

Exome capture sequencing reveals new insights into hepatitis B virus-induced hepatocellular carcinoma at the early stage of tumorigenesis

YONG CHEN^{1,2}, LIJUAN WANG¹, HEXIANG XU¹, XINGXIANG LIU¹ and YINGREN ZHAO²

¹Institute of Liver Diseases, Huai'an Fourth People's Hospital, Huai'an, Jiangsu 223002; ²Department of Infectious Diseases, The First Affiliated Hospital of the Medical College of Xi'an Jiaotong University, Xi'an, Shaanxi 710061, P.R. China

Received May 23, 2013; Accepted July 15, 2013

DOI: 10.3892/or.2013.2652

Abstract. Hepatocellular carcinoma (HCC), the most common type of liver cancer, is the third primary cause of cancer-related mortality worldwide. The molecular mechanisms underlying the initiation and formation of HCC remain obscure. In the present study, we performed exome sequencing using tumor and normal tissues from 3 hepatitis B virus (HBV)-positive BCLC stage A HCC patients. Bioinformatic analysis was performed to find candidate protein-altering somatic mutations. Eighty damaging mutations were validated and 59 genes were reported to be mutated in HBV-related HCCs for the first time here. Further analysis using whole genome sequencing (WGS) data of 88 HBV-related HCC patients from the European Genome-phenome Archive database showed that mutations in 33 of the 59 genes were also detected in other samples. Variants of two newly found genes, *ZNF717* and *PARP4*, were detected in more than 10% of the WGS samples. Several other genes, such as *FLNA* and *CNTN2*, are also noteworthy. Thus, the exome sequencing analysis of three BCLC stage A patients provides new insights into the molecular events governing the early steps of HBV-induced HCC tumorigenesis.

Introduction

Hepatocellular carcinoma (HCC), the most common type of liver cancer, is the third primary cause of cancer mortality

worldwide (1). The approximately 750,000 new liver cancer cases and 700,000 deaths worldwide each year are a relatively high burden (2). Among primary liver cancer, HCC, which arises from hepatocytes, represents the major histological subtype, accounting for 70-85% of all cases worldwide (3). The common risk factors for HCC include viral hepatitis, alcohol, non-alcoholic fatty liver disease and toxin such as aflatoxin, hemochromatosis, α 1-antitrypsin deficiency (4). Chronic hepatitis B virus (HBV) infection accounts for approximately half of all cases of HCC (1), and it is the most common etiology of HCC in Asian countries.

Based on the current understanding of tumorigenesis, cancer is a genetic disease that arises from a single clone of cells expanding in an unregulated manner due to genomic instability and somatically acquired mutations (5,6). The pathogenesis of liver cancer is considered a multistep process. HCC may develop through a multistep process involving multiple genetic events over decades of chronic liver disease, which could facilitate the progressive accumulation of somatically genetic alterations. These newly somatic mutations, which trigger oncogenes and/or inactivate tumor-suppressor genes (TSGs), and preexisting endogenous virus- or chemical-induced mutations may lead to liver cancer, in particular HCC.

Recently, with markedly increased throughput, next-generation sequencing (NGS) technologies provide an efficient tool to identify the somatic mutations at exome and even the whole genome level (7). Several reports involving whole-genome and/or whole-exome sequencing on human HCC samples with viral infection were recently published, and provide a catalog of somatic mutations in the cancer genome, including candidate cancer genes in HCC (8-11). Although previous studies have revealed that certain genetic alterations, such as *TP53* and β -catenin mutations, occur in HCC cells, the molecular mechanisms underlying the initiation and formation of HCC remain obscure.

In the present study, with a whole exome capturing approach, we sequenced the exome of tumor and normal tissues from 3 HBV-induced early stage HCC patients, which was diagnosed to be in stage A of Barcelona Clinic Liver Cancer (BCLC) system. Further analysis using whole genome sequencing data of 88 HBV-related HCC patients from the European Genome-phenome Archive database were carried

Correspondence to: Professor Yong Chen, Institute of Liver Diseases, Huai'an Fourth People's Hospital, 128 Yan'an East Road, Huai'an, Jiangsu 223002, P.R. China
E-mail: xxchenguo@163.com

Professor Yingren Zhao, Department of Infectious Diseases, The First Affiliated Hospital of the Medical College of Xi'an Jiaotong University, 277 Yanta West Road, Xi'an, Shaanxi 710061, P.R. China
E-mail: zhaoyingren@sohu.com

Key words: hepatocellular carcinoma, hepatitis B virus, exome sequencing, somatic mutation

out to discover recurrently mutated genes. Our investigation here may provide new insights into the molecular mechanism of HBV-related HCC.

