Examining the key genes and pathways in hepatocellular carcinoma development from hepatitis B virus-positive cirrhosis

QI-FENG CHEN1*, JIN-GUO XIA2*, WANG LI1, LU-JUN SHEN1, TAO HUANG1 and PEIHONG WU1

1Department of Medical Imaging and Interventional Radiology, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, Guangdong 510060; 2Department of Interventional Radiology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, Jiangsu 210009, P.R. China

Received February 26, 2018; Accepted August 28, 2018

DOI: 10.3892/mmr.2018.9494

Abstract. To identify the key genes and pathways in the development of hepatocellular carcinoma (HCC) from hepatitis B virus (HBV)-positive liver cirrhosis, differentially expressed genes (DEGs) between HCC and liver cirrhosis tissue samples from the GSE17548 gene expression profile dataset were screened. A total of 1,845 DEGs were identified, including 1,803 upregulated and 42 downregulated genes. Gene Ontology, Kyoto Encyclopedia of Genes and Genomes (KEGG) and protein-protein interaction (PPI) network analyses were performed. It was identified that the 'cell cycle' and 'progesterone-mediated oocyte maturation' KEGG pathways were significantly enriched in the DEGs. In addition, the high expression of the hub genes from the PPI network (including cyclin dependent kinase 1, cyclin B1, cyclin B2, mitotic arrest deficient 2 like 1, BUB1 mitotic checkpoint serine/threonine kinase and cyclin A2; P=0.000116, 0.00021, 0.04889, 0.00222, 0.00015 and 0.00647, respectively) was associated with a decrease in overall survival for patients with HCC as identified using survival and expression data from The Cancer Genome Atlas. The identified hub genes and pathways may help to elucidate the molecular mechanisms for HCC progression from HBV-positive liver cirrhosis. Additionally, they may be useful as therapeutic targets or serve as novel biomarkers for HCC prognosis prediction.

Introduction

Hepatocellular carcinoma (HCC) is the most common type of liver cancer and is a principal health problem worldwide (1). The majority of HCC cases are associated with viral infection; hepatitis B virus (HBV) is the principal cause of HCC in Asia (2-4). HBV is epidemiologically associated with the development of liver cirrhosis (5). HCC frequently develops in the background of cirrhosis following chronic HBV infection (6,7).

Bioinformatics analysis combined with microarray technology has provided a novel method for comprehensively studying the alterations in expression associated with disease at the molecular level. Critical tumor-associated genes and cellular signaling pathways have been identified in recent decades using these methods. Therefore, examining the differentially expressed genes (DEGs) between HCC and liver cirrhosis tissues is helpful for identifying the key genes and pathways leading to tumor progression.

However, previous studies have focused primarily on progression from liver cirrhosis to HCC without regard to the etiology of cirrhosis (8,9). Furthermore, analysis regarding the associations between gene expression and patient survival is lacking. At present, there is insufficient knowledge regarding the molecular mechanisms for HCC transformation from liver cirrhosis in patients with HBV infection.

In the present study, expression profiles from HCC and liver cirrhosis tissue samples in a Gene Expression Omnibus (GEO) dataset were compared to identify DEGs. Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment and protein-protein interaction (PPI) network analyses were performed using the DEGs. The prognostic value of the hub genes from the PPI network was validated using survival and expression data from The Cancer Genome Atlas (TCGA) database. The results of the present study may help to identify the genes involved in the progression of HBV-positive liver cirrhosis to HCC, and provide information regarding biomarkers for verification and development in further studies.

Materials and methods

Microarray data. The GSE17548 gene expression profile dataset, based on the GPL570 platform ([HG-U133_Plus_2] Affymetrix
Human Genome U133 Plus 2.0 Array; Affymetrix; Thermo Fisher Scientific, Inc., Waltham, MA, USA] and deposited by Yildiz et al (8), was downloaded from the GEO database (www.ncbi.nlm.nih.gov/geo/). In total, 21 HBV-positive samples, including 11 liver cirrhosis and 10 HCC tissues, were analyzed.

