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

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

¹Peking University School of Basic Medical Sciences
²Guangdong Provincial People’s Hospital
³Capital Medical University Affiliated Beijing Friendship Hospital

October 9, 2023

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.

Methods: Transcriptome long-read sequencing (ONT) was conducted to analyze characterize the transcriptional activity of different HBV DNA integrations in PLC/PRF/5 cells. Additionally, data pertaining to epigenetic regulation such as whole-genome bisulfite sequencing (WGBS), histone chromatin immunoprecipitation sequencing (ChIP-seq), and assay for transposase-accessible chromatin using sequencing (ATAC-seq) were collected to investigate the potential mechanisms associated with the transcriptional regulation of integrated HBV DNA.

Result: Our findings indicate that transcriptional activity of integrated HBV DNA in PLC/PRF/5 cells is influenced by methylation levels of the surrounding host genome near the integration site. The result indicated that elevated methylation of the adjacent host genome adversely impacts transcription activity of integrated HBV DNA. Furthermore, we observed a positive association between histone modification H3K4me3 and the transcription of integrated HBV DNA. These results suggest that host may regulate transcriptional activity of integrated HBV DNA through DNA methylation and histone modifications. Potentially leading to the silencing of integrated HBV DNA.

Conclusion: Our study brought a better understanding on the transcriptional regulation of integrated HBV DNA. This knowledge can be valuable in the development of novel strategy for functional cure of CHB.
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.

Methods: Transcriptome long-read sequencing (ONT) was conducted to analyze characterize the transcriptional activity of different HBV DNA integrations in PLC/PRF/5 cells. Additionally, data pertaining to
epigenetic regulation such as whole-genome bisulfite sequencing (WGBS), histone chromatin immunoprecipitation sequencing (ChIP-seq), and assay for transposase-accessible chromatin using sequencing (ATAC-seq) were collected to investigate the potential mechanisms associated with the transcriptional regulation of integrated HBV DNA.

**Result:** Our findings indicate that transcriptional activity of integrated HBV DNA in PLC/PRF/5 cells is influenced by methylation levels of the surrounding host genome near the integration site. The result indicated that elevated methylation of the adjacent host genome adversely impacts transcription activity of integrated HBV DNA. Furthermore, we observed a positive association between histone modification H3K4me3 and the transcription of integrated HBV DNA. These results suggest that host may regulate transcriptional activity of integrated HBV DNA through DNA methylation and histone modifications. Potentially leading to the silencing of integrated HBV DNA.

**Conclusion:** Our study brought a better understanding on the transcriptional regulation of integrated HBV DNA. This knowledge can be valuable in the development of novel strategy for functional cure of CHB.

**Background**

Hepatitis B virus (HBV) infection is a significant global public health concern. According to the World Health Organization’s 2019 estimates, there were cases of 296 million chronic HBV infections and approximately 820,000 deaths caused by HBV-induced diseases annually\(^1\). 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 (dsDNA) into the human genome\(^2\). The presence of integrated HBV contributes to the instability of the chromosomal structure in patients and also 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 chronic hepatitis B (CHB) patients with low HBV load\(^5\). Previous research conducted by our team has indicated that approximately 72% of CHB patients who have attained a functional cure through interferon treatment continue to experience ongoing HBV RNA transcription in their liver, primarily originating from integrated HBV. Among them, 25% of patients still have ongoing HBsAg translation in the liver\(^6\).

Given these challenges mentioned, it is crucial to prioritize the research on developing drugs to silence or eliminate integrated HBV DNA. We successfully developed a gRNA-miRNA-gRNA Ternary Cassette Combining CRISPR/Cas9 system for targeted cutting of HBV cccDNA sequences and simultaneously silencing cccDNA transcription\(^7\). Furthermore, we have utilized exosome delivery systems to successfully deliver the CRISPR/Cas9 system and successfully cut HBV cccDNA both in vivo and in vitro\(^8\). However, compared to the structurally simple cccDNA, HBV DNA integration exhibits a more complex pattern. Therefore, developing drugs to clear or silence the transcription of integrated HBV DNA depends on models that consider these intricate integration patterns, whether in vitro or in vivo. 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 hepatocellular carcinoma (HCC) cell line integrated with HBV DNA\(^9\), possesses a large number of integration sites, making it the most classical model for studying HBV DNA integration. Furthermore, unlike the cccDNA in primary liver cells infected with HBV, PLC/PRF/5 cells do not exhibit HBV replication, allowing researchers to exclude interference from cccDNA. However, due to the complex integration patterns in PLC/PRF/5 cells and the lack of comprehensive analysis of these cells have significantly impeded progress in the development of relevant drugs and treatment plans.

