LYMPHOID NEOPLASIA

Genetic landscape of hepatitis B virus–associated diffuse large B-cell lymphoma

Weicheng Ren,1* Xiaofei Ye,1,4 Hong Su,2,3,4 Wei Li,1,4,* Dongbing Liu,2,4 Mohammad Pirmoradian,1 Xianhuo Wang,4 Bo Zhang,1 Qiang Zhang,5 Longyun Chen,2,3 Man Nie,1,6 Yao Liu,1 Bin Meng,4 Huiqiang Huang,6 Wenyi Zeng,6 Wenyu Li,7 Kui Wu,2,3 Yong Hou,2,3 Klas G. Wiman,5 Zhiming Li,6 Huilai Zhang,4 Roujun Peng,4 Shida Zhu,2,3 and Qiang Pan-Hammarström1,4,6

1Department of Laboratory Medicine, Karolinska Institutet, Stockholm, Sweden; 2BGI-Shenzhen, Shenzhen, China; 3China National GeneBank, BGI-Research, Shenzhen, China; 4Department of Lymphoma, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center of Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin, China; 5Department of Oncology-Pathology, Cancer Center Karolinska, Karolinska Institutet, Stockholm, Sweden; 6Department of Medical Oncology, Sun Yat-Sen University Cancer Center, State Key Laboratory of Oncology in South China, and Collaborative Innovation Center of Cancer Medicine, Guangzhou, China; 7Lymphoma Division, Cancer Center, Guangdong General Hospital, Guangdong Academy of Sciences, Guangzhou, China; and 8VIP Region, Sun Yat-Sen University Cancer Center, State Key Laboratory of Oncology in South China, and Collaborative Innovation Center of Cancer Medicine, Guangzhou, China

KEY POINTS

- DLBCL patients with concomitant HBV infection are characterized by distinct clinical features.
- Genomic and transcriptomic analyses identified distinct mutation targets and tumorigenic pathways in HBV-associated DLBCLs.
- Hepatitis B virus (HBV) infection is endemic in some parts of Asia, Africa, and South America and remains to be a significant public health problem in these areas. It is known as a leading risk factor for the development of hepatocellular carcinoma, but epidemiological studies have also shown that the infection may increase the incidence of several types of B-cell lymphoma. Here, by characterizing altogether 275 Chinese diffuse large B-cell lymphoma (DLBCL) patients, we showed that patients with concomitant HBV infection (surface antigen positive [HBsAg+]) are characterized by a younger age, a more advanced disease stage at diagnosis, and reduced overall survival. Furthermore, by whole-genome/exome sequencing of 96 tumors and the respective peripheral blood samples and targeted sequencing of 179 tumors from these patients, we observed an enhanced rate of mutagenesis and a distinct set of mutation targets in HBsAg+ DLBCL genomes, which could be partially explained by the activities of APOBEC and activation-induced cytidine deaminase. By transcriptome analysis, we further showed that the HBV-associated gene expression signature is contributed by the enrichment of genes regulated by BCL6, FOXO1, and ZFP36L1. Finally, by analysis of immunoglobulin heavy chain gene sequences, we showed that an antigen-independent mechanism, rather than a chronic antigenic stimulation model, is favored in HBV-related lymphomagenesis. Taken together, we present the first comprehensive genomic and transcriptomic study that suggests a link between HBV infection and B-cell malignancy. The genetic alterations identified in this study may also provide opportunities for development of novel therapeutic strategies. (Blood. 2018;131(24):2670-2681)

Introduction

Diffuse large B-cell lymphoma (DLBCL) is one of the most common lymphoid malignancies, accounting for 30% to 40% of non-Hodgkin lymphomas (NHLs). Based on gene expression profiling, 2 major molecular subtypes with different chromosomal alterations and clinical outcomes have been identified: germinal center B-cell-like (GCB) and activated B-cell–like subtypes.1 Based on the latest 2016 World Health Organization classification, additional DLBCL associated subtypes have been defined, such as large B-cell lymphoma with IRF4 rearrangement, high-grade B-cell lymphoma with MYC, BCL2, and/or BCL6 translocations and Epstein-Barr virus (EBV)–positive DLBCL.2 These new entities highlight that specific chromosomal alterations or infectious agents are associated with different clinical manifestations. Further classification/subtyping of DLBCLs based on genetic characterizations is expected to have a significant impact on etiological studies, clinical management, and development of new therapeutic strategies for these patients.