Materials and methods

Sample and DNA preparation. Three male patients aged 48–52 years with chronic hepatitis B infection were diagnosed as BCLC stage A HCC in 2011 at Huai'an Fourth People's Hospital. There was only 1 tumor nodule in each patient. The tumors were all on the right lobe of liver, with the size of 3x4, 2x4 and 1.5x3 cm², respectively. The matched normal tissues were obtained ~5 cm from the cancer tissues. Written informed consent from the patient was obtained, and the study was reviewed and approved by the Ethics Committees of Huai'an Fourth People's Hospital. DNA of these tissues was extracted through traditional phenol chloroform method.

Exome capture and sequencing. Exome sequencing was performed using SureSelect Human All Exon 38Mb kit (Agilent Technologies Inc., Santa Clara, CA, USA) according to the manufacturer's protocols. The capture region is ~38 Mb and covers 1.22% of the human genome that corresponds to the consensus coding sequence (CCDS) collection of genes. Genomic DNA was randomly fragmented by sonication in to an average size of 500 bp. A pair of adaptors was ligated to each end of DNA fragments. The adaptor-ligated templates were then hybridized to the array to capture fragments in target regions. The captured fragments were amplified, purified and subjected to paired-end sequencing on the Illumina GA IIx platform (Illumina, San Diego, CA, USA). The genomic DNA library preparation, targeted sequence capturing, and massively parallel sequencing were finished by Guangzhou iGenomics Co., Ltd.

Sequence analysis. The raw reads were filtered with Fastx-tools (http://hannonlab.cshl.edu/fastx_toolkit/index.html). Low quality reads were discarded (fractions of N bases ≥10% and fractions of bases with quality less than <50%). Then, software BWA (version 0.5.9) (12) was used to map the paired-end reads to the human reference genome (hg19). The reference genome was downloaded from the UCSC (University of California, Santa Cruz, CA, USA) database (<http://genome.ucsc.edu>). After the alignment, PCR duplications were removed by the SAMtools software package (version 0.1.16) (13). Candidate somatic variants were identified with the VarScan 2 software (version 2.2.8) and filtered with default parameters (14). Function prediction for all the missense single nucleotide variations (SNVs) were carried out using PROVEAN (15) and SIFT (16).

Validation of variants. Somatic indels, nonsense and missense SNVs which were predicted to be damaging with both prediction methods, were validated by PCR and Sanger sequencing. Primers were designed with Primer Premier 5 (Premier Biosoft International, Palo Alto, CA, USA). PCR amplification was performed with 50 µl reaction using the following procedure: 95°C for 2 min, 35 cycles at 95°C for 15 sec, 60°C for 20 sec and 72°C for 30 sec, followed by 72°C for 2 min. The PCR products were purified with E.Z.N.A.[®] Gel Extraction

kit. Sanger sequencing was performed on an ABI 3730 DNA Analyzer. Sequence trace files were manually analyzed.

Whole genome sequence analysis. In order to examine the prevalence of the mutated genes with validated damaging variants in various stages of HCCs, we further checked their mutation frequency in 88 HBV-related HCC patients using whole genome sequencing (WGA) data downloaded from the European Genome-phenome Archive database (<ftp://ftp.sra.ebi.ac.uk/>). The data set was not restricted to BCLC stage A patients. Early-to-advanced stage HCCs were included. Detailed information on the patients was provided in a previous study (8) which focused on HBV integration detection. Sequence alignment and somatic variants calling were carried out using the same procedure of exome sequencing reads analysis.

Results

Three patients with chronic HBV infection were diagnosed as BCLC stage A of HCC without metastasis in 2011. To identify the somatic mutations related to the disease, we performed the exome capture with Agilent SureSelect 38Mb kit for the 3 pairs of cancer and normal tissues. The captured samples with insertion size of 300–400 bp were subjected to paired-end sequencing using Illumina GA IIx. For the 3 pairs of samples, 34.93 Gb clean data were obtained (Table I). A total of 20.76 Gb data were mapped to the target region, achieving a minimum mean depth of 40.14 (Table I). The coverage of target region for each sample was all >95%. The capture rate for each sample ranged from 52.44 to 68.66% (Table I).

