Received: 24 December 2024 | Revised: 28 May 2025 | Accepted: 12 June 2025 | Published online: 25 June 2025

RESEARCH ARTICLE

Medinformatics
2025, Vol. 2(4) 309–317
DOI: 10.47852/bonviewMEDIN52025080

BON VIEW PUBLISHING

HBV Integration into Host Superenhancers Might Not Be a Risk Factor for Hepatocellular Carcinoma

Mengna Zhang[1,2], Ying Ming[1,*], Yunling Du[1], Ziwen Guo[1], Ziyuan Xin[1], Yanjun Li[1], Ge Yang[1] and Zhaoyang Jiang[1]

[1]*Molecular Diagnosis Center, The Affiliated Hospital of Chengde Medical University, China*

[2]*Hebei Key Laboratory of Panvascular Diseases, China*

Abstract: Hepatitis B virus (HBV) DNA integration into the host genome can be found in all phases of chronic HBV infection and plays a role in hepatocarcinogenesis. Integration into CpG islands has been reported to be a risk factor for hepatocellular carcinoma (HCC), but the relationship between HBV integration into other regulatory elements and HCC remains unclear. Superenhancers (SEs) contribute to the cancer cell state by regulating oncogenes. This study aimed to analyze whether integration into host SEs is a risk factor for HCC. We systematically annotated 21,520 HBV integration sites in the human genome obtained from the ViMIC database to determine their distribution in regular elements, including SEs, CpG islands, CCCTC-binding factors (CTCFs)-binding sites, transcription factor-binding sites (TFBSs), and transcription start sites (TSSs). Then, we constructed a logistic regression model to evaluate the relationship between the integration sites and HCC. Integration into CpG islands and TFBS were risk factors for HCC ($P = 0.000$, OR = 2.65 and $P = 0.041$, OR = 1.06, respectively. OR = odds ratio), and integration into SEs and CTCF-binding sites were significantly associated with nontumor tissues ($P = 0.000$, OR = 0.58 and $P = 0.000$, OR = 0.48, respectively). To further investigate the underlying mechanism, we analyzed CpG methylation and histone modifications at flanking integration sites. We found that regions flanking HBV integration sites in SEs were more likely to be hypermethylated ($P = 0.000$). Moreover, the hypermethylated SE regions flanking HBV integration sites showed lower levels of epigenomic markers. Our results suggested that integration into SEs might not be a risk factor for HCC. The protective effect observed for integration into host SEs might be associated with hypermethylation.

Keywords: hepatitis B virus, virus integration, superenhancers, hepatocellular carcinoma

1. Introduction

Chronic infection with hepatitis B virus (HBV) is an important risk factor for the development of hepatocellular carcinoma (HCC) [1]. HCC can be promoted by infection through direct or indirect mechanisms, such as epigenetic remodeling of the host genome, abnormal expression of oncogenes, tumor suppressor genes, or genes related to pathways involved in the regulation of hepatocellular metabolism and viral infection-induced chronic inflammation [2]. Inefficient immune reactions may be the primary oncogenic factor for chronic HBV infection [3]. PD-L1-induced depletion of CD8+ T cells, deficiency of CD4+ CTLs, and alterations in NK cell functions have been reported in patients with HCC [4].

The integration of HBV DNA into the host genome plays an important role in the occurrence of HCC. It can be found in all phases of chronic HBV infection [5]. Approximately 90% of HBV-associated HCCs harbor HBV DNA integrations; thus, virus-host chimera DNA can serve as a biomarker for HBV-related HCC [6]. Generally, the integration sites are located randomly throughout the host genome, but recurrent regions have been reported. TERT and MLL4/KMT2B are the most commonly reported target genes of HBV integration in HCC tissues. Other commonly reported genes included N4BP1, WASHP, and PLEKHG4B. In each study, hundreds of integrated genes were identified [7–11]. These integration target genes are enriched in pathways including axon guidance, positive regulation of Ras protein signal transduction, axonogenesis, and transmitter-gated ion channel activity (not all of them) [7]. More than 10,000 unique HBV integration sites in the human genome have been reported, and approximately 20% of these sites are recurrent [12]. However, integration events are more common in adjacent liver tissues than in HCC tissues [13]. These dispersed locations across host chromosomes and various involved pathways indicate that host genome disruption caused by HBV integration cannot fully explain the mechanisms of HBV-associated HCC. Integrated HBV DNA is also a stable source of viral RNA and proteins, though it does not lead to replication-competent transcripts. Patients with stable hepatitis B surface antigen levels derived predominantly from integrated viral DNA do not respond to nucleos(t)ide analogue therapy [14].

The mechanisms of HBV DNA integration are associated with microhomology between the viral and host genomes [15]. At the 3-dimensional scale, HBV DNA is not randomly positioned in the host genome but instead preferentially establishes contacts with host DNA at active chromatin regions. HBV DNA-host DNA

*Corresponding author: Ying Ming, Molecular Diagnosis Center, The Affiliated Hospital of Chengde Medical University, China. Email: mingying1310@sina.com

contacts are significantly enriched at H3K4me1-marked regions. Chromatin loops formed by integrated HBV DNA and host DNA are also found in transcriptionally active regions [16]. Thus, the integration of HBV DNA into active regulators in the host genome is not rare. Zhao et al. reported that HBV is prone to integrate into rare fragile sites and functional genomic regions, including CpG islands, and that integration into CpG islands is significantly increased in tumors [10]. CpG islands are exceptionally hypomethylated, and many of these hypomethylated regions of DNA function as elements that regulate gene expression, such as promoters and enhancers [17].