Several studies, including the most recent one on 1001 patients, have explored the genome of DLBCL by using next-generation sequencing technologies.3-11 These studies have revealed recurrently occurring somatic mutations or structural variants that affect multiple cellular processes or signaling pathways, including apoptosis, B-cell receptor (BCR), and nuclear factor κB (NF-κB) signaling, epigenetic modifications, immune regulation/ evasion and cell migration. The mutation spectra identified, however, differ somewhat in the published patient cohorts and
the overlap of the mutated genes is still limited between these studies,7,8,12 which may reflect a high genetic heterogeneity in DLBCL. Of note, ethnic background seems to influence the mutational spectra, where some potential cancer-driver genes, such as CD70, CXCR4, DTX1, LYN, and TMSB4X, were mutated at a higher frequency in Chinese DLBCL patients, whereas GNA13 and EZH2 seemed to be more frequently mutated in patients from Western populations.8,13 The genomic changes in DLBCL may also be affected by the age of patients,14 alterations in DNA repair genes,15 exposure to different etiological agents,16 and/or autoimmune/inflammatory immune responses.17

Previous studies have shown that several types of virus, including HIV, EBV, human T-cell leukemia-lymphoma virus, and human herpesvirus-8, may contribute to the development of selected subtypes of lymphoma.18 A recent study has also shown that antiviral treatment can lead to a complete remission of hepatitis C virus (HCV)-associated, low-grade NHL, suggesting a causative role of HCV in these tumors.19 Meta-analyses have further shown that patients infected with hepatitis B virus (HBV) have a two- to threefold higher risk of developing NHL,20-23 including DLBCL.21,24 It also has been reported that DLBCL patients infected with HBV are diagnosed at a young age with a more advanced disease stage and show a significantly worse outcome.24 Thus, epidemiological and clinical studies have both suggested an association of HBV infection with DLBCL. However, the etiological role of HBV in lymphomagenesis remains largely unknown. Here, we performed a comprehensive genetic study on DLBCLs from patients with HBV infection and identified distinct molecular features of these tumors.

**Methods**

**Patient information**

Frozen tumor biopsy specimens from affected lymph nodes of 275 Chinese DLBCL patients were obtained from the Sun Yat-Sen University Cancer Center and Tianjin Medical University Cancer Institute and Hospital. DLBCL samples were classified as GCB or non-GCB based on the Hans algorithm.25 Of these, 266 were de novo and 9 might have resulted from the transformation of follicular lymphoma; 229 were taken at diagnosis, and 46 were taken at relapse (supplemental Table 1, available on the Blood Web site). Detection of HBV surface-antigen (HBsAg) and e-antigen and antibodies against HBV surface-antigen, e-antigen, and core-antigen was performed as routine blood tests in all patients, and the quantitation of viral DNA was performed in a subset of HBsAg+ patients (44 patients). The study was approved by the institutional review boards at the Sun Yat-Sen University Cancer Center, the Tianjin Medical University Cancer Institute, and Hospital and the Karolinska Institute.

**Whole-genome/exome sequencing (WGS/WES) analysis**

DNA was extracted with the DNeasy Tissue and Blood Kit (Qiagen, Venlo, The Netherlands). WGS/WES was performed in BGI-Shenzhen using either HiSeq2500 or HiSeq X10 platforms (Illumina, San Diego, CA) or Complete Genomics (BGI). 60 pairs of DLBCL/control samples were sequenced by WGS, and 47 pairs of samples were sequenced by WES. Among these, 11 pairs of samples have been sequenced by both WGS and WES. Analysis of significance of mutated genes was performed by 3 prediction methods,26-28 and the resulting top gene lists (q value < 0.1) were merged to identify significantly mutated genes in the discovery cohort. The mutational signature analysis was performed based on WGS data using a Bayesian version of the nonmatrix factorization method.29 More details are provided in supplemental Methods.

**Targeted capture sequencing by lymphochip**

Genes included in the targeted sequencing panel lymphochip were selected with the following criteria: (1) mutated in >3 cases within our discovery cohort and expressed in DLBCL samples, (2) recurrently mutated in >3 cases in other B-cell lymphoma cohorts, (3) important for DNA repair, and (4) important for targeted therapy. Together, this consists of the entire coding regions of 212 genes (supplemental Table 2), for a total size of 572 kb, and the lymphochip was synthesized at BGI. 100 ng DNA of each

