



Original Research

Bioinformatic analysis reveals CYP2C9 as a potential prognostic marker for HCC and liver cancer cell lines suitable for its mechanism study

Liu Shuaichen, Wang Guangyi*

Department of Hepatobiliary & Pancreas Surgery, The First Hospital, Jilin University, Changchun, Jilin 130021, China

Correspondence to: 8889571@sina.com

Received April 17, 2018; Accepted May 15, 2018; Published May 30, 2018

Doi: <http://dx.doi.org/10.14715/cmb/2018.64.7.12>

Copyright: © 2018 by the C.M.B. Association. All rights reserved.

Abstract: Hepatocellular carcinoma (HCC) is a common cancer and the sixth most lethal malignancy in the world. We chose gene expression profile of GSE14520 from GEO database aiming to find key genes that affect HCC progression. 22 paired tumor and non-tumor samples were included in this analysis. Differentially expressed genes (DEGs) between tumor and non-tumor were selected using GEO2R. Gene ontology (GO) enrichment and protein-protein interaction (PPI) of the DEGs were done using Metascape. There were 357 DEGs, including 70 up-regulated genes and 287 down-regulated genes. These DEGs were enriched in drug metabolic process, organic acid catabolic process, monocarboxylic metabolic process and etc. Three important modules were detected from PPI network using Molecular Complex Detection (MCODE) algorithm. Moreover, the Kaplan–Meier analysis for overall survival and disease-free survival were applied to those genes in top PPI group. In conclusion, this bioinformatic analysis demonstrated that DEGs, such as CYP2C9, might promote the development of HCC, especially in drug metabolism. It could also be used as a new biomarker for diagnosis.

Key words: HCC; CYPs; Prognosis; Metascape; Differentially expressed genes.

Introduction

Hepatocellular carcinoma is the sixth most common cancer and the third leading cause of cancer-related death worldwide (1). Some risk factors are believed to be related to the development of liver cancer. Excessive alcohol consumption, hepatitis B virus infection, exposure to aflatoxin and patient factors such as nonalcoholic fatty liver disease (NAFLD), type 2 diabetes and gender (2). Although significant progress has been made in the diagnosis and treatment of liver cancer, the five-year survival rate is only about 30% due to the high chemotherapy resistance of advanced liver cancer (3). Cytochrome P450 (CYPs) are an enzyme super family comprised of at least 57 distinct genes, localized to the endoplasmic reticulum (4,5). The major enzymes involved in drug metabolism, accounting for about 75% of the total metabolism (6). Most drugs will be deactivated by CYPs directly or facilitated to excrete from the body. CYPs can also activate many substances to form their active compounds. Changes in the expression of CYP genes will affect the efficiency of detoxification of xenobiotics, as well as messenger molecules that regulate the downstream signal transduction pathways, thereby producing conflicting effects in carcinogenesis (7).

Public available database like Gene Expression Omnibus (GEO), The Cancer Genome Atlas (TCGA) and the Human Protein Atlas (HPA) provide large amount of expression data for a variety of genes for human samples (8). In this study, we chose HCC GSE14520 dataset which contains 22 paired HCC tumor and non-tumor samples for differentially expressed genes (DEGs) ana-

lysis. The DEGs were further enriched for GO, Kyoto Encyclopedia of Genes and Genomes (KEGG) and PPI. Survival plot for the most significant enriched genes were drawn and the expression level of these genes were checked in TCGA and HPA. Another cell line database, the Cancer Cell Line Encyclopedia (CCLE), has genes expression profiles of around 1000 cancer cell lines (9). We use CCLE to select cell lines for further experiment. Our study is useful for elucidating the mechanism of HCC development.

Materials and Methods

Microarray data

GEO (<https://www.ncbi.nlm.nih.gov/geo/>) is a public and free database, containing microarray gene expression data. Gene expression profile of GSE14520 was chosen from GEO database. This dataset was based on Agilent GPL571 platform ([HG-U133A_2] Affymetrix Human Genome U133A 2.0 Array). The GSE14520 dataset cohort 1 included 44 samples, containing 22 HCC tumor samples and 22 non-tumor samples.

Data processing of DEGs

GEO2R (<https://www.ncbi.nlm.nih.gov/geo/geo2r/>) was used to detect differentially expressed genes between HCC tumor and non-tumor samples (10). GEO2R is an online tool that allows researchers to compare two or more groups of samples in a GEO Series. Its analysis speed is fast and the results can be downloaded for different analyzing purposes. The adjusted P values were utilized to reduce the false positive rate using Benjamini

and Hochberg false discovery rate method by default. The adjust P value < 0.05 and $|\log_{2}FC| \geq 2$ were set as the cut off criterion.