Superenhancers (SEs) are large clusters of enhancer elements identified by algorithms such as rank ordering of superenhancers [18, 19] from chromatin immunoprecipitation and sequencing (Chip-seq) data. SEs are enriched in chromatin factors such as cohesin, the histone modification H3K27ac, and the dimethylation of histone H3 at lysine 4 (H3K4me2) and H3K4me1 [20]. SEs play key roles in human cell identity [21] and contribute to the cancer cell state by regulating oncogenes [22]. SEs tend to lie on the surfaces of the 3D chromatin model [23] and are thus presumed targets of viral DNA integration. Viral DNA integration into host SEs is observed in HIV [24] and HPV [25] infections. HPV integration can hijack and multimerize cellular enhancers to generate a viral-cellular SE that drives high viral oncogene expression [26]. However, the frequency and clinical implications of HBV DNA integration into SEs have yet to be fully assessed.

In this study, we systematically annotated HBV integration sites in the human genome to determine their distribution in regular elements and constructed a logistic regression model to evaluate the relationships between integration sites and HCC. Interestingly, integration into SEs was a common event, as we suspected, but was significantly associated with nontumor tissues. Additionally, we investigated the potential epigenomic mechanisms underlying this finding.

2. Materials and Methods

2.1. HBV DNA integration sites and host genome annotation sources

ViMIC is a pilot database of human disease-related virus mutations, integration sites, and cis-effects (v1, released 21-MAY-21) [27]. In this database, the authors identified 39,687 HBV integration sites with host genome locations (GRCh38) and tissue types (http://bmtongji.cn/ViMIC/downloaddata/integration/HBV_integration.csv). We only chose sites without host genome deletions (end location – start location=1) from tumorous/nontumorous tissues. Recurrent integration sites were considered multiple records in this investigation, resulting in a total of 21,520 records, 16,644 of which were unique sites.

Then, we annotated these integration sites with host genome regulatory elements. CpG islands, CTCFs-binding sites, TFBSs, and transcription start sites (TSSs) were obtained from the UCSC Genome Browser (www.genome.ucsc.edu) [28]. Liver tissue SE ranges were obtained from SEdb v2.0 (https://bio.liclab.net/sedb/index.php) [29]. All cell types and tissues from the liver were included in the study.

2.2. Methylation annotation

We computed the mean methylated CpG rate of the flanking regions spanning 100 bp around the integration sites. The whole-genome CpG methylation data for the HepG2 cell line were obtained from the ENCODE project (https://www.encodeproject.org/experiments/ENCSR786DCL/) [30]. The methylation rates of the integration site flanking regions were calculated as described below. CpGs with a minimum coverage of 1 read were included.

For a single CpG, a methylated CpG was considered when more than 40% of the reads were methylated.

The region methylation rate was calculated as (methylated CpG/all CpGs in the region) *100%.

To validate our methylation calculation method, we downloaded methylation panel data for fetal livers from the GEO database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM1014211) and compared the methylation rates calculated as above with the panel results.

2.3. Histone modification of the HepG2 cell line

Chip-seq data were downloaded from the ENCODE project (https://www.encodeproject.org/). Chip-seq data of H3K4me1, H3K4me3, and H3K27ac in HepG2 cells and homo sapiens liver tissue of a female adult (25 years old) were collected. The ENCODE4 histone Chip-seq data standards and processing pipeline was available at https://www.encodeproject.org/chip-seq/histone-encode4/. A brief Chip-seq pipeline could be found at https://www.encodeproject.org/documents/6f6351d4-9310-4a3b-a3c2-70ecac47b28b/@@download/attachment/ChIP-seq_Mapping_Pipeline_Overview.pdf and https://www.encodeproject.org/documents/7009beb8-340b-4e71-b9db-53bb020c7fe2/@@download/attachment/ChIP-seq_pipeline_overview.pdf. We downloaded bigwig files aligned to hg38 with fold change over control. These files contained fold-over control of read depth at each position. Then we used deepTools v3.5.4 (https://github.com/deeptools/deepTools) [31] for Chip-seq data calculation and plotting with parameters reference-point – referencePoint center – *beforeRegionStartLength 10000 – afterRegionStartLength 10000 – skipZeros*.

2.4. Data statistics

We annotated HBV integration sites with host regulatory elements, including SEs, CpGs, CTCFs, TFBSs, and TSSs. Logistic regression was used to assess the associations between host-integrated regulatory elements and HCC. We used a paired Wilcoxon test to compare the methylation rate of 100bp flanking integration in a SE and the mean methylation rate of the same SE. $P < 0.05$ was considered to indicate statistical significance. R v4.2.2 was used for statistical analysis. Default parameters were used for logistic regression and paired Wilcoxon test. Logistic regression and Wilcoxon test functions were in the R base, and R package ggplot2 was used for plotting.

3. Results

3.1. Integration of SEs was significantly associated with HCC

Using this logistic regression model, we investigated the relationships between HCC and integrated host regulatory elements, including SEs, CpG islands, CTCF-binding sites, TFBSs, and TSSs (Table 1). Integration into CpG islands was associated with 634 events and a significantly increased tumor risk, with an odds ratio (OR) of 2.651. An OR >1 indicates an increased risk, and an OR <1 indicates a decreased risk. OR = 1 indicates no effect. Integration into SEs was more frequent than that into CpG islands, with a total of 2,807 events (13.0%), but significantly decreased tumor risk, with an OR = 0.580. The

Table 1. Relationships between HBV integration into host regulatory elements and HCC