---

**Table 1. Clinical characteristics in HBsAg+ and HBsAg− DLBCL patients**

<table>
<thead>
<tr>
<th></th>
<th>HBsAg+ DLBCL</th>
<th>HBsAg− DLBCL</th>
<th>P*</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of patients</td>
<td>56</td>
<td>219</td>
<td>—</td>
</tr>
<tr>
<td>Age, y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;60</td>
<td>10 (17.9)</td>
<td>117 (53.4)</td>
<td>.0001</td>
</tr>
<tr>
<td>≤60</td>
<td>46 (82.1)</td>
<td>102 (46.6)</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>16 (28.6)</td>
<td>86 (39.3)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>40 (71.4)</td>
<td>133 (60.1)</td>
<td>.1391</td>
</tr>
<tr>
<td>Performance status†</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-1</td>
<td>41 (73.2)</td>
<td>176 (81.9)</td>
<td>.1491</td>
</tr>
<tr>
<td>2-4</td>
<td>15 (26.8)</td>
<td>39 (18.1)</td>
<td></td>
</tr>
<tr>
<td>Elevated LDH†</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>32 (59.3)</td>
<td>97 (46.4)</td>
<td>.0922</td>
</tr>
<tr>
<td>No</td>
<td>22 (40.7)</td>
<td>112 (53.6)</td>
<td></td>
</tr>
<tr>
<td>Subtype†</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GCB</td>
<td>21 (37.5)</td>
<td>83 (38.1)</td>
<td>.9372</td>
</tr>
<tr>
<td>Non-GCB</td>
<td>35 (62.5)</td>
<td>135 (61.9)</td>
<td></td>
</tr>
<tr>
<td>Stage†</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I-II</td>
<td>14 (25.9)</td>
<td>112 (54.6)</td>
<td>.0002</td>
</tr>
<tr>
<td>III-IV</td>
<td>40 (74.1)</td>
<td>93 (45.4)</td>
<td></td>
</tr>
<tr>
<td>IPI†</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-2</td>
<td>28 (51.9)</td>
<td>146 (71.2)</td>
<td>.0070</td>
</tr>
<tr>
<td>3-5</td>
<td>26 (48.1)</td>
<td>59 (28.8)</td>
<td></td>
</tr>
<tr>
<td>Spleen involvement†</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>19 (38.0)</td>
<td>34 (17.6)</td>
<td>.0019</td>
</tr>
<tr>
<td>No</td>
<td>31 (62.0)</td>
<td>159 (82.4)</td>
<td></td>
</tr>
<tr>
<td>Liver involvement†</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>7 (14.0)</td>
<td>14 (7.3)</td>
<td>.1303</td>
</tr>
<tr>
<td>No</td>
<td>43 (86.0)</td>
<td>179 (92.7)</td>
<td></td>
</tr>
</tbody>
</table>

Values are reported as n (%) of patients unless indicated otherwise.

**IPI**, international prognostic index; **LDH**, lactate dehydrogenase.

*χ² test was used for comparison. Significant values (P < .05) are highlighted in bold.

†The calculation was based on 271, 263, 274, 259, or 243 samples with available data.

# The overlap of the mutated genes is still limited between these studies, which may reflect a high genetic heterogeneity in DLBCL. Of note, ethnic background seems to influence the mutational spectra, where some potential cancer-driver genes, such as CD70, CXCR4, DTX1, LYN, and TMSB4X, were mutated at a higher frequency in Chinese DLBCL patients, whereas GNA13 and EZH2 seemed to be more frequently mutated in patients from Western populations. The genomic changes in DLBCL may also be affected by the age of patients, alterations in DNA repair genes, exposure to different etiological agents, and/or autoimmune/inflammatory immune responses.

Previous studies have shown that several types of virus, including HIV, EBV, human T-cell leukemia-lymphoma virus, and human herpesvirus-8, may contribute to the development of selected subtypes of lymphoma. A recent study has also shown that antiviral treatment can lead to a complete remission of hepatitis C virus (HCV)-associated, low-grade NHL, suggesting a causative role of HCV in these tumors. Meta-analyses have further shown that patients infected with hepatitis B virus (HBV) have a two- to threefold higher risk of developing NHL, including DLBCL. It also has been reported that DLBCL patients infected with HBV are diagnosed at a young age with a more advanced disease stage and show a significantly worse outcome. Thus, epidemiological and clinical studies have both suggested an association of HBV infection with DLBCL. However, the etiological role of HBV in lymphomagenesis remains largely unknown. Here, we performed a comprehensive genetic study on DLBCLs from patients with HBV infection and identified distinct molecular features of these tumors.

**Methods**

**Patient information**

Frozen tumor biopsy specimens from affected lymph nodes of 275 Chinese DLBCL patients were obtained from the Sun Yat-Sen University Cancer Center and Tianjin Medical University Cancer Institute and Hospital. DLBCL samples were classified as GCB or non-GCB based on the Hans algorithm. Of these, 266 were de novo and 9 might have resulted from the transformation of follicular lymphoma; 229 were taken at diagnosis, and 46 were taken at relapse (supplemental Table 1, available on the Blood Web site). Detection of HBV surface-antigen (HBsAg) and e-antigen and antibodies against HBV surface-antigen, e-antigen, and core-antigen was performed as routine blood tests in all patients, and the quantitation of viral DNA was performed in a subset of HBsAg+ patients (44 patients). The study was approved by the institutional review boards at the Sun Yat-Sen University Cancer Center, the Tianjin Medical University Cancer Institute, and Hospital and the Karolinska Institutet.

**Whole-genome/exome sequencing (WGS/WES) analysis**

DNA was extracted with the DNeasy Tissue and Blood Kit (Qiagen, Venlo, The Netherlands). WGS/WES was performed in BGI-Shenzhen using either HiSeq2500 or HiSeq X10 platforms (Illumina, San Diego, CA) or Complete Genomics (BGI). 60 pairs of DLBCL/control samples were sequenced by WGS, and 47 pairs of samples were sequenced by WES. Among these, 11 pairs of samples have been sequenced by both WGS and WES. Analysis of significance of mutated genes was performed by 3 prediction methods, and the resulting top gene lists (q value < 0.1) were merged to identify significantly mutated genes in the discovery cohort. The mutational signature analysis was performed based on WGS data using a Bayesian version of the nonmatrix factorization method. More details are provided in supplemental Methods.

