Transcriptome profiles associated with selenium-deficiency-dependent oxidative stress identify potential diagnostic and therapeutic targets in liver cancer cells

Damla GÖZEN[1] , **Deniz Cansen KAHRAMAN**[1,]* , **Kübra NARCI**[1] , **Huma SHEHWANA**[2] , **Özlen KONU**[3] ,
Rengül ÇETİN-ATALAY[1]

[1]Cancer Systems Biology Laboratory, Department of Health Informatics, Graduate School of Informatics,
Middle East Technical University, Ankara, Turkey
[2]Department of Biological Sciences, National University of Medical Sciences, Rawalpindi, Pakistan
[3]Department of Molecular Biology and Genetics, Faculty of Science, Bilkent University, Ankara, Turkey

Abstract: Hepatocellular carcinoma (HCC) is one of the most common cancer types with high mortality rates and displays increased resistance to various stress conditions such as oxidative stress. Conventional therapies have low efficacies due to resistance and off-target effects in HCC. Here we aimed to analyze oxidative stress-related gene expression profiles of HCC cells and identify genes that could be crucial for novel diagnostic and therapeutic strategies. To identify important genes that cause resistance to reactive oxygen species (ROS), a model of oxidative stress upon selenium (Se) deficiency was utilized. The results of transcriptome-wide gene expression data were analyzed in which the differentially expressed genes (DEGs) were identified between HCC cell lines that are either resistant or sensitive to Se-deficiency-dependent oxidative stress. These DEGs were further investigated for their importance in oxidative stress resistance by network analysis methods, and 27 genes were defined to have key roles; 16 of which were previously shown to have impact on liver cancer patient survival. These genes might have Se-deficiency-dependent roles in hepatocarcinogenesis and could be further exploited for their potentials as novel targets for diagnostic and therapeutic approaches.

Key words: Hepatocellular carcinoma, selenium, oxidative stress, transcriptome-wide analysis

1. Introduction

Cancer is one of the major public health problems worldwide, being the second most deadly disease, according to the World Health Organization, and nearly 30 million new cases are estimated to occur globally by 2040 (Siegel et al., 2020). Cancer therapeutics has shifted away from the conventional chemotherapeutic drugs towards targeted therapeutic strategies to provide higher efficacy with lower toxicity. For finding novel genes for diagnosis and targeted therapy, it is crucial to determine cancer cell characteristics in more detail and provide more insight into molecular mechanisms behind carcinogenesis.

Hepatocellular carcinoma (HCC) is the fourth most fatal and the fifth most frequent cancer worldwide (Bray et al., 2018). Molecular mechanisms involved in HCC 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. Therefore, it is essential to determine the stress response gene expression profiles of these cells

due to their involvement in hepatocarcinogenesis (Di Maso et al., 2015).

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 (Fulda et al., 2010). One of the stress factors eukaryotic cells try to adapt to is oxidative stress conditions for which cells have evolved different responses. Oxidative stress results from the inability of the biological system to detoxify the reactive intermediates or to compensate the resulting damage that is formed as a result of systemic accumulation of reactive oxygen species (ROS). Disturbances in the normal redox state of cells produce peroxides and free radicals that damage all components of the cell, causing toxic effects. It may also suppress apoptosis and promote proliferation, invasiveness, and metastasis, leading to carcinogenesis (Halliwell et al., 2007).

Selenium (Se) is a trace element which is required for human health, and its deficiency results in various abnormalities. It is found in the structure of selenocysteine

* Correspondence: cansen@metu.edu.tr

149

(SeCys) amino acid, which is the building block of selenoproteins. There are about 25 different selenoproteins identified in humans with various functions, including antioxidant and redox signalling functions (Papp et al., 2007). Normally, deficiency of Se results in oxidative stress, leading to apoptosis. However, in a previous study we performed with HCC cell lines, it was found that 10 of 13 HCC cell lines tolerated Se deficiency to escape apoptosis, and most of these tolerant cell lines had Hepatitis B virus (HBV) sequence integrated into their genome, indicating that this virus might have a role in that acquired tolerance (Irmak et al., 2003). This study was repeated with two isogenic HCC cell lines; HepG2 and HepG2-2.2.15 cells, to test their response to Se deficiency. 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 Se-deficiency-sensitive while HepG2-2.2.15 cells tolerated the absence of Se to survive. Although the underlying mechanism remains unclear, it could be a result of both intrinsic and/or acquired mechanisms.

During the viral genome integration into HepG2-2.2.15 cells, an increase in ROS generation is expected in the cells, which alters the cellular gene expression (Waris et al., 2005). This would result in an intrinsic variation in the gene expression of HepG2-2.2.15 cells compared to HepG2 cells, independent of the Se status of the growth medium. In addition to this intrinsic effect, an acquired mechanism might be at work due to the absence of Se. Understanding the molecular mechanism behind the Se tolerance might give valuable information about cellular response mechanisms gained by some cancer cells to escape from oxidative-stress-dependent apoptosis.

Large-scale expression analysis using microarray or RNAseq data has great potential to enlighten the changes that occur at the molecular level resulting in hepatocellular carcinogenesis (Chen et al., 2020). In this study, systems-based approaches were used to identify genes important in oxidative stress resistance mechanisms, which could be further exploited as novel drug targets or for diagnostic purposes in HCC.

2. Materials and methods

2.1. Cell lines and microarray experiment
As we described in our previous study (Irmak et al., 2003), oxidative-stress-resistant HepG2-2.2.15 cell line (HT (Head-to-Tail)-HBV integration) and -sensitive HepG2 cell line were grown in DMEM (Dulbecco's Modified Eagle's medium) (2×10^5 cells in 10 mm culture dish) (Gibco, Thermo Fisher Scientific, MA, USA) with 0.01% FCS (BioChrom, Berlin, Germany) either supplemented with 0.1 μM Na2SeO3 (Sigma, Taufkircher, Germany) (Se positive) or not (Se negative) for 3 days in duplicates at 37 °C in an incubator containing 5% CO2. They were harvested on the 1st, 2nd, and 3rd days of the treatment,

and RNA isolation was performed. For each day, gene expression data were acquired by Human Genome U133 Plus 2.0 Array Affymetrix Array. The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (Edgar et al., 2002) and are accessible through GEO Series accession number GSE163950[1].

2.2. Preprocessing of microarray data
The qualities of arrays were checked using simpleaffy package available in R[2]. Robust multiarray average (RMA) background correction and quantile normalization of the data were performed.