Identification of DEGs. The raw expression data were preprocessed using the robust multiarray average method (10) in the affy package (11) in Bioconductor. Following preprocessing, HCC tissue samples were compared with liver cirrhosis tissues using an unpaired t-test in the Bioconductor limma package (version 3.10.3; bioconductor.org/biocLite.R). An adjusted P<0.01 and |log fold change| >1 were set as thresholds for the identification of DEGs.

GO term and KEGG pathway enrichment analysis of the DEGs. Subsequent to obtaining the DEGs, GO term and KEGG pathway enrichment analyses were performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID; https://david.ncifcrf.gov/). GO (12) analysis determines the potential functions of a list of genes, including the associated biological process (BP), molecular function (MF) and cellular component (CC) terms. KEGG (http://www.genome/ad.jp/kegg/) is a database of genomes, biological pathways, associated diseases and possible drug targets. P<0.05 was considered to indicate a statistically significant difference for the GO and KEGG enrichment analyses.

PPI network construction. A PPI network of DEGs was constructed using the Search Tool for the Retrieval of Interacting Genes (http://string.embl.de/) database and Cytoscape software (version 3.6.0; www.cytoscape.org). As a large number of DEGs was being analyzed, an interaction confidence score threshold of >0.90 was set to avoid uncertain

Figure 1. Boxplots of sample data. The x axis represents the samples and the y axis represents the alterations in gene expression. Blue boxes represent cirrhotic expression profiles and red boxes represent hepatocellular carcinoma tumor expression profiles.
PPIs. PPI networks were visualized with Cytoscape software. Modules from the PPI network were screened using the Molecular Complex Detection (MCODE) plugin with the following criteria: Degree cutoff, 2; node score cutoff, 0.2; k-core, 2; and max depth, 100. P<0.05 was used as a threshold value. The top 10 genes, as ranked by degree, were considered the hub genes.

Survival validation in TCGA. The mRNA-sequencing expression profiles and clinical data from HCC samples were downloaded from TCGA. For the analysis of hub genes identified in the PPI network, patients with cancer were divided into different groups (high and low) based on the median expression of each hub gene. To investigate the association of each gene with overall survival, survival outcomes between the high and low gene expression groups were compared using the Kaplan-Meier method and log-rank test in the hash package in Bioconductor. P<0.05 was considered to indicate a statistically significant difference in survival.

Results

Data preprocessing and identification of DEGs. Following background correction and normalization, a total of 20,460 genes were annotated in the 21 HCC and liver cirrhosis expression profiles. In total, 1,845 DEGs were identified, including 1,803 upregulated and 42 downregulated genes. A heat map of DEG expression is presented in Fig. 2.

GO term and KEGG pathway enrichment analysis. Using the DAVID tool, the GO term and KEGG pathway enrichment of the identified DEGs was analyzed. GO BP analysis
suggested that the DEGs were primarily associated with the ‘cell division’, ‘mitotic nuclear division’, ‘DNA replication’, ‘sister chromatid cohesion’, ‘DNA replication initiation’, ‘G\textsubscript{1}/S transition of mitotic cell cycle’, ‘telomere maintenance via recombination’, ‘DNA synthesis involved in DNA repair’, ‘chromosome segregation’ and ‘G\textsubscript{2}/M transition of mitotic cell cycle’ terms (Table I).

In the GO CC category, the ‘nucleoplasm’, ‘nucleus’, ‘cytoplasm’, ‘condensed chromosome kinetochore’, ‘chromosome, centromeric region’, ‘cytosol’, ‘kinetochore’, ‘midbody’, 'centrosome', 'spindle', 'ligase activity', 'ubiquitin-protein transferase activity', 'ubiquitin binding', 'identical protein binding' terms were significantly enriched.