Building on our past investigations, we analyzed the HBV DNA integration sites in the PLC/PRF/5 cell line using long-read DNA sequencing technology and successfully cut them using the CRISPR-Cas9 system, thereby reducing the translation of HBsAg from integrated HBV DNA\(^10\). In this study, we aim to investigate the transcriptional activity of each integrated HBV DNA, the translation status of viral proteins, as well as the epigenetic modifications and transcriptional regulation at each integration site in PLC/PRF/5.
cells. We intend to accomplish this by combining DNA long-read sequencing, RNA long-read sequencing, 
ribosome profiling sequencing (Ribo-seq), whole-genome bisulfite sequencing (WGBS), histone chromatin 
immunoprecipitation sequencing (ChIP-seq), and assay for transposase-accessible chromatin using sequencing 
(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

RNA extraction and cDNA preparation

Total RNA was extracted from the tissue using TRIzol reagent (Takara, Kyoto, Japan). RNA purity was 
tested using the Nano Photometer spectrophotometer (IMPLEN, Westlake Village, USA). cDNA libraries 
were constructed from 1 μg of total RNA 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 literature\cite{11}. 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)\cite{12}, and consensus isoforms were generated 
after polishing within each cluster by pinfish. 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 1.33.0 \cite{13}, 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 2.11.4\cite{14}.

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)\cite{15} 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 views (V0.1.11)\cite{16} 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 views was used to visualize the methylation levels of the integrated 
HBV genome. The GlobalMethLev function in views 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)\cite{17} (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 ATACseqQC\textsuperscript{[18]}, and the factorFootprints function was used to detect and visualize the signal changes around the motif regions where transcription factors bind. MACS2\textsuperscript{[19]} 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 deeptools\textsuperscript{[20]}. 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\textsubscript{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.

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% CO\textsubscript{2}.

Real-time Reverse Transcription (RT)-PCR

Real-time RT-PCR was performed as described previously\textsuperscript{[21]}. β-Actin was used as the reference gene to determine gene expression. The primers used for real-time RT-PCR are listed in Supplemental Table 1.

Western blot analysis

Western blot analysis was performed as previously described\textsuperscript{[21]}. 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 carcinoma. Additionally, the availability of various high-throughput sequencing technologies and publicly accessible data in PLC/PRF/5 cells facilitate combine analysis of multi-omics data when investigating the transcriptional regulation mechanism of HBV DNA integration. Therefore, we initially collected a total of 208 high-throughput sequencing datasets associated with PLC/PRF/5 cells from GEO and SRA databases.

The expression of integrated HBV DNA follows the central dogma of molecular biology, indicating that integrated HBV DNA is transcribed to HBV RNA and subsequently translated to viral or virus-host fusion proteins. Consistent with this genetic central dogma, our analysis strategy involved utilizing whole-genome, transcriptome and translationomics sequencing technology to examine the transcriptional activity and characteristics of various HBV DNA integration events. Our previous work has identified the multiple HBV integration in PLC/PRF/5 cells using long-read whole-genome sequencing technology (PacBio)\[10\]. Furthermore, we conducted transcriptome long read sequencing (ONT) to identify transcriptional characteristic of integrated HBV DNA in PLC/PRF/5 cells. Next, we utilized Ribo-seq data to investigate the translation characteristics of HBV transcripts derived from integrated HBV in PLC/PRF/5 cells. Based on transcriptomic analysis results, we further employed epigenetic-related sequencing data, including whole-genome bisulfite sequencing (WGBS), histone chromatin immunoprecipitation sequencing (ChIP-seq), and assay for transposase-accessible chromatin sequencing (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 is illustrated in Fig. 1. The datasets used in this study are summarized in Table S1.

