IDENTIFICATION OF CELLULAR STRESS RELATED BIOMOLECULES FOR EVENTUAL USE IN TARGETED THERAPIES OF HEPATOCELLULAR CARCINOMA


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


DAMLA GÖZEN


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


SEPTEMBER 2021

Approval of the thesis:

**IDENTIFICATION OF CELLULAR STRESS RELATED BIOMOLECULES FOR EVENTUAL USE IN TARGETED THERAPIES OF HEPATOCELLULAR CARCINOMA**

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

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

_____

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

_____

Assoc. Prof. Yeşim Aydın Son
Supervisor, **Health Informatics Dept., METU**

_____

Prof. Dr. Rengül Çetin-Atalay
Co-Supervisor,  **Health Informatics Dept., METU**

_____

**Examining Committee Members:**

Prof. Dr. Ayşe Elif Erson Bensan
Biology Dept., METU

_____

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

_____

Assist. Prof. Aybar Can Acar
Health Informatics Dept., METU

_____

Assoc. Prof. Özlen Konu
Mol. Biology and Genetics Dept., Bilkent University

_____

Assoc. Prof. Nurcan Tunçbağ
Faculty of Engineering, Koç University

_____

**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 :   Damla Gözen

Signature           :   _____

# ABSTRACT

IDENTIFICATION OF CELLULAR STRESS RELATED BIOMOLECULES FOR
EVENTUAL USE IN TARGETED THERAPIES OF HEPATOCELLULAR
CARCINOMA

GÖZEN, Damla

Ph.D., Department of Medical Informatics

Supervisor: Assoc. Prof. Yeşim AYDIN SON

Co-supervisor: Prof. Dr. Rengül ÇETIN-ATALAY

September 2021, 102 pages

Hepatocellular carcinoma (HCC) is one of the most common and deadly cancer types. HCC cells generally display increased resistance to various stress conditions such as oxidative stress. There are conventional therapies used in treatment of HCC but their efficacies are low due to resistance gained by cancer cells and off-target effects. In this thesis study, the aim is to analyze oxidative stress-related gene expression profiles of HCC cell lines to determine genes that could be targeted in novel diagnostic and therapeutic strategies. Selenium (Se) deficiency dependent model of oxidative stress was utilized to identify the genes that are involved in resistance to oxidative stress. The results of a transcriptome-wide gene expression data were analyzed in which differentially expressed genes (DEGs) were identified between HCC cells that are either resistant or sensitive to Se-deficiency dependent oxidative stress. They were further investigated for their importance by cell signaling network analysis. 27 genes were defined to have key roles; 16 of which were previously shown to have impact on patient survival with primary liver cancer. Moreover, the expression of the majority of these genes were found to be correlated with p53-MDM2 pathway. Hence cytotoxic effects of novel small molecules targeting p53-MDM2 protein-protein interaction were tested on 4 HCC cells. Two compounds were shown to induce apoptosis in HCC cells and led to nuclear localization of p53. Altogether, genes identified in this study are proposed to be novel targets for diagnostic and therapeutic approaches and the p53-MDM2 inhibitors have potentials in HCC treatment.

# ÖZ

HEPATOSELLÜLER KARSİNOM HEDEFLİ TEDAVİLERİNDE
KULLANILMAK ÜZERE HÜCRESEL STRES KAYNAKLI
BİYOMOLEKÜLLERİN BELİRLENMESİ

GÖZEN, Damla

Doktora, Tıp Bilişimi Bölümü

Tez Yöneticisi: Doç. Dr. Yeşim AYDIN SON

Eş-Danışman: Prof. Dr. Rengül ÇETİN-ATALAY

Eylül 2021, 102 sayfa

Hepatosellüler karsinom (HCC) en sık görülen ve ölümcül kanser türlerindendir. HCC hücreleri oksidatif stres gibi farklı stres durumlarına karşı genellikle dirençlidirler. HCC tedavisinde kullanılan geleneksel yöntemlerin etkinliği kanser hücrelerinin kazandığı direnç ve hedef dışı etkilerden dolayı düşüktür. Bu tez çalışmasındaki amaç, HCC hücre hatlarının oksidatif stresle alakalı gen ifade profillerini inceleyerek yeni tanı ve tedavi yöntemlerinde hedeflenebilecek genleri belirlemektir. Reaktif oksijen türlerine dirençte kritik olan genlerin belirlenmesi için Selenyum eksikliğine bağlı oksidatif stres modeli incelenmiştir. Transkriptom gen ifade analizi verileri incelenerek Selenyum eksikliğine bağlı oksidatif strese duyarlı ve dirençli HCC hücre hatları arasında farklı ifade edilen genler belirlenmiştir. Bu genler ağ analizi yöntemleri ile incelenmiş ve 27 genin anahtar rollerinin olduğu bulunmuştur; bunlardan 16'sının daha önceki çalışmalarda karaciğer kanser hastalarının sağ kalımında etkili olduğu gösterilmiştir. Ayrıca, bu genlerin çoğunun ifadesinin p53-MDM2 yolağında görevli genlerle ilişkili olduğu bulunmuş ve p53-MDM2 interaksiyon inhibitörlerinin 4 HCC hücre hattı üzerindeki sitotoksik etkileri potansiyel tedavi olarak test edilmiştir. İki bileşiğin HCC hücre apoptozuna yol açtığı ve p53'ün nükleusa lokalizasyonunu sağladığı gösterilmiştir. Sonuç olarak, bu çalışmada belirlenen genler yeni tanı ve tedavi yaklaşımlarında hedeflenmek üzere önerilmektedir ve p53-MDM2 inhibitörleri HCC tedavisinde potansiyel taşımaktadır.

Anahtar Sözcükler: Hepatosellüler karsinom, oksidatif stres, selenyum, Hepatit B virüsü, transkriptom analizi

To My Mom and Dad

# ACKNOWLEDGMENTS

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| **2.2.15** | HepG2-2.2.15 |
| **8-OHdG** | 8-hydroxyl-2-deoxyguanosine |
| **BCL** | Between Cell Line |
| **BSA** | Bovine Serum Albumin |
| **cccDNA** | covalently closed circular DNA |
| **CCLE** | Cancer Cell Line Encyclopedia |
| **cDNA** | complementary DNA |
| **D1** | Day 1 |
| **D2** | Day 2 |
| **D3** | Day 3 |
| **DEG** | Differentially Expressed Gene |
| **DMEM** | Dulbecco's Modified Eagle Medium |
| **DMSO** | Dimethyl Sulfoxide |
| **DNA** | Deoxyribonucleic acid |
| **ES** | Enrichment Score |
| **FCS** | Fetal Calf Serum |
| **FDR** | False Discovery Rate |
| **GO** | Gene Ontology |
| **GPX** | Glutathione Peroxidase |
| **GSEA** | Gene Set Enrichment Analysis |
| **HBV** | Hepatitis B Virus |
| **HBx** | Hepatitis B virus X protein |
| **HCV** | Hepatitis C virus |
| **HCC** | Hepatocellular Carcinoma |
| **HM** | Heatmap |
| **HT-HBV** | Head-to-Tail HBV |
| **IC50** | half maximal inhibitory concentration |
| **KEGG** | Kyoto Encyclopedia of Genes and Genomes |
| **limma** | linear models for microarray data |
| **MDM2** | Mouse double minute 2 homolog |

| | |
|---|---|
| **mRNA** | Messenger RNA |
| **MsigDB** | Molecular Signatures Database |
| **mtROS** | Mitochondrial ROS |
| **NOM p-val** | nominal p value |
| **OS** | Oxidative Stress |
| **PCR** | Polymerase Chain Reaction |
| **PCST** | Prize Collecting Steiner Tree |
| **PPI** | Protein Protein Interaction |
| **rcDNA** | relaxed circular DNA |
| **RMA** | Robust Multi-Array Average |
| **RNA** | Ribonucleic acid |
| **RNA-seq** | RNA sequencing |
| **ROS** | Reactive Oxygen Species |
| **Se** | Selenium |
| **SeCys** | Selenocysteine |
| **SOD** | Superoxide Dismutase |
| **SRB** | Sulforhodamine B |
| **TACE** | Transarterial Chemoembolization |
| **TARE** | Transarterial Radioembolization |
| **Tp53** | Tumor protein 53 |
| **TRX** | Thioredoxin Reductase |
| **UV** | Ultraviolet |
| **WCL** | Within Cell Line |

# CHAPTER 1

# INTRODUCTION

Cancer is one of the major public health problems worldwide, being the second most fatal disease according to World Health Organization. The number of patients diagnosed with cancer has been increasing each year (Siegel et al., 2012). Recently, many studies are conducted to develop novel treatment strategies with increased efficacy and decreased side effects, due to the inadequacy of the existing methods. Along these studies, many of them focuses on novel biomarker discovery that could be used as targeted treatment strategies. The determination of cancer cell characteristics in more detail would therefore be important to give more insight about molecular mechanisms behind carcinogenesis and could be used to find novel genes for targeted therapy.

This chapter aims to give detailed information about hepatocellular carcinoma (HCC), common external stress conditions including oxidative stress, stress response, importance of Selenium element and the importance of p53 protein as a result of genotoxic stress with the intention of providing background information and insight about the main motivation of this thesis study. Moreover, the use of transcriptome-wide analysis and bioinformatics approaches used to obtain the results would be discussed for their power and significance in the discovery of novel biomarkers as potential diagnostic and therapeutic targets in the further parts.

## 1.1 Hepatocellular Carcinoma

HCC is the fourth most fatal and the sixth most frequent cancer worldwide accounting for >80% among all liver cancer types (Bray et al., 2018; Yang et al., 2019). There are several risk factors related to hepatocarcinogenesis including hepatitis B (HBV) and hepatitis C (HCV) viral infection, alcohol and tobacco consumption, toxin exposure such as aflatoxin B1, obesity, chronic liver disease or cirrhosis. It is counted as one of the most aggressive cancers; hence the early diagnosis of it is of significance in the efficacy of the therapy.

HCC frequently develops as a result chronic liver disease associated with cycles of inflammation and hepatocyte regeneration. Liver regeneration is maintained by hepatic progenitor cells which carry stem cell properties. Continuous activation of

these cells induces liver fibrosis, chronic inflammation, and ultimately hepatocellular carcinogenesis. (Wu et al., 2020)

### 1.1.1 Hepatitis B Virus Infection

Chronic HBV infection accounts for more than 50% of the HCC incidences worldwide and the number of infected people is about 350 million (Y. Xie, 2017). HBV is a DNA virus which is composed of a double-stranded relaxed circular DNA (rcDNA) inside a nucleocapsid and an envelope. In case of HBV infection, virus enters into the cell to release its rcDNA which is delivered into the nucleus to be converted to covalently closed circular DNA (cccDNA); which is a more stable form of the viral genome. Viral mRNA synthesis could then be performed from this template DNA. (M. Xie et al., 2018)

HBV-dependent hepatocarcinogenesis occurs depending on various direct and indirect mechanisms. Rapid replication of HBV genome, its integration site through the host genetic material, viral genotype and its mutant type and the expression of virus-driven oncogenes are the most important factors in the oncogenic transformation. Chronic HBV infection also results in a continuous host response which generates liver inflammation, and it might lead to liver fibrosis and even to cirrhosis if not treated. This might be another mechanism that increase mutation rates resulting in hepatocarcinogenesis (Y. Xie, 2017).

After the delivery of the HBV DNA into host nucleus, viral DNA is integrated into the host DNA resulting in genomic instability and insertional mutagenesis (Levrero & Zucman-Rossi, n.d.) Many previous studies focused on determining the presence of any preferential integration sites; however sequencing studies showed that integration sites were randomly distributed through the host genome with a little bias through chromosome 10 and 17 (An et al., 2018).

The expression and translation of viral genes to viral proteins for an extended period including HBx protein disrupts normal cellular transcription functions to decrease proliferation control of the cells leading to carcinogenesis (Figure 1). HBx also have control over host epigenetic regulations to alter the expression of tumor suppressor genes (Levrero & Zucman-Rossi, n.d.). In addition to these, HBx has multiple other oncogenic roles by affecting cellular activities related to DNA repair, cell cycle progression and apoptosis. (Yan et al., 2017).
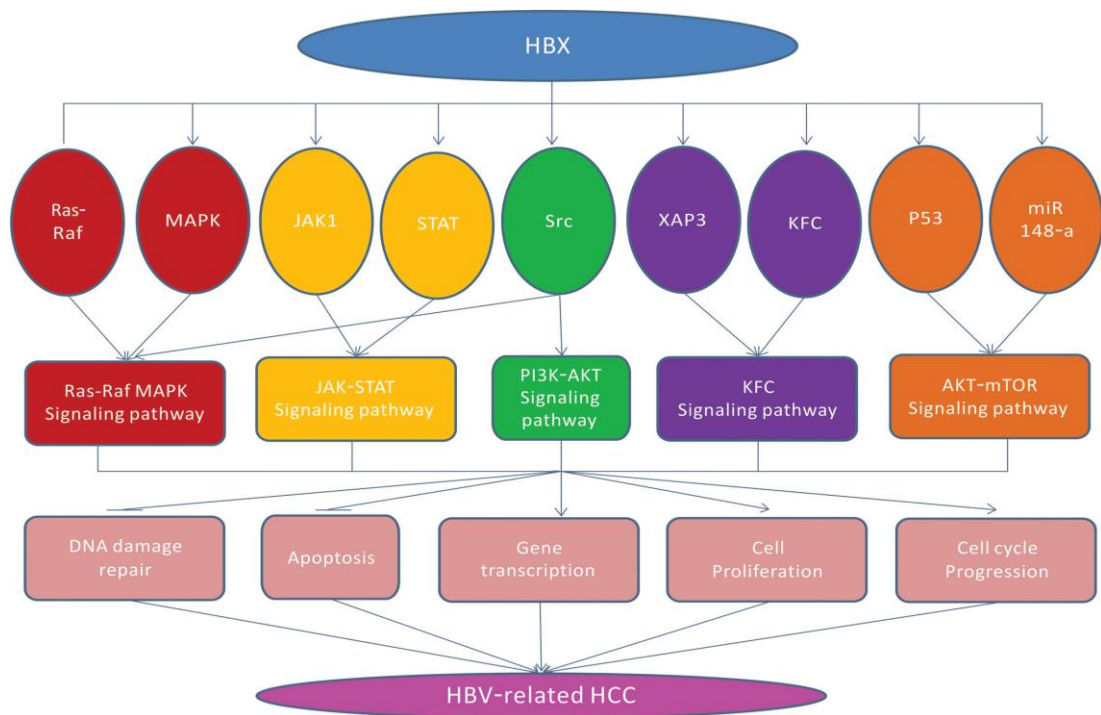
Figure 1: The molecular mechanisms that are possible drivers for HBx induced hepatocarcinogenesis (Yan et al., 2017)

### 1.1.2 TP53 and Hepatocellular Carcinoma

TP53 is a tumor suppressor gene with various transcription factor functions including the regulation of cell cycle arrest, inhibition of angiogenesis, senescence and apoptosis induced as a cellular response to various stress factors as well as DNA damage repair mechanisms. Previous studies identified that TP53 mutations are the most detected alterations in HCC patients, and they are associated to poor prognosis and survival. (Woo et al., 2011).

The regulation of cell differentiation by TP53 gene is another widely studied subject, and its expression levels are known to be very low in human embryonic stem cells and the activation of p53 results in differentiation of the stem cells. The researchers showed the correlation between lack of p53 in HCC cells with stem cell reprogramming. (Hong et al., 2009)

In normal cells, p53 is degraded to keep the levels low. Mdm2 protein is responsible for this degradation process. It binds to p53 to transport it from nucleus to cytosol, provides it ubiquitylation to prepare its degradation by proteasome. Mdm2 dependent degradation of P53 is reversible and in response to a stress condition p53 levels might increase to regulate the cells response to the stressor and eventually the cells faith. (Purvis et al., 2012) Mdm2 known to be overexpressed in many cancer types resulting

in the degradation of p53. Therefore, inhibition of p53-MDM2 interaction might be a promising therapeutic strategy for the treatment of HCC patients. (Chène, 2003).

### 1.1.3 Current Treatment Strategies for Hepatocellular Carcinoma

There are several treatment options currently used for the treatment of HCC and the choice of one is highly linked to the tumor stage. Surgical operations, liver transplantation and other curative therapies are only effective when the tumor is in its early stage. For the treatment of more advanced stages, pharmacological methods are currently the most effective options; but when the low overall survival of HCC patients are taken into consideration, there still is a need for better methods. (B. Chen et al., 2020b; Subramaniam et al., 2013).

#### 1.1.3.1 Conventional Treatment Methods for Hepatocellular carcinoma

For patients with early stages of HCC, hepatic resection is a treatment option when they have well-preserved liver function. Liver transplantation is another highly used option and it has the advantage of removing both the tumor and cirrhotic tissues to provide 70% 5-year survival rate.   For the treatment of patients where resection and transplantation are not possible, ablation could be considered; especially when the tumor is in its early stage. (Cabebe, 2021; B. Chen et al., 2020a)

For the HCC patients with an intermediate staging, transarterial chemoembolization (TACE) treatment is a suitable option, which is consist of the administration of a chemotherapeutic agent to the artery of tumor nodule. This method was shown to significantly increase the survival time of patients and therefore is one of the best non-surgical treatment options. Transarterial radioembolization (TARE) is another similar method which is done by the injection of radioactive compounds into the tumor nodule artery. TARE seems to have more targeted effects compared to TACE and therefore is a more suitable option for the treatment of HCC patients with more advanced stage; but both options were shown to increase the overall survival of patients significantly. (Cabebe, 2021; B. Chen et al., 2020a)

Most of the patients have HCC in advanced stage due to the late diagnosis; and in that case the commonly used treatment strategy is chemotherapy. As mentioned in previous sections, liver cancer cells are highly heterogeneous structures with the involvement of hepatic stem cell population; this mixed nature of HCC make them gain resistance too conventional methods to invade or reoccur; which leads to decreased treatment efficacy. Chemotherapy also has many side effects; since it targets all proliferating cells, not specifically cancer cells. For the treatment of the patients with HCC in advanced stage, the systemic administration of pharmacological agents is the most efficient strategy. (Cabebe, 2021; B. Chen et al., 2020a).

4

### 1.1.3.2 Targeted Therapy

Recently, many studies have been performed to find novel diagnostic strategies and to develop new therapies. For the HCC patients in which case resection or transplantation is not an option, first-line systemic therapy is recommended by various guidelines. Sorafenib is one of the most efficient, approved drugs for advanced HCC treatment which extends patient survival for only 2.8 months. Lenvatinib is another approved targeted first-line therapy in advanced HCC cases. But the efficiency of these first-line treatment strategies remains to be low due to the tolerance gained by the tumor cells against them. Cancer cells escape from the effects of these agents by using alternative proliferative molecular pathways.

In order to develop better strategies, using Sorafenib in combination with other compounds targeting those alternative pathways such as MEK/ERK, mTOR, EGFR, c-MET pathways is an option. To solve this problem the use of second line agents is widely searched for their potentials in HCC treatment. In previous trials, Regorafenib was found to show effective results in the treatment of HCC patients who gained resistance to Sorafenib, extending patient survival for 4 months. However, it was also found that only ~30% of the patients got benefit from the second line Regorafenib treatment strategy and well-preserved liver functioning is the factor that directly effects the treatment efficacy. Another second-line agent use to treat HCC with resistance to Sorafenib is cabozantinib; which is another tyrosine kinase inhibitor. However, discovery of novel agents that could be used either alone or in combinational treatment strategies is still a need (A. Huang et al., 2020).

To fill this gap and find novel therapy strategies with increased efficacy, novel target discovery has great importance to be used as a target to inhibit cancer cell proliferation and resistance. The determination of cancer cell characteristics in more detail would give us more insight about molecular mechanisms behind carcinogenesis and could be used to find novel genes for diagnosis and targeted therapy.

## 1.2 Cellular Stress Response

Investigation of how cancer cells respond to various cellular stress conditions is one of the most important characteristics in order to understand their molecular mechanism of action. Generally, cells respond to stress in a variety of ways from activation of pathways that promote survival to initiating programmed cell death to eliminate damaged cells; but many HCC cells scavenge mechanisms to resist stress-induced apoptosis and they survive (Fulda et al., 2010). Determination of stress response gene expression profiles is, therefore, essential for establishing the nature of cancer cells.

A stress factor could be any factor that alters cells' optimum living environment such as temperature, pH and osmotic pressure changes or over accumulation of Reactive oxygen species (ROS). These factors initiate cellular responses; either intrinsic- within the cell or through the immune system. (Ak & Levine, 2010). When cells sense a stress condition, they try to develop an adaptive response by altering their gene expression

regulations in order to maintain cellular balance after the stress factor is removed. Many transcriptional studies were conducted to determine the stress response gene expression profiles.

Sensing the stress factor is the first step and the cells have diverse sensors in order to be able to adapt despite all the environmental changes. After sensing this factor, cells should initiate a rapid response to eliminate or adapt it and survive. They take the inputs and send the information to effectors through signal transduction. Immediately the stress signal reaches to effectors, rapid alterations occur inside the cell such as translational downregulations, channel and transporter protein regulations and activation of stress-response gene expressions. (de Nadal et al., 2011).

For the initiation of the extrinsic immune response of a cell against a stressor; NF-KB is the main regulatory pathway for the modulation of the expression of many downstream genes; which could also act like an oncogene since it is responsible for cellular proliferation activities. For the initiation of intrinsic stress, the presence of a factor such as disturbance in the balance of ROS, amino acid or glucose levels, hypoxia, damage to the DNA, the presence of toxins etc. is required. In such cases cells responds to these stress factors through p53 protein to prevent the effects of the factor and repair the damage. (Milisav, n.d.; Sengupta et al., 2010) Cells respond stress inducers by trying to repair the effects of damage for the recovery of normal state, adapt to the stress factor or induce autophagy or apoptosis. If the stressor could not be removed from the environment and the effects of it are accumulated, this might result in damage to all parts of a cell including DNA, protein and lipids which in turn cause various disease such as diabetes, cardiac or neurodegenerative disease and even cancer. (Milisav, n.d.)

**1.3 Oxidative Stress**

One of the stress inducing factor in eukaryotic cells is try to adapt to the oxidative stress conditions for which cells have evolved different responses. Oxidative stress occurs when cells are unable to detoxify the reactive agents or to repair the resulting damage that formed because of systemic accumulation of ROS. Disturbances in the normal redox state of cells produce peroxides and free radicals that might give damage to all components of the cell, including nucleic acids, proteins and lipids, resulting in toxic effects. It may also suppress apoptosis and promote proliferation, invasiveness and metastasis leading to carcinogenesis (Halliwell, 2007).

1.3.1   Reactive Oxygen Species and Their Cellular Effects

Reactive oxygen species are composed of Superoxide radicals ($O_2^{\bullet-}$), hydrogen peroxide ($H_2O_2$), hydroxyl radicals ($^{\bullet}OH$), and singlet oxygen element and the sources responsible for their production might be both exogenous and endogenous (Figure 2). Mitochondrial oxidative phosphorylation is the primary source of the endogenous ROS production in cells, which is generated by the reduction of oxygen molecule throughout the respiratory event. Peroxisomes and Cytochrome P450 are the other

endogenous sources leading to ROS production. Exogenous sources include UV light and ionizing radiation exposure, inflammatory cytokine accumulation, etc. (Pizzino et al., 2017; Waris & Ahsan, 2006).
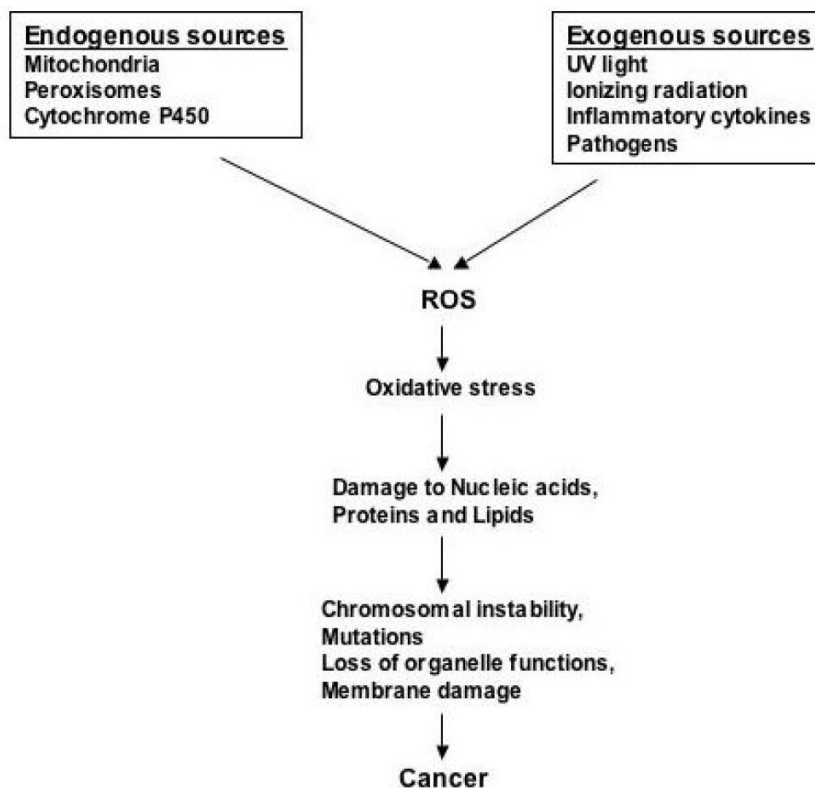


Figure 2: The endogenous and exogenous sources that result in ROS production, its cellular effects and role in carcinogenesis are depicted (Waris & Ahsan, 2006)

It is essential for a cell to keep the levels of ROS low. When cells lose the balance between the levels of ROS and antioxidants, the accumulation of extensive amounts of ROS results in oxidative stress which give damage to nucleic acids, proteins and lipids which might result in mutations, instability in the chromosomal structure, damage to membranes and in turn loss of normal functioning of various organelles (Figure 2). Cells must maintain the balance in order to protect themselves from these oxidant effects. (Pizzino et al., 2017). Inability to re-maintain the balance results in various oxidant effects such as the damage given to the mitochondrial elements prevents the organelle to function normally, resulting in improper metabolic activities. It might also result in extensive cytochrome c release which in turn activates apoptosis. (Murphy, 2009). Overall, those effects might result in transformation of a normal cell into cancer if not prevented.

1.3.2    Cellular Mechanisms and Antioxidants Against Oxidative Stress

Cells have adapted different enzymatic or non-enzymatic mechanisms to reverse the cellular effects of oxidative stress. Enzymatic strategies based on the activities of various enzymes, including SOD and GPX (Figure 3). When $O_2^-$ is generated inside the mitochondria, it is reduced to $H_2O_2$ by superoxide dismutase (SOD) enzyme in the mitochondrial intermembrane space and matrix. It is then fully reduced to water molecule by glutathione peroxidase (GPX). The non-enzymatic mechanisms use antioxidant molecules such as lipoic acid, glutathione, coenzyme Q10, etc. to reduce ROS molecules water molecules.  (Pizzino et al., 2017).
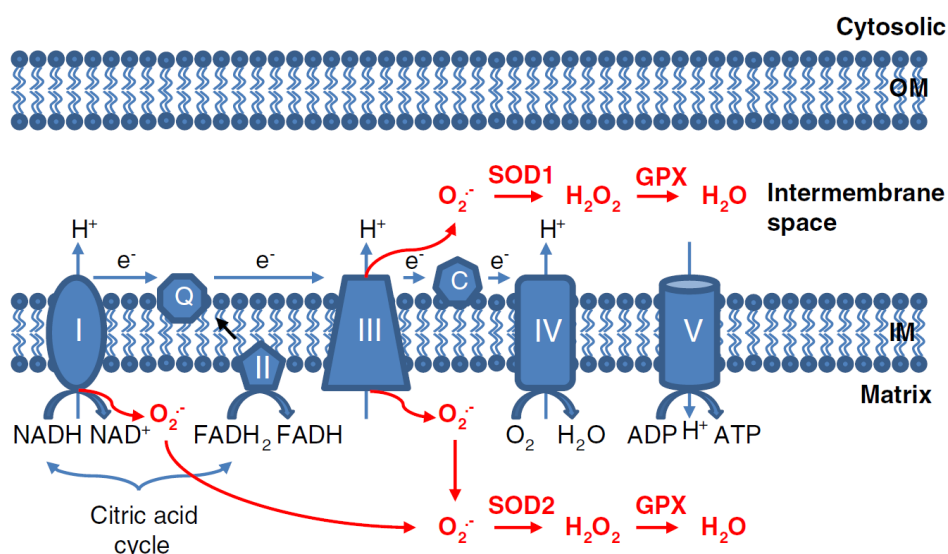


Figure 3: mtROS poductions mechanism and the cellular mechanisms to reverse the ROS effects inside the mitochondria. (Li et al., 2013).

**1.4 Cellular Antioxidant Role of Selenium**

Selenium (Se) is a trace element which is required for human health. It is found in various nutrients such as milk, meat, fruits & vegetables, fish and especially Brazil nuts and the deficiency of it results in various abnormalities. Se is found in soil, so it is inserted to the food chain through the plants; therefore, Se concentration in the soil of a region is the primary determinant of its regional consumption levels. (S. Darvesh & Bishayee, 2010)

Se is found in the structure of selenocysteine (SeCys) amino acid, which is the building block of selenoproteins. There are about 25 different selenoproteins identified in human with various functions including antioxidant and redox signaling function (Figure 4).

**HUMAN SELENOPROTEOME**

| Antioxidant functions | Redox signalling | Thyroid hormone metabolism | Selenocysteine synthesis | Se transport and storage | Protein folding | Unknown |
|---|---|---|---|---|---|---|
| GPx1 | TrxR1 | DIO1 | SPS2 | SelP | Sep15 | SelH |
| GPx2 | TrxR2 | DIO2 | | | SelN | SelI |
| GPx3 | TrxR3 | DIO3 | | | SelM | SelO |
| GPx4 | | | | | SelS | SelT |
| GPx6 | | | | | | SelV |
| SelK | | | | | | |
| SelR | | | | | | |
| SelW | | | | | | |

Figure 4: The known human selenoproteins and their functions. (from (Papp et al., 2007)

## 1.4.1 The Importance of Selenium to Human Health

As mentioned in the previous section, Se is an essential element in human health and the deficiency of it was shown to be related to the development of many diseases including metabolic, neurodegenerative, cardiovascular disorders and even cancer. Among the selenoproteins identified for today, many of them have key roles in signaling, antioxidant defense and redox signaling-related functions to maintain the balance. A mentioned in section 1.3.2, GPX proteins have key roles in the reduction of both hydrogen and lipid peroxides; while TXNRD proteins are necessary elements for the maintenance of balance in thiol system. Many others are responsible for the proper protein folding, thyroid hormone metabolism and endoplasmic reticulum stress handling. There are still many identified selenoproteins with unknown functions and should be further studied. (Guillin et al., 2019; Papp et al., 2007).

## 1.4.2 Selenium-Dependent Glutathione Peroxidases

GPX family selenoproteins constitutes one of the most important elements as antioxidant defense mechanisms. Until now, 8 different GPX paralogs were identified the catalytic unit of 5 of which are composed of selenocysteine and therefore categorized as selenoenzymes; while the remaining 3 are composed of cysteine. GPX1 and GPX4 are the most widely expressed selenoproteins among humans. (Guillin et al., 2019).

GPX proteins uses glutathione as cofactor to provide the reduction of $H_2O_2$ to a $H_2O$ molecule. During this transformation, the oxidation of glutathione tripeptide occurs to form oxidized glutathione and it is transformed back to glutathione with the use of NADPH by glutathione reductase. The concentrations of NADPH and GPX is

therefore important for a cell to maintain the oxidant-antioxidant balance and the deficiency of them were shown to result in inflammation and atherosclerosis in mice. (Li et al., 2013).

### 1.4.3   Selenium-Dependent Thioredoxin Reductases

In addition to the glutathione reductive system, thioredoxin is another key mechanism that is responsible for the reduction of molecules in human cells. Thioredoxin enzymes are composed of 3 selenoproteins (TXNRD1-3) so their functioning is strictly dependent on the presence of Se. They have essential roles in the reduction of protein sulfides as well as $H_2O_2$, which make them important elements in antioxidant defense mechanisms, DNA synthesis and formation of disulfide. (Guillin et al., 2019).

In humans, TRX protein is found in the mitochondria, and they play oxidoreductase roles by interacting with other proteins. In previous studies performed on mice, it was found that the absence of Trx2 resulted in increased apoptosis and vascular abnormalities. It was also shown that Trx2 functioning have regulated the permeability of mitochondria to protect the cells against oxidant-dependent apoptosis. (Li et al., 2013)

## 1.5 Oxidative Stress and Hepatocellular Carcinoma

One of the major driving forces for carcinogenesis is the DNA damage caused by ROS; which is either a result of improper repair functioning or the accumulation of extensive amounts of ROS. Oxidant elements were shown to have both mutagenic and carcinogenic effects on DNA which make them one of the potential accountable for the initiation, promotion and progression of carcinogenesis. (Waris & Ahsan, 2006).