Features	Tumor	Nontumor	P value	OR	95%CI
Total	13016	8504	–	–	–
CpG islands	512	122	0.000	2.651	[1.542,3.760]
SE	1375	1432	0.000	0.580	[−0.462,1.623]
CTCF	49	63	0.000	0.479	[−0.735,1.693]
TFBS	4741	3033	0.041	1.064	[0.033,2.095]
TSS	12	5	0.414	1.554	[−0.161,3.269]

Note: SE: superenhancer; CTCF: CCCTC-binding factor; TFBS: transcription factor-binding site; TSS: transcription start site; OR: odds ratio; 95%CI: 95% confidence interval.

proportion of SEs-integrations found in tumor tissues was 10.6%, while in nontumor tissues, this proportion was 16.8%. Integration into CTCF-binding sites was also showed to be a protective factor against HCC ($P = 0.000$, OR = 0.479); however, integration into CTCF-binding sites showed to be relativity rare events (0.5%). The most integrated host regulatory element was the TFBS, with 7,774 events as a mild tumor risk factor ($P = 0.041$ and OR = 1.064).

3.2. In the host genome, SE regions flanking HBV integration sites are hypermethylated

As integration into CpG islands has been reported to be an HCC risk factor closely associated with methylation, we analyzed whether the lower HCC rate associated with SE integration was associated with host genome methylation. The methylation rate of CpG islands with 100bp flanking integration was calculated as described in the Methods section. To validate our methylation calculation method, we compared the calculated methylation rate with the result of a fetal liver methylation panel. The calculated methylation rate was consistent with the panel results (Figure S1(A)), and the histogram of the methylation rate of CpG islands had a high count in the low-methylation region (Figure S1(B)). Thus, our methylation calculation method could reflect the true methylation state of the host genome.

Regions flanking HBV integration sites in the SE were hypermethylated compared to non-SE integration flanking regions ($P = 0.000$). The methylation conditions were classified into three groups based on the percentage of methylated CpG: low methylation [0,0.25], moderate methylation (0.25,0.75], and high methylation (0.75,1.00]. Both tumor and nontumor SE patients showed a higher rate of high methylation (Figure S2).

Figure 1 shows the histogram of methylation rates of flanking regions of integration in SEs (a) and CpG islands (b). The methylation histogram showed comparable patterns between host whole CpG islands (Figure S1(B)) and flanking regions of integration in CpG islands. However, for SEs, there are distinct patterns (Figure 1(A) and (C)). Regions flanking SEs were more likely to be hypermethylated than those outside SEs (Figure 1(D)).

We then evaluated the methylation rate of 100 bp flanking integration in SEs and the average methylation rate of integrated