All uniquely mapped sequences (target regions and adjacent regions) were used for subsequent variant detection. Using VarScan 2 with default parameters, 461, 524 and 638 somatic variants were identified for the 3 samples, respectively (Table II). According to the annotation results, a total of 266 variants were found to be protein-altering, including 16 exonic indels, 12 nonsense SNVs, 227 missense SNVs and 12 splicing site mutations (Table II). Among these indels, we found 10 mutations leading to frameshift and 6 mutations causing amino acid residue deletion or insertion (Table II).

Function prediction results for missense SNVs and splicing sites showed that 82 missense SNVs (29 in S1, 16 in S2 and 37 in S3, respectively) were predicted to be damaging by both prediction methods. These SNVs, together with the indels and nonsense mutations, were further validated with PCR and Sanger sequencing. A total of 80 (74.1%, 80/108) variants in 78 genes were validated. The distribution pattern of these validated variants across the whole genome is shown in Fig. 1. Detailed information on these variants is listed Table III. Of the variants, 85% (68/80) have not been reported in the dbSNP database (version 137) before. Nineteen of the 78 genes were already reported to mutate in HCC by previous studies according to the Catalogue of Somatic Mutations in Cancer (COSMIC) database (version v64). The other 59 genes were first reported to mutate in HCC in this study. All the validated, mutated genes were annotated through the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis. Thirty genes are involved in various pathways, such as MAPK signaling pathway, Wnt signaling pathway and cell adhesion molecules (CAMs).

Table I. Data summary of exome sequencing.

Samples	S1		S2		S3	
	Tumor	Normal	Tumor	Normal	Tumor	Normal
Raw data (Gb)	4.91	4.32	7.76	7.64	7.62	7.96
Read length (bp)	100	100	100	100	100	100
Clean data (Gb)	3.64	3.16	7.01	6.96	6.92	7.24
Uniquely mapped on the genome (Gb)	3.17	2.76	6.43	6.23	6.45	6.57
Mapped on target region (Gb)	1.77	1.45	4.37	4.30	4.36	4.51
Capture rate (%)	55.98	52.44	67.99	68.90	67.66	68.66
Mean depth of target region (x)	49.03	40.14	85.01	83.53	84.82	87.69
Coverage of target region (%)	95.82	96.17	99.90	99.90	99.80	99.80

Table II. Statistics of the identified somatic variants for each sample.

Annotation	S1	S2	S3
Splicing	4	3	5
ncRNA_splicing	1	0	0
Downstream	3	7	6
UTR5	3	5	18
ncRNA_exonic	6	14	10
Intergenic	87	93	102
ncRNA_UTR5	0	1	0
Intronic	208	269	290
ncRNA_UTR3	0	0	1
Exonic	111	90	160
UTR3	10	14	26
Upstream	2	7	9
ncRNA_intronic	29	12	15
Frameshift insertion	0	2	1
Unknown	2	1	3
Nonframeshift insertion	1	0	1
Frameshift deletion	1	0	6
Stopgain SNV	2	3	7
Nonframeshift deletion	2	2	0
Nonsynonymous SNV	71	54	102
Synonymous SNV	32	28	40

Variants are annotated based on transcripts, i.e., one variant may be assigned as different types according to different transcripts. Therefore, overlap exists in different cells from the same column in this table.

These 80 somatic mutations would cause protein changes. Genes with these variants were defined as candidate HCC-related genes and were selected for further analysis. Whole genome sequence analysis was carried out to analyze the prevalence of the 78 mutated genes in 88 HBV-related HCC patients. For these patients, a minimum 751.49 million reads were mapped to the genome for each sample, yielding a minimum mean depth of 23.3

