Research Article

Seven Key Genes Correlated with Immune Infiltration Reveal Prognostic Significance of Hepatocellular Carcinoma

Tian M and Dong L*

Department of Gastroenterology and Hepatology, Shanghai Institute of Liver Diseases, Zhongshan Hospital of Fudan University, Shanghai, China

*Corresponding author: Ling Dong, Department of Gastroenterology and Hepatology, Shanghai Institute of Liver Diseases, Zhongshan Hospital of Fudan University, Shanghai, 200032, China

Abstract

Background: The poor prognosis of Hepatocellular Carcinoma (HCC) is mainly due to late diagnosis, rapid progression and high recurrence rate. Reliable biomarkers for the diagnosis and prognosis of HCC are urgently needed.

Methods: We selected four public datasets from the GEO database to identify Differentially Expressed Genes (DEGs) and Differentially Expressed miRNAs (DE miRNAs). GO functional annotation, KEGG pathway enrichment analysis, and a Protein-Protein Interaction (PPI) network were constructed to explore the functions and importance of DEGs. To determine the target genes of 33 DE miRNAs screened from GSE10694, the miRNet WebServer was utilized. The key genes were screened out by mRNA-microRNA interaction analysis. Then, those highly expressed genes were verified in parallel databases (ONCOMINE, GEPIA and HPA databases). Further prognostic analysis by Kaplan Meier and diagnostic analysis based on TCGA were conducted. Furthermore, we investigated the association between key genes and immune infiltration in HCC tissues using the TIMER database.

Results: We identified seven key genes, including CCNA2, DLGAP5, MAD2L1, MELK, NCAPG, PRC1, and RRM2 based on public databases. The overexpression of these key genes has been demonstrated in HCC tissues and is associated with advanced disease and poor prognosis of patients with HCC. Furthermore, we found these key genes affect various of infiltrating immune cells and be positively correlated with the expression of gene markers of immune escape in HCC.

Conclusion: These seven key genes might be used as biomarkers for the diagnosis, prognosis, and prediction response to immunotherapy for patients with HCC, as well as the therapeutic targets of HCC.

Keywords: Hepatocellular carcinoma; Key genes; Immune infiltration; Immunotherapy; Biomarker

Abbreviations

AFP: α-Fetoprotein; ANTTs: Adjacent Non-Tumor Tissues; AUC: Area Under the Curve; BP: Biological Process; BRD9: Bromodomain-Containing Protein 9; CC: Cellular Component; DCs: Dendritic Cells; DEGs: Differentially Expressed Genes; DE miRNAs: Differentially Expressed miRNAs; DFS: Disease-Free Survival; FC: Fold Change; GEO: Gene Expression Omnibus; GO: Gene Ontology; GPC3: Glypican-3; HCC: Hepatocellular Carcinoma; HPA: Human Protein Atlas; LIHC: Liver Hepatocellular Carcinoma; MCODE: Molecular Complex Detection; MF: Molecular Function; OS: Overall Survival; PFS: Progression-Free Survival; PPI: Protein-Protein Interaction; RNAi: RNA interference; ROC: Receiver Operating Characteristic; TAMs: Tumor-Associated Macrophages; TCGA: The Cancer Genome Atlas; TILs: Tumor-Infiltrating Lymphocytes; TIMER: Tumor Immune Estimation Resource

Introduction

Liver cancer, of which Hepatocellular Carcinoma (HCC) accounts for almost 85-90 %, is the fourth leading cause of cancer-related deaths and the second most lethal tumor with 42,810 estimated new cases and 30,160 deaths in 2020 [1]. Despite clinical improvements in the therapeutic strategies for HCC, such as chemotherapy, targeted therapy, radiotherapy and immunotherapy, the prognosis remains unsatisfactory, with the 5-year-relative survival rate as low as 18%, mainly due to the concealed symptoms and late diagnosis [2-4]. Currently, the clinical screening and diagnosis of HCC mainly relies on imaging methods (computed tomography and ultrasound) and serum biomarkers. Nevertheless, imaging diagnosis of Small HCC (S-HCC) remains challenging. Besides, α-fetoprotein (AFP), the most used serological biomarker, has a sensitivity of only 65% for the diagnosis of clinical HCC, while its prediction for preclinical HCC is less than 40% [5]. Thus, there is an urgent need to identify more accurate molecular biomarkers for prognosis and diagnosis of HCC.

Progression of HCC is considered a multi-step process, in which gene aberrations, cellular context, and environmental factors play a role in the initiation, evolution, and metastasis of the tumor [6].

Studies have been performed to define the genetic mechanisms and pathways involved in HCC, revealing that many molecules such as HepatoScore-14 [7], Bromodomain-Containing Protein 9 (BRD9) [8], Glypican-3 (GPC3) [9], and METTL14 [10] have a significant association with the initiation or/and progression of HCC, and could serve as diagnostic or prognostic biomarkers. However, most of the existing studies are single-center and have been limited by the sample size and validation.

The liver can stimulate immune responses to prevent the invasion of harmful pathogens and the origination of tumors. On the other hand, immune escape plays an indispensable role in HCC [11]. The proliferation of Tumor-Infiltrating Lymphocytes (TILs) is considered representative of the host's defense to cancer. However, a growing body of evidence reveals that high levels of TILs cannot control the development of the tumor. [12]. Notably, immune checkpoint pathways, including enrichment of many inhibitory co-stimulatory molecules, such as PD-1, LAG-3, CTLA-4, and TIM-3, have been shown to reflect Cytotoxic T Lymphocyte (CTLs) exhaustion [13-15]. Clinical trials using immune checkpoint inhibitors in HCC, such as tremelimumab (anti-CTLA4), nivolumab (anti-PD1), and durvalumab (anti-PDL1), have shown positive responses. Unfortunately, the therapeutic effect is very limited in some patients with HCC [16]. Therefore, markers and drug targets are urgently needed to predict and improve the efficacy of immunotherapy for HCC.

In recent years, with the development of genome-wide spectrum analysis, our knowledge of the genetic background and driving pathways resulting in HCC has dramatically improved [6]. Microarray technology, the monitoring of genome-wide expression levels of genes in specific organisms, has become an indispensable tool for the classification, diagnosis, prognosis, and detection of therapeutic targets for different types of tumors. In addition, combined with bioinformatics, it has been widely used to describe overall gene expression patterns of multiple types of tumors to obtain a better understanding of the mechanisms behind carcinogenesis [17-19]. In our study, four public datasets consisting of 498 HCC tissues and 357 paired adjacent non-tumor tissues (ANTTs) were selected from the Gene Expression Omnibus (GEO) database [20] to identify reliable biomarkers for HCC diagnosis and prognosis prediction. We also added a series of in-depth analyses based on parallel databases (2,481 samples) to validate the expression of the key genes. The association of key genes with immune infiltration levels and the expression of gene markers in T cell exhaustion was analyzed in HCC.

Materials and Methods

Data source and identification of DEGs and DE miRNAs

We searched 'Hepatocellular carcinoma' in the GEO database, a global public library for archiving and free distribution of high-throughput gene expression and other functional genomics datasets [20], with filters activated for 'DataSets', 'Homo sapiens', and 'Expression profiling by array,' to obtain the expression data of mRNA and miRNA of HCC tissues. Finally, three gene expression profiles (GSE54236 based on GPL6480, GSE14520 based on GPL3921, GSE76427 based on GPL10558) and a miRNA expression profile (GSE10694 based on GPL6542) were retrieved after a systematic review. All data were freely accessible.

GEO2R was applied to recognize the DEGs and DE miRNAs between HCC tissues and paired ANTTs in each dataset. The adjusted P values (adj. P) and log Fold Change (FC) in expression were determined. To select DEGs and DEmiRNAs, we set the adj. P and |logFC| for selection as <0.05 and >1.0, respectively. We obtained the overlap of DEGs identified from the three datasets using the Funrich webserver [21] (http://funrich.org/index.html).

Functional pathway enrichment analysis of DEGs and target key genes selection

To explore the functions of DEGs, we conducted GO functional annotation and KEGG pathway enrichment analysis utilizing the Database for annotation, visualization, and integrated discovery database (DAVID database) [22]. Functional annotation consists of a Biological Process (BP), Cellular Component (CC), and Molecular Function (MF). An adj. P<0.05 was statistically significant.