Molecular mechanisms involved in liver cancer are more complex than other cancers. One of the apparent characteristics of liver cancer cells is their increased resistance to various stress conditions such as chronic viral infections or toxins.  It is essential to determine the stress response gene expression profiles of these cells due their involvement in hepatocarcinogenesis.(di Maso et al., 2015)

### 1.5.1   HBV-HCV Infection and ROS Production

As explained in the previous sections, ROS concentrations has various roles in cellular activities including metabolic regulations. When a virus infects a cell; it uses host mechanisms to replicate and maintain its metabolic activities; so, the concentration of ROS also directly affects viral growth and host defense mechanisms against the infection (Figure 5). The above-mentioned defense mechanism includes the initiation of phagocyte activity which is linked to ROS production and the phagocyte response itself also results in cytokine release which serves as a pro-oxidant. When HBV or HCV virus enters human liver cell, they increase the ROS levels by driving the production of pro-oxidants through the host defense mechanism as well as preventing the functioning and production of antioxidant enzymes. The increased metabolic by-

products due to viral replication activity is another factor that result in elevated ROS production. The increased ROS levels in liver cells might accelerate and enhance the viral replication. (Waris & Ahsan, 2006)
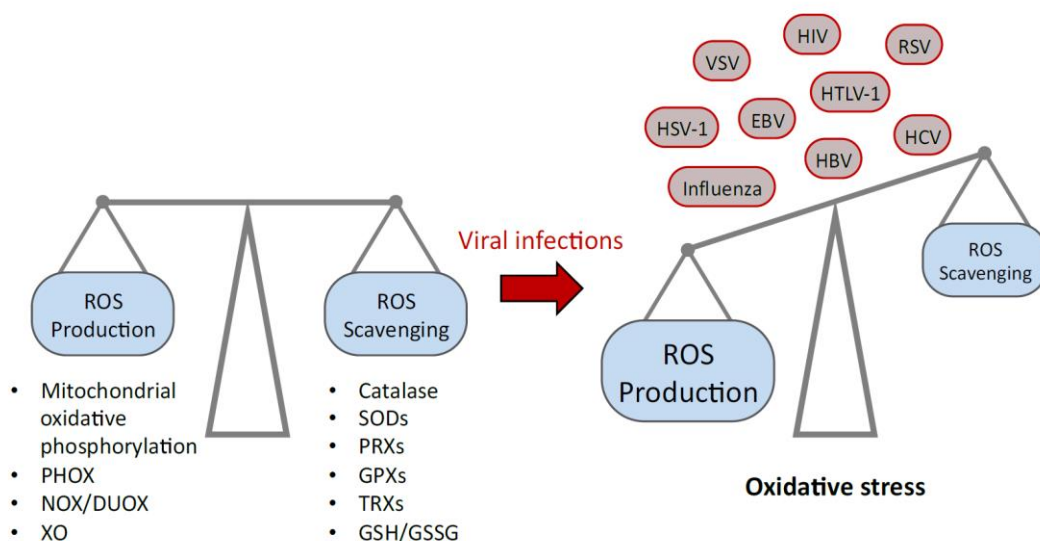


Figure 5: The factors providing the balance between ROS and antioxidant mechanisms are depicted. This balance might be disrupted during a viral infection, which in turn results in oxidative stress. (Guillin et al., 2019).

### 1.5.2 Oxidative Stress-Dependent Hepatocarcinogenesis

One of the most well-known biomarkers of oxidative stress is 8-hydroxyl-2-deoxyguanosine (8-OHdG), which is a frequently seen form of ROS-induced oxidative lesions. They are also accepted as indirect cancer biomarkers because of their potential mutagenic effects. (Valavanidis et al. 2009). They are accounted as one of the potential responsible for GC →TA transversions and CC → TT substitutions occurred in Ras oncogene and p53 tumor suppressor gene in HCC. (Cooke et al., 2003).

Transition of a hepatic cell to liver cancer frequently occur as a result of HBV or HCV infection or aflatoxin exposure which is linked to the increased oxidative stress levels after the cells are infected. The viral replication is promoted by higher ROS concentrations inside the host cells through the initiation of expression of oncogenic factors and suppression of tumor suppressor genes leading to hepatocyte survival and carcinogenesis. (Waris & Ahsan, 2006).

In previous studies, the initiation of NRF2 expression was shown to have important roles in the survival of HCC cells despite the increased ROS levels inside cells. It was found that in 5-15% of HCC cases, NRF2 or its inhibitory gene KEAP1 were mutated resulting in activation of NRF2 gene leading to tumor cell survival. (Levrero & Zucman-Rossi, n.d.).

**1.6 Selenium and Hepatocellular Carcinoma**

There are various *in vitro* and *in vivo* studies that show the inverse correlation between Selenium levels and the risk of HCC development. Many investigators have focused on the molecular mechanism behind this correlation and several of these studies have found that deficiency of Selenium resulted in reduction of antioxidant functions which lead to carcinogenesis. (Casaril et al., 1994; Czeczot et al., 2006). There are various meta-analyses studies that further show the correlation between the Selenium-deficiency and the HCC development incidence in humans. (Gong et al., 2019; Z. Zhang et al., 2016) In addition to these studies, Yu et al. focused on the effects of Selenium supplementation on HBV+ patients for the risk of hepatocarcinogenesis and found a decreased risk of HCC development in HBV-carriers after Selenium supplementation. (M.-W. Yu et al., 1999; Yu Yu et al., 1997). The protective effects of Selenium were not limited to HCC development, and it was shown that low Selenium levels were associated to cirrhosis and other viral liver disorders as well. (S. Darvesh & Bishayee, 2010)

**1.7 High Throughput Techniques in Biomarker Discovery**

One of the widely used techniques for the discovery of disease-related target genes to propose as candidate biomarkers is high-throughput expression methods and their analysis. Researchers gather extensive amount of expression profile information from various transcriptome-wide gene expression analysis methods including RNA-seq and microarray. The next challenge is the analysis of this expression result to extract meaningful data. (Nacu et al., 2007).

Generally, the analysis starts with the normalization and background correction of the raw expression data. Then, to determine differentially expressed genes (DEGs) between different phenotypes various statistical methods are available and chosen depending on experimental design. The DEGs could be filtered by their p-values and fold-change values and the most significant DEGs could be further used to investigate for their molecular importance and association to disease. (Nacu et al., 2007).

A single gene expression value itself has limited value, since genes and their products function in interaction with others and with DNA in complex ways. Because of this reason, most researchers have been focused on strategies that takes groups of genes and proteins into consideration including network analysis methods, Gene Ontology (GO) analysis and Gene Set Enrichment Analysis (GSEA). (Grimes et al., 2019; Nacu et al., 2007).

1.7.1   Expression Microarrays

High-throughput expression analysis methods such as transcriptome analysis by microarray or RNA-seq studies have great potential to determine the changes occur at

the molecular level in a cell to enlighten initiation and development of disorders including hepatocellular carcinogenesis. By using such methods, the molecular properties of a cancer cell could be detected by comparing gene expression values in the two different cell types- such as diseased and healthy or benign and malignant, etc. to determine thousands of DEGs in a very short time to further investigate gene expression profiles. (S. Xie et al., 2019). This information can guide target-based approaches to develop novel therapy strategies. (B. Chen et al., 2020b).

Microarrays are useful tools to identify expression levels of many genes simultaneously. Oligonucleotide probes are printed on known positions on a chip to function as probes for the detection of complementary binding of messenger RNA (mRNA) transcripts isolated from samples of interest. mRNA of the reference and experimental samples are isolated and complementary DNA (cDNA) synthesis is performed for each sample, which are then labelled with fluorescent dye either two-channeled or single-channeled. For the two-channel analysis; reference and experimental samples are labelled with separate colors of dye and their mixture is put on array chip for hybridization to compare expression levels relative to each other while one-channel analysis uses the same dye color for all sample types (Figure 6). Microarray is than scanned to detect gene expression levels of each gene in each sample group. (Slonim & Yanai, 2009).
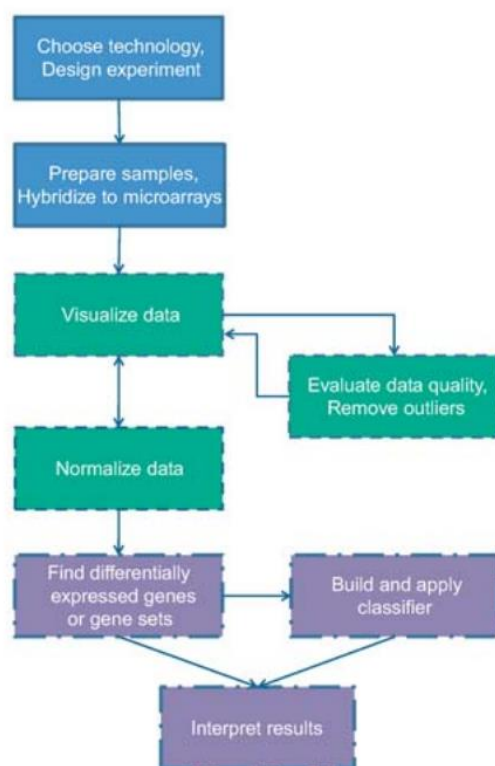


Figure 6: The experimental and data analysis steps of an expression microarray.

Examination of data quality and elimination of the low-quality results and the normalization of the raw data are the first steps in the analysis of the microarray reads (Figure 6). Normalization aims to clarify biological variations by preventing the effects of technical differences (Slonim & Yanai, 2009). The next step is to define the differentially expressed genes by comparing the samples of interest depending on the biological question and interpret the results by using various tools available which will be mentioned in more detail in the following parts.

### 1.7.2 Differential Gene Expression Analysis and Linear Models for Microarray Data Package

In the analysis of the gene expression data, one of the challenges that researchers encounter during statistical comparison is the multifactor designed experiments and the small sample size. In order to overcome this problem, novel statistical techniques have been developed. Linear Models for Microarray Data (limma) is an R programming language package for the analysis of gene expression data such as microarray, RNA-seq and quantitative PCR from the beginning- pre-processing steps- to the end. (Ritchie et al., 2015).

limma uses expression values in a matrix format, each gene is represented in the rows while different samples are collected in each column. Each gene in the rows is fitted to a linear model providing a flexible design for the analysis of such complex data. The linear model used by this approach has the strength to analyze all data together instead of pairwise comparisons of samples allowing information sharing across genes.

The statistical approach used in limma is empirical Bayes method which uses the distribution of all standard variation values to estimate a moderated standard deviation to be used in the t-test denominator to perform moderated t-test. This provides the power of limma in the analysis of small sample numbers by including information between genes.

The first step in limma analysis is to fit all the expression values to a linear model which fully models the systematic part of the data. Each row of the design matrix corresponds to a gene in the experiment and each column corresponds to a sample. Then, the contrast step allows the fitted coefficients to be compared in as many ways as the biological questions to be answered.

### 1.7.3 Network Analysis

In order to gather meaningful results from the gene expression data obtained from sequencing or microarray techniques that were mentioned in the previous section, investigation of the functional properties and relations between DEGs and the related

proteins is needed. As already explained, genes interact with each other in biological pathways; and those gene expression levels are correlated either because of regulation by a common transcription factor or because of their regulatory roles over each other. Therefore, analyzing the genes as a whole in a network with the topology information gives more insight about their functioning. For instance, given a differential network, hub genes in the network with higher node degree values might have key roles in the regulation of expression of the connected genes or the connected genes in a branch might be responsible for roles in a common pathway; therefore, might have key roles in the disease under study. (Grimes et al., 2019; Zuo et al., 2016).

Protein-protein interaction (PPI) networks are also powerful tools to study molecular mechanisms behind diseases. They could be used to gain insight about the pathways that are associated to the disease to determine biomarkers for the diagnosis and treatment of it. They are especially widely used for the discovery of targeted therapy strategies in cancer. (Zuo et al., 2016). KEGG pathways, Reactome pathways, String database are all used to integrate the pathway information into DEG analysis (Grimes et al., 2019).

### 1.7.3.1 String Database

STRING database collects protein-protein interaction information that are proved as well as predicted. These protein-protein interaction might be either directly through physical attraction or indirectly because of their functional association. This interaction information comes from various sources, which could be selected and filtered by users depending on the confidence interval of interest. The main sources for the data are experimental results, co-expression, previously known interactions from other databases, genomic predictions, and automated text-mining results. The database contains 24584628 proteins information to date, coming from 5090 organisms. (Szklarczyk et al., 2019).

### 1.7.3.2 Prize Collecting Steiner Tree (PCST) Algorithm

It is generally difficult to interpret the network data due to the noise and large size of it and various strategies were developed to solve this problem. Solving PCST problem is one of these approaches to reach more optimized interaction data. PCST constructs optimum trees from the given DEGs or proteins (terminal nodes) by using human interactome data as a reference to connect them to each other directly or through hidden, undetected nodes (Steiner nodes) through the shortest path between the nodes. All the terminal nodes do not have to be included in the tree, which is one of the strengths of the algorithm to provide higher-confidence trees. (Tuncbag et al., 2012). The algorithm constructs trees based on cost of interactions showing whether the interaction is real and prizes to exclude any terminal determined by confidence in proteomic/transcriptomic data. It is designed to maximize the prize by including most of the genes from the given gene list and goes through the nodes of interactome data

by taking the edge weight values as cost to find the shortest path by minimizing the cost. (Tuncbag et al., 2012)

In order to construct meaningful trees, some parameters have to be optimized first; β, ω, and μ are the most important input parameters that must be chosen correctly. ω is the value that determines the number of the generated trees by giving a penalty for the construction of each new tree. β is the value that determines the hub frequency in the tress; the higher it is the more hubs there are and the bigger the generated trees are. Similar to this value μ also controls the hubs generated, by competing the degree of nodes to prevent any biased information. It clarifies a gene/protein being a hub whether because it is a highly studies gene/protein so have many determined interaction information, or it really has a greater degree in the interactome data. The smaller this parameter is, the less control we get over the biased hub information.

The optimization of the aforementioned input parameters has great importance for the generation of accurate trees. Forest-tuner script is a publicly available script in github ((https://github.com/gungorbudak/forest-tuner) which runs forest for every combination of parameters within the pre-determined interval to make it possible to choose the most optimum values for each DEG lists through these options.

### 1.7.4   Gene Set Enrichment Analysis (GSEA)

GSEA is a computational tool to test the presence of any significant difference in the enrichment of gene sets of interest such as known pathways or GO ontologies between two different groups of samples (Subramanian et al., 2005). These gene sets are composed of pre-defined genes of interest and the average value of the test-statistics of its individual components is used to determine a group score. The higher the group score, the more likely they are to be expressed differentially between the compared samples. (Nacu et al., 2007).

### 1.8 The Aim and Motivation of the Thesis

The first part of this thesis focuses on the discovery of novel biomarkers by the examination of oxidative stress response genes to be used as potential targets in HCC treatment. Normally, deficiency of Selenium results in oxidative stress, leading to apoptosis. However, in a previous study performed with HCC cell lines, it was found that 10 of 13 HCC cell lines tolerate Selenium-deficiency to escape apoptosis and most of these tolerant cell lines had HBV sequence integrated in their genome, indicating that this virus might have a role in that acquired tolerance (Irmak et al., 2003). A new experiment was designed with two isogenic HCC cell lines; HepG2 and HepG2-2.2.15 cells to test their response to Selenium-deficiency (Figure 7). These two cell lines have the same genome, except the HBV genomic integration in HepG2-2.2.15 cells and it was found that HepG2 cells were Selenium-deficiency sensitive while HepG2-2.2.15

cells tolerated the absence of Selenium to survive. It could be a result of a cryptic mechanism acquired during hepatocarcinogenesis under viral stress.

During the viral genome integration into HepG2-2.2.15 cells; an increase in ROS generation is expected to be seen in the cells which alters the cellular gene expression (Waris & Ahsan, 2006). This would result in an intrinsic variation in the gene expression of HepG2-2.2.15 cells compared to HepG2 cells; independent of the Selenium status of the medium that the cells are grown. In addition to this intrinsic effect, we also suspect the presence of an acquired mechanism depending on the absence of Selenium, which controls the differential response of the two cell lines to Selenium-deficiency. Understanding the molecular mechanism behind this tolerance might give some knowledge about cellular response mechanisms gained by some cancer cells to escape from oxidative stress dependent apoptosis. The resistance genes that we identified can be further exploited as drug targets or diagnostic purposes.
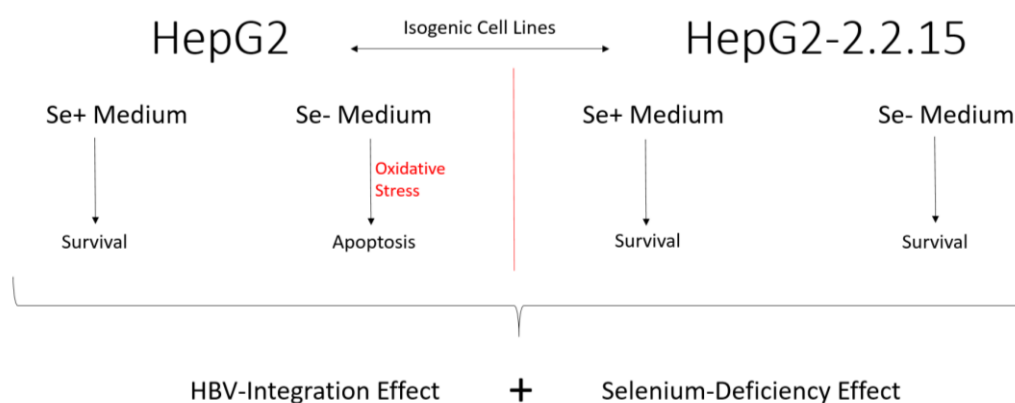


Figure 7: Experimental model design. HepG2 and HepG2-2.2.15 cells are two isogenic hepatocellular carcinoma cell lines with the difference of HBV genome integration in 2.2.15 cells.

In the first part of our study, FOXA1 gene was identified as one of the potential biomarkers for HCC. From previous studies, it was known that FOXA1 binds to MDM2 chromatin region to regulate its transcription resulting in decreased p53 levels (Swetzig et al., 2016). There were also many correlations detected between the expression levels of DEGs identified in the first part and the genes that have roles in p53-MDM2 pathway in our samples of interest. Depending on this association targeting p53-MDM2 mechanism might be a potential therapeutic for HCC.

In the second part of this thesis, we aimed to identify anticancer activities of potential inhibitors of p53-Mdm2 interaction on HCC cell lines with various p53 mutation status and define the underlying mechanisms. HepG2 cells has wild-type gene whereas,

Huh7 and Mahlavu cells have codon Y220C and R249S mutations respectively, and Hep3B cells do not express p53 protein at all.

# CHAPTER 2

# MATERIALS AND METHODS

In this section, the experimental and bioinformatics analysis methodologies are presented. First part focuses on the wet lab experiments used in the determination of biomarkers and the testing of the effects of the compounds that inhibits Mdm2-P53 interaction on HCC cell lines. Then, in the second part, bioinformatics approach used to analyze the gene expression data was explained.

## 2.1 Wet Lab Experiments

### 2.1.1   Cell Lines and the Microarray Experiment

Gene expression analysis data by Human Genome U133 Plus 2.0 Array Affymetrix Array was previously acquired within the framework of TUBITAK 104S333 from the microarray experiment performed in Ankara University, Institute of Biotechnology.

For the verification of cell confluency in the presence or absence of Se, the oxidative stress resistant HepG2-2.2.15 cell line (HT-HBV integration) and sensitive HepG2 cell line were grown in DMEM medium (Gibco, Thermo Fisher Scientific, MA, USA) with 0.01% FCS (BioChrom, Berlin, Germany) either supplemented with 0.1 μM $Na_2SeO_3$ (Sigma, Taufkircher, Germany) (Se positive) or not (Se negative) for 3 days in duplicates at 37°C under 5% CO2. The live images of each cell line grown in both conditions were taken under inverted microscope separately for each day.

For the second part of the study, four HCC cell lines, HepG2, Huh7, Hep3B and Mahlavu were cultured in DMEM medium (Gibco, Thermo Fisher Scientific, MA, USA) supplemented with 10% fetal bovine serum and 1% Penicillin/Streptomycin antibiotic and kept in cell culture incubator at 37 °C under 5% CO2.

### 2.1.2   Cytotoxicity Assay for the Testing of p53-MDM2 Interaction Inhibitors

Cells were cultured on 96-well plates (1000-3000 cell/well) to test the cytotoxic activities of compounds. After 24 hours of incubation at 37 °C, treatments with compounds or DMSO control were done in 5 different concentrations (40 μM, 20 μM,

10 µM, 5 µM and 2,5 µM) of each compound in triplicates. After 72 hours of treatment, 10% TCA was used to fix the cells. The cells were stained with 0.4% sulforhodamine B solution. The absorbance values for each sample were determined at 515 nm in a microplate reader and percent inhibition values were calculated.

### 2.1.3   Cytotoxicity Assessment with Real-Time Cell Analyzer

Cytotoxicity Assessment with Real-Time Cell Analyzer: RT-CA, xCELLigence System (Roche Applied Sciences) was used to perform real-time cell analysis. Cells were cultured on 96-e-plate (Roche Applied Sciences). The proliferation of cells was monitored for 24 hours in a cell culture incubator at 37 °C. Treatments with AM139 compound were done in triplicates at 3 different concentrations (20 µM, 10 µM and 5 µM). The cell index values were recorded for 72 hours to calculate the cell growth ratios.

### 2.1.4   Cell Cycle Analysis

Cells were cultured on 100 mm$^2$ culture dishes. After 24 hours of incubation at 37 °C, they were treated with compounds at IC100 concentrations, or DMSO as control for 48 hours. Then, they were fixed with ice-cold 70% ethanol. MUSE Cell Analyzer was used to perform cell cycle analysis according to the manufacturer's recommendations (Millipore).

### 2.1.5   Immunofluorescence Assay and Assessment of Cell Morphology

Cells were cultured on coverslips in 6-well plates. After 24 hours of incubation, cells were treated with the compounds at IC50 concentrations for 72 hours. The morphologies of live cells were analyzed under phase-contrast microscope at 24, 48 and 72 hours of treatment. They were fixed with 4% paraformaldehyde. After blocking with 3% BSA, immunofluorescence assay was performed by using 7D3 monoclonal α-p53 antibody and Alexa-488 conjugated α-mouse (Invitrogen) as the secondary antibody. Counterstaining was done by Hoechst 33258 (Sigma-Aldrich) and the slides were mounted on glass by glycerol. The results were examined under fluorescence microscope.

## 2.2 Bioinformatics Methodology

### 2.2.1   Pre-processing of Microarray Data

The qualities of arrays were checked using simpleaffy package available in R (Wilson & Miller, 2005). Robust Multi-Array Average (RMA) background correction and quantile normalization of the data were performed using the R-script provided in Appendix A. The distribution of expression intensities before and after the normalization for each sample is given in Figure 8.
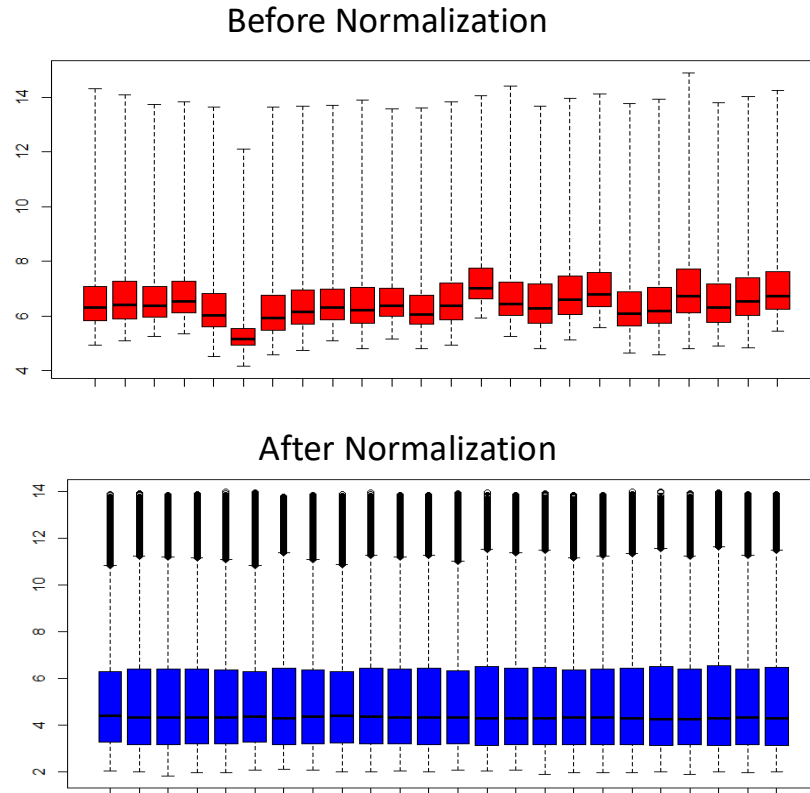
Figure 8: The boxplot showing the distribution of gene expression intensities before and after the normalization.

### 2.2.2    Determination of Differentially Expressed Genes

In order to identify DEGs between HepG2 and HepG2-2.2.15 cells in the presence or absence of Se, we used limma package (Ritchie et al., 2015) in R programming language (Appendix A). We organized the design matrix so that it involves both the cell line (HepG2 or HepG2-2.2.15) and Se status (present or absent) information, separately for each day. The first step was to fit all the data to a linear model which fully models the systematic part of the data. Each row of the design matrix corresponds to an array in the experiment each belong to one of the 24 different samples and each column corresponds to a coefficient with the cell line and Se status information. In Table 1, the design matrix for only one day was shown for simplicity (8 of the 24 total arrays) and each sample (each column) includes two arrays due to the replicas.

Table 1: The design matrix used in limma analysis showing the arrays and the samples belong to them.

| Sample No | HepG2 Se- | HepG2-2.2.15 Se- | HepG2 Se+ | HepG2- 2.2.15 Se+ |
|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 1 |
| 2 | 0 | 0 | 0 | 1 |
| 3 | 0 | 1 | 0 | 0 |
| 4 | 0 | 1 | 0 | 0 |
| 5 | 0 | 0 | 1 | 0 |
| 6 | 0 | 0 | 1 | 0 |
| 7 | 1 | 0 | 0 | 0 |
| 8 | 1 | 0 | 0 | 0 |

In limma analysis, the contrast step, provides the comparison of the fitted coefficients in as many ways as there are questions to be answered, which is independent of the number of distinct RNA samples. In line with the design matrix, we designed the contrast matrix in a specific way to determine the effects of cell line and Se status on gene expression separately for each day (Table 2);

1. **Se+** HepG2 vs HepG2-2.2.15 = Se+ HepG2 – Se+ HepG2-2.2.15
2. **Se-** HepG2 vs HepG2-2.2.15 = Se- HepG2 – Se- HepG2-2.2.15
3. **HepG2** Se- vs Se+ = Se- HepG2 – Se+ HepG2
4. **HepG2-2.2.15** Se- vs Se+ = Se- HepG2-2.2.15 – Se+ HepG2-2.2.15

Table 2: The contrast matrix used in limma analysis showing the comparisons performed in each of the 4 analyses.

| | 1st Analysis | 2nd Analysis | 3rd Analysis | 4th Analysis |
|---|---|---|---|---|
| | Se+ I HepG2 vs HepG2-2.2.15 | Se- I HepG2 vs HepG2-2.2.15 | HepG2 I Se+ vs. Se- | HepG2-2.2.15 I Se+ vs. Se- |
| **HepG2 Se-** | 0 | 1 | -1 | 0 |
| **HepG2-2.2.15 Se-** | 0 | -1 | 0 | -1 |
| **HepG2 Se+** | 1 | 0 | 1 | 0 |
| **HepG2-2.2.15 Se+** | -1 | 0 | 0 | 1 |

In our design, the first two comparisons were interested in the identification of the genes that were differentially expressed between HepG2 and HepG2-2.2.15 cells in the presence or absence of Se, respectively and named as *between cell line* comparisons; while the last two comparisons aimed to answer the question of which gene expressions were altered within a cell line in the presence or absence of Se for HepG2 or HepG2-2.2.15 cells respectively and called as *within cell line* comparisons. The p-value was limited to values lower than 0.01 while selecting the significant DEG lists.

The generated DEG list were compared to each other to find the shared DEGs by drawing venn diagrams. Online Venn Diagrams web-tool by Van de Peer Lab that is available at http://bioinformatics.psb.ugent.be/webtools/Venn/ was used.

## 2.2.3    Clustering of Differentially Expressed Genes

Heatmaps are one of the most useful technique for the visualization and clustering of large datasets. Samples with more similar profiles clustered closer separately from others with distinct profiles. Firstly, in order to draw heatmaps with our DEG lists, we calculated Z-score of each gene in the defined DEG lists (Kreyszig, 1979). To calculate the Z-scores, the average of log2 expression value for each gene were calculated and subtracted from gene expression value of each individual sample. The calculated values were then divided by standard deviation values of each gene. heatmaply package (Galili et al., 2018) in R Programming was used to draw the heatmaps by default parameters. Hierarchical clustering method was used, and the distance was computed by Euclidean measure. R script is provided in Appendix A.

## 2.2.4    DEG Scores

We normalized *within cell line* comparison result with respect to each other by assigning each gene a score- called DEG score- according to the change they showed upon Se-deficiency in comparison of the LIMMA results of the two isogenic cell lines; as shown in the equations in Figure 9 (Cavga et al., 2019):

$$\left| \log_2 \left( \frac{\dfrac{\text{HepG2-2.2.15 Se-}}{\text{HepG2-2.2.15 Se+}}}{\dfrac{\text{HepG2 Se-}}{\text{HepG2 Se+}}} \right) \right| > 1$$

$$\frac{\dfrac{\text{HepG2-2.2.15 Se-}}{\text{HepG2-2.2.15 Se+}}}{\dfrac{\text{HepG2 Se-}}{\text{HepG2 Se+}}} > 2 \qquad \log_2 \left( \frac{\dfrac{\text{HepG2-2.2.15 Se-}}{\text{HepG2-2.2.15 Se+}}}{\dfrac{\text{HepG2 Se-}}{\text{HepG2 Se+}}} \right) > 1$$

Relatively upreguated in HepG2-2.2.15
normalized to HepG2

$$\frac{\dfrac{\text{HepG2-2.2.15 Se-}}{\text{HepG2-2.2.15 Se+}}}{\dfrac{\text{HepG2 Se-}}{\text{HepG2 Se+}}} < 0.5 \qquad \log_2 \left( \frac{\dfrac{\text{HepG2-2.2.15 Se-}}{\text{HepG2-2.2.15 Se+}}}{\dfrac{\text{HepG2 Se-}}{\text{HepG2 Se+}}} \right) < -1$$

Relatively downregulated in HepG2-2.2.15
normalized to HepG2
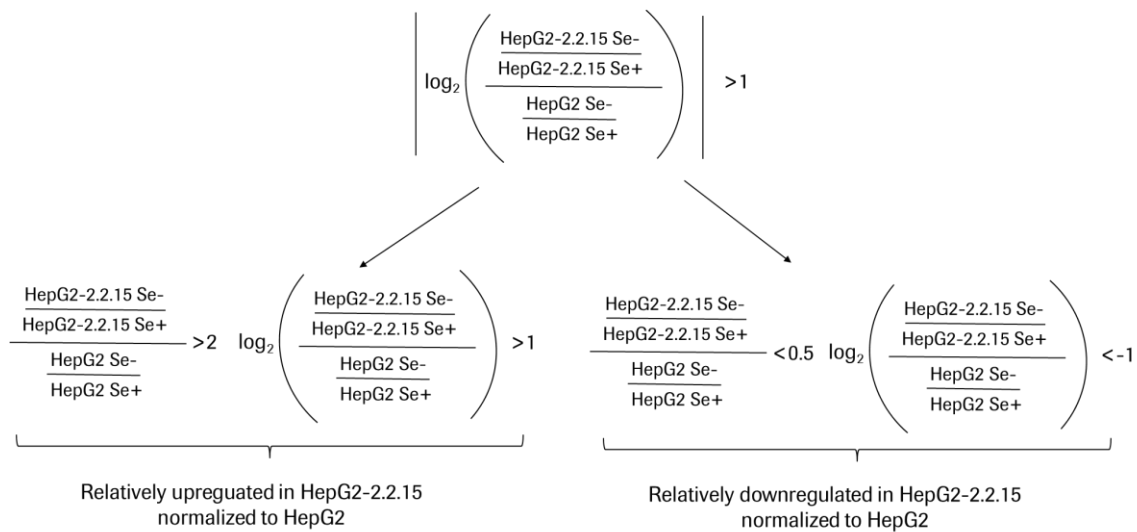
Figure 9: Calculation of DEG scores to normalize comparison results with respect to each other.

