Abstract. Hepatitis B virus (HBV) infection is a critical factor for the initiation and progression of hepatocellular carcinoma (HCC). Gene expression profiles for HBV-associated HCC may provide valuable insight for the diagnosis and treatment of this type of HCC. The present study aimed to screen the differential genes in human HCC tissues based on high-throughput sequencing and to predict the potential therapeutic targets. Total mRNA was extracted from human HCC tissues and paracancerous tissues and sequenced using the Hiseq4000 sequencing platform. Differential gene expressions were screened and further analyzed using quantitative PCR and immunohistochemistry. A total of 2,386 differentially expressed genes were screened. Of these, 1,119 were upregulated and 1,267 were downregulated in paracancerous tissues compared with tumor tissues. Gene Ontology term analysis demonstrated that differentially expressed genes were involved in carboxylic acid catabolism, monocarboxylic acid metabolic processes and α-amino acid metabolic processes. Molecular functional analysis revealed that the differentially expressed genes functioned in oxidoreductase activity, for example acting on CH-OH group of donors and permitting identical protein binding, anion binding, coenzyme binding and monocarboxylic acid transporter activity. The Kyoto Encyclopedia of Genes and Genomes analysis reported that the differentially expressed genes were primarily concentrated in 20 signaling pathways, such as valine, leucine and leucine degradation, retinol metabolism and the cell cycle. Differential expression of proteins regulating the cell cycle, including stratifin, cyclin B1 and cyclin-dependent kinase 1, were significantly higher in tumor tissue compared with those in paracancerous tissue at both the mRNA and protein levels. These results were consistent with those obtained from high-throughput sequencing, indicating the reliability of the high-throughput sequencing. Together, these results identified differentially expressed genes and predicted the subsequent signaling pathways, which may be involved in the occurrence and development of HCC. Therefore, the present study may provide novel implications in the therapeutic and diagnosis of HCC.

Introduction

Hepatocellular carcinoma (HCC) is the most common type of primary hepatocellular carcinoma, accounting for 90% of primary hepatocellular carcinoma cases, and is the leading cause of cancer-associated death worldwide (1). Globally, ~350 million people are infected with hepatitis B virus (HBV), and HBV infection accounts for at least 50% of HCC cases (2). HBV-associated HCC is therefore frequently observed. Notable progress has been made in the diagnosis and treatment of HCC, and overall treatment efficacy has been improved, especially following the development of multi-modality therapy, such as hepatectomy, minimally invasive treatment and liver transplantation (3). However, most patients with HCC are diagnosed during the middle and advanced stages, during which traditional radiotherapy and chemotherapy have limited clinical benefit. This is primarily due to high rates of recurrence and metastasis (4), therefore, it is important to identify novel diagnostic and prognostic markers to aid earlier intervention.

The occurrence and development of HCC are caused by genetic and epigenetic variations, including gene mutations, copy number variations and abnormal methylation (5).
High-throughput technology has identified a large number of tumor driver gene mutations and epigenetic alterations associated with the pathogenesis of HCC (6-10), which may have potential as diagnostic and prognostic molecular markers (11,12). Mutations in key driver genes lead to uncontrolled proliferation and clonal amplification of cancer cells (13-15). However, differentially expressed genes in HCC have not been screened and the subsequent signaling pathways involved in the development of HCC have not been validated.

The present study aimed to identify potentially novel therapeutic targets for HCC using high-throughput-based gene expression analysis in human HCC tissues. To prevent false-positive results from high-throughput screening, immunohistochemical analysis was also performed to verify differential gene expression.

