0000287	magnesium ion binding	F	IDA	9
GO:0016773	phosphotransferase activity, alcohol group as acceptor	F	IDA	6
GO:0003756	protein disulfide isomerase activity	F	IDA	8
GO:0045216	cell-cell junction organization	P	IMP	8
GO:0006621	protein retention in ER lumen	P	IMP	8
GO:0035437	maintenance of protein localization in endoplasmic reticulum	P	IMP	8
GO:0034329	cell junction assembly	P	IMP	8
GO:0045454	cell redox homeostasis	P	IDA	8

Table 20: Cluster specific gene ontologies for Sorafenib treated Huh7 cells (continued)

GO:0006974	response to DNA damage stimulus	P	IDA	7
GO:0071501	cellular response to sterol depletion	P	IDA	7
GO:0043240	Fanconi anaemia nuclear complex	C	IDA	7
GO:0032933	SREBP-mediated signaling pathway	P	IMP	7
GO:0002102	podosome	C	IDA	6
GO:0046777	protein autophosphorylation	P	IDA	6
GO:0005913	cell-cell adherens junction	C	IDA	4
GO:0030154	cell differentiation	P	IDA	4
GO:0006694	steroid biosynthetic process	P	IDA	3
GO:0006695	cholesterol biosynthetic process	P	IDA	3
GO:0044255	cellular lipid metabolic process	P	IDA	3
GO:0055114	oxidation-reduction process	P	IDA	3
GO:0016491	oxidoreductase activity	F	IDA	3
GO:0032787	monocarboxylic acid metabolic process	P	IDA	3
GO:0016628	oxidoreductase activity, acting on the CH-CH group of donors, NAD or NADP as acceptor	F	IDA	3
GO:0015030	Cajal body	C	IDA	16
GO:0015031	protein transport	P	IMP	15
GO:0045184	establishment of protein localization	P	IDA	15
GO:0008104	protein localization	P	IMP	15
GO:0003924	GTPase activity	F	IDA	15
GO:0016192	vesicle-mediated transport	P	IDA	15
GO:0033036	macromolecule localization	P	IDA	15
GO:0006879	cellular iron ion homeostasis	P	IMP	12
GO:0050750	low-density lipoprotein particle receptor binding	F	IDA	12
GO:0042157	lipoprotein metabolic process	P	IMP	12
GO:0001954	positive regulation of cell-matrix adhesion	P	IMP	12
GO:0042592	homeostatic process	P	IMP	12
GO:0032872	regulation of stress-activated MAPK cascade	P	IMP	10
GO:0005248	voltage-gated sodium channel activity	F	IDA	10
GO:0043666	regulation of phosphoprotein phosphatase activity	P	IMP	10

Table 21: Cluster specific gene ontologies for PIK-75 treated Mahlavu cells

GOID	Description	Which	Evidence	Cluster
GO:0016272	prefoldin complex	C	IDA	9
GO:0005829	cytosol	C	IDA	9
GO:0047710	bis(5'-adenosyl)-triphosphatase activity	F	IDA	9
GO:0031941	filamentous actin	C	IDA	8
GO:0005884	actin filament	C	IDA	8
GO:0015629	actin cytoskeleton	C	IDA	8
GO:0005179	hormone activity	F	IDA	5
GO:0006260	DNA replication	P	IGI	4
GO:0006259	DNA metabolic process	P	IDA	4
GO:0006325	chromatin organization	P	IMP	11
GO:0051276	chromosome organization	P	IDA	11
GO:0000785	chromatin	C	IDA	11
GO:0004965	G-protein coupled GABA receptor activity	F	IDA	10
GO:0007214	gamma-aminobutyric acid signaling pathway	P	IDA	10
GO:0016917	GABA receptor activity	F	IMP	10
GO:0030054	cell junction	C	IDA	10
GO:0008066	glutamate receptor activity	F	IDA	10
GO:0032376	positive regulation of cholesterol transport	P	IDA	1
GO:0032374	regulation of cholesterol transport	P	IDA	1
GO:0032488	Cdc42 protein signal transduction	P	IMP	1
GO:0090084	negative regulation of inclusion body assembly	P	IDA	1
GO:0034380	high-density lipoprotein particle assembly	P	IDA	1
GO:0065005	protein-lipid complex assembly	P	IMP	1
GO:0070325	lipoprotein particle receptor binding	F	IPI	1
GO:0008285	negative regulation of cell proliferation	P	IDA	1
GO:0043691	reverse cholesterol transport	P	IDA	1
GO:0001558	regulation of cell growth	P	IEP	1
GO:0007266	Rho protein signal transduction	P	IMP	1
GO:0042632	cholesterol homeostasis	P	IMP	1