DEG scores have provided the normalization of the results for each DEG relative to each other which were calculated by the formula provided above to determine the relative changes between HepG2 and HepG2-2.2.15 *within cell line* comparison results. The DEGs with absolute DEG scores higher than 1; which indicates more than 2-fold change were identified. DEG score with value more than 1 is related to increased response to Se-deficiency in HepG2-2.2.15 cells while DEG score that are lower than 1 are associated to decreased response in HepG2-2.2.15 compared to HepG2 cells.

### 2.2.5   Gene Set Enrichment Analysis

Gene set enrichment analysis (GSEA) (Subramanian et al., 2005) was performed using the DEG lists by taking DEG score values as a pre-ranked input for each gene. Enriched gene sets were searched from either the MSigDB hallmark or GO-Biological Process data collection using the default "weighted" enrichment statistic parameter and "meandiv" normalization. The MSigDB hallmark gene sets are curated, highly reliable gene sets since they are composed of genes that show consistent expression patterns in specific biological pathways based on publications. Enriched gene sets with a FDR q-values < 0.25 were considered to be significant.

### 2.2.6   Network Analysis

### 2.2.6.1 Prize Collecting Steiner Tree

We used Prize Collecting Steiner Tree (PCST) (Tuncbag et al., 2016) approach in order to identify the interactions through our DEG lists by using the information provided by human interactome data (Appendix A). For running the approach, we used

Omics Integrator software's Forest module to determine the sub-network in our set of DEGs.

PCST constructs optimum trees from the given DEGs (terminal nodes) by using human interactome data as a reference to find the shortest path between the nodes. It is designed to maximize the prize by including most of the genes from the given gene list taking gene expression differences (log fold change values) as prize and goes through the nodes of interactome data by taking the edge weight values as cost to find the shortest path by minimizing the cost.

In order to construct meaningful trees, some parameters have to be optimized first. β, ω, and μ are the input parameters that must be chosen correctly which determine the number and size of trees, number of hubs, etc. For each DEG list, the optimum values of those input parameters must be determined separately. Forest-tuner algorithm (https://github.com/gungorbudak/forest-tuner) was used to find the optimum parameters which gives trees with higher number of prices and smallest mean degrees of nodes.

In our analysis, STRING protein-protein interaction database v10 was used to extract the interactome reference set. In this database, each edge is given a confidence score between 0 and 1 according to the reliability of the data source. We have chosen the edges that have confidence scores higher than 0.7. These scores were used to determine the costs by OmicsIntegrator. The network data generated was visualized by Cytoscape (Shannon, 2003).

2.2.6.2 String

Protein-protein interaction (PPI) networks are powerful tools to define some key proteins within known networks. The data needed for these networks can be extracted from databases such as STRING. Some values such as degree of a node and betweenness centrality might be used to determine biologically important hub proteins. The degree of a node is the total number of edges linked to that node; so the higher the degree the more central the protein is (Cavga et al., 2019). The betweenness centrality of a node shows whether that node acts as a bridge between two other nodes and higher betweenness centrality values indicate a key role in connecting different sub-networks. (H. Yu et al., 2007).

In this study, the identified DEG lists were further analyzed to gain more insight about their biological mechanisms by obtaining experimentally validated PPI information from STRING database (Jensen et al., 2009). Hub proteins with high degree or betweenness centrality values were further analyzed.

2.2.7   Correlation Analysis:

In order to evaluate whether the expression level changes of DEGs through our samples of interest are correlated to that of the genes playing roles in p53-MDM2

pathway, we performed Spearman correlation analysis. A matrix was generated for the DEG lists vs p53-MDM2 pathway genes and Rho values were calculated for each gene pair. Rho could be any value between 1 and -1; 1 being 100% positive correlation while -1 meaning 100% negative correlation between the gene expression values in samples of interest. p-values showing the significance of the correlation were also estimated and the values lower than 0.05 was taken as significant.

# CHAPTER 3

## RESULTS

Results of this thesis are presented in this chapter. The results of the bioinformatics methods that were used to analyze the microarray experiment were explained in the first part. We have used various strategies to interpret the microarray data to propose 27 genes that could be further searched for their potential to be used in the targeted therapy of HCC.

The remaining part of the chapter focused on experimental data for the testing of cytotoxicity of p53-MDM2 interaction inhibitor compounds and their mechanisms of action.

Figure 10 shows the general steps followed in the microarray study; starting from the experimental setup followed by the computational methods used in its assessment for the identification of potent biomarkers. Each topic was explained in detail in the related coming sections.
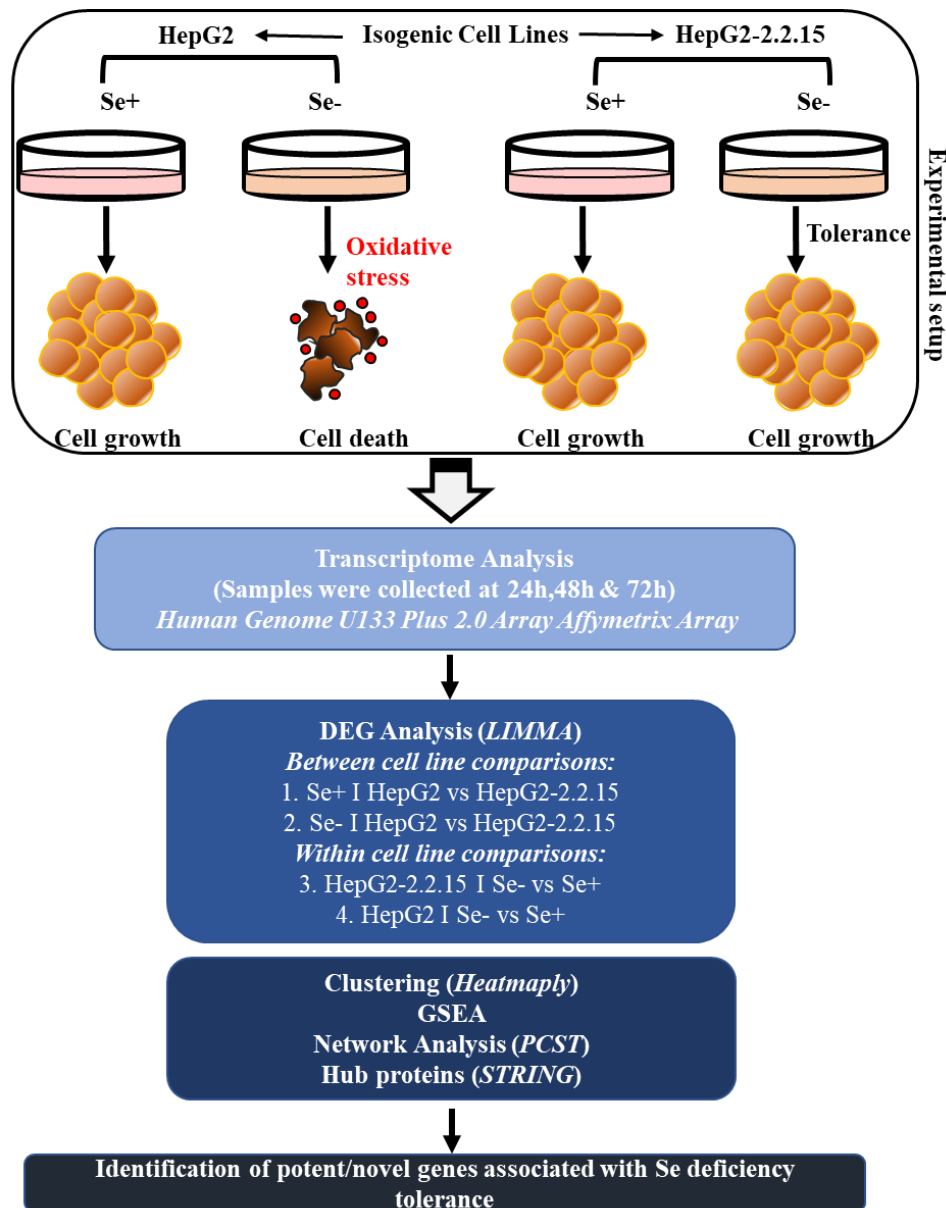
Figure 10: Experimental design. HepG2 and HepG2-2.2.15 cells- two isogenic HCC lines with the difference of HBV genome integration in HepG2-2.2.15 cells- were grown in the presence or absence of Se to perform transcriptome analysis. The results were examined by further bioinformatics methods to enlighten the differential response mechanisms.

## 3.1 Identification of DEGs by Linear Modelling of the Effect of Selenium-Deficiency Induced Cell Death in HCC Cells' Transcriptome Data

As mentioned previously, limma is a package used in the determination of DEGS in the analysis of microarray experiments. A linear model is fitted for each gene. Our transcriptome wide gene expression analysis result had three independent factors that

affected gene expression levels: 2 cell lines, 2 Se status and 3 days. In order to analyze the results of such a complex data and determine DEGs that are important for the tolerance to Se-deficiency dependent oxidative stress, we used limma package.

In this method, two matrices have to be identified: the design matrix provides the information of different RNA samples used in the analysis and the contrast matrix allows the coefficients defined by the design matrix to be compared to answer the biological question of interest. Each column of the contrast matrix corresponds to a comparison of interest between different RNA samples.

In our study, as explained in methods, four comparisons were performed to answer different questions separately for 24 hours (D1), 48 hours (D2) and 72 hours (D3): first two analysis aimed to identify the DEGs between cell lines; HepG2 vs HepG2-2.215 cells in the presence or absence of Se, respectively, while the last two analysis have given the DEGs within each cell line depending on Se status.

Figure 11 shows the live images of HepG2 and HepG2-2.2.15 cells grown in Se+ or Se- media for 72 hours. Cell confluency at the first 24 hours was comparable for both cell lines independent of the Se status, showing the Se deficiency effect was not apparent yet. In the 48$^{th}$ hour, the HepG2 cell numbers in Se- media was lower compared to Se+ control, while HepG2-2.2.15 cell numbers were similar in both conditions. The most dramatic effect of Se-deficiency was detected on the third day and the HepG2 cell numbers decreased dramatically compared to those of Se+ control, while HepG2-2.2.15 cell numbers were comparable in both conditions.

Numbers of DEGs for each condition were identified per comparison respectively as indicated in Figure 12 and 13. Venn diagrams show the individual or shared DEGs in different comparison lists separately for D1, D2 and D3. By the first two limma analyses *between cell line* comparisons (Figure 12) the DEGs that were only differentially expressed in Se- HepG2 vs HepG2-2.2.15 and not in Se+ HepG2 vs HepG2-2.2.15 were further named as 'Se-deficiency effect genes' since they were altered depending on the absence of Se. The shared DEGs in both Se+ and Se- comparisons were named as 'HBV-integration effect genes' in this study (Figure 11), as their expression differences were independent of the Se status, which is thought to be about the integration of HBV viral genome.
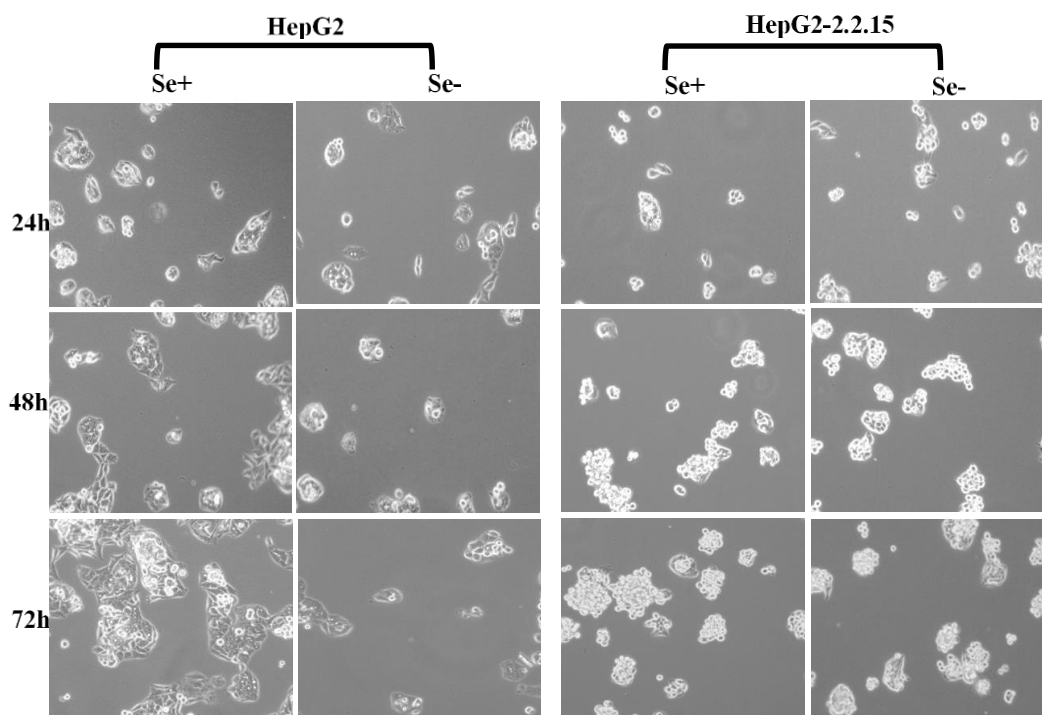
Figure 11: Experimental groups and DEG analysis of HepG2 and HepG2-2.2.15 cell lines. Live images of HepG2 and HepG2-2.2.15 cells grown in Se+ or Se- media for 72 hours.
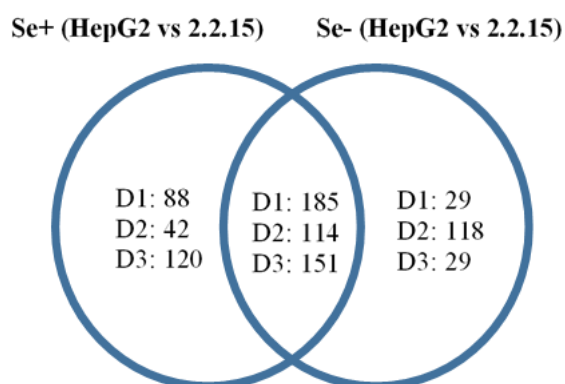


Figure 12: DEG numbers identified as a result of the between cell line comparisons by limma analysis. Venn diagrams were drawn in order to indicate the common and unique DEGs within indicated comparison groups for each day. 2.2.15: HepG2-2.2.15, D1: 24h, D2: 48h; D3: 72h.

By the last two limma analyses within each cell line comparisons (Figure 13) the expression of GPX and SEPW1, the two selenoproteins, decreased in the absence of Se in both cell lines. This could be expected due to the direct dependence of their

30

expression on the presence of Se. The DEGs that were not shared in common were named as 'Se-deficiency effect genes' and further examined since they were directly related to the different reactions each cell line gave to the deficiency of Se.
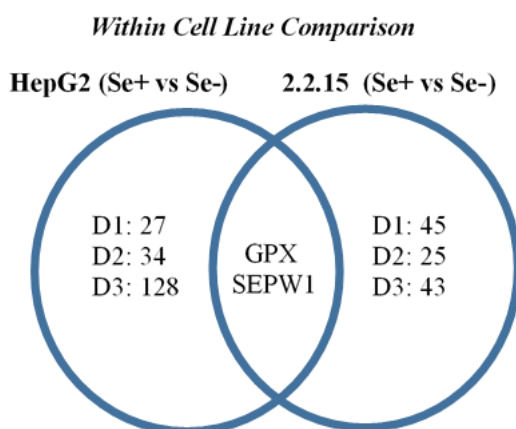


Figure 13: DEG numbers identified as a result of the within cell line comparisons by limma analysis. Venn diagrams were drawn in order to indicate the common and unique DEGs within indicated comparison groups for each day. 2.2.15: HepG2-2.2.15, D1: 24h, D2: 48h; D3: 72h.

## 3.2 Clustering Analysis of Between Cell Line Comparison DEGs under Selenium-Deficiency

For *between cell line* comparisons, the z-scores for expressions of the Se-deficiency effect genes in each sample were calculated by the method explained in the Materials and Methods chapter and used to draw heat map that clusters sample all together for D1, D2, and D3 data (Figure 14). As expected, HepG2 and HepG2-2.2.15 cells were clustered in two distinct groups due to their separate DEG expression profiles. PPAP2A, HOXD1, and CLYBL genes were found to be the most DEGs between the two cell lines because of the smallest p-values they have. HepG2-2.2.15 Se+ and Se- samples were clustered together in a day-wise fashion, while HepG2 Se+ and Se- samples were clustered in two distinct groups in a Se-status-wise fashion especially when the 2nd and 3rd day data is taken into consideration. This might be an indication of a more dramatic change in the gene expression profiles of HepG2 cells occur with respect to Se-deficiency; while HepG2-2.2.15 cells have similar DEG expression levels in both Se status.

31
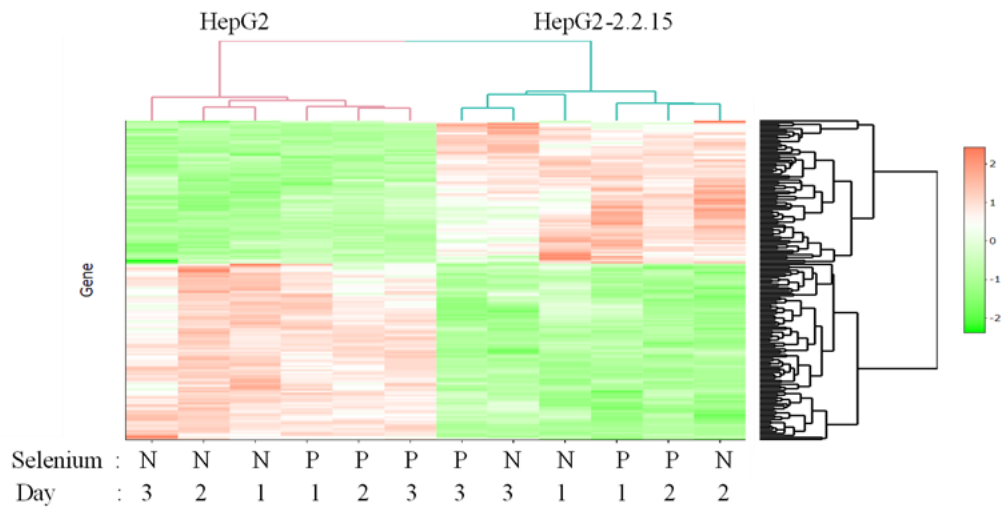
Figure 14: Heat map drawn with z-scores calculated for the Se-deficiency effect gene expression values.

A similar clustering was performed for the HBV-integration effect genes. HepG2 and HepG2-2.2.15 cells were clustered separately again; but this time both cell lines were clustered in a day-wise fashion within themselves (Figure 15). FGF13, GPC3 and MAP7D2 genes were identified as the DEGs with the smallest p-values between HepG2 and HepG2-2.2.15.
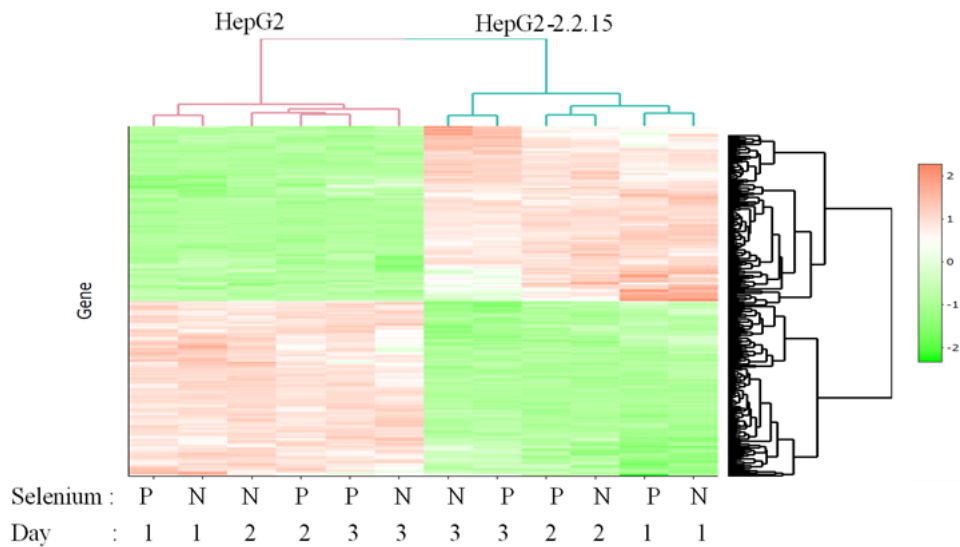


Figure 15: Heat map drawn with the z-scores calculated for the HBV-integration effect gene expression values.

The same DEG list was used to draw heat maps for the expression levels of 8 different cell lines taken from CCLE results in order to test whether the clustering of these cell lines for the expression levels of HBV-integration effect DEGs are dependent on the

presence of HBV in their genome. Table 3 indicates the HBV integration status of these cell lines as well as their Se-deficiency tolerance. Six of these HCC cell lines have HBV genomic integration and were found to be resistant to Se-deficiency dependent oxidative stress (SNU182, SNU475, SNU423, SNU387, SNU449 and PLC.PRF.5) in a previous study (Irmak et al., 2003)) while two of them (HepG2 and HUH7) were known to be virus-free and sensitive to Se-deficiency dependent oxidative stress.

Table 3: HBV-integration and Se-deficiency tolerance conditions of 8 cell lines were depicted in the table.

| Cell Line | HBV-Integration | Se-Deficiency Tolerance |
| --- | --- | --- |
| SNU182 | + | + |
| SNU449 | + | + |
| SNU423 | + | + |
| SNU387 | + | + |
| SNU182 | + | + |
| PLC/PRF/5 | + | + |
| HUH7 | - | - |
| HEPG2 | - | - |

Dendrograms showed clustering of the two Se-deficiency sensitive cell lines together; distinctly from the other 6 cell lines further supporting the HBV-integration effect hypothesis stating that differential expression of these genes was indeed related to the HBV genome integration independent of the cell's Se status. (Figure 16)
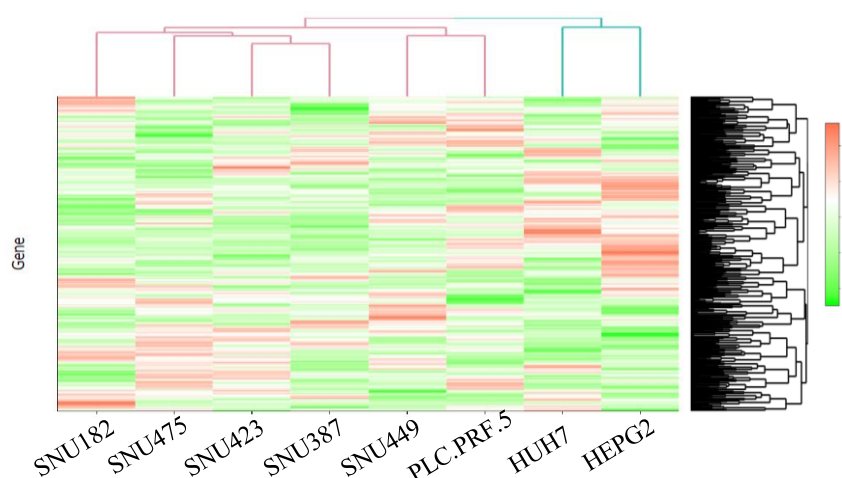


Figure 16: Heat map drawn with Z scores of HBV-integration effect genes calculated for the expression levels of 8 different cell lines taken from CCLE results.

## 3.3 GSEA of Isogenic HepG2 and HepG2-2.2.15 Cells

With the *within cell line* comparison results, we have identified the DEG lists for the response of the two cell lines to Se-deficiency separately; so we got the lists of DEGs whose expression levels were altered in Se- condition compared to that in Se+ control for each cell line. In order to combine these separate results and determine the differential response of the two isogenic cell lines to Se-deficiency with respect to each other we have calculated DEG scores (Cavga et al., 2019) as explained in methods (Figure 17a). DEG scores were used to determine the enriched pathways by GSEA; pathways enriched in lists with positive DEG scores (DEG scores > 1) indicate that the relevant genes were relatively up-regulated in HepG2-2.2.15 when normalized to HepG2 cells within Se- vs Se+ comparisons. Accordingly, the activities of these pathways were higher in HepG2-2.2.15 cells in response to Se-deficiency. In contrast, pathways enriched in lists with negative DEG scores (DEG scores < 1) indicate a lower activity in HepG2-2.2.15 cells in response to Se-deficiency compared to HepG2.

GSEA results have revealed that the relatively up-regulated genes in HepG2-2.2.15 cells were found to be more related to DNA-Repair, G2M checkpoint, Oxidation Reduction and MTORC1 signaling pathways, which might be key pathways for the acquired tolerance gained by HepG2-2.2.15 cells to Se-deficiency dependent oxidative stress as they all seem to have linked to survival mechanisms. The enriched pathways in the DEG lists with the DEG scores higher than 0, were found to be also related to stress response, repair mechanisms and cell cycle supporting the previous finding (Table 4). On the contrary, the relatively down-regulated genes in HepG2-2.2.15 when normalized to those in HepG2 were found to be enriched in IL2-STAT5 Signaling pathway, which is known to have functions related to apoptosis (Longmore et al., 1998) and might indicate lower apoptotic phenotype in HepG2-2.2.15 comparison relative to HepG2 in comparison. The same DEG lists were further used to generate a network of predicted associations between proteins of interest by STRING, and enriched pathways found on this network was consistent with GSEA results (Figure 17b).

**A**

$$\log_2\left(\frac{2.2.15 \; \frac{Se-}{Se+}}{HepG2 \; \frac{Se-}{Se+}}\right)$$

> 1 ↙        ↘ < -1

| NAME | SIZE | ES | NOM p-val | FDR q-val |
|---|---|---|---|---|
| E2F_TARGETS | 49 | 0.59 | 0.00 | 0.00 |
| G2M_CHECKPOINT | 34 | 0.58 | 0.00 | 0.00 |
| MTORC1_SIGNALING | 16 | 0.54 | 0.01 | 0.01 |
| ESTROGEN RESPONSE LATE | 18 | 0.43 | 0.06 | 0.08 |
| DNA_REPAIR | 16 | 0.40 | 0.12 | 0.15 |

| NAME | SIZE | ES | NOM p-val | FDR q-val |
|---|---|---|---|---|
| IL2 STAT5 SIGNALING | 17 | -0.29 | 0.19 | 0.20 |

**B**



| pathway | description | count in gene set | false discovery rate | |
|---|---|---|---|---|
| hsa04150 | mTOR signaling pathway | 8 of 148 | 0.00011 | ● |
| hsa03030 | DNA replication | 5 of 36 | 0.00011 | ● |
| hsa03420 | Nucleotide excision repair | 5 of 46 | 0.00020 | ● |
| hsa04110 | Cell cycle | 6 of 123 | 0.00098 | ● |
| hsa03410 | Base excision repair | 4 of 33 | 0.00098 | ● |
| KW-0227 | DNA damage | 9 of 379 | 0.0075 | ● |

Figure 17: Enrichment results of within cell line comparison DEG list (a) DEG scores were given for each gene for Day 3 according to the indicated formula and the GSEA was performed to find the enriched Hallmark pathways with their enrichment scores (b) the STRING was used to perform pathway analysis and to find the enriched pathways. ES: enrichment score, NOM p-val: nominal p value, FDR: false discovery rate.

Table 4: DEG scores were given for each gene in Day 3 and the GSEA was performed to find the enriched GO BP_pathways with their enrichment scores.

| NAME | SIZE | ES | NOM p-val | FDR q-val |
|---|---|---|---|---|
| CELL_CYCLE | 156 | 0.49 | 0.00 | 0.00 |
| DNA_DEPENDENT_DNA_REPLICATION | 40 | 0.64 | 0.00 | 0.00 |
| CELL_CYCLE_PHASE_TRANSITION | 63 | 0.57 | 0.00 | 0.00 |
| DNA_REPLICATION | 51 | 0.59 | 0.00 | 0.00 |
| CELL_CYCLE_PROCESS | 125 | 0.50 | 0.00 | 0.00 |
| CELL_CYCLE_G1_S_PHASE_TRANSITION | 32 | 0.65 | 0.00 | 0.00 |
| MITOTIC_CELL_CYCLE | 96 | 0.50 | 0.00 | 0.00 |
| DNA_METABOLIC_PROCESS | 90 | 0.51 | 0.00 | 0.00 |
| CELLULAR_RESPONSE_TO_DNA_DAMAGE_STIMULUS | 72 | 0.50 | 0.00 | 0.00 |
| DNA_REPLICATION_INITIATION | 17 | 0.74 | 0.00 | 0.00 |
| REGULATION_OF_CELL_CYCLE_PHASE_TRANSITION | 41 | 0.57 | 0.00 | 0.00 |
| CELL_CYCLE_DNA_REPLICATION | 23 | 0.63 | 0.00 | 0.00 |
| REGULATION_OF_CELL_CYCLE | 89 | 0.45 | 0.00 | 0.00 |
| REGULATION_OF_MITOTIC_CELL_CYCLE | 52 | 0.49 | 0.00 | 0.00 |
| DNA_REPAIR | 56 | 0.47 | 0.00 | 0.00 |
| POSITIVE_REGULATION_OF_MITOTIC_CELL_CYCLE | 18 | 0.62 | 0.00 | 0.00 |
| TELOMERE_ORGANIZATION | 22 | 0.59 | 0.00 | 0.00 |
| NUCLEAR_DNA_REPLICATION | 21 | 0.60 | 0.00 | 0.00 |
| REGULATION_OF_CELL_CYCLE_PROCESS | 68 | 0.45 | 0.00 | 0.00 |
| POSITIVE_REGULATION_OF_CELL_CYCLE_PROCESS | 29 | 0.54 | 0.00 | 0.00 |
| POSITIVE_REGULATION_OF_CELL_CYCLE_PHASE_TRANSITION | 15 | 0.66 | 0.00 | 0.00 |
| POSITIVE_REGULATION_OF_CELL_CYCLE | 35 | 0.50 | 0.00 | 0.00 |
| CHROMOSOME_ORGANIZATION | 106 | 0.40 | 0.00 | 0.00 |
| REGULATION_OF_DNA_REPLICATION | 17 | 0.60 | 0.00 | 0.00 |
| CELL_DIVISION | 57 | 0.43 | 0.00 | 0.00 |
| NEGATIVE_REGULATION_OF_CELL_CYCLE | 47 | 0.45 | 0.00 | 0.00 |
| ORGANELLE_LOCALIZATION | 28 | 0.51 | 0.00 | 0.01 |
| RESPONSE_TO_RADIATION | 32 | 0.50 | 0.00 | 0.01 |
| SIGNAL_TRANSDUCTION_BY_P53_CLASS_MEDIATOR | 22 | 0.55 | 0.00 | 0.01 |
| CELL_CYCLE_G2_M_PHASE_TRANSITION | 25 | 0.52 | 0.00 | 0.01 |
| NEGATIVE_REGULATION_OF_CELL_CYCLE_PHASE_TRANSITION | 22 | 0.54 | 0.00 | 0.01 |
| ORGANELLE_FISSION | 50 | 0.43 | 0.00 | 0.01 |
| DNA_GEOMETRIC_CHANGE | 17 | 0.57 | 0.00 | 0.01 |
| NEGATIVE_REGULATION_OF_CELL_CYCLE_PROCESS | 30 | 0.48 | 0.00 | 0.01 |
| ESTABLISHMENT_OF_ORGANELLE_LOCALIZATION | 23 | 0.51 | 0.00 | 0.01 |
| DNA_INTEGRITY_CHECKPOINT | 18 | 0.56 | 0.00 | 0.01 |
| CHROMOSOME_SEGREGATION | 35 | 0.45 | 0.00 | 0.02 |
| REGULATION_OF_CELL_CYCLE_G2_M_PHASE_TRANSITION | 21 | 0.52 | 0.00 | 0.02 |
| DNA_BIOSYNTHETIC_PROCESS | 19 | 0.53 | 0.00 | 0.02 |
| ANATOMICAL_STRUCTURE_HOMEOSTASIS | 35 | 0.45 | 0.00 | 0.02 |
| ORGANIC_ACID_METABOLIC_PROCESS | 42 | 0.43 | 0.00 | 0.02 |
| DOUBLE_STRAND_BREAK_REPAIR | 26 | 0.47 | 0.00 | 0.03 |
| RECOMBINATIONAL_REPAIR | 17 | 0.54 | 0.00 | 0.03 |
| CELLULAR_RESPONSE_TO_ENDOGENOUS_STIMULUS | 81 | 0.36 | 0.00 | 0.03 |
| OXIDATION_REDUCTION_PROCESS | 57 | 0.39 | 0.00 | 0.03 |

36

## 3.4 Pathway Analysis of Selenium-Deficiency Dependent Differentially Expressed Genes in Isogenic HCC Cell Lines

PCST algorithm was used to construct trees for Se-deficiency effect genes by taking STRING as the reference dataset (Figure 18-20). PCST constructs optimum trees from the given DEGs (terminal nodes) by using human interactome data as a reference to find the shortest path between the nodes. The Steiner nodes were determined by the algorithm (shown by diamond in the figures) and the nodes that have high betweenness centrality values were identified in order to investigate in more detail (biggest nodes in each tree) since they might play some key roles considering their connecting positions between different branches of the tree. Most of the genes with high betweenness centrality (both DEGs and Steiner nodes) were found to have roles related to oxidative stress response either with impacts in oxidative stress dependent apoptosis or antioxidant pathways (Figure 18-20).

