The prognostic value of differentially expressed CYP3A subfamily members for hepatocellular carcinoma

Tingdong Yu, Xiangkun Wang, Guangzhi Zhu, Chuangye Han, Hao Su, Xiwen Liao, Chengkun Yang, Wei Qin, Ketuan Huang, and Tao Peng

Department of Hepatobiliary Surgery, The First Affiliated Hospital of Guangxi Medical University, Nanning, China

Objective: The activities of four cytochrome P3A (CYP3A) subfamily members (CYP3A4, CYP3A5, CYP3A7, and CYP3A43) are well documented in drug metabolism. However, the association between CYP3A subfamily members and hepatocellular carcinoma (HCC) remains unclear. This study investigated the prognostic value of CYP3A subfamily mRNA expression levels with HCC prognosis.

Materials and methods: Data from a total of 360 HCC patients were retrieved from The Cancer Genome Atlas database, and data from 231 HCC patients were retrieved from the Gene Expression Omnibus database. Kaplan–Meier analysis and Cox regression models were utilized to determine median survival, overall survival, and recurrence-free survival. Hazard ratios and 95% CI were calculated.

Results: Low expression of CYP3A4, CYP3A5, and CYP3A43 in the tumor tissue was associated with short median survival (crude p = 0.004, 0.001, and 0.0001; adjusted p = 0.022, 0.005, and 0.013, respectively). Joint-effects combination analysis of CYP3A4, CYP3A5/CYP3A4, CYP3A43/CYP3A5, and CYP3A43 revealed that high expression groups of two genes (group C, group c, group 3) were associated with a reduced risk of death, as compared to low expression of two genes (group A, group a, group 1), and the adjusted p values were 0.001, 0.004, and 0.001, respectively. Joint-effects analysis of CYP3A4, CYP3A5, and CYP3A43 showed that groups III and IV had a reduced risk of death, as compared to group I (adjusted p = 0.024 and 0.002, respectively).

Conclusion: CYP3A4, CYP3A5, and CYP3A43 mRNA expression levels are potential prognostic markers of HCC.

Keywords: mRNA expression, CYP3A subfamily, hepatocellular carcinoma, prognosis, biomarker

Introduction

Liver cancer is the second leading cause of cancer-related death worldwide with an estimated 782,500 new diagnoses and 745,500 deaths occurring worldwide in 2012. Of these cases, roughly 50% were reported in China alone.1 In 2015, liver cancer was the fourth most commonly diagnosed cancer in Chinese men.2 Hepatocellular carcinoma (HCC) is the most common type of liver cancer, accounting for about 85%–90% of cases, and has a very low 5-year survival rate of only 7%.3,4 Many etiological factors have been associated with HCC, including infection with hepatitis B and C viruses, cirrhosis, non-alcoholic fatty liver diseases, aflatoxin exposure, diabetes mellitus, obesity, excessive alcohol ingestion, hemochromatosis, and various other metabolic factors.5
Despite the significant benefits of many curative procedures (e.g., surgical resection, liver transplantation, transarterial chemoembolization, radiofrequency ablation, percutaneous ethanol injection, transarterial radiation, microwave ablation, and systemic therapy), the long-term survival rate of HCC remains unsatisfactory.\(^6\) Hence, further studies are required to identify new biomarkers to better assess survival and tumor progression in HCC.\(^6\)

The cytochrome P450 (CYP450) enzymes are a group of membrane-bound proteins that catalyze the oxidation of endobiotics and xenobiotics.\(^7\) The CYP3A subfamily is the most important group of enzymes of the CYP450 superfamily in humans as they metabolize a variety of clinically available drugs.\(^8\) The CYP3A subfamily consists of four differentially regulated members: CYP3A4, CYP3A5, CYP3A7, and CYP3A43.\(^7\) Expression levels of CYP3A5, CYP3A7, and CYP3A43 are usually lower than that of CYP3A4.\(^5\) CYP3A4 is the most predominately expressed CYP in the human liver in response to the exposure to several drugs.\(^8\) CYP3A7 is the major form of CYP in the fetal liver and its role has been explored in adverse drug reactions and interindividual differences in drug metabolism.\(^10\) CYP3A43 is the most recently identified member of the CYP3A subfamily and is, thus, less well studied than the other three members.\(^11\) Owing to its low expression levels in both the fetal and adult liver, CYP3A43 cannot be isolated by conventional protein purification approaches.\(^11\)

