Received: 26 February 2020 | Revised: 2 April 2020 | Accepted: 17 April 2020

DOI: 10.1002/ijc.33027

INFECTIOUS CAUSES OF CANCER

IJC International Journal of Cancer

Characterization of hepatitis B virus infection and viral DNA integration in non-Hodgkin lymphoma

Mengge Li[1] | Yuling Shen[1,2] | Yiming Chen[3] | Haifeng Gao[4] | Jiaqin Zhou[2] | Qing Wang[1] | Chunsun Fan[1,5] | Wei Zhang[1] | Jin Li[1] | Hui Cong[1] | Jinyang Gu[6] | Yu Gan[1] | Hong Tu[1]

[1]State Key Laboratory of Oncogenes and Related Genes, Shanghai Cancer Institute, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[2]Department of Head and Neck Surgery, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[3]Department of Pathology, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

[4]Department of Immunology, School of Basic Medical Sciences, Fudan University, Shanghai, China

[5]Department of etiology, Qidong People's Hospital/Qidong Liver Cancer Institute, Qidong, China

[6]Department of Transplantation, Xinhua Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Correspondence
Hong Tu, State Key Laboratory of Oncogenes and Related Genes, Shanghai Cancer Institute, Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China.
Email: tuhong@shsci.org

Funding information
Chinese State Key Laboratory of Oncogenes and Related Genes, Grant/Award Numbers: 91-14-16, 91-15-06; Doctoral Innovation Fund of Shanghai Jiao Tong University School of Medicine, Grant/Award Number: BXJ201922; Fourth Round of Three-year Public Health Program and Key Disciplines, Grant/Award Number: 15GWZK0801; National Key Projects Specialized in Infectious Diseases, Grant/Award Number: 2017ZX10201201-008-003; National Key R&D Program of China, Grant/Award Numbers: 2017YFC0908103, 2017YFC0908104; National Natural Science Foundation of China, Grant/Award Numbers: 81572312, 81872505

Abstract

Hepatitis B virus (HBV) infection has been reported to be associated with non-Hodgkin lymphoma (NHL). However, the evidence is limited to the seroepidemiological study. There is a lack of evidence showing the HBV infection and integration in NHL cells. Here, we reported that in the Shanghai area, the positive rates of serum HBsAg (OR: 3.11; 95% CI: 2.20-4.41) and HBeAg (OR: 3.99; 95% CI: 1.73-9.91) were significantly higher in patients with NHL. HBsAg, HBcAg and HBV DNA were detected in 34.4%, 45.2% and 47.0% of the NHL tissues, respectively. Furthermore, by using a high-throughput viral integration detection approach (HIVID), integrated HBV DNA was identified from 50% (6/12) HBV-related NHL tissues. There were a total of 313 HBV integration sites isolated from the NHL tissues, among which four protein-coding genes (*FAT2*, *SETX*, *ITGA10* and *CD63*) were interrupted by HBV DNA in their exons. Seven HBV preferential target genes (*ANKS1B*, *HDAC4*, *EGFLAM*, *MAN1C1*, *XKR6*, *ZBTB38* and *CCDC91*) showed significantly altered expression levels in NHL, suggesting a potential role of these genes in NHL development. Taken together, HBV integration is a common phenomenon in NHL. This finding opens up a new direction of research into the mechanistic link between HBV infection and NHL.

KEYWORDS

hepatitis B virus, integration, non-Hodgkin lymphoma

Abbreviations: *CCNE1*, Cyclin E1; DLBCL, diffuse large B-cell lymphoma; HBV, hepatitis B virus; HCC, hepatocellular carcinoma; HIVID, high-throughput viral integration detection; IPI, International Prognostic Index; *MLL4*, myeloid/lymphoid or mixed-lineage leukemia 4; NHL, non-Hodgkin lymphoma; *TERT*, telomerase reverse transcriptase.

Mengge Li and Yuling Shen contributed equally to this study.

1 | INTRODUCTION

Non-Hodgkin lymphoma (NHL) accounts for more than 90% of the lymphoma cases and represents 3% of all new cancer cases worldwide.[1] Viruses,[2] including hepatitis B virus (HBV),[3-7] have been suggested to contribute to the development of NHL.

HBV is a hepatotropic DNA virus with lymphotropic property that can infect and replicate in lymphocytes.[8,9] Our recent meta-analysis indicated that the risk of NHL is significantly increased in HBV-infected individuals (summary odds ratio [sOR]: 2.52; 95% confidence interval [CI]: 2.22-2.86) regardless of the study design.[10] Overall, B-cell NHL showed a stronger association with HBV infection than T-cell NHL. It should be noted that the previous studies investigating the association of NHL with HBV mostly relied on seroepidemiology. Whether HBV infects NHL cells remains debated. While some studies have reported that HBV antigens could be detected in NHL tissues at a rate of 10%-30%,[11,12] other study does not find any NHL case that was positive for HBsAg and HBcAg in tumor tissues.[13] The integrated state of HBV DNA in NHL cells has never been revealed before.

The integration of HBV DNA into the host genome has been considered as one of the mechanisms contributing to the development of hepatocellular carcinoma (HCC).[14] HBV integration can result in genomic instability or insertional mutagenesis of diverse cancer-related genes. An earlier study found that approximately 50% woodchuck hepatitis virus (WHV) integrations affect the MYC family of proto-oncogenes in woodchuck model of HCC with chronic infection of WHV.[15] After that, many studies have proved that HBV integration is a common event that can be observed in 80%-90% of HBV-associated HCC cases.[16-19] Genes such as the telomerase reverse transcriptase (*TERT*),[20] Myeloid/lymphoid or mixed-lineage leukemia 4 (*MLL4*)[21] and Cyclin E1 (*CCNE1*)[22] have been repeatedly identified as preferential integration sites of HBV in HCC. According to a pooled analysis, one-third of the genes recurrently targeted by HBV integration are known as cancer-related genes.[17] Interestingly, 32% of the recurrently targeted genes showed abnormal expression in HCC tissues, and 44% of the recurrently targeted genes exhibited copy number variations in ~20% of the HCC cases,[17] suggesting that HBV integration plays a crucial *cis*-activation role in HCC carcinogenesis.

Globally, there are about 248 million individuals chronically infected with HBV.[23] In China, although the overall prevalence of HBV has decreased in the past two decades[24] owing to the universal HBV vaccination, because of its vast population, the absolute number of HBV-infected individuals remains very large.[23] In recent years, the occurrence of NHL has been steadily increasing in China, making NHL a crucial contributor to the whole cancer burden. To investigate the association and the underlying mechanism between HBV infection and NHL, we carried out a comprehensive study involving not only seroepidemiology, but also molecular and genetic analysis to provide direct evidence supporting that HBV infection is an etiology factor of NHL.

What's new?