To investigate the importance of these DEGs, a protein-protein interaction (PPI) network was constructed by the online database STRING (https://string-db.org/) [23], followed by the visual investigation of the interaction networks among different biomolecules in Cytoscape. Then, to determine the hub genes, we used the plugin Molecular Complex Detection (MCODE) with the method of MCC performs. Finally, the top 25 genes with the highest degree values were regarded as hub genes.

To determine the target genes of 33 DE miRNAs screened from GSE10694, we used the miRNet WebServer (https://www.mirnet.ca/) [24] were utilized. The intersection of the target genes and the 25 hub genes were the final identified key genes.

Verification of expression levels based on parallel databases of key genes

The mRNA expression levels of key genes in HCC tissues vs. ANTTs were verified among the five selected datasets (Chen Liver with 197 samples [25], Mas Liver with 115 samples [26], Roessler Liver with 43 samples [27], Roessler Liver 2 with 445 samples [27] and Wurmbach Liver with 75 samples [28]) based on the ONCOMINE database (https://www.oncomine.org) [29].

The mRNA expression level of key genes in liver hepatocellular carcinoma (LIHC) and normal liver tissues were visualized by the GEPIA database (http://gepia.cancer-pku.cn/) [30].

The protein expression levels of the key genes in liver cancer tissues and non-tumor liver tissues were presented by the Human Protein Atlas (HPA, https://www.proteinatlas.org/) [31].

The prognostic and diagnostic value analysis of key genes

The prognostic significance of key genes for HCC were analyzed by Kaplan Meier survival analysis (https://kmplot.com/analysis/index.php?p=service&cancer=liver_rnaseq). The correlation between Overall Survival (OS)/Progression-Free Survival (PFS) and the expression level of key genes was evaluated in 364/370 patients with liver cancer. Presently, patients with liver cancer were categorized into two groups according to the median expression levels of key genes. P<0.05 was considered as statistically significant.

In addition, the genomic alterations of the key genes were investigated by cBioPortal [32] with a LIHC dataset consisting of 442
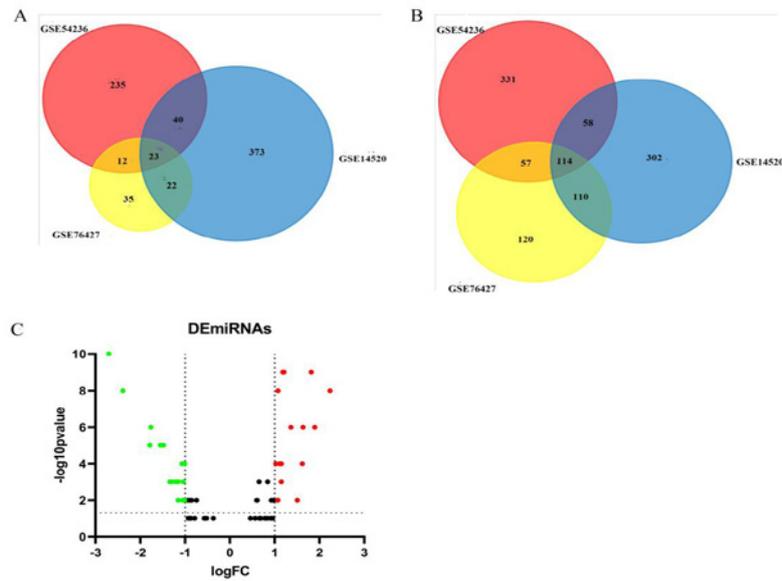
Figure 1: DEGs and DE miRNAs identified from the selected dataSets.
A: Venn diagram of up-regulated DEGs based on the three GEO datasets (GSE54236, GSE14520, GSE76427).
B: Venn diagram of down-regulated DEGs based on the three GEO datasets (GSE54236, GSE14520, GSE76427).
C: Volcano plot of the identified 33 DE miRNAs base on GSE10694 dataset.
Red, upregulation; green, downregulation. The t-test was used to analyze DEGs, with the cut-off criteria of |logFC|>1.0 and adj. P<0.05.
DEG, differentially expressed gene; DE miRNA, differentially expressed miRNA; GEO, Gene Expression Omnibus; logFC, log-fold change.

samples. We assessed the association of alterations of the key genes with the Disease-Free Survival (DFS) and OS of patients with LIHC.

We downloaded the mRNA-sequencing data and clinical signatures of LIHC and non-cancerous tissues from The Cancer Genome Atlas (TCGA) (https://portal.gdc.cancer.gov/) (N=423 samples). The expression values were normalized by log2 transformation and imported into medcalc software. Receiver Operating Characteristic (ROC) curve analyses were performed to estimate the diagnostic value of key genes. The Area Under the Curve (AUC) represented the diagnostic efficacy of key genes for HCC. P<0.05 was regarded as statistically significant.

Relationship between the expression of key genes and immune infiltration in HCC

TIMER (Tumor IMmune Estimation Resource, https://cistrome.shinyapps.io/timer/) [33] is used to analyze the immune infiltration of various types of cancers in TCGA to explore the influence of key genes on the tumor immune infiltration. Herein, we investigated the association of key genes with the infiltrating levels of immune cells in HCC, including B cells, CD8+ T cells, CD4+ T cells, macrophages, neutrophils, and dendritic cells (DCs). Furthermore, we investigated the relationship between the expression of genetic markers (PD-1, PDL-1, CTLA4, TIM-3, and LAG3) of T cell exhaustion with the expression of key genes.

Association of key genes with clinicopathological profiles of patients with HCC

We determined the correlation of different clinical stages, alcohol consumption and hepatitis virus status of patients with HCC by Kaplan–Meier analysis to explore the association between key genes and clinicopathological features of HCC. In addition, we downloaded

the mRNA-sequencing data and clinical signatures of patients with HCC from TCGA. After normalizing the expression values of key genes by log2 transformation, we compared the expression of key genes in different tumor stages (S1, S2 and S3-4) and differentiation grades (G1, G2 and G3-4) with the one-way ANOVA analysis by GraphPad primer. P<0.05 was considered statistically significant.
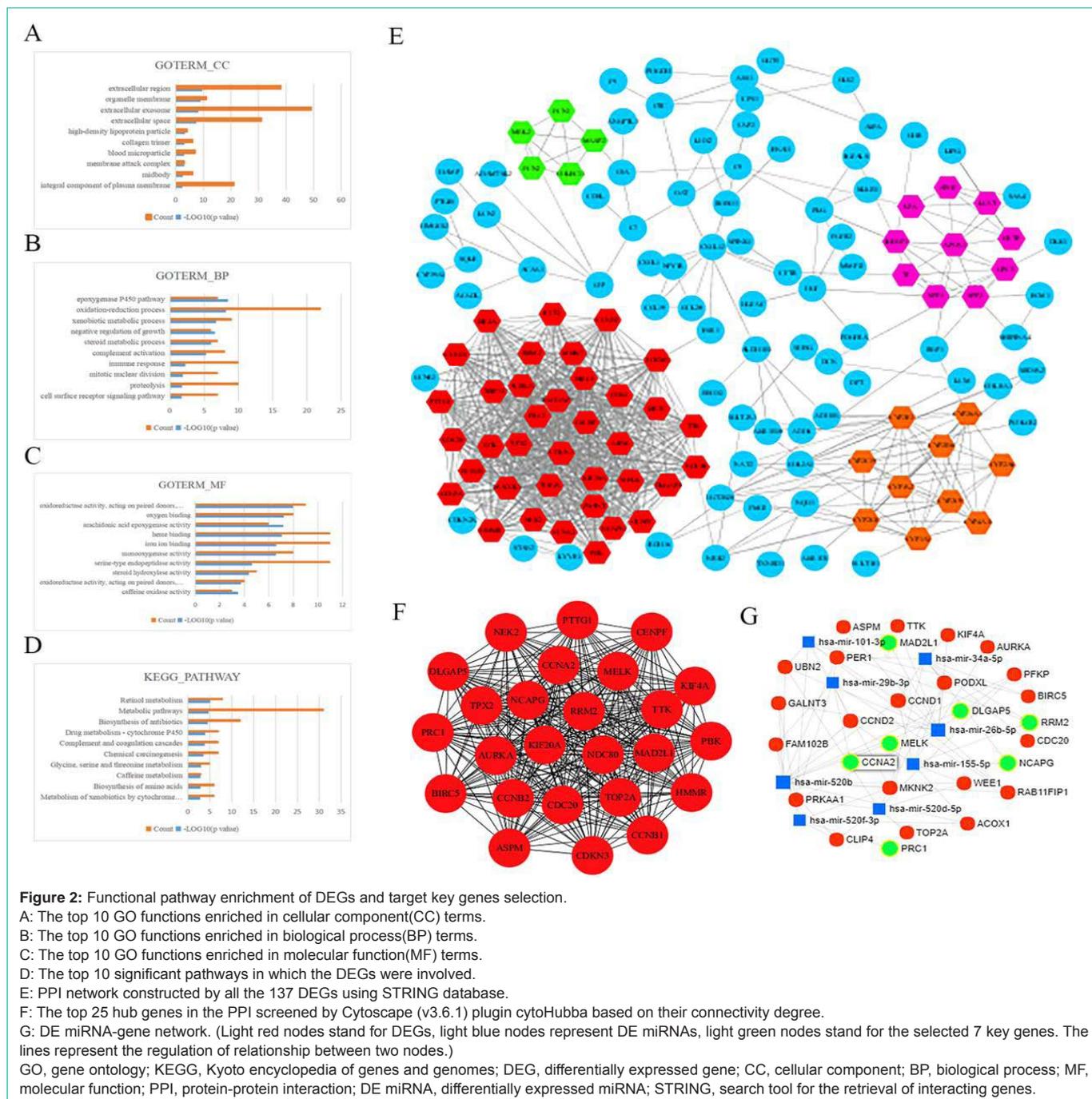
Results