In addition, many investigations have explored the relationships between these genes and many cancers. Specifically, high expression of CYP3A4 may be associated with metastasis of Ewing’s sarcoma.\(^12\) CYP3A4 expression greater than the median level was associated with increased neuroblastoma mortality, and homozygous mutants of CYP3A5 *3/*3 were associated with a 4.3-fold greater risk of neuroblastoma mortality.\(^13\) CYP3A5 acts as a tumor suppressor gene in HCC via regulation of the mTORC2/Akt signaling pathway.\(^14\) CYP3A7 is reportedly overexpressed in HCC,\(^15,16\) and genetic variants of CYP3A7*1C have been associated with adverse outcomes in chronic lymphocytic leukemia (CLL), breast cancer, and lung cancer.\(^17\) Carriers of the CYP3A43 G-allele exhibited a significant 5-fold increase in mortality of early-onset prostate cancer.\(^18\)

Of the four members of the CYP3A subfamily, previous studies confirmed that only CYP3A5 and CYP3A7, but not CYP3A4 and CYP3A43, were associated with HCC. Therefore, the aim of the present study was to evaluate the prognostic values of the mRNA expression levels of all members of the CYP3A subfamily in HCC.

Materials and methods
Patient data collection
First, the Metabolic gEn RApid Visualizer database (http://merav.wi.mit.edu/) was accessed on October 15, 2017 to determine whether any of the four members of the CYP3A subfamily are differentially expressed between normal liver tissues and primary liver tumors.\(^19\) Then, the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database was accessed on September 10, 2017 to construct protein–protein interaction networks between CYP3A subfamily members and other proteins.\(^20\)

Then, the OncoLnc (http://www.oncolnc.org/; accessed October 15, 2017) and The Cancer Genome Atlas (TCGA; http://tcga-data.nci.nih.gov/tcga; accessed October 15, 2017) databases were accessed to obtain the expression levels of CYP3A4, CYP3A5, CYP3A7, and CYP3A43 at 50% cutoff values. The presented results are based, in part, on a previous study of the TCGA database.\(^21\)

Data from a total of 360 HCC patients, which included sex, race, age, body mass index, TNM stage, survival time, and survival status, were collected. Gene expression data were downloaded from the GSE14520 dataset of the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE14520; accessed October 15, 2017).\(^22\) The above dataset included gene expression levels that originated from the [HT_HG-U133A] Affymetrix HT Human Genome U133A\(^23\) and [HT_HG-U133A_2] Affymetrix HT Human Genome U133A_2\(^24\) arrays. To prevent batch effects, the former array, which had more patients and samples (231 patients and 455 tissues: 225 HCC tumor tissues and 220 liver tissues, respectively), was selected.

Enrichment analysis of the CYP3A subfamily
The online database Database for Annotation, Visualization, and Integrated Discovery (DAVID) ver. 6.7 (https://david-d.ncifcrf.gov/; accessed October 15, 2017)\(^24,25\) was employed for enrichment analysis. The database contains gene ontology (GO) functional analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis.\(^26\) In the present study, GO analysis was performed to obtain molecular function and cellular component terms, while KEGG pathways were drawn between CYP3A and other subfamilies.

Survival analysis
The data of 360 HCC patients were retrieved from the TCGA database and two groups were formed (n=180 each)
according to the 50% cutoff values. The median survival time (MST) was utilized to evaluate patient prognosis and a Cox regression model adjusted for TNM stage, age, and sex was used to identify factors significantly associated with HCC. In order to assure a rational comparison between the above two databases, the 50% cutoff was also used for the GEO database. Overall survival (OS) and recurrence-free survival (RFS) were applied to estimate patient prognosis. A Cox regression model was also adjusted for statistically significant factors, which included age, sex, alanine aminotransferase level, multinodular status, hepatitis B virus infection, main tumor size, alpha-fetoprotein (AFP), cirrhosis, and Barcelona Clinic Liver Cancer (BCLC) stage.

Joint-effects analysis of CYP3A4, CYP3A5, and CYP3A43

In the TCGA database, there were significant differences in the expression levels of only CYP3A4, CYP3A5, and CYP3A43 between tumor and non-tumor tissues. Joint-effects analysis was conducted with the following combinations: 1) CYP3A4 and CYP3A5; 2) CYP3A4 and CYP3A43; 3) CYP3A5 and CYP3A43; and 4) CYP3A4, CYP3A5, and CYP3A43.