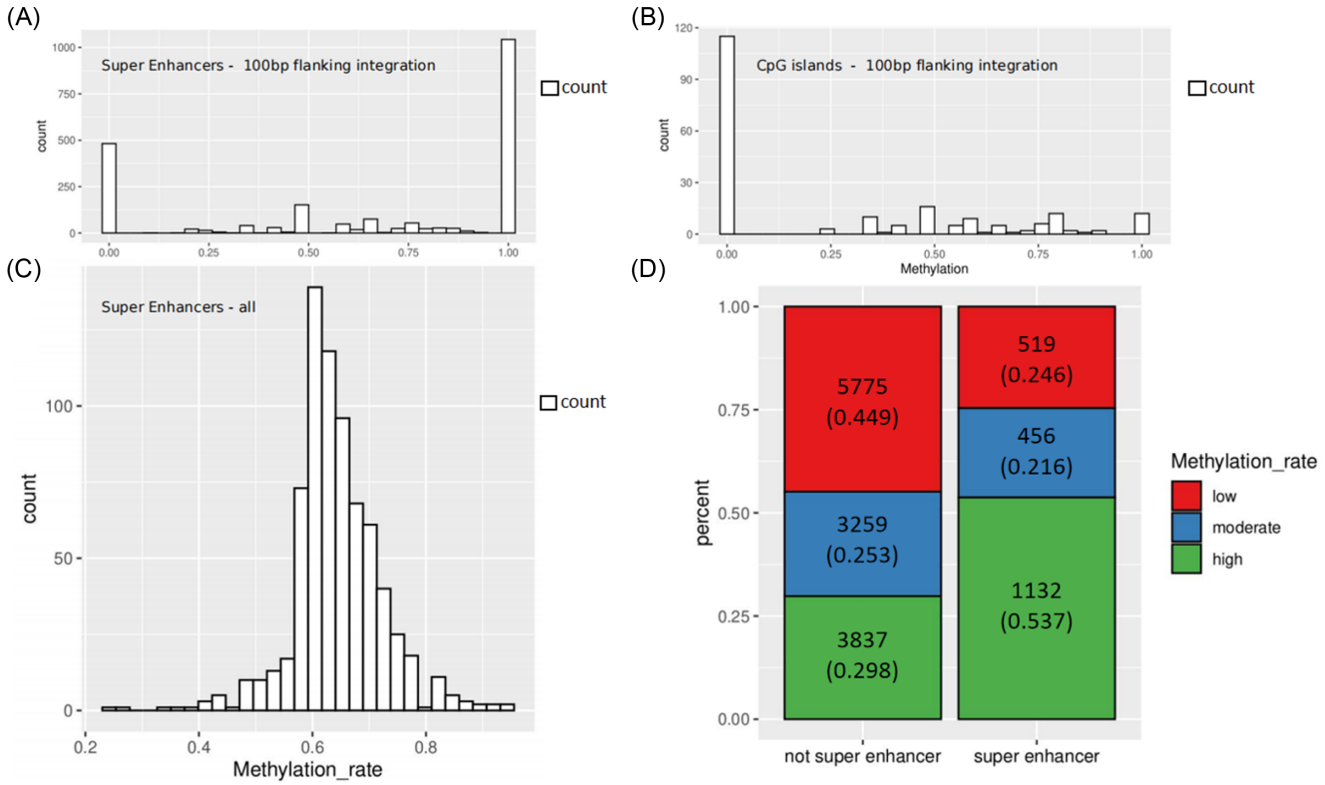


Figure 1. Different methylation distributions in CpG islands and superenhancers flanking HBV integration regions. (A) Histogram of methylation rate flanking integration in SEs. The majority of sites were hypermethylated. (B) Histogram of methylation rate flanking integration in CpGs. The majority of sites were hypomethylated. The methylation rate distribution of integration in CpGs was similar to that of CpGs (Figure S1(B)). (C) Histogram of average methylation rate in SEs. A normal distribution was found. Different from the similar distributions of methylation rate in CpG integrations and CpGs, distinct distributions of methylation rate were found in total SE regions and integration sites in SEs. (D) The methylation conditions were classified into three groups based on the percentage of methylated CpG: low methylation [0,0.25], moderate methylation (0.25,0.75], and high methylation (0.75,1.00]. Regions flanking HBV integration sites in the SE were hypermethylated compared to regions flanking non-SE integration ($P = 0.000$).

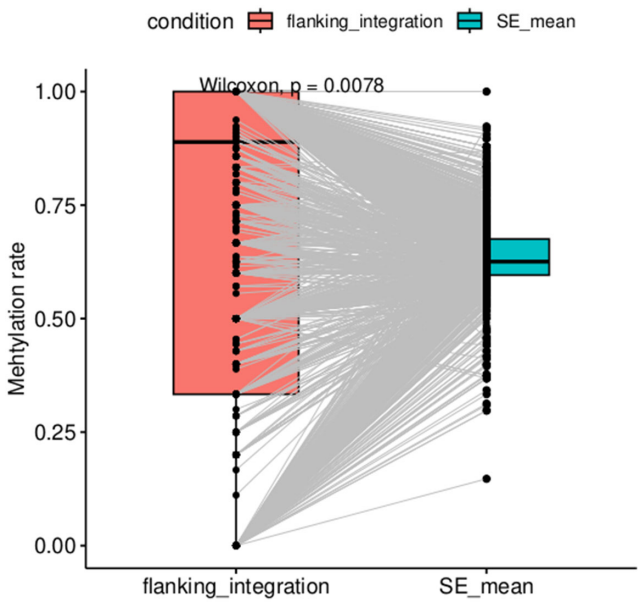


Figure 2. Flanking regions of HBV integration in superenhancers showed a higher methylation rate than the mean methylation rate of superenhancers ($P = 0.008$, paired Wilcoxon test). The flanking regions of HBV integration in SEs showed a higher methylation rate than the mean methylation rate in the same SE ($P = 0.008$).

SEs with a paired Wilcoxon test (Figure 2). The flanking regions of HBV integration in SEs showed a higher methylation rate than the mean methylation rate in SEs ($P = 0.008$).

These results indicated that methylation in SEs was not uniform and that HBV DNA integration in SEs was more likely to occur in hypermethylated regions.

3.3. The hypermethylated SE region flanking HBV integration sites showed lower levels of epigenomic markers

SEs are characterized as clusters with high levels of H3K4me1, H3K27ac, or other master transcription factors. We then evaluated the relationship between CpG methylation and histone modification in the SE region flanking HBV integration sites. Integration sites in SEs were grouped into 3 classes according to the methylation rate of the flanking host genomic region (low: 0–0.25, moderate: 0.25–0.75, high: 0.75–1). The histone modification levels of HepG2 were overall higher than that of human liver tissue. However, these reported integration sites in SEs showed similar histone modification level patterns when aligned to HepG2 and human tissue Chip-seq data. Compared with the centers of SEs, all integration sites showed low H3K4me1 values. Integration sites with high and moderate methylation levels exhibited relatively low H3K4me3 and H3K27ac levels, whereas integration sites with low methylation levels had much greater H3K4me3 and H3K27ac levels than those with moderate or high methylation (Figure 3).

Histone modification markers in liver tissue or in vitro cultured cell lines showed some differences; however, regions annotated as SEs in liver tissue (right lobe of liver)/hepatocyte (in vitro differentiated cell)/HepG2 were enriched with H3K4me1, a marker of SEs, in human tissue Chip-seq data (Figure S3). This result indicated that SEs annotated in cell lines were also likely to be annotated as SEs in human liver tissue.

4. Discussion

In this study, we constructed a logistic regression model to evaluate the relationships between host regulatory elements integrated with HBV DNA and HCC. SEs are newly defined

