

Multi-Omics Panoramic Analysis of HBV Integration, Transcriptional Regulation, and Epigenetic Modifications in PLC/PRF/5 Cell Line

Guiwen Guan¹, Abudurexiti Abulaiti¹, Chia-Chen Chen¹, Zhiqiang Gu¹, Jing Yang², Ting Zhang¹, Xiaojie Chen³, Zhao Zhou¹, Fengmin Lu¹, and Xiangmei Chen¹

¹Peking University School of Basic Medical Sciences

²Guangdong Provincial People's Hospital

³Capital Medical University Affiliated Beijing Friendship Hospital

April 23, 2024

Abstract

Background: The clearance or transcriptional silencing of integrated HBV DNA is crucial for achieving a functional cure in patients with chronic hepatitis B (CHB) and reducing the risk of hepatocellular carcinoma (HCC) development. The PLC/PRF/5 cell line is commonly used as an in vitro model for studying HBV integration. In this study, we employed a range of multi-omics techniques to gain a panoramic understanding of the characteristics of HBV integration in PLC/PRF/5 cells and to reveal the transcriptional regulatory mechanisms of integrated HBV DNA.

Methods: Transcriptome long-read sequencing (ONT) was conducted to analyze and characterize the transcriptional activity of different HBV DNA integration sites in PLC/PRF/5 cells. Additionally, we collected data related to epigenetic regulation, including whole-genome bisulfite sequencing (WGBS), histone chromatin immunoprecipitation sequencing (ChIP-seq), and assays for transposase-accessible chromatin using sequencing (ATAC-seq), to explore the potential mechanisms involved in the transcriptional regulation of integrated HBV DNA.

Result: Long-read RNA sequencing analysis revealed significant transcriptional differences at various integration sites in the PLC/PRF/5 cell line, with higher HBV DNA transcription levels at integration sites on chr11, chr13, and the chr13/chr5 fusion chromosome t (13:5). Combining long-read DNA and RNA sequencing results, we found that transcription of integrated HBV DNA generally starts downstream of the SP1, SP2, or XP promoters. ATAC-seq data confirmed that chromatin accessibility has limited influence on the transcription of integrated HBV DNA in the PLC/PRF/5 cell line. Analysis of whole-genome bisulfite sequencing data showed that the methylation intensity of integrated HBV DNA was highly negatively correlated with its transcription level ($r = -0.8929$, $p = 0.0123$). After AzaD treatment, the transcription level of integrated HBV DNA significantly increased, especially for the integration chr17, which had the highest level of methylation. Through ChIP-seq data, we observed the association between histone modification of H3K4me3 and H3K9me3 with the transcription of integrated HBV DNA.

Authors

Guiwen Guan^{1#*}, Abudurexiti Abulaiti^{1,2#}, Chia-Chen Chen^{1,3#}, Zhiqiang Gu¹, Jing Yang⁴, Ting Zhang¹, Xiaojie Chen^{5,6}, Zhao Zhou¹, Fengmin Lu^{1,7*}, Xiangmei Chen^{1*}

Affiliations

¹Department of Microbiology & Infectious Disease Center, School of Basic Medical Sciences, Peking University, Beijing, 100191, China.

²Department of Cell Biology, School of Basic Medical Sciences, Peking University, Beijing, 100191, China.

³ National Heart and Lung Institute Faculty of Medicine (NHLI), Imperial College
22 London, Hammersmith campus, W12 0NN, London, UK

⁴ Medical Research Institute, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, 510080, China.

⁵ Liver Transplantation Center, National Clinical Research Center for Digestive Diseases, Beijing Friendship Hospital, Capital Medical University, Beijing 100050, China.

⁶ Clinical Research Center for Pediatric Liver Transplantation of Capital Medical University, Beijing 100050, China.

⁷ Beijing Key Laboratory of Hepatitis C and Immunotherapy for Liver Diseases, Peking University Hepatology Institute, Peking University People's Hospital, Beijing, 100044, China.

These authors contributed equally to this study

* **Corresponding author:**

Xiangmei Chen:

Department of Microbiology & Infectious Disease Center, School of Basic Medical Sciences, Peking University, Beijing, 100191, China.

Tel: +86-10-82802807; Fax: +86-10-82805136;

E-mail: xm_chen6176@bjmu.edu.cn

Fengmin Lu:

Department of Microbiology & Infectious Disease Center, School of Basic Medical Sciences, Peking University, Beijing, 100191, China.

Beijing Key Laboratory of Hepatitis C and Immunotherapy for Liver Diseases, Peking University Hepatology Institute, Peking University People's Hospital, Beijing, 100044, China.

Tel: 86-10-8280-5136; Fax: 86-10-8280-5136

E-mail: lu.fengmin@hsc.pku.edu.cn

Guiwen Guan

Department of Microbiology & Infectious Disease Center, School of Basic Medical Sciences, Peking University, Beijing, 100191, China.

Tel: 86-188-1017-2170

E-mail: guanguiwen@bjmu.edu.cn

Keywords: Hepatitis B virus; Viral integration; PLC/PRF/5; Multi-omics; Cripsr-cas9; RNA Interference.

Financial Support

This work was supported by the National Natural Science Foundation of China [grant numbers: 82272315] and Beijing Municipal Natural Science Foundation [grant numbers: 7222108].

Background

Globally, approximately 296 million individuals suffered from chronic HBV infection in 2019, resulting in over 820,000 deaths¹. Upon infecting the liver cells of the host, HBV not only generates covalently closed circular DNA (cccDNA) within the cells' nuclei but also integrates its double-stranded linear DNA (dslDNA) into the human genome². The presence of integrated HBV contributes to the instability of the chromosomal structure in patients and allows for the transcription and translation of viral proteins, including HBsAg and

truncated HBx^{3,4}. Research has demonstrated that integration is the primary contributor to the presence of HBsAg in CHB patients with HBeAg negative and low HBV load^{5,6}. The clearance or transcriptional silencing of integrated HBV DNA is crucial for achieving a functional cure in patients with CHB and reducing the risk of HCC. However, the regulation mechanism of the transcription of the integrated HBV DNA has not yet been fully elucidated.

Although some researchers have tried to simulate HBV DNA integration through DNA insertion techniques, these methods significantly differ from natural HBV DNA integration. Among cell models with integrated HBV DNA, Hep3B and Huh-1 have only a few integration sites, while PLC/PRF/5, the first discovered cell line integrated with HBV DNA⁷, possesses a large number of integration sites, making it the most classical model for studying HBV DNA integration. Furthermore, PLC/PRF/5 cells, unlike primary liver cells infected with HBV, do not exhibit HBV replication. This characteristic enables researchers to eliminate interference from cccDNA. However, the lack of comprehensive understanding of the complex HBV DNA integration patterns and transcription characteristics within PLC/PRF/5 cells limited the research on HBV DNA integration based on the PLC/PRF/5 cell line.

The availability of various high-throughput sequencing technologies and publicly accessible data in PLC/PRF/5 cells facilitate combined analysis of multi-omics data when investigating the transcriptional regulation mechanism of HBV DNA integration. In our previous study, we analyzed the HBV DNA integration sites in the PLC/PRF/5 cell line using long-read DNA sequencing technology⁸ and constructed a map of integrated HBV DNA in PLC/PRF/5. In this study, we aim to investigate the transcriptional activity of each integrated HBV DNA and explore the transcriptional regulatory mechanisms of integrated HBV DNA by combining DNA long-read sequencing, RNA long-read sequencing, WGBS, ChIP-seq, and ATAC-seq. The findings from this study have the potential to enhance our understanding of integrated HBV in PLC/PRF/5 cells and provide a theoretical foundation for the future development of drugs targeting HBV DNA integration.

Materials and Methods

Cell line and Cell culture

PLC/PRF/5 cells were purchased from the American Type Culture Collection (Manassas, VA, USA). Cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% fetal bovine serum (Gibco, Life Technologies, Carlsbad, CA, USA) in a humidified incubator maintained at 37°C and 5% CO₂.