Gene ontology and KEGG pathway analysis of DEGs

Gene ontology analysis (GO) is a useful method to annotate high throughput genome or transcriptome data. Researchers can get hint of the grouping of DEGs, identifying the related biological processes (11). Kyoto Encyclopedia of Genes and Genomes (KEGG) is a database containing systematic analysis for genes, from a molecular network perspective. It involves most known metabolic and regulatory pathways and will annotate a list of genes according to these pathways (12) Metascape (<http://metascape.org/>) is a relatively new online tool that can carry out GO and KEGG enrichment. For each given gene list, all genes in the genome were used as the enrichment background. Terms with p-value < 0.01 , minimum count 3, and enrichment factor > 1.5 (enrichment factor is the ratio between observed count and the count expected by chance) are collected and grouped into clusters based on their membership similarities. More specifically, p-values are calculated based on accumulative hypergeometric distribution. Kappa scores were used as the similarity metric when performing hierarchical clustering on the enriched terms and then sub-trees with similarity > 0.3 are considered a cluster. The most statistically significant term within a cluster is chosen as the one representing the cluster (13).

Survival analysis of differentially expressed CYPs

Survival analysis of differentially expressed CYPs was carried out with GEPIA (14). GEPIA performs overall survival (OS) or disease-free survival (DFS) analysis based on gene expression. GEPIA uses Log-rank test (Mantel–Cox test) for hypothesis test. The cox proportional hazard ratio was included in the survival plot.

Comparison of the CYP2C9 expression level

The mRNA expression level of CYP2C9 was examined by GEPIA (14). GEPIA has the RNA sequencing expression data of 9,736 tumors and 8,587 normal samples from the TCGA and the GTEx projects. We compared the CYP2C9 mRNA level in this dataset and drew the boxplot with the website function. Immunohistochemistry (IHC) images of CYP2C9 expression in HCC tissues and in normal liver tissues were downloaded from the HPA (<http://www.proteinatlas.org/>) (15,16). The mRNA expression of CYP2C9 in cancer cell lines were examine using CCLE (<https://portals.broadinstitute.org/ccle>)(9).

Protein-protein Interaction Enrichment Analysis

PPI network was detected using Metascape (13). For each given gene list, protein-protein interaction enrichment analysis was carried out with the following databases: BioGrid, InWeb_IM, OmniPath. The resultant network contains the subset of proteins that form physical interactions with at least another list member. If the network contains 3 to 500 proteins, Molecular Complex Detection (MCODE) algorithm was further applied to identify densely connected network components.

Results

Identification of differentially expressed genes

There were 22 paired tumor and non-tumor samples in this study. We utilize GEO2R online analysis tool to detect the DEGs, using GSE14520. Adjusted P value < 0.05 and $|\log_{2}FC| \geq 2$ were used as cut-off criteria. A total of 357 differential expressed genes were detected, among which 70 were up-regulated and 287 were down-regulated.

GO function enrichment analysis

In order to understand the potential function of the DEGs, we performed GO function and KEGG pathway enrichment using Metascape. The enrichment summary was shown in Figure 1 and Table 1. The bar chart was colored by p-value. Across the input DEGs, the most significant enriched GO terms were drug metabolic process, organic acid catabolic process, monocarboxylic acid metabolic process and steroid metabolic process. The most significant enriched KEGG pathways were fatty acid degradation, complement and coagulation cascades, tyrosine metabolism and tryptophan metabolism. Interestingly, all the highly significant enriched results were metabolism related.

Protein-protein interaction analysis

PPI analysis was carried out by Metascape. Within the input DEGs, the top 3 interaction networks were Cytochrome P450 - arranged by substrate type (Figure 2A), Terminal pathway of complement (Figure 2B) and Resolution of Sister Chromatid Cohesion (Figure 2C). The genes related to each network were listed in Table 2. Cytochrome P450 (CYPs) was the most significant interaction network and involved CYPs included CYP1A1, CYP1A2, CYP2A6, CYP2A7, CYP2B6, CYP2C19, CYP2C8, CYP2C9, CYP2E1, CYP3A4, CYP26A1, CYP4F2, CYP39A1 and CYP3A43. All these genes were significantly down-regulated in tumors with more than 2 folds change, compared to non-tumor tissues. We