Materials and methods

**Tissue collection.** The samples were collected from eight patients with HBV-associated HCC at the Department of Hepatobiliary Gland Extracorporeal Surgery of Guigang People's Hospital (Guigang, China) between January 2017 and March 2017. The age of the patients ranged from 38 to 61 years (median age, 48.5 years; male: Female ratio of 5:3), and the primary HCC tissue specimens and matched paracancerous liver tissues were collected during hepatectomy. Clinicopathological characteristics of patients, including age and serum levels of bilirubin, alkaline phosphatase, albumin, alanine transaminase (ALT), aspartate transaminase (AST), alphafetoprotein (AFP), hepatitis B surface antigen (HBsAg), hepatitis B surface antibody (HBsAb), hepatitis B e antigen (HBeAg), hepatitis B e antibody (HBeAb) and hepatitis B core antibody (HBcAb) were collected. HCC was confirmed by a pathological report from Guigang People's Hospital based on the evaluation of differentiation and metastasis as previously described (16). The tissues were divided into two groups: Paracancerous tissue group and hepatocellular carcinoma tissue group. The inclusion criteria were as follows: i) No etiological evidence of hepatitis C virus, type 2 diabetes mellitus, fatty liver, alcoholic liver or aflatoxin exposure history; ii) no preoperative radiotherapy, digital subtraction angiography intervention, antineoplastic drug therapy or and radiofrequency ablation; and iii) no other history of cancer. Patients who did not meet one of the aforementioned criteria were excluded. Three tumor tissues and corresponding paracancerous tissues were analyzed using high-throughput sequencing. Three differentially expressed genes, stratin (SFN), cyclin B1 (CCNB1) and cyclin-dependent kinase 1 (CDK1) were verified in the total 8 tumor tissues and corresponding paracancerous tissues.

After the specimens were removed, tumors and paracancerous liver tissues were cut into small pieces (~2 mm) on ice, then put into 1.5 ml nucleic acid-free centrifugal tubes and stored in liquid nitrogen. Cancer tissues and paracancerous tissues were collected from the non-necrotic area of the cancer, and paracancerous liver tissues were taken from the non-invasive area 2 cm away from the edge of the obvious mass, avoiding the burning site of the scalpel and preventing the loss of DNA and RNA due to high temperature. Patients provided informed written consent before organizing the collection, and the study was approved by the Ethics Committee of Guigang City People's Hospital.

**Construction of the Illumina sequencing library.** Total RNA was extracted from the samples using TRIzol® (Invitrogen; Thermo Fisher Scientific, Inc.). The concentration and purity of the RNA were detected using Nanodrop 2000 (Thermo Fisher Scientific, Inc.), the integrity of RNA was detected by agarose gel electrophoresis (0.5% gel; ethidium bromide staining and visualization at UV light), and the value of RNA integrity was determined using Agilent 2100 (Agilent Technologies, Inc.). The total amount of RNA required for single library construction was >5 µg, the concentration was >200 ng/µl, and optical density 260/280 was between 1.8 and 2.2. Then, a Ribo-Zero Magnetic kit (Epicentre; Illumina, Inc.) was used to remove ribosomal RNA, RNase R (Epicentre; Illumina, Inc.) was used to remove linear RNA and a TruSeq™ Stranded Total RNA Library Prep kit (Illumina, Inc.) was used to construct Paired-End sequencing library. The analysis of mRNA was performed using the Hiseq4000 sequencing platform (Illumina, Inc.). Principal component analysis (PCA) was used to analyze multidimensional data.

**Sequence data analysis.** SepPrep and Silkle software (linux-64 v1.3.2; Anaconda Clound) were used to examine the data quality and the data obtained after quality control (Phred quality score) with the reference genome data were compared using Bowtie version 2. KNIEF and CIRCexplorer2 were used to predict the expression of RNA, and then levels of RNA in the samples were calculated.

**Differentially expressed gene enrichment analysis.** Gene Ontology software was used for enrichment analysis of GO functions, including molecular function, cell components, and biological processes, and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was carried out using the KEGG Orthology-Based Annotation System and KEGG annotation (https://www.genome.jp/kegg-bin/get_htext?ko00001).

**Prediction of target genes.** Differential analysis of gene expression was performed using DESeq version 2 (Bioconductor) to predict target genes of RNA.

**Quantitative (q)PCR.** According to the results of high-throughput sequencing, three differentially expressed RNAs (SFN, CCNB1 and CDK1) were analyzed using fluorescence qPCR. The primer sequences used are listed in Table I. Total mRNA was amplified using a one-step RT-PCR kit (00081405, CWBio). The primers were added into a 25-µl ULtraSYBR mixture (01170, CWBio) PCR reaction system according to the manufacturer's protocol. The thermocycling conditions were as follows: Pre-denaturation at 95°C for 10 min, 95°C denaturation for 10 sec, 58.5°C annealing for 30 sec and 72°C extension for 30 sec (40 cycles). The Cq value for each gene was determined and expression levels of target genes were calculated using the 2^-ΔΔCq method (17). mRNA expression of target genes were normalized to GAPDH.