Resolution of integrated HBV DNA transcriptional architecture.

Previous researches have reported complex HBV DNA integration in PLC/PRF/5 cells\[10,22\]. However, 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 these cells. The emergence of long-read sequencing technology makes it possible to generate long sequence containing full-length HBV and flanking host regions within a single read, enabling the acquisition of comprehensive full-length HBV transcripts\[11\]. Therefore, we employed third-generation long read transcriptome sequencing (Oxford Nanopore Technology, ONT) to sequence the transcripts from PLC/PRF/5 cells and investigate 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) 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, chr13, and t (5;13) (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.

Quantification 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. Fig. 3A illustrates 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), 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.

Our subsequent analysis will 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 transcriptional terminal signals, generating PreS1, PreS2/S and HBx transcripts (Fig. 3B). While the HBV integrated fragment contains multiple promoters, the transcription of integrated HBV exhibits promoter selectivity (Fig. S2C). In addition, the RNA originating from the integrated HBV DNA further exhibited splicing properties, forming fusion transcripts with host exons (Fig. 3C). Furthermore, we observed that certain integrated HBV DNA initiate transcription with the host promoter, and transcription ceased via HBV canonical polyA (Fig 3D). Besides, multiple transcriptional modes can arise from the same integrated HBV DNA (Fig. S2A, B). 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 S2E, F).

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 S3A. 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 S3B). 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 consist with ONT long-read sequencing (Fig S3C), further validating the accuracy of the results of PLC/PRF/5 cell transcriptome long-read sequencing.

In summary, our findings demonstrate that integrated HBV DNA on chr11, chr12, chr13, chr16, chr17, and t (5;13) retains complete or partial SP1 and SP2 promoters, enabling initiate transcription. In contrast, integrated HBV DNA on chr4 cannot initiate transcription due to the absence of HBV promoters. This implies that 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, in addition to factors intrinsic to HBV, other host factors may also be involved in the transcriptional regulation of integrated HBV DNA. Consequently, we focus on to explore the mechanism of host factors on the transcriptional regulation of integrated HBV DNA in PLC/PRF/5 cells.

Translation characteristic of chimeric transcripts derived from integrated HBV in PLC/PRF/5 cells.

Proteins are essential components for executing gene functions. Ribo-seq technology, which involves deep sequencing, can identify translated regions, providing comprehensive insights into the viral proteome. To investigate the translation characteristics of transcripts originated from integrated HBV DNA in PLC/PRF/5 cells, we obtained Ribo-seq data (GSM4446981) from these cells [23]. Firstly, we assessed the lengths of ribosome-protected fragments (RPFs). As depicted in Fig S4A, the result revealed a peak in the number of RPFs with lengths of 29-30 nt, consistent with previous reports [24]. The RPF with length of 28 nt, 29 nt, and 30 nt, exhibited a prominent 3-nt periodicity signature within the coding sequence (CDS) (Fig. S4C). Furthermore, we employed the bioinformatics software Ribocode [25] to identify translating ORFs in PLC/PRF/5 cells. As shown in Figure 4.26, we identified a total of 40,106 ORFs, comprising 27,840 annotated human ORFs and 12,266 unannotated ORFs. Among the unannotated ORFs, there were 8,634 novel ORFs which include the HBV surface antigen HBsAg (Fig. 3E). We performed a Protein Blast on the NCBI website for the amino acid sequences of ORFs identified by Ribocode and found that they aligned with the amino acid sequence of the HBV genotype A L-HBsAg (Fig. S4D). Additionally, we performed Western Blot experiments to detect three type HBsAg, including L-HBsAg (Fig. S4E).

Analysis of chromatin accessibility around HBV integration sites in PLC/PRF/5 cells and prediction of related transcription factors
In cell nuclei, DNA is highly compacted, wrapped around nucleosomes to form chromatin, which then forms chromosomes\(^ {[26]} \). The compaction of nucleosomes affects the accessibility of large biomolecules, such as transcription factors and RNA polymerases, that need to bind to corresponding DNA sequences to function during gene transcription. Chromatin accessibility refers to the ease with which large biomolecules can approach chromatin and can undergo dynamic changes\(^ {[27]} \). This accessibility is an essential mechanism for cells to regulate gene expression.