There were three groups of CYP3A4 and CYP3A5 combinations according to the expression levels: group A (low CYP3A4 and low CYP3A5), group B (low CYP3A4/high CYP3A5 and high CYP3A4/low CYP3A5), and group C (high CYP3A4 and high CYP3A5).

Likewise, there were three groups of CYP3A4 and CYP3A43 combinations: group a (low CYP3A4 and low CYP3A43 expressions), group b (low CYP3A4/high CYP3A43 and high CYP3A4/low CYP3A43 expressions), and group c (high CYP3A4 and high CYP3A43 expressions).

There were three groups of CYP3A5 and CYP3A43 combinations: group 1 (low CYP3A5 and low CYP3A43 expressions), group 2 (low CYP3A5/high CYP3A43 and high CYP3A5/low CYP3A43 expressions), and group 3 (high CYP3A5 and high CYP3A43 expressions).

There were three groups of CYP3A4, CYP3A5, and CYP3A43 combinations: group I (low CYP3A4, low CYP3A5, and low CYP3A43); group II (high CYP3A4/low CYP3A5/low CYP3A43, low CYP3A4/high CYP3A5/low CYP3A43, and low CYP3A4/low CYP3A5/high CYP3A43); group III (high CYP3A4/high CYP3A5/low CYP3A43, high CYP3A4/low CYP3A5/high CYP3A43, and low CYP3A4/high CYP3A5/ high CYP3A43); and group IV (high CYP3A4, high CYP3A5, and high CYP3A43). The Cox regression model was adjusted for statistically significant factors (i.e., TNM stage, age, and sex) in keeping with the above combinations.

In the GEO database, only CYP3A45 and CYP3A43 were statistically significant. Joint-effects analysis was conducted with the following three combinations of CYP3A5 and CYP3A43: group i (low CYP3A5 and low CYP3A43); group ii (low CYP3A5/high CYP3A43 and high CYP3A5/low CYP3A43), and group iii (high CYP3A5 and high CYP3A43).

Statistical analysis

The Pearson correlation coefficient was used to identify correlations among the CYP3A4, CYP3A5, CYP3A7, and CYP3A43 genes. Correlation plots were depicted with R ver. 3.2.0 software (https://www.r-project.org/). Interaction networks among the aforementioned four genes were constructed with Cytoscape ver. 3.5.1 software (http://www.cytoscape.org/). Interaction networks between the four proteins of interest (i.e., CYP3A4, CYP3A5, CYP3A7, and CYP3A43) and other proteins were depicted using the STRING database of known and predicted protein–protein interactions (https://string-db.org/; accessed October 15, 2017). MST and probability (p) values were calculated with Kaplan–Meier survival analysis and the log-rank test. Univariate survival analyses were performed using the Cox hazards regression model. Scatter diagrams and survival curves were constructed with GraphPad Prism software ver. 7 (GraphPad Software, Inc., La Jolla, CA, USA). All statistical analyses were performed using SPSS software ver. 16 (SPSS Inc., Chicago, IL, USA).

Results

Baseline patient characteristics

According to the detailed characteristics of 360 HCC patients from the TCGA database (Table 1), only TNM stage was significantly associated with MST (p<0.001). The detailed characteristics of 231 patients from the GEO database, sex, multinodular status, main tumor size, BCLC stage, cirrhosis, and AFP were related to OS (p<0.048, p=0.003, p<0.001, p<0.001, p=0.004, and p=0.001, respectively), while sex, cirrhosis status, main tumor size, and BCLC stage were associated with RFS (p=0.001, p=0.019, p=0.020, and p<0.001, respectively; Table 2).

Analysis of CYP3A subfamily gene expression levels in tumor and non-tumor tissues

As shown by the box diagrams downloaded from an online website (Figure 1A–D, respectively), the expression levels of CYP3A4, CYP3A5, CYP3A7, and CYP3A43 were high in normal liver tissues and low in liver primary tumors. Scatter
diagrams from a search of the GEO database showed that all of the above four genes generated significant results between tumor and non-tumor tissues (all \( p < 0.0001 \); Figure 1E).

**Analysis of GO and KEGG pathways of the CYP3A subfamily**

GO analysis of the biological functions of *CYP3A4*, *CYP3A5*, *CYP3A7*, and *CYP3A43* returned the cellular component and the molecular function terms of “integral component of membrane”, “monooxygenase activity”, “oxidoreductase activity”, “iron ion binding”, and “heme binding” (Figure 1F). In the KEGG pathway analysis, DAVID determined the associations between CYP3A subfamily members and other genes.