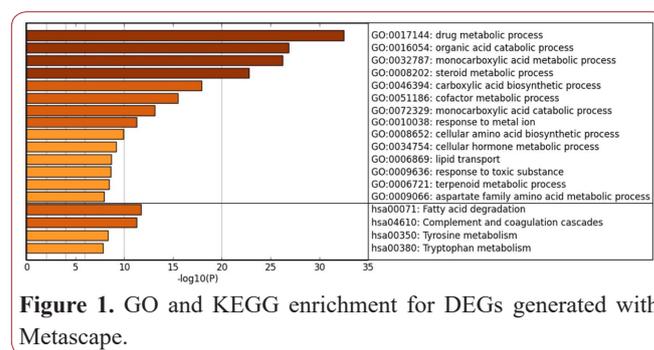


Figure 1. GO and KEGG enrichment for DEGs generated with Metascape.

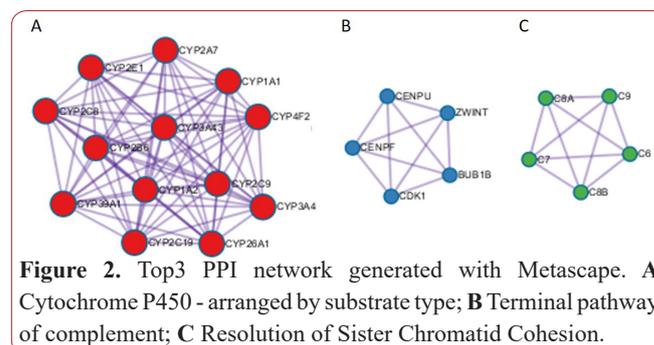


Figure 2. Top3 PPI network generated with Metascape. **A** Cytochrome P450 - arranged by substrate type; **B** Terminal pathway of complement; **C** Resolution of Sister Chromatid Cohesion.

Table 1. GO and KEGG enrichment for DEGs.

GO	Category	Description	Count	%	Log10(P)
GO:0017144	GO Biological Processes	drug metabolic process	60	23.4375	-32.49
GO:0016054	GO Biological Processes	organic acid catabolic process	35	13.671875	-26.88
GO:0032787	GO Biological Processes	monocarboxylic acid metabolic process	48	18.75	-26.25
GO:0008202	GO Biological Processes	steroid metabolic process	34	13.28125	-22.80
GO:0046394	GO Biological Processes	carboxylic acid biosynthetic process	33	12.890625	-17.92
GO:0051186	GO Biological Processes	cofactor metabolic process	35	13.671875	-15.52
GO:0072329	GO Biological Processes	monocarboxylic acid catabolic process	17	6.640625	-13.16
GO:0010038	GO Biological Processes	response to metal ion	23	8.984375	-11.30
GO:0008652	GO Biological Processes	cellular amino acid biosynthetic process	12	4.6875	-9.90
GO:0034754	GO Biological Processes	cellular hormone metabolic process	13	5.078125	-9.22
GO:0006869	GO Biological Processes	lipid transport	20	7.8125	-8.67
GO:0009636	GO Biological Processes	response to toxic substance	24	9.375	-8.63
GO:0006721	GO Biological Processes	terpenoid metabolic process	12	4.6875	-8.47
GO:0009066	GO Biological Processes	aspartate family amino acid metabolic process	9	3.515625	-7.96
hsa00071	KEGG Pathway	Fatty acid degradation	11	4.296875	-11.74
hsa04610	KEGG Pathway	Complement and coagulation cascades	13	5.078125	-11.28
hsa00350	KEGG Pathway	Tyrosine metabolism	8	3.125	-8.35
hsa00380	KEGG Pathway	Tryptophan metabolism	8	3.125	-7.86

Table 2. Top 3 PPI networks.

GO	Description	LogP	Z-score	Genes
R-HSA-211897	Cytochrome P450 - arranged by substrate type	-17.69	21.82	CYP1A1,CYP1A2,CYP2A6,CYP2A7,CYP2B6,CYP2C19,CYP2C8,CYP2C9,CYP2E1,CYP3A4,CYP26A1,CYP4F2,CYP39A1,CYP3A43
R-HSA-166665	Terminal pathway of complement	-9.42	22.78	C6, C7, C8A, C8B, C9
R-HSA-2500257	Resolution of Sister Chromatid Cohesion	-3.01	4.91	BUB1B, CDK1, CENPF, ZWINT, CENPU

can infer that the down regulation of CYPs has a potential role in HCC development.