Increasing sero-epidemiological evidence has suggested a significant association between chronic hepatitis B virus (HBV) infection and non-Hodgkin lymphoma (NHL). However, whether HBV infects NHL cells is still debated. Here, using a serological, histological, and genetic approach, the authors show that HBV integration is a common phenomenon in NHL. Expression analysis of the HBV preferential target genes identified seven novel candidate genes, which may have potential functions in NHL development. Taken together, the results provide direct evidence that HBV infection is an etiological factor of NHL. The information on HBV integration breakpoints provides new clues for mechanistic investigation of HBV-induced NHL.

2 | MATERIALS AND METHODS

2.1 | Patients and samples

A total of 411 cytological or pathological confirmed NHL patients from Shanghai Renji Hospital between April 2014 and December 2018 were included in our study for the measurement of HBV seromarkers. Diagnostic criteria for the subtype of lymphoma were based on the Revised European American Lymphoma (REAL)/World Health Organization (WHO). Blood samples were collected before any treatment for the disease. Serum samples from 957 age- and sex-matched healthy controls were randomly obtained from individuals who visited Shanghai Renji Hospital for routine medical examination between 2014 and 2018. In both the NHL and the healthy control groups, none of the subjects had any other types of primary cancer. Paraffin-embedded of 117 NHL tissues were also collected from Renji Hospital. The study was approved by the institutional ethics review committee of Renji Hospital, Shanghai Jiao-Tong University School of Medicine and conducted according to the principles of the Declaration of Helsinki. Informed consent was obtained from all patients.

2.2 | ELISA and immunohistochemistry

The presence of HBV seromarkers was detected using an AXSYM HBV reagent pack (Abbott Laboratories, Abbott Park, Illinois). The laboratory technician who performed these assays was blind to the subjects' clinical status. Two tissue microarray chips, each containing 93 diffuse large B-cell lymphoma (DLBCL) and 5 lymph node sections were purchased from Super Biotek (Shanghai, China). The detailed experimental procedure was performed as previously described.[25] The monoclonal antibodies against HBsAg were purchased from Zhong Shan (Beijing, China), and the polyclonal antibodies against HBcAg were obtained from Abnova (Taiwan, China).

TABLE 1 Prevalence of HBV infection in non-Hodgkin lymphoma patients and controls

	Control (n = 957)	NHL Total (n = 411)			B-cell NHL (n = 323)			T-cell NHL (n = 50)		
Variables	**n (%)**	**n (%)**	**OR (95% CI)**	***P***	**n (%)**	**OR (95% CI)**	***P***	**n (%)**	**OR (95% CI)**	***P***
Age	58.09 ± 13.33	58.04 ± 14.59		.623	56.90 ± 13.82		.653	58.81 ± 12.21		.575
Sex										
Male	560 (58.51)	233 (56.70)	1		183 (56.66)	1		32 (64.00)	1	
Female	397 (41.49)	178 (43.30)	1.08 (0.85-1.36)	.531	140 (43.34)	1.08 (0.84-1.39)	.558	18 (36.00)	0.79 (0.44-1.43)	.442
HBsAg										
Negative	889 (92.89)	322 (78.35)	1		257 (79.57)	1		43 (86.00)	1	
Positive	68 (7.11)	79 (19.22)	3.11 (2.20-4.41)	.000	66 (20.43)	3.36 (2.33-4.84)	.000	7 (14.00)	2.13 (0.92-4.92)	.070
Anti-HBs										
Negative	497 (51.93)	224 (54.50)	1		174 (53.87)	1		30 (60.00)	1	
Positive	460 (48.07)	187 (45.50)	0.90 (0.72-1.14)	.083	149 (46.13)	0.93 (0.72-1.19)	.547	20 (40.00)	0.72 (4.40-1.29)	.266
HBeAg										
Negative	948 (99.06)	396 (96.35)	1		309 (95.67)	1		47 (94.00)	1	
Positive	9 (0.94)	15 (3.65)	3.99 (1.73-9.91)	.000	14 (4.33)	4.77 (2.05-11.00)	.000	3 (6.00)	6.72 (1.75-25.70)	.011
Anti-HBe										
Negative	735 (76.80)	280 (68.13)	1		219 (67.80)	1		36 (72.00)	1	
Positive	222 (23.20)	131 (31.87)	1.55 (1.20-2.00)	.001	104 (32.20)	1.57 (1.19-2.08)	.001	14 (28.00)	1.04 (0.68-2.43)	.435
Anti-HBc										
Negative	473 (49.43)	164 (39.90)	1		124 (38.39)	1		20 (40.00)	1	
Positive	484 (50.57)	247 (60.10)	1.47 (1.16-1.86)	.000	199 (61.61)	1.57 (1.22-2.03)	.000	30 (60.00)	1.47 (0.83-2.62)	.0194

Abbreviations: CI, confidence interval; Control, health examination individuals; NHL, non-Hodgkin lymphoma; OR, odds ratio.

TABLE 2 Clinical features of HBV-related and nonrelated non-Hodgkin lymphoma

Characteristics	HBV infection cases (n = 79)	HBV non-infection cases (n = 79)	*P*
Sex			1.000
Male	48	48	
Female	31	31	
Age, years			.870
≤60	49	48	
>60	30	31	
LDH			.000
Normal	37	61	
Abnormal	42	18	
Ann Arbor stage			.001
I-II	20	41	
III-IV	59	38	
Extranodal involvement			.750
≤1 site	38	40	
>1 sites	41	39	
ECOG performance status			.029
≤2	9	2	
>2	70	77	
IPI			.023
Low	21	41	
Low-intermediate	23	21	
High-intermediate	25	14	
High	10	3	
Tumor marker			
AFP			.021
Positive	5	0	
Negative	70	77	
CEA			.725
Positive	3	4	
Negative	72	73	
CA199			.045
Positive	10	3	
Negative	64	70	
CA125			.033
Positive	26	14	
Negative	46	56	
Liver function			
ALT			.247
Positive	6	3	
Negative	73	76	
AST			.002
Positive	19	5	
Negative	60	74	
Albumin			.166
Positive	28	20	
Negative	51	59	
TDIL			.005
Positive	19	6	
Negative	60	73	
DBIL			.003
Positive	28	12	
Negative	51	67	

Abbreviations: AFP, alpha fetal protein; ALT, alanine transaminase; AST, glutamic-oxalacetic transaminase; DBIL, direct bilirubin; ECOG performance status, Eastern Cooperative Oncology Group performance status, describes a patient's level of functioning in terms of their ability to care for themself, daily activity and physical ability (walking, working, etc.); HBV infection cases, positive for $HBsAg^+$ in NHL patients; HBV noninfection cases, negative for all HBV serologic markers in NHL patients; IPI, The International Prognostic Index; LDH, lactate dehydrogenase; TBIL, total bilirubin.