**Targeted capture sequencing by lymphochip**

Genes included in the targeted sequencing panel lymphochip were selected with the following criteria: (1) mutated in >3 cases within our discovery cohort and expressed in DLBCL samples, (2) recurrently mutated in >3 cases in other B-cell lymphoma cohorts, (3) important for DNA repair, and (4) important for targeted therapy. Together, this consists of the entire coding regions of 212 genes (supplemental Table 2), for a total size of 572 kb, and the lymphochip was synthesized at BGI. 100 ng DNA of each
The full length of the TAp73 sequence (tumor suppressor isoform of p73, a structural homology of p53) was cloned into a pcDNA 3.1 vector (Thermo Fisher Scientific, San Diego, CA). TAp73 mutants were generated by site-directed mutagenesis. TAp73 vectors and a reporter construct containing a p53-binding site in front of a GFP reporter (p53-responsive EGFP reporter) were cotransfected into H1299 cells (lacking p53 expression) with Lipofectamine 2000 (Thermo Fisher Scientific) in a 6-well plate. After 24 to 48 hours, cells were harvested and GFP level was measured by flow cytometry (Accuri C6; BD Biosciences).

Functional characterization of TP73 mutants

The full length of the TAp73 sequence (tumor suppressor isoform of p73, a structural homology of p53) was cloned into a pcDNA 3.1 vector (Thermo Fisher Scientific, San Diego, CA). TAp73 mutants were generated by site-directed mutagenesis. TAp73 vectors and a reporter construct containing a p53-binding site in front of a GFP reporter (p53-responsive EGFP reporter) were cotransfected into H1299 cells (lacking p53 expression) with Lipofectamine 2000 (Thermo Fisher Scientific) in a 6-well plate. After 24 to 48 hours, cells were harvested and GFP level was measured by flow cytometry (Accuri C6; BD Biosciences).

Statistics

P values were calculated using appropriate statistical tests (χ²-test, Fisher’s exact test or Mann-Whitney U test)
Clinical features in HBsAg\(^+\) DLBCL patients

In our DLBCL cohort, 20% of patients were HBsAg\(^+\) (56/275), which is notably higher than the positive rate in the general population (7%).\(^3\)\(^5\) There was no significant difference between the HBsAg\(^+\) and HBsAg\(^-\) groups in the percentage of GCB vs non-GCB patients or sex (Table 1). However, compared with HBsAg\(^-\) cases, the HBsAg\(^+\) DLBCL patients displayed significantly younger age at diagnosis (median age, 42 vs 60 years; \(P < .0001\)), higher international prognostic index, more advanced disease at diagnosis (stage III/IV), and more frequent involvement of spleen (Table 1). Furthermore, they showed significantly reduced overall survival as compared with the HBsAg\(^-\) patients (supplemental Figure 1A). When the 2 cell-of-origin subtypes or the treatment were considered separately, a significantly reduced overall survival was observed in HBsAg\(^+\) patients with the GCB subtype (supplemental Figure 1B-C). Patients with a relatively high HBV viral load (\(\geq 2000\) IU/mL) had clinical characteristics similar to those of the remaining HBsAg\(^+\) patients, except that they appeared to be younger at diagnosis (median age, 32 vs 48 years; \(P = .0281\); supplemental Figure 2A).

Genetic landscape of HBV-associated DLBCLs

Identification of significantly mutated genes in the discovery cohort

We next performed WGS/WES on DNAs prepared from tumor biopsy specimens and matched peripheral blood samples from 65 Chinese DLBCL patients. By including previously sequenced 31 Chinese DLBCL samples,\(^8\)\(^,\)\(^3\)\(^6\) altogether, 96 paired tumor/control samples were analyzed as a discovery cohort, including 20 HBsAg\(^+\) tumors. A median of 162 (11-460) or 10,835 (635-36,052) somatic mutations per tumor were observed in samples sequenced by WES or WGS, respectively. Among these, a median of 67 (5-160, WES) or 79 (9-159, WGS) nonsilent mutations in the coding region per tumor were observed.

Figure 2. Enhanced mutagenesis in HBsAg\(^+\) DLBCLs.

(A-B) Comparison of the mutation load in the whole genome (A) or coding genome (B) between HBsAg\(^+\) and HBsAg\(^-\) DLBCLs. (C) Major mutational signatures were identified according to the 96-substitution classifications from 60 Chinese DLBCLs characterized by WGS. (D) Signatures 2, 6, and 7 were enriched in the HBsAg\(^+\) samples. Error bars represent standard deviation. A Mann-Whitney \(U\) test was used to calculate the \(P\) value. Sig, signature.
Enhanced mutagenesis in HBV-associated DLBCLs