2.3. Determination of differentially expressed genes (DEG)s
To identify DEGs between HepG2 and HepG2-2.2.15 cells in the presence or absence of Se, limma package in R was used (Ritchie et al., 2015). The design matrix was organized so that it involved both cell lines (HepG2 or HepG2-2.2.15) and Se status (present or absent) information, separately for each day to fit the data to a linear model in the first step.

In line with the design matrix, the contrast matrix was designed in a specific way to determine the effects of cell line and Se status on gene expression separately for each day.

The first two comparisons were based on the identification of the genes that were differentially expressed between HepG2 and HepG2-2.2.15 cells in the presence of Se (Se+ HepG2 vs. Se+ HepG2-2.2.15) or absence of Se (Se−HepG2 vs. Se−HepG2-2.2.15) and named "between cell line comparisons". 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 (Se−HepG2 vs. Se+ HepG2) or HepG2-2.2.15 (Se−HepG2-2.2.15 vs. Se+ HepG2-2.2.15) and called "within cell line" comparisons. The p-value threshold was taken as 0.01 while selecting the DEG lists.

2.4. Clustering analysis of differentially expressed genes
Z-score of each gene in the defined DEG list was calculated for clustering of the samples (Kreyszig, 1979). Heatmaply package (Galili et al., 2017) in R-Bioconductor was used to draw the heat maps.

2.5. Gene set enrichment analysis with DEGs
A score called 'DEG score' was calculated (Cavga et al., 2019), providing the normalization of the results for each DEG relative to each other. DEG scores were calculated by subtracting $\log_2$ fold change values of HepG2 DEGs in within cell line comparison results from that of HepG2-2.2.15 DEGs and taking the absolute values for each. Gene set enrichment analysis (GSEA) (Subramanian et al., 2005) was performed using DEG lists and for each

gene, preranked inputs were determined using DEG score values. The Molecular Signatures Database (MSigDB) hallmark or GO-Biological Process data collection was used to search for enriched gene sets using the default "weighted" enrichment statistic parameter and "meandiv" normalization. Enriched gene sets with FDR q-values < 0.25 were considered significant.

2.6. Network analysis

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

2.6.2. Prize-collecting Steiner tree
Prize-collecting Steiner tree (PCST) (Tuncbag et al., 2016) approach was used to identify the interactions through the DEG lists by using the information provided by human interactome data. OmicsIntegrator software's Forest module was used to determine the subnetwork in the set of DEGs.

To construct meaningful trees, β, ω, and μ input parameters must be chosen correctly, and for each DEG list, the optimum values of those input parameters were determined separately by the Forest-tuner algorithm[3] to generate 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 (Jensen et al., 2009). In this database, each edge is given a confidence score between 0 and 1 according to the reliability of the data source. The edges that have confidence scores higher than 0.7 were chosen. These scores were used to determine the costs by OmicsIntegrator.

2.7. Survival data
Kaplan–Meier survival analysis was performed by generating plots for each gene of interest by KM plotter[4], which plots the survival data of liver cancer patients according to liver cancer RNA-seq results from TCGA including all pathology and patient types. The results with logrank p-values smaller than 0.05 were accepted to be significant.

3. Results

3.1. Identification of differentially expressed selenium deficiency resistance and sensitivity genes
Transcriptome-wide gene expression analysis was performed in two different cell lines with or without selenium treatment in a time-dependent manner (three

[3]https://github.com/gungorbudak/forest-tuner accessed on 25.06.2020
[4]Menyhart et al. (2018). Kaplan-Meier Plotter [online]. Website https://kmplot.com

time-points; day 1, 2, and 3), and the workflow of the analysis, which was followed throughout the study, is summarized in Figure 1.

To analyze the results and determine DEGs that are important for the tolerance to Se-deficiency-dependent oxidative stress, limma package in R was used. As explained in methods, four comparisons were performed separately for day 1, 2, and 3.

Representative live images of HepG2 and HepG2-2.2.15 cells grown in Se+ or Se– media for 72 h showed that the effects of Se deficiency on cell confluency of HepG2 cells was detected most significantly on the 3rd day when compared to those incubated in Se+ media (Figure 2a). Meanwhile, HepG2-2.2.15 cells were able to grow under both conditions. These results were in parallel with our previous findings (Irmak et al., 2003), where HepG2 cells underwent apoptotic cell death under selenium deficiency, while HepG2 2.2.15 cells displayed tolerance and similar growth rates under Se– and Se+ culture conditions.

Numbers of DEGs for each condition were identified per comparison respectively as indicated in Figures 2b and 2c. By the first two limma analyses between cell line comparisons (Figure 2b), 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 'Se-deficiency-dependent effect genes' since they were altered depending on the absence of Se. The shared DEGs in both Se+ and Se– comparisons were named 'HBV-integration effect genes' in this study (Figure 2b), as their expression differences were independent of the Se status; which is thought to be about the integration of HBV viral genome. By the last two limma analyses within each cell line comparisons (Figure 2c), 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 expression on the presence of Se. The DEGs that were not shared in common were named 'Se-deficiency-dependent effect genes' and further examined since they were directly related to the different reactions each cell line gave to the deficiency of Se.