DEGs and DE miRNAs

In this study, four independent datasets (GSE54236, GSE14520, GSE76427, and GSE10694), including a total of 498 HCC tissues and 357 ANTTs were selected from the GEO database (Table 1). Consequently, 23 overlapping upregulated genes (360 in GSE54236, 572 in GSE14520, and 115 in GSE76427) and 114 overlapping down-regulated genes (668 in GSE54236, 755 in GSE14520, and 450 in GSE76427) were identified using a Venn diagram (Figure 1A,1B). In addition, after analyzing the GSE10694 dataset, we identified 33 (16 upregulated and 17 downregulated) DE miRNAs, as displayed in Figure 1C.

Finally, 137 Differentially Expressed Genes (DEGs) and 33 differentially expressed miRNAs (DE miRNAs) were identified for subsequent analysis.

Functional pathway enrichment of DEGs and target key genes selection

We analyzed 137 DEGs separately with GO and KEGG pathway analyses, as shown in Figure 2A-2D. The top ten functions/pathways revealed extracellular exosomes in the CC group, oxidation-reduction processes in the BP group, and serine-type endopeptidase activity in the MF group. Pathway enrichment analysis revealed the involvement of metabolic pathways.

Figure 2: Functional pathway enrichment of DEGs and target key genes selection.

A: The top 10 GO functions enriched in cellular component(CC) terms.

B: The top 10 GO functions enriched in biological process(BP) terms.

C: The top 10 GO functions enriched in molecular function(MF) terms.

D: The top 10 significant pathways in which the DEGs were involved.

E: PPI network constructed by all the 137 DEGs using STRING database.

F: The top 25 hub genes in the PPI screened by Cytoscape (v3.6.1) plugin cytoHubba based on their connectivity degree.

G: DE miRNA-gene network. (Light red nodes stand for DEGs, light blue nodes represent DE miRNAs, light green nodes stand for the selected 7 key genes. The lines represent the regulation of relationship between two nodes.)

GO, gene ontology; KEGG, Kyoto encyclopedia of genes and genomes; DEG, differentially expressed gene; CC, cellular component; BP, biological process; MF, molecular function; PPI, protein-protein interaction; DE miRNA, differentially expressed miRNA; STRING, search tool for the retrieval of interacting genes.

As shown in Figure 2E, these overlapping DEGs with combined scores higher than 0.4 showed significant interactions and networks in the PPI network. Four modules were selected out and the top 25 genes displayed by the method of MCC and sequenced as follows: MELK, CDKN3, RRM2, PRC1, ASPM, KIF4A, CENPF, TOP2A, PBK, HMMR, TTK, BIRC5, TPX2, PTTG1, CCNB1, DLGAP5, AURKA, NDC80, NEK2, NCAPG, KIF20A, CCNA2, CCNB2, CDC20, and MAD2L1 (Figure 2F).

To identify the reliable key genes, target genes of the 33 DE miRNAs were predicted by the miRNet webserver. Then, the above target genes were compared with the selected 25 hub genes, and those

at the intersection were considered as the key genes. Finally, seven key genes were selected, including CCNA2, DLGAP5, MAD2L1, MELK, NCAPG, PRC1, and RRM2 (Figure 2G).

Verification of expression levels based on parallel databases of key genes

To verify the expression of key genes, we used Oncomine and GEPIA database to confirm the mRNA expression levels of seven key genes, which were markedly upregulated in HCC tissues (Figure 3A,3B). Additionally, results were similar for the protein expression levels of these key genes, as these were not detected or expressed at low levels in normal liver tissues. In contrast, medium or high expression
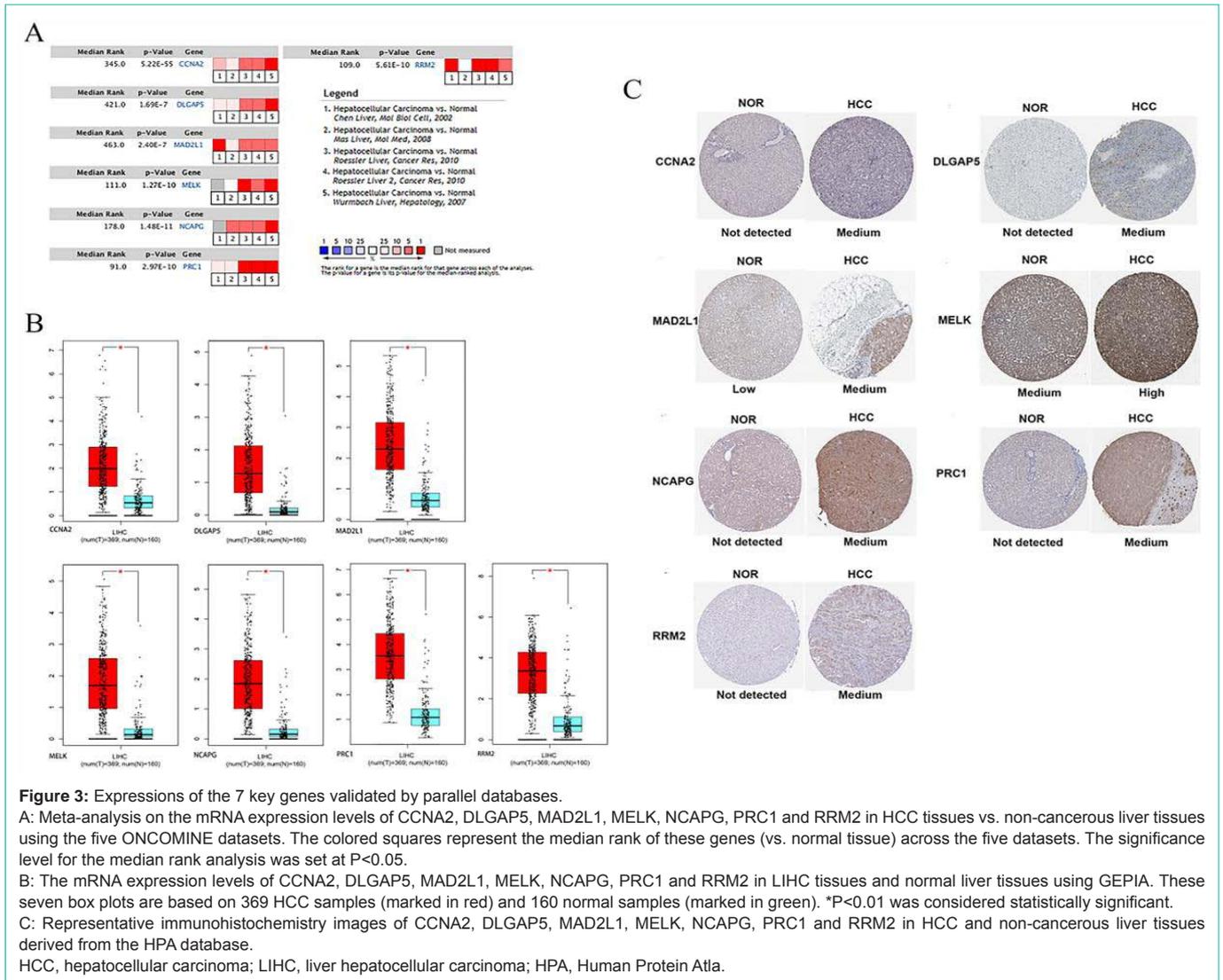
Figure 3: Expressions of the 7 key genes validated by parallel databases.