2.3 | DNA extraction, PCR amplification and DNA sequencing

Genomic DNA was extracted by a standard protocol using GeneRead FFPE DNA Kit (QIAGEN, Hilden, Germany). The extracted tissue DNA was amplified for detecting the presence of HBV S, C and P by semi-nested PCR. The primer sequences that were used for PCR are listed in Table S1. The amplification was carried out as previously described.[25] The PCR products were examined by gel electrophoresis and were subsequently sent for automated sequencing (Biosune, Shanghai, China).

2.4 | HBV capture sequencing and bioinformatics analysis

The high-throughput viral integration detection (HIVID), a novel experimental method based on NGS technique, has been proved as an efficient method for identification of HBV integration in HCC.[16,22] The detailed experimental procedure has been described in previous literature.[22] HBV capture sequencing of NHL tissues was performed by MyGenostics (Beijing, China) and BGI (Shenzhen, China). Genes near the breakpoints in the human genome were annotated using the University of California Santa Cruz (UCSC) Genome Browser (GRCh37/hg19). The gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses of HBV targeted genes were performed using the web-accessible functional annotation tool from the Database for Annotation, Visualization and Integrated Discovery (DAVID) version 6.7 (http:// david.abcc.ncifcrf.gov). The mRNA expression data as well as the clinicopathological annotations of NHL and normal lymph tissues were available from The

Cancer Genome Atlas (TCGA, http://cancergenome.nih.gov)[26] and GTEx (https://www.gtexportal.org/home/)[27] database, respectively.

2.5 | Statistical analysis

The chi-square test or Fisher's exact test was used to analyze the categorical variables that were expressed as proportions. Student's *t*-test was used to compare continuous variables. A two-tailed *P*-value of less than .05 was considered as statistically significant. Binomial test was used to analyze the HBV integration frequency between observed and expected groups. The statistical analysis was performed using SPSS software version 22.0 (SPSS Inc., Chicago, Illinois) or GraphPad Prism (GraphPad Software, La Jolla, California).

3 | RESULTS

3.1 | Association of HBV infection and NHL

As the association between HBV infection and NHL varies among areas,[10] we first carried out an epidemiology study in Shanghai with 411 NHL patients and 957 age- and sex-matched healthy controls (Tables 1 and S2). The prevalence of HBsAg (OR: 3.11; 95% CI: 2.20-4.41) and HBeAg (OR: 3.99; 95% CI: 1.73-9.91) was significantly higher in patients with NHL than that in the healthy controls. We then stratified NHL into B-cell NHL (n = 323) and T-cell NHL (n = 50) subgroups and found that the prevalence of HBsAg was significantly increased in B-cell NHL (OR: 3.36; 95% CI: 2.33-4.84), but not in T-cell NHL (OR: 2.13; 95% CI: 0.92-4.92). Similarly, anti-HBc, a seromarker for past exposure to HBV, showed a significant association with the B-cell NHL but not the T-cell NHL.

Fifteen clinical parameters including the International Prognostic Index (IPI)[28] were compared between 79 HBsAg positive NHL cases and 79 HBV negative NHL cases which were randomly selected from the above cohort (Table 2). Results showed that patients who were positive for HBV infection had a significantly higher level of serum LDH, a more advanced stage of NHL, a worse ECOG performance status and a less favorable prognosis. Besides, there were significant differences in positive rates of tumor markers AFP, CA199 and CA125 between HBV infected and noninfected patients with NHL. HBV-related NHL also showed a worse liver function as indicated by higher levels of AST, total bilirubin and direct bilirubin.

3.2 | Detection of HBV antigens and DNA in NHL tissues

To search for the direct evidence of HBV infection in NHL, we performed immunohistochemical staining of HBsAg and HBcAg on a pair of tissue microarray sections each containing 93 DLBCL and 5 lymph node tissues. As illustrated in Figure 1, HBsAg was generally located in

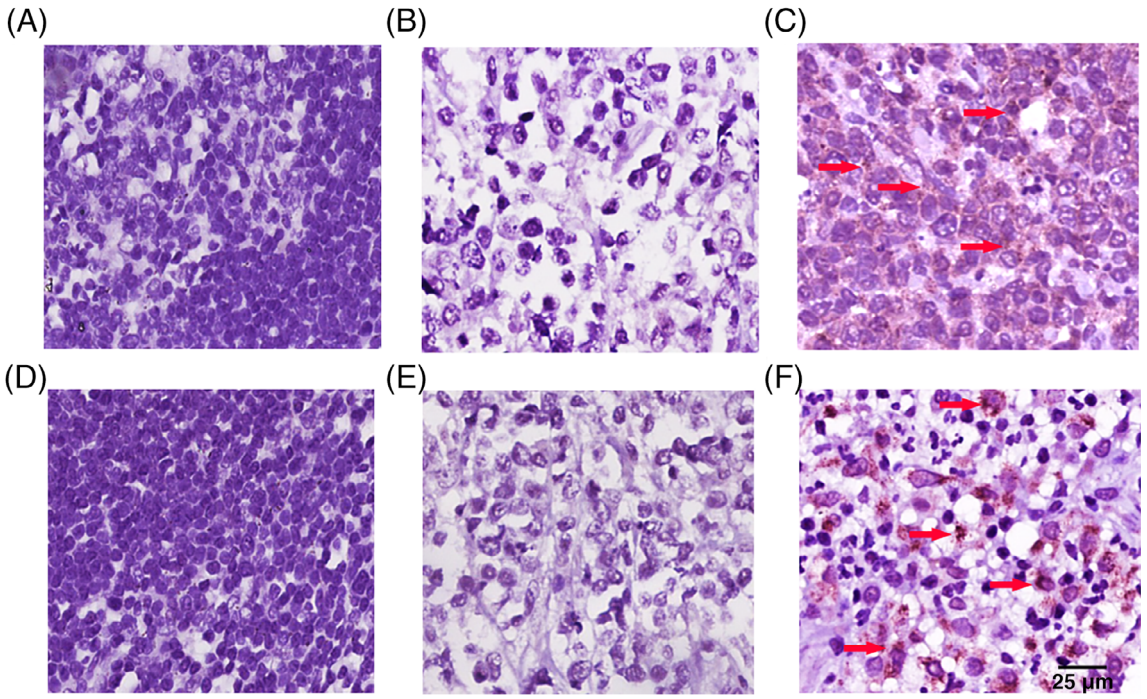

FIGURE 1 Immunohistochemical analysis of the HBV-specific antigens in non-Hodgkin lymphoma and lymph node. Immunohistochemical staining of HBsAg in lymph node and non-Hodgkin lymphoma (NHL, A-C). A, Negative immunoreactivity of HBsAg in lymph node. B, Negative immunoreactivity of HBsAg in NHL. C, Positive immunoreactivity of HBsAg in NHL. Immunohistochemical staining of HBcAg in lymph node and NHL (D-F). D, Negative immunoreactivity of HBcAg in lymph node. E, Negative immunoreactivity of HBcAg in NHL. F, Positive immunoreactivity of HBcAg in NHL (magnification, ×40)