3.2. Differentially expressed gene clustering analysis of the between cell line comparisons
For between cell line comparisons, the z-scores for expressions of the Se-deficiency-dependent effect genes in each sample were calculated and used to draw a heat map clustering samples all together for day 1 (D1), day 2 (D2), and day 3(D3) data (Figure 3a). As expected, HepG2 and HepG2-2.2.15 cells were clustered in two distinct groups. PPAP2A, HOXD1, and CLYBL genes were found to be the most DEGs between the two cell lines. A similar clustering was performed for the HBV-integration effect genes. HepG2 and HepG2-2.2.15 cells were clustered separately again (Figure 3b).
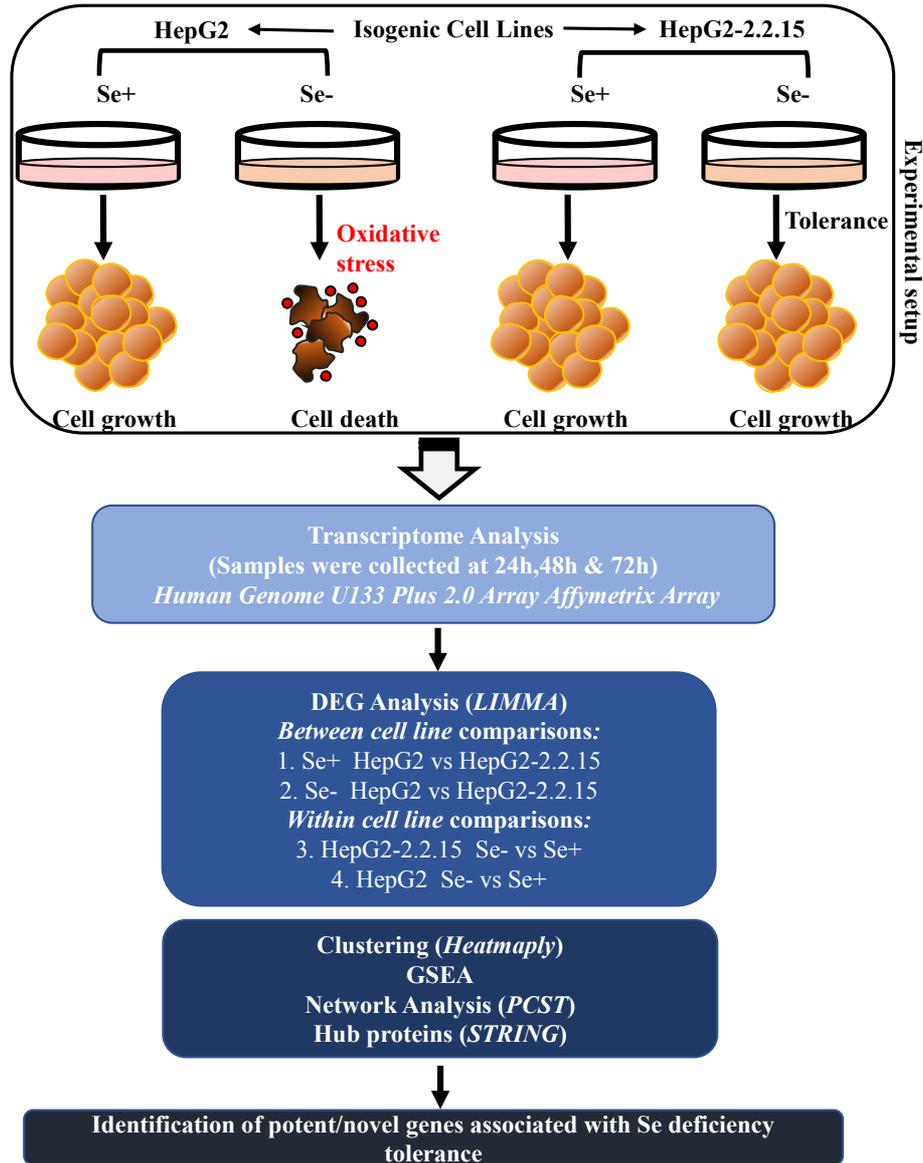
Figure 1. Illustration of the experimental design used to generate and analyze the data. HepG2 and HepG2-2.2.15 cells (two isogenic HCC lines with the difference of HBV genome integration in HepG2-2.2.15 cells) were grown in the presence or absence of Se to perform transcriptome analysis. The results were examined by further bioinformatics methods to determine the differential response mechanisms.

FGF13, GPC3, and MAP7D2 genes were identified as the DEGs with the smallest p-values between HepG2 and HepG2-2.2.15. In parallel, we also used the HBV-integration effect gene lists to draw heat maps for the comparative expression levels of these DEGs in 8 different hepatocellular cancer cell lines taken from Cancer cell line encyclopedia database[5]. Six of these 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) are known to be virus-free and sensitive to Se-deficiency-dependent oxidative stress. 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 3c).

[5]https://portals.broadinstitute.org/ccle

A.

B.

Between Cell Line **Comparison**

Se+ (HepG2 vs 2.2.15)   Se- (HepG2 vs 2.2.15)

D1: 88
D2: 42
D3: 120

D1: 185
D2: 114
D3: 151

D1: 29
D2: 118
D3: 29

C.

Within Cell Line **Comparison**

HepG2 (Se+ vs Se-)   2.2.15 I (Se+ vs Se-)

D1: 27
D2: 34
D3: 128

GPX
SEPW1

D1: 45
D2: 25
D3: 43
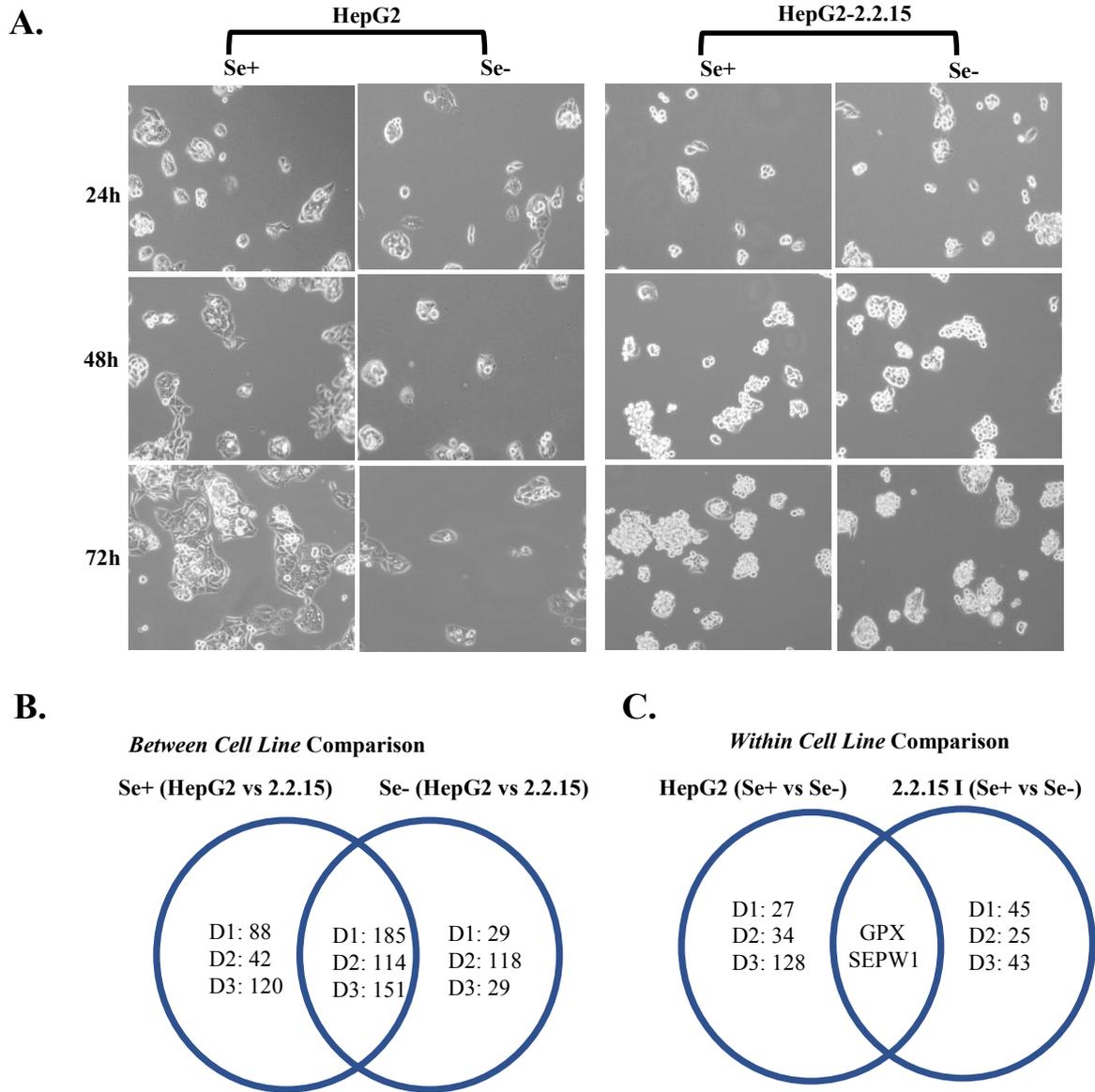
Figure 2. Experimental groups and differential expression analysis of HepG2 and HepG2-2.2.15 cell lines. (a) Representative live images of HepG2 and HepG2-2.2.15 cells grown in Se+ or Se– media for 72 h. The DEG numbers were identified as a result of the (b) between cell line and (c) within cell line 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: 24 h, D2: 48 h; D3: 72 h.