In the metabolism of aromatic hydrocarbons, benzo(a)pyrene and B(a)P-4,5-oxide are metabolized by CYP3A4 into B(a)P-7,8-oxide and B(a)P-9,10-oxide, respectively, and finally transformed into the DNA adduct (+)-trans-BPDE-N2-dG, which is a known cause of cancers of the skin, lung, and stomach. In the metabolism of azo dyes, Sudan I is metabolized by CYP3A subfamily members into the benzenediazonium ions naphthalene-1,2-diol, 4’-OH-Sudan I and 6-OH-Sudan I, and finally transformed into 8-phenylazo-guanine in DNA and DNA, RNA, and protein adducts, which have been associated with the occurrence of liver and bladder cancers. In the metabolism of natural carcinogens, aflatoxin B1 is transformed by CYP3A4 into AFB1-exo-8,9-epoxide, which can be metabolized to the DNA adducts AFB1-N7-Gua and AFB1-FAPPY, which have both been associated with the occurrence of lung and liver cancers (Figure 2).

**Correlation analysis and interaction analysis among CYP3A subfamily members**

Pearson correlation coefficients among the four CYP3A members were calculated. In the TCGA database, *CYP3A4* was negatively correlated with *CYP3A7* (\( r = -0.12, p < 0.05 \)), but not significantly associated with *CYP3A5* (\( r = 0.07, p > 0.05 \)). The other genes were positively and significantly correlated with each other (all \( p < 0.05 \); Figure 3A). In the GEO database, all four genes were positively and statistically significantly correlated with the three other genes (all \( p < 0.05 \); Figure 3B).

The gene–gene interactions between CYP3A subfamily members and other genes were further analyzed. Four genes were associated with other CYP subfamily members (*CYP1A2*, *CYP2A6*, *CYP2A7*, *CYP2D6*, *CYP2C8*, *CYP2C9*, *CYP2C18*, *CYP2C19*, *CYP2D6*, *CYP2J2*, *CYP3A7–CYP3A51P*, and *CYP51A1*) and other genes, including *TM4SF5*, *XDH*, *SLC6A4*, *UGT2B7*, *UGT1A6*, and so on (Figure 3C). Moreover, protein–protein interaction networks...
Table 2 Demographic and clinical characteristics of 231 HCC patients in GEO database

<table>
<thead>
<tr>
<th>Variables</th>
<th>Patients (n=231)</th>
<th>Overall survival</th>
<th>Recurrence–free survival</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MST (months)</td>
<td>HR (95% CI)</td>
<td>Log-rank test, p</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>191</td>
<td>NA</td>
<td>Ref.</td>
</tr>
<tr>
<td>Female</td>
<td>30</td>
<td>NA</td>
<td>0.59 (0.34–1.00)</td>
</tr>
<tr>
<td>Missing</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Age, years</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤60</td>
<td>181</td>
<td>NA</td>
<td>Ref.</td>
</tr>
<tr>
<td>&gt;60</td>
<td>40</td>
<td>NA</td>
<td>0.96 (0.65–1.44)</td>
</tr>
<tr>
<td>Missing</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>HBV status</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AVR-CC</td>
<td>56</td>
<td>NA</td>
<td>Ref.</td>
</tr>
<tr>
<td>CC+NO</td>
<td>162</td>
<td>NA</td>
<td>0.78 (0.56–1.09)</td>
</tr>
<tr>
<td>Missing</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ALT</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤50 U/L</td>
<td>130</td>
<td>NA</td>
<td>Ref.</td>
</tr>
<tr>
<td>&gt;50 U/L</td>
<td>91</td>
<td>NA</td>
<td>1.06 (0.78–1.44)</td>
</tr>
<tr>
<td>Missing</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Main tumor size</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤5 cm</td>
<td>140</td>
<td>NA</td>
<td>Ref.</td>
</tr>
<tr>
<td>&gt;5 cm</td>
<td>80</td>
<td>53</td>
<td>1.87 (1.38–2.55)</td>
</tr>
<tr>
<td>Missing</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Multinodular</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>45</td>
<td>48</td>
<td>Ref.</td>
</tr>
<tr>
<td>No</td>
<td>176</td>
<td>NA</td>
<td>0.59 (0.42–0.84)</td>
</tr>
<tr>
<td>Missing</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cirrhosis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>203</td>
<td>NA</td>
<td>Ref.</td>
</tr>
<tr>
<td>No</td>
<td>18</td>
<td>NA</td>
<td>0.23 (0.09–0.63)</td>
</tr>
<tr>
<td>Missing</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>BCLC stage</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0+A</td>
<td>168</td>
<td>NA</td>
<td>Ref.</td>
</tr>
<tr>
<td>B+C</td>
<td>51</td>
<td>20</td>
<td>3.63 (2.64–5.00)</td>
</tr>
<tr>
<td>Missing</td>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>AFP</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤300 ng/mL</td>
<td>100</td>
<td>NA</td>
<td>Ref.</td>
</tr>
<tr>
<td>&gt;300 ng/mL</td>
<td>118</td>
<td>NA</td>
<td>1.67 (1.23–2.27)</td>
</tr>
<tr>
<td>Missing</td>
<td>13</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: *Missing, information of sex, age, ALT, multinodular status, cirrhosis was unavailable in 10 patients. *Missing, information of HBV status and AFP was unavailable in 13 patients. *Missing, information of main tumor size was unavailable in 11 patients. *Missing, information of BCLC stage was unavailable in 12 patients. CC+NO contains two categories of people with CC and not HBV infection status (NO). Bold font indicates p≤0.05.