RNA extraction and cDNA preparation

Total RNA was extracted from the tissue using TRIzol reagent (Takara, Kyoto, Japan). The purity of RNA was tested using the Nano Photometer spectrophotometer (IMPLEN, Westlake Village, USA). cDNA libraries were constructed from 1 µg of total RNA of PLC/PRF/5 cells using a cDNA-PCR Sequencing Kit (SQK-PCS109) according to the manufacturer's protocol. Briefly, reverse transcriptase was used to enrich full-length cDNAs and add defined PCR adapters to both ends of the first-strand cDNA, followed by 14 cycles of cDNA PCR using LongAmp Taq DNA polymerase (New England Biolabs, Ipswich, MA, USA) with an 8-minute elongation time. The PCR products were then subjected to ONT adaptor ligation using T4 DNA ligase (New England Biolabs, Ipswich, MA, USA). Agencourt XP beads (Beckman Coulter, Brea, USA) were used for DNA purification. The final cDNA libraries were loaded onto FLO-MIN109 flow cells and analyzed on a PromethION platform at Biomarker Technology Company (Beijing, China).

Oxford Nanopore Technologies long read processing

The data analysis methods are consistent with previous literature⁹. Briefly, raw reads were first filtered with a minimum average read quality score of 7 and a minimum read length of 500 bp. Ribosomal RNA was discarded after mapping to the rRNA database. Next, full-length, non-chimeric (FLNC) transcripts

were identified by searching for the primer at both ends of the reads. Clusters of FLNC transcripts were obtained after mapping to the reference genome with minimap2 (V2.24)¹⁰, and consensus isoforms were generated after polishing within each cluster by pifish. The raw FASTQ data was subjected to quality control and trimming using the built-in quality control tool of Porechop, with default parameters. Low-quality sequences and adapter contamination were removed from the data. The quality of ONT sequencing reads was evaluated using Nanoplot V 1.33.0¹¹, with default parameters. FLNC transcripts were aligned to 2× genotype A (PLC/PRF/5 cells) HBV genome and human genome reference hg19 using minimap2. Transcripts aligned to both HBV and human genomes were considered to originate from integrated HBV DNA. Finally, the identified HBV transcripts were visualized using Integrative Genomics Viewer software, version 2.11.4¹².

Processing of WGBS data

Raw sequencing FASTQ files were assessed for quality, adapter content, and duplication rates with FastQC v0.11.7, then trimmed using trim_galore. Data alignment was performed using Bismark (V0.24.0)¹³ to align the sequencing data to the human hg19 and HBV reference genomes, generating SAM files with default parameters. For methylation analysis, the bismark_methylation_extractor was used to generate a report on global genomic cytosine methylation. Then, the GlobalMethLev function in the viewBS (V0.1.11)¹⁴ was used to compare the overall methylation levels of the integrated HBV DNA with the host genome in PLC/PRF/5 cells. The MethOneRegion function in viewBS was used to visualize the methylation levels of the integrated HBV genome. The GlobalMethLev function in viewBS was used to calculate the average genomic methylation levels of different regions around the integrated sites, and the R software (V4.2.2) pheatmap was used to visualize the methylation results.

Processing of ATAC-seq data

Raw sequencing FASTQ files were assessed for quality, adapter content, and duplication rates with FastQC v0.11.7, trimmed using trim_galore, and aligned with bowtie2 (v2.6.0)¹⁵ (parameters: -N 1 -X 2000) to the HBV and human genomes. PCR duplicates bias was removed using the samtools markdup. ATAC-seq data were assessed for quality control using the ATACseqQC¹⁶, and the factorFootprints function was used to detect and visualize the signal changes around the motif regions where transcription factors bind. MACS2¹⁷ software was used for peak calling with parameters set as '-shift -100 -extsize 200'. BAM files were converted to BigWig format, with key parameters '-binSize 50 -normalizeUsing RPM', by deeptools¹⁸. The deeptools plotProfile tool was used to calculate ATAC-seq signal and visualize the chromatin accessibility levels. The HINT-ATAC tool was used to predict potential transcription factor information in PLC/PRF/5 cells.

Processing of ChIP-seq data

Raw sequencing FASTQ files were assessed for quality, adapter content, and duplication rates with FastQC v0.11.7, trimmed using trim_galore, and aligned with bowtie2 (default parameters) to the HBV and human genomes to generate BAM files. The markdup function in Samtools was used to remove PCR-amplified duplicate sequences from the BAM files. Next, chimeric reads were extracted using Samtools and aligned to the integrated HBV DNA using bowtie2 with the '-no-softclip' parameter to calculate the number of chimeric reads from different integration sites. Chimeric read counts were normalized using RPM, and differences between different samples were compared.

Comparison of histone modification levels between groups

The GSE113879 dataset WIG files were downloaded, representing the histone modification levels of each

position of HBV DNA normalized to RPM. A custom script was used to correct the start site to the EcoRI site. Next, the WIG file was converted to BigWig format using the UCSC wigToBigWig software. The average signal values of the replicate experiments were calculated using the deeptools bigwigCompare function with the ‘—operation mean’ parameter. Deeptools bigwigCompare function was further used to calculate fold enrichment over the input of averaged HBV RPM, setting the ‘—operation’ parameter set to \log_2 ratio. Calculation of histone modification levels near the integration site: The binary BigWig files of the GSM6341171, GSM6341172, and GSM6341173 datasets were downloaded from the GEO database. The deeptools plotProfile function was used to calculate the histone modification levels in the region near the integration site.

Real-time Reverse Transcription (RT)-PCR

Real-time RT-PCR was performed as described previously¹⁹. β -Actin was used as the reference gene to determine gene expression. The primers used for real-time RT-PCR are listed in **Table S1**.

Western blot analysis

Western blot analysis was performed as previously described¹⁹. Briefly, the lysed cell supernatant was run on an SDS-PAGE and blotted with antibodies. The antibodies used in western blot are anti-HBs (ab9193, Abcam, Cambridge, UK), anti-GAPDH (AP7873b, Abcepta, Suzhou, China), and anti- β -tubulin (AM1020b, Abgent, CA, USA).

Quantification of HBsAg

The supernatant of the cultured cells was harvested at 48 h post-treatment. Levels of hepatitis B surface antigen (HBsAg) in the supernatant were quantified using an HBsAg quantitative determination kit (Shenzhen New Industries Biomedical Engineering Co., Ltd., Shenzhen, China) on a MAGLUMI X3 series automatic chemiluminescence immunoassay analyzer according to the manufacturer’s instructions.

Results

Research design

The PLC/PRF/5 cell line is a well-established *in vitro* cell model for studying hepatitis B virus (HBV) integration and hepatocellular carcinoma⁷. Therefore, we initially collected the high-throughput sequencing datasets associated with PLC/PRF/5 cells from GEO (<https://www.ncbi.nlm.nih.gov/geo/>) and SRA databases (<https://www.ncbi.nlm.nih.gov/sra>).

Our previous work has identified the multiple HBV integration in PLC/PRF/5 cells using long-read whole-genome sequencing technology (PacBio)⁸. By integrating our previous PLC/PRF/5 DNA sequencing data obtained with the PacBio platform and the data acquired through targeted DNA long-read sequencing techniques in studies by Meng, Ricardo Ramirez’s studies^{20,21}, we constructed a reference genome for HBV integration in the PLC/PRF/5 cell line. Furthermore, we conducted transcriptome long-read sequencing (ONT) to identify transcriptional characteristics of integrated HBV DNA in PLC/PRF/5 cells. Based on transcriptomic analysis results, we further employed epigenetic-related sequencing data, including WGBS, ChIP-seq, and ATAC-seq, to determine the relationship between the epigenetic modification and the transcription of integrated HBV DNA. Cellular experiments were conducted to validate the regulatory mechanisms exerting integrated HBV DNA transcription in PLC/PRF/5 cells. The analysis flow chart of this study is illustrated in **Fig. 1**. The datasets used in this study are summarized in **Table S2**.