3.3. Gene set enrichment analysis of isogenic HepG2 and HepG2-2.2.15 cells

For the within cell line comparison results, DEG scores were calculated (Cavga et al., 2019) as explained in methods (Figure 4a). The pathways enriched in lists with positive DEG scores (DEG scores > 1) indicate that the relevant genes were relatively upregulated 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, the pathways enriched in lists with negative DEG scores (DEG scores < 1) indicated a lower activity in HepG2-2.2.15 cells in response to Se-deficiency compared to HepG2. GSEA results have revealed that the relatively upregulated 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

A.

B.

C.

Cell Line	HBV-Integration	Se–Deficiency Tolerance
SNU182	+	+
SNU449	+	+
SNU423	+	+
SNU387	+	+
SNU182	+	+
PLC/PRF/5	+	+
HUH7	-	-
HEPG2	-	-
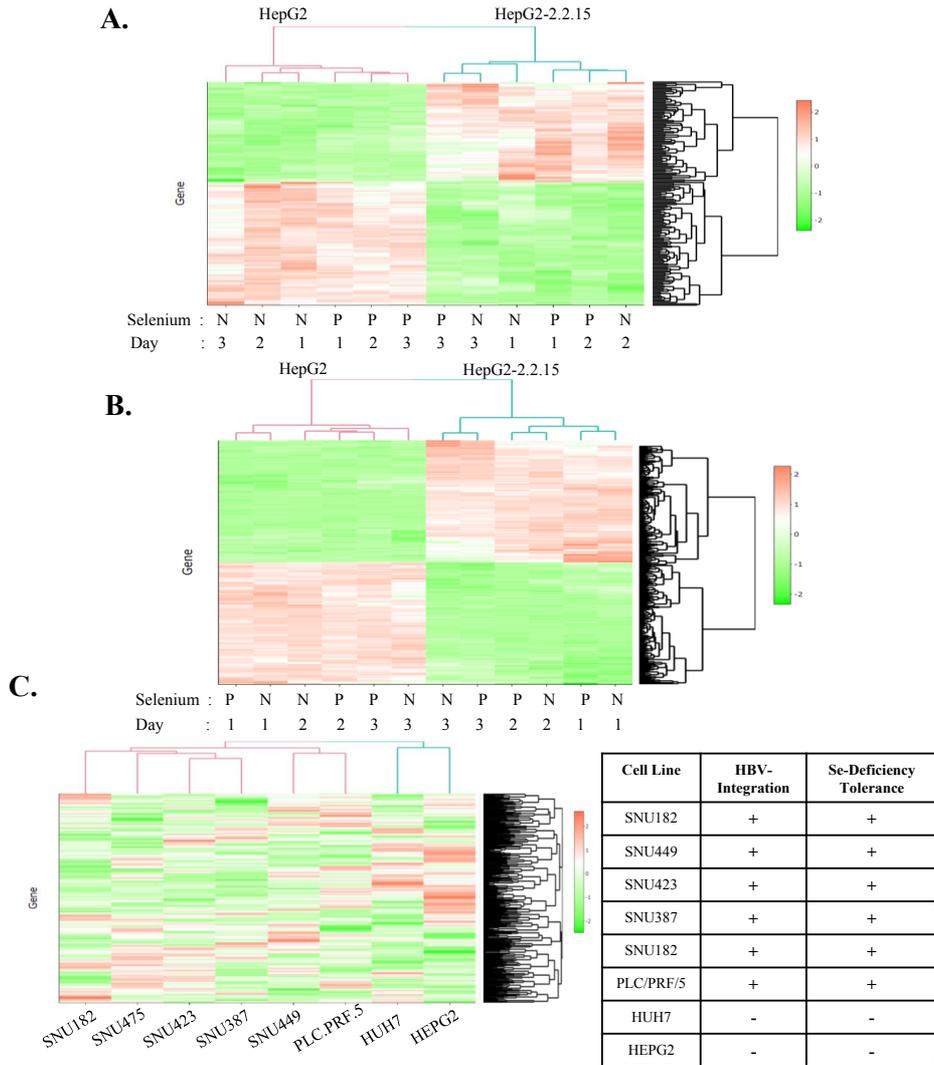
Figure 3. The heat map drawn with (a) Se-deficiency effect genes and (b) HBV-integration effect genes. (c) The heat map drawn with Z scores of HBV-integration effect genes calculated for the expression levels of 8 different cell lines taken from CCLE results; HBV-integration and Se-deficiency tolerance conditions of cell lines were depicted in the table.