In theory, after HBV DNA fragments integrated into the host chromosome, their transcriptional activity may be affected by the around host genome’s chromatin accessibility. 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 100kb upstream and downstream). 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 (\textbf{Fig. 4A}). Furthermore, we used the \(\log_2\)CPM 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 (\(r = 0.3739, p = 0.3216\)) (\textbf{Fig. 4B}). Therefore, the role of chromatin accessibility in regulating integrated HBV DNA transcription is limited. The lack of statistical difference in the correlation analysis may be related to the small number of integration sites.

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 footprints. For instance, in the case of the transcription factor NFYC, Tn5 cleavage in these footprints decreases significantly within peak regions of high cleavage probability (\textbf{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 100kb 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\textbf{Fig. 4D}, NFYC, KLF15, FOS, JUN, and other transcription factors are enriched around the HBV integration site compared to randomly selected background regions. This indicates that these enriched transcription factors can bind to promoters in the open chromatin regions around the HBV integration site. In addition, previous studies have reported that the aforementioned transcription factors can bind to the HBV cccDNA promoter sequence and initiate HBV gene transcription\(^ {28-30} \). Based on these findings, 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.

In summary, our findings suggest that chromatin accessibility around HBV integration sites in PLC/PRF/5 cells may play a limited role in the transcriptional regulation of integrated HBV DNA. However, the enrichment of specific transcription factors such as NFYC, KLF15, FOS, and JUN near these integration sites indicates that these factors may be involved in the transcriptional regulation of integrated HBV DNA by binding to HBV promoters. Future studies should investigate the role of these transcription factors in the transcriptional regulation of integrated HBV DNA, as well as the impact of host factors on the transcriptional activity of integrated HBV DNA in PLC/PRF/5 cells.

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

We first compared the average methylation levels of integrated HBV DNA and the human genome in PLC/PRF/5 cells. As \textbf{Fig. 5D} shows that the average methylation level of integrated HBV DNA was lower than that of the human genome (HBV 19.7% vs. human 32.9%), consistent with the remained transcriptional activity of integrated HBV DNA. 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)[31] 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 DNA were higher than those on the corresponding CpG islands of cccDNA (Fig. 5A, 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.

Previous studies have found that the methylation level of integrated HBV DNA is strongly correlated with the methylation level of the adjacent host genome[32], suggesting that the transcriptional activity of integrated HBV DNA may be influenced by the methylation level of its neighboring region. Therefore, we first analyzed the methylation characteristics of the adjacent regions of the HBV integration sites in PLC/PRF/5 cells. As shown in Fig. 5E, 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. This may be because the hypomethylated HBV genome itself affects the methylation level of the host genome around the integration site, resulting in a low methylation state. On the other hand, it cannot be ruled out that HBV dsDNA is more likely to integrate into hypomethylated host genome sequences with higher transcriptional activity. 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. 5F, the transcription level of integrated HBV DNA was strongly negatively correlated with the methylation level \( r = -0.8929, p = 0.0123 \), indicating that a hypermethylation of the host genome around the integration site is not conducive to the transcription of integrated HBV DNA. Consistent with this, we also found that, host genome around the HBV integration sites on chr5, chr11, chr12, and chr13, was relatively hypomethylated (Fig. 5E). Correspondingly, these integration sites of HBV DNA had high transcriptional activity, whereas the methylation levels around the integration sites on chr3, chr16, and chr17 were relatively higher (Fig. 5E), 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.

To verify the effect of DNA methylation on the transcription of integrated HBV DNA, PLC/PRF/5 cells were treated with the DNA methylation inhibitor AzaD. After 48 hours of treatment, 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 \( \mu M \) and 4 \( \mu M \) (Fig. 5G). Consistently with this result, the Western blot results also confirmed that AzaD treatment upregulated the intracellular HBsAg protein levels in a dose-dependent manner (Fig. 5H). In addition, we observed that HBsAg levels in the cell culture supernatant of AzaD-treated PLC/PRF/5 cells also showed an increasing trend with increasing drug concentration, especially at a treatment concentration of 4 \( \mu M \), where the HBsAg level was significantly upregulated \( t = 3.507, p = 0.0247 \) (Fig. 5I). However, the changes in HBsAg levels in the cell culture supernatant were not as apparent as those in intracellular protein levels, which may be related to the low secretion efficiency of HBsAg derived from integration. These results further confirm that the transcription of integrated HBV in PLC/PRF/5 cells is regulated by methylation.