Table 22: Cluster specific gene ontologies for TGX-221 treated Mahlavu cells

GOID	Description	Which	Evidence	Cluster
GO:0002237	response to molecule of bacterial origin	P	IDA	3
GO:0042742	defense response to bacterium	P	IMP	3
GO:0001932	regulation of protein phosphorylation	P	IGI	3
GO:0009617	response to bacterium	P	IDA	3
GO:0032755	positive regulation of interleukin-6 production	P	IDA	3
GO:0050830	defense response to Gram-positive bacterium	P	IDA	3
GO:0032675	regulation of interleukin-6 production	P	IMP	3
GO:0007420	brain development	P	IMP	2
GO:0031016	pancreas development	P	IEP	2
GO:0004180	carboxypeptidase activity	F	IDA	5
GO:0004181	metallocarboxypeptidase activity	F	IDA	5
GO:0008235	metalloexopeptidase activity	F	IDA	5

Table 23: Cluster specific gene ontologies for PIK-75 and Sorafenib treated Mahlavu cells

GOID	Description	Which	Evidence	Cluster
GO:0004558	alpha-glucosidase activity	F	IDA	10
GO:0046483	heterocycle metabolic process	P	IDA	9
GO:0015085	calcium ion transmembrane transporter activity	F	IDA	9
GO:0000421	autophagic vacuole membrane	C	IDA	9
GO:0016327	apicolateral plasma membrane	C	IDA	8
GO:0043296	apical junction complex	C	IDA	8
GO:0005911	cell-cell junction	C	IDA	4, 8
GO:0008509	anion transmembrane transporter activity	F	IDA	5
GO:0005452	inorganic anion exchanger activity	F	IDA	5
GO:0015301	anion:anion antiporter activity	F	IDA	5
GO:0016323	basolateral plasma membrane	C	IDA	5
GO:0004965	G-protein coupled GABA receptor activity	F	IDA	4
GO:0030695	GTPase regulator activity	F	IDA	4

Table 23: Cluster specific gene ontologies for PIK-75 and Sorafenib treated Mahlavu cells (continued)

GO:0005355	glucose transmembrane transporter activity	F	IDA	31
GO:0070776	MOZ/MORF histone acetyltransferase complex	C	IDA	3
GO:0016592	mediator complex	C	IDA	3
GO:0010887	negative regulation of cholesterol storage	P	IDA	3
GO:0010875	positive regulation of cholesterol efflux	P	IDA	3
GO:0017127	cholesterol transporter activity	F	IDA	3
GO:0032376	positive regulation of cholesterol transport	P	IDA	3
GO:0006414	translational elongation	P	IDA	29
GO:0022627	cytosolic small ribosomal subunit	C	IDA	29
GO:0001775	cell activation	P	IDA	28
GO:0001819	positive regulation of cytokine production	P	IDA	28
GO:0016174	NAD(P)H oxidase activity	F	IDA	28
GO:0032727	positive regulation of interferon-alpha production	P	IDA	28
GO:0006952	defense response	P	IDA	28
GO:0001774	microglial cell activation	P	IDA	28
GO:0032481	positive regulation of type I interferon production	P	IDA	28
GO:0006885	regulation of pH	P	IDA	26
GO:0015078	hydrogen ion transmembrane transporter activity	F	IDA	26
GO:0006814	sodium ion transport	P	IDA	26
GO:0006024	glycosaminoglycan biosynthetic process	P	IDA	26
GO:0006390	transcription from mitochondrial promoter	P	IDA	24
GO:0042645	mitochondrial nucleoid	C	IDA	24
GO:0006139	nucleobase-containing compound metabolic process	P	IDA	24
GO:0051607	defense response to virus	P	IDA	23
GO:0002684	positive regulation of immune system process	P	IDA	22
GO:0050863	regulation of T cell activation	P	IDA	22
GO:0032673	regulation of interleukin-4 production	P	IDA	22

Table 23: Cluster specific gene ontologies for PIK-75 and Sorafenib treated Mahlavu cells (continued)