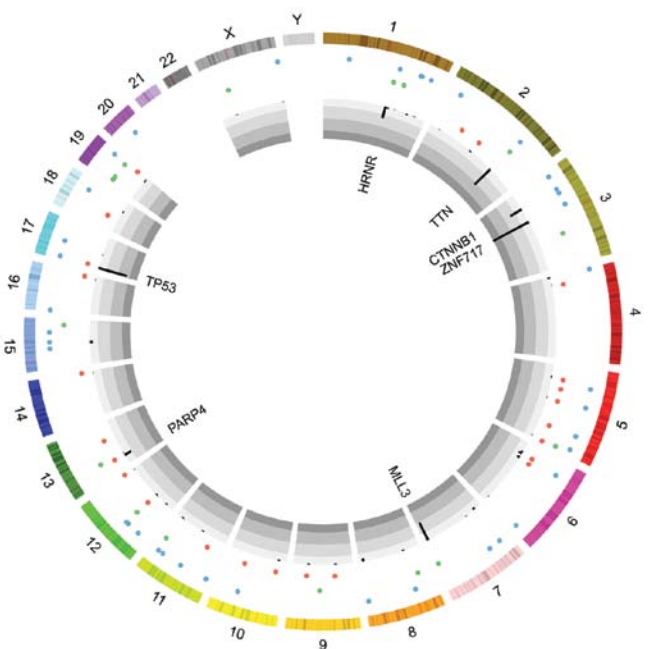


Figure 1. Distribution of validated protein-altering variants across the whole genome. Variants in S1, S2 and S3 are shown in red, green and blue dots respectively. The black bars show the mutation frequency in 88 HCC patients for genes with validated variants. Seven genes mutated in at least 9 patients (over 10%) are indicated. HCC, hepatocellular carcinoma.

and a minimum coverage of 99.03% of the whole genome. Forty-seven genes were found to be mutated in at least one patient. Thirty-three of them have not previously been reported in the COSMIC database.

Genes mutated in at least 9 patients of the 88 samples (>10%) are indicated in Fig. 1. Two of these seven genes, *ZNF717* and *PARP4* have not been reported in the COSMIC database before. *ZNF717* showed the highest frequency as variants of this gene were found in 47 of the 88 samples. Furthermore, the missense SNV in this gene was not reported in the dbSNP database before. This SNV was also detected in another patient of the 88 HCC samples. The variant in *PARP4* was also reported for the first time. According to the KEGG database, the protein encoded by *PARP4* is involved in the base excision repair pathway. The remaining 5 genes,

Table III. Information on the validated somatic mutations.