Three prediction methods were performed to define cancer-driver genes in our DLBCL discovery cohort.24-28 Altogether, 131 genes were affected by somatically occurring, nonsilent mutations in at least 5 samples (5%) and among these, 66 genes were considered to be statistically significantly mutated (Figure 1). The most frequently mutated genes (≥10%) were PIM1, BTG2, TP53, HIST1H1E, KMT2D/MLL2, B2M, BTG1, FAS, CD70, DTX1, SGK1, TMSB4X, KLF2, MYD88, BCL6, CD79B, ZFP36L1, SOCS1, HIST1H1C, and OSBPL10. Among the 66 potential cancer-driver genes, KLF2 (15%), ZFP36L1 (13%), OSBPL10 (10%), VMP1 (6%), TP73 (6%), MSL2 (6%), and MEOX2 (5%) have not been appreciated previously as being significant mutation targets in DLBCLs. Genes significantly mutated in our cohort were further assigned to the DAVID database for KEGG pathway analysis. Notably, consistent with the high HBV infection rate in our DLBCL cohort, one of the most significantly mutated pathways is the HBV-infection associated pathway, which has not been described previously. Additional significant pathways identified include BCR, JAK-STAT, and NF-κB signaling pathways (supplemental Table 3).

Enhanced mutagenesis in HBV-associated DLBCLs

The general mutation profile in the genome of HBsAg+ and HBsAg- DLBCLs was subsequently compared. The total mutation load in the whole genome was higher in HBsAg+ DLBCLs, although this difference did not reach statistical significance (median, 15,036 vs 9902; Figure 2A). In the coding genome, significantly more nonsilent mutations were observed in HBsAg+ DLBCLs (median, 99 vs 66; Figure 2B).

The genome-wide mutational signatures of 60 cases with WGS data were subsequently characterized based on the 96 possible mutation types. Seven highly confident mutational signatures were extracted from our cohort (Figure 2C), and signatures 2, 6, and 7 were significantly enriched in HBsAg+ tumors (Figure 2D). Signatures 2 and 6 have previously been described,29 whereas signature 7, which was mainly characterized by T to V mutations (71%, V = A/C/G), was novel. Of these 3 mutation signatures, signature 2 has been linked to APOBEC enzymes, which belong to a family of proteins that usually function in the defense against viral infections, such as HIV and HBV.37,38 Taken together, our data suggest that HBV-associated DLBCLs display overall enhanced mutagenesis and are associated with selected mutational signatures, which may be partially resulted from APOBEC enzyme activity.

### Figure 3. Somatic mutation spectrum in HBsAg+ and HBsAg- DLBCLs

The frequency of nonsilent mutations between HBsAg+ (n = 56) and HBsAg- (n = 219) DLBCLs that were sequenced by WES/WGS and/or lymphochip were compared by χ² test or Fisher’s exact test. Significant values (P < .05) are highlighted in bold. Potential AID off-targets are marked by an asterisk, and P values from Fisher’s exact test are marked by a number sign.
were analyzed by lymphochip, which targeted coding regions of 212 genes. The WES/WGS and lymphochip data were subsequently combined and analyzed in altogether 275 DLBCL samples (56 HBsAg\(^+\)). 75 genes were affected by nonsilent mutations in at least 3 HBsAg\(^+\) DLBCLs (5% of cases), representing the most frequently mutated targets in these tumors (supplemental Table 5). The mutation profiles in HBsAg\(^+\) and HBsAg\(^-\) DLBCLs were subsequently compared (Figure 3; supplemental Table 6). Firstly, 14 genes were confirmed to be preferentially mutated in the HBsAg\(^+\) group, including KLF2, TMSB4X, CD70, BCL6, FAS, TNFRSF14, UBE2A, CDS8, SGK1, ZFP36L1, CXCR4, FOXO1, CSK, and MSL2. Of these, 11 genes are potentially off-targets of activation-induced cytidine deaminase (AID), a B-cell–specific factor that initiates somatic hypermutation (SHM) in the immunoglobulin genes (marked with stars in Figure 3; supplemental Table 3). Secondly, well-known mutation targets in DLBCL, such as B2M, GNA13, BCL7A, CIITA, and GNAI2, were less frequently mutated in HBsAg\(^+\) tumors. However, there are some recurrently mutated genes that are equally distributed in HBsAg\(^-\) and HBsAg\(^+\) DLBCLs, such as PIM1, KMT2D, ITPKB, DTX1, MYD88, BTG1, HIST1H1E, TBL1XR1, and TNFAIP3. Among these, 6 genes are potentially AID off-targets (Figure 3; supplemental Table 7). Patients with a higher HBV viral load (supplemental Figure 2B-C) or occult HBV infection (supplemental Figure 3) might also be associated with certain frequently mutated genes. However, sample size for each specific group is notably limited. Taken together, these data showed a distinctive set of mutated genes in HBsAg\(^+\) DLBCLs, affecting multiple key pathways involving lymphomagenesis (Figure 4). Aberrant SHM mediated by AID activity may have contributed to the mutations observed in a subset of the highly mutated genes in HBsAg\(^+\) DLBCLs.