GO:0004674	protein serine/threonine kinase activity	F	IDA	22
GO:0050870	positive regulation of T cell activation	P	IDA	22
GO:0031461	cullin-RING ubiquitin ligase complex	C	IDA	21
GO:0000151	ubiquitin ligase complex	C	IDA	21
GO:0032446	protein modification by small protein conjugation	P	IDA	21
GO:0008630	DNA damage response, signal transduction resulting in induction of apoptosis	P	IDA	20
GO:0042770	signal transduction in response to DNA damage	P	IDA	20
GO:0002098	tRNA wobble uridine modification	P	IDA	20
GO:0004594	pantothenate kinase activity	F	IDA	20
GO:0030056	hemidesmosome	C	IDA	20
GO:0043065	positive regulation of apoptotic process	P	IDA	20
GO:0043068	positive regulation of programmed cell death	P	IDA	20
GO:0090083	regulation of inclusion body assembly	P	IDA	2
GO:0016628	oxidoreductase activity, acting on the CH-CH group of donors, NAD or NADP as acceptor	F	IDA	2
GO:0090084	negative regulation of inclusion body assembly	P	IDA	2
GO:0007264	small GTPase mediated signal transduction	P	IDA	19
GO:0001738	morphogenesis of a polarized epithelium	P	IDA	19
GO:0012506	vesicle membrane	C	IDA	18
GO:0030659	cytoplasmic vesicle membrane	C	IDA	18
GO:0030658	transport vesicle membrane	C	IDA	18
GO:0012505	endomembrane system	C	IDA	18
GO:0030121	AP-1 adaptor complex	C	IDA	18
GO:0016272	prefoldin complex	C	IDA	17
GO:0030983	mismatched DNA binding	F	IDA	17
GO:0006298	mismatch repair	P	IDA	17
GO:0030968	endoplasmic reticulum unfolded protein response	P	IDA	15

Table 23: Cluster specific gene ontologies for PIK-75 and Sorafenib treated Mahlavu cells (continued)

GO:0034976	response to endoplasmic reticulum stress	P	IDA	15
GO:0000278	mitotic cell cycle	P	IDA	15
GO:0009411	response to UV	P	IDA	13
GO:0045948	positive regulation of translational initiation	P	IDA	13
GO:0001523	retinoid metabolic process	P	IDA	12
GO:0055114	oxidation-reduction process	P	IDA	12
GO:0047035	testosterone dehydrogenase (NAD+) activity	F	IDA	12
GO:0009890	negative regulation of biosynthetic process	P	IDA	11
GO:0009083	branched chain family amino acid catabolic process	P	IDA	1
GO:0009308	amine metabolic process	P	IDA	1
GO:0006695	cholesterol biosynthetic process	P	IDA	1
GO:0004080	biotin-[propionyl-CoA-carboxylase (ATP-hydrolyzing)] ligase activity	F	IDA	1
GO:0071110	histone biotinylation	P	IDA	1
GO:0043170	macromolecule metabolic process	P	IEP	17, 21
GO:0010942	positive regulation of cell death	P	IGI	20
GO:0034645	cellular macromolecule biosynthetic process	P	IGI	13
GO:0006164	purine nucleotide biosynthetic process	P	IMP	9
GO:0003993	acid phosphatase activity	F	IMP	7
GO:0030833	regulation of actin filament polymerization	P	IMP	4
GO:0031333	negative regulation of protein complex assembly	P	IMP	4
GO:0032488	Cdc42 protein signal transduction	P	IMP	3
GO:0008203	cholesterol metabolic process		IMP	2, 3
GO:0016192	vesicle-mediated transport	P	IMP	3, 18
GO:0016569	covalent chromatin modification	P	IMP	3
GO:0003735	structural constituent of ribosome	F	IMP	29
GO:0032728	positive regulation of interferon-beta production	P	IMP	28
GO:0050776	regulation of immune response	P	IMP	28
GO:0042116	macrophage activation	P	IMP	28
GO:0050650	chondroitin sulfate proteoglycan biosynthetic process	P	IMP	26

Table 23: Cluster specific gene ontologies for PIK-75 and Sorafenib treated Mahlavu cells (continued)