Abbreviations: AFP, alpha fetoprotein; ALT, alanine aminotransferase; AVR-CC, active viral replication chronic carrier; BCLC, Barcelona Clinic Liver Cancer; CC, chronic carrier; GEO, Gene Expression Omnibus; HBV, hepatitis B virus; HCC, hepatocellular carcinoma; HR, hazard ratio; MST, median survival time; Ref., reference; NA, not available.

drawn by STRING showed that six CYP family member proteins (CYP1A1, CYP1A2, CYP2B6, CYP2C8, CYP2C9, and CYP2E1) were also associated with CYP3A4, CYP3A5, CYP3A7, and CYP3A43 (Figure 3D).

Survival analysis of CYP3A subfamily members

A multivariate Cox regression model adjusted for prognostic-related characteristics in the TCGA database (i.e., sex, age, and TNM stage) revealed significant relationships between MSTs and the CYP genes CYP3A4, CYP3A5, and CYP3A43 (adjusted p=0.022, hazard ratio [HR]=0.64, 95% CI=0.44–0.94; adjusted p=0.005, HR=0.58, 95% CI=0.39–0.84; and adjusted p=0.013, HR=0.61, 95% CI=0.41–0.90, respectively; Table 3). A multivariate Cox regression model adjusted for prognostic-related characteristics in the GEO database (i.e., sex, age, hepatitis B virus, alanine aminotransferase, main tumor size, multinodular status, BCLC stage, AFP, and cirrhosis status) showed that CYP3A5 was significantly associated with OS and RFS (adjusted p=0.001, HR=0.59, 95% CI=0.42–0.81; adjusted p=0.017, HR=0.73, 95% CI=0.56–0.94 respectively; Table 4) and CYP3A43 was
significantly associated with OS (adjusted $p=0.046$, HR=0.73, 95% CI=0.53–0.99; Table 4).

Survival curves of the genes of interest in the TCGA database are presented in Figure 4A–D and survival curves of these genes in the GEO database are shown in Figure 5A–F. In addition, scatter diagrams of the expression levels of these genes in both databases are shown in Figure 4E, F.

Joint-effects analysis of CYP3A subfamily members

Analysis of CYP3A4 and CYP3A5 combinations in the TCGA database showed that group A had the poorest MST of 931 days (adjusted $p=0.005$) and group C had the best MST of 2456 days (adjusted $p=0.001$). In regard to the effects of the CYP3A4 and CYP3A43 combinations in the TCGA database, group a had the poorest MST of 931 days (adjusted $p=0.011$) and group b had the best MST of 2531 days (adjusted $p=0.061$). Surprisingly, MST was not determined for group c, which contained the most favorable patient factors, possibly due to the influence of other potential elements. Analysis of the CYP3A5 and CYP3A43 combinations in the TCGA database showed that group 1 had the poorest MST of 931 days (adjusted $p=0.003$) and group 3 had the best MST of 2456 days (adjusted $p=0.001$). Detailed joint-effects analysis results are shown in Table 5 and associated survival curves are shown in Figure 6A–C.

Analysis of the CYP3A4, CYP3A5, and CYP3A43 combinations in the GEO database revealed that group I had the poorest MST of 931 days (adjusted $p=0.004$) and group IV had the best MST of 2456 days (adjusted $p=0.002$; Table 6). Associated survival curves are shown in Figure 6D.