The transcriptional landscape of integrated HBV DNA in PLC/PRF/5 cells

Due to limitations of low throughput and short read lengths, neither Sanger sequencing nor second-generation high-throughput sequencing can accurately characterize the transcriptional features of integrated HBV DNA in PLC/PRF/5 cells. Therefore, we employed long-read transcriptome sequencing based on the ONT platform to sequence the transcripts from PLC/PRF/5 cells and investigated the transcriptional features of integrated HBV DNA.

A total of 7,238,100 cleaning reads was obtained, with an average length of 806 bp. Among these reads, 1630 were identified as chimeric fragments of the HBV DNA and human genome (**Fig. S1**). Initially, chimeric reads were aligned to the genome of each integrated HBV (previously acquired using PacBio WGS sequencing by our research group, as well as in the studies by Meng and Ricardo Ramirez) to assign a transcriptional signature to each integration event. As depicted in **Fig. 2**, the integrated HBV DNA on chr4 and t (1;8) (chromosomal fusion between chromosome 1 and 8) did not initiate transcription, while the remaining integrated HBV DNA fragments did. Specifically, the integrated HBV DNA on chr11, chr12, chr13, chr16, and chr17 utilized HBV promoters, while those on chr3 and chr17 employed host promoters.

The characteristic of HBV chimeric transcripts derived from different integrated HBV DNA

To investigate the transcriptional activity of HBV DNA at various integration sites in PLC/PRF/5 cells, we quantified chimeric transcripts originating from distinct integration sites using long-read transcriptome sequencing. The results showed that the transcription levels of HBV DNA on different integration sites vary substantially. The highest number of transcripts were derived from integrated HBV DNA on chr11, chr13, and t (5;13) (**Fig. 3A**), suggesting that these three integration sites exhibit more active HBV DNA transcription and are the primary sources of viral protein expression in the PLC/PRF/5 cells.

Next, we focus on the transcripts from integrated HBV DNA on different chromosomes in PLC/PRF/5 cells, revealing transcriptional features of the integrated HBV DNA. We discovered that the majority of integrated HBV DNA utilized promoters SP1, SP2, and XP to initiate transcription. Additionally, human polyA served as transcription terminal signals, generating chimeric PreS1, PreS2/S, and HBx transcripts (**Fig. 3B**). While the HBV integrated fragment contains multiple promoters (chr13, t (13;5)), the transcription of integrated HBV exhibits promoter selectivity (**Fig. 3C & Fig. S2B**). In addition, the RNA originating from the integrated HBV DNA further exhibited splicing properties (chr12), forming fusion transcripts with host exons (chr12. **Fig. 3D**). Furthermore, we observed that certain integrated HBV DNA initiate transcription with the host promoter (chr3), and transcription ceased via HBV canonical poly(A) (**Fig 3E**). Besides, multiple transcriptional modes can arise from the same integrated HBV DNA (**Fig. 3F & Fig. S2A**). Due to the presence of host homologous sequences between different integration sites, the origins of two types of transcripts cannot be determined with certainty (**Fig. S2C & S2D**).

To verify the accuracy of long-read sequencing results, we conducted reverse transcription-polymerase chain reaction (RT-PCR) experiments to detect HBV chimeric transcripts with the highest expression from chr13 & t (5;13) and the lowest expression from chr16. The locations of the primers and the sizes of the amplified fragments are shown in **Fig. 3G and Table S3**. The RT-PCR results reveal that a prominent amplification band of approximately 500 bp is visible for the chimeric transcript from chr13 & t (5;13), while the chimeric transcript from chr16 is nearly undetectable (**Fig. 3H**). This confirms that the transcription level of integrated HBV DNA on chr13 & t (5;13) is higher than that on chr16. Furthermore, we performed Sanger sequencing on the PCR amplification product of the HBV chimeric transcript from chr13 & t (5;13), and the obtained sequence consists of ONT long-read sequencing (**Fig. 3I**), further validating the accuracy of the results of PLC/PRF/5 cell transcriptome long-read sequencing.

Taken together, these results suggest significant differences of transcriptional level and characteristic of integrated HBV DNA on various sites and the preservation of promoters on integrated HBV DNA fragments is important for the transcription of integrated HBV DNA.

The effect of chromatin accessibility on the transcription of integrated HBV DNA is limited in PLC/PRF/5 cells

The accessibility of chromatin is an essential mechanism to regulate gene transcription²². To explore the correlation between chromatin states around HBV integration sites in PLC/PRF/5 cells and the transcriptional activity of integrated HBV DNA, we further analyzed the ATAC-seq dataset (GSM4217243) from PLC/PRF/5 cells. We divided the integrated HBV into two groups based on transcriptional activity and compared the chromatin accessibility levels around the integration sites (within 100 kb upstream and downstream). HBV DNA integrated into chr3, chr11, chr12, and chr13 is comparatively active in transcription and has been defined as the transcriptionally active group, whereas HBV DNA integrated into chr4 and chr8 shows low levels of transcription and has been defined as the transcriptionally inactive group. The results show that in some regions around transcriptional active integration sites, ATAC-seq signal are significantly higher than those around transcriptional inactive integration sites (**Fig. 4A**). Furthermore, we used the log2CPM of mapped reads in a certain region to normalize and measure the state of chromatin accessibility in that region. The correlation analysis between this value and the transcription level of HBV DNA at each integration site showed a positive correlation trend, but the correlation was not statistically significant ($r = 0.3739$, $p = 0.3216$) (**Fig. 4B**). Therefore, the role of chromatin accessibility in regulating integrated HBV DNA transcription is limited in PLC/PRF/5 cells.

The principle of ATAC-seq technology is based on the Tn5 enzyme binds and cuts open chromatin. When transcription factors were bound to DNA, it will prevent Tn5 cleavage in an otherwise nucleosome-free regions, resulting in small regions called footprints²³. For instance, in the case of the transcription factor NFYC, Tn5 cleavage in these footprints decreases significantly within peak regions of high cleavage probability (**Fig. 4C**). Based on this, we used HINT-ATAC to predict transcription factors that potentially play a role in transcription of integrated HBV DNA in PLC/PRF/5 cells. We selected the 100 kb region upstream and downstream of the HBV integration site to predict and enrich transcription factors that may play a role in the vicinity of the integration site. As shown in **Fig. 4D**, NFYC, KLF15, FOS, JUN, and other transcription factors are enriched around the HBV integration site compared to randomly selected background regions. Since previous studies have reported that those transcription factors can bind to the HBV cccDNA promoter sequence and initiate HBV gene transcription²⁴⁻²⁶, we hypothesize that these transcription factors may bind to integrated HBV DNA or host DNA in the vicinity of the integration site, participating in the initiation of integrated HBV DNA transcription.

Methylation modification inhibits the transcription of integrated HBV DNA in PLC/PRF/5 cells

The methylation level of the genome is one of the important factors affecting gene transcription²⁷. In order to explore whether the transcription of integrated HBV DNA is influenced by methylation levels, we firstly compared the average methylation levels of integrated HBV DNA with the average methylation levels of the whole human genome in PLC/PRF/5 cells. As shown in **Fig. 5A**, the average methylation level of integrated HBV DNA was lower than that of the human genome (HBV 19.7% vs. human 32.9%). Considering that the transcriptional activity of cccDNA in the form of mini-chromosomes in hepatocyte nuclei is also regulated by DNA methylation modifications, we further analyzed the methylation levels of cccDNA extracted from HBV-infected HepG2-NTCP cells (from dataset SRR10426842)²⁸ and compared them with the methylation characteristics of integrated HBV DNA in the PLC/PRF/5 cells. The results showed that the methylation levels of CpG island 1, CpG island 2, and CpG island 3 on integrated HBV DNA²⁹ were higher than those on the corresponding CpG islands of cccDNA (**Fig. S3A, B, C**). This result suggests that the episomal HBV genome itself may be at a low methylation level, and when HBV DNA is integrated into the human genome, it is methylated by the host methylation modification system, resulting in suppressed expression. This high methylation modification of foreign integrated DNA may be one of the mechanisms of host cell

self-protection.