TABLE 3 Recurrent HBV integration sites in non-Hodgkin lymphoma

Sample ID	Chromosome location	Insertion site (distance, bp)	Gene	Gene type	Gene function
FFPE-2629	chr12:99801186	Intron 11	*ANKS1B*	Protein coding	Brain development/Involved in Alzheimer's disease
FFPE-2629	chr12:100348169	Intron 1	*ANKS1B*		
FFPE-2629	chr12:99184682	Intron 21	*ANKS1B*		
A-001	chr1:19756532	Intron 1	*CAPZB*	Protein coding	Cell morphology maintenance
A-001	chr1:19756535	Intron 1	*CAPZB*		
FFPE-2629	chr10:68465134	Intron 9	*CTNNA3*	Protein coding	Formation of stretch-resistant cell-cell adhesion complexes
FFPE-2629	chr10:69146347	Intron 5	*CTNNA3*		
FFPE-2629	chr5:38441060	Intron 17	*EGFLAM*	Protein coding	Cell adhesion
5152	chr5:37985429	Intergenic (273082)	*EGFLAM*		
FFPE-2629	chr18:33881067	Intron 1	*FHOD3*	Protein coding	Stress fiber formation together with cell elongation
FFPE-2629	chr18:34004350	Intron 3	*FHOD3*		
FFPE-2629	chr2:240043010	Intron 12	*HDAC4*	Protein coding	Transcriptional regulation and cell cycle progression
FFPE-2629	chr2:239945356	Intergenic (24508)	*HDAC4*		
FFPE-2629	chr11:132551095	Intron 2	*OPCML*	Protein coding	Cell contraction
lymphomaDNA-6	chr11:133165274	Intron 1	*OPCML*		
FFPE-2629	chr15:24826304	Intron 8	*PWRN1*	ncRNA	Establishing paternal imprinting in the Prader-Willi syndrome region
lymphomaDNA-6	chr15:24793840	Intergenic (9464)	*PWRN1*		
FFPE-2629	chr4:139287839	Intron 2	*LINC00499*	ncRNA	—
FFPE-2629	chr4:139298000	Intron 2	*LINC00499*		
FFPE-2629	chr5:39611373	Intergenic (441020)	*LINC00603*	ncRNA	—
C-025	chr5:40372297	Intergenic (318871)	*LINC00603*		
FFPE-18223	chr1:73793409	Intron 3	*LINC01360*	ncRNA	Involved in major depressive disorder
FFPE-2629	chr1:73576805	Intergenic (195048)	*LINC01360*		
FFPE-2629	chr11:97892788	Intergenic (401893)	*MIR7976*	ncRNA	—
FFPE-2629	chr11:97952042	Intergenic (461147)	*MIR7976*		
C-025	chr2: 33141554	Intron 4	*LINC00486*	ncRNA	—
C-025	chr2: 33141629	Intron 4	*LINC00486*		

Note: —, no reports related to this gene so far.

cell membrane and cytoplasm, whereas HBcAg was mainly detected in the cytoplasm and nucleus. HBsAg and HBcAg were positive in 34.4% (32/93) and 45.2% (42/93) of NHL tissues, respectively. Co-existence of HBsAg and HBcAg was observed in 33.3% (31/93) of NHL tissues.

We then examined the presence of HBV S, C or P gene in 117 paraffin-embedded HBV-associated NHL tissues. Of the 117 NHL tissue samples, 55 (47.0%) had HBV DNA sequence in tissue specimen (Tables S3 and S4). Seven and 14 NHLs, respectively, displayed positive PCR results only for one or two HBV genes, suggesting the existence of interrupted HBV DNA sequences in NHL tissues.

3.3 | Identification of HBV DNA integration in NHL

Of the 34 NHL cases which showed co-existence of HBV S, C and P genes, 12 had genomic DNA ≥3 μg that enabled us to perform HIVID analysis (Figure S1). The basic characteristics of these 12 samples are summarized in Table S5. Totally, 313 integration breakpoints were identified from 6/12 NHL specimens (support reads ≥2, Tables S6 and S7). A significant heterogeneity of HBV integration number was found in these six samples. Case FFPE-2629 harbored the largest number of HBV integration breakpoints at 262. Each of the rest five samples had the number of integration breakpoints between 2 and 16 (Table S6). Around half of the HBV integration occurred in the intergenic regions (155/313, 49.5%), and the remaining took place in the introns (140/313, 44.7%), 3′-UTR (5/313, 1.6%), gene upstream region (4/313, 1.3%) and gene downstream region (4/313, 1.3%). Four protein-coding genes, namely, *FAT2*, *SETX*, *ITGA10* and *CD63*, and one noncoding RNA namely *LOC79160* were interrupted by HBV DNA in their exons.

HBV integration sites, designated as HBV targeting the same gene or inserting into the vicinal intergenic sequences in different tissue samples or different cell populations of the same tissue, are summarized in Table 3. As integrated HBV is regarded as a strong *cis*-activator of flanking genes which can influence the expression of their target genes

TABLE 4 Pooled analysis of HBV cotargeted genes in non-Hodgkin lymphoma and hepatocellular carcinoma