Figure 18 shows the tree generated with the D1 data. FOXA1 an oxidoreductase, which has a proapoptotic role and CYP7A1 another oxidoreductase whose over expression was shown to decrease oxidative stress in mice (H. Liu et al., 2016) were the DEGs with high betweenness centrality values. ONECUT1 known to have roles in cell cycle regulation and PITX2 involved in oxidative stress response were detected as Steiner nodes, which are not differentially expressed but might be critical in the differential response due to their connecting positions between different branches of the tree. (Iizuka et al., 2003; Strungaru et al., 2011).
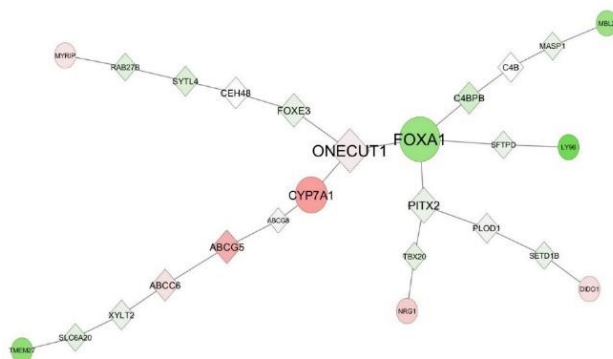


Figure 18: Network constructed by PCST algorithm for acquired HepG2 vs HepG2-2.2.15 D1. STRING database was used as the reference database. Circles indicate the terminal nodes and diamonds indicate the Steiner nodes. Node colors indicate expression level difference of each gene between cell lines, green and red indicating negative and positive fold changes respectively.

In D2 results, the DEGs TXNRD1 and ALDH1L2 and the Steiner nodes ACLY, TXNIP, SCD5, MTR and TXNDC17 were the genes with the highest betweenness centrality values (Figure 19). They were all shown to be exhibiting oxidative stress and redox homeostasis related functions; the decreased expression of them other than

TXNIP was found to be associated to increased ROS levels resulting in oxidative stress.
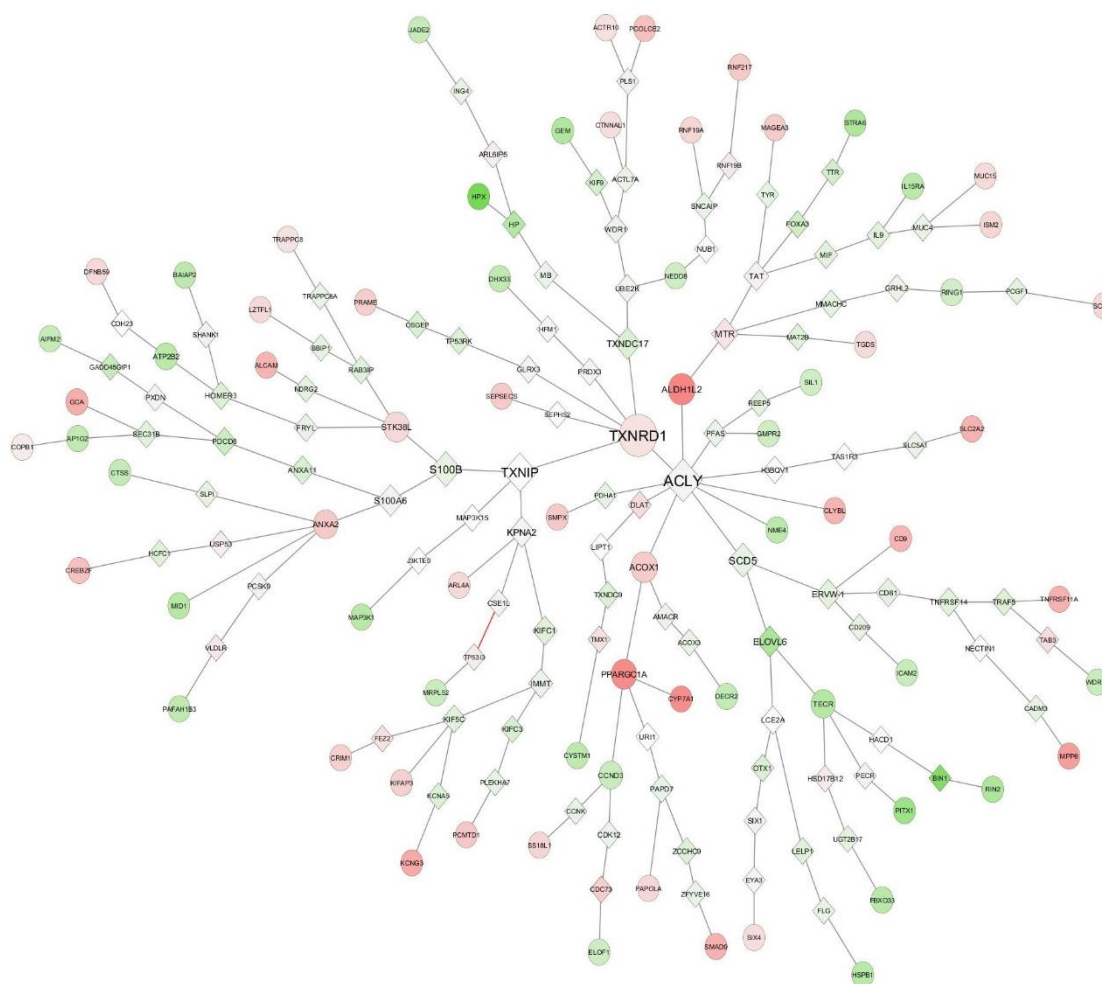


Figure 19: Network constructed by PCST algorithm for acquired HepG2 vs HepG2-2.2.15 D2. STRING database was used as the reference database. Circles indicate the terminal nodes and diamonds indicate the Steiner nodes. Node colors indicate expression level difference of each gene between cell lines, green and red indicating negative and positive fold changes respectively.

In D3 results, the Steiner nodes with high betweenness centrality values included LSM4 that is associated stress response and CNBP, DMPK, QDPR with antioxidant functions (Figure 20). This has supported the idea that these DEGS could be the key elements in the acquired tolerance gained by HepG2-2.2.15 cells to Se-deficiency dependent oxidative stress.
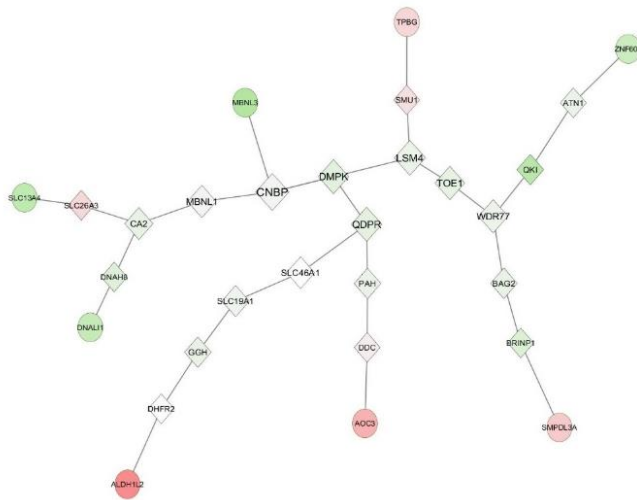
Figure 20: Network constructed by PCST algorithm for acquired HepG2 vs HepG2-2.2.15 Day 3. STRING database was used as the reference database. Circles indicate the terminal nodes and diamonds indicate the Steiner nodes. Node colors indicate expression level difference of each gene between cell lines, green and red indicating negative and positive fold changes respectively.

## 3.5 Definition and Clinical Significance of Selected Genes Related with Oxidative Stress Resistance

In this thesis study, various bioinformatics approaches were used to determine potentially important biomarkers that have functions in oxidative stress resistance. Overall, all the genes that were defined to be important in our study by different analysis methods were summarized in Table 5 and the extended list is included in the Appendix. 27 Genes were selected based on *within* or *between cell line* comparisons that were associated with either Se or HBV effect. While most of the genes were previously identified in HCC (17 genes) and oxidative stress (20 genes), HOXD1 and CLYBL are proposed to be critical for the first time by this study.

Fifteen DEGs with high betweenness centrality values were identified by the PCST analysis for the *between cell line* comparison results, and eleven of these genes were Steiner nodes which were not differentially expressed but were determined to be on key positions on pathways that might have effects on the expression of the determined DEGs having indirect effects on differential response. With the GSEA, six DEGs (DUT, POLD3, E2F2, GINS2, PIK3R3, TMEM97) as a result of within cell line comparisons were determined to be key elements for the Se effect according to their highest enrichment scores. Three genes from each HBV and Se effect DEGs were selected from the heatmap analysis results with the smallest p-values that have the most significant impact on the differential response to Se-deficiency between the isogenic HepG2 and HepG2-2.2.15 cell lines.

39

Table 5: Genes identified by *within* or *between cell line* comparisons related with either Se or cell line (HBV) effect. The associations of each gene with oxidative stress and/or HCC in previous studies were indicated. Se: Selenium-deficiency effect, HBV: HBV-integration effect, BCL: Between cell line, WCL: Within cell line, (E): Existing DEG, (S): Steiner node, HM: Heatmap, OS: Oxidative stress, r: Reported.

| Gene | Effect | Comparison | Analysis | OS | HCC | Literature |
|---|---|---|---|---|---|---|
| FOXA1 | Se | BCL | PCST (E) | r | r | (L. Song et al., 2009; H. Zhang et al., n.d.) |
| CYP7A1 | Se | BCL | PCST (E) | r | | Liu et al. (2016) |
| ONECUT1 | Se | BCL | PCST (S) | | r | (Iizuka et al., 2003) |
| PITX2 | Se | BCL | PCST (S) | r | r | (Archer et al., 2010; Strungaru et al., 2011) |
| TXNRD1 | Se | BCL | PCST (E) | r | r | (Kiermayer et al., 2007; Lee et al., 2019) |
| ALDH1L2 | Se | BCL | PCST (E) | r | r | (Lee et al., 2017; Sarret et al., 2019) |
| ACLY | Se | BCL | PCST (S) | r | r | (Migita et al., 2013; Pope et al., 2019) |
| TXNIP | Se | BCL | PCST (S) | r | | (Zhou & Chng, 2013) |
| SCD5 | Se | BCL | PCST (S) | r | r | (G. I. Yu et al., 2018) |
| MTR | Se | BCL | PCST (S) | r | | (Si et al., 2016) |
| TXNDC17 | Se | BCL | PCST (S) | r | | (Liyanage et al., 2019) |
| LSM4 | Se | BCL | PCST (S) | r | | (L. Chen & Liu, 2017) |
| CNBP | Se | BCL | PCST (S) | r | | (de Peralta et al., 2016) |
| DMPK | Se | BCL | PCST (S) | r | | (Pantic et al., 2013) |

Table 5: Genes identified by *within* or *between cell line* comparisons related with either Se or cell line (HBV) effect. The associations of each gene with oxidative stress and/or HCC in previous studies were indicated. Se: Selenium-deficiency effect, HBV: HBV-integration effect, BCL: Between cell line, WCL: Within cell line, (E): Existing DEG, (S): Steiner node, HM: Heatmap, OS: Oxidative stress, r: Reported. ( continued )

| QDPR | Se | BCL | PCST (S) | r | r | (Gu et al., 2017; Nwosu et al., 2017) |
|---|---|---|---|---|---|---|
| DUT | Se | WCL | GSEA | | r | (Takatori et al., 2010) |
| POLD3 | Se | WCL | GSEA | r | r | (Jiang et al., 2019; Tan et al., 2020) |
| E2F2 | Se | WCL | GSEA | r | r | (Castillo et al., 2015; Y.-L. Huang et al., 2019) |
| GINS2 | Se | WCL | GSEA | r | r | (Lian et al., 2018; C. Liu et al., 2019) |
| PIK3R3 | Se | WCL | GSEA | r | r | (Engedal et al., 2018; Ibrahim et al., 2018) |
| TMEM97 | Se | WCL | GSEA | r | | (J.-H. Wang et al., 2020) |
| FGF13 | HBV | BCL | HM | r | r | (Bublik et al., 2017; Grose et al., 2014) |
| GPC3 | HBV | BCL | HM | | r | (Akutsu, 2010; Guo et al., 2020) |
| MAP7D2 | HBV | BCL | HM | | r | (Nishida et al., 2014) |
| PPAP2A | Se | BCL | HM | | r | (Jenkins et al., 2012; Nwosu et al., 2017) |
| HOXD1 | Se | BCL | HM | | | |
| CLYBL | Se | BCL | HM | | | |

Finally, to find out the clinical relevance of the selected genes; Kaplan Meier Plots were drawn for the liver cancer RNA-seq results (Figure 21). The expression levels of 16 out of 27 genes were found to be related to survival time of the HCC patients.
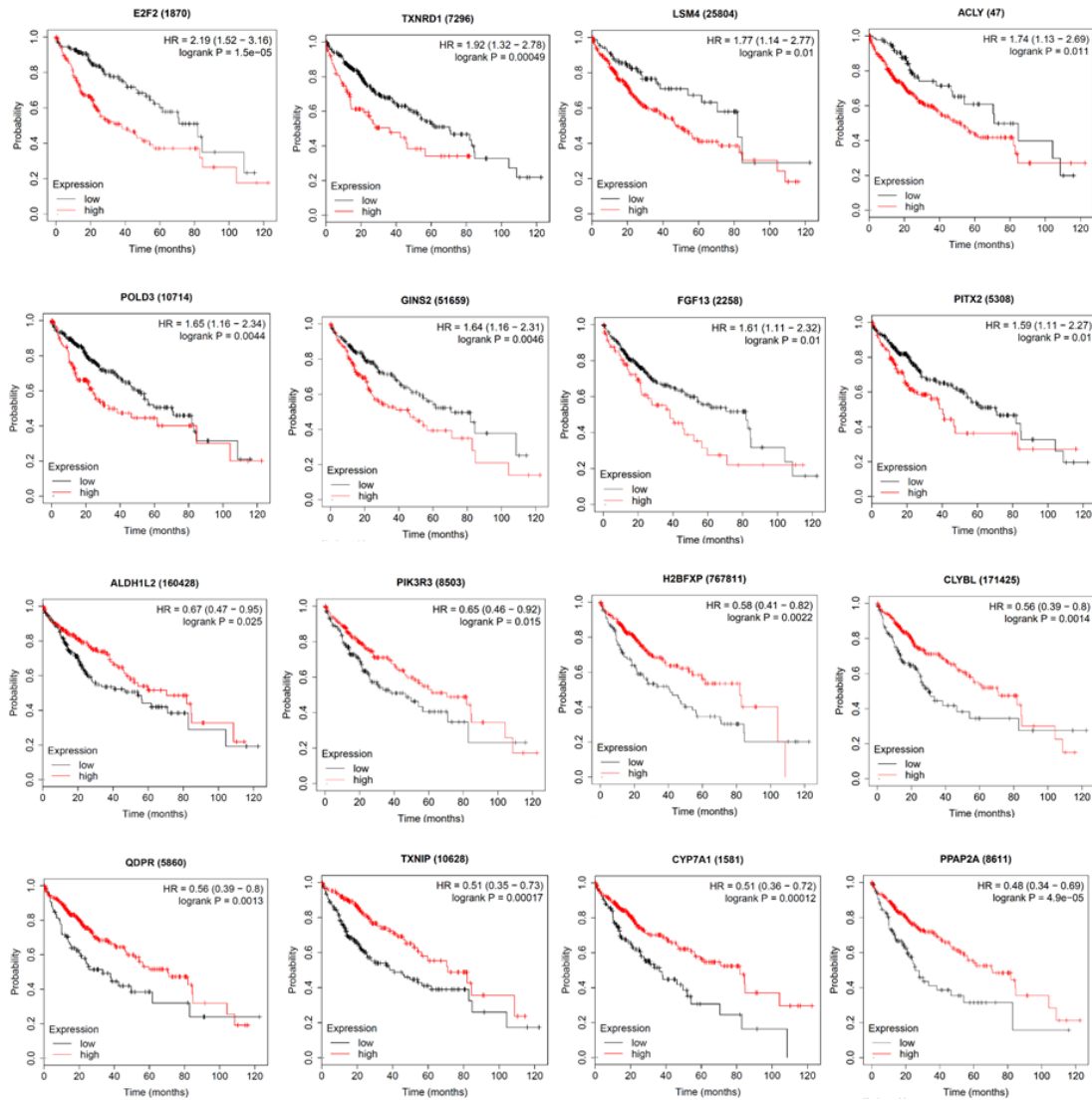
Figure 21:Kaplan Meier plots were generated for selected DEGs in HCC patients in liver cancer RNA-seq dataset.

The expression levels of these genes in our 12 samples were used to generate expression graphs (Figure 22). The pattern of differential expression is apparent when these graphs are examined in parallel to the methods used. For instance, for the expression of genes identified by *between cell line* comparisons, the expression level difference between six HepG2 samples and six HepG2-2.2.15 samples could clearly be seen. Similarly, the expression levels of genes identified by *within cell line* comparison have a different expression pattern and do not change depending on the cell line but Se status. For the expression of Steiner node genes identified by PCTS algorithm, the patterns seem to be more irregular, since their expression is not differentially expressed in compared samples.
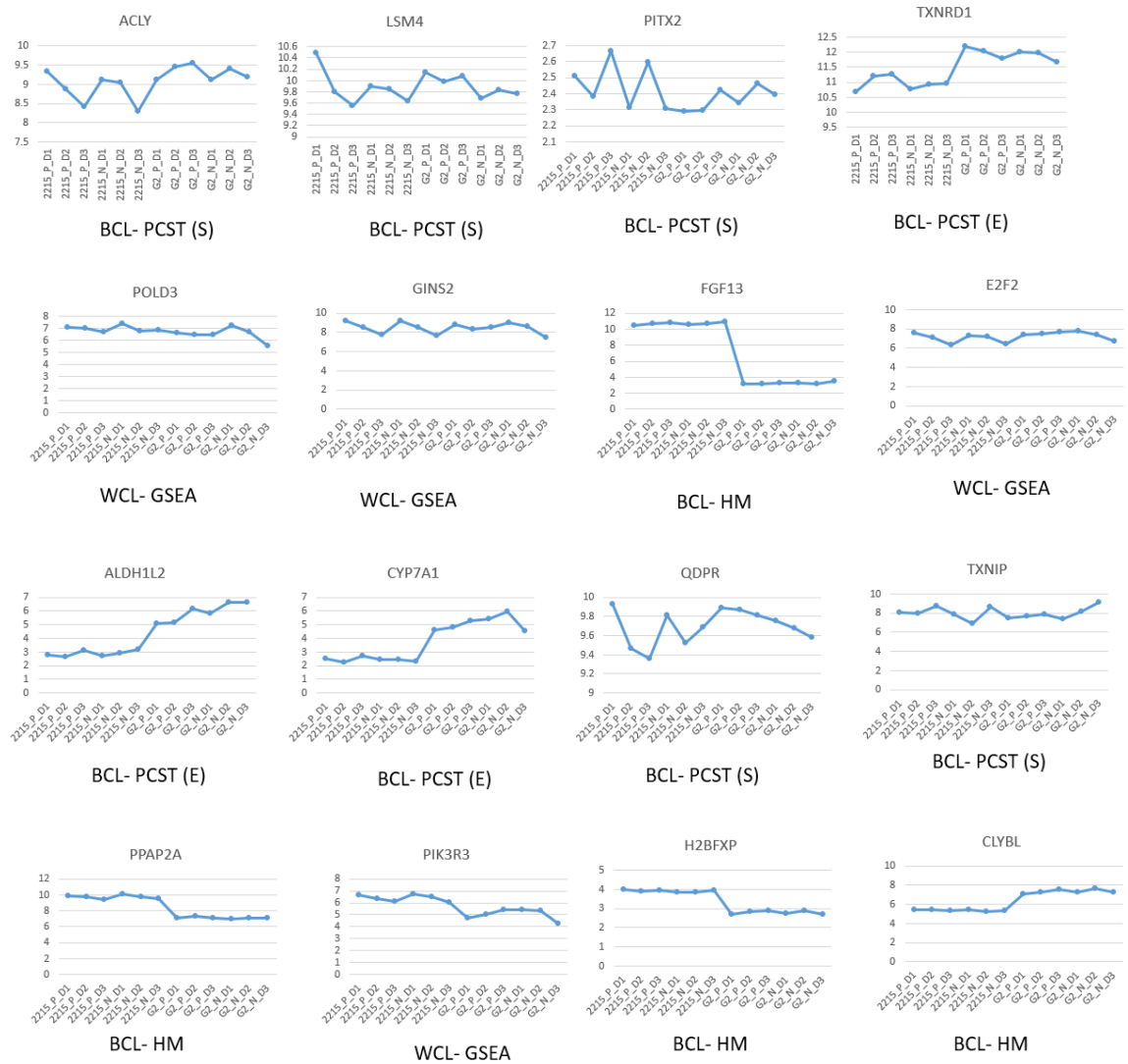
Figure 22: Log2 expression values of the identified genes in our 12 different samples.

## 3.6 Correlation Analysis of the Identified DEGs with p53-MDM2 Pathway Gene Expression Values

In previous studies, FOXA1 gene amplification and alterations in expression levels in various cancer types including breast, prostate, lung and liver cancers were studied and it was also proposed as potential target for cancer therapy in parallel to our findings (Bernardo et al., 2013; Rahman et al., 2020; Y. Wang et al., 2019). As already explained in the Introduction chapter, FOXA1 is also binds to DNA to regulate the transcription of various genes including P53 and MDM2 (Swetzig et al., 2016). By taking this information into consideration, we tried to understand whether there is any association between p53-MDM2 pathway genes and our DEGs of interest.

We first used STRING database to test whether our DEGs of interest have any Protein-Protein interactions with P53 and MDM2 proteins. Figure 23 shows the generated network and the interactions between all proteins. TP53 protein was found to be interacting with 7 of the genes that were identified in the first part of our study, FOXA1, CNBP, TXNIP, FGF13, TXNRD1, E2F2 and DUT.
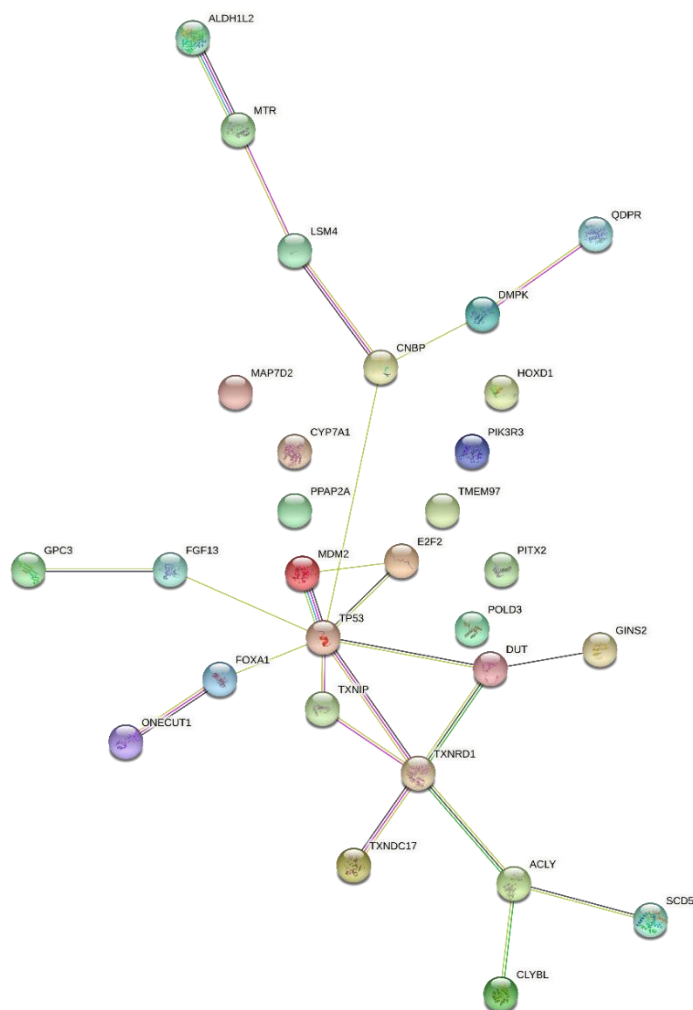


Figure 23: Network generated by STRING showing the interactions between the genes we identified in the first part of the study and p53-MDM2 genes.

We than extended the p53-MDM2 pathway list to include other genes that have regulatory roles in that pathway (Nag et al., 2013) and tested the presence of any potential correlations between the expression levels of these genes and our DEGs by analysing the results of our microarray experiment. Many significant correlations were detected between the two lists, including FOXA1 and p53-MDM2 expression levels which is in parallel with the previous findings. Table 6 and Table 7 show the Rho and p values estimated by Spearman correlation analysis, respectively. The closer the Rho values to +/- 1, the more correlated the gene expression levels between the relevant

genes. The significance of the correlation was further shown by the estimates p-values (Table 7). TP53 and MDM2 gene expressions were found to be correlated to the expression of 12 and 8 genes from our DEG lists, respectively; while FOXA1 from our list was found to have correlated expression with 10 of the genes that functions in p53-MDM2 pathway. In general, 21 of our 27 DEGs were found to have correlated expression to at least 5 of the genes in p53-MDM2 pathway.

Depending on the aforementioned association, which is also supported by the previous studies, targeting p53-MDM2 mechanism might be a potential strategy for the treatment of HCC patients. With this motivation we have tested the cytotoxicity of 9 different p53-MDM2 interaction inhibitor compounds on 4 different HCC cell lines and the mechanisms behind their cytotoxic effects. The remaining sections will focus on the results of this study.

Table 6: Table showing the Rho values estimated by Spearman correlation analysis for the expression of p53-MDM2 pathway genes (columns) and our DEGs (rows). Blue indicates positive correlation while red indicates negative correlation and the darker the color, the more correlated expression of the two genes.

| | TP53 | MDM2 | ARF1 | ARF3 | ARF4 | ARF5 | ARF6 | E2F1 | HIPK2 | IGF1R | JMY | NPM1 | NPM2 | NPM3 | NUMB | RNF2 | RPL26L1 | SIVA1 | USP7 | YY1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GPC3 | 0.83 | -0.57 | -0.27 | 0.87 | -0.20 | 0.60 | -0.73 | 0.00 | 0.55 | 0.01 | 0.30 | 0.27 | -0.73 | 0.70 | -0.72 | 0.25 | 0.61 | 0.40 | -0.62 | -0.19 |
| TMEM97 | 0.80 | -0.38 | -0.36 | 0.80 | -0.33 | 0.64 | -0.62 | 0.14 | 0.48 | 0.05 | 0.22 | 0.17 | -0.56 | 0.66 | -0.66 | 0.17 | 0.45 | 0.38 | -0.54 | -0.36 |
| HOXD1 | 0.78 | -0.44 | -0.30 | 0.80 | -0.17 | 0.53 | -0.70 | -0.08 | 0.48 | 0.06 | 0.38 | 0.36 | -0.78 | 0.66 | -0.76 | 0.17 | 0.59 | 0.34 | -0.65 | -0.14 |
| PIK3R3 | 0.74 | -0.49 | -0.31 | 0.81 | -0.26 | 0.57 | -0.72 | -0.04 | 0.45 | -0.01 | 0.29 | 0.31 | -0.74 | 0.68 | -0.71 | 0.31 | 0.64 | 0.36 | -0.59 | -0.15 |
| TXNDC17 | 0.72 | -0.45 | -0.38 | 0.77 | -0.22 | 0.55 | -0.73 | -0.10 | 0.40 | 0.02 | 0.32 | 0.38 | -0.76 | 0.68 | -0.71 | 0.29 | 0.62 | 0.33 | -0.57 | -0.03 |
| DMPK | 0.71 | -0.50 | 0.21 | 0.68 | -0.08 | -0.07 | 0.01 | -0.31 | 0.23 | 0.08 | 0.66 | 0.35 | -0.62 | 0.71 | -0.59 | -0.56 | 0.48 | -0.27 | -0.73 | -0.23 |
| PPAP2A | 0.64 | -0.51 | -0.11 | 0.68 | -0.14 | 0.48 | -0.38 | -0.27 | 0.41 | 0.16 | 0.52 | 0.44 | -0.75 | 0.86 | -0.59 | 0.05 | 0.43 | 0.11 | -0.73 | -0.21 |
| FOXA1 | 0.62 | -0.66 | -0.13 | 0.84 | -0.15 | 0.68 | -0.79 | 0.02 | 0.69 | 0.26 | 0.28 | 0.15 | -0.71 | 0.65 | -0.74 | 0.34 | 0.43 | 0.46 | -0.66 | -0.20 |
| POLD3 | 0.59 | -0.37 | -0.51 | 0.57 | -0.03 | 0.47 | -0.82 | -0.08 | 0.30 | -0.08 | 0.07 | 0.29 | -0.61 | 0.41 | -0.47 | 0.41 | 0.58 | 0.45 | -0.33 | 0.24 |
| FGF13 | 0.52 | -0.67 | 0.27 | 0.72 | -0.10 | -0.05 | -0.27 | -0.54 | 0.34 | 0.31 | 0.75 | 0.39 | -0.80 | 0.72 | -0.78 | -0.20 | 0.64 | -0.30 | -0.85 | -0.06 |
| PITX2 | 0.37 | 0.02 | 0.03 | 0.32 | -0.34 | -0.09 | -0.01 | -0.04 | -0.10 | -0.24 | 0.34 | 0.27 | -0.36 | 0.22 | -0.38 | -0.27 | 0.34 | -0.12 | -0.24 | -0.07 |
| GINS2 | 0.24 | 0.07 | -0.62 | 0.14 | 0.02 | 0.72 | -0.65 | 0.54 | 0.24 | -0.36 | -0.57 | -0.14 | -0.02 | -0.11 | 0.08 | 0.64 | 0.06 | 0.90 | 0.22 | -0.05 |
| DUT | 0.19 | 0.13 | -0.62 | 0.01 | 0.08 | 0.66 | -0.55 | 0.52 | 0.16 | -0.48 | -0.66 | -0.17 | 0.10 | -0.19 | 0.27 | 0.67 | 0.01 | 0.87 | 0.34 | -0.03 |
| LSM4 | 0.10 | 0.01 | -0.12 | 0.09 | -0.13 | 0.76 | -0.17 | 0.80 | 0.44 | -0.08 | -0.57 | -0.53 | 0.37 | -0.07 | 0.24 | 0.33 | -0.49 | 0.74 | 0.20 | -0.64 |
| TXNIP | 0.00 | 0.12 | -0.09 | -0.09 | 0.38 | -0.62 | 0.48 | -0.59 | -0.39 | 0.27 | 0.52 | 0.50 | -0.37 | 0.23 | -0.16 | -0.89 | 0.18 | -0.61 | -0.41 | 0.28 |
| E2F2 | -0.03 | 0.43 | -0.52 | -0.22 | -0.27 | 0.39 | -0.21 | 0.81 | -0.12 | -0.41 | -0.92 | -0.58 | 0.62 | -0.50 | 0.52 | 0.39 | -0.34 | 0.66 | 0.69 | -0.17 |
| QDPR | -0.06 | 0.00 | -0.27 | -0.05 | 0.20 | 0.51 | -0.04 | 0.67 | 0.23 | -0.04 | -0.71 | -0.53 | 0.39 | -0.14 | 0.42 | 0.19 | -0.37 | 0.65 | 0.22 | -0.46 |
| SCD5 | -0.17 | -0.36 | 0.54 | 0.08 | 0.13 | -0.41 | 0.34 | -0.58 | 0.10 | 0.65 | 0.71 | 0.26 | -0.19 | 0.36 | -0.38 | -0.39 | 0.01 | -0.75 | -0.43 | 0.10 |
| ACLY | -0.22 | 0.71 | -0.42 | -0.52 | -0.20 | 0.10 | 0.29 | 0.59 | -0.44 | -0.45 | -0.75 | -0.38 | 0.65 | -0.50 | 0.71 | 0.03 | -0.47 | 0.32 | 0.66 | -0.22 |
| MTR | -0.45 | 0.62 | -0.29 | -0.69 | 0.10 | 0.05 | 0.08 | 0.51 | -0.31 | -0.46 | -0.84 | -0.43 | 0.75 | -0.81 | 0.81 | 0.36 | -0.50 | 0.42 | 0.88 | 0.15 |
| ALDH1L2 | -0.53 | 0.67 | 0.04 | -0.80 | 0.24 | -0.65 | 0.73 | 0.01 | -0.69 | -0.40 | -0.35 | -0.13 | 0.63 | -0.67 | 0.76 | -0.34 | -0.31 | -0.34 | 0.64 | 0.17 |
| CLYBL | -0.53 | 0.91 | -0.38 | -0.83 | -0.02 | -0.27 | 0.37 | 0.31 | -0.64 | -0.26 | -0.64 | -0.27 | 0.76 | -0.78 | 0.75 | -0.04 | -0.56 | 0.01 | 0.84 | 0.23 |
| CYP7A1 | -0.57 | 0.71 | -0.22 | -0.72 | 0.07 | -0.30 | 0.42 | 0.40 | -0.60 | -0.45 | -0.66 | -0.31 | 0.71 | -0.76 | 0.71 | 0.03 | -0.35 | 0.05 | 0.82 | 0.14 |
| MAP7D2 | -0.58 | 0.82 | -0.28 | -0.80 | 0.00 | -0.29 | 0.13 | 0.28 | -0.54 | -0.35 | -0.69 | -0.33 | 0.71 | -0.95 | 0.69 | 0.23 | -0.41 | 0.12 | 0.88 | 0.34 |
| CNBP | -0.63 | 0.51 | 0.13 | -0.83 | 0.38 | -0.64 | 0.80 | -0.12 | -0.66 | -0.32 | -0.25 | -0.03 | 0.58 | -0.54 | 0.79 | -0.29 | -0.34 | -0.41 | 0.59 | 0.27 |
| ONECUT1 | -0.72 | 0.51 | 0.01 | -0.67 | 0.13 | -0.17 | 0.30 | 0.50 | -0.31 | -0.15 | -0.77 | -0.59 | 0.73 | -0.83 | 0.71 | 0.09 | -0.57 | 0.24 | 0.70 | -0.01 |
| TXNRD1 | -0.84 | 0.50 | 0.13 | -0.76 | 0.10 | -0.37 | 0.41 | 0.27 | -0.38 | 0.04 | -0.52 | -0.46 | 0.83 | -0.81 | 0.62 | 0.08 | -0.61 | -0.14 | 0.78 | 0.22 |