Gene	Chr	Position	Ref	Alter	Ref_aa	Alter_aa	Type	Sample	COSMIC	Num in 88 ^a	dbSNP
<i>ABCC9</i>	12	21954112	G	A	R	*	Nonsense	S3	No	2	NA
<i>ACTR5</i>	20	37394933	T	G	I	S	Missense	S3	Yes	0	rs34733925
<i>ACVRL1</i>	12	52309919	A	T	E	V	Missense	S1	No	0	NA
<i>ADAMTS12</i>	5	33576162	G	T	S	R	Missense	S1	Yes	2	NA
<i>ALPK2</i>	18	56149264	C	G	G	R	Missense	S3	No	2	NA
<i>ATG2A</i>	11	64676833	T	C	Y	C	Missense	S3	No	2	rs148849835
<i>ATP2C2</i>	16	84495388	T	A	D	E	Missense	S1	No	1	NA
<i>BCOR</i>	X	39933872	G	A	R	C	Missense	S2	No	1	NA
<i>C9orf40</i>	9	77563010	G	A	T	I	Missense	S2	No	0	NA
<i>CAT</i>	11	34482880	G	T	R	L	Missense	S1	No	0	rs140071186
<i>CD5L</i>	1	157805636	G	T	A	E	Missense	S3	No	2	NA
<i>CDCA2</i>	8	25319663	C	A	A	D	Missense	S2	No	0	NA
<i>CDHR2</i>	5	176017571	AGGAGA	.	QES	R	Deletion	S2	Yes	0	NA
<i>CDHR2</i>	5	176017568	C	CT	-	-	Frameshift	S2			NA
<i>CNTN2</i>	1	205034321	C	A	F	L	Missense	S3	No	2	NA
<i>CNTNAP3B</i>	9	43849812	T	G	C	G	Missense	S1	No	2	rs3864761
<i>COL22A1</i>	8	139825238	A	T	Y	N	Missense	S3	No	4	NA
<i>CTNNB1</i>	3	41266125	C	T	T	I	Missense	S3	Yes	16	rs121913413
<i>CUL9</i>	6	43152422	G	T	R	M	Missense	S1	Yes	4	NA
<i>DPP10</i>	2	116548913	T	A	Y	N	Missense	S1	Yes	2	NA
<i>EXD1</i>	15	41507990	G	A	Q	*	Nonsense	S3	Yes	0	NA
<i>FAM193A</i>	4	2673948	A	G	H	R	Missense	S3	No	4	NA
<i>FAM196B</i>	5	169309669	G	A	L	F	Missense	S3	No	0	NA
<i>FLNA</i>	X	153594965	C	A	V	F	Missense	S3	No	1	NA
<i>GPR125</i>	4	22390134	T	A	R	*	Nonsense	S1	No	1	NA
<i>HADHA</i>	2	26426943	T	G	Q	P	Missense	S3	No	1	NA
<i>HAPLN1</i>	5	82940402	A	T	D	E	Missense	S1	Yes	1	NA
<i>HLA-DRB5</i>	6	32489813	G	C	T	R	Missense	S3	No	5	rs79606458
<i>HNRNPH1</i>	5	179046361	C	A	G	W	Missense	S1	No	1	NA
<i>HRNR</i>	1	152188109	T	A	Q	L	Missense	S2	Yes	14	rs61814940
<i>HSPH1</i>	13	31724120	C	T	E	K	Missense	S2	No	1	NA
<i>KAT6A</i>	8	41791810	G	A	Q	*	Nonsense	S3	No	2	NA
<i>KIAA1257</i>	3	128706628	C	T	W	*	Nonsense	S2	No	0	NA
<i>KLF5</i>	13	73636078	G	T	R	L	Missense	S1	No	0	NA
<i>KRT32</i>	17	39619088	T	C	D	G	Missense	S3	No	1	NA
<i>LILRA5</i>	19	54823629	C	G	G	R	Missense	S1	No	1	NA
<i>MACC1</i>	7	20193850	G	A	P	L	Missense	S3	No	3	NA
<i>MAP3K12</i>	12	53876968	T	G	D	A	Missense	S3	Yes	2	NA
<i>MAP3K4</i>	6	161470088	G	A	E	K	Missense	S3	No	1	NA
<i>MESPI</i>	15	90294304	C	CG	-	-	Frameshift	S2	No	0	NA
<i>MLL3</i>	7	151945225	T	C	E	G	Missense	S2	Yes	23	rs77957001
<i>MROH2B</i>	5	41061715	C	T	W	*	Nonsense	S3	No	0	rs1023840
<i>NACAD</i>	7	45124694	A	C	L	R	Missense	S3	No	1	NA
<i>NEBL</i>	10	21074726	T	C	T	A	Missense	S1	No	2	NA
<i>NPAT</i>	11	108056150	G	C	N	K	Missense	S3	No	0	NA
<i>NVL</i>	1	224455814	C	A	V	L	Missense	S3	No	1	NA
<i>OR13C2</i>	9	107367609	C	A	Q	H	Missense	S1	No	2	NA
<i>OR111</i>	19	15198054	T	A	Y	N	Missense	S2	No	0	NA
<i>OR2D3</i>	11	6942894	T	C	V	A	Missense	S3	No	0	NA
<i>OR2H2</i>	6	29556464	T	A	L	H	Missense	S1	No	3	NA
<i>PARP4</i>	13	25051949	T	C	Y	C	Missense	S1	No	9	NA
<i>PCDHA12</i>	5	140255941	T	C	I	T	Missense	S3	Yes	0	NA

Table III. Continued.