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 related to stress response, repair mechanisms, and cell cycle (Table S1). On the contrary, the relatively downregulated 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 (Zamorana et al., 1998) and might indicate lower apoptotic phenotype in within HepG2-2.2.15 comparison relative to that of HepG2. 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 were consistent with GSEA results (Figure 4b).

3.4. Pathway analysis to identify key genes related to selenium-dependent oxidative stress tolerance

PCST algorithm was used to construct trees for Se–deficiency effect genes by taking STRING as the reference dataset (Figure 5). 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 5). In day 1 results, FOXA1, an oxidoreductase which has a proapoptotic role, and CYP7A1 were differentially expressed while ONECUT1

with roles in cell cycle regulation and PITX2 involved in oxidative stress response were detected as Steiner nodes, i.e. differentially expressed but might be critical in the differential response (Figure 5a). In day 2 results, the DEGs TXNRD1 and ALDH1L2 and the Steiner nodes ACLY, TXNIP, SCD5, MTR, and TXNDC17 all exhibiting oxidative stress and redox homeostasis-related functions (Figure 5b) emerged. In day 3 results, the Steiner nodes with high betweenness centrality values included LSM4 that is associated stress response and CNBP, DMPK, QDPR with antioxidant functions (Figure 5c). This has supported

A.

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

> 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 4. 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. Only top 5 KEGG pathways with the smallest FDR values, and Uniprot annotated keyword related to damage was shown for clarity. ES: enrichment score, NOM p-val: nominal p value, FDR: false discovery rate.

Figure 5. The trees constructed by PCST algorithm for acquired HepG2 vs. HepG2-2.2.15 DEG lists at day 1 (a), day 2, (b) and day 3 (c). 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.

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.

3.5. Definition and clinical significance of selected genes related with oxidative stress resistance
In this study, various bioinformatics approaches were used to determine potentially important biomarkers that have functions in Se-dependent oxidative stress resistance.

Overall, all the genes that were defined to be important in our study by different analysis methods are summarized in Table. Twenty-seven genes were selected based on within or between cell line comparisons that were associated with either selenium or HBV effect. While most of the genes were previously identified in HCC (17 genes) and oxidative stress (20 genes), HOXD1 and CLYBL were shown 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 heat map 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.

Finally, to find out the clinical relevance of the selected genes; Kaplan–Meier plots were drawn for the liver cancer RNA-seq results (Figure 6). The expression levels of 16 out of 27 genes were found to be associated with overall survival time of the HCC patients.

4. Discussion
In this study, the aim was to determine Se-deficiency-dependent oxidative-stress-related gene expression profiles of two isogenic HCC cell lines varying with

Table. 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: Heat map, OS: Oxidative stress, r: Reported.

Gene	Effect	Comparison	Analysis	OS	HCC	Literature
FOXA1	Se	BCL	PCST (E)	r	r	Zhang et al. (2005), Song et al. (2009)
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 (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 et al. (2013)
SCD5	Se	BCL	PCST (S)	r	r	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		Chen et al. (2017)
CNBP	Se	BCL	PCST (S)	r		de Peralta et al. (2016)
DMPK	Se	BCL	PCST (S)	r		Pantic et al. (2013)
QDPR	Se	BCL	PCST (S)	r	r	Gu et al. (2017), Nwosu et al. (2017)
DUT	Se	WCL	GSEA		r	Takatori (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), Huang et al. (2019)
GINS2	Se	WCL	GSEA	r	r	Lian et al. (2018), Liu et al. (2019)
PIK3R3	Se	WCL	GSEA	r	r	Engedal et al. (2018), Ibrahim et al. (2018)
TMEM97	Se	WCL	GSEA	r		Wang et al. (2020)
FGF13	HBV	BCL	HM	r	r	Coleman et al. (2014), Bublik et al. (2017)
GPC3	HBV	BCL	HM		r	Akutsu et al. (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			

respect to HBV integration, to identify genes that could be targeted with novel diagnostic and therapeutic strategies. With this aim, the results of an Affymetrix expression array were analyzed and the DEG lists were identified using limma analysis. 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 (Figure 2). 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 protein nodes within cell signaling networks. Some values such as degree of a node and betweenness centrality might be used to
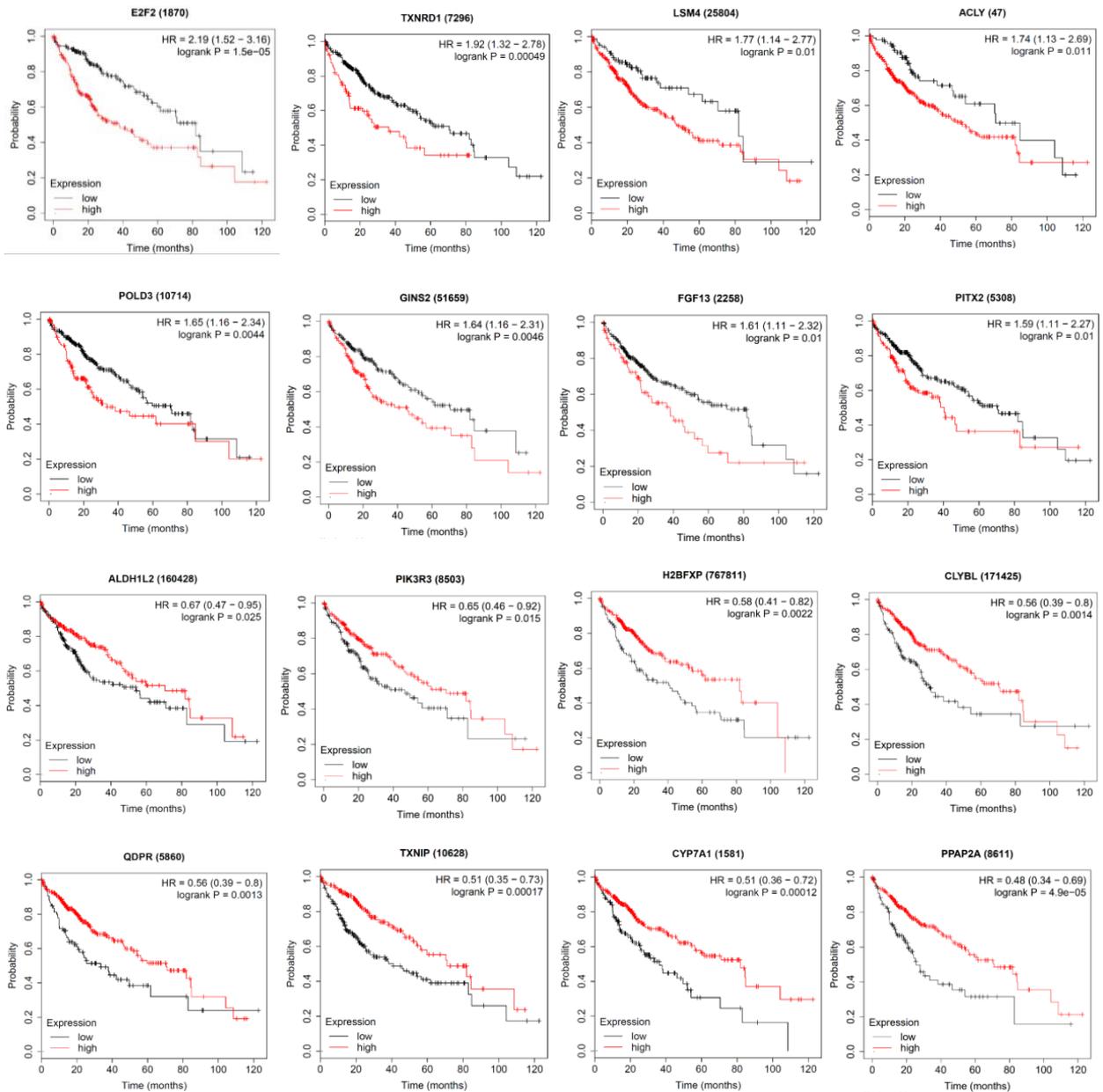
Figure 6. The Kaplan Meier plots were generated for selected DEGs in HCC patients in liver cancer RNA-seq dataset by considering Overall Survival data.