ATP, adenosine triphosphate.

| Term                                      | Description                                      | Count | P-value        |
|-------------------------------------------|--------------------------------------------------|-------|---------------|
| GO:00051301                               | Cell division                                    | 96    | 1.40x10\textsuperscript{-26} |
| GO:0007067                                | Mitotic nuclear division                          | 70    | 2.80x10\textsuperscript{-20} |
| GO:0006260                                | DNA replication                                   | 52    | 9.60x10\textsuperscript{-19} |
| GO:0007062                                | Sister chromatid cohesion                         | 37    | 1.69x10\textsuperscript{-14} |
| GO:0006270                                | DNA replication initiation                        | 18    | 5.43x10\textsuperscript{-11} |
| GO:0000082                                | G\textsubscript{1}/S transition of mitotic cell cycle | 32    | 7.00x10\textsuperscript{-11} |
| GO:0000722                                | Telomere maintenance via recombination            | 16    | 7.09x10\textsuperscript{-9}   |
| GO:0000731                                | DNA synthesis involved in DNA repair              | 15    | 2.63x10\textsuperscript{-7}   |
| GO:0007059                                | Chromosome segregation                            | 21    | 2.99x10\textsuperscript{-7}   |
| GO:0000086                                | G\textsubscript{2}/M transition of mitotic cell cycle | 31    | 5.38x10\textsuperscript{-7}   |

| Term                                      | Description                                      | Count | P-value        |
|-------------------------------------------|--------------------------------------------------|-------|---------------|
| GO:0005654                                | Nucleoplasm                                      | 354   | 2.06x10\textsuperscript{-21} |
| GO:0005634                                | Nucleus                                          | 565   | 6.15x10\textsuperscript{-15} |
| GO:0005737                                | Cytoplasm                                        | 523   | 2.26x10\textsuperscript{-10} |
| GO:0000777                                | Condensed chromosome kinetochore                 | 26    | 9.69x10\textsuperscript{-9}   |
| GO:0000775                                | Chromosome, centromeric region                    | 20    | 4.06x10\textsuperscript{-8}   |
| GO:0005829                                | Cytosol                                          | 343   | 5.61x10\textsuperscript{-8}   |
| GO:0000776                                | Kinetochore                                      | 23    | 2.22x10\textsuperscript{-7}   |
| GO:0030496                                | Midbody                                          | 30    | 2.56x10\textsuperscript{-7}   |
| GO:0005813                                | Centrosome                                       | 65    | 6.05x10\textsuperscript{-7}   |
| GO:0005819                                | Spindle                                          | 26    | 8.38x10\textsuperscript{-6}   |

| Term                                      | Description                                      | Count | P-value        |
|-------------------------------------------|--------------------------------------------------|-------|---------------|
| GO:0005515                                | Protein binding                                  | 842   | 2.67x10\textsuperscript{-11} |
| GO:00199901                               | Protein kinase binding                            | 60    | 1.10x10\textsuperscript{-4}   |
| GO:0003677                                | DNA binding                                      | 180   | 1.19x10\textsuperscript{-4}   |
| GO:0005524                                | ATP binding                                      | 163   | 1.33x10\textsuperscript{-4}   |
| GO:0004674                                | Protein serine/threonine kinase activity          | 50    | 9.04x10\textsuperscript{-4}   |
| GO:0004842                                | Ubiquitin-protein transferase activity            | 45    | 9.53x10\textsuperscript{-4}   |
| GO:0003697                                | Single-stranded DNA binding                      | 18    | 0.001422       |
| GO:0016874                                | Ligase activity                                  | 38    | 0.001512       |
| GO:0043130                                | Ubiquitin binding                                | 16    | 0.00157        |
| GO:0042802                                | Identical protein binding                        | 85    | 0.002182       |
'centrosome' and 'spindle' terms were enriched in the DEGs (Table I).

In the GO MF category, the 'protein binding', 'protein kinase binding', 'DNA binding', 'ATP binding', 'protein serine/threonine kinase activity', 'ubiquitin-protein transferase activity', 'single-stranded DNA binding', 'ligase activity', 'ubiquitin binding' and 'identical protein binding' terms were enriched (Table I).

In addition, DEGs were primarily associated with the 'cell cycle', 'DNA replication', 'fanconi anemia pathway', 'mismatch repair', 'homologous recombination', 'nucleotide excision repair', 'progesterone-mediated oocyte maturation', 'oocyte meiosis' and 'base excision repair' pathways, determined by KEGG analysis (Fig. 3).