Gene	Chr	Position	Ref	Alter	Ref_aa	Alter_aa	Type	Sample	COSMIC	Num in 88 ^a	dbSNP
<i>PCDHB12</i>	5	140588823	A	T	Q	L	Missense	S1	No	0	rs199513227
<i>PHLDB1</i>	11	118527435	C	T	R	W	Missense	S3	Yes	1	NA
<i>PLA2G4C</i>	19	48601455	G	A	P	L	Missense	S2	No	0	NA
<i>PLEK</i>	2	68613658	T	G	V	G	Missense	S1	No	0	NA
<i>PPP1R13B</i>	14	104224075	C	A	R	L	Missense	S1	No	1	NA
<i>PRAMI</i>	19	8555622	C	A	V	F	Missense	S2	No	0	NA
<i>PRPH</i>	12	49690017	G	A	R	H	Missense	S3	No	1	NA
<i>RARG</i>	12	53606940	C	T	R	Q	Missense	S2	No	0	NA
<i>RNPEP</i>	1	201970804	C	A	F	L	Missense	S3	No	0	NA
<i>SEMA3F</i>	3	50225285	C	T	R	W	Missense	S3	No	0	NA
<i>SEMA7A</i>	15	74707184	C	G	G	R	Missense	S3	No	0	NA
<i>SIGLEC7</i>	19	51645658	G	T	W	L	Missense	S3	No	2	NA
<i>SKIV2L2</i>	5	54640368	A	T	T	S	Missense	S1	No	0	NA
<i>SNCG</i>	10	88719421	C	A	S	R	Missense	S3	No	0	NA
<i>SORL1</i>	11	121429524	C	A	P	H	Missense	S2	Yes	2	NA
<i>TAMM41</i>	3	11885549	T	C	Y	C	Missense	S3	No	0	rs149208464
<i>TBL3</i>	16	2028215	C	T	Q	*	Nonsense	S3	No	0	NA
<i>TNKS2</i>	10	93579726	A	T	N	Y	Missense	S1	No	0	NA
<i>TNN</i>	1	175048774	G	A	G	S	Missense	S2	Yes	1	NA
<i>TP53</i>	17	7577534	C	A	R	S	Missense	S1;S3	Yes	34	rs28934571
<i>TRAPPC8</i>	18	29444630	C	A	R	L	Missense	S1	Yes	0	NA
<i>TTN</i>	2	179402119	A	T	L	H	Missense	S2	Yes	26	NA
	2	179581852	C	CT	-	-	Frameshift	S3			NA
<i>UNC13C</i>	15	54793167	A	C	R	S	Missense	S3	No	4	NA
<i>ZCCHC11</i>	1	52961155	T	G	T	P	Missense	S3	No	0	NA
<i>ZFYVE16</i>	5	79768637	G	A	R	Q	Missense	S3	Yes	0	NA
<i>ZNF605</i>	12	133503218	T	C	K	E	Missense	S1	No	1	NA
<i>ZNF717</i>	3	75786430	C	T	G	R	Missense	S3	No	47	NA

All missense SNVs were predicted to be damaging or deleterious by the 2 methods. Chr, chromosome; ^aNumber of individuals with mutated corresponding genes in the 88 HBV-related HCC patients; SNVs, single nucleotide variations.

HRNR, *CTNNB1*, *MLL3*, *TTN* and *TP53* have previously been reported in the COSMIC database. However, the variants identified in *TTN* were reported for the first time here.

Discussion

We performed exome sequencing of HCCs and corresponding normal tissues from 3 Chinese male BCLC stage A HCC patients with chronic HBV infection to identify key genetic lesions contributing to the initial stage of the disease. Further whole genome sequence analysis in 88 patients at various stages (across early to advanced stage) was also carried out to check the prevalence of the mutated genes.

Eighty protein-altering somatic mutations, including exonic indels, nonsense SNVs and missense SNVs which were predicted to be damaging by 2 function prediction methods were detected and validated in this study. Fifty-nine genes were first reported to mutate in HCC in the present study (Table III).

For the 78 genes with these variants, 47 of them were also mutated at least once in the 88 WGS samples. Among them, 7 genes, *ZNF717* and *PARP4*, *HRNR*, *CTNNB1*, *MLL3*, *TTN* and *TP53* were mutated in at least 9 patients (over 10%, Fig. 1). *HRNR*, *CTNNB1*, *MLL3*, *TTN* and *TP53* were reported to be related to HCC according to the COSMIC database.

TP53 is a famous tumor suppressor gene. The missense SNV identified on *TP53* here was also called R249S, and it is frequently found in 10-61% of HCC cases (17,18). Consistently, we identified the mutation in sample S1 and S3 and further WGS analysis also detected this mutation in 10 samples. In HCC, *TP53* mutations also vary in different geographic areas, presumably reflecting differences in both etiological agents and host susceptibility factors (19). In some areas, such as sub-Saharan Africa and China, aflatoxin B1 exposure and chronic viral hepatitis are responsible for a very high incidence of HCC (with up to 100/100,000 cases/year), where a high proportion of an R249S point mutation was found (20-22). Furthermore, it was shown that adduction of AFB₁ metabolites in the third base of codon 249 in the *TP53*

leading to the R249S mutation (23), indicating that this is an early mutational event in hepatocarcinogenesis. Recently, it was found that *TP53* with mutation R249S can interact with HBx, which may contribute to cell proliferation and survival of HBx-expressing HCC cells (24). Thus, this mutation may play an important role in the cancer development of HBV-induced HCC patients.