Analysis of the CYP3A5 and CYP3A43 combinations in the GEO database showed that group i had the poorest MST of 58 months (adjusted $p=0.014$). MSTs were not calculated

Figure 1 Gene expression levels of CYP3A4 (A), CYP3A5 (B), CYP3A7 (C), and CYP3A43 (D) in normal liver tissue and primary liver tumors. Expression levels in the GEO database (E) and GO analysis (F) of the four genes.

Abbreviations: CYP3A, cytochrome P3A; GEO, Gene Expression Omnibus; GO, gene ontology.
for the other groups. Detailed joint-effects analysis results are shown in Table 5 and associated survival curves are shown in Figure 6E.

**Discussion**

The present investigation of the associations between the gene expression levels of the CYP3A subfamily members in the TCGA database showed that low expression levels of *CYP3A4*, *CYP3A5*, and *CYP3A43* were associated with a poor prognosis of HCC, while in the GEO database, low expression levels of *CYP3A5* were associated with a poor prognosis of HCC. Joint-effects analysis of the aforementioned three genes in the TCGA database showed that groups with the poorest prognostic factors had the poorest prognosis. Thus, the expression levels of *CYP3A4*, *CYP3A5*, and *CYP3A43* both alone and in combination may serve as potential biomarkers of HCC.

It is well established that the members of the CYP3A subfamily are predominantly associated with drug metabolism. It has been reported that up to 37% of drugs are metabolized by CYP3A subfamily members. The expression levels of CYP3A5, CYP3A7, and CYP3A43 are usually lower than that of CYP3A4. CYP3A7 expression is higher in the fetal liver than in the adult liver; whereas CYP3A43 is hardly detectable. CYP3A4 and CYP3A5 have been reported as a mycobacterial for the treatment of multidrug-resistant tuberculosis. CYP3A4 and CYP3A5 have been reported to be involved in the metabolism of bedaquiline, a mycobacterial drug. CYP3A4, together with CYP2C8 and CYP2C19, is involved in the metabolism of bedaquiline, a mycobacterial drug. CYP3A4 and CYP3A5 have been reported to be involved in the metabolism of bedaquiline, a mycobacterial drug.
locus of CYP3A is reportedly associated with age at menarche and breast cancer risk.\textsuperscript{32} CYP3A4 expression is related to breast cancer development\textsuperscript{13} and an increased risk of prostate cancer,\textsuperscript{14} and is a new biomarker for predicting poor prognosis of HCC.\textsuperscript{35} High expression of CYP3A4 may be associated with metastasis of Ewing’s sarcoma,\textsuperscript{12} and a significant association was reported between the single-nucleotide polymorphism rs2246709 of CYP3A4 and survival of patients with acute lymphoblastic leukemia.\textsuperscript{36} A linkage between CYP3A4 and CYP3A5 was found to increase the risk of prostate cancer.\textsuperscript{37} CYP3A5 acts as a tumor suppressor in HCC via regulation of the mTORC2/Akt signaling pathway.\textsuperscript{14} Also, a mild association of CYP3A4/CYP3A5 genotypes and expression levels with neuroblastoma has been reported.\textsuperscript{13} The results of a meta-analysis indicated that the CYP3A5*3 genetic polymorphism may play a significant role in the development of both acute and chronic leukemia, as well as colorectal cancer, especially among Asian and Caucasian

\textbf{Figure 3} Matrix graphs of Pearson correlations of CYP3A4, CYP3A5, CYP3A7, and CYP3A43 mRNA expression levels in TCGA database (A) and GEO database (B). Gene–gene interaction networks among the four genes of interest with other genes (C) and protein–protein interaction networks among the four proteins of interest with other proteins (D).

\textbf{Abbreviations:} CYP3A, cytochrome P3A; GEO, Gene Expression Omnibus; TCGA, The Cancer Genome Atlas.

\textbf{Note:} *p≤0.05, **p≤0.01, ***p≤0.001.
The CYP3A4*3 genetic polymorphism may play a role in the risk of childhood acute lymphoblastic leukemia. The single-nucleotide polymorphism rs45446698 of CYP3A4 was found to be associated with breast cancer mortality (HR=1.74, p=0.009), all-cause mortality in lung cancer patients (HR=1.43, p=0.03), and CLL (HR=1.62, p=0.03). The CYP3A4*1C allele, which leads to adult expression of the fatal CYP3A7 gene, is likely to be a functional allele that influences the levels of circulating endogenous sex hormones and the subsequent outcomes of CLL, breast cancer, and lung cancer. CYP3A7 was found to be overexpressed in HCC. An increased probability of developing prostate cancer was observed in the G allele of rs2740574 and the C allele of rs501275 of CYP3A43 in a combined ethnic group analysis.