PPI network and module selection. The PPI network consisted of 685 nodes and 4,603 edges (Fig. 4A). According to the degree for each gene, the top 10 hub genes were cyclin dependent kinase 1 (CDK1), cell division cycle 20, cyclin B1 (CCNB1), cyclin B2 (CCNB2), mitotic arrest deficient 2 like 1 (MAD2L1), aurora kinase B, BUB1 mitotic checkpoint serine/threonine kinase (BUB1), cyclin A2 (CCNA2), centromere protein E and kinesin family member 2C (Table II). Using the MCODE Cytoscape plugin, four significant modules were identified (Fig. 4B-E). An enrichment analysis of the genes involved in the top significant modules was performed. The results identified that the DEGs in modules were principally related to ubiquitin mediated proteolysis, cell cycle, oocyte meiosis, progesterone-mediated oocyte maturation, HTLV-I infection, p53 signaling pathway and Epstein-Barr virus infection.

Validation with TCGA data. Information regarding 57,000 genes from 424 patients was included in the TCGA mRNA-sequencing expression data. Patients were sorted into high or low expression groups according to the median expression value. As CDK1, CCNB1, CCNB2, MAD2L1, BUB1 and CCNA2 were enriched in the 'cell cycle' and the 'progesterone-mediated oocyte maturation' pathways, the significance of these six hub genes in prognosis with TCGA data was subsequently validated with the Kaplan-Meier method. It was identified that the high expression of CDK1, CCNB1, CCNB2, MAD2L1, BUB1 and CCNA2 was associated with a decreased overall survival for patients with HCC (P=0.00116, 0.00021, 0.04889, 0.00222, 0.00015 and 0.00647, respectively; Fig. 5).

Discussion

Different from the previous studies conducted by Yildiz et al (8) and He et al (9), which analyzed the key genes and pathways for HCC developing from liver cirrhosis regardless of etiology, the present study was conducted specifically with HBV-positive samples. Therefore, drugs targeting these identified key genes and pathways may be considered for HCC therapy.

The hub genes CDK1, CCNB1, CCNB2, MAD2L1, BUB1 and CCNA2 may be suitable for the diagnosis and treatment of HBV-associated HCC. Garnier et al (15) identified that CDK1 may enhance liver regeneration by promoting rapid and efficient hepatocyte proliferation in a rat model, which could promote mutations in the β-catenin gene, potentially promoting...
Figure 4. PPI network of DEGs. (A) Constructed PPI network of DEGs.
Figure 4. Continued. (B) Module 1 and (C) Module 2.
the progression to HCC (16). Zhang et al (17) suggested that miR-582-5p inhibits the proliferation of HCC cells by targeting CDK1. However, in a comparative study of HCC and cirrhosis, Masaki et al (18) observed that the expression of CDK1 did not differ between HCC and the adjacent cirrhotic tissues in all the tested patients with hepatitis C virus-RNA. Therefore, alterations of the role of CDK1 in HCC, depending on different viral types, require further evaluation. Wang et al (19) demonstrated that the increased expression of CCNB1 and CCNB2 was associated with the restoration of the young regenerating liver phenotype. Another previous study identified that CCNB1 was associated with G2/M phase cell cycle arrest-induced growth inhibition, suggesting that this gene may be suitable as an anti-cancer agent in future therapy (20). Sze et al (21) suggested that mitotic arrest deficient (MAD)1 interacted with MAD2, leading to its retention in the cytoplasm, which may oppose mitotic checkpoint control in hepatocarcinogenesis. Similarly, Sze et al (22) suggested that MAD2 deficiency may cause a mitotic checkpoint defect in hepatoma cells. Wills et al (23) suggested that BUB1, a loss-of-heterozygosity driver gene, may be implicated in the germline causes of genetic instability.