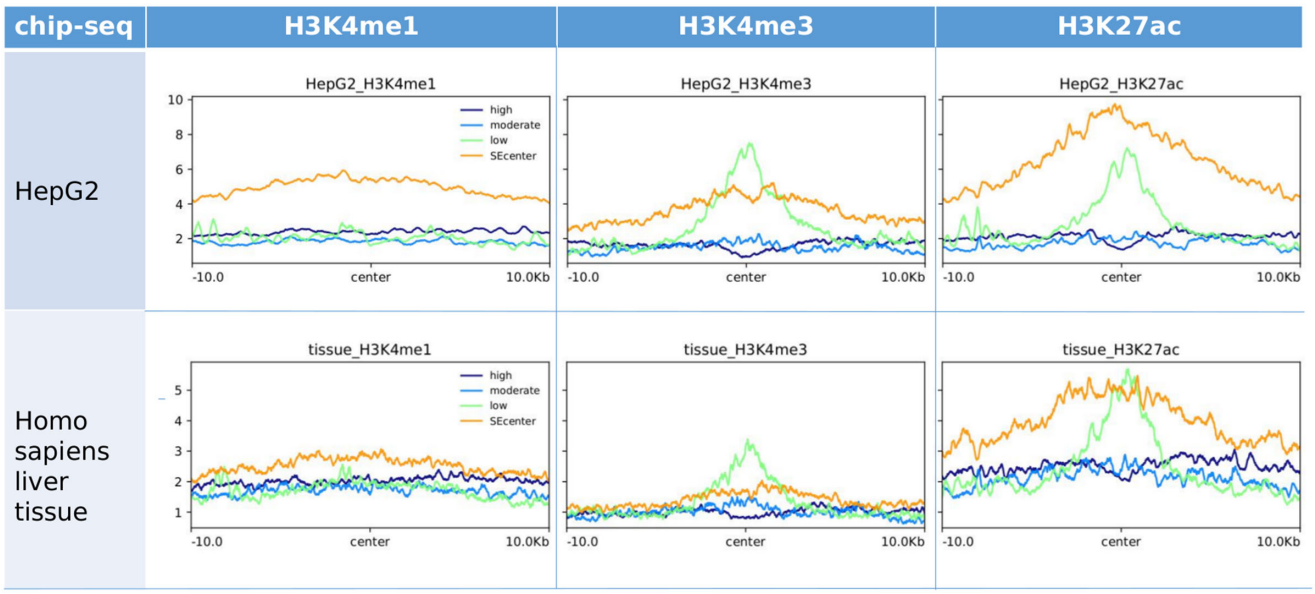


Figure 3. Histone modifications flanking HBV integration sites with high, moderate, and low methylation levels in superenhancers (SEs). These figures showed H3K4me1, H3K4me3, and H3K27ac modification levels in HepG2 cells and human liver tissue from a female adult. All Chip-seq data were collected from ENCODE database. The y-lab was the fold-over control and the x-lab showed the distance from the integration site or the center of the SEs. Integration sites were grouped into 3 classes according to the methylation rate of the flanking host genomic region (low: 0–0.25, moderate: 0.25–0.75, high: 0.75–1). High and moderate methylated integration sites showed low histone modification levels, while low methylated integration sites showed a much higher H3K4me3 level. The H3K27ac level of low methylated integration sites in SEs was also higher than that of high and moderate methylation sites, but were not higher than that of SE centers.

regulatory elements that play important roles in cell identity and oncogenes. Interestingly, we found that integration into SEs was frequent but was significantly associated with a nontumor status ($P = 0.000$, OR = 0.580). This protective effect might be attributed to hypermethylation. Integration into CTCF was also showed to be a protective factor against HCC in our logistic model ($P = 0.000$, OR = 0.479). This protein cooperates with host CTCF, repressing viral transcription of HBV [32].

HBV DNA integration into the host genome is common in chronic HBV infection. Viral DNA preferentially establishes contacts at active chromatin regions. HBV DNA-host DNA contacts are significantly enriched at H3K4me1-marked regions [16], indicating that HBV integration in SEs could be a frequent event. In this study, we annotated 21,520 integration sites from the ViMIC database. Among them, 13.0% occurred in SEs. Among the integrations found in tumor tissues, 10.6% were located in SEs. However, in nontumor tissues, this proportion was 16.8%. The absolute value of the OR of integration into SEs was approximately half of that of integration into CpG islands, a reported risk factor for HCC [10]. The most important feature of CpG islands is that they are hypomethylated. The methylation level of integrated HBV DNA was associated with that of the flanking host genome [33]. Therefore, we analyzed the methylation conditions of 100 bp flanking HBV integration sites in SEs and found that more than half of these host genome regions were hypermethylated. This finding indicates that HBV DNA integrated into SEs was more likely to be hypermethylated. The methylated viral genome had decreased levels of hepatitis B e antigen in patients with HBV genotypes B and C [34]. Moreover, hypomethylation near HBV integration sites was reported to be a risk factor for HCC [35].

However, hypermethylation is not a feature of SEs. The methylation distributions of host whole SEs and flanking regions of integration were markedly different, indicating that methylation in SEs was not uniform and that HBV DNA integration in SEs was more likely to occur in hypermethylated regions. In enhancers, DNA methylation is inversely correlated with H3K4me1 and H3K4me3 enrichment [36]. High DNA methylation was detected in regions with low H3K4me1 and H3K4me3 levels. This finding was consistent with our results showing that the regions of integration in SEs with high methylation rates had low H3K4me1, H3K4me3, and H3K27ac levels. The majority of integration sites in SEs were hypermethylated, indicating that these sites were enriched in regions with low levels of histone modifications.

DNA with low histone modifications seems to be more vulnerable to integration in SEs. Histone chaperones are histone-binding proteins that can regulate nucleosome assembly and shield histone surfaces [37, 38]. Whether histone modification or histone chaperones influence HBV integration into the host genome remains unclear. Histone and chaperones regulating HBV viral chromatin assembly [39] and HIV integration into the host genome [40] have been reported.

Ideally, the features in a logistic regression model should be independent of one another. The presence of multicollinearity among features can lead to unstable estimates and an expansion of the 95% confidence interval (CI) for OR. In this study, we included five features. Integration sites within SEs were found to be more likely to be hypermethylated, whereas hypomethylation was a key characteristic of CpG islands. These findings suggest that SEs and CpG islands may exhibit multicollinearity, which could explain why the lower bound of the 95% CI for the OR associated with SEs was less than 0. In normal conditions, OR values should be greater than 0. A similar pattern in OR 95% CI was also observed for CTCF-binding sites, which can similarly be attributed to feature multicollinearity. CTCF bindings were reported to be related with genome methylation [41]. Furthermore, TSS located near or within TFBS regions could contribute to feature multicollinearity. However, complete avoidance of multicollinearity is challenging in real-world medical datasets. Transcription factors could upregulate or downregulate downstream genes either in cis or trans. In this study, we did not differentiate between these mechanisms since integration into TFBS was not our primary focus.