The results of the present study demonstrated that the expression levels of CYP3A4, CYP3A5, and CYP3A43 are associated with HCC. Moreover, low expression levels of CYP3A4, CYP3A5, and CYP3A43 are linked to a poor prognosis of HCC, which is consistent with the findings of Jiang et al who reported that CYP3A4 acts as a tumor suppressor gene in HCC via regulation of the mTORC2/Akt signaling pathway. Although CYP3A7 was not associated with HCC in the two independent datasets employed in this study, the expression levels of CYP3A7 in HCC tumor and non-tumor tissues, as demonstrated by statistically significant p values,
suggesting that both may serve as tumor suppressor genes in HCC. The expression levels of CYP3A5 and CYP3A43 were consistently significantly different in the TCGA and GEO databases, while only CYP3A4 was significantly different in the TCGA database. This phenomenon may partly be explained as the standardized expression levels of genes in the GEO database, and the adjusted $p$ value of CYP3A4 was the nearest to the significant $p$ cutoff value. To the best of our knowledge, this is the first report of these findings, thus further validations are warranted in other populations.

Meanwhile, in the metabolism of aromatic hydrocarbons, benzo(a)pyrene is metabolized by CYP3A4 and finally transformed into DNA adducts, which are known to induce cancers of the skin, lung, and stomach. In azo dye metabolism, Sudan I is metabolized by CYP3A subfamily members and finally into DNA, RNA, and protein adducts, which could lead to the onset of liver and bladder cancers. In the metabolism of natural carcinogens, aflatoxin B1 is transformed by CYP3A4 and the final metabolite could lead to oncogenesis of the lung and liver. The gene–gene interaction networks showed that CYP3A4 is related to CYP1A2 in predicted, co-expression, and shared protein domains aspects; CYP3A5 and CYP3A7 are related to CYP1A2 in shared protein domains aspects; and CYP3A43 is related to CYP1A2 in co-expression and shared protein domains aspects. The protein–protein interaction networks showed that CYP3A4 is linked to CYP1A2 in text-mining, already

Figure 4 Kaplan–Meier survival curves of the CYP3A4 (A), CYP3A5 (B), CYP3A7 (C), and CYP3A43 (D) genes in TCGA database. Scatter plots of CYP3A4, CYP3A5, CYP3A7, and CYP3A43 mRNA expression levels in TCGA database (E) and GEO database (F). Abbreviations: CYP3A, cytochrome P3A; GEO, Gene Expression Omnibus; TCGA, The Cancer Genome Atlas.
known interaction, co-expression, gene co-occurrence, and protein homology aspects, while CYP3A5 and CYP3A7 are linked to CYP1A2 in text-mining, already known interaction, gene co-occurrence, and protein homology aspects. In the metabolism of aromatic amines and amides, IQ and MeIQx are metabolized by CYP1A2 and finally transformed into DNA adducts, which have been linked to cancers of the liver, colon, lung, and breast. Given the foregoing results, we speculate that CYP3A4, CYP3A5, and CYP3A43 may be related to HCC due to the complicated relationships with substrate metabolism and gene expression, including CYP1A2 expression and enzyme activity.

However, there were some limitations to this study that should be addressed. First, larger population studies are required to increase the credibility of the present findings. Second, other potential prognosis-related factors regarding
Table 5  Joint-effects analysis of the combinations of CYP3A4 and CYP3A5, CYP3A4 and CYP3A43, and CYP3A5 and CYP3A43 in two databases