**Immunohistochemistry.** Cancer tissues were fixed with 4% paraformaldehyde overnight at 4°C. The tissues were then dehydrated using 70, 80 and 90% ethanol, and mixed with anhydrous ethanol and xylene for 15 min, xylene 1 for 15 min and
xylene II for 15 min (until transparent). Tissues were embedded in paraffin and sliced into 10-µm sections. The paraffin slices were dewaxed and hydrated in 70, 75, 80, 85 and 95% alcohol. Then, 3% (v/v) H₂O₂ was used to block endogenous peroxidase activity for 5 min at room temperature. Immunostaining of the slides was performed using the following antibodies overnight at 4˚C: Rabbit polyclonal anti-SFN (1:50, bs-20373R, BIOSS), rabbit polyclonal anti-CCNB1 (1:400, bs-20373R, BIOSS) and rabbit polyclonal anti-CDK1 (1:100, bs-20373R, BIOSS). The slides were then washed with PBS and incubated with horse-radish peroxidase-labeled goat anti-rabbit IgG secondary antibody (1:10,000; A16104SAMPLE; Thermo Fisher Scientific, Inc.) for 30 min at room temperature, and visualized with 3,3′-diaminobenzidine chromogen for 3 min at room temperature. At least four fields were taken from each image using a light microscope (magnification, x200; BX51, Olympus Corporation). Staining intensity was analyzed by Image-Pro Plus software (National Institutes of Health). The relative expressions of target proteins were normalized to the negative control (without primary antibody).

**Results**

**Basic information of the patients.** Eight patients with HCC were enrolled in the present study. The age of the patients ranged from 38 to 61 years. The mean values of bilirubin, alkaline phosphatase, albumin, ALT, AST and AFP serum levels were 15.0 µM, 92.1 U/l, 44.6 g/l, 31.5 U/l, 33.6 U/l and 785.3 ng/ml, respectively. HBsAg, HBeAb and HBcAb were positive in all patients. The patients were in a stage of medium- and low-differentiation without lymph node involvement and distal metastasis. All patients were diagnosed with cirrhosis and portal vein embolism, but without ascites (data not shown).

**Quality analysis of the data.** The data quality control analysis showed that the total number of reference reads no. Bases, bp Q30, bp N, % Q20, % Q30, %

<table>
<thead>
<tr>
<th>Sample</th>
<th>Controls 1</th>
<th>Controls 2</th>
<th>Controls 3</th>
<th>Cancer 1</th>
<th>Cancer 2</th>
<th>Cancer 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reads no.</td>
<td>43,704,194</td>
<td>41,440,988</td>
<td>42,235,972</td>
<td>40,926,974</td>
<td>44,150,718</td>
<td>42,218,858</td>
</tr>
<tr>
<td>Bases, bp</td>
<td>6,599,333,294</td>
<td>6,257,589,188</td>
<td>6,377,631,772</td>
<td>6,179,973,074</td>
<td>6,666,758,418</td>
<td>6,375,047,558</td>
</tr>
<tr>
<td>Q30, bp</td>
<td>6,226,745,352</td>
<td>5,914,658,053</td>
<td>6,099,059,319</td>
<td>5,744,415,761</td>
<td>6,218,604,371</td>
<td>5,979,826,518</td>
</tr>
<tr>
<td>N, %</td>
<td>0.001709</td>
<td>0.001422</td>
<td>0.001376</td>
<td>0.000725</td>
<td>0.001765</td>
<td>0.001859</td>
</tr>
<tr>
<td>Q20, %</td>
<td>97.73</td>
<td>97.81</td>
<td>98.28</td>
<td>97.11</td>
<td>97.24</td>
<td>97.5</td>
</tr>
<tr>
<td>Q30, %</td>
<td>94.35</td>
<td>94.51</td>
<td>95.63</td>
<td>92.95</td>
<td>93.27</td>
<td>93.82</td>
</tr>
</tbody>
</table>

Controls 1, 2 and 3 were the paracancerous liver tissues of cancer samples 1, 2 and 3, respectively.

**Table II. Quality analysis of the original data.**

**Statistical analysis.** All data were analyzed using SPSS version 19.0 (IBM Corp.) The differences between groups were analyzed using unpaired Student's t-tests and P<0.05 was considered to indicate a statistically significant difference.