determine biologically important hub proteins (Yu et al., 2007). The eleven nodes identified by PCST analysis were Steiner nodes and previous studies have shown that such hidden nodes are able to connect the pathways and also indicate cross talks (Huang et al., 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). Among the six genes identified by GSEA, four (POLD3, E2F2, GINS2, PIK3R3) were previously shown to have oxidative stress and hepatocellular carcinogenesis related functions (Table). Moreover, the association of four of the six DEGs identified by heat map analysis with HCC was shown in previous studies. These results suggest that our statistical methodology has revealed cancer pathways that could be related with hepatocarcinogenesis and extrinsic factors. HOXD1 and CLYBL genes were not associated with any oxidative stress or hepatocellular-carcinogenesis-related function to our knowledge.

GSEA results also indicated that the expression of 15 of those genes was regulated by different combinations of six transcription factors as shown in Table S2, i.e. HSD17B8, CHX10, ZBTB5, TFCP2, LYF1, and E2F2. These transcription factors might be the main targets that should be further investigated for their potential in novel diagnostics and therapeutic strategies. The expression of 16 genes among 27 genes was 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 (Figure 6). Since various analysis methods were used, it is difficult to make a straightforward association between the survival data and our overall analysis results. However, it is worth exploiting the mechanism of action of these genes/pathways to verify their clinical significance in HCC.

The identification of genes that could be used to predict the prognosis of HCC patients, or lead to the discovery of new therapeutic strategies 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. Moreover, the effects of selenium-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.

Conflict of interest
The authors declare that there is no conflict of interest.

References

Akutsu N, Yamamoto H, Sasaki S, Taniguchi H, Arimura Y et al. (2010). Association of glypican-3 expression with growth signaling molecules in hepatocellular carcinoma. World Journal of Gastroenterology 16 (28): 3521-3528. doi: 10.3748/wjgv16.i28.3521

Archer KJ, Mas VR, Maluf DG, Fisher RA (2010). High-throughput assessment of CpG site methylation for distinguishing between HCV-cirrhosis and HCV-associated hepatocellular carcinoma. Molecular Genetics and Genomics 283 (4): 341-349. doi: 10.1007/s00438-010-0522-y

Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA et al. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. A Cancer Journal of Clinicians 68 (6): 394-424. doi: 10.3322/caac.21492

Bublik DR, Bursać S, Sheffer M, Oršolić I, Shalit T et al. (2017). Regulatory module involving FGF13, miR-504, and p53 regulates ribosomal biogenesis and supports cancer cell survival. Proceedings of the National Academy of Sciences of the United States of America 114 (4): E496-E505. doi: 10.1073/pnas.1614876114

Castillo DS, Campalans A, Belluscio LM, Carcagno AL, Radicella JP et al. (2015). E2F1 and E2F2 induction in response to DNA damage preserves genomic stability in neuronal cells. Cell Cycle 14 (8): 1300-1314. doi: 10.4161/15384101.2014.985031

Cavga AD, Tardu M, Korkmaz T, Keskin O, Ozturk N et al. (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): 729-742. doi: 10.1007/s10142-019-00680-5

Chen B, Garmire L, Calvisi DF, Chua MS, Kelley RK et al. (2020). Harnessing big 'omics' data and AI for drug discovery in hepatocellular carcinoma. Nature reviews. Gastroenterology & Hepatology 17 (4): 238-251. doi: 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: 1809592. doi: 10.1155/2017/1809592

Coleman SJ, Grose RP, Kocher HM (2014). Fibroblast growth factor family as a potential target in the treatment of hepatocellular carcinoma. Journal of Hepatocellular Carcinoma 1: 43-54. doi: 10.2147/JHC.S48958

de Peralta MS, Mouguelar VS, Sdrigotti MA, Ishiy FA, Fanganiello RD et al. (2016). Cnbp ameliorates Treacher Collins Syndrome craniofacial anomalies through a pathway that involves redox-responsive genes. Cell Death and Disease 7 (10): e2397. doi: 10.1038/cddis.2016.299

Di Maso V, Mediavilla MG, Vascotto C, Lupo F, Baccarani U et al. (2015). Transcriptional up-regulation of APE1/Ref-1 in hepatic tumor: role in hepatocytes resistance to oxidative stress and apoptosis, PLoS One 10 (12): e0143289. doi: 10.1371/journal.pone.0143289

Edgar R, Domrachev M, Lash AE (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Research 30 (1): 207-210.