The results demonstrate a negative correlation between the methylation level of the host genome around the integration site and the transcription level of integrated HBV DNA. We also confirmed methylation modification affects the transcriptional expression of integrated HBV DNA by the cellular experiment. However, we observed that the host genome around the integration of HBV DNA on chr4 and chr8 with no transcriptional activity was hypomethylated. The integrated HBV DNA sequence on chr4 has no promoter, which can explain the lack of transcription. However, the integrated HBV DNA on chr8 retains the complete SP1, SP2, and XP, and surrounded by a hypomethylated host genome, contradicts the expected transcriptional activity. These findings suggest that, apart from host methylation, other host factors may also regulate integrated HBV DNA.
Transcription of integrated HBV DNA is associated with histone modifications in PLC/PRF/5 cells

In eukaryotic cells, DNA resides within the nucleus as part of chromatin. Nucleosomes, the fundamental units of chromatin, consist of 146 to 147 base pairs of DNAs wrapped around a histone octamer, which is composed of an H2A-H2B tetramer and two H3-H4 dimers. The interaction between histone N- and C-termini and DNA can be altered by post-translational modifications (PTMs), which in turn can influence the state of chromatin and gene expression. Previous studies have demonstrated that HBV cccDNA associate with host cell histones into episomal chromatin, with cccDNA transcription being regulated by host histone modifications. Once HBV DNA integrates into the host genome, its transcription may be influenced by histone modifications. Consequently, we used PLC/PRF/5 histone ChIP-seq data from the GEO database, along with identified HBV DNA integration sites and transcription levels, to investigate the relationship between histone modification levels and transcriptional activity at each integration site.

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 transcription. Flecken, Tobias, et al. conducted a histone ChIP-seq study on CHB patient’s liver biopsy samples and PLC/PRF/5 cells. Although they investigated the correlation between histone PTM deposition on HBV cccDNA and intrahepatic HBV RNA levels in liver biopsy samples, they did not focus on the association between histone PTM deposition on integrated HBV DNA and transcription. In order to comprehend the distinct deposition of histone 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). 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. In summary, 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.

Furthermore, 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. Analyzing the correlation between HBV DNA transcription levels and histone PTM levels at each integration site (Fig. 6B,C) revealed 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) These findings suggest that histone modifications may play a crucial role in regulating integrated HBV DNA transcription.

In the previous sections, we examined the enrichment of histone PTMs on integrated HBV DNA sequences. Next, we analyzed the histone modification levels of the host genome within 5kb 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 results provide further evidence suggesting that histone PTM H3K4me3 may promote the active transcription of integrated HBV DNA.

**Discussion**

Prior studies have shown that integrated HBV DNA, aside from causing insertion mutations and chromosomal instability in the host while disrupting gene expression in the integration vicinity, has the potential to express HBsAg and truncated HBx protein, ultimately triggering hepatocellular carcinoma (HCC). Clinical studies demonstrate that, despite eliminating or rendering cccDNA transcriptionally silent, integrated HBV DNA is still capable of expressing HBsAg persistently, hampering the achievement of a functional cure in CHB patients[3]. 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). This suggests that host factors may influence the transcriptional activity of integrated HBV DNA beyond viral promoter sequences.

In infected liver cells, HBV cccDNA combines with histones and non-histones to form mini-chromosomes. Its transcriptional activity is regulated by host epigenetic changes such as DNA methylation, histone modification, and chromatin remodeling. However, the current research on the transcriptional regulation mechanism of integrated HBV DNA is limited. Thus, we further investigated the effect of epigenetic modifications on the transcriptional regulation of integrated HBV by combining multi-omics data.