Table 7: The p-values estimated for Spearman correlation analysis for the expression of p53-MDM2 pathway genes (columns) and our DEGs (rows). The values highlighted by red indicates the numbers lower than 0.05; so accepted as significant.

| | TP53 | MDM2 | ARF1 | ARF3 | ARF4 | ARF5 | ARF6 | E2F1 | HIPK2 | IGF1R | JMY | NPM1 | NPM2 | NPM3 | NUMB | RNF2 | RPL26L1 | SIVA1 | USP7 | YY1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TXNRD1 | 0.00 | 0.10 | 0.70 | 0.01 | 0.77 | 0.24 | 0.19 | 0.39 | 0.23 | 0.90 | 0.08 | 0.13 | 0.00 | 0.00 | 0.04 | 0.80 | 0.04 | 0.67 | 0.00 | 0.48 |
| GPC3 | 0.00 | 0.06 | 0.39 | 0.00 | 0.54 | 0.04 | 0.01 | 1.00 | 0.07 | 0.99 | 0.34 | 0.40 | 0.01 | 0.01 | 0.01 | 0.43 | 0.04 | 0.20 | 0.03 | 0.56 |
| TMEM97 | 0.00 | 0.23 | 0.26 | 0.00 | 0.30 | 0.03 | 0.04 | 0.67 | 0.12 | 0.89 | 0.50 | 0.59 | 0.06 | 0.02 | 0.02 | 0.59 | 0.14 | 0.23 | 0.07 | 0.25 |
| HOXD1 | 0.00 | 0.15 | 0.34 | 0.00 | 0.60 | 0.08 | 0.01 | 0.80 | 0.12 | 0.87 | 0.23 | 0.25 | 0.00 | 0.02 | 0.01 | 0.59 | 0.05 | 0.29 | 0.03 | 0.67 |
| PIK3R3 | 0.01 | 0.11 | 0.32 | 0.00 | 0.42 | 0.06 | 0.01 | 0.90 | 0.15 | 0.97 | 0.35 | 0.33 | 0.01 | 0.02 | 0.01 | 0.33 | 0.03 | 0.25 | 0.05 | 0.64 |
| ONECUT1 | 0.01 | 0.09 | 0.99 | 0.02 | 0.68 | 0.59 | 0.34 | 0.10 | 0.32 | 0.64 | 0.01 | 0.05 | 0.01 | 0.00 | 0.01 | 0.78 | 0.06 | 0.46 | 0.01 | 0.99 |
| TXNDC17 | 0.01 | 0.14 | 0.23 | 0.01 | 0.50 | 0.07 | 0.01 | 0.77 | 0.20 | 0.96 | 0.31 | 0.23 | 0.01 | 0.02 | 0.01 | 0.37 | 0.03 | 0.30 | 0.06 | 0.94 |
| DMPK | 0.01 | 0.10 | 0.51 | 0.02 | 0.82 | 0.83 | 0.97 | 0.33 | 0.47 | 0.82 | 0.02 | 0.27 | 0.03 | 0.01 | 0.05 | 0.06 | 0.12 | 0.39 | 0.01 | 0.47 |
| PPAP2A | 0.03 | 0.09 | 0.73 | 0.02 | 0.67 | 0.12 | 0.23 | 0.40 | 0.19 | 0.62 | 0.09 | 0.15 | 0.01 | 0.00 | 0.05 | 0.89 | 0.17 | 0.73 | 0.01 | 0.51 |
| CNBP | 0.03 | 0.09 | 0.68 | 0.00 | 0.23 | 0.03 | 0.00 | 0.72 | 0.02 | 0.31 | 0.43 | 0.92 | 0.05 | 0.07 | 0.00 | 0.37 | 0.29 | 0.18 | 0.05 | 0.39 |
| FOXA1 | 0.04 | 0.02 | 0.68 | 0.00 | 0.64 | 0.02 | 0.00 | 0.96 | 0.02 | 0.42 | 0.38 | 0.64 | 0.01 | 0.03 | 0.01 | 0.28 | 0.16 | 0.13 | 0.02 | 0.53 |
| POLD3 | 0.05 | 0.24 | 0.09 | 0.06 | 0.94 | 0.13 | 0.00 | 0.82 | 0.34 | 0.80 | 0.83 | 0.37 | 0.04 | 0.19 | 0.13 | 0.18 | 0.05 | 0.15 | 0.30 | 0.46 |
| MAP7D2 | 0.05 | 0.00 | 0.38 | 0.00 | 1.00 | 0.37 | 0.68 | 0.38 | 0.07 | 0.27 | 0.02 | 0.30 | 0.01 | 0.00 | 0.02 | 0.47 | 0.19 | 0.72 | 0.00 | 0.29 |
| CYP7A1 | 0.06 | 0.01 | 0.48 | 0.01 | 0.83 | 0.34 | 0.18 | 0.20 | 0.04 | 0.15 | 0.02 | 0.33 | 0.01 | 0.01 | 0.01 | 0.92 | 0.27 | 0.89 | 0.00 | 0.67 |
| ALDH1L2 | 0.08 | 0.02 | 0.90 | 0.00 | 0.44 | 0.03 | 0.01 | 0.97 | 0.02 | 0.20 | 0.27 | 0.68 | 0.03 | 0.02 | 0.01 | 0.29 | 0.32 | 0.29 | 0.03 | 0.60 |
| CLYBL | 0.08 | 0.00 | 0.22 | 0.00 | 0.96 | 0.40 | 0.24 | 0.32 | 0.03 | 0.42 | 0.03 | 0.39 | 0.01 | 0.00 | 0.01 | 0.90 | 0.06 | 0.97 | 0.00 | 0.47 |
| FGF13 | 0.08 | 0.02 | 0.40 | 0.01 | 0.77 | 0.89 | 0.40 | 0.07 | 0.28 | 0.32 | 0.01 | 0.21 | 0.00 | 0.01 | 0.00 | 0.53 | 0.03 | 0.34 | 0.00 | 0.85 |
| MTR | 0.15 | 0.04 | 0.35 | 0.02 | 0.75 | 0.89 | 0.82 | 0.09 | 0.32 | 0.13 | 0.00 | 0.16 | 0.01 | 0.00 | 0.00 | 0.25 | 0.10 | 0.18 | 0.00 | 0.65 |
| PITX2 | 0.24 | 0.96 | 0.92 | 0.31 | 0.28 | 0.78 | 0.97 | 0.90 | 0.77 | 0.44 | 0.28 | 0.40 | 0.26 | 0.48 | 0.22 | 0.40 | 0.28 | 0.72 | 0.44 | 0.83 |
| GINS2 | 0.46 | 0.83 | 0.04 | 0.67 | 0.96 | 0.01 | 0.03 | 0.07 | 0.44 | 0.25 | 0.06 | 0.67 | 0.96 | 0.73 | 0.80 | 0.03 | 0.87 | 0.00 | 0.48 | 0.89 |
| ACLY | 0.48 | 0.01 | 0.18 | 0.09 | 0.54 | 0.75 | 0.37 | 0.05 | 0.15 | 0.15 | 0.01 | 0.22 | 0.03 | 0.10 | 0.01 | 0.92 | 0.13 | 0.31 | 0.02 | 0.50 |
| DUT | 0.56 | 0.68 | 0.04 | 0.99 | 0.82 | 0.02 | 0.07 | 0.09 | 0.62 | 0.12 | 0.02 | 0.60 | 0.77 | 0.56 | 0.39 | 0.02 | 0.97 | 0.00 | 0.28 | 0.92 |
| SCD5 | 0.59 | 0.26 | 0.07 | 0.80 | 0.68 | 0.19 | 0.28 | 0.05 | 0.77 | 0.03 | 0.01 | 0.42 | 0.56 | 0.26 | 0.23 | 0.21 | 0.97 | 0.01 | 0.17 | 0.77 |
| LSM4 | 0.77 | 0.99 | 0.72 | 0.78 | 0.68 | 0.01 | 0.59 | 0.00 | 0.15 | 0.82 | 0.06 | 0.08 | 0.24 | 0.83 | 0.44 | 0.30 | 0.11 | 0.01 | 0.54 | 0.03 |
| QDPR | 0.85 | 1.00 | 0.39 | 0.89 | 0.53 | 0.09 | 0.90 | 0.02 | 0.47 | 0.90 | 0.01 | 0.08 | 0.21 | 0.67 | 0.18 | 0.56 | 0.24 | 0.03 | 0.48 | 0.13 |
| E2F2 | 0.92 | 0.17 | 0.08 | 0.50 | 0.40 | 0.21 | 0.51 | 0.00 | 0.72 | 0.18 | 0.00 | 0.05 | 0.04 | 0.10 | 0.09 | 0.21 | 0.28 | 0.02 | 0.02 | 0.60 |
| TXNIP | 1.00 | 0.72 | 0.97 | 0.78 | 0.22 | 0.03 | 0.12 | 0.05 | 0.21 | 0.39 | 0.09 | 0.10 | 0.24 | 0.47 | 0.62 | 0.00 | 0.57 | 0.04 | 0.18 | 0.38 |

### 3.7 Testing the Cytotoxic Effects of the P53-MDM2 Interaction Inhibitors

The cytotoxic effects of all compounds were initially tested by Sulforhodamine B (SRB) assay on 4 HCC cell lines carrying various p53 mutations as explained in the Introduction part and given in the Table 8.

Table 8: p53 mutation statuses of HCC cell lines.

| Cell Line | p53 mutation status |
|-----------|---------------------|
| Mahlavu   | R249S               |
| Huh7      | Y220C               |
| HepG2     | p53-wt              |
| Hep3B     | p53-null            |

Percent cell growth inhibition values were calculated to determine IC50 values (Table 9). Although all 9 compounds exhibited cytotoxic effects on cell lines, 2 of them, AM137 and AM139 were chosen to be further studied; due to their significantly lower IC50 values in HepG2 cell line, which is known to have wild type p53 expression.

Table 9: Cytotoxic bioactivities (IC50 µM) of compounds in for primary liver cancer cells.

|       | HepG2 | Huh7 | Hep3B | Mahlavu |
|-------|-------|------|-------|---------|
| AM118 | 5.1   | 7.6  | 6.2   | 3.6     |
| AM119 | 6.2   | 8.6  | 7.8   | 5.6     |
| AM129 | 14.8  | 10.1 | 12.7  | 15.1    |
| AM130 | 4.7   | 8.5  | 5.7   | 3.1     |
| AM136 | 6.7   | 8.6  | 9.2   | 4.1     |
| AM137 | **2.3** | 9.5 | 5.5  | 5.0     |
| AM139 | **0.4** | 8.2 | 6.6  | 5.8     |
| AM63c | 6.8   | 5.3  | 7.6   | 1.2     |
| AM87a | 10.9  | 11.3 | 10.5  | 5.5     |

SRB cytotoxicity assay was done in triplicates $R2 \geqslant 0.8$

The cytotoxic activity of AM139 was also monitored by Real-Time cell analyser; treatment of which was shown to result in complete inhibition of growth for all cell lines at 20 μM, while percent inhibition values decrease as the concentration values decrease (Figure 24). These result show that all compounds have cytotoxic activities on all 4 cell lines, among which AM137 and AM139 showing the most cytotoxic activities.
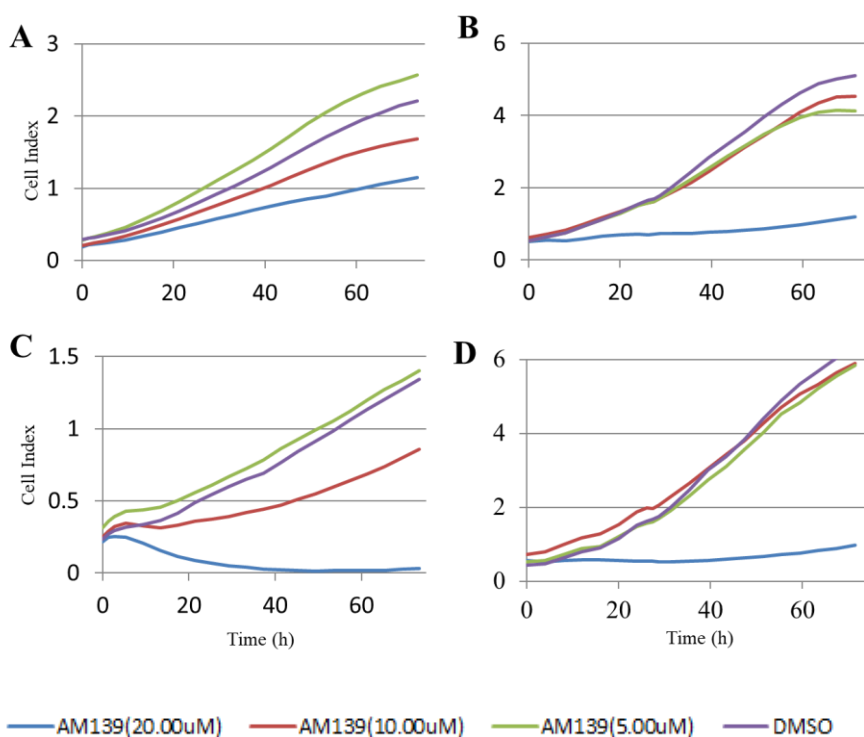


Figure 24: Cell index vs time graphs of HepG2 (A) Huh7 (B) Hep3B (C) and Mahlavu (D) cells obtained from real time cell analyzer; which were treated with AM139 at 3 different concentrations; 20.00 μM (Blue), 10.00 μM (Red), 5.00 μM (Green) or DMSO (Purple) for 72 hours.

### 3.8 Determination of the Cytotoxicity Mechanism

Next, further analyses were performed in order to understand the underlying mechanisms behind these cytotoxic effects. It is known that cell morphology is an indicator of health status of cells; so we examined the morphology differences of cells after the treatment with AM137, AM139 or DMSO control (Figure 25). From these images, the changes in cell shapes after the treatment with AM137 and AM139 when compared to healthy DMSO controls could clearly be seen. The small and round shaped cells in treated samples could be an indicator of presence of apoptotic cells.
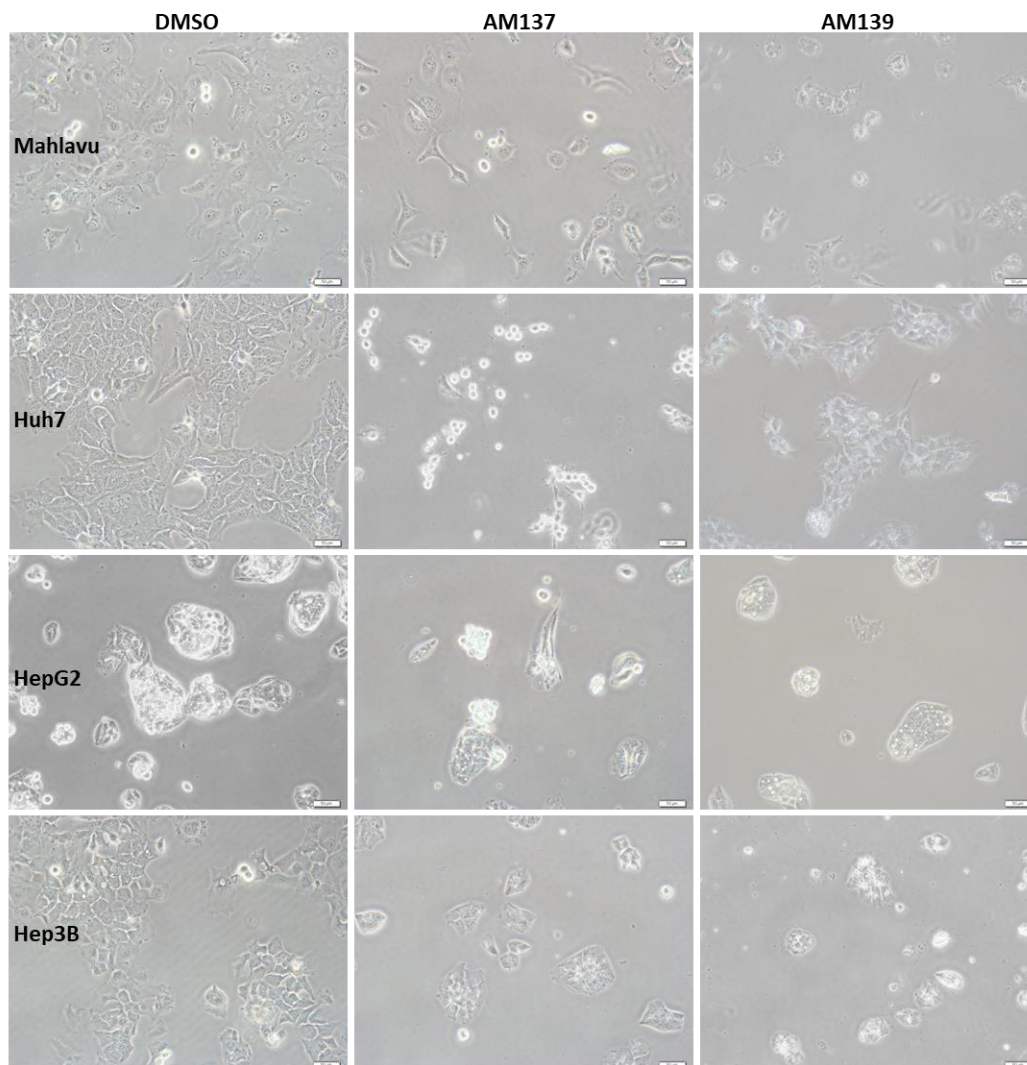
Figure 25: Morphologic changes in Mahlavu, Huh7, HepG2 and Hep3B cells in the presence of AM137 or AM139 (IC50) or DMSO control for 48 hours. All images were captured at 40x under an inverted phase-contrast microscope.

In order to further test this hypothesis, cell cycle analysis was performed (Figure 26) and percent cell population in each cell cycle stage was calculated for all samples. There was significant increase in cell percentages in sub-G1stage for especially Huh7 and Hep3B cell lines which shows that treatment with these compounds initiates apoptotic process.
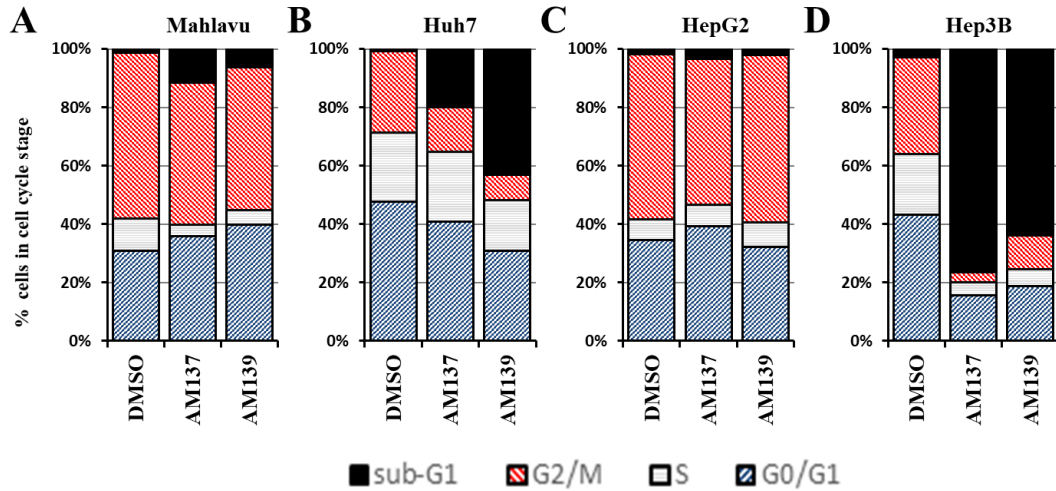
Figure 26: Histograms of cell cycle analysis results of HepG2 (A) Huh7 (B) Hep3B (C) and Mahlavu (D) cells after the treatment of AM137, AM139 (IC100 μM) or DMSO control for 48 hours. Percentages of sub-G1, G1, S, G2/M are given for each condition.

Previously, it was shown that Mdm2 inhibits p53 activity by initiating its nuclear export to stimulate protosomal degradation (Davis et al., 2013). To test whether treatment with AM137 and AM139 inhibits Mdm2-p53 binding, therefore preventing nuclear export and degradation of p53, we performed immunofluorescence analysis to determine the subcellular localization of p53 protein in treated and control cells (Figure 27). The results of immunofluorescence analysis revealed potential nuclear localization of p53 after treatment with compounds, when compared to the localization of p53 protein in both cytosol and nucleus in DMSO control cells. Since it is p53-null cell line, no p53 stain was detected in Hep3B cells, as expected. The nuclear localization of p53 after treatment will further be studied by confocal microscopy.

A

DMSO          4g          4i

B

DMSO          4g          4i
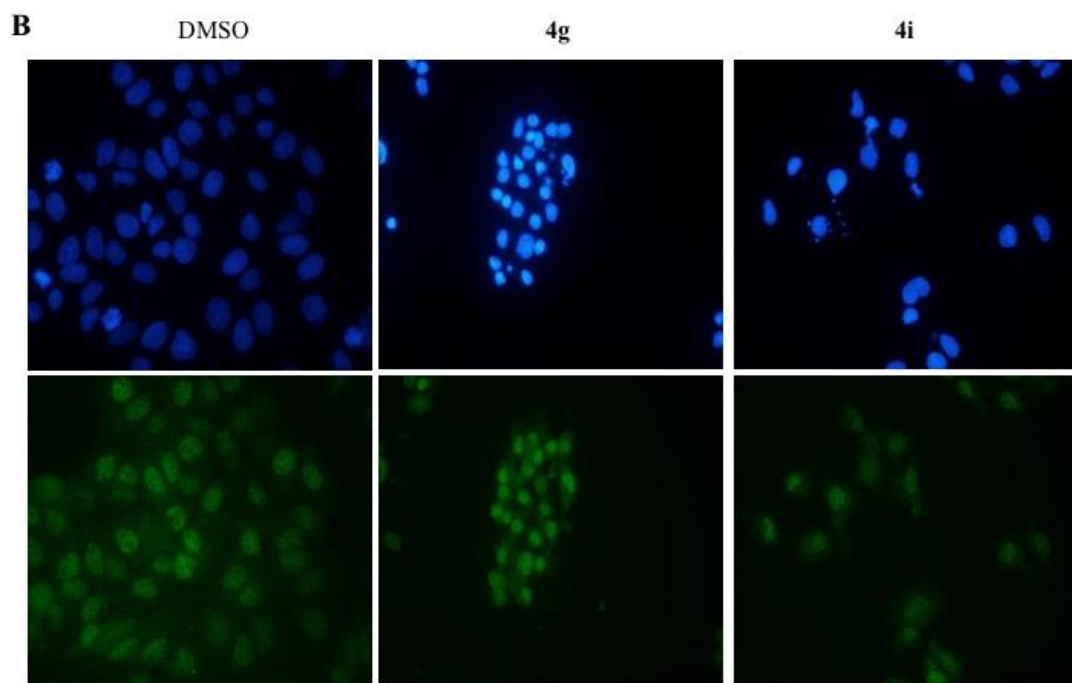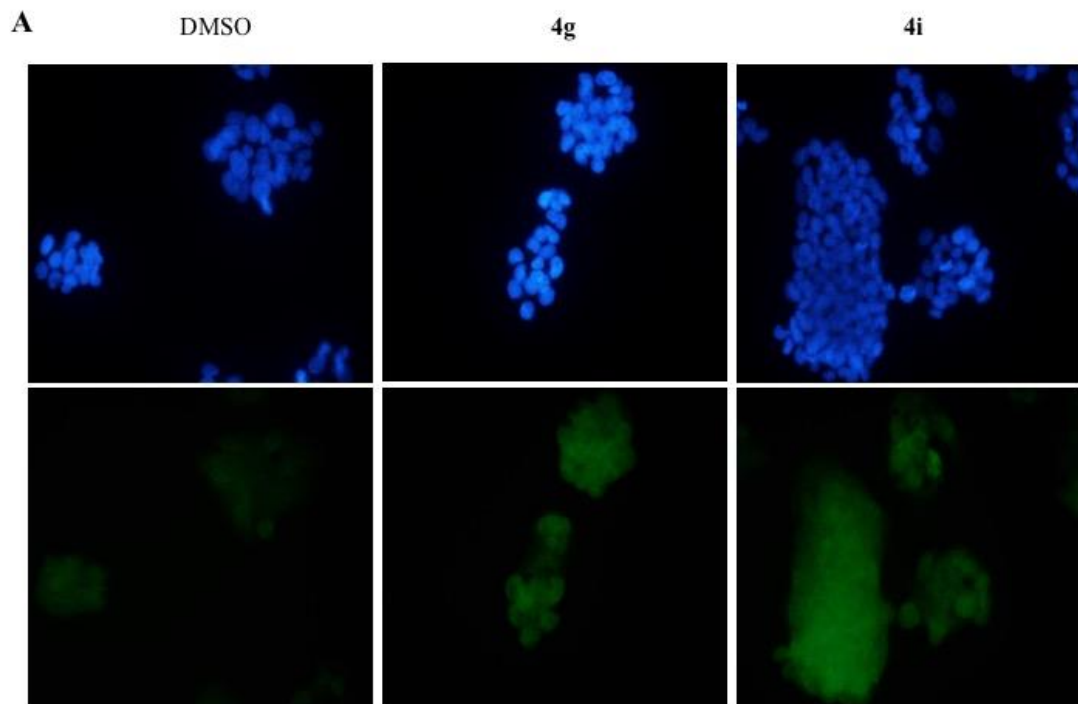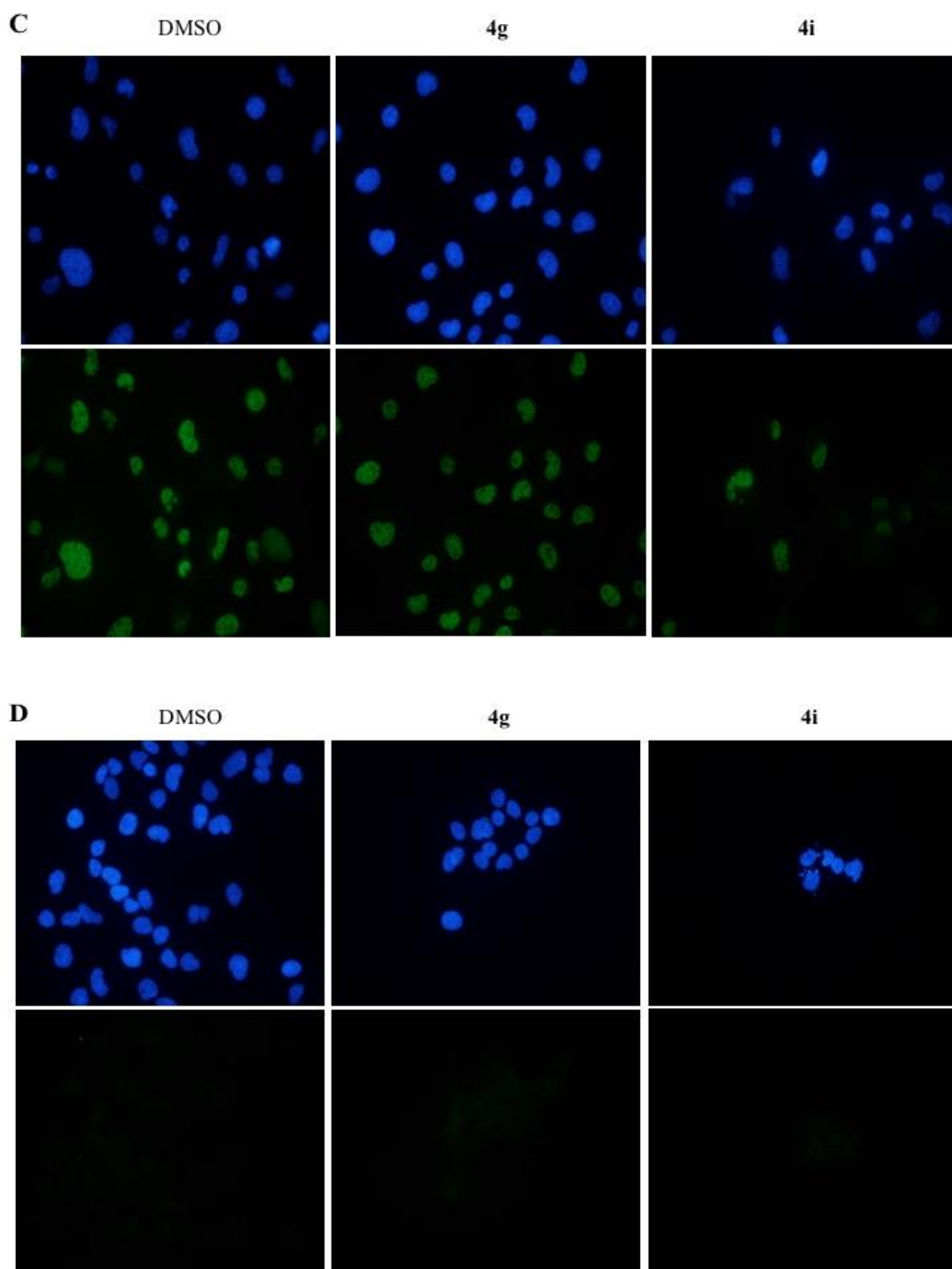
Figure 27: Immunofluorescence (bottom) and Hoechst (top) staining of HepG2 (A) Huh7 (B) Mahlavu (C) and Hep3B (D) cells after the treatment of AM137 (4g), AM139 (4i) (IC50 μM) or DMSO control for 48 hours to determine the subcellular localization of p53 protein in green fluorescence. Hoechst dye (blue) stains DNA and therefore nucleus. All images were captured at 40x under fluorescence microscope.

# CHAPTER 4

## DISCUSSION

In this thesis study, the aim was to determine novel biomarkers that could be used in the diagnosis and treatment of HCC and to test the potential cytotoxic effects of novel compounds on HCC cell lines. With this aim, Se deficiency-dependent oxidative stress related gene expression profiles of two isogenic HCC cell lines varying with respect to their HBV integration status were analyzed to identify genes that could be targeted as novel diagnostic and therapeutic strategies. Then the effects of a novel P53-MDM2 interaction inhibitor was tested as potential therapeutics. This chapters discuss the results obtained by the wet-lab and bioinformatics analysis methods.

### 4.1 The Examination of the Differentially Expressed Gene Lists

The results of an Affymetrix expression array were analyzed and the DEG lists were identified using limma analysis. DEG lists were further sub-categorized as 'Selenium-deficiency effect genes' or 'HBV-integration effect genes'.

### 4.1.1 The Comparison of Statistical Methods: t-test vs limma

t-test is one of the most commonly used statistical methods; but it could lead to some false results depending on the variance distribution of the samples and the sample size. When it is the subject of a microarray analysis, the use of t-test might give deceptive results because the genes with low variance values might distort variance estimates and could have large t-statistics to be falsely chosen as differentially expressed. Moreover, t-test is not the most suitable option when the sample sizes are low (Jeanmougin et al., 2010).

It was previously shown in various studies that limma gives better results in the identification of true DEGs, especially for the small sample sizes. This power of limma comes from the empirical Bayes estimate used in the calculation of a moderate standard deviation which takes the distribution of all the standard deviations into account (Jeanmougin et al., 2010). This feature of limma make it one of the most optimum statistical method to be used in our study when our small sample size was taken into consideration.