**Enriched BCL6 alterations in HBV-associated DLBCLs**

BCL6, a well-known proto-oncogene involved in the development of B-cell lymphoma, encodes a transcriptional repressor that is required for germinal center reactions. Close to half of the BCL6 mutations identified in our cohort were splicing mutations located at a hotspot position (chr3:187463196), which affects the splicing between the non-coding exons 1 and 2, resulting in retention of intron 1 (Figure 5A). The latter may affect RNA stability of the gene or lead to the disruption of the negative regulation loop.\(^{39,40}\) In samples characterized by WGS, the frequency of chromosomal translocation involving BCL6 was also significantly increased in HBsAg\(^+\) DLBCLs (57% vs 28%; \(P = 0.0422\), \(\chi^2\) test), with the BCL6 breakpoints occurring in a 3-kb region located at exon 1 and intron 1, the known major breakpoint region (Figure 5A). Taken together, the genetic alterations (nonsilent mutations and/or translocations) in BCL6 were significantly enriched in the HBsAg\(^+\) DLBCLs (79% vs 28%; \(P = 0.0013\), Fisher’s exact test), suggesting a critical role of BCL6 dysregulation in the development of HBV-associated DLBCL.

**Enriched mutations in KLF2 or ZFP36L1 in HBV-associated DLBCLs**

KLF2 encodes a transcription factor which is important for maintaining follicular B-cell identity, and its absence in mice leads to an expansion of marginal zone B cells.\(^{41}\) It is highly mutated (20% to 42%) in splenic marginal zone lymphoma\(^ {42,43}\) but was found to be either not recurrently mutated\(^ {39,42}\) or mutated at a relatively lower frequency in DLBCLs.\(^ {43}\) In our cohort, KLF2 was affected by nonsilent mutations in 17% of all tumors analyzed, and it was one of the most significantly mutated genes in HBsAg\(^+\) DLBCLs (29%; Figures 3 and 5B).

It has recently been shown that KLF2 can be directly regulated by the messenger RNA decay activator protein ZFP36L1, which is required for the development and maintenance of marginal zone B-cell compartment.\(^ {44}\) However, a significant frequency of mutation in ZFP36L1 has not been reported in splenic marginal zone lymphoma or DLBCL.\(^ {12}\) In our DLBCL cohort, ZFP36L1 was mutated in 11% of all samples and preferentially mutated in HBsAg\(^+\) DLBCLs (Figures 3 and 5C). The KLF2 and ZFP36L1 mutations were
mutually exclusive in most of the cases, and 41% of HBsAg+ DLBCL patients carried at least one nonsilent mutation in either gene, which was significantly higher than in HBsAg− DLBCL patients (41% vs 20%; P = .0011, χ² test). This suggests that KLF2 and ZFP36L1 are functionally linked in lymphomagenesis and that mutations in these molecules may affect a common process or pathway.

Genetic alterations in the p53 signaling pathway
Approximately one-third of the HBsAg+ DLBCLs carried nonsilent, almost mutually exclusive mutations in TP53, TP63, or TP73 (Figure 6A-C). The DNA-binding domains (DBD) of p53, p63, and p73 show >63% identity in amino-acid sequence and are structurally highly conserved. Eight TP63 and 7 TP73 mutations were located in the DBD domain of the respective
proteins, and most of these alterations are located at or close to TP53 mutation hotspots (Figure 6D). Furthermore, mutations in TP63 and TP73, but not in TP53, were enriched in HBsAg+ DLBCLs (18% vs 6%; \( P = .0072 \), \( \chi^2 \) test). Using a p53-responsive EGFP reporter assay, we further investigated whether TP73 mutations can affect the DNA-binding activity of p73. Cotransfection of a p53 binding site containing EGFP reporter and a wild-type (WT) TAp73-expressing construct into a p53-deficient cell line showed that WT TAp73 could bind to p53-responsive sites and transactivate GFP expression, although less efficiently as compared with WT p53 (Figure 6E). Two mutations (K192E and R300C) in the DBD domain displayed a significantly impaired ability to transactivate GFP expression. Two truncating mutants (G449fs and S492fs) showed a significantly increased ability of transactivation in the current assay (Figure 6E). However, the loss of the SAM domain and the C terminus in these mutants may abolish the physical interaction between p73 and other proteins such as PTEN, resulting in an inability to induce apoptosis in response to DNA damage. The other four mutations (M266I and P270L in the DBD domain and A530D and P534T in the SAM domain) did not affect the transcriptional activity of TAp73. Taken together, these results suggest that a subset of the identified TP73 mutations may impact the DNA-binding activity of p73.

**Gene expression analysis in HBsAg+ and HBsAg− DLBCLs**

We next characterized the transcriptome of 108 Chinese DLBCLs, including 24 HBsAg+ samples. A distinctive gene expression profile was observed in HBsAg+ tumors compared with HBsAg− tumors, including a group of upregulated genes (\( n = 377 \), such as MDM2, PIK3CD, SGK1, BCL2L1, CCND1, TP63, and several major histocompatibility complex class II molecule–related genes and a group of downregulated genes (\( n = 324 \), such as TLR9, CD320, and MYC (Figure 7A; supplemental Table 8). Furthermore, GSEA showed that 2 gene sets were significantly upregulated in HBsAg+ DLBCLs, including the antigen processing and presentation and p53