According to previous research, HBV is more likely to integrate into transcriptionally active regions[39], 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 in this study, we discovered that the chromatin around transcriptionally active integration sites were more accessible than that around integration sites without transcription. Furthermore, some transcription factors were predicted that may function around HBV integration sites by ATAC-seq data. Most of the enriched transcription factors, which have been reported to be involved in HBV replication, still need further experimental verification to validate their actual functions on the transcription of integrated HBV.

Genomic DNA methylation is a crucial epigenetics modification that regulates gene transcription. In this study, we compared the average methylation levels of integrated HBV DNA, host genome in PLC/PRF/5, and nuclear HBV cccDNA from HepG2-NTCP cells. The methylation level of integrated HBV DNA was higher than that of cccDNA but lower than that of the host genome. These results suggest that unmethylated HBV genome is methylated to inhibit transcription of integrated HBV DNA after it is integrated into the host genome. We found a significant negative correlation between methylation of host genome around integration sites and HBV DNA transcription levels. This indicates that, HBV DNA which integrated on highly methylated host regions may be transcriptionally silenced due to methylated after it is integrated into the host genome. Furthermore, we observed that the methylation level of host genomic regions approximately 500bp to 1kb away from HBV integration sites was lower than that of the distal region. Similar findings were reported in a recent study[40].

Histone PTMs is another epigenetic modification which plays an important role in regulating gene transcription, occurring in both cccDNA and the host genome. We compared the differences in the deposition
of histone PTMs between on cccDNA from liver specimens of CHB patients and on integrated HBV DNA from PLC/PRF/5 cells. Our results showed 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 histone PTM H3K4me3 may promote transcription. We also found that enrichment level of H3K4me3 on integrated HBV DNA and flanking host genome were positively correlated with the transcriptional levels of HBV DNA on each integration sites. Tropberger and Flecken's research further supported these findings by highlighting the positive correlation between active histone PTM H3K4me3 and HBV DNA transcription[38], indicating that the histone PTM H3K4me3 may promote both integrated HBV DNA and cccDNA transcription. Additionally, we observed a highly enrichment of inhibitory histone PTM H3K9me3 in 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.

Through comprehensive analysis, we explored the integration, transcriptional characteristics, and potential regulatory mechanisms of integrated HBV DNA in PLC/PRF/5 cell line. The findings revealed a moderate correlation between the transcriptional activity of integrated HBV DNA and the DNA methylation and histone modification levels of HBV DNA and flanking host genome regions. 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 Supplementary Table S1.

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, XD and XC, and all authors commented on.

Conflict of Interest statement
The authors declare that they have no competing interests

REFERENCES

12


Figure legend

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

Fig. 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.

Fig. 3 Transcriptional and translational 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. (C) 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. (D) IGV snapshot of HBV DNA transcription integrated on chr3. (E) PLC/PRF/5 cell RPF density at HBsAg ORF.

Fig. 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 100kb 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.

Fig. 5 DNA methylation inhibits integrated HBV DNA transcription

(A) Distribution of HBV CpG islands. (B) Methylation levels of integrated HBV DNA in PLC/PRF/5 cells. (C) Methylation levels of HBV cccDNA from HepG2-NTCP cells. The x-axis represents the position of HBV, and the y-axis represents the percentage of methylated cytosine (C) bases. (D) 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). (E) Heat map of host genome methylation levels around integration sites in the PLC/PRF/5 cell line. (F) Correlation between host genome methylation levels within 500 bp upstream and downstream of HBV integration sites and the transcriptional level of integrated HBV DNA. (G) RT-qPCR was used to detect the relative expression of intracellular HBs RNA using β-actin as the reference gene. * Indicates P < 0.05, **** indicates P < 0.0001.

Fig. 6 H3K4me3 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, 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 log2 ratio of ChIP RPM/input RPM. The vertical axis represents the transcriptional level of each integration site, which were represented by taking the log2 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 (5kb 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 log₂ value of the counts of reads from HBV chimeric reads from each respective integration site.

Long read RNA sequencing

- Reads mapping to Human and HBV genome
- Map to integrated HBV genome
- HBV chimeric transcripts
- Ascribe transcripts to specific HBV integration

ATAC-seq + ChIP-seq + WGS + Ribo-seq + Western blot

mechanisms of integrated HBV DNA transcriptional regulation

translation characteristics of transcripts