Next, we analyzed the differences in methylation levels of the host chromosome at various distances from the HBV integration sites in PLC/PRF/5 cells (500 bp, 1 kb, 2kb, 5 kb, 10kb, 20kb, and 50 kb). As shown in **Fig. 5B**, the average methylation level within the 500 bp region nearby HBV integration sites were relatively low compared to the genome at a distance from the integration site. Subsequently, we performed a correlation analysis between the average methylation level of the host genome within the 500 bp region nearby HBV integration sites and the number of HBV chimeric transcripts from each integration site. As shown in **Fig. 5C**, the transcription level of integrated HBV DNA was strongly negatively correlated with the methylation level ($r = -0.8929$, $p = 0.0123$). The host genome around the HBV integration sites on chromosomes 5, 11, 12, and 13 was relatively hypomethylated, as shown in **Fig. 5B**, and these sites exhibited high transcriptional activity of HBV DNA. Conversely, the methylation levels around the integration sites on chromosomes 3, 16, and 17 were relatively higher (see **Fig. 5B**), correlating with low transcriptional activity of the integrated HBV DNA. These results suggest that methylation status of the adjacent host genome affect the transcriptional level of integrated HBV DNA. 5-aza-2'-deoxycytidine (AzaD) is a cytidine deoxynucleo-9 side analogue. It inhibits DNA methyltransferases (DNMTs), leading to DNA hypomethylation including gene promoter regions

³⁰. To further verify the effect of DNA methylation on the transcription of integrated HBV DNA, we used AzaD to inhibit the DNA methylation and detected the effects of AzaD on the expression of integrated HBV DNA. The results showed that AzaD treatment significantly upregulated the HBs RNA levels in PLC/PRF/5 cells, presenting a dose-dependent effect, with significantly higher HBs RNA levels at treatment concentrations of 1 μ M and 4 μ M (**Fig. 5D**). Consistent with this result, the Western blot results also confirmed that AzaD treatment upregulated the intracellular HBsAg protein levels in a dose-dependent manner (**Fig. 5E**). In addition, we also analyzed the transcription levels of HBV RNA at each integration site before and after AzaD treatment using quantitative reverse transcriptase PCR (qRT-PCR). The results showed that the transcription level of HBV DNA integrated into chr17 significantly increased after AzaD treatment (**Fig. 5F**), and this integration site has the highest level of methylation among HBV integration sites in the PLC/PRF/5 cell line (**Fig. 5B**). These results further confirm that the transcription of integrated HBV in PLC/PRF/5 cells is regulated by methylation.

Transcription of integrated HBV DNA is associated with histone modifications in PLC/PRF/5 cells

Previous studies have demonstrated that H3K4Me3 is presented at the transcription start site (TSS) of actively transcribed genes, promoting transcription by rapidly recruiting RNA polymerase for mRNA synthesis, while H3K9Me3 is predominantly found in transcriptionally silent heterochromatin, hindering RNA polymerase access to the promoter region and inhibiting mRNA transcription³¹. To investigate whether the transcription of integrated HBV DNA is associated with histone modifications in PLC/PRF/5 cells, we analyzed the effect of modifications of H3K9Me3 and H3K4Me3 on transcription of integrated DNA in the PLC/PRF/5 using Chip-seq data from GEO database.

Firstly, to comprehend the distinct deposition of histone post-translational modifications (PTMs) on cccDNA and integrated HBV DNA, we compared the deposition patterns of two histone modifications on integrated HBV DNA in PLC/PRF/5 cells and cccDNA from liver biopsy samples of HBeAg-positive (HBeAg⁺) and HBeAg-negative (HBeAg⁻) CHB patient's liver biopsy samples (**Fig. 6A & Fig. S4**). We observed three deposition features: first, the inhibitory H3K9me3 modification is less enriched on integrated HBV DNA, HBeAg⁺ cccDNA, and HBeAg⁻ cccDNA compared to the H3K4me3; second, the deposition of H3K4me3 is relatively highly enriched and exhibits a similar pattern on integrated HBV DNA and HBeAg⁺ patient cccDNA, while its enrichment level is lower on HBeAg⁻ patient cccDNA; third, on integrated HBV DNA and HBeAg⁺ cccDNA, the activating histone PTM H3K4me3 is highly enriched around promoter SP1 and SP2 and the S gene ORF, while the inhibitory histone PTM H3K9me3 is relatively less enriched. Taken

together, the deposition pattern of histone PTM on integrated HBV DNA and HBeAg⁺ cccDNA is similar, with highly enriched activating histone PTM H3K4me3 and relatively low enrichment of inhibitory histone PTM H3K9me3, promoting active transcription of HBV DNA.

Next, we explored the correlation between the activating histone PTM H3K4me3 and the inhibitory histone PTM H3K9me3 on different integration sites of HBV DNA and integrated HBV DNA transcriptional activity. Since the ChIP-seq library was sequenced by NGS, limited by read length and shared overlapping sequences of integrated HBV DNA, we could only select histone PTM levels at the junction of HBV and the human genome to represent the histone PTM status at each integration site, and use the normalized read count of the HBV-host junction reads to represent the enrichment level of histone PTM. The result showed that the activating histone PTM H3K4me3 was positively correlated with integrated HBV DNA transcription levels, although not statistically significant ($r = 0.6109$, $p = 0.0878$), while the inhibitory histone PTM H3K9me3 enrichment level was negatively correlated with transcription levels ($r = -0.7806$, $p = 0.0223$) (**Fig. 6B & 6C**).

Furthermore, we analyzed the histone modification levels of the host genome within 5 kb upstream and downstream of the integration sites. The results revealed that integration sites with active transcription exhibited higher levels of activating histone PTMs, such as H3K4me1, H3K4me3, and H3K27ac, compared to transcriptionally inactive sites. Compared to H3K27ac, H3K4me1 and H3K4me3 had higher levels of activating histone modifications around integration sites (**Fig. 6D**). Furthermore, we analyzed the correlation between histone modification levels around each integration site and integrated HBV DNA transcription levels at each site. The analysis revealed a significant positive correlation between H3K4me3 modification levels near integration sites and integrated HBV DNA transcription levels ($r = 0.809$, $p = 0.0083$), while no significant correlation was found for H3K4me1 and H3K27ac modifications (**Fig. 6E**).

These findings suggest that histone modifications may play a crucial role in regulating integrated HBV DNA transcription.

Discussion

Previous studies have shown that integrated HBV DNA, has the potential to express HBsAg and truncated HBx protein, ultimately triggering HCC³. Despite numerous systemic studies revealed the transcriptional characteristics of HBV cccDNA, detail research regarding the transcriptional patterns of integrated HBV DNA remains insufficient. This study is based on genomic architecture of integrated HBV DNA in PLC/PRF/5 cells identified previously by WGS long read sequencing, and is the first to comprehensively determine the transcriptional landscape of each integrated HBV DNA in PLC/PRF/5 cells through RNA long-read sequencing.

We conducted transcriptome sequencing and obtained the transcription spectrum of integrated HBV DNA in PLC/PRF/5 cells for the first time. Ascribing HBV chimeric transcripts to the integrated HBV genome, we found that, the transcriptional activity of different integrated HBV DNA varies significantly. The most active transcription of HBV DNA occurred in those on chr11, chr13, and t (5;13). The integrated HBV DNA on chr4 cannot initiate transcription due to the absence of HBV promoters. This implies that and the preservation of promoters on integrated HBV DNA fragments is essential for the transcription of integrated HBV DNA. However, even though the integrated HBV DNA on chr8 retains complete SP1 and SP2 promoters, it does not undergo transcription. Furthermore, significant differences exist in the transcription levels of integrated HBV DNA on various sites, even with complete promoter preservation. This suggests that host factors may influence the transcriptional activity of integrated HBV DNA beyond viral promoter sequences.