### 4.1.2 The Overall Interpretation of the Differentially Expressed Gene Lists

The DEGs were sub-categorized either as 'HBV-integration effects genes' or 'Se-deficiency effect genes' depending on being shared in HepG2 vs HepG2-2.2.15 comparisons for both Se+ and Se- conditions or only found in Se- comparison, respectively. The shared DEGs were thought to be related to HBV-integration effect since they were differentially expressed independent of the Se status. Our CCLE clustering results verified these DEGs were as a result of HBV-integration effect, since the two cell lines lacking HBV were clustered together distinctly from the cell lines that had HBV genomic integration when the expression levels of 'HBV-integration effect genes' were considered. The expression levels of some of the genes from our HBV-integration effect gene list were also previously shown to be altered depending on HBV infection by various studies. (Jagya et al., 2014; Lamontagne et al., 2016). GSE52755 dataset in GEO database analyze the gene expression changes in human liver cells after HBV infection, so comparing our HBV-integration effect DEGs with their findings were also enlightening to further verify the association between HBV-integration and alterations in the expression of these genes; 34 common genes (KLHL14, GDA, DPYD, IL6R, SETBP1, TUBB2B, FOXN3, GPC3, RBMS1, GPR137B, RPS23, PLP2, RBM23, BHMT2, HOXA10, FBXL21, LGI1, NR5A2, SLC38A4, CYB561, SPP1, DKK4, L3MBTL4, HABP2, SEMA4F, PRTG, DEFB1, GSTO2, STXBP6, VNN1, RBM47, CTDSPL, ARL15, NME3) were identified; but the significance of this association have to be verified by statistical methods.

We found that HBV integration had a larger effect on differential gene expression compared to that of Se deficiency when the DEG numbers were taken into consideration. The most dramatic Se-deficiency effect was seen on the second day which might be due to the use of Se from the cellular stocks in the first 24 hours and the deprivation occurs afterwards altering the gene expression levels.
For the *within cell line* comparisons, GPX and SEPW1 expressions was lost in both cell lines under Se-deficiency; which was an expected result as the dependence of the synthesis of these two selenoproteins on the presence of Se is considered. There was no other shared DEGs between HepG2 and HepG2-2.2.15 within cell line comparisons, which might be an indicator of difference between the reactions of these two cell lines to Se-deficiency.

### 4.2 The Determination of Key Genes in the Differential Response to Selenium Deficiency

The genes thought to play key roles in the differential response to Se deficiency-dependent oxidative stress were determined by clustering, GSEA as well as network analysis methods and 27 genes were identified to be the most significant ones and can provide important leads in further studies of HCC diagnosis and therapy.

Protein-protein interaction (PPI) networks are powerful tools to define some key proteins within known networks. Some values such as degree of a node and betweenness centrality might be used to determine biologically important hub proteins (H. Yu et al., 2007).

Fifteen DEGs with high betweenness centrality values were identified by the PCST analysis for the *between cell line* comparison results, and eleven of the genes were Steiner nodes which were not differentially expressed but were determined to be on key positions on pathways that might have effects on the expression of the determined DEGs having indirect effects on differential response. Previous studies have shown that such hidden nodes are able to connect the pathways and also indicate cross talks (S. -s. C. Huang & Fraenkel, 2009). Fourteen of the genes identified by PCST were previously shown to have functions in oxidative stress response and seven of those were also known to be related to HCC (Table 5). This suggest that our statistical methodology has revealed cancer pathways that could be related with hepatocarcinogenesis and extrinsic factors.

With the GSEA, six DEGs (DUT, POLD3, E2F2, GINS2, PIK3R3, TMEM97) as a result of within cell line comparisons were determined to be key elements for the Se effect and four (POLD3, E2F2, GINS2, PIK3R3) were previously shown to have oxidative stress and hepatocellular carcinogenesis related functions (Table 5).

Three genes from each HBV and Se effect DEGs were selected from the heatmap analysis results with the smallest p-values that have the most significant impact on the differential response to Se deficiency between the isogenic HepG2 and HepG2-2.2.15 cell lines. PPAP2A, one of the three genes with lowest p-values was determined to have effects on cell proliferation a previous study but the other two, HOXD1 and CLYBL genes were not associated with any oxidative stress or hepatocellular carcinogenesis related function but the association of HOXD1 with ovarian cancer was previously reported (Table 5).

GSEA results also indicated that the expression of 15 of those genes were regulated by different combinations of six transcription factors as shown in Table 10. HSD17B8 overexpression was found to be associated with endometrial and breast cancers; so proposed as potential target in previous studies (Cornel et al., 2012; D. Song et al., 2006). ZBTB5 genes' role as transcriptional repressor of p21, which is a cell cycle arrest regulator makes it a stimulator of cell cycle; so known to possess a proto-oncogenic role (Koh et al., 2009). TFCP2 is another gene with proto-oncogenic roles in HCC, pancreatic and breast cancers (Kotarba et al., 2018) while the overexpression of LYF1 transcription factor was shown to be associated to the prevention of hepatocarcinogenesis (Huo et al., 2017). E2F2 is another transcription factor whose roles in HCC was studied. The overexpression of E2F2 results in HCC progression through affecting various pathways such as cell cycle, DNA replication and p53 signaling thus it was proposed as promising target for HCC (Zeng et al., 2020). In line with our findings and the previous studies, these transcription factors might be the main targets that should be further investigated for their potential in novel diagnostics and therapeutic strategies.

Table 10: The common transcription factors that were associated with the regulation of genes-of interest identified. r: reported.

| GENE NAME | HSD17B8 | CHX10 | ZBTB5 | TFCP2 | LYF1 | E2F2 |
|---|---|---|---|---|---|---|
| FOXA1 | r | | r | r | | |
| CYP7A1 | | | | | | |
| ONECUT1 | | | | | | |
| PITX2 | | r | | r | r | r |
| TXNRD1 | | | | | | |
| ALDH1L2 | | | | | | |
| ACLY | | | | | | |
| TXNIP | | | r | | | |
| SCD5 | | | | | | |
| MTR | | | r | | | r |
| TXNDC17 | | | | r | | |
| LSM4 | | | | | | |
| CNBP | | | | | | |
| DMPK | | | | | r | |
| QDPR | | | | | | |
| DUT | r | | | | | r |
| POLD3 | r | | | | | r |
| E2F2 | r | | | | | r |
| GINS2 | r | | | | | |
| PIK3R3 | | r | r | | | r |
| TMEM97 | r | | | | | r |
| FGF13 | | | r | | r | |
| GPC3 | | r | | r | r | |
| MAP7D2 | | | | | | |
| PPAP2A | | r | | | | |
| HOXD1 | | | | | | |
| CLYBL | | | | | | |

**4.3 The Association of the Identified Key Genes in the Differential Response to Selenium-Deficiency With Their Clinical Relevance**

The expression of 16 genes among 27 genes were found to be significantly related to clinical results; the lower expression of ACLY, LSM4, PITX2, TXNRD1, POLD3, GINS2, FGF13, E2F2 and higher expression of ALDH1L2, CYP7A1, QDPR, TXNIIP, PPAP2A, PIK3R3, H2BFXP, CLYBL genes were found to have significant effects on liver cancer patient life spans as shown by the Kaplan Meier Plots. This suggests that especially these 16 genes might be important in the differential response to ROS – dependent effects and should be further searched for their roles in oxidative stress defense mechanisms.

Among the 16 genes associated with significant clinical results, FGF13 gene was the only DEG proposed as 'HBV-integration effect gene' in our study. When we separated the effect of expression of FGF13 on the life span of liver cancer patients either dependent on HBV or not, we found that the effect of gene was only significant for the life span of HBV-dependent liver cancer patients and not for the ones that lack HBV (Appendix C). This might further support the proposed association of this gene to HBV-integration effect.

## 4.4 The Impact of the Identified Genes

The identification of genes that could be used to predict the prognosis of the HCC patient or lead to the discovery of new therapeutic strategy is important since cancer cells find alternative pathways to compensate the effects of targeted therapies that are currently being used. For this reason, the identification of novel biomarkers that play key roles on those compensatory pathways is very critical. The use of computational methods to analyze the regulation of gene expression values in cancer cells is a powerful guide to develop novel targeted-therapeutic strategies.

The outputs of this study emphasize the role of DEGs regardless of Se status and as a result of HBV-integration effect as potential targets in HBV-dependent HCC treatments. This information might be important since approximately half of the HCC incidences are known to be dependent on viral infection. HBV virus encodes oncogenic HBx protein after infecting cells and the expression of HBx is related to many outcomes by altering host gene expressions to modulate proliferative mechanisms in hepatocytes and stimulate viral replication. WU_HBX_TARGETS_3_UP gene set is found in the Molecular Signatures Database that consist of 18 genes up-regulated by expression of HBx protein both in SK-Hep-1 cells and primary hepatocytes (gene list provided in Appendix D, Table 17). In GSEA, this list was found to be enriched in HepG2-2.2.15 compared to HepG2 cells that were grown in the presence of Se (Appendix D, Figure 29) indicating the active role of HBx protein in HepG2-2.2.15 cells. Through these genes, 5 of them (CCNI, CDKN3, MYB, PDCD2 and SLC2A5) were identified to be either directly present or known to be interacting with the DEGS in our extended list (Appendix B) suggesting the importance of HBx dependent gene expression alterations in oxidative stress resistance consistently with our findings.

Moreover, the effects of Se rich diet on the treatment of HCC patients might further be studied to reveal genes identified as potential drivers of Se deficiency effect in this study. The role of dietary Se in carcinogenesis is a controversial issue. There are various studies focusing on anti-cancer roles of Se through several mechanisms, including selenoproteins' antioxidant functions, tumor cell growth inhibitions, stimulatory roles in cell cycle and apoptosis, and effect on DNA repair. However, another study showed the effect of dietary Se on progression of malignant mesothelioma tumors, by using Se for an increased reducing capacity to stimulate their redox metabolism. (Papp et al., 2007) Moreover, Se might also has toxic effects if consumed in excessive amounts resulting in poisoning. Although previous *in vitro* analysis results with HCC cell lines did not detect any Se-dependent toxic effects (Irmak et al., 2003), Se consumptions should be carefully adjusted,

and the effects should be further studied to unveil the mechanisms behind the conflictive effects.

## 4.5 Correlations Between our DEGs and p53-MDM2 Pathway Gene Expression Values

The role of P53 gene in stress conditions for the regulation of the reaction of cells to the stress factors such as ROS was explained in the Introduction chapter. FOXA1 was found to be differentially expressed in all 3 days in between cell line comparisons and it is known to have regulatory roles on the transcription of P53 and MDM2 genes. P53 protein was also shown to be interacting with 7 of the DEGs that were identified to be important in differential response to Se-deficiency. Through the common transcription factors identified that are associated with the expression of our DEGs of interest, some has functions related to P53 signaling, such as E2F2 gene. We further searched for any possible correlations between the expression of our 27-gene DEG list and the genes that are parts of p53-MDM2 pathway and many significant correlations were identified. All these results might highlight significant role of p53-MDM2 pathway in the differential response to Se-deficiency dependent oxidative stress; although the expression of the genes that constitute that pathway was not differentially expressed themselves. Analysis of the p53-MDM2 pathway genes at the protein level might give differential results between HepG2 and HepG2-2.2.15 cells under Se-deficiency, unlike our results that revealed the mRNA level results and might be further tested.

## 4.6 p53-MDM2 Interaction Inhibitors Gave Promising Results in HCC Cell Lines With Stem-Like Properties

In this part of the study, 9 different p53-Mdm2 binding inhibitors were tested for their cytotoxic effects on HCC. 2 of them were determined to be the most promising ones due to their lowest IC50 values in HepG2 cell line, which is known to have wild type p53 expression. By the examination of morphology changes and cell cycle analysis experiments, these inhibitors were shown to induce apoptosis. The nuclear localization of p53 should further be studied by confocal microscopy.

The tested p53-MDM2 interaction inhibitors were spiropyrazoline oxindoles and their cytotoxic effects were shown in MCF-7 breast cancer cell and cancer stem cells previously (Amaral et al., 2019; Monteiro et al., 2014). The p53 null status of Hep3B cell line suggests that these cells might gain stem cell like properties through stem cell reprogramming. The significantly increased Hep3B cell percentages in sub-G1 stage after treatment with AM137 and AM139 in cell cycle analysis results highly correlates with the increased cytotoxic effects of these compounds to cancer stem cells. This further shows that there might be a preferential activity of these compounds against cancer stem cells.

The immunofluorescence data on liver cancer cells with HepG2 cells indicate the activation and the nuclear localization of p53 in order to initiate mechanisms involved in DNA repair. Therefore, we do not observe SubG1 arrest due to active p53 protein in

HepG2 cells. The cell death induced AM137 and AM139 must be due to other pathways in these cells that must be studied in the future. However, in the cells with unfunctional (Huh7 and Mahlavu) or null (Hep3B) for p53 protein we do not observe any difference in p53 localization upon treatment with AM137 and AM139, which indicates that functional p53 protein is not required for strong SubG1 arrest. HBV infection results in accumulation of ROS through various mechanisms leading to oxidative stress if could not be balanced (Wang et al., 2016). In such a stress condition one of the known regulated pathways is p53 tumor suppressor gene but HBx protein is known to be associated with its loss of function mutation (Urso et al., 2015). Both Huh7 and Mahlavu cells have p53 loss of function mutations on Y220C and R249S respectively based on HBx or aflatoxin as stress-inducing factors. Therefore their response to p53-MDM2 interaction inhibitors were different compared to p53 wild-type HepG2 cells' when immunofluorescence results are considered.

As stated above, loss of p53 function is correlated with the stemness of cells. In correlation with previous studies, the overall results support that AM137 and AM139 are active on cells having stem cell like properties. It might be interesting to test the effects of AM137 and AM139 on HepG2-2.2.15 cell line to compare the cytotoxicity results with that of HepG2. HepG2-2.2.25 is expected to have more stem-like property with respect to the HBV infection (Mani & Andrisani, 2018) and depending on our findings, we might expect a higher cytotoxic activity of AM137 and AM139 compounds on HepG2-2.2.15 cells.

# CHAPTER 5

## CONCLUSION

Hepatocellular carcinoma (HCC) is one of the most common and fatal cancer types. There are some treatment options used for the treatment of HCC patients; but their efficacies are low due to the resistance gained by cancer cells and there is still a need for the identification of novel biomarkers that could be used in diagnosis and novel targeted therapy strategies. This brings the necessity for the investigation of HCC cell characteristics driving their increased resistance. One of the known properties of HCC cells is their resistance to various stress conditions to survive from stress-induced apoptosis, such as their resistance to increased reactive oxygen species (ROS) unlike normal cells, which use various mechanisms to reduce ROS and the inability of it results in apoptosis.

Normal cells use Selenium trace element in their defense mechanisms against oxidative stress and the lack of Selenium results in oxidative stress-dependent apoptosis. However, some HCC cell lines were shown to be resistant to Selenium-deficiency dependent oxidative stress and the mechanism behind this tolerance was unknown. In this thesis study, we have analyzed the results of a transcriptome data for the determination of the resistance mechanisms between the sensitive and resistant HCC cell lines. We have identified 27 genes that were differentially expressed between Selenium-deficiency dependent oxidative stress sensitive and resistant cell lines which were subcategorized as "HBV-integration effect genes" or "Se-deficiency effect genes" and proposed to have key roles in oxidative stress resistance. The expression of 15 of these genes were found to be regulated by 6 common transcription factors and their potential as HCC biomarkers might be searched in further studies. The genes that were identified as HBV-integration effect genes might be promising targets for the inhibition of the resistance gained by cancer cells. Moreover, the application of Selenium rich diet for HCC patients might further be studied to reveal genes identified as potential drivers of Se deficiency effect in this study.

FOXA1 gene was proposed to be one of the most important DEGs in the oxidative stress-resistance due to its regulatory role in p53-MDM2 pathway and many significant correlations between the expressions of our 27 genes and p53-MDM2 pathway genes were identified. This might indicate a role of p53-MDM2 pathway in the resistance gained to selenium-deficiency dependent oxidative stress. Analysis of the abundance of these genes at the protein level in the resistant and sensitive cell lines in the presence or absence of Selenium might be very informative for understanding the role of p53-MDM2 pathway in the resistance mechanism.

We also revealed the significant cytotoxic effects of two p53-MDM2 interaction inhibitors which gave promising results in HCC cell lines with stem-like properties due to either unfunctional or lack of p53 activity. We have shown that the application of these compounds resulted in nuclear localization of p53 in HepG2 cell line with wild type p53; while no difference in the localization of p53 in cell lines that lack functional p53 was detected resulting in sub-G1 arrest. The nuclear localization of p53 in HepG2 cells might be related to the activation of DNA repair mechanisms and this should be further tested by experimental techniques. Overall, this two spiropyrazoline oxindoles compounds can be proposed to be promising agents for HCC therapy.

# REFERENCES

Ak, P., & Levine, A. J. (2010). p53 and NF-κB: different strategies for responding to stress lead to a functional antagonism. *The FASEB Journal*, *24*(10). https://doi.org/10.1096/fj.10-160549

Akutsu, N. (2010). Association of glypican-3 expression with growth signaling molecules in hepatocellular carcinoma. *World Journal of Gastroenterology*, *16*(28). https://doi.org/10.3748/wjg.v16.i28.3521

Amaral, J. D., Silva, D., Rodrigues, C. M. P., Solá, S., & Santos, M. M. M. (2019). A Novel Small Molecule p53 Stabilizer for Brain Cell Differentiation. *Frontiers in Chemistry*, *7*. https://doi.org/10.3389/fchem.2019.00015

An, P., Xu, J., Yu, Y., & Winkler, C. A. (2018). Host and viral genetic variation in HBV-related hepatocellular carcinoma. In *Frontiers in Genetics* (Vol. 9, Issue JUL). Frontiers Media S.A. https://doi.org/10.3389/fgene.2018.00261

Archer, K. J., Mas, V. R., Maluf, D. G., & Fisher, R. A. (2010). High-throughput assessment of CpG site methylation for distinguishing between HCV-cirrhosis and HCV-associated hepatocellular carcinoma. *Molecular Genetics and Genomics*, *283*(4). https://doi.org/10.1007/s00438-010-0522-y

Bernardo, G. M., Bebek, G., Ginther, C. L., Sizemore, S. T., Lozada, K. L., Miedler, J. D., Anderson, L. A., Godwin, A. K., Abdul-Karim, F. W., Slamon, D. J., & Keri, R. A. (2013). FOXA1 represses the molecular phenotype of basal breast cancer cells. *Oncogene*, *32*(5). https://doi.org/10.1038/onc.2012.62

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *68*(6), 394–424. https://doi.org/10.3322/caac.21492

Bublik, D. R., Bursać, S., Sheffer, M., Oršolić, I., Shalit, T., Tarcic, O., Kotler, E., Mouhadeb, O., Hoffman, Y., Fuchs, G., Levin, Y., Volarević, S., & Oren, M. (2017). Regulatory module involving FGF13, miR-504, and p53 regulates

ribosomal biogenesis and supports cancer cell survival. *Proceedings of the National Academy of Sciences*, *114*(4). https://doi.org/10.1073/pnas.1614876114

Cabebe, E. C. (2021). Hepatocellular Carcinoma (HCC) Guidelines. *Medscape*.

Casaril, M., Corso, F., Bassi, A., Capra, F., Gabrielli, G. B., Stanzial, A. M., Nicoli, N., & Corrocher, R. (1994). Decreased activity of scavenger enzymes in human hepatocellular carcinoma, but not in liver metastases. *International Journal of Clinical & Laboratory Research*, *24*(2). https://doi.org/10.1007/BF02593907

Castillo, D. S., Campalans, A., Belluscio, L. M., Carcagno, A. L., Radicella, J. P., Cánepa, E. T., & Pregi, N. (2015). E2F1 and E2F2 induction in response to DNA damage preserves genomic stability in neuronal cells. *Cell Cycle*, *14*(8). https://doi.org/10.4161/15384101.2014.985031

Cavga, A. D., Tardu, M., Korkmaz, T., Keskin, O., Ozturk, N., Gursoy, A., & Kavakli, I. H. (2019). Cryptochrome deletion in p53 mutant mice enhances apoptotic and anti-tumorigenic responses to UV damage at the transcriptome level. *Functional & Integrative Genomics*, *19*(5). https://doi.org/10.1007/s10142-019-00680-5

Chen, B., Garmire, L., Calvisi, D. F., Chua, M. S., Kelley, R. K., & Chen, X. (2020a). Harnessing big 'omics' data and AI for drug discovery in hepatocellular carcinoma. In *Nature Reviews Gastroenterology and Hepatology* (Vol. 17, Issue 4, pp. 238–251). Nature Research. https://doi.org/10.1038/s41575-019-0240-9

Chen, B., Garmire, L., Calvisi, D. F., Chua, M.-S., Kelley, R. K., & Chen, X. (2020b). Harnessing big 'omics' data and AI for drug discovery in hepatocellular carcinoma. *Nature Reviews Gastroenterology & Hepatology*, *17*(4). https://doi.org/10.1038/s41575-019-0240-9

Chen, L., & Liu, B. (2017). Relationships between Stress Granules, Oxidative Stress, and Neurodegenerative Diseases. *Oxidative Medicine and Cellular Longevity*, *2017*. https://doi.org/10.1155/2017/1809592

Chène, P. (2003). Inhibiting the p53–MDM2 interaction: an important target for cancer therapy. *Nature Reviews Cancer*, *3*(2). https://doi.org/10.1038/nrc991

Cooke, M. S., Evans, M. D., Dizdaroglu, M., & Lunec, J. (2003). Oxidative DNA damage: mechanisms, mutation, and disease. *The FASEB Journal*, *17*(10), 1195–1214. https://doi.org/10.1096/fj.02-0752rev

Cornel, K. M. C., Kruitwagen, R. F. P. M., Delvoux, B., Visconti, L., van de Vijver, K. K., Day, J. M., van Gorp, T., Hermans, R. J. J., Dunselman, G. A., & Romano, A. (2012). Overexpression of 17β-Hydroxysteroid Dehydrogenase

Type 1 Increases the Exposure of Endometrial Cancer to 17β-Estradiol. *The Journal of Clinical Endocrinology & Metabolism*, *97*(4). https://doi.org/10.1210/jc.2011-2994

Czeczot, H., Ścibior, D., Skrzycki, M., & Podsiad, M. (2006). *Glutathione and GSH-dependent enzymes in patients with liver cirrhosis and hepatocellular carcinoma* (Vol. 53, Issue 1). www.actabp.pl

Davis, J. R., Mossalam, M., & Lim, C. S. (2013). Controlled Access of p53 to the Nucleus Regulates Its Proteasomal Degradation by MDM2. *Molecular Pharmaceutics*, *10*(4). https://doi.org/10.1021/mp300543t

de Nadal, E., Ammerer, G., & Posas, F. (2011). Controlling gene expression in response to stress. In *Nature Reviews Genetics* (Vol. 12, Issue 12, pp. 833–845). https://doi.org/10.1038/nrg3055

de Peralta, M. S. P., Mouguelar, V. S., Sdrigotti, M. A., Ishiy, F. A. A., Fanganiello, R. D., Passos-Bueno, M. R., Coux, G., & Calcaterra, N. B. (2016). Cnbp ameliorates Treacher Collins Syndrome craniofacial anomalies through a pathway that involves redox-responsive genes. *Cell Death & Disease*, *7*(10). https://doi.org/10.1038/cddis.2016.299

di Maso, V., Mediavilla, M. G., Vascotto, C., Lupo, F., Baccarani, U., Avellini, C., Tell, G., Tiribelli, C., & Crocè, L. S. (2015). Transcriptional Up-Regulation of APE1/Ref-1 in Hepatic Tumor: Role in Hepatocytes Resistance to Oxidative Stress and Apoptosis. *PLOS ONE*, *10*(12). https://doi.org/10.1371/journal.pone.0143289

Engedal, N., Žerovnik, E., Rudov, A., Galli, F., Olivieri, F., Procopio, A. D., Rippo, M. R., Monsurrò, V., Betti, M., & Albertini, M. C. (2018). From Oxidative Stress Damage to Pathways, Networks, and Autophagy via MicroRNAs. *Oxidative Medicine and Cellular Longevity*, *2018*. https://doi.org/10.1155/2018/4968321

Fulda, S., Gorman, A. M., Hori, O., & Samali, A. (2010). Cellular Stress Responses: Cell Survival and Cell Death. *International Journal of Cell Biology*, *2010*. https://doi.org/10.1155/2010/214074

Galili, T., O'Callaghan, A., Sidi, J., & Sievert, C. (2018). heatmaply: an R package for creating interactive cluster heatmaps for online publishing. *Bioinformatics*, *34*(9). https://doi.org/10.1093/bioinformatics/btx657

Gong, Y., Dong, F., Geng, Y., Zhuang, K., Ma, Z., Zhou, Z., Huang, B., Su, Z., & Hou, B. (2019). Selenium concentration, dietary intake and risk of hepatocellular carcinoma – A systematic review with meta-analysis. *Nutrición Hospitalaria*. https://doi.org/10.20960/nh.02776

Grimes, T., Potter, S. S., & Datta, S. (2019). Integrating gene regulatory pathways into differential network analysis of gene expression data. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-41918-3

Grose, R., Coleman, S., & Kocher, H. (2014). Fibroblast growth factor family as a potential target in the treatment of hepatocellular carcinoma. *Journal of Hepatocellular Carcinoma*. https://doi.org/10.2147/JHC.S48958

Gu, Y., Wang, Y., Zhang, H., Zhao, T., Sun, S., Wang, H., Zhu, B., & Li, P. (2017). Protective effect of dihydropteridine reductase against oxidative stress is abolished with A278C mutation. *Journal of Zhejiang University-SCIENCE B*, *18*(9). https://doi.org/10.1631/jzus.B1600123

Guillin, O. M., Vindry, C., Ohlmann, T., & Chavatte, L. (2019). Selenium, selenoproteins and viral infection. In *Nutrients* (Vol. 11, Issue 9). MDPI AG. https://doi.org/10.3390/nu11092101

Guo, M., Zhang, H., Zheng, J., & Liu, Y. (2020). Glypican-3: A New Target for Diagnosis and Treatment of Hepatocellular Carcinoma. *Journal of Cancer*, *11*(8). https://doi.org/10.7150/jca.39972

Halliwell, B. (2007). Oxidative stress and cancer: have we moved forward? *Biochemical Journal*, *401*(1). https://doi.org/10.1042/BJ20061131

Hong, H., Takahashi, K., Ichisaka, T., Aoi, T., Kanagawa, O., Nakagawa, M., Okita, K., & Yamanaka, S. (2009). Suppression of induced pluripotent stem cell generation by the p53–p21 pathway. *Nature*, *460*(7259). https://doi.org/10.1038/nature08235

Huang, A., Yang, X. R., Chung, W. Y., Dennison, A. R., & Zhou, J. (2020). Targeted therapy for hepatocellular carcinoma. In *Signal Transduction and Targeted Therapy* (Vol. 5, Issue 1). Springer Nature. https://doi.org/10.1038/s41392-020-00264-x

Huang, S. -s. C., & Fraenkel, E. (2009). Integrating Proteomic, Transcriptional, and Interactome Data Reveals Hidden Components of Signaling and Regulatory Networks. *Science Signaling*, *2*(81), ra40–ra40. https://doi.org/10.1126/scisignal.2000350

Huang, Y.-L., Ning, G., Chen, L.-B., Lian, Y.-F., Gu, Y.-R., Wang, J.-L., Chen, D.-M., Wei, H., & Huang, Y.-H. (2019). Promising diagnostic and prognostic value of E2Fs in human hepatocellular carcinoma. *Cancer Management and Research*, *Volume 11*. https://doi.org/10.2147/CMAR.S182001

Huo, Q., Ge, C., Tian, H., Sun, J., Cui, M., Li, H., Zhao, F., Chen, T., Xie, H., Cui, Y., Yao, M., & Li, J. (2017). Dysfunction of IKZF1/MYC/MDIG axis

contributes to liver cancer progression through regulating H3K9me3/p21 activity. *Cell Death & Disease*, *8*(5). https://doi.org/10.1038/cddis.2017.165

Ibrahim, S., Li, G., Hu, F., Hou, Z., Chen, Q., Li, G., Luo, X., Hu, J., & Feng, Y. (2018). PIK3R3 promotes chemotherapeutic sensitivity of colorectal cancer through PIK3R3 /NF-kB/TP pathway. *Cancer Biology & Therapy*, *19*(3). https://doi.org/10.1080/15384047.2017.1416936

Iizuka, N., Oka, M., Yamada-Okabe, H., Mori, N., Tamesa, T., Okada, T., Takemoto, N., Hashimoto, K., Tangoku, A., Hamada, K., Nakayama, H., Miyamoto, T., Uchimura, S., & Hamamoto, Y. (2003). Differential gene expression in distinct virologic types of hepatocellular carcinoma: association with liver cirrhosis. *Oncogene*, *22*(19). https://doi.org/10.1038/sj.onc.1206401

Irmak, M. B., Ince, G., Ozturk, M., & Cetin-Atalay, R. (2003). Acquired Tolerance of Hepatocellular Carcinoma Cells to Selenium Deficiency: A Selective Survival Mechanism? 1. In *CANCER RESEARCH* (Vol. 63).

Jagya, N., Varma, S. P. K., Thakral, D., Joshi, P., Durgapal, H., & Panda, S. K. (2014). RNA-Seq Based Transcriptome Analysis of Hepatitis E Virus (HEV) and Hepatitis B Virus (HBV) Replicon Transfected Huh-7 Cells. *PLoS ONE*, *9*(2). https://doi.org/10.1371/journal.pone.0087835

Jeanmougin, M., de Reynies, A., Marisa, L., Paccard, C., Nuel, G., & Guedj, M. (2010). Should We Abandon the t-Test in the Analysis of Gene Expression Microarray Data: A Comparison of Variance Modeling Strategies. *PLoS ONE*, *5*(9). https://doi.org/10.1371/journal.pone.0012336

Jenkins, E., Brenner, M., Laragione, T., & Gulko, P. S. (2012). Synovial expression of Th17-related and cancer-associated genes is regulated by the arthritis severity locus Cia10. *Genes & Immunity*, *13*(3). https://doi.org/10.1038/gene.2011.73

Jensen, L. J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., Doerks, T., Julien, P., Roth, A., Simonovic, M., Bork, P., & von Mering, C. (2009). STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Research*, *37*(Database). https://doi.org/10.1093/nar/gkn760

Jiang, L., Wu, C.-L., Wu, J.-Z., Yang, X., Wang, H., & Li, G.-J. (2019). Microarray-based measurement of microRNA-449c-5p levels in hepatocellular carcinoma and bioinformatic analysis of potential signaling pathways. *Pathology - Research and Practice*, *215*(1). https://doi.org/10.1016/j.prp.2018.10.007

Kiermayer, C., Michalke, B., Schmidt, J., & Brielmeier, M. (2007). Effect of selenium on thioredoxin reductase activity in Txnrd1 or Txnrd2 hemizygous mice. *Biological Chemistry*, *388*(10). https://doi.org/10.1515/BC.2007.133

Koh, D.-I., Choi, W.-I., Jeon, B.-N., Lee, C.-E., Yun, C.-O., & Hur, M.-W. (2009). A Novel POK Family Transcription Factor, ZBTB5, Represses Transcription of p21CIP1 Gene. *Journal of Biological Chemistry*, *284*(30). https://doi.org/10.1074/jbc.M109.025817

Kotarba, G., Krzywinska, E., Grabowska, A. I., Taracha, A., & Wilanowski, T. (2018). TFCP2/TFCP2L1/UBP1 transcription factors in cancer. *Cancer Letters*, *420*. https://doi.org/10.1016/j.canlet.2018.01.078

Kreyszig, E. (1979). *Advanced engineering mathematics*. Wiley.

Lamontagne, J., Mell, J. C., & Bouchard, M. J. (2016). Transcriptome-Wide Analysis of Hepatitis B Virus-Mediated Changes to Normal Hepatocyte Gene Expression. *PLOS Pathogens*, *12*(2). https://doi.org/10.1371/journal.ppat.1005438

Lee, D., Xu, I. M., Chiu, D. K., Leibold, J., Tse, A. P., Bao, M. H., Yuen, V. W., Chan, C. Y., Lai, R. K., Chin, D. W., Chan, D. F., Cheung, T., Chok, S., Wong, C., Lowe, S. W., Ng, I. O., & Wong, C. C. (2019). Induction of Oxidative Stress Through Inhibition of Thioredoxin Reductase 1 Is an Effective Therapeutic Approach for Hepatocellular Carcinoma. *Hepatology*, *69*(4). https://doi.org/10.1002/hep.30467

Lee, D., Xu, I. M.-J., Chiu, D. K.-C., Lai, R. K.-H., Tse, A. P.-W., Lan Li, L., Law, C.-T., Tsang, F. H.-C., Wei, L. L., Chan, C. Y.-K., Wong, C.-M., Ng, I. O.-L., & Wong, C. C.-L. (2017). Folate cycle enzyme MTHFD1L confers metabolic advantages in hepatocellular carcinoma. *Journal of Clinical Investigation*, *127*(5). https://doi.org/10.1172/JCI90253

Levrero, M., & Zucman-Rossi, J. (n.d.). *Mechanisms of HBV-induced hepatocellular carcinoma*.