GO:0030166	proteoglycan biosynthetic process	P	IMP	26
GO:0007005	mitochondrion organization	P	IMP	24
GO:0002682	regulation of immune system process	P	IMP	22
GO:0080008	CUL4 RING ubiquitin ligase complex	C	IMP	21
GO:0034227	tRNA thio-modification	P	IMP	20
GO:0042026	protein refolding	P	IMP	2
GO:0051056	regulation of small GTPase mediated signal transduction	P	IMP	19
GO:0007059	chromosome segregation	P	IMP	15
GO:0006986	response to unfolded protein	P	IMP	15
GO:0009416	response to light stimulus	P	IMP	13
GO:0006776	vitamin A metabolic process	P	IMP	12
GO:0008265	Mo-molybdopterin cofactor sulfurase activity	F	IMP	12
GO:0030031	cell projection assembly	P	IMP	11
GO:0006091	generation of precursor metabolites and energy	P	IMP	10
GO:0018271	biotin-protein ligase activity	F	IMP	1
GO:0008536	Ran GTPase binding	F	IPI	7
GO:0031625	ubiquitin protein ligase binding	F	IPI	2
GO:0030742	GTP-dependent protein binding	F	IPI	19

Table 24: Cluster specific gene ontologies for PI3Ki-beta and Sorafenib treated Mahlavu cells

GOID	Description	Which	Evidence	Cluster
GO:0006695	cholesterol biosynthetic process	P	IMP	1, 6
GO:0008610	lipid biosynthetic process	P	IMP	1, 6
GO:0032933	SREBP-mediated signaling pathway	P	IDA	1
GO:0012507	ER to Golgi transport vesicle membrane	C	IDA	1
GO:0016628	oxidoreductase activity, acting on the CH-CH group of donors, NAD or NADP as acceptor	F	IDA	1
GO:0030135	coated vesicle	C	IDA	1
GO:0005324	long-chain fatty acid transporter activity	F	IGI	9
GO:0045218	zonula adherens maintenance	P	IMP	7

Table 24: Cluster specific gene ontologies for PI3Ki-beta and Sorafenib treated Mahlavu cells (continued)

GO:0005815	microtubule organizing center	C	IDA	7
GO:0045217	cell-cell junction maintenance	P	IMP	7
GO:0090136	epithelial cell-cell adhesion	P	IMP	7
GO:0003774	motor activity	F	IDA	7
GO:0051908	double-stranded DNA specific 5'-3' exodeoxyribonuclease activity	F	IDA	7
GO:0016327	apicolateral plasma membrane	C	IDA	7
GO:0016787	hydrolase activity	F	IDA	7
GO:0006720	isoprenoid metabolic process	P	IDA	6
GO:0033257	Bcl3/NF-kB2 complex	C	IDA	4
GO:0010608	posttranscriptional regulation of gene expression	P	IMP	4
GO:0015858	nucleoside transport	P	IDA	3

Table 25: Cluster specific gene ontologies for Sorafenib treated Mahlavu cells

GOID	Description	which	Evidence	Cluster
GO:0006469	negative regulation of protein kinase activity	P	IMP	5
GO:0033673	negative regulation of kinase activity	P	IDA	5
GO:0042325	regulation of phosphorylation	P	IMP	5
GO:0005947	mitochondrial alpha-ketoglutarate dehydrogenase complex	C	IDA	6
GO:0006550	isoleucine catabolic process	P	IMP	6
GO:0009083	branched chain family amino acid catabolic process	P	IDA	6
GO:0016779	nucleotidyltransferase activity	F	EXP	2

APPENDIX F

COPYRIGHT PERMISSIONS

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS

Aug 11, 2021

This Agreement between METU -- Kübra Narcı ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	5125910218003
License date	Aug 11, 2021
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	CA: Cancer Journal for Clinicians
Licensed Content Title	Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries
Licensed Content Author	Ahmedin Jemal, Lindsey A. Torre, Rebecca L. Siegel, et al
Licensed Content Date	Sep 12, 2018
Licensed Content Volume	68
Licensed Content Issue	6
Licensed Content Pages	31

Figure 53. Copyright permission of the image used in the Figure 1.

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Aug 11, 2021

This Agreement between METU -- Kübra Narcı ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5125911297839
License date	Aug 11, 2021
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Investigational New Drugs
Licensed Content Title	Inhibition of Akt signaling in hepatoma cells induces apoptotic cell death independent of Akt activation status
Licensed Content Author	Francesca Buontempo et al
Licensed Content Date	Jul 14, 2010
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Will you be translating?	no

Figure 54. The copyright permission of the image used in Figure 11.