According to previous research, HBV is more likely to integrate into transcriptionally active regions³², indicating that chromatin accessibility around integration sites may impact the transcription of integrated HBV DNA. Through analyzing the ATAC-seq data of PLC/PRF/5 cells, we discovered that the chromatin around transcriptionally active integration sites were more accessible than that around integration sites

without transcription. However, correlation analysis showed that the accessibility of chromatin near the integrated HBV DNA does not have a significant correlation with its transcription level, suggesting that the accessibility of chromatin has limited regulatory effects on the transcription of integrated HBV DNA in the PLC/PRF/5 cell line. Furthermore, some transcription factors were predicted that may function around HBV integration sites, including NFYC, KLF15, FOS, JUN, and JUNB by ATAC-seq data. Previous research has shown that these transcription factors can bind to HBV promoters and enhance their transcriptional activity. For instance, NFYC can bind to HBV's Sp1 promoter²⁶, KLF15 can bind to HBV's Sp1 and CP promoters²⁴, and JUN-FOS can bind to the Sp2 promoter²⁵. These results suggest that transcription factors may regulate the transcription of integrated HBV DNA in the PLC/PRF/5 cell line, but still need further experimental verification to validate their actual functions on the transcription of integrated HBV.

In eukaryotic cells, CpG islands in the promoter regions of host genes are modified under the action of methyltransferases, thereby inhibiting gene transcription³³. The literature reports that cccDNA contains three CpG islands, which also undergo methylation and affect virus replication³⁴. Watanabe's study has found that the methylation level of integrated HBV DNA is strongly correlated with the methylation level of the adjacent host genome³⁵, Watanabe's study has found that the methylation level of integrated HBV DNA is strongly correlated with the methylation level of the adjacent host genome, but it is still unclear whether the methylation of integrated HBV is regulated by the host. Our study found that the methylation level of integrated HBV DNA is higher than that of cccDNA. This result suggests that the episomal HBV genome itself may be at a low methylation level, and when HBV DNA is integrated into the human genome, it is methylated by the host methylation modification system, resulting in suppressed expression. This high methylation modification of foreign integrated DNA may be one of the mechanisms of host cell self-protection. However, it has not yet been elucidated whether methylation affects the transcription of integrated HBV. In this study, we found that the methylation level of integrated HBV DNA was higher than that of cccDNA but lower than that of the host genome. In addition, we observed a significant negative correlation between methylation of host genome around integration sites and HBV DNA transcription levels. These results suggest that the HBV genome is methylated to inhibit transcription after it is integrated into the host genome, and methylation modification is an important mechanism affecting the replication of integrated HBV.

Histone PTMs is another epigenetic modification which plays an important role in regulating gene transcription, occurring in both cccDNA and the host genome³⁶. In this study, we found that the H3K4me3 in the S gene region of integrated HBV DNA in PLC/PRF/5 cells had a similar deposition pattern to that of cccDNA from HBeAg⁺ patients, which suggests both integrated HBV DNA in PLC/PRF/5 cells and cccDNA from HBeAg⁺ patients have a higher transcription activity. Furthermore, we observed a highly enrichment of inhibitory histone PTM H3K9me3 in integrated HBV DNA without transcription on t (1;8), indicating that the chromatin around this integration site was in a state of heterochromatin and consequently inhibiting the transcription of HBV DNA. Additionally, we observed that on the integrated HBV DNA in the PLC/PRF/5 cell line, the deposition of H3K4me3 and H3K9me3 seems to be mutually exclusive. However, in the liver tissues of patients with chronic hepatitis B, H3K4me3 and H3K9me3 are not always mutually exclusive, indicating that the epigenetic modifications of HBV cccDNA in the human liver are more complex.

In addition to the promoter of integrated HBV, methylation levels, and histone modification levels, we also focused on the impact of the cell cycle on HBV transcription in PLC/PRF/5 cells. We found that, compared to PLC/PRF/5 cells cultured with 10% FBS, those cultured with 0.5% FBS experienced cell cycle arrest, and at the same time, the transcription level of integrated HBV was increased (**Fig. S5**).

This study also has many limitations. In our analysis of the correlation between histone modifications of integrated HBV DNA and its transcription levels, we only included data for H3K4me3 and H3K9me3 histones. Although these two histones are representative, future studies will need to analyze correlations with more histone modifications to comprehensively validate the impact of epigenetic modifications on the transcription of integrated HBV DNA. Additionally, HBV DNA fragments that are too short or transcribed from homologous regions cannot be accurately sourced, which will affect the precision of result analysis. In

this study, we focused solely on the transcriptional regulation of HBV integration in the PLC/PRF/5 cell line. In the future, it is necessary to analyze the transcriptional regulation of HBV integration in primary hepatocytes under natural infection conditions using single-cell multi-omics technologies.

In summary, through comprehensive analysis, we successfully constructed an integration and transcription map of HBV integration in the PLC/PRF/5 cell line. Our research confirms that the structure of integrated HBV DNA, methylation levels, and the modification levels of H3K4me3 and H3K9me3 histones affect the transcription of integrated HBV DNA. These results suggest that the regulation of epigenetic modifications in integrated HBV DNA could lead to transcriptional silencing and provide novel insights to improve the functional cure rate for CHB patients while reducing the incidence of liver cancer.

Data Availability Statement

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request. Other datasets from public databases are listed in Table S2.

Author contributions

Study concept and design was performed by XC, FL, GG and AA. Analysis and interpretation of data were performed by AA, GG, CC, ZG, JY, TZ, ZZ. Administrative, technical, or material support was performed by XC and FL. Study supervision was performed by XC. The first draft of the manuscript was written by XC, AA, GG, CC, and XC, and all authors commented on.

Conflict of Interest statement

The authors declare that they have no competing interests

REFERENCES

1. Jeng WJ, Papatheodoridis GV, Lok ASF. Hepatitis B. *Lancet (London, England)*. 2023;401(10381):1039-1052.
2. Howell J, Seaman C, Wallace J, et al. Pathway to global elimination of hepatitis B: HBV cure is just the first step. *Hepatology (Baltimore, Md)*. 2023;78(3):976-990.
3. Wooddell CI, Yuen MF, Chan HL, et al. RNAi-based treatment of chronically infected patients and chimpanzees reveals that integrated hepatitis B virus DNA is a source of HBsAg. *Science translational medicine*. 2017;9(409).
4. Hu B, Wang R, Fu J, et al. Integration of hepatitis B virus S gene impacts on hepatitis B surface antigen levels in patients with antiviral therapy. *Journal of gastroenterology and hepatology*. 2018;33(7):1389-1396.
5. Meier MA, Calabrese D, Suslov A, Terracciano LM, Heim MH, Wieland S. Ubiquitous expression of HBsAg from integrated HBV DNA in patients with low viral load. *Journal of hepatology*. 2021;75(4):840-847.
6. Manesis EK, Papatheodoridis GV, Tiniakos DG, et al. Hepatitis B surface antigen: relation to hepatitis B replication parameters in HBeAg-negative chronic hepatitis B. *J Hepatol*. 2011;55(1):61-68.
7. Chakraborty PR, Ruiz-Opazo N, Shouval D, Shafritz DA. Identification of integrated hepatitis B virus DNA and expression of viral RNA in an HBsAg-producing human hepatocellular carcinoma cell line. *Nature*. 1980;286(5772):531-533.
8. Chen CC, Guan G, Qi X, et al. Pacbio Sequencing of PLC/PRF/5 Cell Line and Clearance of HBV Integration Through CRISPR/Cas-9 System. *Frontiers in molecular biosciences*. 2021;8:676957.
9. Guan G, Zou J, Zhang T, Lu F, Chen X. A global survey of alternative splicing of HBV transcriptome using long-read sequencing. *Journal of hepatology*. 2022;76(1):234-236.