Li, X., Fang, P., Mai, J., Choi, E. T., Wang, H., & Yang, X. F. (2013). Targeting mitochondrial reactive oxygen species as novel therapy for inflammatory diseases and cancers. In *Journal of Hematology and Oncology* (Vol. 6, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/1756-8722-6-19

Lian, Y.-F., Li, S.-S., Huang, Y.-L., Wei, H., Chen, D.-M., Wang, J.-L., & Huang, Y.-H. (2018). Up-regulated and interrelated expressions of GINS subunits predict poor prognosis in hepatocellular carcinoma. *Bioscience Reports*, *38*(6). https://doi.org/10.1042/BSR20181178

Liu, C., Wang, R., & Zhang, Y. (2019). GINS complex subunit 2 (GINS2) plays a protective role in alcohol-induced brain injury. *Artificial Cells, Nanomedicine, and Biotechnology*, *47*(1). https://doi.org/10.1080/21691401.2018.1540425

Liu, H., Pathak, P., Boehme, S., & Chiang, JohnY. L. (2016). Cholesterol 7α-hydroxylase protects the liver from inflammation and fibrosis by maintaining cholesterol homeostasis. *Journal of Lipid Research, 57*(10). https://doi.org/10.1194/jlr.M069807

Liyanage, D. S., Omeka, W. K. M., Yang, H., Godahewa, G. I., Kwon, H., Nam, B.-H., & Lee, J. (2019). Identification of thioredoxin domain-containing protein 17 from big-belly seahorse Hippocampus abdominalis: Molecular insights, immune responses, and functional characterization. *Fish & Shellfish Immunology, 86*. https://doi.org/10.1016/j.fsi.2018.11.040

Longmore, G. D., Keegan, A. D., Zamorano, J., Wang, H. Y., Wang, R., & Shi, Y. (1998). *Cell Cycle Progression STAT5 in Protection from Apoptosis But Not in Regulation of Cell Growth by IL-2: Role of.* http://www.jimmunol.org/content/160/7/3502

Mani, S., & Andrisani, O. (2018). Hepatitis B Virus-Associated Hepatocellular Carcinoma and Hepatic Cancer Stem Cells. *Genes, 9*(3). https://doi.org/10.3390/genes9030137

Migita, T., Okabe, S., Ikeda, K., Igarashi, S., Sugawara, S., Tomida, A., Taguchi, R., Soga, T., & Seimiya, H. (2013). Inhibition of ATP Citrate Lyase Induces an Anticancer Effect via Reactive Oxygen Species. *The American Journal of Pathology, 182*(5). https://doi.org/10.1016/j.ajpath.2013.01.048

Milisav, I. (n.d.). *Cellular Stress Responses.* www.intechopen.com

Monteiro, Â., Gonçalves, L. M., & Santos, M. M. M. (2014). Synthesis of novel spiropyrazoline oxindoles and evaluation of cytotoxicity in cancer cell lines. *European Journal of Medicinal Chemistry, 79*. https://doi.org/10.1016/j.ejmech.2014.04.023

Murphy, M. P. (2009). How mitochondria produce reactive oxygen species. In *Biochemical Journal* (Vol. 417, Issue 1, pp. 1–13). https://doi.org/10.1042/BJ20081386

Nacu, Ş., Critchley-Thorne, R., Lee, P., & Holmes, S. (2007). Gene expression network analysis and applications to immunology. *Bioinformatics, 23*(7), 850–858. https://doi.org/10.1093/bioinformatics/btm019

Nag, S., Qin, J., Srivenugopal, K. S., Wang, M., & Zhang, R. (2013). The MDM2-p53 pathway revisited. In *Journal of Biomedical Research* (Vol. 27, Issue 4, pp. 254–271). Nanjing Medical University. https://doi.org/10.7555/JBR.27.20130030

Nishida, N., Chishina, H., Arizumi, T., Takita, M., Kitai, S., Yada, N., Hagiwara, S., Inoue, T., Minami, Y., Ueshima, K., Sakurai, T., & Kudo, M. (2014). Identification of Epigenetically Inactivated Genes in Human Hepatocellular Carcinoma by Integrative Analyses of Methylation Profiling and Pharmacological Unmasking. *Digestive Diseases*, *32*(6). https://doi.org/10.1159/000368015

Nwosu, Z. C., Megger, D. A., Hammad, S., Sitek, B., Roessler, S., Ebert, M. P., Meyer, C., & Dooley, S. (2017). Identification of the Consistently Altered Metabolic Targets in Human Hepatocellular Carcinoma. *Cellular and Molecular Gastroenterology and Hepatology*, *4*(2). https://doi.org/10.1016/j.jcmgh.2017.05.004

Pantic, B., Trevisan, E., Citta, A., Rigobello, M. P., Marin, O., Bernardi, P., Salvatori, S., & Rasola, A. (2013). Myotonic dystrophy protein kinase (DMPK) prevents ROS-induced cell death by assembling a hexokinase II-Src complex on the mitochondrial surface. *Cell Death & Disease*, *4*(10). https://doi.org/10.1038/cddis.2013.385

Papp, L. V., Lu, J., Holmgren, A., & Khanna, K. K. (2007). From Selenium to Selenoproteins: Synthesis, Identity, and Their Role in Human Health. *Antioxidants & Redox Signaling*, *9*(7). https://doi.org/10.1089/ars.2007.1528

Pizzino, G., Irrera, N., Cucinotta, M., Pallio, G., Mannino, F., Arcoraci, V., Squadrito, F., Altavilla, D., & Bitto, A. (2017). Oxidative Stress: Harms and Benefits for Human Health. In *Oxidative Medicine and Cellular Longevity* (Vol. 2017). Hindawi Limited. https://doi.org/10.1155/2017/8416763

Pope, E. D., Kimbrough, E. O., Vemireddy, L. P., Surapaneni, P. K., Copland, J. A., & Mody, K. (2019). Aberrant lipid metabolism as a therapeutic target in liver cancer. *Expert Opinion on Therapeutic Targets*, *23*(6). https://doi.org/10.1080/14728222.2019.1615883

Purvis, J. E., Karhohs, K. W., Mock, C., Batchelor, E., Loewer, A., & Lahav, G. (2012). p53 Dynamics Control Cell Fate. *Science*, *336*(6087). https://doi.org/10.1126/science.1218351

Rahman, F., Mahmud, P., Karim, R., Hossain, T., & Islam, F. (2020). Determination of novel biomarkers and pathways shared by colorectal cancer and endometrial cancer via comprehensive bioinformatics analysis. *Informatics in Medicine Unlocked*, *20*. https://doi.org/10.1016/j.imu.2020.100376

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*(7). https://doi.org/10.1093/nar/gkv007

S. Darvesh, A., & Bishayee, A. (2010). Selenium in the Prevention and Treatment of Hepatocellular Carcinoma. *Anti-Cancer Agents in Medicinal Chemistry*, *10*(4). https://doi.org/10.2174/187152010791162252

Sarret, C., Ashkavand, Z., Paules, E., Dorboz, I., Pediaditakis, P., Sumner, S., Eymard-Pierre, E., Francannet, C., Krupenko, N. I., Boespflug-Tanguy, O., & Krupenko, S. A. (2019). Deleterious mutations in ALDH1L2 suggest a novel cause for neuro-ichthyotic syndrome. *Npj Genomic Medicine*, *4*(1). https://doi.org/10.1038/s41525-019-0092-9

Sengupta, S., Peterson, T. R., & Sabatini, D. M. (2010). Regulation of the mTOR Complex 1 Pathway by Nutrients, Growth Factors, and Stress. *Molecular Cell*, *40*(2). https://doi.org/10.1016/j.molcel.2010.09.026

Shannon, P. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, *13*(11). https://doi.org/10.1101/gr.1239303

Si, M., Zhao, C., Zhang, B., Wei, D., Chen, K., Yang, X., Xiao, H., & Shen, X. (2016). Overexpression of Mycothiol Disulfide Reductase Enhances Corynebacterium glutamicum Robustness by Modulating Cellular Redox Homeostasis and Antioxidant Proteins under Oxidative Stress. *Scientific Reports*, *6*(1). https://doi.org/10.1038/srep29491

Siegel, R., DeSantis, C., Virgo, K., Stein, K., Mariotto, A., Smith, T., Cooper, D., Gansler, T., Lerro, C., Fedewa, S., Lin, C., Leach, C., Cannady, R. S., Cho, H., Scoppa, S., Hachey, M., Kirch, R., Jemal, A., & Ward, E. (2012). Cancer treatment and survivorship statistics, 2012. *CA: A Cancer Journal for Clinicians*, *62*(4). https://doi.org/10.3322/caac.21149

Slonim, D. K., & Yanai, I. (2009). Getting started in gene expression microarray analysis. In *PLoS Computational Biology* (Vol. 5, Issue 10). https://doi.org/10.1371/journal.pcbi.1000543

Song, D., Liu, G., Luu-The, V., Zhao, D., Wang, L., Zhang, H., Xueling, G., Li, S., Désy, L., Labrie, F., & Pelletier, G. (2006). Expression of aromatase and 17β-hydroxysteroid dehydrogenase types 1, 7 and 12 in breast cancer. *The Journal of Steroid Biochemistry and Molecular Biology*, *101*(2–3). https://doi.org/10.1016/j.jsbmb.2006.06.015

Song, L., Wei, X., Zhang, B., Luo, X., liu, J., Feng, Y., & Xiao, X. (2009). Role of Foxa1 in regulation of bcl2 expression during oxidative-stress-induced apoptosis in A549 type II pneumocytes. *Cell Stress and Chaperones*, *14*(4). https://doi.org/10.1007/s12192-008-0095-4

Strungaru, M. H., Footz, T., Liu, Y., Berry, F. B., Belleau, P., Semina, E. v., Raymond, V., & Walter, M. A. (2011). PITX2 Is Involved in Stress Response in Cultured Human Trabecular Meshwork Cells through Regulation of SLC13A3. *Investigative Opthalmology & Visual Science*, *52*(10). https://doi.org/10.1167/iovs.10-6967

Subramaniam, S., Kelley, R. K., & Venook, A. P. (2013). A review of hepatocellular carcinoma (HCC) staging systems. In *Chinese Clinical Oncology* (Vol. 2, Issue 4). AME Publishing Company. https://doi.org/10.3978/j.issn.2304-3865.2013.07.05

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

Swetzig, W. M., Wang, J., & Das, G. M. (2016). Estrogen receptor alpha (ERα/ESR1) mediates the p53-independent overexpression of MDM4/MDMX and MDM2 in human breast cancer. *Oncotarget*, *7*(13), 16049–16069. https://doi.org/10.18632/oncotarget.7533

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., & Mering, C. von. (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, *47*(D1). https://doi.org/10.1093/nar/gky1131

Takatori, H., Yamashita, T., Honda, M., Nishino, R., Arai, K., Yamashita, T., Takamura, H., Ohta, T., Zen, Y., & Kaneko, S. (2010). dUTP pyrophosphatase expression correlates with a poor prognosis in hepatocellular carcinoma. *Liver International*, *30*(3). https://doi.org/10.1111/j.1478-3231.2009.02177.x

Tan, J., Duan, M., Yadav, T., Phoon, L., Wang, X., Zhang, J.-M., Zou, L., & Lan, L. (2020). An R-loop-initiated CSB–RAD52–POLD3 pathway suppresses ROS-induced telomeric DNA breaks. *Nucleic Acids Research*, *48*(3). https://doi.org/10.1093/nar/gkz1114

Tuncbag, N., McCallum, S., Huang, S. -s. C., & Fraenkel, E. (2012). SteinerNet: a web server for integrating "omic" data to discover hidden components of response pathways. *Nucleic Acids Research*, *40*(W1). https://doi.org/10.1093/nar/gks445

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). https://doi.org/10.1038/srep28668

Wang, J.-H., Urrutia-Cabrera, D., Mora, S. M., Nguyen, T., Hung, S., Hewitt, A., Edwards, T., & Wong, R. (2020). *Functional study of the AMD-associated gene TMEM97 in retinal pigmented epithelium using CRISPR interference*. https://doi.org/10.1101/2020.07.10.198143

Wang, Y., Yang, L., Chen, T., Liu, X., Guo, Y., Zhu, Q., Tong, X., Yang, W., Xu, Q., Huang, D., & Tu, K. (2019). A novel lncRNA MCM3AP-AS1 promotes the growth of hepatocellular carcinoma by targeting miR-194-5p/FOXA1 axis. *Molecular Cancer*, *18*(1). https://doi.org/10.1186/s12943-019-0957-7

Waris, G., & Ahsan, H. (2006). Reactive oxygen species: role in the development of cancer and various chronic conditions. *Journal of Carcinogenesis*, *5*(1). https://doi.org/10.1186/1477-3163-5-14

Wilson, C. L., & Miller, C. J. (2005). Simpleaffy: a BioConductor package for Affymetrix Quality Control and data analysis. *Bioinformatics*, *21*(18). https://doi.org/10.1093/bioinformatics/bti605

Woo, H. G., Wang, X. W., Budhu, A., Kim, Y. H., Kwon, S. M., Tang, Z., Sun, Z., Harris, C. C., & Thorgeirsson, S. S. (2011). Association of TP53 mutations with stem cell-like gene expression and survival of patients with hepatocellular carcinoma. *Gastroenterology*, *140*(3), 1063-1070.e8. https://doi.org/10.1053/j.gastro.2010.11.034

Wu, Y., Zhang, J., Zhang, X., Zhou, H., Liu, G., & Li, Q. (2020). Cancer Stem Cells: A Potential Breakthrough in HCC-Targeted Therapy. *Frontiers in Pharmacology*, *11*. https://doi.org/10.3389/fphar.2020.00198

Xie, M., Yang, Z., Liu, Y., & Zheng, M. (2018). The role of HBV-induced autophagy in HBV replication and HBV related-HCC. In *Life Sciences* (Vol. 205, pp. 107–112). Elsevier Inc. https://doi.org/10.1016/j.lfs.2018.04.051

Xie, S., Jiang, X., Zhang, J., Xie, S., Hua, Y., Wang, R., & Yang, Y. (2019). Identification of significant gene and pathways involved in HBV-related hepatocellular carcinoma by bioinformatics analysis. *PeerJ*, *2019*(7). https://doi.org/10.7717/peerj.7408

Xie, Y. (2017). *Hepatitis B Virus-Associated Hepatocellular Carcinoma*. https://doi.org/10.1007/978-981-10-5765-6_2

Yan, S.-Y., Fan, J.-G., & Qio, L. (2017). Hepatitis B Virus (HBV) Infection and Hepatocellular Carcinoma- New Insights for an Old Topic. *Current Cancer Drug Targets*, *17*(6). https://doi.org/10.2174/1568009616666160926124530

Yang, J. D., Hainaut, P., Gores, G. J., Amadou, A., Plymoth, A., & Roberts, L. R. (2019). A global view of hepatocellular carcinoma: trends, risk, prevention and management. In *Nature Reviews Gastroenterology and Hepatology* (Vol. 16, Issue 10, pp. 589–604). Nature Publishing Group. https://doi.org/10.1038/s41575-019-0186-y

Yu, G. I., Mun, K. H., Yang, S. H., Shin, D. H., & Hwang, J. S. (2018). Polymorphisms in the 3′-UTR of SCD5 gene are associated with hepatocellular carcinoma in Korean population. *Molecular Biology Reports*, *45*(6). https://doi.org/10.1007/s11033-018-4313-6

Yu, H., Kim, P. M., Sprecher, E., Trifonov, V., & Gerstein, M. (2007). The Importance of Bottlenecks in Protein Networks: Correlation with Gene Essentiality and Expression Dynamics. *PLoS Computational Biology*, *3*(4). https://doi.org/10.1371/journal.pcbi.0030059

Yu, M.-W., Horng, I.-S., Hsu, K.-H., Chiang, Y.-C., Liaw, Y. F., & Chen, C.-J. (1999). Plasma Selenium Levels and Risk of Hepatocellular Carcinoma among Men with Chronic Hepatitis Virus Infection. *American Journal of Epidemiology*, *150*(4). https://doi.org/10.1093/oxfordjournals.aje.a010016

Yu Yu, S., Zhu, Y. J., & Li, W. G. (1997). Protective role of selenium against hepatitis B virus and primary liver cancer in Qidong. *Biological Trace Element Research*, *56*(1). https://doi.org/10.1007/BF02778987

Zeng, Z., Cao, Z., & Tang, Y. (2020). Increased E2F2 predicts poor prognosis in patients with HCC based on TCGA data. *BMC Cancer*, *20*(1). https://doi.org/10.1186/s12885-020-07529-2

Zhang, H., Dong, Y., Zhao, H., Brooks, J. D., Hawthorn, L., Nowak, N., Marshall, J. R., Gao, A. C., & Ip, C. (n.d.). *Microarray Data Mining for Potential Selenium Targets in Chemoprevention of Prostate Cancer*. http://www.ncbi.nlm.nih.gov/

Zhang, Z., Bi, M., Liu, Q., Yang, J., & Xu, S. (2016). Meta-analysis of the correlation between selenium and incidence of hepatocellular carcinoma. *Oncotarget*, *7*(47). https://doi.org/10.18632/oncotarget.12804

Zhou, J., & Chng, W.-J. (2013). Roles of thioredoxin binding protein (TXNIP) in oxidative stress, apoptosis and cancer. *Mitochondrion*, *13*(3). https://doi.org/10.1016/j.mito.2012.06.004

Zuo, Y., Cui, Y., di Poto, C., Varghese, R. S., Yu, G., Li, R., & Ressom, H. W. (2016). INDEED: Integrated differential expression and differential network analysis of omic data for biomarker discovery. *Methods*, *111*, 12–20. https://doi.org/10.1016/j.ymeth.2016.08.015

# APPENDICES

## APPENDIX A

### SCRIPTS

Table 11: R programming script to read and transform raw probe intensities to expression values.

```
pData <- read.table(DataFile,row.names=1, header=TRUE, sep="\t")
library("affy")
Data <- ReadAffy(phenoData=pData)
write.exprs(Data, file="raw_data")
eset<- expresso(Data)
```

Table 12: R programming script used in limma analysis to construct design and contrast matrix and determine DEGs.

```
library("limma")
data=read.table('MyDATA',header=T,row.names = 1)
pdata=read.table('phenodata',header=T,sep='\t')
data=data[,match(pdata$celfile.name, colnames(data))]
day=as.numeric(pdata$Days)
sc=factor(paste(pdata$Selenium,pdata$Cells, pdata$Days, sep='.'))
design<-model.matrix(~0+sc)
ContMatrix<-makeContrasts(G2vs2215_P_D1=scSelP.G2.D1-scSelP.G2215.D1,
            G2vs2215_P_D2=scSelP.G2.D2-scSelP.G2215.D2,
            G2vs2215_P_D3=scSelP.G2.D3-scSelP.G2215.D3,
            G2vs2215_N_D1=scSelN.G2.D1-scSelN.G2215.D1,
            G2vs2215_N_D2=scSelN.G2.D2-scSelN.G2215.D2,
            G2vs2215_N_D3=scSelN.G2.D3-scSelN.G2215.D3,
            G2_PvsN_D1=scSelN.G2.D1-scSelP.G2.D1,
            G2_PvsN_D2=scSelN.G2.D2-scSelP.G2.D2,
            G2_PvsN_D3=scSelN.G2.D3-scSelP.G2.D3,
```

Table 11: R programming script to read and transform raw probe intensities to expression values. ( continued )

```
                2215_PvsN_D1=scSelN.2215.D1-scSelP.2215.D1,
                2215_PvsN_D2=scSelN.2215.D2-scSelP.2215.D2,
                2215_PvsN_D1=scSelN.2215.D1-scSelP.2215.D1,levels=design)
fit<-lmFit(data,design)
fit2<-contrasts.fit(fit,ContMatrix)
fit2<-eBayes(fit2)
```

Table 13: R programming script for heatmaply function to draw dendograms.

```
library(heatmaply)
#for html file
heatmaply(matrix,
margins = c(10, 400, 1, 1),
dedogram= "raw",
scale_fill_gradient_fun = scale_fill_gradient2(midpoint = 0.5, low = "green", mid="white", high = "red"),
k_col = 2, grid_gap = 0.1,ylab= "Gene",
fontsize_row = 5,dendrogram = TRUE,
fontsize_col = 8,
showticklabels = c(TRUE, FALSE),
file= "Heatmap.html")
```

Table 14: Script used to run the Forest algorithm.

```
python /home/knarci/OmicsIntegrator/scripts/forest.py -p Diff_G2vs2215_daydependent_.txt -
e /home//home/knarci/msgsteiner-1.3/msgsteiner -c /home/knarci/damla/configs/config_Diff
_G2vs2215_daydependent_w20.0b16.0mu0.2.txt
```

Table 15: Script used to run the PCST algorithm.

```
~/damla/configs/config_Diff_G2vs2215_daydependent_w12.0b30.0mu0.2.txt  .
nohup python ../forest-tuner.py --forestPath /home/knarci/OmicsIntegrator/scripts/
forest.py --msgsteinerPath /home/knarci/msgsteiner-1.3/msgsteiner --prizePath Omics_
OnlyN_D2.txt --edgePath /home/knarci/forest_interaction_alper2.txt -w 2,20,2 -b 2,30,2 -m 0.2
--minNodes 10 --outputsDirName OnlyN_D2 --dataPath OnlyN_D2.tsv --processes 8 &
```

# APPENDIX B

## COMPLETE DEG LIST

Table 16: All the DEGs identified in this study by the indicated analysis and comparison methods.

| Gene Name | Effect | Comparison | Analysis |
|-----------|--------|------------|----------|
| FOXA1 | Se | BCL | PCST |
| ONECUT1 | Se | BCL | PCST |
| CYP7A1 | Se | BCL | PCST |
| PITX2 | Se | BCL | PCST |
| ABCG5 | Se | BCL | PCST |
| FOXE3 | Se | BCL | PCST |
| ABCC6 | Se | BCL | PCST |
| C4BPB | Se | BCL | PCST |
| C4B | Se | BCL | PCST |
| PLOD1 | Se | BCL | PCST |
| SYTL4 | Se | BCL | PCST |
| XYLT2 | Se | BCL | PCST |
| MASP1 | Se | BCL | PCST |
| RAB27B | Se | BCL | PCST |
| SETD1B | Se | BCL | PCST |
| SFTPD | Se | BCL | PCST |
| SLC6A20 | Se | BCL | PCST |
| TBX20 | Se | BCL | PCST |
| ACLY | Se | BCL | PCST |
| TXNRD1 | Se | BCL | PCST |
| S100B | Se | BCL | PCST |
| SCD5 | Se | BCL | PCST |
| TXNDC17 | Se | BCL | PCST |
| ALDH1L2 | Se | BCL | PCST |
| S100A6 | Se | BCL | PCST |
| MTR | Se | BCL | PCST |

Table 16: All the DEGs identified in this study by the indicated analysis and comparison methods. ( continued )

| ELOVL6 | Se | BCL | PCST |
|---|---|---|---|
| ACOX1 | Se | BCL | PCST |
| KPNA2 | Se | BCL | PCST |
| STK38L | Se | BCL | PCST |
| PPARGC1A | Se | BCL | PCST |
| ERVW-1 | Se | BCL | PCST |
| TAT | Se | BCL | PCST |
| KIFC1 | Se | BCL | PCST |
| ANXA2 | Se | BCL | PCST |
| IMMT | Se | BCL | PCST |
| TECR | Se | BCL | PCST |
| ANXA11 | Se | BCL | PCST |
| CD81 | Se | BCL | PCST |
| LCE2A | Se | BCL | PCST |
| PDCD6 | Se | BCL | PCST |
| TNFRSF14 | Se | BCL | PCST |
| WDR1 | Se | BCL | PCST |
| FRYL | Se | BCL | PCST |
| UBE2K | Se | BCL | PCST |
| CNBP | Se | BCL | PCST |
| LSM4 | Se | BCL | PCST |
| QDPR | Se | BCL | PCST |
| DMPK | Se | BCL | PCST |
| TOE1 | Se | BCL | PCST |
| WDR77 | Se | BCL | PCST |
| MBNL1 | Se | BCL | PCST |
| CA2 | Se | BCL | PCST |
| SLC19A1 | Se | BCL | PCST |
| HOXB9 | Se | BCL | PCST |
| INTS12 | Se | BCL | PCST |
| DUT | Se | WCL | GSEA |
| POLD3 | Se | WCL | GSEA |
| RFC5 | Se | WCL | GSEA |
| PRIM1 | Se | WCL | GSEA |
| LIG1 | Se | WCL | GSEA |
| POLA2 | Se | WCL | GSEA |
| DUT | Se | WCL | GSEA |
| E2F8 | Se | WCL | GSEA |

Table 16: All the DEGs identified in this study by the indicated analysis and comparison methods. ( continued )

| CENPM | Se | WCL | GSEA |
|---|---|---|---|
| POLD3 | Se | WCL | GSEA |
| TIMELESS | Se | WCL | GSEA |
| RAD51AP1 | Se | WCL | GSEA |
| UBR7 | Se | WCL | GSEA |
| CDCA3 | Se | WCL | GSEA |
| ASF1B | Se | WCL | GSEA |
| MCM2 | Se | WCL | GSEA |
| ESPL1 | Se | WCL | GSEA |
| UNG | Se | WCL | GSEA |
| CCNE1 | Se | WCL | GSEA |
| MCM4 | Se | WCL | GSEA |
| CHEK2 | Se | WCL | GSEA |
| SHMT1 | Se | WCL | GSEA |
| TCF19 | Se | WCL | GSEA |
| WDR90 | Se | WCL | GSEA |
| MCM6 | Se | WCL | GSEA |
| GINS1 | Se | WCL | GSEA |
| TACC3 | Se | WCL | GSEA |
| LIG1 | Se | WCL | GSEA |
| CDC25B | Se | WCL | GSEA |
| POLA2 | Se | WCL | GSEA |
| PRIM2 | Se | WCL | GSEA |
| RAD51C | Se | WCL | GSEA |
| MCM3 | Se | WCL | GSEA |
| SPC25 | Se | WCL | GSEA |
| MYBL2 | Se | WCL | GSEA |
| KIF18B | Se | WCL | GSEA |
| SUV39H1 | Se | WCL | GSEA |
| CDKN2C | Se | WCL | GSEA |
| RNASEH2A | Se | WCL | GSEA |
| CHEK1 | Se | WCL | GSEA |
| STMN1 | Se | WCL | GSEA |
| HMMR | Se | WCL | GSEA |
| RFC2 | Se | WCL | GSEA |
| DLGAP5 | Se | WCL | GSEA |
| SLBP | Se | WCL | GSEA |
| RRM2 | Se | WCL | GSEA |

Table 16: All the DEGs identified in this study by the indicated analysis and comparison methods. ( continued )

| BUB1B | Se | WCL | GSEA |
|---|---|---|---|
| DNMT1 | Se | WCL | GSEA |
| BRCA1 | Se | WCL | GSEA |
| FOS | Se | WCL | GSEA |
| IMPA2 | Se | WCL | GSEA |
| SLC27A2 | Se | WCL | GSEA |
| GINS2 | Se | WCL | GSEA |
| FDFT1 | Se | WCL | GSEA |
| RAB31 | Se | WCL | GSEA |
| SERPINA5 | Se | WCL | GSEA |
| DHRS2 | Se | WCL | GSEA |
| RNASEH2A | Se | WCL | GSEA |
| E2F2 | Se | WCL | GSEA |
| GINS2 | Se | WCL | GSEA |
| MCM2 | Se | WCL | GSEA |
| ESPL1 | Se | WCL | GSEA |
| EXO1 | Se | WCL | GSEA |
| MCM6 | Se | WCL | GSEA |
| E2F1 | Se | WCL | GSEA |
| TACC3 | Se | WCL | GSEA |
| KIF15 | Se | WCL | GSEA |
| CDC25B | Se | WCL | GSEA |
| POLA2 | Se | WCL | GSEA |
| CDC7 | Se | WCL | GSEA |
| PRIM2 | Se | WCL | GSEA |
| RAD54L | Se | WCL | GSEA |
| PBK | Se | WCL | GSEA |
| MCM3 | Se | WCL | GSEA |
| TRAIP | Se | WCL | GSEA |
| CENPA | Se | WCL | GSEA |
| CDC45 | Se | WCL | GSEA |
| MYBL2 | Se | WCL | GSEA |
| SUV39H1 | Se | WCL | GSEA |
| CDKN2C | Se | WCL | GSEA |
| CHEK1 | Se | WCL | GSEA |
| MT2A | Se | WCL | GSEA |
| STMN1 | Se | WCL | GSEA |
| HMMR | Se | WCL | GSEA |

Table 16: All the DEGs identified in this study by the indicated analysis and comparison
methods. ( continued )

| IDI1 | Se | WCL | GSEA |
|---|---|---|---|
| PIK3R3 | Se | WCL | GSEA |
| DHFR | Se | WCL | GSEA |
| TMEM97 | Se | WCL | GSEA |
| MCM2 | Se | WCL | GSEA |
| UNG | Se | WCL | GSEA |
| MCM4 | Se | WCL | GSEA |
| EBP | Se | WCL | GSEA |
| CYB5B | Se | WCL | GSEA |
| CCNE1 | Se | WCL | GSEA |
| CXCL10 | Se | WCL | GSEA |
| GSTO1 | Se | WCL | GSEA |
| SLC29A2 | Se | WCL | GSEA |
| MAP6 | Se | WCL | GSEA |
| GPX4 | Se | WCL | GSEA |
| IRF4 | Se | WCL | GSEA |
| TNFRSF18 | Se | WCL | GSEA |
| PPAP2A | Se | BCL | HEATMAP |
| H2BFXP | Se | BCL | HEATMAP |
| LINC00958 | Se | BCL | HEATMAP |
| CLYBL | Se | BCL | HEATMAP |
| HOXD1 | Se | BCL | HEATMAP |
| RNF217 | Se | BCL | HEATMAP |
| FAM20A | Se | BCL | HEATMAP |
| KCNG3 | Se | BCL | HEATMAP |
| LACC1 | Se | BCL | HEATMAP |
| CD9 | Se | BCL | HEATMAP |
| TMEM170B | Se | BCL | HEATMAP |
| PVRL3 | Se | BCL | HEATMAP |
| SMPDL3A | Se | BCL | HEATMAP |
| CHFR | Se | BCL | HEATMAP |
| RIN2 | Se | BCL | HEATMAP |
| PPARGC1A | Se | BCL | HEATMAP |
| PDLIM3 | Se | BCL | HEATMAP |
| Sep5 | Se | BCL | HEATMAP |
| PRAME | Se | BCL | HEATMAP |
| MCAM | Se | BCL | HEATMAP |
| MAGEA6 | Se | BCL | HEATMAP |

Table 16: All the DEGs identified in this study by the indicated analysis and comparison methods. ( continued )

| | | | |
|---|---|---|---|
| CYSTM1 | Se | BCL | HEATMAP |
| SLC2A2 | Se | BCL | HEATMAP |
| FAM198B | Se | BCL | HEATMAP |
| MBNL3 | Se | BCL | HEATMAP |
| ISM2 | Se | BCL | HEATMAP |
| SCPEP1 | Se | BCL | HEATMAP |
| MRPL52 | Se | BCL | HEATMAP |
| MPP6 | Se | BCL | HEATMAP |
| PM20D2 | Se | BCL | HEATMAP |
| ZC4H2 | Se | BCL | HEATMAP |
| SLC16A7 | Se | BCL | HEATMAP |
| TMEM27 | Se | BCL | HEATMAP |
| GPR19 | Se | BCL | HEATMAP |
| NDNL2 | Se | BCL | HEATMAP |
| CYP7A1 | Se | BCL | HEATMAP |
| GLCCI1 | Se | BCL | HEATMAP |
| ANXA2 | Se | BCL | HEATMAP |
| MAGEA3 | Se | BCL | HEATMAP |
| PCOLCE2 | Se | BCL | HEATMAP |
| IL15RA | Se | BCL | HEATMAP |
| FBXO33 | Se | BCL | HEATMAP |
| HECTD2 | Se | BCL | HEATMAP |
| MID1 | Se | BCL | HEATMAP |
| SLC13A4 | Se | BCL | HEATMAP |
| DNALI1 | Se | BCL | HEATMAP |
| ALDH1L2 | Se | BCL | HEATMAP |
| LY96 | Se | BCL | HEATMAP |
| HPX | Se | BCL | HEATMAP |
| INSIG2 | Se | BCL | HEATMAP |
| SMIM3 | Se | BCL | HEATMAP |
| FMR1 | Se | BCL | HEATMAP |
| LINC00889 | Se | BCL | HEATMAP |
| RING1 | Se | BCL | HEATMAP |
| SS18L1 | Se | BCL | HEATMAP |
| TUBBP5 | Se | BCL | HEATMAP |
| DIRAS1 | Se | BCL | HEATMAP |
| DHX33 | Se | BCL | HEATMAP |
| NRG1 | Se | BCL | HEATMAP |