Gene	Chromosome location	NHL		HCC		Gene type	Gene function
		Total integration (n)	Insertion site (n)	Total integration (n)	Insertion site (n)		
AKAP13	15q	1	Intron 9 (1)	3	Intron 2 (2), intron 8 (1)	Protein coding	Signal transduction
ANKRD11	16q	1	Intron 5 (1)	2	Intron 1 (1), intron 3 (1)	Protein coding	Chromatin regulation
BTBD11	12q	1	Intron 1 (1)	1	Intron 1 (1)	Protein coding	Regulation of neurite outgrowth and cell adhesion
CCDC91	12p	1	Intron 13 (1)	2	Intron 9 (1), intron 4 (1)	Protein coding	Membrane traffic regulation
CDH4	20q	1	Intron 2 (1)	2	Intron 1 (1), intron 2 (1)	Protein coding	Cell adhesion
DOCK3	3p	1	Intron 9 (1)	1	Intron 1 (1)	Protein coding	Cell adhesion
GPC5	13q	1	Intron 7 (1)	2	Intron 5 (1), intron 6 (1)	Protein coding	Control of cellular response to growth factors and adhesion factors
GSG1L	16p	1	Intron 2 (1)	1	Intron 1 (1)	Protein coding	AMPA receptor gating modification
KCNN3	1q	1	Intron 2 (1)	1	Intron 3 (1)	Protein coding	Voltage-independent potassium channel regulation
MAN1C1	1p	1	Intron 2 (1)	1	Intron 6 (1)	Protein coding	Maturation of Asn-linked oligosaccharides
MED13L	12q	1	Utr3 (1)	2	Exon 4 (2)	Protein coding	Regulation the transcription of RNA polymerase II-dependent genes
NTM	11q	1	Intron 3 (1)	2	Intron 1 (2)	Protein coding	Neural cell adhesion
PTPRD	9p	1	Intron 10 (1)	6	Intron 9 (2), intron 3 (2), intron 2 (1)	Protein coding	Presynaptic differentiation
SNTG2	2p	1	Intron 14 (1)	2	Intron 4 (1), intron 15 (1)	Protein coding	Organization of subcellular localization with varies proteins
SNX27	1q	1	Intron 1 (1)	1	Intron 7 (1)	Protein coding	Transportation from endosome to plasma membrane
SOX5	12p	1	Intron 4 (1)	2	Intron 11 (2)	Protein coding	Binds specifically to the DNA sequence 5′-AACAAT-3′
SRGAP1	12q	1	Intron 1 (1)	2	Intron 1 (2)	Protein coding	GTPase-activating protein
XKR6	8p	1	Intron 1 (1)	1	Intron 1 (1)	Protein coding	Involved in the autoimmune diseases
ZBTB38	3q	1	Intergenic (1)	1	Intron 3 (1)	Protein coding	Transcriptional regulation
ZNF605	12q	1	Intergenic (1)	2	Intron 1 (2)	Protein coding	Transcriptional regulation
LINC00486	2p	2	Intron 4 (2)	10	Intron 4 (4), intron 2 (2), intron 10 (3)	ncRNA	—
LINC00971	3p	1	Intron 19 (1)	2	Intron 10 (1)	ncRNA	—
MIR5009	15q	1	Intron 2 (1)	1	Intron 3 (1)	ncRNA	—

Abbreviations: —, no reports related to this gene so far; HCC, hepatocellular carcinoma; NHL, non-Hodgkin Lymphoma.

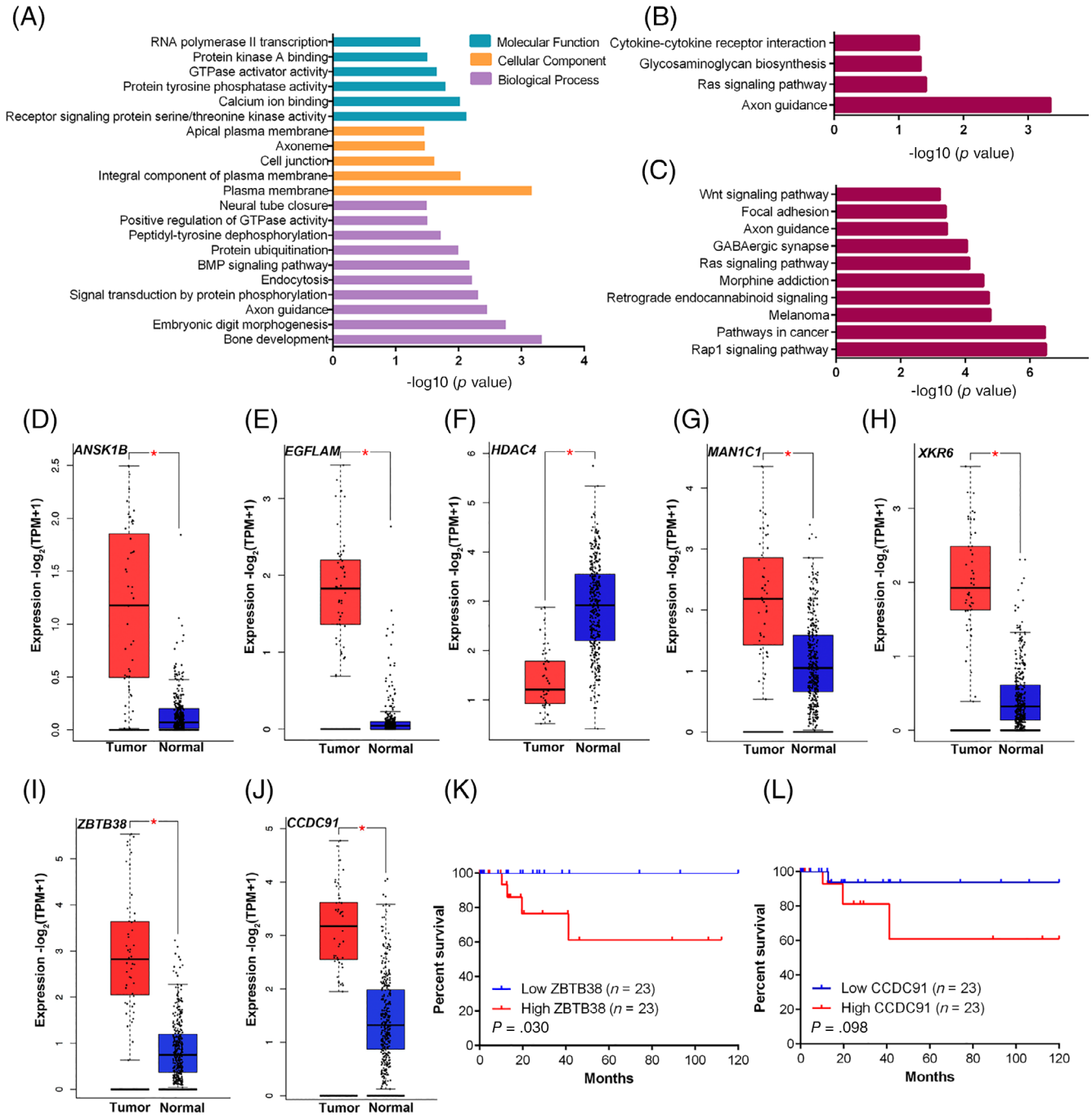

FIGURE 2 Annotation and potential function of HBV integrated genes. A, Gene ontology annotation analysis (cellular component, biological process and molecular function) of 406 integration genes in non-Hodgkin lymphoma (NHL, $P < .05$ and count >5). B, KEGG pathway enrichment analysis of HBV integration genes in NHL ($P < .05$). C, Top-ranked terms of KEGG pathway enrichment analysis of 3852 HBV integration genes in hepatocellular carcinoma (HCC, $P < .05$). D-J, Expression analysis of *ANKS1B* (D), *EGFLAM* (E), *HDAC4* (F), *MAN1C1* (G), *XKR6* (H), *ZBTB38* (I) and *CCDC91* (J) in 47 NHL and 337 normal tissues from TCGA and GTEx database, respectively (*$P < .05$). K-L, Kaplan-Meier analysis of overall survival based on the expression of *ZBTB38* (K) and *CCDC91* (L) in 46 patients with NHL from TCGA database

over a long distance,[29] breakpoints located <500 kb from annotated genes were included in our analysis. Seven protein-coding genes namely *ANKS1B*, *CAPZB*, *CTNNA3*, *EGFLAM*, *FHOD3*, *HDAC4* and *OPCML* were found to be repeatedly targeted by HBV DNA. Additionally, six noncoding RNA genes, that is, *LINC00499*, *LINC00603*, *LINC01360*, *LINC00486*, *MIR7976* and *PWRN1* were also recurrently targeted by HBV DNA. Interestingly, for *CAPZB*, *LINC00499* and *LINC00486* genes, the HBV DNA was consistently inserted into the same intron of the respective gene. It should be noted that in the intergenic region, HBV had very few common integration sites.