The missense SNV of *ZNF717* was detected in sample S3. The mutation was also detected in other WGS samples. Protein-altering variants of *ZNF717* were detected in 47 WGS samples at various stages. This suggested that this gene may be related to various stages of the disease. Protein encoded by this gene belongs to the zinc-finger family, which is known to play key roles in regulating expression of genes important for cell growth, proliferation, differentiation and apoptosis (25,26). However, the function of most zinc-finger proteins in tumor occurrence and development is currently unknown. As a zinc-finger protein containing 8 conserved C2H2-type zinc-finger motifs, *ZNF717* may function in the transcriptional regulation. Further investigations should be carried out to confirm its roles.

PARP4 was also reported for the first time here. The missense SNV is detected in sample S1. Other protein-altering mutations of this gene were also detected in 9 other early-stage WGS samples. The protein encoded by this gene is involved in the base excision repair pathway (hsa03410). A previous study showed that inhibition of *PARP4* interfered with DNA base excision repair but can sensitize cells to apoptotic stimuli, resulting in increased tumor cell apoptosis *in vivo*. Interruption of this gene may prevent tumor formation through this mechanism (27). Thus, mutations of this gene at early-stage of the disease may serve as a self-compensation.

Several other newly discovered genes are also noteworthy. For example, the protein encoded by *FLNA* is an actin-binding protein that crosslinks actin filaments and links actin filaments to membrane glycoprotein. It is involved in several important pathways, such as the MAPK signaling pathway (hsa04010) and proteoglycans in cancer (hsa05205). Although mutation of this gene in HCCs has not previously been reported, comparative proteomics analysis showed that *FLNA* protein was significantly differentially expressed in HCC cell lines and may influence the metastasis of HCC cells (28). Mutation of this gene was also detected in another early-stage HBV-induced HCC patient of the WGS samples, indicating that interruption of this gene at early stage may contribute to the initial formation of the disease. Moreover, *FLNA* protein is also involved in the focal adhesion (hsa04510) pathway. Another newly reported gene here, *CNTN2*, is also related to cell adhesion as it is a member of the cell adhesion molecules pathway. The cell adhesion pathway is important for cancer cell invasion and metastasis. Abnormality of the cell adhesion pathway is considered a characteristic of the advanced stage of cancer. Our findings here thus support the hypothesis that the cancer cells may have acquired the capacity for metastasis at the early stage of development (29). Four genes involved in olfactory transduction including *OR13C2*, *OR111*, *OR2D3* and *OR2H2* were also found for the first time in this study to be mutated in HBV-induced HCCs, suggesting these genes warrant further investigation.

In summary, exome sequencing of HCCs and adjacent normal tissues from 3 BCLC stage A HCC patients was carried out to identify key genetic lesions contributing to the initial stage of the disease. Eighty damaging mutations were validated and 59 genes were first reported to be mutated in HBV-related HCCs here. Further WGS analysis showed that mutations in 33 of the 59 genes were also detected in other samples. Variants of 2 newly found genes, *ZNF717* and *PARP4*, were detected in more than 10% of the WGS samples. Several other genes, such as *FLNA* and *CNTN2*, are also noteworthy. Thus, the exome sequencing analysis of 3 BCLC stage A patients provided new insights into the molecular events governing the early step of HBV-induced HCC tumorigenesis.