A: Meta-analysis on the mRNA expression levels of CCNA2, DLGAP5, MAD2L1, MELK, NCAPG, PRC1 and RRM2 in HCC tissues vs. non-cancerous liver tissues using the five ONCOMINE datasets. The colored squares represent the median rank of these genes (vs. normal tissue) across the five datasets. The significance level for the median rank analysis was set at P<0.05.

B: The mRNA expression levels of CCNA2, DLGAP5, MAD2L1, MELK, NCAPG, PRC1 and RRM2 in LIHC tissues and normal liver tissues using GEPIA. These seven box plots are based on 369 HCC samples (marked in red) and 160 normal samples (marked in green). *P<0.01 was considered statistically significant.

C: Representative immunohistochemistry images of CCNA2, DLGAP5, MAD2L1, MELK, NCAPG, PRC1 and RRM2 in HCC and non-cancerous liver tissues derived from the HPA database.

HCC, hepatocellular carcinoma; LIHC, liver hepatocellular carcinoma; HPA, Human Protein Atla.

in liver cancer tissues was found in the HPA database (Figure 3C).

In conclusion, both mRNA and protein expression levels of these seven key genes were significantly increased in patients with HCC.

The prognostic and diagnostic value of key genes

To assess the prognostic value of key genes, we performed Kaplan–Meier analyses. A total of 364 patients with HCC were available for OS analysis and 370 patients for PFS analysis. Overexpression of CCNA2, DLGAP5, MAD2L1, MELK, NCAPG, PRC1, and RRM2 in patients with HCC might be responsible for poor OS (Figure 4A) and PFS (Figure 4B). Meanwhile, the alteration frequencies of the key genes were assessed in 442 patients with LIHC using the cBioPortal database and we analyzed the association between the key gene alterations and prognosis of patients with LIHC. As revealed in Supplementary Figure 1, LIHC patients with key gene alterations had significantly worse OS and DFS than those without key gene alterations.

The diagnostic value of these key genes in HCC was analyzed with ROC curves. As demonstrated in Figure 4C, CCNA2 (AUC=0.969,

$p<0.001$), DLGAP5 (AUC=0.965, $p<0.001$), MAD2L1 (AUC=0.945, $p<0.001$), MELK (AUC=0.973, $p<0.001$), NCAPG (AUC=0.972, $p<0.001$), PRC1 (AUC=0.976, $p<0.001$), and RRM2 (AUC=0.961, $p<0.001$) were able to distinguish HCC tissues from normal para-tumor mucosa efficiently.

In brief, all seven key genes emerged as having prognostic and diagnostic capabilities for patients with HCC.

Relationship between key genes expression and immune infiltration in HCC

Given the critical role of immunity in tumorigenesis and progression, we investigated the relationship between key genes and the microenvironment of HCC using the TIMER database. The results suggested that the expression levels of CCNA2 ($r = 0.155$, $p = 3.9E-03$), MELK ($r = 0.16$, $p = 2.9E-03$), NCAPG ($r = 0.15$, $p = 5.04E-03$), PRC1 ($r = 0.199$, $p = 2E-04$), and RRM2 ($r = 0.133$, $p = 1.31E-02$) were significantly increased with tumor purity in HCC. In addition, infiltration levels of B cells showed the greatest significant direct correlation with DLGAP5($r = 0.478$, $p = 4.41E-21$), followed by CD8+ T cells with DLGAP5($r = 0.334$, $p = 2.42E-10$), CD4+ T
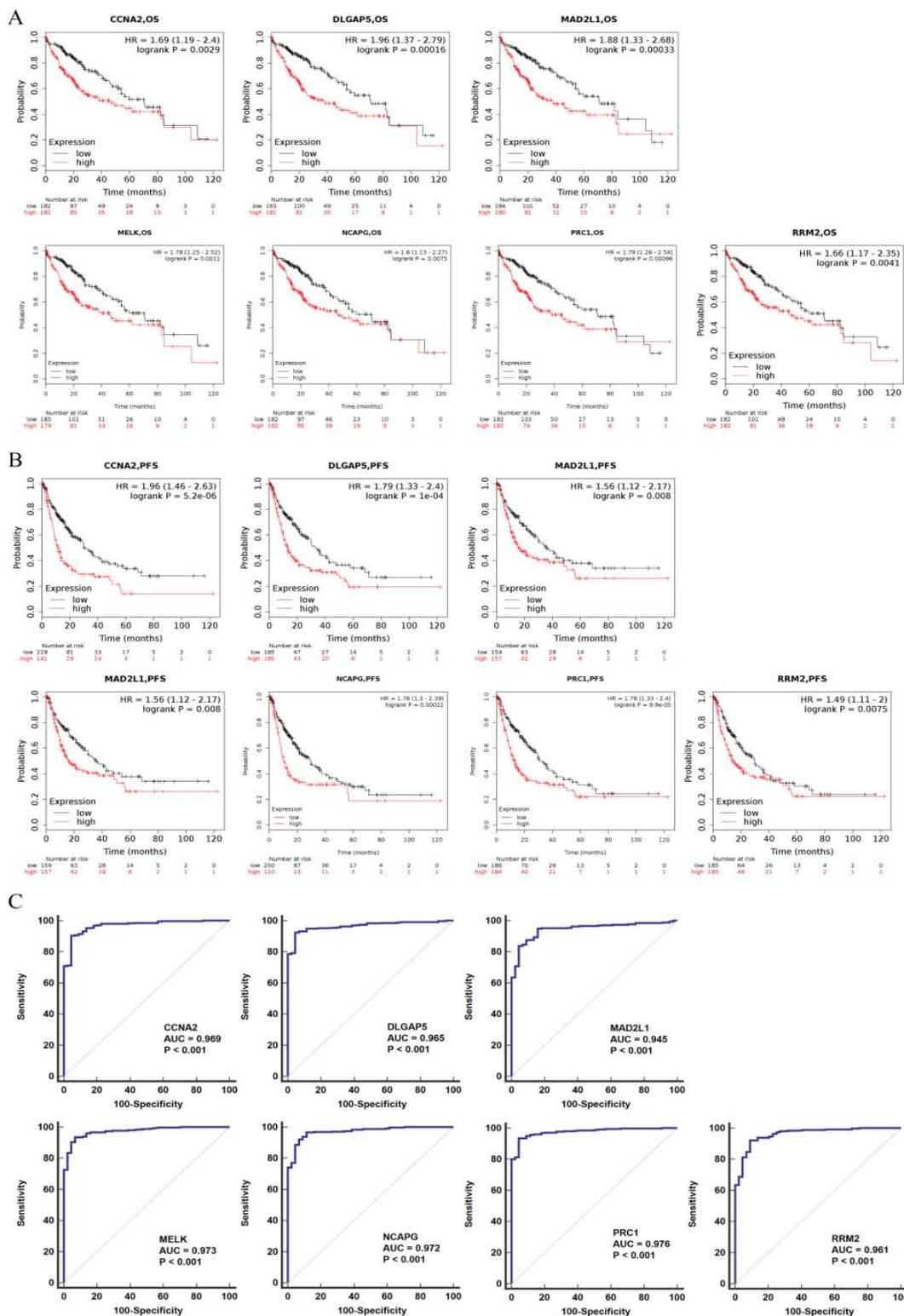
Figure 4: The prognostic and diagnostic value analysis of key genes.