OXFORD UNIVERSITY PRESS LICENSE
TERMS AND CONDITIONS

Aug 11, 2021

This Agreement between METU -- Kübra Narcı ("You") and Oxford University Press ("Oxford University Press") consists of your license details and the terms and conditions provided by Oxford University Press and Copyright Clearance Center.

License Number	5125920198346
License date	Aug 11, 2021
Licensed content publisher	Oxford University Press
Licensed content publication	Bioinformatics
Licensed content title	edgeR: a Bioconductor package for differential expression analysis of digital gene expression data
Licensed content author	Robinson, Mark D.; McCarthy, Davis J.
Licensed content date	Nov 11, 2009
Type of Use	Thesis/Dissertation
Institution name	
Title of your work	SYSTEMS BIOLOGY ANALYSIS OF KINASE INHIBITORS IN LIVER CANCER CELLS USING NEXT GENERATION SEQUENCING DATA
Publisher of your work	METU

Figure 55. Copyright permission of the figure used in Figure 14.

CURRICULUM VITAE

PERSONAL INFORMATION

EDUCATION

Degree	Institution	Graduation
PHD	METU, Medical Informatics	2021
MS	METU, Bioinformatics	2014
BS	METU, Molecular Biology and Genetics	2012
High School	Erzincan Milli Piyango Anatolium	2007

WORK EXPERIENCE

2017-Present Seven Bridges Genomics Bioinformatic Analyst

- Research, analysis, and optimization of various bioinformatics softwares and algorithms for NGS analysis. Used batch scripting, Common Workflow Language (CWL), cloud computing for dockerizing and deployment the tools and workflows in different languages on different Seven Bridges platforms.

- Construction of the procedures and pipeline, NGS analysis, populational genomics analysis and reporting of the initial phase of TGP (Turkey Genome Project) which was a government-funded project.
- Benchmarking of Seven Bridges GRAF aligner and variant caller, graph-based pipeline construction for the analysis of population-specific analysis. Method investigations and benchmarks for structural variant analysis.

2014-2017 INTERGEN Bioinformatics Analyst

- Analysis of Miseq and Sanger data using IGV tool, web services like; Ensembl and HGMD databases, SNP prediction tools; PolyPhen, Sift, Mutation Tester, and HSF.
- Primer and oligo design for PCR, MLPA data analysis using Coffanalyser.
- Construction of a pipeline for WGS analysis of truseq data using several software packages. Analysis and management of patient WGS data.
- Next Generation Sequence (NGS) analysis using software packages like; BWA, samtools, Picard-tools, GATK, SNPeff, VEP, and Gemini.

RESEARCH EXPERIENCE

2011 Fly lab of Prof. Dr. Annette Schenck at the Human Genetics Department for Molecular Life Sciences in Nijmegen, Netherlands

Screen of about 350 ID-genes within a high-throughput behavioral assay using the UAS-Gal4-system in combination with RNA-interference, so that each ID-gene could be silenced one by one specifically within lateral glial cells.

2012 Assos. Prof. Dr. Yeşim Son Aydın, Bioinformatics Department at METU.

Microarray analysis of breast cancer samples using R packages and METU-SNP.

2012 IGEM (International Genetically Engineered Machines) competition as IGEM-METU team

Worked to develop biological corbon monixide filter, supervision of the modeling part and web design.

2013 DKFZ-Husar Bioinformatics, Heidelberg, Germany, supervision of Karl-Heinz Glatting and Dr. Agnes Hotz-Wagenblatt

Develop an integrated package; "compaRNA" which is designed to compare two sets of ncRNA sequence data using DESeq, EdgeR, SAMSeq R packages, worked in Unix and Linux environments, used R and Perl.

2012-2014 Master studies with Prof. Dr. Mahinur S. Akkaya's Plant Molecular Genetics Lab.

Thesis Title: "Design of a Sequence Based miRNA Clustering Method; Analysis of Fungal milRNA and Host Organism Target Genes.
Designed a sequence similarity-based clustering approach on human miRNA dataset, and applied them onto the analysis of Fungal milRNA predictions.

TEACHING EXPERIENCE

- Lectured the course **'Introduction to Bioinformatics'** at Biyoakademi between 2017-2020.