10. Li H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics (Oxford, England)*. 2021;37(23):4572-4574.
11. De Coster W, D'Hert S, Schultz DT, Cruets M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics (Oxford, England)*. 2018;34(15):2666-2669.
12. Robinson JT, Thorvaldsdóttir H, Wenger AM, Zehir A, Mesirov JP. Variant Review with the Integrative Genomics Viewer. *Cancer research*. 2017;77(21):e31-e34.
13. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics (Oxford, England)*. 2011;27(11):1571-1572.
14. Huang X, Zhang S, Li K, Thimmapuram J, Xie S, Wren J. ViewBS: a powerful toolkit for visualization of high-throughput bisulfite sequencing data. *Bioinformatics (Oxford, England)*. 2018;34(4):708-709.
15. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature methods*. 2012;9(4):357-359.
16. Ou J, Liu H, Yu J, et al. ATACseqQC: a Bioconductor package for post-alignment quality assessment of ATAC-seq data. *BMC genomics*. 2018;19(1):169.
17. Zhang Y, Liu T, Meyer CA, et al. Model-based analysis of ChIP-Seq (MACS). *Genome biology*. 2008;9(9):R137.
18. Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic acids research*. 2014;42(Web Server issue):W187-191.
19. Yan Y, Allweiss L, Yang D, et al. Down-regulation of cell membrane localized NTCP expression in proliferating hepatocytes prevents hepatitis B virus infection. *Emerging microbes & infections*. 2019;8(1):879-894.
20. Meng G, Tan Y, Fan Y, et al. TSD: A Computational Tool To Study the Complex Structural Variants Using PacBio Targeted Sequencing Data. *G3 (Bethesda)*. 2019;9(5):1371-1376.
21. Ramirez R, van Buuren N, Gamelin L, et al. Targeted Long-Read Sequencing Reveals Comprehensive Architecture, Burden, and Transcriptional Signatures from Hepatitis B Virus-Associated Integrations and Translocations in Hepatocellular Carcinoma Cell Lines. *J Virol*. 2021;95(19):e0029921.
22. Kornberg RD. Chromatin structure: a repeating unit of histones and DNA. *Science (New York, NY)*. 1974;184(4139):868-871.
23. Tsai IK, Yen MY, Ho IC, Yu KW, Liu CY, Cheng DL. Clostridium perfringens septicemia with massive hemolysis. *Scand J Infect Dis*. 1989;21(4):467-471.
24. Zhou J, Tan T, Tian Y, et al. Kruppel-like factor 15 activates hepatitis B virus gene expression and replication. *Hepatology*. 2011;54(1):109-121.
25. Lauer U, Weiss L, Lipp M, Hofschneider PH, Kekule AS. The hepatitis B virus preS2/St transactivator utilizes AP-1 and other transcription factors for transactivation. *Hepatology*. 1994;19(1):23-31.
26. Lu CC, Yen TS. Activation of the hepatitis B virus S promoter by transcription factor NF-Y via a CCAAT element. *Virology*. 1996;225(2):387-394.
27. Kass SU, Pruss D, Wolffe AP. How does DNA methylation repress transcription? *Trends Genet*. 1997;13(11):444-449.
28. Liu Y, Cheng J, Siejka-Zielińska P, et al. Accurate targeted long-read DNA methylation and hydroxymethylation sequencing with TAPS. *Genome biology*. 2020;21(1):54.
29. Zhang Y, Li C, Zhang Y, et al. Comparative analysis of CpG islands among HBV genotypes. *PLoS one*. 2013;8(2):e56711.

30. Christman JK. 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. *Oncogene*. 2002;21(35):5483-5495.
31. Butler JS, Dent SY. The role of chromatin modifiers in normal and malignant hematopoiesis. *Blood*. 2013;121(16):3076-3084.
32. 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(5):975-984.
33. Doerfler W. In pursuit of the first recognized epigenetic signal—DNA methylation: a 1976 to 2008 synopsis. *Epigenetics*. 2008;3(3):125-133.
34. Zhang Y, Mao R, Yan R, et al. Transcription of hepatitis B virus covalently closed circular DNA is regulated by CpG methylation during chronic infection. *PLoS One*. 2014;9(10):e110442.
35. Watanabe Y, Yamamoto H, Oikawa R, et al. DNA methylation at hepatitis B viral integrants is associated with methylation at flanking human genomic sequences. *Genome Res*. 2015;25(3):328-337.
36. Flecken T, Meier MA, Skewes-Cox P, et al. Mapping the Heterogeneity of Histone Modifications on Hepatitis B Virus DNA Using Liver Needle Biopsies Obtained from Chronically Infected Patients. *Journal of virology*. 2019;93(9).

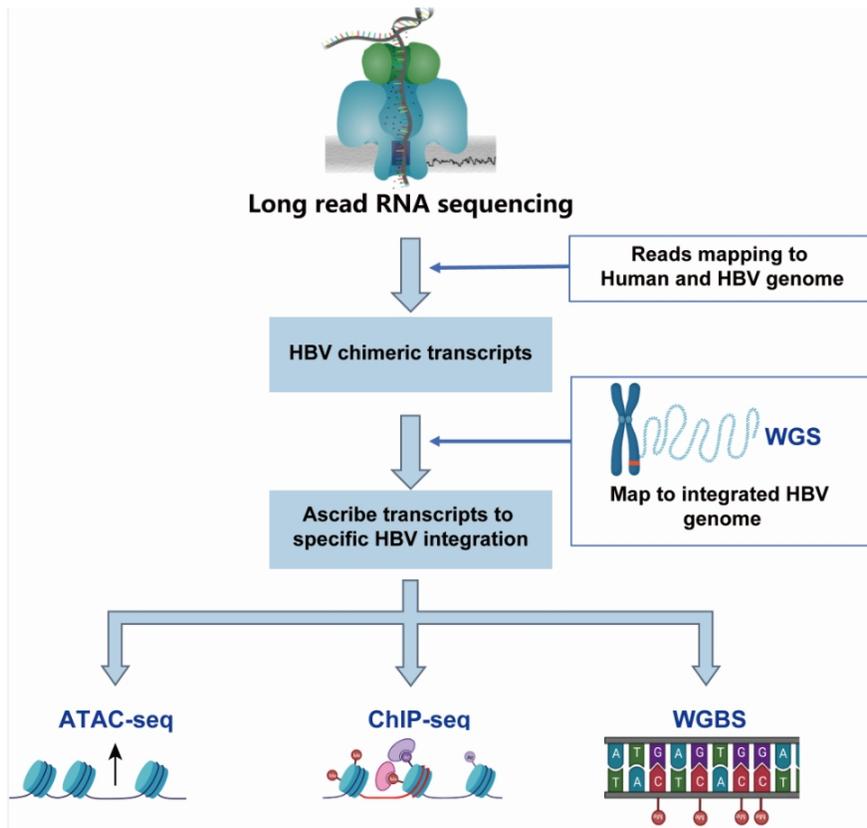


Figure 1: Analysis flow chart of HBV integration in PLC/PRF/5 cells.