A: Kaplan-Meier OS curves of 364 patients with HCC according to the expression levels of CCNA2, DLGAP5, MAD2L1, MELK, NCAPG, PRC1 and RRM2.

B: Kaplan-Meier PFS curves of 370 patients with HCC according to the expression levels of CCNA2, DLGAP5, MAD2L1, MELK, NCAPG, PRC1 and RRM2.

C: ROC curve analysis CCNA2, DLGAP5, MAD2L1, MELK, NCAPG, PRC1 and RRM2 in HCC.

Data are presented as the hazard ratio with a 95% confidence interval. Log-rank P<0.01 was regarded as statistically significant.

OS, overall survival; PFS, progression-free survival; LIHC, liver hepatocellular carcinoma; AUC: area under the curve; HCC: hepatocellular carcinoma.
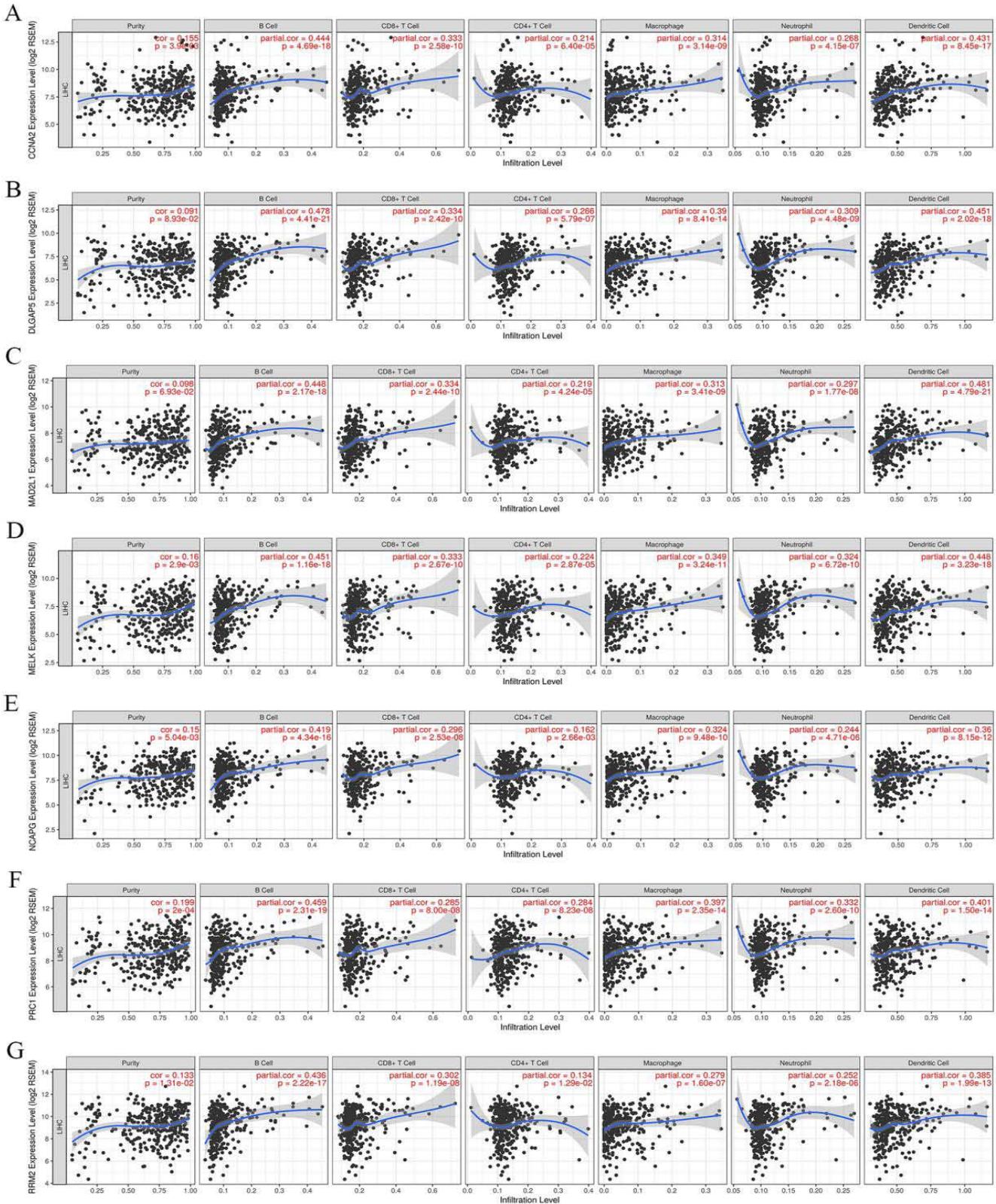
Figure 5: Relationship between expression levels of key genes and the infiltrating immunocytes in HCC.

The expression levels of (A) CCNA2, (B) DLGAP5, (C) MAD2L1, (D) MELK, (E) NCAPG, (F) PRC1 and (G) RRM2 were significantly correlated with tumor purity and infiltration levels of B cells, CD8+ T cells, CD4+ T cells, macrophages, neutrophils, and dendritic cells.

TIMER, Tumor Immune Estimation Resource; HCC, hepatocellular carcinoma.

Table 1: Statistics of the datasets derived from the Gene Expression Omnibus (GEO) database.

Data Sets	Hepatocellular carcinoma tissues	Paired adjacent non-tumor tissues	Total no.
GSE54236	88	72	160
GSE14520	223	223	446
GSE76427	115	52	167
GSE10694	72	10	82

GSE54236 datasets contained 88 HCC tissues and 72 paired adjacent non-tumor tissues. GSE14520 datasets contained 223 HCC tissues and 223 paired adjacent non-tumor tissues. GSE76427 included 115 HCC tissues and 52 paired adjacent non-tumor tissues. The miRNA expression profile of GSE10694 included 78 HCC tissues and 10 paired adjacent non-tumor tissues.

Table 2: Correlation analysis between key genes and related immune markers of exhausted T cells.

Gene markers		PD-1/PDCD1	CTLA-4	TIM-3/HAVCR2	LAG3	PDL1/CD274
CCNA2	Cor	0.324	0.335	0.348	0.293	0.278
	P	1.59E-10	3.55E-11	6.75E-12	1.09E-08	5.23E-08
DLGAP5	Cor	0.34	0.368	0.364	0.314	0.297
	P	1.79E-11	2.53E-13	6.59E-13	7.63E-10	5.15E-09
MAD2L1	Cor	0.319	0.345	0.375	0.279	0.303
	P	3.26E-10	7.76E-12	1.10E-13	5.11E-08	2.47E-09
MELK	Cor	0.287	0.322	0.316	0.275	0.322
	P	1.78E-08	2.10E-10	6.16E-10	8.38E-08	2.03E-10
NCAPG	Cor	0.312	0.344	0.309	0.299	0.216
	P	8.63E-10	8.95E-12	1.55E-09	5.30E-09	2.77E-05
PRC1	Cor	0.28	0.281	0.3	0.263	0.306
	P	4.34E-08	3.81E-08	4.38E-09	3.00E-07	1.67E-09
RRM2	Cor	0.297	0.357	0.345	0.326	0.296
	P	5.36E-09	1.28E-12	8.37E-12	1.22E-10	5.80E-09

cells with PRC1 (r = 0.284, p = 8.23E-8), macrophages with PRC1 (r = 0.397, p = 2.35E-14), neutrophils with PRC1 (r = 0.332, p = 2.60E-10), and dendritic cells with MAD2L1(r = 0.481, p = 4.79E-21), as demonstrated in Figure 5. Interestingly, positive correlations between the expression levels of key genes and the expression of gene markers in exhausted T cells, such as PD-1, PDL-1, CTLA4, TIM-3, and LAG3, were observed in HCC, as shown in Table 2.