**Table I. Primer sequences.**

<table>
<thead>
<tr>
<th>Genes</th>
<th>Primers, 5'-3'</th>
<th>Primer length, bp</th>
<th>Product length, bp</th>
<th>Annealing, °C</th>
</tr>
</thead>
<tbody>
<tr>
<td>SFN</td>
<td>Forward GGTGACTACTACCGCTACCTGG</td>
<td>22</td>
<td></td>
<td>61.4</td>
</tr>
<tr>
<td></td>
<td>Reverse GGCATCTCCTCCTTGCTGAGC</td>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCNB1</td>
<td>Forward TTGAGGAAAGAGCAAGACGTC</td>
<td>20</td>
<td></td>
<td>56.9</td>
</tr>
<tr>
<td></td>
<td>Reverse AACCGATCAATAATGGAGACAG</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDK1</td>
<td>Forward AAGATGTGCTTATGCAAGATTTC</td>
<td>22</td>
<td></td>
<td>57.5</td>
</tr>
<tr>
<td></td>
<td>Reverse CATGTACTGACCAGGGG</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAPDH</td>
<td>Forward CAATGACCCCTTCATTGACC</td>
<td>20</td>
<td></td>
<td>57.2</td>
</tr>
<tr>
<td></td>
<td>Reverse GAGAAGCTCCCGTCTCAG</td>
<td>20</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
genome sequences was >90% (Table III), indicating that the quality control of sequencing data was good and could be used in subsequent data analysis.

**PCA analysis.** PCA analysis showed that there were obvious differences between the control group and cancer group. By contrast, there was no inner difference among the data in each group. Thus, the data could be used in the subsequent analysis (Fig. 1).

**Differential gene expression analysis.** Differential gene expression analysis is presented in Fig. 2. 2,386 differentially expressed genes were screened (multiple of differentially expressed genes between samples was >1, and the value of differentially expressed genes in samples was <0.05). There were 1119 downregulated and 1,267 upregulated genes in cancer tissues compared with paracancerous tissues.

**GO functional enrichment analysis.** GO functional enrichment analysis of differentially expressed genes was used to identify GO functional items. Firstly, all genes were mapped to each term in GO database and the number of differentially expressed genes was calculated. Then, significantly enriched differentially expressed genes with the whole genome were analyzed. The GO analysis showed that differentially expressed genes were enriched in the following GO molecular functions: GO: 0046395 carboxylic acid catabolic process, GO: 0032787 monocarboxylic acid metabolic process and GO: 1901605 α-amino acid metabolic processes. Genes were also enriched in the following cellular component functions: GO: 0070062 extracellular exosome, GO: 0044444 cytoplasmic part and GO: 0044459 plasma membrane part. Differentially expressed genes were enriched in the following biological processes: GO: 0016614 oxidoreductase activity, acting on CH-OH group of donors, GO: 0042802 identical protein binding, GO: 0043168 anion binding, GO: 0050662 coenzyme binding and GO: 0008028 monocarboxylic acid transmembrane transporter activity (data not shown).

**Enrichment analysis of KEGG signaling pathways of differentially expressed genes.** Enrichment analysis of KEGG signaling pathways of differentially expressed genes was based on the whole genome and is shown in Fig. 3. The metabolic pathways and signaling pathways in which differentially expressed genes functioned were determined. The results showed that the differentially expressed genes were mainly concentrated in 20 signaling pathways, including degradation of valine, leucine and isoleucine, retinol metabolism and the cell cycle.
Validation of differentially expressed genes. The validation of three differentially expressed genes is shown in Fig. 4. The results showed that the expression of SFN, CCNB1 and CDK1 genes in tumors were upregulated in cancer tissues compared with control (P<0.05; Figs. 4A and B, S1), which was consistent with the results of high-throughput sequencing. The results were further confirmed using qPCR and a consistent result was obtained (vs. control; P<0.05; Fig. 4C).

Discussion

In the present study, differential gene expression between HCC and paracancerous tissues was analyzed. The differentially expressed genes were mainly concentrated in 20 signaling pathways, such as valine, leucine and isoleucine degradation, retinol metabolism and the cell cycle. It was then further confirmed that cell cycle-associated genes (SFN, CCNB1 and CDK1) were abnormally expressed in HCC tissues. The present study may provide novel therapeutic targets for HCC.