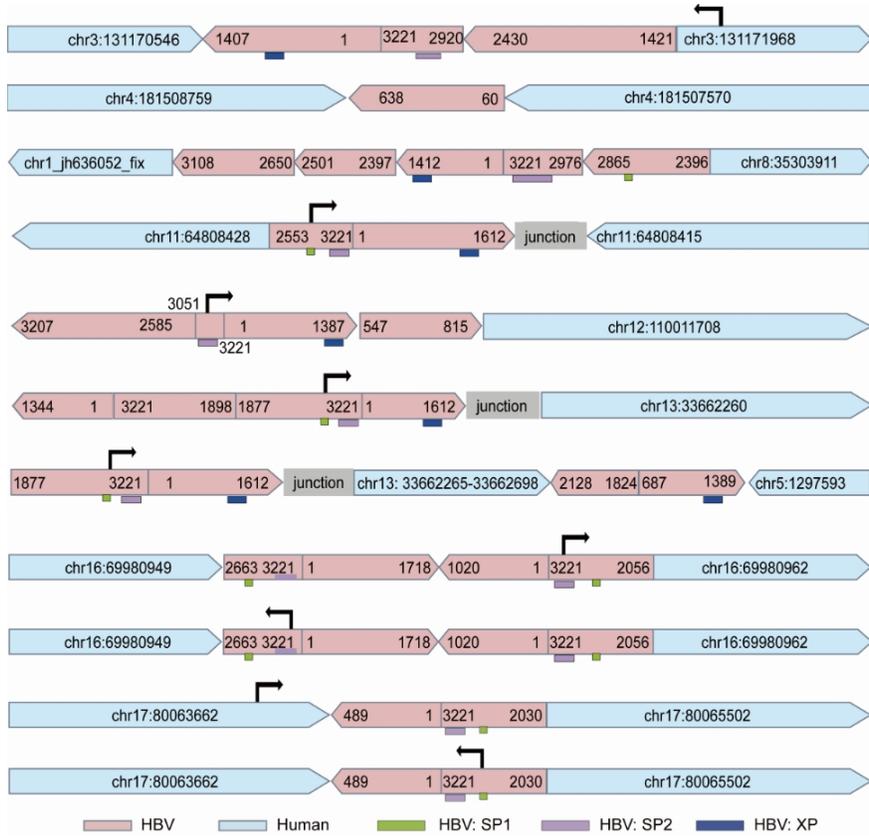


Figure 2: Transcriptional profiles of HBV integrations in PLC/PRF/5 cells. The light blue and pink colors representing the human and HBV genomes, respectively. The numbers depict the base position of the HBV genome. The black arrows indicate the transcription start site and direction, and the HBV promoters are represented by different colored boxes. The integrated HBV DNA in PLC/PRF/5 cells on chr11, chr13, and t (5;13) shares a nearly 230 bp homologous junction sequence of at the flanking host genome sequence. We labeled the homologous sequence as junction.

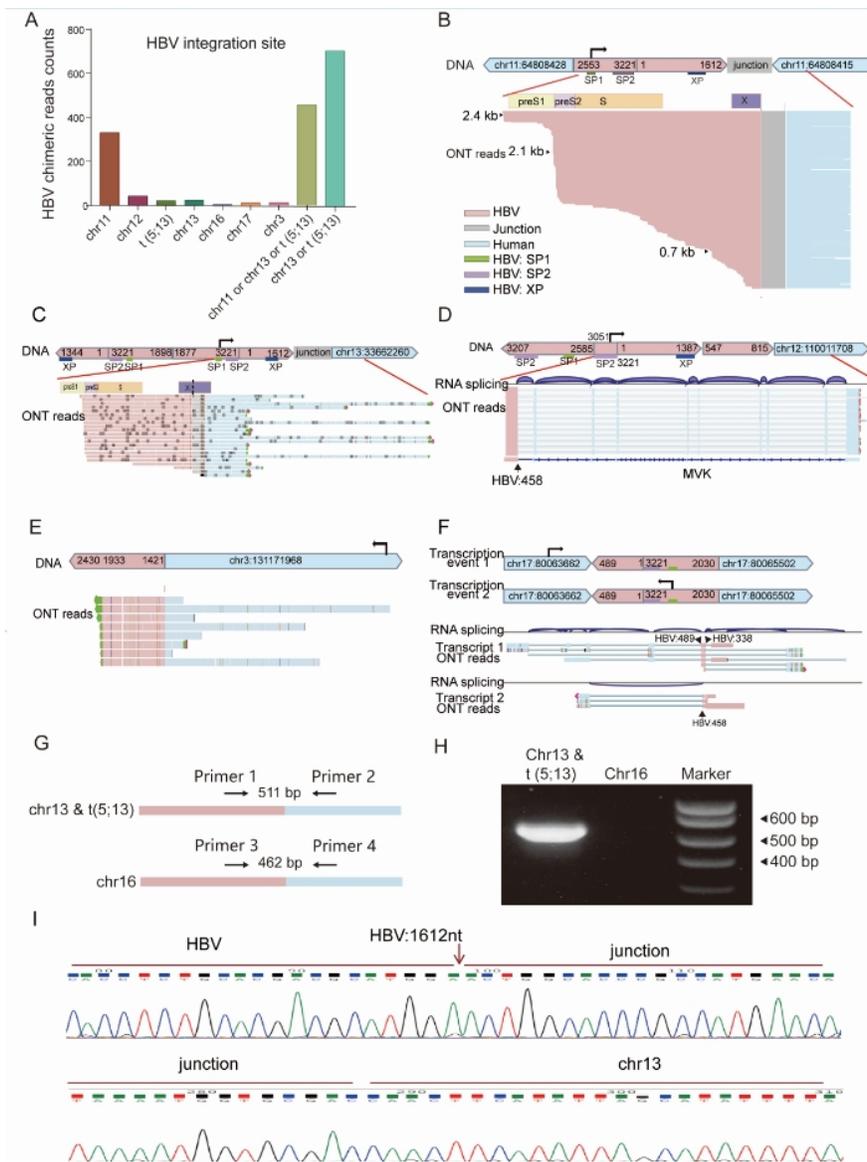


Figure 3: Transcriptional characteristics of integrated HBV DNA in PLC/PRF/5. (A) Quantification of chimeric transcripts derived from different integrated HBV DNA. The counts of chimeric HBV RNA reads originating from different integration sites was used to represent the transcription level of HBV DNA at different integration sites. (B) IGV snapshot of HBV DNA transcription integrated on chr11. Pink is the HBV sequence, light blue is the host sequence, and gray is the junction sequence. Integration of HBV transcription is usually accompanied by the addition of a poly(A) tail in humans, taking HBV integration in chromosome 11 as an example. (C) IGV snapshot of transcription patterns of HBV DNA on chr13. (D) IGV snapshot of HBV DNA transcription integrated on chr12. The pink color in ONT reads is the HBV sequence (3196~3221/1~458nt), the light blue is the MVK gene exon sequence, the green represents the polyA, and the dark blue is the RNA splicing information. (E) IGV snapshot of HBV DNA transcription integrated on chr3. (F) Schematic representation of two transcription patterns of integrated HBV DNA on chr17. (G) Primer pairs for PCR amplification of chr13/t(5;13) and chr16. (H) Gel electrophoresis image showing PCR products. (I) Sanger sequencing chromatogram of the HBV:1612nt junction.

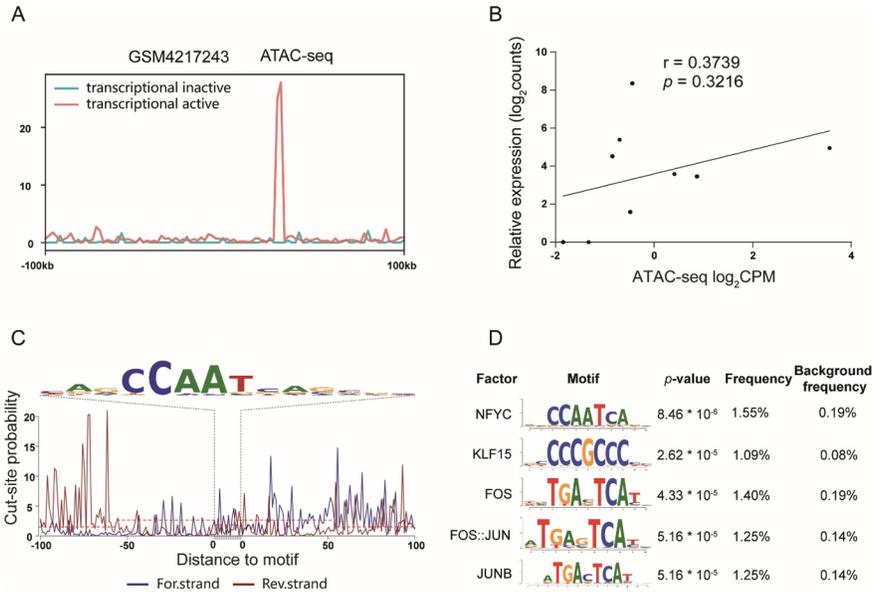


Figure 4: Analysis of Chromatin Accessibility Around HBV Integration Sites in PLC/PRF/5 Cells and Prediction of Related Transcription Factors. (A) The ATAC-seq signal density plots within 100 kb upstream and downstream of transcriptional active (green) and transcriptional inactive (blue) integration sites. **(B)** Dot plot of the ATAC-seq accessibility and integrated HBV DNA transcription level. **(C)** The NFYC transcription factor aggregates on chromosome 1 of PLC/PRF/5 cells and can impede Tn5 enzyme cleavage when it binds to DNA sequences, reducing the likelihood of cutting in regions marked by NFYC. **(D)** Predicted transcription factors that bind to the genome at 100 kb upstream and downstream near the HBV integration site.