These results suggested that these key genes affect various of infiltrating immune cells and may be associated with tumor immune escape.

Association of key genes with clinicopathological profiles of patients with HCC

To explore the association between key genes and clinicopathological features of HCC, we determined the correlation with different clinical stages, alcohol consumption and hepatitis virus status of patients with HCC by Kaplan–Meier analysis. As shown in Table 3, CCNA2, DLGAP5, MAD2L1, MELK, and NCAPG were significantly correlated with clinical stage III-IV. Only MELK and NCAPG were significantly associated with clinical stage II. Table 4 shows that CCNA2, DLGAP5, and PRC1 were significantly correlated with the alcohol consumption. Additionally, all seven key genes were significantly associated with HCC without the hepatitis virus. We compared the expression levels of key genes at different tumor stages (S1, S2, and S3-4) and differentiation grades (G1, G2, and G3-G4) with one-way ANOVA analysis. The data suggested that the expression levels of key genes, except for CCNA2, were significantly

correlated with HCC progression (Supplementary Figure 2).

In conclusion, these seven key genes might serve as biomarkers for the diagnosis of HCC without hepatitis virus. DLGAP5, MAD2L1, MELK, and NCAPG were significantly increased in HCC of advanced stages, suggesting that these genes may be involved in the progression of HCC.

Discussion

Even though the mortality of liver cancer has reduced in women and stabilized in men, there will be approximately 1.1 million cases in 2030 and 1.6 million patients in 2060 dying from liver cancer, as estimated by the World Health Organization [1]. To cope with this worldwide concern, studies into the genetic background and signaling pathways of HCC are performed to find biomarkers that can be used for the diagnosis, prognosis, or as therapeutic targets for HCC [7-10]. However, most of the performed studies were single-center and have been limited by their sample size, undermining the results. We selected four independent datasets with 855 subjects and several datasets from parallel databases with 2481 samples to identify and verify reliable biomarkers for HCC. As this analysis was based on publicly available data described by other investigators. Finally, we identified seven key genes: CCNA2, DLGAP5, MAD2L1, MELK, NCAPG, PRC1, and RRM2.

These seven key genes have been demonstrated in several studies to be involved in various types of solid tumors. CCNA2 affects the initiation and progression of breast tumors, and the absence of

Table 3: Expression levels of key genes with clinical status of liver cancer patients.

Gene	Clinical stages	HR (95% CI)	Log Rank P
CCNA2	Stage I	1.49 (0.81-2.75)	0.2011
	Stage II	1.97 (0.89-4.34)	0.088
	Stage III-IV	2.1 (1.16-3.83)	0.0126*
DLGAP5	Stage I	1.21 (0.66-2.22)	0.5437
	Stage II	2.07 (0.91-4.69)	0.0747
	Stage III-IV	2.14 (1.18-3.88)	0.0108*
MAD2L1	Stage I	1.73 (0.92-3.23)	0.0829
	Stage II	1.41 (0.65-3.06)	0.3778
	Stage III-IV	2.26 (1.26-4.05)	0.0051*
MELK	Stage I	1.84 (0.99-3.43)	0.051
	Stage II	2.38 (1.05-5.4)	0.0332*
	Stage III-IV	2.01 (1.12-3.6)	0.017*
NCAPG	Stage I	1.15 (0.63-2.1)	0.6574
	Stage II	2.34 (1.04-5.25)	0.0348*
	Stage III-IV	1.86 (1.04-3.33)	0.0337*
PRC1	Stage I	1.3 (0.71-2.4)	0.3965
	Stage II	1.71 (0.77-3.79)	0.1783
	Stage III-IV	1.66 (0.93-2.95)	0.083
RRM2	Stage I	1.41 (0.77-2.61)	0.2649
	Stage II	1.69 (0.75-3.83)	0.204
	Stage III-IV	1.19 (0.66-2.14)	0.5609

Table 4: Expression levels of key genes with risk factors of liver cancer patients.

Gene	Risk Factors		HR (95% CI)	Log Rank P
CCNA2	Alcohol consumption	Yes	2.07 (1.07-4)	0.027*
		None	1.78 (1.11-2.86)	0.015*
	Hepatitis virus	Yes	1.19 (0.62-2.27)	0.604
		None	2.17 (1.36-3.47)	0.0009*
DLGAP5	Alcohol consumption	Yes	2.71 (1.38-5.3)	0.0025*
		None	1.52 (0.95-2.43)	0.0763
	Hepatitis virus	Yes	1.34 (0.7-2.57)	0.3824
		None	2.36 (1.47-3.78)	0.0002*
MAD2L1	Alcohol consumption	Yes	1.7 (0.89-3.25)	0.105
		None	1.96 (1.22-3.14)	0.0049*
	Hepatitis virus	Yes	1.49 (0.78-2.86)	0.2263
		None	2.36 (1.47-3.81)	0.0003*
MELK	Alcohol consumption	Yes	1.47 (0.78-2.79)	0.232
		None	1.75 (1.1-2.8)	0.0174*
	Hepatitis virus	Yes	1.09 (0.57-2.07)	0.7961
		None	2.66 (1.65-4.27)	3.0e-5*
NCAPG	Alcohol consumption	Yes	1.59 (0.84-3.03)	0.1525
		None	1.85 (1.16-2.96)	0.0086*
	Hepatitis virus	Yes	1.02 (0.53-1.94)	0.9638
		None	2.19 (1.37-3.49)	0.0008*
PRC1	Alcohol consumption	Yes	1.97 (1.03-3.75)	0.036*
		None	1.57 (0.98-2.5)	0.058
	Hepatitis virus	Yes	1.05 (0.55-2)	0.8821
		None	2.44 (1.53-3.91)	0.0001*
RRM2	Alcohol consumption	Yes	1.33 ()0.7-2.51)	0.3763
		None	1.72 (1.07-2.75)	0.0226*
	Hepatitis virus	Yes	0.91 (0.47-1.73)	0.7646
		None	2.34 (1.46-3.74)	0.0003*

CCNA2 could lead to impaired ERK activity and reduced breast tumor metastasis [34]. DLGAP5 transcription was demonstrated to be elevated in the S phase of the cell cycle and maintained in the G2 and M phases, consistent with involvement in the initiation and development of cancer. Studies have shown the diagnostic and prognostic values of DLGAP5 in patients with lung and colorectal cancer [35-37]. Dysregulation of MAD2L1 could lead to chromosomal instability and aneuploidy, which might promote the formation of human cancers. The expression level of MAD2L1 has been shown to be increased in breast, lung, and gastric cancer [38-40]. MELK has been shown to play an essential role in various cellular and biological processes, such as cell proliferation and tumorigenesis [41]. Overexpression of MELK was found in breast cancer, melanoma and ovarian cancer, and in these cancers, it was associated with the poor prognosis [42-44]. NCAPG may contribute to the pathogenesis of various types of tumors, including prostate cancer [45], renal cell carcinoma [46], multiple myeloma [47], and melanoma [48]. Interestingly, researchers demonstrated that NCAPG was a key gene in liver carcinogenesis and the expression level of NCAPG in patients with HCC was associated with recurrences and patient survival [49,50]. PRC1 has been exerted a vital role in the Wnt/β-catenin signaling pathway and may have a carcinogenic effect by promoting the occurrence, proliferation and metastasis of tumors [51]. Upregulation of RRM2 has been shown to promote rapid cell division by increasing dNTP accumulation, and as such is involved in a variety of cancers in which it is also associated with poor prognosis [52-55].

In the present study, survival curves and ROC analyses suggested that these seven key genes may have predictive potential for HCC. Increased expressions of CCNA2, DLGAP5, MAD2L1, MELK, and NCAPG was significantly correlated with advanced stages of HCC, consistent with the association with poor prognosis. CCNA2, DLGAP5, and PRC1 were significantly correlated with alcohol consumption, indicating that they might serve as predictors for HCC associated with alcohol consumption. Additionally, all the key genes showed significant association with HCC patients without hepatitis virus, suggesting that these genes could be used as markers in HCC patients without hepatitis virus and not in those with hepatitis virus.