Engedal N, Žerovnik E, Rudov A, Galli F, Olivieri F et al. (2018). From oxidative stress damage to pathways, networks, and autophagy via MicroRNAs. Oxidative Medicine and Cellular Longevity 2018: 4968321. doi: 10.1155/2018/4968321

Fulda S, Gorman AM, Hori O, Samali A (2010). Cellular stress responses: cell survival and cell death. International Journal of Cell Biology 2010: 214074. doi: 10.1155/2010/214074

Galili T, O'Callaghan A, Sidi J, Sievert C (2017). heatmaply: an R package for creating interactive cluster heatmaps for online publishing. Bioinformatics 34 (9): 1600-1602. doi: 10.1093/bioinformatics/btx657

Gu YT, Wang YC, Zhang HJ, Zhao TT, Sun SF (2017). Protective effect of dihydropteridine reductase against oxidative stress is abolished with A278C mutation. Journal of Zhejiang University. Science B 18 (9): 770-777. doi: 10.1631/jzusB1600123

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): 2008-2021. doi: 10.7150/jca.39972

Halliwell B (2007). Oxidative stress and cancer: have we moved forward? The Biochemical Journal 401 (1): 1-11. doi: 10.1042/BJ20061131

Huang SS, Fraenkel E (2009). Integrating proteomic, transcriptional, and interactome data reveals hidden components of signaling and regulatory networks. Science Signaling 2 (81): ra40. doi: 10.1126/scisignal.2000350

Huang YL, Ning G, Chen LB, Lian YF, Gu YR et al. (2019). Promising diagnostic and prognostic value of E2Fs in human hepatocellular carcinoma. Cancer Management and Research 11: 1725-1740. doi: 10.2147/CMAR.S182001

Ibrahim S, Li G, Hu F, Hou Z, Chen Q et al. (2018). PIK3R3 promotes chemotherapeutic sensitivity of colorectal cancer through PIK3R3/NF-kB/TP pathway. Cancer Biology and Therapy 19 (3): 222-229. doi: 10.1080/15384047.2017.1416936

Iizuka N, Oka M, Yamada-Okabe H, Mori N, Tamesa T et al. (2003). Differential gene expression in distinct virologic types of hepatocellular carcinoma: association with liver cirrhosis. Oncogene 22 (19): 3007-3014. doi: 10.1038/sj.onc.1206401

Irmak MB, Ince G, Ozturk M, Cetin-Atalay R (2003). Acquired tolerance of hepatocellular carcinoma cells to selenium deficiency: a selective survival mechanism? Cancer Research 63: 6707-6715.

Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C et al. (2009) STRING 8--a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Research D412-6. doi: 10.1093/nar/gkn760

Jenkins E, Brenner M, Laragione T, Gulko PS (2012). Synovial expression of Th17-related and cancer-associated genes is regulated by the arthritis severity locus Cia10. Genes and Immunity 13 (3): 221-231. doi: 10.1038/gene.2011.73

Jiang L, Wu CL, Wu JZ, Yang X, Wang HL et al. (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): 68-81. doi: 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): 1091-1097. doi: 10.1515/BC.2007.133

Kreyszig E (1979). Advanced Engineering Mathematics. 4th ed. Hoboken, NJ, USA: Wiley.

Lee D, Xu IM, Chiu DK, Leibold J, Tse AP et al. (2019). Induction of oxidative stress through inhibition of thioredoxin reductase 1 is an effective therapeutic approach for hepatocellular carcinoma. Hepatology 69 (4): 1768-1786. doi: 10.1002/hep.30467

Lee D, Xu IM, Chiu DK, Lai RK, Tse AP et al. (2017). Folate cycle enzyme MTHFD1L confers metabolic advantages in hepatocellular carcinoma. The Journal of Clinical Investigation 127 (5): 1856-1872. doi: 10.1172/JCI90253

Lian YF, Li SS, Huang YL, Wei H. Chen DM et al. (2018). Up-regulated and interrelated expressions of GINS subunits predict poor prognosis in hepatocellular carcinoma. Bioscience Reports 38 (6): BSR20181178. doi: 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): 1-9. doi: 10.1080/21691401.2018.1540425

Liu H, Pathak P, Boehme S, Chiang JY (2016). Cholesterol 7α-hydroxylase protects the liver from inflammation and fibrosis by maintaining cholesterol homeostasis. Journal of Lipid Research 57 (10): 1831-1844. doi: 10.1194/jlr.M069807

Liyanage DS, Omeka WKM, Yang H, Godahewa GI, Kwon H et al. (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: 301-310. doi: 10.1016/j.fsi.2018.11.040

Menyhárt O, Nagy Á, Győrffy B (2018). Determining consistent prognostic biomarkers of overall survival and vascular invasion in hepatocellular carcinoma. Royal Society Open Science 5 (12): 181006. doi: 10.1098/rsos.181006

Migita T, Okabe S, Ikeda K, Igarashi S, Sugawara S et al. (2013). Inhibition of ATP citrate lyase induces an anticancer effect via reactive oxygen species: AMPK as a predictive biomarker for therapeutic impact. The American Journal of Pathology 182 (5): 1800-1810. doi: 10.1016/j.ajpath.2013.01.048

Nwosu ZC, Megger DA, Hammad S, Sitek B, Roessler S et al. (2017). Identification of the Consistently Altered Metabolic Targets in Human Hepatocellular Carcinoma. Cellular and Molecular Gastroenterology and Hepatology, 4 (2): 303-323.e1. doi: 10.1016/j.jcmgh.2017.05.004.