Survival analysis for CYPs

Gene Expression Profiling Interactive Analysis (GEPIA) online survival tool was used to examine the prognostic information of the 13 CYPs in the PPI network. It was found that lower expression of CYP2C9 (HR 0.33, p=4.4E-5) was associated with better overall survival (OS) (Figure 3A) and lower expression of CYP2C9 (HR 0.44, p=2.2E-4) was associated with better disease-free survival (DFS) (Figure 3B). All the rest CYPs have nothing to do with survival. These results showed CYP2C9 might be a prognostic marker for HCC patient.

Expression level of Cytochrome P450

Both mRNA and protein expression level of CYP2C9

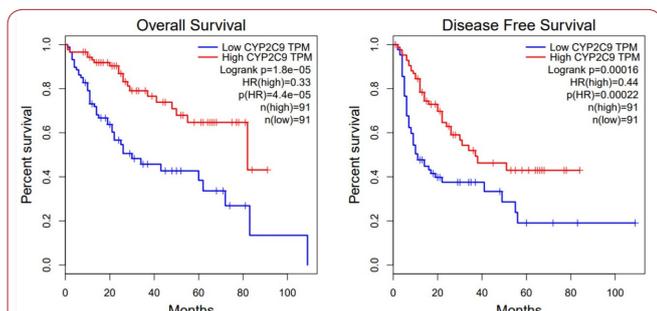


Figure 3. Overall survival and disease-free survival for CYP2C9.

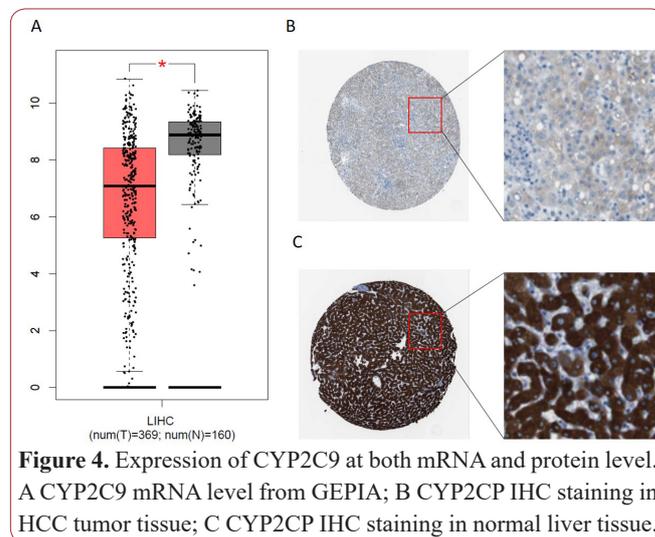


Figure 4. Expression of CYP2C9 at both mRNA and protein level. A CYP2C9 mRNA level from GEPIA; B CYP2C9 IHC staining in HCC tumor tissue; C CYP2C9 IHC staining in normal liver tissue.

were examined, using GEPIA and HPA database separately. Compared to normal liver sample, the mRNA level of CYP2C9 was significantly decreased in HCC tumor sample (Figure 4A). The protein level of CYP2C9 was also dramatically lower in HCC tumor (Figure 4B) than normal tissue (Figure 4C). The mRNA level in liver cancer cell lines was examined with the Cancer Cell Line Encyclopedia (CCLE). HUH1 has the highest CYP2C9 level and SKHEP1 has the lowest (Figure 5).

Discussion

CYPs are an enzyme super family comprised of at

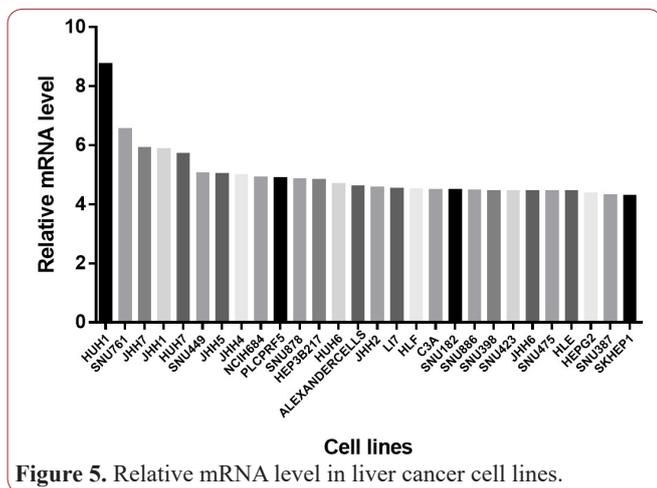


Figure 5. Relative mRNA level in liver cancer cell lines.