The immune system affects the development, and recovery of HCC. Specifically, the tumor microenvironment, including Tumor-Associated Macrophages (TAMs) [56] and TILs [57], determines the development of HCC. Consequently, we further investigated the correlation between the key genes and the levels of immune infiltrates in HCC.

We found that the expression levels of key genes were significantly associated with various immune cells in HCC. Furthermore, positive correlations between the expression levels of key genes and the

expression of gene markers in exhausted T cells, such as PD-1, PDL-1, CTLA4, TIM-3, and LAG3 were observed in HCC. Clinical trials of immune checkpoint inhibitors in HCC have shown positive responses, but the clinical response of some HCC patients remains unsatisfactory [16]. Our results indicate a worse prognosis with higher expression levels of key genes and a potential mechanism for inducing immune escape in HCC. We suggest that these genes can be used to evaluate the infiltration levels of immune cells in HCC and predict the response to immunotherapy.

There is moderate evidence that a non-canonical D box (D2) is essential for the ubiquitination of CCNA2 *in vitro* and its degradation *in vivo* [58]. The silencing of DLGAP5 by siRNA can inhibit the proliferation and invasion of HCC significantly [59]. In addition, knockdown of MELK by RNA interference (RNAi) can inhibit the proliferation of cancer cells and activates mitosis catastrophes [60,61]. Evidence from another study suggests that miR-181c could exhibit tumor-suppression *via* the downregulation of NCAPG levels [62]. It is reasonable to propose further that the combination of inhibitors of these key genes and immunotherapy may improve the survival time of patients with HCC. However, our study lacked clinical experiments to verify.

In conclusion, CCNA2, DLGAP5, MAD2L1, MELK, NCAPG, PRC1, and RRM2 were identified based on public databases. The overexpression of these key genes has been demonstrated in HCC tissues and is associated with advanced disease and poor prognosis of patients with HCC. However, more experimental evidence with larger sample sizes is needed to verify our findings and explore the role of key genes in the development of HCC and their relationship with tumor immunity. The expression of key genes was positively correlated with the expression of gene markers of immune escape in HCC, suggesting these genes might be used to evaluate the immune infiltration levels in HCC and predict the response to immunotherapy.

Data Availability Statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article.

Author Contributions

TMM and DL contributed to the design of this study. TMM contributed to the collection, analysis, collation, and validation of the data, as well as the drafting of the manuscript. DL contributed to the review and the revision of the manuscript. Both authors gave final approval to this manuscript for publication.

Funding

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA: a cancer journal for clinicians. 2020; 70: 7-30.

2. Llovet JM, Montal R, Sia D, Finn RS. Molecular therapies and precision medicine for hepatocellular carcinoma. Nature reviews Clinical oncology. 2018; 15: 599-616.

3. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians. 2018; 68: 394-424.

4. Czauderna C, Marquardt JU, Galle PR, Worns MA. [Hepatocellular carcinoma]. Der Internist. 2017; 58: 469-79.

5. Marrero JA, Lok AS. Newer markers for hepatocellular carcinoma. Gastroenterology. 2004; 127: S113-9.

6. Villanueva A. Hepatocellular Carcinoma. The New England journal of medicine. 2019; 380: 1450-62.

7. Morris JS, Hassan MM, Zohner YE, Wang Z, Xiao L, Rashid A, et al. HepatoScore-14: Measures of biological heterogeneity significantly improve prediction of hepatocellular carcinoma risk. Hepatology (Baltimore, Md). 2020.

8. Dou C, Sun L, Wang L, Cheng J, Wu W, Zhang C, et al. Bromodomain-containing protein 9 promotes the growth and metastasis of human hepatocellular carcinoma by activating the TUFT1/AKT pathway. Cell death & disease. 2020; 11: 730.

9. Qin Y, Cheng S, Li Y, Zou S, Chen M, Zhu D, et al. The development of a Glypican-3-specific binding peptide using *in vivo* and *in vitro* two-step phage display screening for the PET imaging of hepatocellular carcinoma. Biomaterials science. 2020.

10. Zhou T, Ren Z, Chen C. [METTL14 as a predictor of postoperative survival outcomes of patients with hepatocellular carcinoma]. Nan fang yi ke da xue xue bao = Journal of Southern Medical University. 2020; 40: 567-72.

11. Shlomai A, de Jong YP, Rice CM. Virus associated malignancies: the role of viral hepatitis in hepatocellular carcinoma. Seminars in cancer biology. 2014; 26: 78-88.

12. Gao Q, Qiu SJ, Fan J, Zhou J, Wang XY, Xiao YS, et al. Intratumoral balance of regulatory and cytotoxic T cells is associated with prognosis of hepatocellular carcinoma after resection. Journal of clinical oncology: official journal of the American Society of Clinical Oncology. 2007; 25: 2586-93.

13. Woo SR, Turnis ME, Goldberg MV, Bankoti J, Selby M, Nirschl CJ, et al. Immune inhibitory molecules LAG-3 and PD-1 synergistically regulate T-cell function to promote tumoral immune escape. Cancer research. 2012; 72: 917-927.

14. Kudo M. Immune Checkpoint Inhibition in Hepatocellular Carcinoma: Basics and Ongoing Clinical Trials. Oncology. 2017; 92: 50-62.

15. Li H, Wu K, Tao K, Chen L, Zheng Q, Lu X, et al. Tim-3/galectin-9 signaling pathway mediates T-cell dysfunction and predicts poor prognosis in patients with hepatitis B virus-associated hepatocellular carcinoma. Hepatology (Baltimore, Md). 2012; 56: 1342-1351.

16. Greten TF, Lai CW, Li G, Staveley-O'Carroll KF. Targeted and Immune-Based Therapies for Hepatocellular Carcinoma. Gastroenterology. 2019; 156: 510-524.

17. Trevino V, Falciani F, Barrera-Saldaña HA. DNA microarrays: a powerful genomic tool for biomedical and clinical research. Molecular medicine (Cambridge, Mass). 2007; 13: 527-541.

18. Li Y, Chen R, Yang J, Mo S, Quek K, Kok CH, et al. Integrated Bioinformatics Analysis Reveals Key Candidate Genes and Pathways Associated With Clinical Outcome in Hepatocellular Carcinoma. Frontiers in Genetics. 2020; 11: 814.

19. Ji Y, Yin Y, Zhang W. Integrated Bioinformatic Analysis Identifies Networks and Promising Biomarkers for Hepatitis B Virus-Related Hepatocellular Carcinoma. International journal of genomics. 2020; 2020: 2061024.

20. Clough E, Barrett T. The Gene Expression Omnibus Database. Methods in molecular biology (Clifton, NJ). 2016; 1418: 93-110.

21. Pathan M, Keerthikumar S, Ang CS, Gangoda L, Quek CY, Williamson NA, et al. FunRich: An open access standalone functional enrichment and interaction network analysis tool. Proteomics. 2015; 15: 2597-2601.

22. Huang DW, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, et al. DAVID

Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. Nucleic acids research. 2007; 35: W169-W175.

23. von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. STRING: a database of predicted functional associations between proteins. Nucleic acids research. 2003; 31: 258-261.

24. Fan Y, Xia J. miRNet-Functional Analysis and Visual Exploration of miRNA-Target Interactions in a Network Context. Methods in molecular biology (Clifton, NJ). 2018; 1819: 215-233.

25. Chen X, Cheung ST, So S, Fan ST, Barry C, Higgins J, et al. Gene expression patterns in human liver cancers. Molecular biology of the cell. 2002; 13: 1929-1939.

26. Mas VR, Maluf DG, Archer KJ, Yanek K, Kong X, Kulik L, et al. Genes involved in viral carcinogenesis and tumor initiation in hepatitis C virus-induced hepatocellular carcinoma. Molecular medicine (Cambridge, Mass). 2009; 15: 85-94.

27. Roessler S, Jia HL, Budhu A, Forgues M, Ye QH, Lee JS, et al. A unique metastasis gene signature enables prediction of tumor relapse in early-stage hepatocellular carcinoma patients. Cancer research. 2010; 70: 10202-12.