Finally, we will discuss the annotation sources of CpG islands and SEs used in this study. CpG island regions were obtained from UCSC Genome Browser and were predicted by genome sequence (https://genome.ucsc.edu/cgi-bin/hgTables). It was not tissue-specific but still worked in our logistic model. Integration into host CpG islands showed to be a significant risk factor of HCC with OR∈ [1.542, 3.760]. This finding aligns with previous clinical studies by Zhao et al. [10]. Liver SEs were collected from SEdb, as described in Methods. SEs predicted by all cell types and tissues from the liver were included in the study, reflecting a comprehensive approach. All SE regions derived from liver tissue (right lobe of liver), hepatocytes (in vitro differentiated cells), and HepG2 cells exhibited H3K4me1 enrichment according to human tissue Chip-seq data and HepG2 Chip-seq data, as illustrated in Figure S3.

5. Conclusion

Overall, our results suggested that HBV DNA integration into host SEs was not a rare event and might not be a risk factor for HCC. Integration sites in SEs were more likely to be hypermethylated and to have lower H3K4me1 values. However, further studies are needed to validate the impact of HBV integration into SEs on HCC risk. Investigations on the mechanisms underlying HBV integration enrichment in hypermethylated regions in SEs are also needed.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings of this study are available in public repositories. HBV integration sites in ViMIC: http://bmtongji.cn/ViMIC/downloaddata/integration/HBV_integration.csv. Human genome regulatory elements in hg38: www.genome.ucsc.edu. SEdb: http://www.licpathway.net:8081/sedb/. CpG methylation data of HepG2: https://www.encodeproject.org/experiments/ENCSR786DCL/. methylation panel data for fetal livers: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM1014211. Chip-seq data from the ENCODE project: https://www.encodeproject.org/.

Author Contribution Statement

Mengna Zhang: Methodology, Software, Formal analysis, Investigation, Writing – original draft, Visualization. **Ying Ming:** Conceptualization, Writing – review & editing, Supervision. **Yunling Du:** Validation. **Ziwen Guo:** Formal analysis. **Ziyuan Xin:** Data curation, Investigation. **Yanjun Li:** Data curation. **Ge Yang:** Resources. **Zhaoyang Jiang:** Visualization.

References

[1] Yu, M. C., Yuan, J. M., Govindarajan, S., & Ross, R. K. (2000). Epidemiology of hepatocellular carcinoma. *Canadian Journal of Gastroenterology*, *14*(8), 703–709. https://doi.org/10.1155/2000/371801

[2] Jiang, Y., Han, Q., Zhao, H., & Zhang, J. (2021). The mechanisms of HBV-induced hepatocellular carcinoma. *Journal of Hepatocellular Carcinoma*, *8*, 435–450. https://doi.org/10.2147/JHC.S307962

[3] Farbod, Y., Kankouni, H., Moini, M., & Fung, S. (2025). Hepatitis B-induced hepatocellular carcinoma: Understanding viral carcinogenesis and disease management. *Journal of Clinical Medicine*, *14*(7), 2505. https://doi.org/10.3390/jcm14072505

[4] Chen, Y., & Tian, Z. (2019). HBV-induced immune imbalance in the development of HCC. *Frontiers in Immunology*, *10*, 2048. https://doi.org/10.3389/fimmu.2019.02048

[5] Pollicino, T., & Caminiti, G. (2021). HBV-integration studies in the clinic: Role in the natural history of infection. *Viruses*, *13*(3), 368. https://doi.org/10.3390/v13030368

[6] Yeh, S. H., Li, C. L., Lin, Y. Y., Ho, M. C., Wang, Y. C., Tseng, S. T., & Chen, P. J. (2023). Hepatitis B virus DNA integration drives carcinogenesis and provides a new biomarker for HBV-related HCC. *Cellular and Molecular Gastroenterology and Hepatology*, *15*(4), 921–929. https://doi.org/10.1016/j.jcmgh.2023.01.001

[7] Lin, S. Y., Zhang, A., Lian, J., Wang, J., Chang, T. T., Lin, Y. J., . . . , & Su, Y. H. (2021). Recurrent HBV integration targets as potential drivers in hepatocellular carcinoma. *Cells*, *10*(6), 1294. https://doi.org/10.3390/cells10061294

[8] Cui, X., Li, Y., Xu, H., Sun, Y., Jiang, S., & Li, W. (2023). Characteristics of Hepatitis B virus integration and mechanism of inducing chromosome translocation. *NPJ Genomic Medicine*, *8*(1), 11. https://doi.org/10.1038/s41525-023-00355-y

[9] Sung, W. K., Zheng, H., Li, S., Chen, R., Liu, X., Li, Y., . . . , & Luk, J. M. (2012). Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nature Genetics*, *44*(7), 765–769. https://doi.org/10.1038/ng.2295

[10] Zhao, L. H., Liu, X., Yan, H. X., Li, W. Y., Zeng, X., Yang, Y., . . . , & Wang, H. Y. (2016). Genomic and oncogenic preference of HBV integration in hepatocellular carcinoma. *Nature Communications*, *7*(1), 12992. https://doi.org/10.1038/ncomms12992