Table II. Core identified differentially expressed genes and their corresponding degrees.

| Gene         | Degree |
|--------------|--------|
| CDK1         | 129    |
| CDC20        | 101    |
| CCNB1        | 93     |
| CCNB2        | 81     |
| MAD2L1       | 77     |
| AURKB        | 76     |
| BUB1         | 68     |
| CCNA2        | 67     |
| CENPE        | 65     |
| KIF2C        | 64     |

CDK1, cyclin dependent kinase 1; CDC20, cell division cycle 20; CCNB1, cyclin B1; CCNB2, cyclin B2; MAD2L1, mitotic arrest deficient 2 like 1; AURKB, aurora kinase B; BUB1, BUB1 mitotic checkpoint serine/threonine kinase; CCNA2, cyclin A2; CENPE, centromere protein E; KIF2C, kinesin family member 2C.
in polycystic liver disease. Furthermore, Wang et al. (24) identified that BUB1B activated parathyroid hormone like hormone to promote the G-protein-coupled receptor-induced loss of cell adhesion. CCNA functions in the S and G2-M phases of the cell cycle, and CCNA2 overexpression is associated with carcinogenesis in the liver (25-27). Masaki et al. (18) suggested that the activation of CCNA expression may serve an essential role in the transformation of cirrhosis to HCC. Wang et al. (28)
additionally identified that viral insertion was able to contribute to tumorigenesis in HBV-associated HCC by disrupting the control of the CCNA gene. Furthermore, Nault et al (29) identified that adeno-associated virus type 2 was associated with oncogenic insertional mutagenesis in human HCC, including in CCNA2. Yang et al (30) observed that bile acid-activated farnesoid X receptor stimulated the miR-22 silencing of CCNA2, exerting a protective effect in HCC cells. It is noteworthy that CDK1, CCNB1, CCNB2 and MAD2L1 were enriched in the ‘p53 signaling pathway’. Cellular tumor antigen p53 (p53), discovered in 1979, suppresses cell growth and oncogenic transformation (31). The disruption of p53 protein expression is an important factor underlying tumorigenesis in humans. Reportedly, ≤50% of all malignant tumors exhibit p53 mutations (32). Furthermore, the p53 response pathway is frequently defective in HCC (33), and p53 status is associated with the prognosis of patients with HCC (34). This is consistent with the poor prognosis associated with the high expression of CDK1, CCNB1, CCNB2 and MAD2L1 in the present analysis.

The present study demonstrated that dysregulation of the cell cycle and progesterone-mediated oocyte maturation pathways was closely associated with the development and progression of HCC. It was established that the dysregulation of the cell cycle pathway is an initiating event in cancer (35). Cell cycle checkpoints ensure normal genetic stability, the dysregulation of which may cause deviant cell proliferation leading to tumorigenic mutations (36). Wong et al (37) identified that an abnormal cell cycle is crucial in the development of cancer, subsequent to analyzing cancer samples from early to late stage. Furthermore, the dysfunction of cell cycle regulators and checkpoint mediators in HCC was discussed in previous studies (38-40). Dituri et al (39) demonstrated that alterations in the expression of cell cycle checkpoint-associated proteins frequently occur during HCC development. Yang et al (41) identified that short hairpin RNAs were able to suppress HBV-associated HCC cell proliferation through an effect on the cell cycle pathway. Previous studies concerning the function of the progesterone-mediated oocyte maturation pathway in liver cancer development are limited. In agreement with the study conducted by Jin et al (42), the progesterone-mediated oocyte maturation pathway was significantly enriched in the DEGs in the present study. In addition, Duckworth et al (43) identified that protein kinase A negatively regulates cell division cycle 25C during oocyte maturation. Further studies investigating the function of the progesterone-mediated oocyte maturation pathway in liver cancer are required.

In summary, the present study provided a comprehensive analysis of the DEGs between HCC and cirrhotic liver tissue samples through bioinformatics analysis to identify the potential genes and pathways involved in the transformation of HBV-positive liver cirrhosis to HCC. The identified genes and pathways may be suitable for use as therapeutic targets or novel biomarkers for prognosis. However, further studies are required to confirm these results and determine the underlying mechanisms.

Acknowledgements

Not applicable.

Funding

The present study was supported by the Science and Technology Planning Project of Guangdong (Guangdong, China; grant no. 2014A020212532).

Availability of data and materials

The datasets used and/or analyzed during the present study are available from the corresponding author on reasonable request.