3.4 | Comparison of HBV integration pattern between NHL and HCC

In order to compare the HBV integration pattern between NHL and HCC, we pooled the viral-host junctions of HCC from 29 studies to form an ensemble that contained a total of 8037 integration sites that have been reported thus far (Table S8). In HCC, HBV integration was most frequently located in chromosomes 5, 8, 16, 17, 18 and 19 ($P < .05$, Figure S2A,B), whereas in NHL, HBV integration was most frequently located in chromosomes 1, 5 and 12 ($P < .05$, Figure S2C, D). Chromosomes 7 and 15 were found to have significantly lower integration rate than the expected value in both HCC and NHL. In the HBV genome, 40% of the integration breakpoints occurred within the nt. 1400-1900, which include 3′-end of the X gene ($P < .05$, Figure S2E,F) in HCC. However, in NHL, the breakpoints were more prone to be located in 3′-part of the S gene, especially at the sites located within the nt. 273-465 ($P < .05$, Figure S2G,H).

There were 23 genes targeted by HBV in both HCC and NHL (Table 4). Interestingly, certain introns such as intron 4 of *LINC00486*, intron 2 of *CDH4*, intron 1 of *XKR6*, intron 1 of *BTBD11* and intron 1 of *SRGAP1* were repeatedly targeted by HBV integration in both HCC and NHL. Notably, intron 4 of *LINC00486* was repeatedly targeted by HBV DNA for two times in NHL and 10 times in HCC, suggesting a common insertional mutagenesis mechanism occurring in both HCC and NHL.

3.5 | Annotation and expression analysis of HBV-targeted genes in NHL

GO analysis of the 406 HBV-targeted genes in NHL revealed that terms related to developmental process and cell differentiation, signal transduction, cell junction and transcriptional regulation were significantly enriched ($P < .05$ and counts >5, Figure 2A). In the KEGG pathway analysis, axon guidance was the top-ranked pathway ($P < .0001$, Figure 2B), followed by Ras signaling, glycosaminoglycan biosynthesis and cytokine-cytokine receptor interaction ($P < .05$, Figure 2B). Notably, KEGG analysis of the HBV-targeted genes in HCC also revealed axon guidance and Ras signaling as two of the top 10 most significantly enriched pathways ($P < .05$, Figure 2C), implying that some pathways are commonly affected by HBV integration in these two malignancies.

To understand the biological impacts of HBV preferential target genes in NHL development, we analyzed the mRNA expression of these genes using the data from TCGA and GTEx. Six genes namely *ANKS1B*, *EGFLAM*, *ZBTB38*, *CCDC91*, *XKR6* and *MAN1C1*, and one gene namely, *HDAC4* were found to be significantly upregulated and downregulated, respectively, in NHL tissues (n = 47) when compared to the normal lymph tissues (n = 337, $P < .05$, Figure 2D-J). Of note, NHL patients with high *ZBTB38* expression had poorer prognosis compared to those with low *ZBTB38* expression ($P < .05$, Figure 2K). Additionally, the patients with high expression of *CCDC91* had a shorter overall survival with borderline significance ($P = .098$, Figure 2L). These results suggested that HBV target genes might have oncogenic functions in NHL.

4 | DISCUSSION

Increasing evidence has suggested a significant association between chronic HBV infection and NHL. However, considerable heterogeneity exists across studies. While HBV infection was significantly associated with NHL in Asia and Europe, no such association was found in Africa, Oceania and America.[10] Therefore, in our study, we first carried out a large-scale epidemiology study in Shanghai, China. Consistent with a previous report,[12] our study demonstrated significantly higher positive rate of HBsAg in patients with B-cell NHL when compared to that of the controls in the same area. The strength of the association between HBeAg and NHL was even more robust. These results confirmed that in Shanghai, HBV infection, particularly the active chronic infection, is associated with NHL.

The evidence for the infection of HBV in lymphoma tissue is scarce. Hogfeldtt et al had reported the presence of HBV genome in lymphoma tissues, but the positive rate was not mentioned in their paper.[30] Recently, Wang et al examined the HBV DNA sequences and HBx protein in 10 HBsAg-positive patients with DLBCL.[12] Our results clearly demonstrated that HBV-related antigens and HBV DNA exist in NHL tissues. Since PCR is a very sensitive method for DNA detection, in some circumstances, HBV gene fragments may be amplified with the templates derived from circulating viruses rather than HBV DNA inside the NHL cells. Considering this, in addition to HBV DNA, we also tested the viral antigens in NHL by immunohistochemistry. The positive rates of HBsAg (34.4%) and HBcAg (45.2%) in NHL tissues supported the reliability of our PCR results. To the best of our knowledge, our study is the largest of its kind to search for the evidence of HBV infection in NHL. Moreover, it was revealed for the first time that HBV DNA can integrate into the genome of NHL cell. The identification of integrated HBV sequences largely eliminates the miss amplification of viral sequences originated from blood contamination, thus provides convincing evidence to support the infection of HBV in NHL.

Similar to the situation in HCC, HBV integration also has its preferential targets in NHL. For instance, *ANKS1B* was repeatedly interrupted by HBV DNA for three times. *CAPZB*, *LINC00499* and *LINC00486* were recurrently interrupted by HBV DNA not only in the same gene, but also in the same intron of the respective genes. Genes commonly targeted by HBV in HCC usually play a role in liver carcinogenesis.[17] Interestingly, with the TCGA and GTEx analysis, the seven genes which were preferentially targeted by HBV DNA demonstrated differential expression between NHL and normal lymph tissues. It is noted that among these seven genes, six showed an increased expression in NHL, suggesting that HBV integration causes *cis*-activation of protumor genes rather than inactivation of tumor suppressor genes. Regarding the oncogenic function of these genes, only *HDAC4* had been reported before to be involved in lymphoma and leukemia.[31,32] Our results thus provide novel candidate genes for research of their function in NHL, particularly those genes like *ANKS1B*,[33-35] *ZBTB38*[36-39] and *CCDC91*[40] that have been reported to play an oncogenic role in other cancers. Given their increased expressions in NHL, the pathological functions of these three genes in NHL are worth

further investigation. *TERT*, *MLL4* and *CCNE1* are frequently reported to be integrated by HBV DNA in HCC. However, in our study, we did not find these genes were interrupted by HBV DNA in NHL cells. It is widely accepted that the gene recurrently targeted by viral integration usually plays a role in carcinogenesis. TCGA analysis showed that *TERT*, *MLL4* and *CCNE1* have aberrant expressions in HCC, and their biological functions in hepatocarcinogenesis have been extensively studied.[21,41,42] On the contrary, except for *TERT*, the impacts of *MLL4* and *CCNE1* on NHL have never been documented. It is possible that the integration pattern of HBV DNA in NHL is different from that in HCC due to the heterogeneity of the cancer evolution in distinct cancer type. Another possibility is that the sample size of NHL used for target sequencing in our study (n = 12) is very limited compared to that of HCC tested before (n > 1500). So, the real conclusion whether NHL shares the same "hot" HBV integration genes with HCC awaits future large-scale investigation.