28. Wurmbach E, Chen YB, Khitrov G, Zhang W, Roayaie S, Schwartz M, et al. Genome-wide molecular profiles of HCV-induced dysplasia and hepatocellular carcinoma. Hepatology (Baltimore, Md). 2007; 45: 938-947.

29. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia. 2004; 6: 1-6.

30. 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.

31. Ponten F, Jirstrom K, Uhlen M. The Human Protein Atlas--a tool for pathology. The Journal of pathology. 2008; 216: 387-393.

32. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer discovery. 2012; 2: 401-404.

33. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, et al. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. Cancer research. 2017; 77: e108-e110.

34. Gao T, Han Y, Yu L, Ao S, Li Z, Ji J. CCNA2 is a prognostic biomarker for ER+ breast cancer and tamoxifen resistance. PloS one. 2014; 9: e91771.

35. Bassal S, Nomura N, Venter D, Brand K, McKay MJ, van der Spek PJ. Characterization of a novel human cell-cycle-regulated homologue of Drosophila dlg1. Genomics. 2001; 77: 5-7.

36. Shi YX, Yin JY, Shen Y, Zhang W, Zhou HH, Liu ZQ. Genome-scale analysis identifies NEK2, DLGAP5 and ECT2 as promising diagnostic and prognostic biomarkers in human lung cancer. Scientific reports. 2017; 7: 8072.

37. Branchi V, Garcia SA, Radhakrishnan P, Gyorffy B, Hissa B, Schneider M, et al. Prognostic value of DLGAP5 in colorectal cancer. International journal of colorectal disease. 2019; 34: 1455-1465.

38. Guo Y, Zhang X, Yang M, Miao X, Shi Y, Yao J, et al. Functional evaluation of missense variations in the human MAD1L1 and MAD2L1 genes and their impact on susceptibility to lung cancer. Journal of medical genetics. 2010; 47: 616-622.

39. Scintu M, Vitale R, Prencipe M, Gallo AP, Bonghi L, Valori VM, et al. Genomic instability and increased expression of BUB1B and MAD2L1 genes in ductal breast carcinoma. Cancer letters. 2007; 254: 298-307.

40. Wang Y, Wang F, He J, Du J, Zhang H, Shi H, et al. miR-30a-3p Targets MAD2L1 and Regulates Proliferation of Gastric Cancer Cells. OncoTargets and therapy. 2019; 12: 11313-11324.

41. Jiang P, Zhang D. Maternal Embryonic Leucine Zipper Kinase (MELK): a novel regulator in cell cycle control, embryonic development, and cancer. International journal of molecular sciences. 2013; 14: 21551-21560.

42. Janostiak R, Rauniyar N, Lam TT, Ou J, Zhu LJ, Green MR, et al. MELK Promotes Melanoma Growth by Stimulating the NF-kappaB Pathway. Cell reports. 2017; 21: 2829-2841.

43. Kohler RS, Kettelhack H, Knipprath-Meszaros AM, Fedier A, Schoetzau A, Jacob F, et al. MELK expression in ovarian cancer correlates with poor outcome and its inhibition by OTSSP167 abrogates proliferation and viability of ovarian cancer cells. Gynecologic oncology. 2017; 145: 159-166.

44. Speers C, Zhao SG, Kothari V, Santola A, Liu M, Wilder-Romans K, et al. Maternal Embryonic Leucine Zipper Kinase (MELK) as a Novel Mediator and Biomarker of Radioresistance in Human Breast Cancer. Clinical cancer research : an official journal of the American Association for Cancer Research. 2016; 22: 5864-5875.

45. Goto Y, Kurozumi A, Arai T, Nohata N, Kojima S, Okato A, et al. Impact of novel miR-145-3p regulatory networks on survival in patients with castration-resistant prostate cancer. Br J Cancer. 2017; 117: 409-420.

46. Yamada Y, Arai T, Kojima S, Sugawara S, Kato M, Okato A, et al. Regulation of antitumor miR-144-5p targets oncogenes: Direct regulation of syndecan-3 and its clinical significance. Cancer science. 2018; 109: 2919-2936.

47. Cohen Y, Gutwein O, Garach-Jehoshua O, Bar-Haim A, Kornberg A. The proliferation arrest of primary tumor cells out-of-niche is associated with widespread downregulation of mitotic and transcriptional genes. Hematology (Amsterdam, Netherlands). 2014; 19: 286-292.

48. Ryu B, Kim DS, Deluca AM, Alani RM. Comprehensive expression profiling of tumor cell lines identifies molecular signatures of melanoma progression. PloS one. 2007; 2: e594.

49. Zhang Q, Su R, Shan C, Gao C, Wu P. Non-SMC Condensin I Complex, Subunit G (NCAPG) is a Novel Mitotic Gene Required for Hepatocellular Cancer Cell Proliferation and Migration. Oncology research. 2018; 26: 269-276.

50. Zhang L, Huang Y, Ling J, Zhuo W, Yu Z, Shao M, et al. Screening and function analysis of hub genes and pathways in hepatocellular carcinoma via bioinformatics approaches. Cancer biomarkers: section A of Disease markers. 2018; 22: 511-521.

51. Chen J, Rajasekaran M, Xia H, Zhang X, Kong SN, Sekar K, et al. The microtubule-associated protein PRC1 promotes early recurrence of hepatocellular carcinoma in association with the Wnt/beta-catenin signalling pathway. Gut. 2016; 65: 1522-1534.

52. Liu X, Zhou B, Xue L, Yen F, Chu P, Un F, et al. Ribonucleotide reductase subunits M2 and p53R2 are potential biomarkers for metastasis of colon cancer. Clinical colorectal cancer. 2007; 6: 374-381.

53. Morikawa T, Maeda D, Kume H, Homma Y, Fukayama M. Ribonucleotide reductase M2 subunit is a novel diagnostic marker and a potential therapeutic target in bladder cancer. Histopathology. 2010; 57: 885-892.

54. Xia G, Wang H, Song Z, Meng Q, Huang X, Huang X. Gambogic acid sensitizes gemcitabine efficacy in pancreatic cancer by reducing the expression of ribonucleotide reductase subunit-M2 (RRM2). Journal of experimental & clinical cancer research: CR. 2017; 36: 107.

55. Rahman MA, Amin AR, Wang D, Koenig L, Nannapaneni S, Chen Z, et al. RRM2 regulates Bcl-2 in head and neck and lung cancers: a potential target for cancer therapy. Clinical cancer research: an official journal of the American Association for Cancer Research. 2013; 19: 3416-3428.

56. Coussens LM, Werb Z. Inflammation and cancer. Nature. 2002; 420: 860-867.

57. Chen DS, Mellman I. Oncology meets immunology: the cancer-immunity cycle. Immunity. 2013; 39: 1-10.

58. Zhang S, Tischer T, Barford D. Cyclin A2 degradation during the spindle assembly checkpoint requires multiple binding modes to the APC/C. Nat Commun. 2019; 10: 3863.

59. Liao W, Liu W, Yuan Q, Liu X, Ou Y, He S, et al. Silencing of DLGAP5 by siRNA significantly inhibits the proliferation and invasion of hepatocellular carcinoma cells. PloS one. 2013; 8: e80789.

60. Beke L, Kig C, Linders JT, Boens S, Boeckx A, van Heerde E, et al. MELK-T1, a small-molecule inhibitor of protein kinase MELK, decreases DNA-damage tolerance in proliferating cancer cells. Bioscience reports. 2015; 35.

61. Gray D, Jubb AM, Hogue D, Dowd P, Kljavin N, Yi S, et al. Maternal embryonic leucine zipper kinase/murine protein serine-threonine kinase 38 is a promising therapeutic target for multiple cancers. Cancer research. 2005; 65: 9751-9761.

62. Ai J, Gong C, Wu J, Gao J, Liu W, Liao W, et al. MicroRNA-181c suppresses growth and metastasis of hepatocellular carcinoma by modulating NCAPG. Cancer management and research. 2019; 11: 3455-3467.