<table>
<thead>
<tr>
<th>Databases</th>
<th>Group</th>
<th>CYP3A4 expression</th>
<th>CYP3A5 expression</th>
<th>CYP3A43 expression</th>
<th>Samples</th>
<th>MST</th>
<th>Crude HR (95% CI)</th>
<th>Adjusted HR (95% CI)</th>
<th>Adjusted p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCGA</td>
<td>A</td>
<td>Low</td>
<td>Low</td>
<td>Ref.</td>
<td>108</td>
<td>931</td>
<td>&lt;0.001</td>
<td>Ref.</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>Low</td>
<td>High</td>
<td>Ref.</td>
<td>144</td>
<td>1622</td>
<td>0.38 (0.23–0.61)</td>
<td>0.036 (0.50–1.18)</td>
<td>0.229</td>
</tr>
<tr>
<td></td>
<td></td>
<td>High</td>
<td>High</td>
<td>Ref.</td>
<td>108</td>
<td>2456</td>
<td>0.38 (0.23–0.61)</td>
<td>0.036 (0.50–1.18)</td>
<td>0.229</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Low</td>
<td>High</td>
<td>Ref.</td>
<td>124</td>
<td>931</td>
<td>&lt;0.001</td>
<td>Ref.</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td>High</td>
<td>High</td>
<td>Ref.</td>
<td>112</td>
<td>2531</td>
<td>0.38 (0.23–0.61)</td>
<td>0.036 (0.50–1.18)</td>
<td>0.229</td>
</tr>
<tr>
<td>GEO</td>
<td>i</td>
<td>Low</td>
<td>Low</td>
<td>Ref.</td>
<td>176</td>
<td>58</td>
<td>0.35 (0.22–0.57)</td>
<td>0.049 (0.34–0.88)</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td>ii</td>
<td>Low</td>
<td>High</td>
<td>Ref.</td>
<td>84</td>
<td>NA</td>
<td>0.35 (0.22–0.57)</td>
<td>0.049 (0.34–0.88)</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td>iii</td>
<td>High</td>
<td>High</td>
<td>Ref.</td>
<td>171</td>
<td>NA</td>
<td>0.35 (0.22–0.57)</td>
<td>0.049 (0.34–0.88)</td>
<td>0.049</td>
</tr>
</tbody>
</table>

Notes: *Adjustment for sex, age, TNM stage in TCGA database and adjustment for sex, age, HBV status, ALT, main tumor size, multinodular status, cirrhosis, AFP, and BCLC stage in GEO database. Bold fonts indicate \( p \leq 0.05 \). Each group stands for a combination in a row or combinations in two continuous rows. Detailed descriptions are shown in “Joint-effects analysis of CYP3A4, CYP3A5, and CYP3A43” section.

Abbreviations: AFP, alpha fetoprotein; ALT, alanine aminotransferase; BCLC, Barcelona Clinic Liver Cancer; CYP3A, cytochrome P3A; GEO, Gene Expression Omnibus; HBV, hepatitis B virus; HR, hazard ratio; MST, median survival time; TCGA, The Cancer Genome Atlas.

Figure 6 Survival curves of the joint-effects analysis of the combination of CYP3A4 and CYP3A5 (A), CYP3A4 and CYP3A43 (B), CYP3A5 and CYP3A43 (C), and CYP3A4, CYP3A5, and CYP3A43 (D) in TCGA database and CYP3A5 and CYP3A43 (E) in GEO database.

Abbreviations: CYP3A, cytochrome P3A; GEO, Gene Expression Omnibus; TCGA, The Cancer Genome Atlas.
tumor evolution and prognosis, including drinking status, smoking status, cirrhosis status, Child–Pugh score, BCLC stage, tumor numbers, main tumor size, tumor capsule status, vascular invasion, AFP levels, and antiviral therapy, should be included in future analysis to better evaluate the relationships between CYP3A subfamily members and HCC prognosis. Third, further well-designed studies concentrating on functional validation are warranted with a greater number of research centers and more racially diverse countries.

To summarize, the results of the present study demonstrated that CYP3A4, CYP3A5, and CYP3A43 present potential serum biomarkers for the early diagnosis of HCC and combination analysis revealed significant interactions that could serve as better prognostic indicators of HCC. However, because of the incomplete clinical data and small sample size in this study, further studies are needed to validate these findings.

**Acknowledgments**

This work was supported in part by the National Nature Science Foundation of China (Nos. 81560535, 81072321, 30760243, 30460143, and 30560133), 2009 Program for New Century Excellent Talents in University, Guangxi Nature Sciences Foundation (No. GuiKeGong 1104003A-7), and Guangxi Health Ministry Medicine Grant (Kcy-Scientific Research-Grant Z201018), and self-funded Scientific Research Fund of the Health and Family Planning Commission of Guangxi Zhuang Autonomous Region (Z2016318). The authors also acknowledge the support by the National Key Clinical Specialty Programs (General Surgery and Oncology) and the Key Laboratory of Early Prevention and Treatment for Regional High-Incidence-Tumor (Guangxi Medical University), Ministry of Education, China. The authors thank Prof Cheryl A Winkler and Stephen J O’Brien for their invaluable help to our study.

**Disclosure**

The authors report no conflicts of interest in this work.

**References**