least 57 distinct genes, localized to the endoplasmic reticulum (4,5). The major enzymes involved in drug metabolism, accounting for about 75% of the total metabolism (6). Most drugs will be deactivated by CYPs directly or facilitated to excrete from the body. CYPs can also activate many substances to form their active compounds. Changes in the expression of CYP genes will affect the efficiency of detoxification of xenobiotics, as well as messenger molecules that regulate the downstream signal transduction pathways, thereby producing conflicting effects in carcinogenesis (7). CYP2C9 is one of the most abundant CYP2C proteins, accounting for 20% of liver CYP content, and contributes to the metabolism of many carcinogens and drugs (17,18).

A previous study about the CYPs expression was carried out in Chinese population with 93 end-stage liver disease samples. It was found that CYPs were generally decrease in end-stage liver disease, which was the same as in GSE14520. Among them, CYP2C9 had a about 50% decrease. The mRNA expression level was measured by real-time PCR (19). Besides expression differences, CYPs also have different clearance rate in cells. Previous studies showed that the clearance for CYP3A4 in HCC tumor tissues was significantly reduced compared with adjacent non-cancerous tissues (20). In patients with cirrhosis, the clearance of CYP3A4/5 and CYP2C19 were significantly decreased (21-23). The assessment of changes in CYPs clearance may be useful not only for the design of personalized liver cancer treatment, but also for the identification of patients with liver cancer who have other diseases for the treatment of drug dose regimens (24).

Previous DEGs enrichment analysis always show that cell cycle is the most affected pathways in HCC (25,26). However, in this study, we found that drug metabolic process was the most significant enriched term (-LogP=-32.49) and protein-protein interaction network among CYPs was most significantly affected (-LogP=-17.69). The local expression of CYPs in tumors is very important for their treatment, because these functionally related enzymes may be involved in the development of liver cancer, and in the determination of anti-cancer drug-sensitive tumors (27). In this study, we saw consistent down regulation of 13 CYPs (CYP1A1, CYP1A2, CYP2A6, CYP2A7, CYP2B6, CYP2C19, CYP2C8, CYP2C9, CYP2E1, CYP3A4,

CYP26A1, CYP4F2, CYP39A1 and CYP3A43). Then we checked how many of these CYPs associated with survival. Surprisingly, only low expression of CYP2C9 was associated with better OS (HR 0.33, $p=4.4E-5$) and DFS (HR 0.44, $p=2.2E-4$). Since our dataset was only 22 pairs of tumor and non-tumor samples, we examined the mRNA expression level of CYP2C9 using TCGA and GTEx samples. The result was consistent that tumor samples had lower CYP2C9 expression. Then we further checked HPA database for the protein expression level, tumor samples have much lower IHC staining intensity, which was accordant with mRNA expression.

Based on the expression and prognostic value of CYP2C9, it might have important roles in hepatocarcinogenesis. However, currently the mechanism studies for CYP2C9 and HCC were limited (4,24). By exploring CCLE database, we plot the mRNA expression of CYP2C9 in liver cancer cell lines. We can see HUH1 (8.72) has the highest CYP2C9 level, which might be suitable to knock down CYP2C9. SKHEP1 has the lowest (4.25) CYP2C9 level, nearly half of HUH1, might be suitable to overexpress CYP2C9. However, these cell line expression still need to be validated by real-time PCR.

In conclusion, we found that CYP2C9 had lower expression at both mRNA and protein level in HCC tumor samples. Lower expression of CYP2C9 was associated with better overall survival and disease-free survival. HUH1 and SKHEP1 cell line might be suitable for the mechanism study for CYP2C9 in HCC.