Due to the lack of early diagnostic measures, the incidence of recurrence and metastasis after operation in HCC is high, and drug treatments have limited therapeutic benefit (18,19). Therefore, early diagnosis and prediction are important to improve the treatment of HCC. Similar to other malignant tumors, HCC is often caused by the activation of oncogenes (myc, ras) or the inactivation of tumor suppressor genes (e.g., p53) (20,21). The aim of the present study was to detect differentially expressed genes in HCC and paracancerous tissues using high-throughput sequencing.

Tumor development is usually associated with multiple signaling pathways (22). KEGG analysis showed that the differentially expressed genes were mainly concentrated in 20 signaling pathways, such as valine, leucine and isoleucine degradation, retinol metabolism and the cell cycle. In addition, 2,386 differentially expressed genes were screened. The differentially expressed genes of SFN, CCNB1 and CDK1 were selected from the high throughput results for further verification. SFN, CCNB1 and CDK1 genes expressed differently in high-throughput results, were associated to the cell cycle (23).
Experiments have shown that knocking out the SFN gene can lead to cell failure to maintain stable G2/M cycle after DNA damage and sensitizes cells to DNA damage. This method is therefore used in the treatment of neuroblastic tumor (24). SFN can arrest cell cycle and induce apoptosis of human tumor cells (25). Increased expression of SFN results in the chelation of CDK1/cyclinB1 in cytoplasm, thus blocking the interaction between cell division cycle 2 (CDC2) gene and CDK1 and preventing cells from entering into cell mitosis (26-28).

CDK1 is a member of the protein kinase family and is encoded by the CDC2 gene. At the late stage of G2, CDK1 and cyclin B1 lotus root synthesize complex, forming mitotic promoter factor, which can promote cell cycling from G2 to M (29). CDK1 is an important regulator of mitotic initiation, cell cycling and metastasis. High expression of active CDK1 promotes G2/M expression and accelerates cancer cell growth (30). Some studies have shown that specific inhibitors of CDK1 can induce reversible dormancy of human cells in G2/M phase, leading to the apoptosis of cancer cells, suggesting that selective inhibitors of CDK1 may play a pivotal role in the treatment of cancer (30,31).

CCNB1 is an important regulator of the cell cycle associated with the detection point of G2/M phase. Interaction of CCNB1 and CDK1 phosphorylates the substrate cell division homologous protein cyclin 25, initiates cell cycle progression from G1/S phase to G2/M phase and promotes mitosis (32). A previous study reported that the abnormal expression of CCNB1 is associated with abnormal cell proliferation and tumorigenesis (33). A number of studies have shown that CCNB1 is highly expressed in breast, lung and gastrointestinal cancer (34-36). In the present study, the results of qPCR and immunohistochemistry demonstrated that the expression of SFN, CCNB1 and CDK1 was elevated in cancer tissues and that these results were consistent with the high-throughput results.

There were still limitations of the present study. First, only three tumor tissues and corresponding paracancerous tissues were analyzed using high-throughput sequencing. Three differentially expressed genes, SFN, CCNB1 and CDK1 were verified in the total 8 tumor tissues and corresponding paracancerous tissues. Consistent results were obtained from high-throughput sequencing, PCR and western blotting; however, the differentially expressed genes and their proteins need external validation in a larger cohort of patients with cirrhosis and different stages of HCC. Second, although differentially expressed genes in HCC were screened, the potential value of these genes in the treatment and prognosis of HCC were not confirmed. Third, there were still some discrepancies regarding the differentially expressed genes reported in the present study and previous studies. These might be caused by different patient ethnicities, types of HCC and disease stages. However, larger sample size is crucial and laboratory experiments should be carried out to validate these findings and disclose potential mechanisms.
In conclusion, differentially expressed genes were screened using high-throughput technology and enriched in tumor-associated signaling pathways and functions. The next step is to investigate the functions of these differentially expressed genes.

Acknowledgements

Not applicable.

Funding

Construction Funding of Key Clinical Specialties Guangxi Zhuang Autonomous Region (Z20190320).

Availability of data and materials

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

Authors' contributions

HZ, YH, WQ, PC, LH, WZ, LL and HL performed the experiments and analyzed the data. HZ and XQ designed the study and wrote the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

All patients provided informed consent before tissue collection, and the study was approved by the Ethics Committee of Guigang City People's Hospital (approval no. 013142017).

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References