The profile of HBV sequence adjacent to viral-host junction displays a different pattern in NHL and HCC. In HCC, the HBV integration mostly occurs at the 3′-end of the X gene, while in NHL, the integrated HBV sequences are significantly enriched in S gene, particularly the region between nt. 273 and nt. 465. The integrated DNA of HBV in HCC usually contains the viral basal core promoter (BCP)/enhancer II (Enh II), and the sequence encoding C-terminal truncated HBx which is a pleiotropic transactivator.[43] Interestingly, C-terminal truncated middle S protein (MSt) also displays transcriptional activity.[44] In most cases, the integrated preS/S sequence isolated from NHL encodes a MSt protein with 123-187 aa truncation at its C-terminal. This MSt is supposed to possess the trans-activation function. It is speculated that sustained expression of MSt trans-activator may render the growth advantage to host cells and facilitate the clonal expansion during the tumor initiation and progression stages.

There are several limitations to our study. First, due to the challenge of DNA quantity and quality, we only had 12 HBV-associated NHL samples that could be used for HIVID in our study. Therefore, the real pattern of HBV integration in NHL awaits further larger-scale investigations. Second, from the current study, it is uncertain whether HBV integrations confer the host cell a growth advantage over its neighbors. Previously, integrated HBV DNA has been found in the lymphatic tissue of one HBV carrier, suggesting that HBV integration can be an early event occurring soon after HBV infection in lymphocytes.[45] Therefore, comparison of the HBV integration pattern between NHL and nontumorous lymphatic tissues will help to address whether HBV integration causes clonal expansion in NHL. Thirdly, in the current study, we only used the public cancer database to analyze the changes in expression of genes recurrently targeted by HBV in NHL. In future, we will confirm these results by using tumor samples from Chinese patients, and conduct in vitro and in vivo function studies to evaluate the contribution of HBV integration in NHL development.

Collectively, our work provides serological, histological and molecular evidence to support that HBV infection is closely associated with NHL. The information on HBV integration breakpoints provides new clues for mechanistic investigation on HBV-induced NHL.

ACKNOWLEDGEMENTS

Our study was funded by grants from the National Key Projects Specialized in Infectious Diseases (2017ZX10201201-008-003), the National Key R&D Program of China (2017YFC0908103, 2017 YFC0908104), the National Natural Science Foundation of China (81872505, 81572312), Chinese State Key Laboratory of Oncogenes and Related Genes (91-14-16, 91-15-06), Doctoral Innovation Fund of Shanghai Jiao Tong University School of Medicine (BXJ201922) and the Fourth Round of Three-year Public Health Program and Key Disciplines (15GWZK0801).

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

DATA ACCESSIBILITY

HBV capture sequencing data were uploaded to the NCBI database with an accession number as PRJNA603601. Other data supporting the findings in our study are available from the authors upon reasonable request.

ORCID

Hong Tu https://orcid.org/0000-0002-8298-5207

REFERENCES

1. Ferlay J, Colombet M, Soerjomataram I, et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int J Cancer*. 2019;144:1941-1953.
2. Castillo JJ, Reagan JL, Bishop KD, Apor E. Viral lymphomagenesis: from pathophysiology to the rationale for novel therapies. *Br J Haematol*. 2014;165:300-315.
3. Fwu CW, Chien YC, You SL, et al. Hepatitis B virus infection and risk of intrahepatic cholangiocarcinoma and non-Hodgkin lymphoma: a cohort study of parous women in Taiwan. *Hepatology*. 2011;53:1217-1225.
4. Huang CE, Yang YH, Chen YY, et al. The impact of hepatitis B virus infection and vaccination on the development of non-Hodgkin lymphoma. *J Viral Hepat*. 2017;24:885-894.
5. Su TH, Liu CJ, Tseng TC, et al. Chronic hepatitis B is associated with an increased risk of B-cell non-Hodgkin's lymphoma and multiple myeloma. *Aliment Pharmacol Ther*. 2019;49:589-598.
6. Bassig BA, Willhauck-Fleckenstein M, Shu XO, et al. Serologic markers of viral infection and risk of non-Hodgkin lymphoma: a pooled study of three prospective cohorts in China and Singapore. *Int J Cancer*. 2018;143:570-579.
7. Zhou X, Wuchter P, Egerer G, et al. Role of virological serum markers in patients with both hepatitis B virus infection and diffuse large B-cell lymphoma. *Eur J Haematol*. 2019;103:410-416.
8. Gu JR, Chen YC, Jiang HQ, et al. State of hepatitis B virus DNA in leucocytes of hepatitis B patients. *J Med Virol*. 1985;17:73-81.
9. Trippler M, Meyer zum Buschenfelde KH, Gerken G. HBV viral load within subpopulations of peripheral blood mononuclear cells in HBV infection using limiting dilution PCR. *J Virol Methods*. 1999;78:129-147.
10. Li M, Gan Y, Fan C, et al. Hepatitis B virus and risk of non-Hodgkin lymphoma: an updated meta-analysis of 58 studies. *J Viral Hepat*. 2018;25:894-903.
11. Wang F, Yuan S, Teng KY, et al. High hepatitis B virus infection in B-cell lymphoma tissue and its potential clinical relevance. *Eur J Cancer Prev*. 2012;21:261-267.