---

**Figure 6. Genetic alterations in TP53, TP63, and TP73 genes.** (A) Mutations identified in TP53, TP63, and TP73 in HBsAg+ and HBsAg- DLBCLs. (B-C) Diagram showing mutations identified in TP63 and TP73, respectively. (D) Alignment of the DNA-binding domain (DBD) of TP53, TP63, and TP73. OD, oligomerization domain; SAM, sterile α motif; TA, transactivation. (E) Characterization of transcriptional activity of mutant TP73. Data shown are from 3 independent experiments. Error bars represent standard deviation. *** \( P < .0001 \).
Figure 7. HBsAg⁺ DLBCLs displays a distinctive gene expression pattern. (A) HBsAg⁺ DLBCLs showed a distinctive gene expression profile. The normalized expression levels were analyzed using the Qlucore Omics Explorer software. Significantly and differentially expressed genes (P < .05) between HBsAg⁺ (n = 24) and HBsAg⁻ (n = 84) DLBCLs were used to draw the heatmap. (B) GSEA revealed pathway alterations in HBsAg⁺ DLBCLs compared with that in HBsAg⁻ DLBCLs. The GSEA was performed with 1000 sample permutations. Enrichments were considered significant at false discovery rate q < .05. (C) GSEAPreranked tool was used to analyze BCL6 core transcriptional signatures (120 genes, 47 BCL6 targeted genes derived from DLBCL, 40 ZFP36L1-bound genes, 44 and FOXO1-targeted genes along the preranked genes differentially expressed in HBsAg⁺ vs HBsAg⁻ DLBCLs. The enrichment was tested with 1000 gene set permutations. Enrichments were considered significant at false discovery rate q < .05. ES, enrichment score; NES, normalized enrichment score.
signaling pathways (Figure 7B). The GSEAPeranked tool was further applied, and the BCL6-targeted, ZFP36L1-bound, and FOXO1-bound genes were significantly enriched in genes differentially expressed between HBsAg⁺ and HBsAg⁻ DLBCLs (Figure 7C), which is consistent with the observation that these transcription factors were preferentially mutated in HBsAg⁺ DLBCLs.

Characterization of V(D)J region of immunoglobulin heavy chain (IgH) in HBsAg⁺ and HBsAg⁻ DLBCLs

By sequencing the V(D)J regions of IgH from 52 DLBCL samples using polymerase chain reaction cloning and/or high-throughput sequencing methods, we identified the major clone from 27 DLBCL samples, including 15 HBsAg⁺ samples (supplemental Table 9). The most frequent used heavy-chain variable (VH) region genes were VH4-34 (20%) and VH3-23 (20%) in HBsAg⁺ DLBCLs and VH4-34 in HBsAg⁻ DLBCLs (30%). All VH region genes were mutated (6-86 mutations; 3% to 31%), and there was no significant difference in mutation frequency between the 2 groups. The amino acid sequences deduced from CDR3 sequences are shown in supplemental Table 9, and no sequences was qualified as stereotyped according to the criteria described previously. Furthermore, we did not find any significant sequence homology between the CDR3 region in HBsAg⁺ samples and the anti-HBsAg or HBsAg binding protein using the BLAST program.

Discussion

The genetics of DLBCL has been studied by various NGS approaches, and the landscape of somatic mutation of these tumors is emerging. However, most of the DLBCL patients previously studied belonged to Western populations. By analyzing a small cohort of patients, we have previously shown a different mutation spectrum in Chinese DLBCLs. Here, by integrating WGS, WES, targeted resequencing, and RNA sequencing, we performed a comprehensive genetic characterization of 275 DLBCL samples and identified 7 new potential cancer drivers for Chinese patients, including KLF2, ZFP36L1, OSBPL10, VMP1, TP73, MSL2, and MEOX2. Furthermore, 20% of our patients were HBsAg⁺, which provided an opportunity to explore the genetic alternations in HBV-associated DLBCLs. We observed overall enhanced mutagenesis in the HBsAg⁺ DLBCL genomes and identified a unique set of genetic alterations and a distinct gene expression profile in these tumors.

Together with the epidemiology and clinical evidence, our genomic data strongly suggest a causative role of HBV infection in the development of DLBCL. A previous study has suggested a chronic antigenic stimulation as a likely mechanism for the oncogenic role of HBV for lymphoma. This is supported by a history of chronic HBV infection, a frequent involvement of spleen and retroperitoneal lymph nodes, and biased usages of certain variable regions of IgH and κ light-chain genes (IGHV4-34 and IGKV4-1) in HBV-associated DLBCLs, which is reminiscent of HCV-associated NHL, where most lymphoma cells express IGHV1-69 and VK3-A27 genes that encode antibodies specific for HCV-E2 antigens.