Table 16: All the DEGs identified in this study by the indicated analysis and comparison methods. ( continued )

| RNF19A | Se | BCL | HEATMAP |
|---|---|---|---|
| FKBPL | Se | BCL | HEATMAP |
| CTNNAL1 | Se | BCL | HEATMAP |
| NOTCH1 | Se | BCL | HEATMAP |
| SMAD9 | Se | BCL | HEATMAP |
| LOC389831 | Se | BCL | HEATMAP |
| DIDO1 | Se | BCL | HEATMAP |
| AOC3 | Se | BCL | HEATMAP |
| HSPB1 | Se | BCL | HEATMAP |
| TNFRSF11A | Se | BCL | HEATMAP |
| C9orf40 | Se | BCL | HEATMAP |
| NEDD8 | Se | BCL | HEATMAP |
| SV2A | Se | BCL | HEATMAP |
| CRIM1 | Se | BCL | HEATMAP |
| GCA | Se | BCL | HEATMAP |
| KIFAP3 | Se | BCL | HEATMAP |
| MYRIP | Se | BCL | HEATMAP |
| CCND3 | Se | BCL | HEATMAP |
| DDAH2 | Se | BCL | HEATMAP |
| TM2D3 | Se | BCL | HEATMAP |
| WFDC21P | Se | BCL | HEATMAP |
| AVPI1 | Se | BCL | HEATMAP |
| MUC15 | Se | BCL | HEATMAP |
| ICAM2 | Se | BCL | HEATMAP |
| TXNRD1 | Se | BCL | HEATMAP |
| ARHGEF26 | Se | BCL | HEATMAP |
| ARL4A | Se | BCL | HEATMAP |
| LZTFL1 | Se | BCL | HEATMAP |
| MAP3K1 | Se | BCL | HEATMAP |
| ALCAM | Se | BCL | HEATMAP |
| STRA6 | Se | BCL | HEATMAP |
| FRMD8 | Se | BCL | HEATMAP |
| FZD5 | Se | BCL | HEATMAP |
| C7orf26 | Se | BCL | HEATMAP |
| ANTXR1 | Se | BCL | HEATMAP |
| CTSS | Se | BCL | HEATMAP |
| JADE2 | Se | BCL | HEATMAP |
| NHP2L1 | Se | BCL | HEATMAP |

Table 16: All the DEGs identified in this study by the indicated analysis and comparison methods. ( continued )

| ACTR10 | Se | BCL | HEATMAP |
|---|---|---|---|
| CAMK2N2 | Se | BCL | HEATMAP |
| IGF2 | Se | BCL | HEATMAP |
| SCML2 | Se | BCL | HEATMAP |
| CCDC91 | Se | BCL | HEATMAP |
| RAB30-AS1 | Se | BCL | HEATMAP |
| FBXL22 | Se | BCL | HEATMAP |
| CREBL2 | Se | BCL | HEATMAP |
| ELOF1 | Se | BCL | HEATMAP |
| C14orf28 | Se | BCL | HEATMAP |
| SMPX | Se | BCL | HEATMAP |
| RIMS3 | Se | BCL | HEATMAP |
| POP7 | Se | BCL | HEATMAP |
| PCMTD1 | Se | BCL | HEATMAP |
| SEZ6L2 | Se | BCL | HEATMAP |
| RMDN1 | Se | BCL | HEATMAP |
| SEPSECS | Se | BCL | HEATMAP |
| TRIQK | Se | BCL | HEATMAP |
| GMPR2 | Se | BCL | HEATMAP |
| TRAPPC8 | Se | BCL | HEATMAP |
| ACOX1 | Se | BCL | HEATMAP |
| MCTP2 | Se | BCL | HEATMAP |
| FAM184A | Se | BCL | HEATMAP |
| PAFAH1B3 | Se | BCL | HEATMAP |
| NVL | Se | BCL | HEATMAP |
| WDR89 | Se | BCL | HEATMAP |
| FAM65C | Se | BCL | HEATMAP |
| AIFM2 | Se | BCL | HEATMAP |
| CNTLN | Se | BCL | HEATMAP |
| C1orf85 | Se | BCL | HEATMAP |
| STK38L | Se | BCL | HEATMAP |
| GEM | Se | BCL | HEATMAP |
| AP1G2 | Se | BCL | HEATMAP |
| SIL1 | Se | BCL | HEATMAP |
| LRRN4 | Se | BCL | HEATMAP |
| TECR | Se | BCL | HEATMAP |
| DECR2 | Se | BCL | HEATMAP |
| PAPOLA | Se | BCL | HEATMAP |

Table 16: All the DEGs identified in this study by the indicated analysis and comparison methods. ( continued )

| ZNF605 | Se | BCL | HEATMAP |
|---|---|---|---|
| SLC25A27 | Se | BCL | HEATMAP |
| NME4 | Se | BCL | HEATMAP |
| PITX1 | Se | BCL | HEATMAP |
| BAIAP2 | Se | BCL | HEATMAP |
| EML1 | Se | BCL | HEATMAP |
| TPBG | Se | BCL | HEATMAP |
| DFNB59 | Se | BCL | HEATMAP |
| ATP2B2 | Se | BCL | HEATMAP |
| COPB1 | Se | BCL | HEATMAP |
| NCKAP5L | Se | BCL | HEATMAP |
| RNF157 | Se | BCL | HEATMAP |
| NAT14 | Se | BCL | HEATMAP |
| TGDS | Se | BCL | HEATMAP |
| PRKCSH | Se | BCL | HEATMAP |
| CAMK2D | Se | BCL | HEATMAP |
| CREBZF | Se | BCL | HEATMAP |
| SIX4 | Se | BCL | HEATMAP |
| ZNF608 | Se | BCL | HEATMAP |
| FOXA1 | Se | BCL | HEATMAP |
| MBL2 | Se | BCL | HEATMAP |
| WDR34 | Se | BCL | HEATMAP |
| LARP1 | Se | BCL | HEATMAP |
| FGF13 | HBV | BCL | HEATMAP |
| GPC3 | HBV | BCL | HEATMAP |
| MAP7D2 | HBV | BCL | HEATMAP |
| NR5A2 | HBV | BCL | HEATMAP |
| HDHD1 | HBV | BCL | HEATMAP |
| PLA2G16 | HBV | BCL | HEATMAP |
| GNG4 | HBV | BCL | HEATMAP |
| FST | HBV | BCL | HEATMAP |
| EPB41L3 | HBV | BCL | HEATMAP |
| WDR72 | HBV | BCL | HEATMAP |
| SETBP1 | HBV | BCL | HEATMAP |
| L3MBTL4 | HBV | BCL | HEATMAP |
| PAQR9 | HBV | BCL | HEATMAP |
| MSRB3 | HBV | BCL | HEATMAP |
| CD24 | HBV | BCL | HEATMAP |

Table 16: All the DEGs identified in this study by the indicated analysis and comparison methods. ( continued )

| CDH17 | HBV | BCL | HEATMAP |
|---|---|---|---|
| ADAMTS2 | HBV | BCL | HEATMAP |
| PLAC8 | HBV | BCL | HEATMAP |
| CPED1 | HBV | BCL | HEATMAP |
| FAM127A | HBV | BCL | HEATMAP |
| HOXA10 | HBV | BCL | HEATMAP |
| SPP1 | HBV | BCL | HEATMAP |
| FAM101B | HBV | BCL | HEATMAP |
| LGI1 | HBV | BCL | HEATMAP |
| FKBP11 | HBV | BCL | HEATMAP |
| ARL15 | HBV | BCL | HEATMAP |
| LOC400043 | HBV | BCL | HEATMAP |
| HABP2 | HBV | BCL | HEATMAP |
| KLHL14 | HBV | BCL | HEATMAP |
| FAM110C | HBV | BCL | HEATMAP |
| STXBP6 | HBV | BCL | HEATMAP |
| GPR137B | HBV | BCL | HEATMAP |
| DGKE | HBV | BCL | HEATMAP |
| CTBP2 | HBV | BCL | HEATMAP |
| VNN1 | HBV | BCL | HEATMAP |
| TMEM261 | HBV | BCL | HEATMAP |
| OTUD1 | HBV | BCL | HEATMAP |
| MAGI2-AS3 | HBV | BCL | HEATMAP |
| PARD3B | HBV | BCL | HEATMAP |
| PLP2 | HBV | BCL | HEATMAP |
| TUBB2B | HBV | BCL | HEATMAP |
| PDGFA | HBV | BCL | HEATMAP |
| DPYD | HBV | BCL | HEATMAP |
| SLCO1B3 | HBV | BCL | HEATMAP |
| CCL16 | HBV | BCL | HEATMAP |
| PRLR | HBV | BCL | HEATMAP |
| FOXN3 | HBV | BCL | HEATMAP |
| MEGF6 | HBV | BCL | HEATMAP |
| ABCG5 | HBV | BCL | HEATMAP |
| CTDSPL | HBV | BCL | HEATMAP |
| TLE4 | HBV | BCL | HEATMAP |
| ZDHHC14 | HBV | BCL | HEATMAP |
| RASGEF1A | HBV | BCL | HEATMAP |

Table 16: All the DEGs identified in this study by the indicated analysis and comparison methods. ( continued )

| FBXL21 | HBV | BCL | HEATMAP |
|---|---|---|---|
| HOXC9 | HBV | BCL | HEATMAP |
| ZAK | HBV | BCL | HEATMAP |
| AKR1C3 | HBV | BCL | HEATMAP |
| EMP2 | HBV | BCL | HEATMAP |
| SERPINI1 | HBV | BCL | HEATMAP |
| CYB561 | HBV | BCL | HEATMAP |
| BCL11A | HBV | BCL | HEATMAP |
| SLC35A4 | HBV | BCL | HEATMAP |
| TEAD2 | HBV | BCL | HEATMAP |
| IRX3 | HBV | BCL | HEATMAP |
| PAG1 | HBV | BCL | HEATMAP |
| LINC00162 | HBV | BCL | HEATMAP |
| PTPRJ | HBV | BCL | HEATMAP |
| RBM23 | HBV | BCL | HEATMAP |
| TMEM42 | HBV | BCL | HEATMAP |
| SEMA4F | HBV | BCL | HEATMAP |
| CHST9 | HBV | BCL | HEATMAP |
| NGFRAP1 | HBV | BCL | HEATMAP |
| BIN1 | HBV | BCL | HEATMAP |
| ATP10D | HBV | BCL | HEATMAP |
| B4GALT6 | HBV | BCL | HEATMAP |
| RBMS1 | HBV | BCL | HEATMAP |
| PRIMA1 | HBV | BCL | HEATMAP |
| RPS23 | HBV | BCL | HEATMAP |
| GDA | HBV | BCL | HEATMAP |
| TNIK | HBV | BCL | HEATMAP |
| PROM1 | HBV | BCL | HEATMAP |
| TBC1D16 | HBV | BCL | HEATMAP |
| GSTO2 | HBV | BCL | HEATMAP |
| SPINK1 | HBV | BCL | HEATMAP |
| SEMA3G | HBV | BCL | HEATMAP |
| SLC38A4 | HBV | BCL | HEATMAP |
| SLC41A1 | HBV | BCL | HEATMAP |
| RBM47 | HBV | BCL | HEATMAP |
| CD163 | HBV | BCL | HEATMAP |
| UGT2B4 | HBV | BCL | HEATMAP |
| CLDN11 | HBV | BCL | HEATMAP |

Table 16: All the DEGs identified in this study by the indicated analysis and comparison methods. ( continued )

| SREK1IP1 | HBV | BCL | HEATMAP |
|----------|-----|-----|---------|
| TSPAN5 | HBV | BCL | HEATMAP |
| LINC01234 | HBV | BCL | HEATMAP |
| IL6R | HBV | BCL | HEATMAP |
| NME3 | HBV | BCL | HEATMAP |
| MARCH3 | HBV | BCL | HEATMAP |
| PPP1R3D | HBV | BCL | HEATMAP |
| MYO10 | HBV | BCL | HEATMAP |
| PRTG | HBV | BCL | HEATMAP |
| ZNF703 | HBV | BCL | HEATMAP |
| CYP39A1 | HBV | BCL | HEATMAP |
| DEFB1 | HBV | BCL | HEATMAP |
| FAM13A | HBV | BCL | HEATMAP |
| C10orf10 | HBV | BCL | HEATMAP |
| BHMT2 | HBV | BCL | HEATMAP |
| DKK4 | HBV | BCL | HEATMAP |

**HBV+ OR HBV- HCC PATIENT SURVIVAL CURVES**
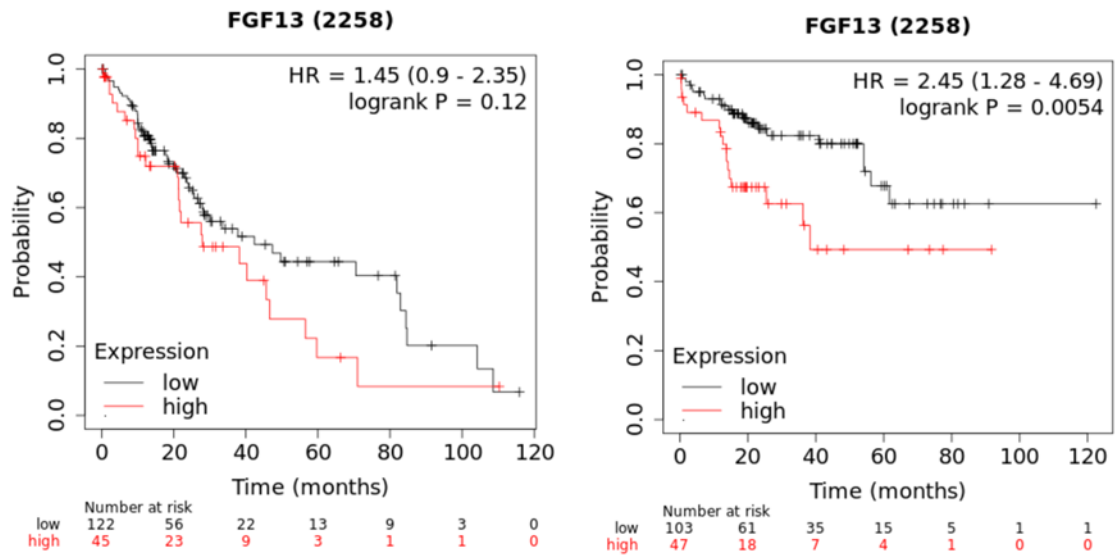


Figure 28: The Kaplan Meier plots that were generated for FGF13 gene with HBV- and HBV+ HCC patient cohorts in liver cancer RNA-seq dataset, respectively.

# APPENDIX D

## WU_HBX_TARGETS_3_UP GENE SET AND GSEA RESULT

Table 17:  The genes that are upregulated in SKHEP1 and primary hepatocytes composing WU_HBX_TARGETS_3_UP gene set in Molecular Signatures Database

|    | GENE SYMBOLS |
|----|--------------|
| 1  | CASP4        |
| 2  | CCNI         |
| 3  | CDK4         |
| 4  | CDKN3        |
| 5  | DAD1         |
| 6  | FAS          |
| 7  | GSTM2        |
| 8  | IFITM1       |
| 9  | IFNAR2       |
| 10 | MFNG         |
| 11 | MYB          |
| 12 | MYC          |
| 13 | PDCD2        |
| 14 | SLC2A5       |
| 15 | TFAP4        |
| 16 | TWF1         |
| 17 | TYMS         |
| 18 | VCL          |

| | |
|---|---|
| Dataset | rma normalized data_GSEA_collapsed_to_symbols.2215P_vs_G2P.cls #2215P_versus_G2P.2215P_vs_G2P.cls #2215P_versus_G2P_repos |
| Phenotype | 2215P_vs_G2P.cls#2215P_versus_G2P_repos |
| Upregulated in class | 2215P |
| GeneSet | WU_HBX_TARGETS_3_UP |
| Enrichment Score (ES) | 0.57088405 |
| Normalized Enrichment Score (NES) | 1.3499917 |
| Nominal p-value | 0.11741683 |
| FDR q-value | 0.1232687 |
| FWER p-Value | 0.138 |
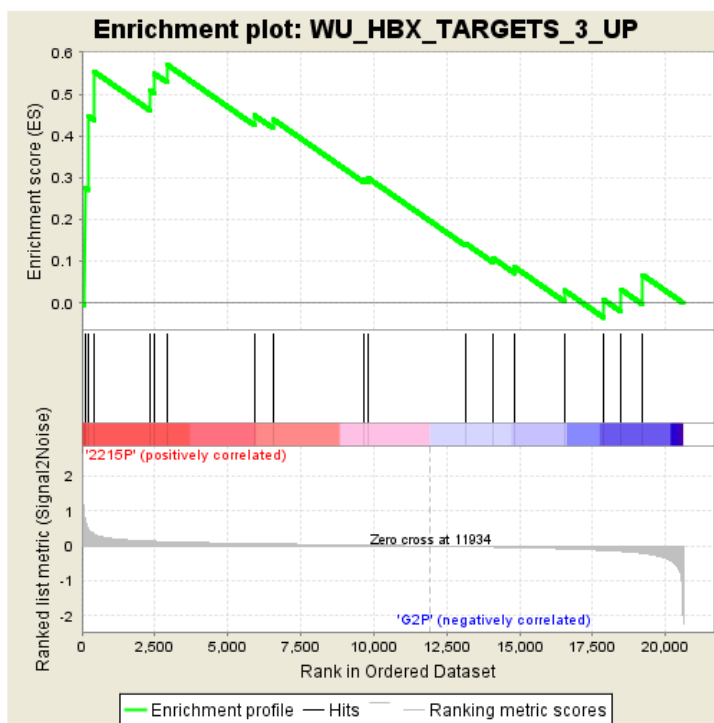


Figure 29: GSEA result of WU_HBX_TARGETS_3_UP gene set in HepG2-2.2.15 vs HepG2 cells grown in presence of Se.

93

# APPENDIX E

# THE COPYRIGHT PERMISSONS

A

## CCC Marketplace™

### EUREKA SCIENCE (FZC) - License Terms and Conditions

This is a License Agreement between Damla Gozen ("You") and EUREKA SCIENCE (FZC) ("Publisher") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by EUREKA SCIENCE (FZC), and the CCC terms and conditions.

All payments must be made in full to CCC.

| | | | |
|---|---|---|---|
| **Order Date** | 07-Jun-2021 | **Type of Use** | Republish in a thesis/dissertation |
| **Order License ID** | 1124025-1 | **Publisher** | BENTHAM SCIENCE PUBLISHERS LTD. |
| **ISSN** | 1568-0096 | **Portion** | Chart/graph/table/figure |

### LICENSED CONTENT

| | | | |
|---|---|---|---|
| **Publication Title** | CURRENT CANCER DRUG TARGETS | **Country** | Netherlands |
| **Date** | 01/01/2001 | **Rightsholder** | EUREKA SCIENCE (FZC) |
| **Language** | English, English | **Publication Type** | Journal |

### REQUEST DETAILS

| | | | |
|---|---|---|---|
| **Portion Type** | Chart/graph/table/figure | **Distribution** | Worldwide |
| **Number of charts / graphs / tables / figures requested** | 1 | **Translation** | Original language of publication |
| **Format (select all that apply)** | Print, Electronic | **Copies for the disabled?** | No |
| **Who will republish the content?** | Academic institution | **Minor editing privileges?** | No |
| **Duration of Use** | Life of current edition | **Incidental promotional use?** | No |
| **Lifetime Unit Quantity** | Up to 499 | **Currency** | EUR |
| **Rights Requested** | Main product | | |

### NEW WORK DETAILS

| | | | |
|---|---|---|---|
| **Title** | Identification of cellular stress related biomolecules for eventual use in targeted therapies of Hepatocellular carcinoma | **Institution name** | Middle East Technical University |
| **Instructor name** | Rengul Cetin-Atalay | **Expected presentation date** | 2021-07-29 |

### ADDITIONAL DETAILS

| | | | |
|---|---|---|---|
| **Order reference number** | N/A | **The requesting person / organization to appear on the license** | Damla Gozen |

94

**B**

## Reactive oxygen species: role in the development of cancer and various chronic conditions

Gulam Waris[1] and Haseeb Ahsan[2]

▸ Author information  ▸ Article notes  ▾ Copyright and License information  Disclaimer

**C**

**D**

# Selenium, Selenoproteins and Viral Infection

by Olivia M. Guillin [1,2,3,4,5], Caroline Vindry [1,2,3,4,5], Théophile Ohlmann [1,2,3,4,5] and Laurent Chavatte [1,2,3,4,5,*]

[1] CIRI, Centre International de Recherche en Infectiologie, CIRI, 69007 Lyon, France
[2] Institut National de la Santé et de la Recherche Médicale (INSERM) Unité U1111, 69007 Lyon, France
[3] Ecole Normale Supérieure de Lyon, 69007 Lyon, France
[4] Université Claude Bernard Lyon 1 (UCBL1), 69622 Lyon, France
[5] Unité Mixte de Recherche 5308 (UMR5308), Centre national de la recherche scientifique (CNRS), 69007 Lyon, France
* Author to whom correspondence should be addressed.

View Full-Text    Download PDF    Browse Figures    Citation Export

## Abstract

Reactive oxygen species (ROS) are frequently produced during viral infections. Generation of these ROS can be both beneficial and detrimental for many cellular functions. When overwhelming the antioxidant defense system, the excess of ROS induces oxidative stress. Viral infections lead to diseases characterized by a broad spectrum of clinical symptoms, with oxidative stress being one of their hallmarks. In many cases, ROS can, in turn, enhance viral replication leading to an amplification loop. Another important parameter for viral replication and pathogenicity is the nutritional status of the host. Viral infection simultaneously increases the demand for micronutrients and causes their loss, which leads to a deficiency that can be compensated by micronutrient supplementation. Among the nutrients implicated in viral infection, selenium (Se) has an important role in antioxidant defense, redox signaling and redox homeostasis. Most of biological activities of selenium is performed through its incorporation as a rare amino acid selenocysteine in the essential family of selenoproteins. Selenium deficiency, which is the main regulator of selenoprotein expression, has been associated with the pathogenicity of several viruses. In addition, several selenoprotein members, including glutathione peroxidases (GPX), thioredoxin reductases (TXNRD) seemed important in different models of viral replication. Finally, the formal identification of viral selenoproteins in the genome of molluscum contagiosum and fowlpox viruses demonstrated the importance of selenoproteins in viral cycle. View Full-Text

Keywords: reactive oxygen species; glutathione peroxidases; thioredoxin reductases; influenza virus; hepatitis C virus; coxsackie virus; human immunodeficiency virus; molluscum contagiosum virus; viral selenoproteins; immunity

▼ Show Figures



Graphical abstract

**E**

Figure 30: The copyright permissions taken for Figures (A) 1, (B) 2, (C) 3, (D) 5, (E) 6 respectively.

Sayın Damla Gözen,
Uygun şekilde atıf verildiği müddetçe sıkıntı yok, ayrıca bir doküman göndermemiz
gerekmiyor.
Saygılarımla

Bedia BAL
Dergi Sorumlusu
Akademik Dergiler Birimi
TÜBİTAK ULAKBİM-CABİM
Yüzüncüyıl, İşçi Blokları Mahallesi
Muhsin Yazıcıoğlu Caddesi
No:51/C 06530 Çankaya / ANKARA - TÜRKİYE
T +90 312 298 96 14
F
[ http://www.tubitak.gov.tr/ | www.tubitak.gov.tr ]
bedia.bal@tubitak.gov.tr
biol@tubitak.gov.tr
..........................................................................

[ http://www.tubitak.gov.tr/ ]
[ http://www.tubitak.gov.tr/disclaimer | Disclaimer ]


Kimden: "damla gozen" <damla.gozen@metu.edu.tr>
Kime: bmys-info@ulak.tubitak.gov.tr, "Core System" <biol@tubitak.gov.tr>
Gönderilenler: 24 Mayıs Pazartesi 2021 7:15:20
Konu: copyright permission

Dear Sir or Madam,

I would like to get permission for the attached article entitled
"Transcriptome profiles associated with selenium-deficiency-dependent
oxidative stress identify potential diagnostic and therapeutic targets in
liver cancer cells" published in 02.2021 to be used in my Ph.D thesis. I
would be glad if you could provide me an official document for the
permission. The figures, tables and information will all be used in my
thesis.

Best regards,

Damla Gözen

Figure 31: The copyright permissions for Figures and Tables taken or reproduced from
Gözen et al 2021.

<div align="center">**CURRICULUM VITAE**</div>

# DAMLA GOZEN

Konutkent 2 Sitesi, A5 Blok, No:6, Cayyolu, Ankara, Turkey
Mobile: +90 555 513 66 21 | E-mail: dgozen90@gmail.com

## PERSONAL INFORMATION

Date and Place of Birth:    April 22, 1990, Ankara
Citizenship:    Republic of Turkey
    Portuguese Republic

## EDUCATION

2014-...    **MIDDLE EAST TECHNICAL UNIVERSITY (METU)**
**Doctor of Philosophy** candidate, Medical Informatics with the cGPA of 4,00/4,00
*TUBITAK Ph.D. Scholarship Student*

2013-2014    **BILKENT UNIVERSITY**
Doctor of Philosophy, Molecular Biology and Genetics with the cGPA of 3,96/4,00

2011-2013    **BILKENT UNIVERSITY**
**Master of Science**, Molecular Biology and Genetics with the cGPA of 3,91/4,00

2007- 2011    **BILKENT UNIVERSITY**
**Bachelor of Science**, Molecular Biology and Genetics with the cGPA of 3,33/4,00

2004-2007    **TED ANKARA COLLEGE FOUNDATION PRIVATE HIGH SCHOOL**
An *International Baccalaureate* diploma

1996-2004    **TED ANKARA COLLEGE FOUNDATION SCHOOL**

## WORK EXPERIENCE

Nov 2019- …    -    **ROCHE DIAGNOSTICS TURKEY A.S.**
Sales Specialist- was responsible for the sales of pathology and molecular diagnostics products. Is now a member of Digital & IT solutions squad representing the commercial chapter to produce projects and develop strategies regarding to meet patient-based healthcare needs.

Sept 2017- Nov 2019 -    **NUCLEUS GENETİK**

Sales and Application Specialist- was responsible for the sales and application of Oxford Nanopore sequencing device as well as TWIST Bioscience and NanoString Technologies products.


**RESEARCH EXPERIENCE & PROFESSIONAL SKILLS**

2014-… - **Ph.D. PROJECT AT METU, DEPARTMENT OF MEDICAL INFORMATICS, CANSYL LABORATORY**

For her Ph.D. project, currently working in Prof. Dr. Rengul Cetin-Atalay's Laboratory, on the oxidative stress related gene expression profiles of liver cancer cells to identify genes that could be targeted in novel diagnostic and therapeutic strategies. Has been using R Programming Language for the analysis of an Affymetrix microarray result and for the determination of differentially expressed genes (DEGs) by various statistical models (LIMMA and ANCOVA), using several online tools for the functional analysis of the identified DEGs, and further searching the results by testing novel compounds for their cytotoxic effects on HCC.


2011-2013- **M.S. PROJECT AT BILKENT UNIVERSITY, DEPARTMENT OF MOLECULAR BIOLOGY AND GENETICS**

For her M.S. project, worked in Prof. Dr. Can Akcali's Laboratory, on the attachment and proliferation features of cancer cell lines with different EMT Status on carbon nanotubes (CNT) and hydrogel scaffolds that were used to mimic cancer niche. Practiced the culturing of primer and cell line cancer cells, the preparation of surface and synthesis of CNT, analysis of the attached cells on CNTS by Scanning Electron Microscopy and the determination of their viability by MTT; Reverse-Transcriptase PCR, SDS-PAGE, Immunofluorescence and Immunohistochemistry for the detection of EMT marker expressions.


2010-2011- **SENIOR PROJECT AT BILKENT UNIVERSITY, DEPARTMENT OF MOLECULAR BIOLOGY AND GENETICS**

For her senior project, worked in Prof. Dr. Tamer Yagci's Laboratory, developed her own project proposal on determining the          SIP1 gene expression patterns in HCC cells, which seems to be up-regulated and result in apoptosis inhibition and SIP1 correlation with miRNA200a, which was found to be down-regulated in many cancer types. Practiced Immunoprecipitation assay to determine protein interactions, protein isolation, SDS-PAGE and Western Blotting to find out gene expression patterns, ELISA and Wound Healing Assay.


2007-2017- **OTHER LABORATORY SKILLS:**

Has practiced the laboratory techniques:  Performing Mendelian genetics experiments with *Drosophila melanogaster,* observing gene expression patterns in *Caenorhabditis elegans,* FACS, bacterial culture techniques, novel drug discovery for cancer treatment (Cytotoxicity assays, Real-time cell proliferation analysis, SRB staining)


**PUBLICATIONS**

•    EXAMINATION OF FABRICATION CONDITIONS OF ACRYLATE-BASED HYDROGEL FORMULATIONS FOR DOXORUBICIN RELEASE AND EFFICACY TEST FOR HEPATOCELLULAR CARCINOMA CELL
Bayramoglu G, **Gozen D**, Ersoy G, Ozalp VC, Akcali KC, Arica MY. J Biomater Sci Polym Ed. 2014; 25(7):657-78.


•    A NEW TOOL FOR DIFFERENTIATING HEPATOCELLULAR CANCER CELLS: PATTERNED CARBON NANOTUBE ARRAYS
Kucukayan-Dogu G, **Gozen D**, Bitirim V, Akcali KC, Bengu E. Appl. Surf. Sci. 2015, 351: 27-32.

- A SERIES OF 2, 4 (1H, 3H)-QUINAZOLINEDIONE DERIVATIVES: SYNTHESIS AND BIOLOGICAL EVALUATION AS POTENTIAL ANTICANCER AGENTS
Akgun H, Us Yilmaz D, Cetin-Atalay R, **Gozen D**. Lett. Drug Des. Discov. 2016, 13, 64–76.

- TRANSCRIPTOME PROFILES ASSOCIATED WITH SELENIUM-DEFICIENCY-DEPENDENT OXIDATIVE STRESS IDENTIFY POTENTIAL DIAGNOSTIC AND THERAPEUTIC TARGETS IN LIVER CANCER CELLS
**Gözen D**, Kahraman DC, Narci K, Shehwana H, Konu Ö, Çetin-Atalay R. Turk J Biol. 2021 20;45(2):149-161.

- INHIBITION OF MDM2-P53 INTERACTION BY NOVEL COMPOUNDS AS THERAPEUTICS OF HEPATOCELLULAR CARCINOMA
**Gozen D**, Santos M, Cetin-Atalay R. COST Meeting, TENERIFE, 2014 (ABSTRACT)

- THE IDENTIFICATION OF OXIDATIVE STRESS RELATED GENE PROFILES IN LIVER CANCER CELLS AS TARGETS FOR DIAGNOSTICS AND THERAPEUTICS
**Gozen D**, Shehwana H, Konu O, Dogan T and Cetin-Atalay R. HIBIT Conference, CYPRUS, 2017 (ABSTRACT)

## TEACHING ASISTANCE EXPERIENCE

**2011-2014- MBG 101 BIOLOGY 1 TEACHING ASSISTANCE**
Worked as one of the four Teaching Assistants in MBG 101 Biology 1 course given by Assoc. Prof. Ali Gure which included a 4-hour-Lab session given by the teaching assistants.

## SCHOLARSHIPS & AWARDS

2013-2021 Has a Ph.D. Scholarship from The Scientific and Technological Research Council of Turkey (TÜBİTAK) awarded for cGPAs through undergraduate and graduate education, as well as the ALES national exam score

2005-2006 Received 80% Merit Scholarship for a yearly GPA in her 2nd year in High School

## LANGUAGE SKILLS

- Turkish (Native Language)
- Advanced English : **TOEFL IBT Score: 102/120 (Aug 16, 2013)**

## COMPUTER SKILLS

- Advanced knowledge of Microsoft Office (data analysis with Excel, figure preparation with PowerPoint, Word)
- Intermediate knowledge of Matlab and R programming
- Intermediate knowledge of web-based molecular biology and genetics databases

## EXTRACURRICULAR ACTIVITIES

2012-2020 **MEMBER OF AKUT SEARCH AND RESCUE ASSOCIATION**
Was trained for natural disaster scenarios, had rescue, firefighting and climbing trainings.
2016-… **FIRST AID CERTIFICATE**

2012-…        **SCUBA DIVING CERTIFICATE**
              CMAS 2 Star Diver
2009-2012     **BILKENT UNIVERSITY INFORMATION AND ADVERTISEMENT OFFICE**

2002-2011     **HORSE RIDING**
              Licensed; 2006, 2007, 2008 3rd degree in Region Championship