HONORS AND AWARDS

- Graduation from High School with first degree.
- TUBITAK scholarship for being 4153rd place in National University Entrance Exam out of 1.8 million candidates (359.00 overall 380.00), during bachelor, master, and doctorate.
- Bronze medal in IGEM 2012 Synthetic Biology Competition as a student in METU Team.
- Full Boarding award on Summer School in Bioinformatics & NGS Data Analysis (#NGSchool2017), which held on September 10-17, 2017 in Warsaw, Poland.

PUBLICATIONS

- Kubra Narci*, Deniz Cansen Kahraman, Tulin Ersahin, Nurcan Tuncbagi Rengul Cetin Atalay (2021), Context dependent isoform specific PI3K inhibition confers drug resistance in Hepatocellular carcinoma cells (on review process at BMC Cancer, biorxiv)
- Serhat Tetikol*, Kubra Narci*, Gungor Budak*, Deniz Turgut*, Ozem Kalay, Elif Arslan, Sinem Demirkaya-Budak, Alexey Dolgoborodov, Amit Jain, Duygu Kabakci-Zorlu, Richard Brown, Vladimir Semenyuk, Brandi Davis-Dusenbery, Population-specific genome graphs improve high-throughput sequencing data analysis: A case study on the African pan-genome (on review process at Nature Com., biorxiv).
- Damla Gozen, Deniz C. Kahraman, Kubra Narci, Huma Shehwna, Ozlen Konu, Rengul Cetin-Atalay (2021). Transcriptome profiles associated with selenium deficiency dependent oxidative stress identify potential diagnostic and therapeutic targets in liver cancer cells, Turk J Biol.
- Olson N et al. (2021), precision FDA Truth Challenge V2: Calling variants from short- and long-reads in difficult-to-map regions (on review process at Cell Genomics)
- K.Narci*, H.Oğul, M.Akkaya, (2016), Sequence-based microRNA clustering, International Conference on Bioinformatics Models, Methods and Algorithms, (BIOSTEC'16) Roma.

CONFERANCES AND PRESENTATIONS

- Kubra Narci, Gungor Budak, Deniz Turgut, Serhat Tetikol (2020), Measuring the effectiveness of population-specific genome graph references on a public qatari dataset, ASHG Virtual Meeting 2020.
- Kubra Narci, Tunca Dogan, Aybar Can Acar, Tulin Ersahin, Nurcan Tuncbag and Rengul Cetin Atalay, (2017), Systems Biology Analysis of Kinase Inhibitors in Cancer Cells Using Next Generation Sequencing Data, 10th International Symposium on Health Informatics and Bioinformatics (HIBIT 2017), METU Northern Cyprus Campus.
- Kubra Narci, A. Caglar Ozketen, Ayse Andac, Burak Demiralay, Bala Gur-Dedeoglu, Yasemin Oztemur and Mahinur S. Akkaya. Hunting down of the endogenic microRNA like small RNAs in plant pathogenic and symbiotic

fungi. XVI International Congress on Molecular Plant-Microbe Interactions, Rhodes, Greece (2014).

- Hasan Oğul, Kübra Narcı, İ. Kutay Özturk, Taylan Beyaztaş,A. Çağlar Özketen, Mahinur S. Akkaya. Is Pathogen miRNAs controlling gene expression of the host organism? In search of miRNAs in Blumeria graminis f. sp. hordei and candidate targets in barley. Vienna Plant symposium (2013).

SKILLS

Languages: English (Advanced) and German (Beginner).

Programming: Python (Numpy, SciPy, Matplotlib, Pandas, pysam), R, Perl and JavaScript.

Bioinformatics Databases: NCBI, ensemble, UCSC, GEO, ArrayExpress, DAVID, ENTREZ, Uniprot, Ensembl,SignalP, PDB, miRBase, TAM, and BioMart.

Bioinformatic Softwares: Bowtie, BWA, Samtools, VEP, SnpEff, Bcftools, GATK, TopHat, HTSeq, edgeR, DESeq, Kallisto, Omics Integrator, HOMER, PCA, Plink, ADMIXTURE, EIGENSOFT and Treemix.

Environments: Windows, MacOS, Unix and Linux environments.

Document Creation: Microsoft Office Suite, LaTex, Markdown, and Jupyter.