References

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA: a cancer journal for clinicians* 2015, 65, 87-108.
2. Parkin DM, The global health burden of infection-associated cancers in the year 2002. *International journal of cancer* 2006, 118, 3030-3044.
3. Tang A, Hallouch O, Chernyak V, Kamaya A, Sirlin CB. Epidemiology of hepatocellular carcinoma: Target population for surveillance and diagnosis. *Abdominal radiology (New York)* 2018, 43, 13-25.
4. Yu D, Green B, Marrone A, Guo Y, Kadlubar S, Lin D et al. Suppression of *cyp2c9* by microRNA hsa-mir-128-3p in human liver cells and association with hepatocellular carcinoma. *Scientific reports* 2015, 5, 8534.
5. Korobkova EA. Effect of natural polyphenols on *cyp* metabolism: Implications for diseases. *Chemical research in toxicology* 2015, 28, 1359-1390.
6. Guengerich FP. Cytochrome p450 and chemical toxicology. *Chemical research in toxicology* 2008, 21, 70-83.
7. Nebert DW, Dalton TP. The role of cytochrome p450 enzymes in endogenous signalling pathways and environmental carcinogenesis. *Nature reviews. Cancer* 2006, 6, 947-960.
8. Zhao X, Chen Y, Mao Q, Jiang X, Jiang W, Chen J et al. Overexpression of *ythdf1* is associated with poor prognosis in patients with hepatocellular carcinoma. *Cancer biomarkers: section A of Disease markers* 2018, 21, 859-868.
9. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S et al. The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 2012, 483, 603-607.
10. Davis S, Meltzer PS. Geoquery: A bridge between the gene ex-

pression omnibus (geo) and bioconductor. *Bioinformatics* (Oxford, England) 2007, 23, 1846-1847.

11. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM et al. Gene ontology: Tool for the unification of biology. The gene ontology consortium. *Nature genetics* 2000, 25, 25-29.

12. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. Kegg: Kyoto encyclopedia of genes and genomes. *Nucleic acids research* 1999, 27, 29-34.

13. Tripathi S, Pohl MO, Zhou Y, Rodriguez-Frandsen A, Wang G, Stein DA et al. Meta- and orthogonal integration of influenza "omics" data defines a role for ubr4 in virus budding. *Cell host & microbe* 2015, 18, 723-735.

14. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. Gepia: A web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic acids research* 2017, 45, W98-w102.

15. Interactive human protein atlas launches. *Cancer discovery* 2015, 5, 339.

16. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A et al. Proteomics. Tissue-based map of the human proteome. *Science (New York, N.Y.)* 2015, 347, 1260419.

17. Goldstein JA, de Morais SM. Biochemistry and molecular biology of the human cyp2c subfamily. *Pharmacogenetics* 1994, 4, 285-299.

18. Miners JO, Birkett DJ. Cytochrome p4502c9: An enzyme of major importance in human drug metabolism. *British journal of clinical pharmacology* 1998, 45, 525-538.

19. Chen H, Shen ZY, Xu W, Fan TY, Li J, Lu YF et al. Expression of p450 and nuclear receptors in normal and end-stage chinese livers. *World journal of gastroenterology* 2014, 20, 8681-8690.

20. Ye L, Yang X, Guo E, Chen W, Lu L, Wang Y et al. Sorafenib metabolism is significantly altered in the liver tumor tissue of hepatocellular carcinoma patient. *PloS one* 2014, 9, e96664.

21. Pentikainen PJ, Valisalmi L, Himberg JJ, Crevoisier C. Pharmacokinetics of midazolam following intravenous and oral administration in patients with chronic liver disease and in healthy subjects. *Journal of clinical pharmacology* 1989, 29, 272-277.

22. MacGilchrist AJ, Birnie GG, Cook A, Scobie G, Murray T, Watkinson G et al. Pharmacokinetics and pharmacodynamics of intravenous midazolam in patients with severe alcoholic cirrhosis. *Gut* 1986, 27, 190-195.

23. Andersson T, Olsson R, Regardh CG, Skanberg I. Pharmacokinetics of [14c] omeprazole in patients with liver cirrhosis. *Clinical pharmacokinetics* 1993, 24, 71-78.

24. Gao J, Zhou J, He XP, Zhang YF, Gao N, Tian X et al. Changes in cytochrome p450s-mediated drug clearance in patients with hepatocellular carcinoma in vitro and in vivo: A bottom-up approach. *Oncotarget* 2016, 7, 28612-28623.

25. Yin L, Cai Z, Zhu B, Xu C. Identification of key pathways and genes in the dynamic progression of hcc based on wgcna. *Genes* 2018, 9.

26. Yin L, Chang C, Xu C. G2/m checkpoint plays a vital role at the early stage of hcc by analysis of key pathways and genes. *Oncotarget* 2017, 8, 76305-76317.

27. Yan T, Lu L, Xie C, Chen J, Peng X, Zhu L et al. Severely impaired and dysregulated cytochrome p450 expression and activities in hepatocellular carcinoma: Implications for personalized treatment in patients. *Molecular cancer therapeutics* 2015, 14, 2874-2886.