Nishida N, Chishina H, Arizumi T, Takita M, Kitai S et al. (2014). Identification of epigenetically inactivated genes in human hepatocellular carcinoma by integrative analyses of methylation profiling and pharmacological unmasking. Digestive Diseases 32 (6): 740-746. doi: 10.1159/000368015

Pantic B, Trevisan E, Citta A, Rigobello MP, Marin O et al. (2013). Myotonic dystrophy protein kinase (DMPK) prevents ROS-induced cell death by assembling a hexokinase II-Src complex on the mitochondrial surface. Cell Death and Disease 4 (10): e858. doi: 10.1038/cddis.2013.385

Papp LV, Lu J, Holmgren A, Khanna KK (2007). From selenium to selenoproteins: synthesis, identity, and their role in human health. Antioxidant and Redox Signaling 9 (7): 775-806. doi: 10.1089/ars.2007.1528

Pope ED, Kimbrough EO, Vemireddy LP, Surapaneni PK, Copland JA et al. (2019). Aberrant lipid metabolism as a therapeutic target in liver cancer. Expert Opinion on Therapeutic Targets 23 (6): 473-483. doi: 10.1080/14728222.2019.1615883

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW et al. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research 43 (7): e47. doi: 10.1093/nar/gkv007

Sarret C, Ashkavand Z, Paules E, Dorboz I, Pediaditakis P et al. (2019). Deleterious mutations in ALDH1L2 suggest a novel cause for neuro-ichthyotic syndrome. NPJ Genomic Medicine 4: 17. doi: 10.1038/s41525-019-0092-9

Siegel R, Kimberly D, Jemal A et al. (2020). Cancer statistics, 2020. CA: A Cancer Journal for Clinicians 70: 7-30. doi: 10.3322/caac.21590

Si M, Zhao C, Zhang B, Wei D, Chenet K et al. (2016). Overexpression of mycothiol disulfide reductase enhances corynebacterium glutamicum robustness by modulating cellular redox homeostasis and antioxidant proteins under oxidative stress. Scientific Reports 6: 29491. doi: 10.1038/srep29491

Song L, Wei X, Zhang B, Luo X, Liu J et al. (2009). Role of Foxa1 in regulation of bcl2 expression during oxidative-stress-induced apoptosis in A549 type II pneumocytes. Cell Stress Chaperones 14 (4): 417-425. doi: 10.1007/s12192-008-0095-4

Strungaru MH, Footz T, Liu Y, Berry FB, Belleau P et al. (2011). PITX2 is involved in stress response in cultured human trabecular meshwork cells through regulation of SLC13A3. Investigative ophthalmology and visual science 52 (10): 7625-7633. doi: 10.1167/iovs.10-6967

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America 102 (43): 15545-50. doi: 10.1073/pnas.0506580102

Takatori H, Yamashita T, Honda M, Nishino R, Arai K et al. (2010) dUTP pyrophosphatase expression correlates with a poor prognosis in hepatocellular carcinoma. Liver International 30 (3): 438-446. doi: 10.1111/j.1478-3231.2009.02177x

Tan J, Duan M, Yadav T, Phoon L, Wang X et al. (2020). An R-loop-initiated CSB-RAD52-POLD3 pathway suppresses ROS-induced telomeric DNA breaks. Nucleic Acids Research 48 (3): 1285-1300. doi: 10.1093/nar/gkz1114

Tuncbag N, Milani P, Pokorny JL, Johnson H, Sio TT et al. (2016). Network modeling identifies patient-specific pathways in glioblastoma. Scientific Reports 29 (6): 28668. doi: 10.1038/srep28668

Wang JH, Urrutia-Cabrera D, Mesa Mora S, Nguyen T, Hung S et al. (2020). Functional study of the AMD-associated gene TMEM97 in retinal pigmented epithelium using CRISPR interference. BioRxiv. doi: 10.1101/2020.07.10.198143

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

Yu H, Kim PM, 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: e59. doi: 10.1371/journal.pcbi.0030059

Yu GI, Mun KH, Yang SH, Shin DH, Hwang JS (2018). Polymorphisms in the 3'-UTR of SCD5 gene are associated with hepatocellular carcinoma in Korean population. Molecular Biology Reports 45 (6): 1705-1714. doi: 10.1007/s11033-018-4313-6

Zamorano J, Wang HY, Wang R, Shi Y, Longmore GD et al. (1998). Regulation of cell growth by IL-2: role of STAT5 in protection from apoptosis but not in cell cycle progression. The Journal of Immunology 160 (7): 3502-3512.

Zhang H, Dong Y, Zhao H, Brooks JD, Hawthorn L et al. (2005). Microarray data mining for potential selenium targets in chemoprevention of prostate cancer. Cancer Genomics Proteomics 2 (2): 97-114.

Zhou J, Chng WJ (2013). Roles of thioredoxin binding protein (TXNIP) in oxidative stress, apoptosis, and cancer. Mitochondrion 13 (3): 163-169.

Supplementary Information

Table S1. DEG scores were given for each gene for day 3 according to the indicated formula and (A) the GSEA was performed to find the enriched GO BP_pathways with their enrichment scores. The highlighted results indicate DNA repair and oxidative-stress-related GO terms.

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

Table S1. (Continued).

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
REGULATION_OF_SIGNAL_TRANSDUCTION_BY_P53_CLASS_MEDIATOR	15	0.57	0.00	0.02
STEROID_METABOLIC_PROCESS	28	0.48	0.00	0.02
MEIOTIC_CELL_CYCLE_PROCESS	23	0.50	0.00	0.02
DNA_RECOMBINATION	31	0.46	0.01	0.02
MEIOTIC_CELL_CYCLE	28	0.47	0.01	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
DNA_CONFORMATION_CHANGE	35	0.43	0.01	0.03
COFACTOR_METABOLIC_PROCESS	30	0.44	0.01	0.03
ESTABLISHMENT_OF_PROTEIN_LOCALIZATION_TO_ORGANELLE	16	0.55	0.00	0.03
CELL_CYCLE_CHECKPOINT	21	0.48	0.01	0.04
MITOTIC_NUCLEAR_DIVISION	32	0.42	0.01	0.04
COENZYME_METABOLIC_PROCESS	15	0.52	0.01	0.04
RESPONSE_TO_LIGHT_STIMULUS	18	0.49	0.01	0.05
SMALL_MOLECULE_METABOLIC_PROCESS	96	0.33	0.00	0.05
SMALL_MOLECULE_BIOSYNTHETIC_PROCESS	49	0.38	0.00	0.05
CELLULAR_RESPONSE_TO_HORMONE_STIMULUS	32	0.42	0.01	0.06
CELL_CYCLE_ARREST	16	0.50	0.01	0.06
PROTEIN_DNA_COMPLEX_SUBUNIT_ORGANIZATION	22	0.46	0.01	0.06
NUCLEAR_CHROMOSOME_SEGREGATION	28	0.42	0.02	0.06
CELLULAR_PROCESS_INVOLVED_IN_REPRODUCTION_IN_MULTICELLULAR_ORGANISM	19	0.49	0.02	0.06
REGULATION_OF_DNA_METABOLIC_PROCESS	19	0.48	0.01	0.07

Table S2. The common transcription factors identified by GSEA analysis and associated with the regulation of genes of interest were marked by circles.

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