In our DLBCL cohort, the clinical presentation is very similar to that described previously, and we also observed that HBsAg⁺ patients had more frequent involvement of the spleen. However, by analyzing the IGHV of a subset of HBsAg⁺ DLBCL in our cohort, we did not find any evidence of a biased usage of IGHV genes or stereotyped CDR3 regions or any homology of CDR3 regions with anti-HBsAg antibodies. The latter was confirmed by reanalyzing the sequences published by Deng et al. Thus, unlike the classical antigen-driven, HCV-related lymphomas, the chronic antigenic stimulation model is less favored for HBV-associated DLBCLs. This is further supported by the clinical observations that HBV-associated DLBCLs do not respond to antiviral therapy.

An alternative mechanism could be that, like in EBV-driven lymphoma, HBV directly infects B cells, leading to the genetic alterations that contribute to the development of malignancies. HBV is by definition a hepatotropic virus, but it can also infect lymphocytes and the lymphoid system that has been shown to be an important reservoir for HBV. As HBV DNA has been previously found to be integrated into chromosomal DNA of cells from lymph nodes, one possible mechanism could be that like in HBV-induced hepatocellular carcinomas, HBV DNA may integrate into the B-cell genome and directly activate oncogenes or disrupt tumor suppressors. However, we were not able to detect HBV gene integration based on our WGS data. Another possibility, which is supported by our mutation signature analysis, is that HBV-induced, APOBEC-mediated mutagenesis contributes to the overall increased mutation burden in HBsAg⁺ DLBCLs. It is also of note that the majority of genes that were preferentially mutated in HBV-associated DLBCLs do not overlap with those in HBV- or HCV-related lymphomas. Taken together, genetic lesions resulting from chronic HBV infection of B cells due to chronic infection, leading to an enhancement of mutational activity mediated by AID and further contributed to the mutation profiles we observed in HBV-associated DLBCLs. Finally, expression of HBV viral protein, in particular HBx, may be directly involved in regulation of the p53 and NF-κB signaling pathways, as well as modulation of transcriptional networks. Taken together, genetic lesions resulting from chronic HBV infection of B cells may be a plausible mechanism underlying the oncogenic role of HBV in lymphomagenesis.

The HBV-associated genetic changes in DLBCLs mainly affect several pathways, including p53 signaling, FOXP3 signaling, and immune evasion. It is notable that this is the first report showing that TP63 and TP73 are frequently mutated in any type of cancer, strongly suggesting a unique role of these 2 tumor suppressors in lymphomagenesis, especially in individuals infected with HBV. Four genes that frequently altered in HBsAg⁺ DLBCLs can be assigned to the FOXO signaling pathway (BCL6, CXCR4, KLF2, and SGK1), which regulates the dark zone program of germinal centers and is a hallmark of tonic BCR signaling in DLBCL. It is possible that the genetic changes in this pathway promote the growth and survival of tumor cells in HBsAg⁺ DLBCLs in an antigen-independent, tonic BCR-signaling–dependent manner. In addition, considering the exceptionally high frequency of BCL6 genetic alternations in HBsAg⁺ tumors, therapies targeting BCL6 may help to suppress the growth of the tumor cells in these patients. Finally, among the genes related to immune regulation/evasion, CD70, TNFRSF14, and CD58 were frequently mutated in HBsAg⁺ DLBCLs, whereas B2M and CIITA, both related to the function of major histocompatibility complex
class II molecule, were less mutated. CD70 is critical for protection against EBV infection, and mutations in CD70 and TNFRSF14 in malignant B cells may also impair T- or natural killer cell-mediated antitumor responses, particularly in a microenvironment affected by HBV infection. Therapies that modulate the CD27-CD70, TNFRSF14-CD272, or CD58-CD2 pathways could thus be an interesting alternative to immunotherapy in HBsAg+ patients.41–43

In summary, using a comprehensive analysis of genetic alterations in HBV-infected DLBCLs, we present the first genetic evidence that suggest a direct link between HBV infection and B-cell lymphomagenesis. Based on the distinct clinical and molecular features, HBsAg+ DLBCL should be classified as a separate subtype of DLBCL. Further studies on larger cohorts of HBV-associated DLBCL, including those potentially with occult HBV infection, are required for identification and validation of candidate driver genes in these tumors and will shed light on the complex mechanism underlying HBV infection and B-cell lymphomagenesis.

Acknowledgments
The authors thank G. Melino for sharing the TP73 plasmid; L. Hammarström, R. Dalla-Favera, and L. Pasqualelli for helpful discussions and critically reading the manuscript; K. Basso for discussions on BCL6 gene signature; and the Science for Life Laboratory, the National Genomics Infrastructure, and Uppmax for providing assistance in sequencing and computational infrastructure.

This work was supported by the Swedish Cancer Society, the Swedish Research Council, the European Research Council (RNAEDIT-649019), the Swedish Childhood Cancer Fund, the Chinese Natural Science Foundation (81670184 and 81611130086), the Shenzhen Peacock Plan (KQTD20150330171505310), STINT (joint China-Sweden Mobility Program), Radiumhemmet, the Center for Innovative Medicine, and the KIDS Program at the Karolinska Institutet.

REFERENCES
19. Macioca N, O’Brien A, Ardeshkina K. Remission of follicular lymphoma after treatment for...
HBV-ASSOCIATED DIFFUSE LARGE B-CELL LYMPHOMA