[11] Péneau, C., Imbeaud, S., La Bella, T., Hirsch, T. Z., Caruso, S., Calderaro, J., . . . , & Zucman-Rossi, J. (2022). Hepatitis B virus integrations promote local and distant oncogenic driver alterations in hepatocellular carcinoma. *Gut*, *71*(3), 616–626. https://doi.org/10.1136/gutjnl-2020-323153

[12] Bousali, M., Papatheodoridis, G., Paraskevis, D., & Karamitros, T. (2021). Hepatitis B virus DNA integration, chronic infections and hepatocellular carcinoma. *Microorganisms*, *9*(8), 1787. https://doi.org/10.3390/microorganisms9081787

[13] Yang, L., Ye, S., Zhao, X., Ji, L., Zhang, Y., Zhou, P., . . . , & Huang, J. (2018). Molecular characterization of HBV DNA integration in patients with hepatitis and hepatocellular carcinoma. *Journal of Cancer*, *9*(18), 3225–3235. https://doi.org/10.7150/jca.26052

[14] Grudda, T., Hwang, H. S., Taddese, M., Quinn, J., Sulkowski, M. S., Sterling, R. K., . . . , & Thio, C. L. (2022). Integrated hepatitis B virus DNA maintains surface antigen production during antiviral treatment. *The Journal of Clinical Investigation*, *132*(18), e161818. https://doi.org/10.1172/JCI161818

[15] Tu, T., Budzinska, M. A., Shackel, N. A., & Urban, S. (2017). HBV DNA integration: Molecular mechanisms and clinical implications. *Viruses*, *9*(4), 75. https://doi.org/10.3390/v9040075

[16] Yang, B., Li, B., Jia, L., Jiang, Y., Wang, X., Jiang, S., . . . , & Yang, P. (2020). 3D landscape of Hepatitis B virus interactions with human chromatins. *Cell Discovery*, *6*(1), 95. https://doi.org/10.1038/s41421-020-00218-1

[17] Nishiyama, A., & Nakanishi, M. (2021). Navigating the DNA methylation landscape of cancer. *Trends in Genetics*, *37*(11), 1012–1027. https://doi.org/10.1016/j.tig.2021.05.002

[18] Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., . . . , & Young, R. A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*, *153*(2), 307–319. https://doi.org/10.1016/j.cell.2013.03.035

[19] Lovén, J., Hoke, H. A., Lin, C. Y., Lau, A., Orlando, D. A., Vakoc, C. R., . . . , & Young, R. A. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell*, *153*(2), 320–334. https://doi.org/10.1016/j.cell.2013.03.036

[20] Pott, S., & Lieb, J. D. (2015). What are super-enhancers? *Nature Genetics*, *47*(1), 8–12. https://doi.org/10.1038/ng.3167

[21] Ma, H., Qu, J., Pang, Z., Luo, J., Yan, M., Xu, W., . . . , & Qu, Q. (2024). Super-enhancer omics in stem cell. *Molecular Cancer*, *23*(1), 153. https://doi.org/10.1186/s12943-024-02066-z

[22] Guo, S., Zhang, L., Ren, J., Lu, Z., Ma, X., Liu, X., . . . , & Li, J. (2025). The roles of enhancer, especially super-enhancer-driven genes in tumor metabolism and immunity. *International Journal of Biological Macromolecules*, *308*, 142414. https://doi.org/10.1016/j.ijbiomac.2025.142414

[23] Abbas, A., He, X., Niu, J., Zhou, B., Zhu, G., Ma, T., . . . , & Zeng, J. (2019). Integrating Hi-C and FISH data for modeling of the 3D organization of chromosomes. *Nature Communications*, *10*(1), 2049. https://doi.org/10.1038/s41467-019-10005-6

[24] Lucic, B., Chen, H. C., Kuzman, M., Zorita, E., Wegner, J., Minneker, V., . . . , & Lusic, M. (2019). Spatially clustered loci with multiple enhancers are frequent targets of HIV-1 integration. *Nature Communications*, *12*(1), 6326. https://doi.org/10.1038/s41467-021-26471-w

[25] Rheinberger, M., Costa, A. L., Kampmann, M., Glavas, D., Shytaj, I. L., Sreeram, S., . . . , & Lusic, M. (2023). Genomic profiling of HIV-1 integration in microglia cells links viral integration to the topologically associated domains. *Cell Reports*, *42*(2), 112110. https://doi.org/10.1016/j.celrep.2023.112110

[26] Warburton, A., Redmond, C. J., Dooley, K. E., Fu, H., Gillison, M. L., Akagi, K., . . . , & McBride, A. A. (2018). HPV integration hijacks and multimerizes a cellular enhancer to generate a viral-cellular super-enhancer that drives high viral oncogene expression. *PLoS Genetics*, *14*(1), e1007179. https://doi.org/10.1371/journal.pgen.1007179

[27] Wang, Y., Tong, Y., Zhang, Z., Zheng, R., Huang, D., Yang, J., . . . , & Zhang, X. (2022). ViMIC: A database of human disease-related virus mutations, integration sites and *cis*-effects. *Nucleic Acids Research*, *50*(D1), D918–D927. https://doi.org/10.1093/nar/gkab779

[28] Nassar, L. R., Barber, G. P., Benet-Pagès, A., Casper, J., Clawson, H., Diekhans, M., . . . , & Kent, W. J. (2023). The UCSC genome browser database: 2023 update. *Nucleic Acids Research*, *51*(D1), D1188–D1195. https://doi.org/10.1093/nar/gkac1072

[29] Wang, Y., Song, C., Zhao, J., Zhang, Y., Zhao, X., Feng, C., . . . , & Li, C. (2023). SEdb 2.0: A comprehensive super-enhancer database of human and mouse. *Nucleic Acids Research*, *51*(D1), D280–D290. https://doi.org/10.1093/nar/gkac968