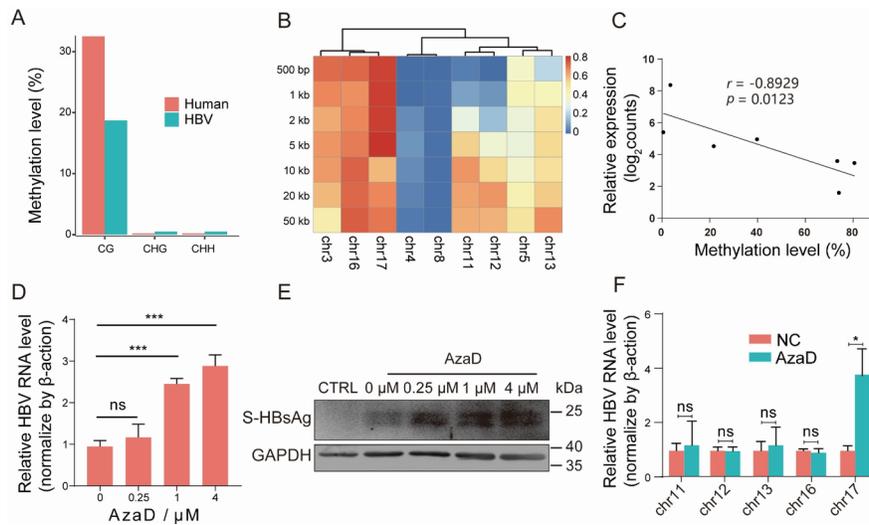


Figure 5: DNA methylation inhibits integrated HBV DNA transcription. (A) genome-wide methylation levels of integrated HBV DNA and host genome. CG, CHG and CHH represent the three forms of C nucleotide distribution on the genome (H represents A, T or C nucleotide). (B) Heat map of host genome methylation levels around integration sites in the PLC/PRF/5 cell line. (C) Correlation between host genome methylation levels within 500 bp upstream and downstream of HBV integration sites and the transcriptional level of integrated HBV DNA. (Methylation levels of genome around Integration sites on chromosome 4 and chromosome 8 with severely outlying were excluded from the correlation analysis.) The y-axis represents log₂-transformed HBV chimeric read counts from each integration site, indicating the transcriptional level of each integration site. (D) RT-qPCR was used to detect the relative expression of intracellular HBV RNA of PLC/PRF/5 cells before and after AzaD treatment using primer that targeting S region (329-451 nt) of HBV. The control group and experimental group were treated with DMSO and different concentrations of AzaD, respectively, for 48 h before harvesting. (E) Western Blot was used to detect the intracellular HBsAg level of PLC/PRF/5 cells before and after AzaD treatment. (F) Quantitative PCR detection of transcription levels at each HBV integration sites in PLC/PRF/5 cells before and after AzaD treatment using primer that targeting the specific sequence of each integrated sites (Table S1). The expression level of genes was normalized using β -actin. * Indicates $P < 0.05$, **** indicates $P < 0.0001$.

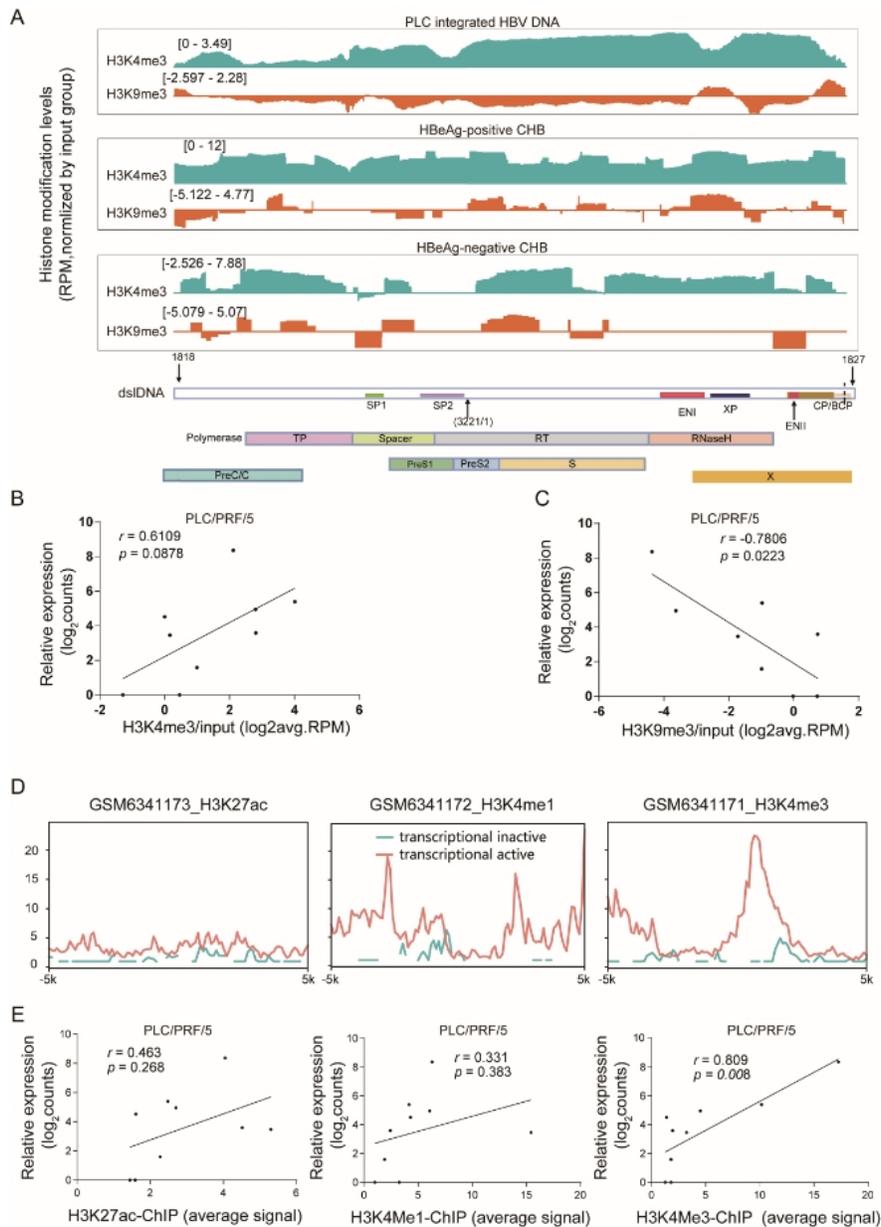


Figure 6: H3K4me3 and H3K9me3 deposition on integrated HBV-DNA correlates with viral transcription. (A) Distribution of H3K4me3 and H3K9me3 along the HBV genome in different samples. (B-C) Pearson correlation was used to evaluate the association between histone PTMs H3K4me3 and H3K9me3 on integrated HBV DNA and its transcription levels. The horizontal axis represents histone modification levels of each HBV integrated site, which were represented by taking the \log_2 ratio of ChIP-seq RPM/input RPM. The vertical axis represents the transcription levels of each integrated site, which were represented by taking the \log_2 of the count of HBV chimeric reads from each integrated site. (D) The green and blue lines represent the ChIP-seq average signal for histone PTM within 5kb upstream and downstream of the integration site, the with and without transcriptional activity, respectively. (E) Pearson correlation between average histone ChIP-seq signal of each integration site (5 kb upstream and downstream) and its corresponding transcriptional activity. The vertical axis represents the transcription levels of each integrated site, which were determined by computing the \log_2 value of the counts of reads from HBV chimeric reads from each respective integration site.