Authors’ contributions

QC, JX and PW conceived and designed the study. QC, JX, WL, LS and TH analyzed the data. TH performed literature searches. QC, JX and TH wrote the paper. QC, WL, LS, TH and PW reviewed and edited the manuscript. All authors read and approved the manuscript.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Chen QF, Jia ZY, Yang ZQ, Fan WL and Shi HB: Transarterial chemoembolization monotherapy versus combined transarterial chemoembolization-microwave ablation therapy for hepatocellular carcinoma tumors ≤5 cm: A propensity analysis at a single center. Cardiovasc Interv Radiol 40: 1748-1755, 2017.
2. El-Serag HB: Epidemiology of viral hepatitis and hepatocellular carcinoma. Gastroenterology 142: 1264-1273.e1, 2012.
3. Levrero M and Zucman-Rossi J: Mechanisms of HBV-induced hepatocellular carcinoma. J Hepatol 64 (1 Suppl): S84-S101, 2016.
4. Parkin DM: The global health burden of infection-associated cancers in the year 2002. Int J Cancer 118: 3030-3044, 2006.
5. Zhou HY, Luo Y, Chen WD and Gong GZ: Hepatitis B virus mutation may play a role in hepatocellular carcinoma recurrence: A systematic review and meta-regression analysis. J Gastroenterol Hepatol 30: 977-983, 2015.
6. Intaraprasong P, Siramolpiwat S and Vilaichone RK: Advances in management of hepatocellular carcinoma. Asian Pac J Cancer Prev 17: 3697-3703, 2016.
7. Sherman M: Hepatocellular carcinoma: Epidemiology, surveillance, and diagnosis. Semin Liver Dis 30: 3-16, 2010.
8. Yildiz G, Arslan-Ergul A, Bagislar S, Konu O, Yuzugullu H, Gursoy-Yuzugullu O, Ozturk N, Ozen C, Ozdag H, Erdal E, et al: Genome-wide transcriptional reorganization associated with senescence-to-immortality switch during human hepatocellular carcinogenesis. PLoS One 8: e64016, 2013.
9. He B, Yin J, Gong S, Gu J, Xiao J, Shi W, Ding W and He Y: Bioinformatics analysis of key genes and pathways for hepatocellular carcinoma transformed from cirrhosis. Medicine (Baltimore) 96: e6938, 2017.
10. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U and Speed TP: Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4: 249-264, 2003.
11. Gautier L, Cope L, Bolstad BM and Irizarry RA: Affy-analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 20: 307-315, 2004.
27. Nault JC, Datta S, Imbeaud S, Franconi A, Mallet M, Couchy G, Letouzé E, Pilati C, Verret B, Blanc JF, et al: Recurrent AAV2-related insertional mutagenesis in human hepatocellular carcinomas. Nat Genet 47: 1187-1193, 2015.

28. Yang F, Hu Y, Liu HX and Wan YJ: MiR-22-silenced cyclin A expression in colon and liver cancer cells is regulated by bile acid receptor. J Biol Chem 290: 6507-6515, 2015.

29. Olson DC and Levine AJ: The properties of p53 proteins selected for the loss of suppression of transformation. Cell Growth Diff 5: 61-71, 1994.

30. Wang L, Huang J, Jiang M, Lin H, Qi L and Diao H: Activated PTHLH coupling feedback phosphoinositide to G-protein coupled receptors. J Hum Genet 24: 1707-1714, 2016.

31. Olivier M, Hollstein M and Hainaut P: TP53 mutations in human cancers: Origins, consequences, and clinical use. Cold Spring Harb Perspect Biol 2: a001008, 2010.

32. Wong YH, Chen RH and Chen BS: Core and specific network markers of carcinogenesis from multiple cancer samples. J Theor Biol 362: 17-34, 2014.

33. Wills ES, Crossen WR, Veltman JA, Woestenek R, Furuta M, Kozaki K, Tanimoto K, Tanaka S, Arii S, Antonaci S, Falasca M and Giannelli G: PI3K class IB controls the cell cycle checkpoint promoting cell proliferation in hepatocellular carcinoma. Cell Cycle 17: 449-456, 2016.