12. Wang Y, Wang H, Pan S, et al. Capable infection of hepatitis B virus in diffuse large B-cell lymphoma. *J Cancer*. 2018;9:1575-1581.
13. Park SC, Jeong SH, Kim J, et al. High prevalence of hepatitis B virus infection in patients with B-cell non-Hodgkin's lymphoma in Korea. *J Med Virol*. 2008;80:960-966.
14. Feitelson MA, Lee J. Hepatitis B virus integration, fragile sites, and hepatocarcinogenesis. *Cancer Lett*. 2007;252:157-170.
15. Hsu T, Moroy T, Etiemble J, et al. Activation of c-myc by woodchuck hepatitis virus insertion in hepatocellular carcinoma. *Cell*. 1988;55: 627-635.
16. Zhao LH, Liu X, Yan HX, et al. Genomic and oncogenic preference of HBV integration in hepatocellular carcinoma. *Nat Commun*. 2016;7: 12992.
17. Li X, Zhang J, Yang Z, et al. The function of targeted host genes determines the oncogenicity of HBV integration in hepatocellular carcinoma. *J Hepatol*. 2014;60:975-984.
18. Murakami Y, Saigo K, Takashima H, et al. Large scaled analysis of hepatitis B virus (HBV) DNA integration in HBV related hepatocellular carcinomas. *Gut*. 2005;54:1162-1168.
19. Duan M, Hao J, Cui S, et al. Diverse modes of clonal evolution in HBV-related hepatocellular carcinoma revealed by single-cell genome sequencing. *Cell Res*. 2018;28:359-373.
20. Paterlini-Brechot P, Saigo K, Murakami Y, et al. Hepatitis B virus-related insertional mutagenesis occurs frequently in human liver cancers and recurrently targets human telomerase gene. *Oncogene*. 2003;22:3911-3916.
21. Saigo K, Yoshida K, Ikeda R, et al. Integration of hepatitis B virus DNA into the myeloid/lymphoid or mixed-lineage leukemia (MLL4) gene and rearrangements of MLL4 in human hepatocellular carcinoma. *Hum Mutat*. 2008;29:703-708.
22. Li W, Zeng X, Lee NP, et al. HIVID: an efficient method to detect HBV integration using low coverage sequencing. *Genomics*. 2013; 102:338-344.
23. Schweitzer A, Horn J, Mikolajczyk RT, Krause G, Ott JJ. Estimations of worldwide prevalence of chronic hepatitis B virus infection: a systematic review of data published between 1965 and 2013. *Lancet*. 2015;386:1546-1555.
24. Liu J, Zhang S, Wang Q, et al. Seroepidemiology of hepatitis B virus infection in 2 million men aged 21-49 years in rural China: a population-based, cross-sectional study. *Lancet Infect Dis*. 2016;16: 80-86.
25. Jin Y, Gao H, Chen H, et al. Identification and impact of hepatitis B virus DNA and antigens in pancreatic cancer tissues and adjacent non-cancerous tissues. *Cancer Lett*. 2013;335:447-454.
26. Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet*. 2013;45:1113-1120.
27. Consortium GT. The genotype-tissue expression (GTEx) project. *Nat Genet*. 2013;45:580-585.
28. International Non-Hodgkin's Lymphoma Prognostic Factors Program. A predictive model for aggressive non-Hodgkin's lymphoma. *N Engl J Med*. 1993;329:987-994.
29. Hu Z, Zhu D, Wang W, et al. Genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism. *Nat Genet*. 2015;47:158-163.
30. Hogfeldt T, Jaing C, Loughlin KM, et al. Differential expression of viral agents in lymphoma tissues of patients with ABC diffuse large B-cell lymphoma from high and low endemic infectious disease regions. *Oncol Lett*. 2016;12:2782-2788.
31. Lee SH, Yoo C, Im S, Jung JH, Choi HJ, Yoo J. Expression of histone deacetylases in diffuse large B-cell lymphoma and its clinical significance. *Int J Med Sci*. 2014;11:994-1000.
32. Sandhu SK, Volinia S, Costinean S, et al. miR-155 targets histone deacetylase 4 (HDAC4) and impairs transcriptional activity of B-cell lymphoma 6 (BCL6) in the emu-miR-155 transgenic mouse model. *Proc Natl Acad Sci USA*. 2012;109:20047-20052.
33. Eckel-Passow JE, Serie DJ, Bot BM, Joseph RW, Cheville JC, Parker AS. ANKS1B is a smoking-related molecular alteration in clear cell renal cell carcinoma. *BMC Urol*. 2014;14:14.
34. Fu X, McGrath S, Pasillas M, Nakazawa S, Kamps MP. EB-1, a tyrosine kinase signal transduction gene, is transcriptionally activated in the t(1;19) subset of pre-B ALL, which express oncoprotein E2a-Pbx1. *Oncogene*. 1999;18:4920-4929.
35. Lin J, Lu C, Stewart DJ, et al. Systematic evaluation of apoptotic pathway gene polymorphisms and lung cancer risk. *Carcinogenesis*. 2012; 33:1699-1706.
36. Marchal C, de Dieuleveult M, Saint-Ruf C, et al. Depletion of ZBTB38 potentiates the effects of DNA demethylating agents in cancer cells via CDKN1C mRNA up-regulation. *Oncogenesis*. 2018;7:82.
37. Oikawa Y, Omori R, Nishii T, Ishida Y, Kawaichi M, Matsuda E. The methyl-CpG-binding protein CIBZ suppresses myogenic differentiation by directly inhibiting myogenin expression. *Cell Res*. 2011;21:1578-1590.
38. Miotto B, Chibi M, Xie P, et al. The RBBP6/ZBTB38/MCM10 axis regulates DNA replication and common fragile site stability. *Cell Rep*. 2014;7:575-587.
39. Jing J, Liu J, Wang Y, et al. The role of ZBTB38 in promoting migration and invasive growth of bladder cancer cells. *Oncol Rep*. 2019;41: 1980-1990.
40. Zeng C, Guo X, Long J, et al. Identification of independent association signals and putative functional variants for breast cancer risk through fine-scale mapping of the 12p11 locus. *Breast Cancer Res*. 2016;18:64.
41. Ma Z-X, Yang C-M, Li M-G, Hong T. Telomerase reverse transcriptase promoter mutations in hepatocellular carcinogenesis. *Hepatoma Research*. 2019;5:8.
42. Aziz K, Limzerwala JF, Sturmlechner I, et al. Ccne1 overexpression causes chromosome instability in liver cells and liver tumor development in mice. *Gastroenterology*. 2019;157:210-26.e12.
43. Tu H, Bonura C, Giannini C, et al. Biological impact of natural COOH-terminal deletions of hepatitis B virus X protein in hepatocellular carcinoma tissues. *Cancer Res*. 2001;61:7803-7810.
44. Kekule AS, Lauer U, Meyer M, Caselmann WH, Hofschneider PH, Koshy R. The preS2/S region of integrated hepatitis B virus DNA encodes a transcriptional transactivator. *Nature*. 1990;343:457-461.
45. Umeda M, Marusawa H, Seno H, et al. Hepatitis B virus infection in lymphatic tissues in inactive hepatitis B carriers. *J Hepatol*. 2005;42:806-812.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Li M, Shen Y, Chen Y, et al. Characterization of hepatitis B virus infection and viral DNA integration in non-Hodgkin lymphoma. *Int. J. Cancer*. 2020; 1–11. https://doi.org/10.1002/ijc.33027