References

1. El-Serag HB and Rudolph KL: Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. *Gastroenterology* 132: 2557-2576, 2007.
2. Jemal A, Bray F, Center MM, Ferlay J, Ward E and Forman D: Global cancer statistics. *CA Cancer J Clin* 61: 69-90, 2011.
3. Perz JF, Armstrong GL, Farrington LA, Hutin YJ and Bell BP: The contributions of hepatitis B virus and hepatitis C virus infections to cirrhosis and primary liver cancer worldwide. *J Hepatol* 45: 529-538, 2006.
4. Schütte K, Bornschein J and Malfertheiner P: Hepatocellular carcinoma - epidemiological trends and risk factors. *Dig Dis* 27: 80-92, 2009.
5. Stratton MR: Exploring the genomes of cancer cells: progress and promise. *Science* 331: 1553-1558, 2011.
6. Stratton MR, Campbell PJ and Futreal PA: The cancer genome. *Nature* 458: 719-724, 2009.
7. Meyerson M, Gabriel S and Getz G: Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet* 11: 685-696, 2010.
8. Sung WK, Zheng H, Li S, *et al*: Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nat Genet* 44: 765-769, 2012.
9. Li M, Zhao H, Zhang X, *et al*: Inactivating mutations of the chromatin remodeling gene *ARID2* in hepatocellular carcinoma. *Nat Genet* 43: 828-829, 2011.
10. Tao Y, Ruan J, Yeh SH, *et al*: Rapid growth of a hepatocellular carcinoma and the driving mutations revealed by cell-population genetic analysis of whole-genome data. *Proc Natl Acad Sci USA* 108: 12042-12047, 2011.
11. Totoki Y, Tatsuno K, Yamamoto S, *et al*: High-resolution characterization of a hepatocellular carcinoma genome. *Nat Genet* 43: 464-469, 2011.
12. Li H and Durbin R: Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26: 589-595, 2010.
13. Li H, Handsaker B, Wysoker A, *et al*: The sequence alignment/map format and SAMtools. *Bioinformatics* 25: 2078-2079, 2009.
14. Koboldt DC, Zhang Q, Larson DE, *et al*: VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 22: 568-576, 2012.
15. Choi Y, Sims GE, Murphy S, Miller JR and Chan AP: Predicting the functional effect of amino acid substitutions and indels. *PLoS One* 7: e46688, 2012.
16. Ng PC and Henikoff S: SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res* 31: 3812-3814, 2003.
17. Hussain SP, Schwank J, Staib F, Wang XW and Harris CC: TP53 mutations and hepatocellular carcinoma: insights into the etiology and pathogenesis of liver cancer. *Oncogene* 26: 2166-2176, 2007.
18. Imbeaud S, Ladeiro Y and Zucman-Rossi J: Identification of novel oncogenes and tumor suppressors in hepatocellular carcinoma. *Semin Liver Dis* 30: 75-86, 2010.
19. Staib F, Hussain SP, Hofseth LJ, Wang XW and Harris CC: TP53 and liver carcinogenesis. *Hum Mutat* 21: 201-216, 2003.
20. Kew MC: Synergistic interaction between aflatoxin B1 and hepatitis B virus in hepatocarcinogenesis. *Liver Int* 23: 405-409, 2003.
21. Parkin DM, Bray F, Ferlay J and Pisani P: Global cancer statistics, 2002. *CA Cancer J Clin* 55: 74-108, 2005.

22. Scorsone KA, Zhou YZ, Butel JS and Slagle BL: p53 mutations cluster at codon 249 in hepatitis B virus-positive hepatocellular carcinomas from China. *Cancer Res* 52: 1635-1638, 1992.
23. Bressac B, Kew M, Wands J and Ozturk M: Selective G to T mutations of p53 gene in hepatocellular carcinoma from southern Africa. *Nature* 350: 429-431, 1991.
24. Gouas DA, Shi H, Hautefeuille AH, *et al*: Effects of the *TP53* p.R249S mutant on proliferation and clonogenic properties in human hepatocellular carcinoma cell lines: interaction with hepatitis B virus X protein. *Carcinogenesis* 31: 1475-1482, 2010.
25. Lodomery M and Dellaire G: Multifunctional zinc finger proteins in development and disease. *Ann Hum Genet* 66: 331-342, 2002.
26. Papavassiliou AG: Transcription factors: structure, function, and implication in malignant growth. *Anticancer Res* 15: 891-894, 1995.
27. Hans MA, Müller M, Meyer-Ficca M, Bürkle A and Küpper JH: Overexpression of dominant negative PARP interferes with tumor formation of HeLa cells in nude mice: evidence for increased tumor cell apoptosis in vivo. *Oncogene* 18: 7010-7015, 1999.
28. Ai J, Huang H, Lv X, *et al*: FLNA and PGK1 are two potential markers for progression in hepatocellular carcinoma. *Cell Physiol Biochem* 27: 207-216, 2011.
29. Scheel C, Onder T, Karnoub A and Weinberg RA: Adaptation versus selection: the origins of metastatic behavior. *Cancer Res* 67: 11476-11480, 2007.