[30] Luo, Y., Hitz, B. C., Gabdank, I., Hilton, J. A., Kagda, M. S., Lam, B., . . . , & Cherry, J. M. (2020). New developments on the Encyclopedia of DNA elements (ENCODE) data portal. *Nucleic Acids Research*, *48*(D1), D882–D889. https://doi.org/10.1093/nar/gkz1062

[31] Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., . . . , & Manke, T. (2016). deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, *44*(W1), W160–W165. https://doi.org/10.1093/nar/gkw257

[32] Watanabe, Y., Yamamoto, H., Oikawa, R., Toyota, M., Yamamoto, M., Kokudo, N., . . . , & Itoh, F. (2015). DNA methylation at hepatitis B viral integrants is associated with methylation at flanking human genomic sequences. *Genome Research*, *25*(3), 328–337. https://doi.org/10.1101/gr.175240.114

[33] D'Arienzo, V., Ferguson, J., Giraud, G., Chapus, F., Harris, J. M., Wing, P. A. C., . . . , & Parish, J. L. (2021). The CCCTC-binding factor CTCF represses hepatitis B virus enhancer I and regulates viral transcription. *Cellular Microbiology*, *23*(2), e13274. https://doi.org/10.1111/cmi.13274

[34] Nakamura, T., Inoue, J., Ninomiya, M., Kakazu, E., Iwata, T., Takai, S., . . . , & Masamune, A. (2020). Effect of viral DNA methylation on expression of hepatitis B virus proteins depends on the virus genotype. *Virus Genes*, *56*(4), 439–447. https://doi.org/10.1007/s11262-020-01761-5

[35] Zhang, H., Dong, P., Guo, S., Tao, C., Chen, W., Zhao, W., . . . , & Zeng, C. (2020). Hypomethylation in HBV integration regions aids non-invasive surveillance to hepatocellular carcinoma by low-pass genome-wide bisulfite sequencing. *BMC Medicine*, *18*(1), 200. https://doi.org/10.1186/s12916-020-01667-x

[36] Sharifi-Zarchi, A., Gerovska, D., Adachi, K., Totonchi, M., Pezeshk, H., Taft, R. J., . . . , & Araúzo-Bravo, M. J. (2017). DNA methylation regulates discrimination of enhancers from promoters through a H3K4me1-H3K4me3 seesaw mechanism. *BMC Genomics*, *18*(1), 964. https://doi.org/10.1186/s12864-017-4353-7

[37] Jonas, F., Vidavski, M., Benuck, E., Barkai, N., & Yaakov, G. (2023). Nucleosome retention by histone chaperones and remodelers occludes pervasive DNA-protein binding. *Nucleic Acids Research*, *51*(16), 8496–8513. https://doi.org/10.1093/nar/gkad615

[38] Hammond, C. M., Bao, H., Hendriks, I. A., Carraro, M., García-Nieto, A., Liu, Y., . . . , & Groth, A. (2021). DNAJC9 integrates heat shock molecular chaperones into the histone chaperone network. *Molecular Cell*, *81*(12), 2533–2548. https://doi.org/10.1016/j.molcel.2021.03.041

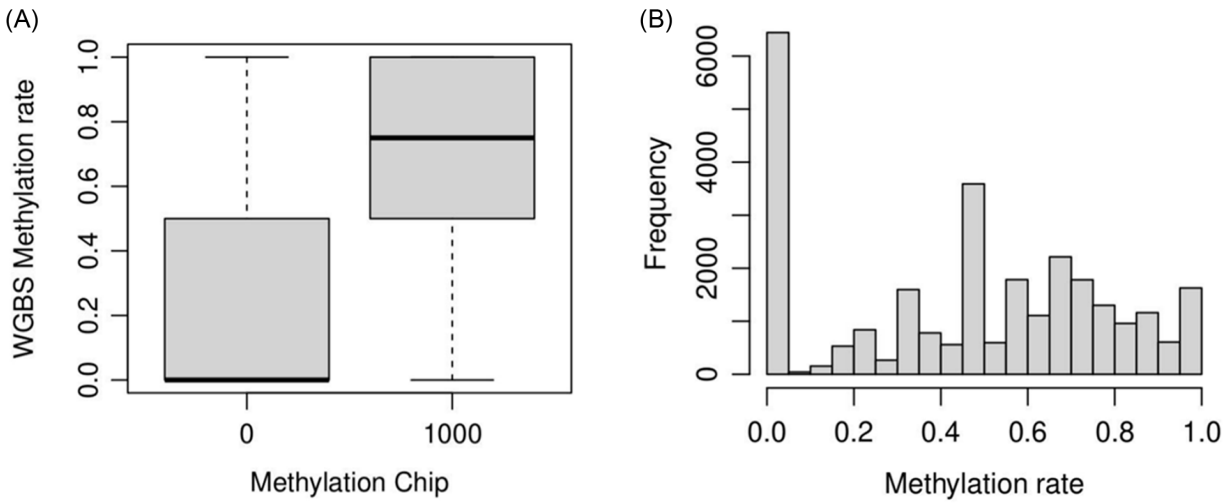
[39] Alvarez-Astudillo, F., Garrido, D., Varas-Godoy, M., Gutiérrez, J. L., Villanueva, R. A., & Loyola, A. (2021). The histone variant H3.3 regulates the transcription of the hepatitis B virus. *Annals of Hepatology*, *21*, 100261. https://doi.org/10.1016/j.aohep.2020.09.005

[40] Matysiak, J., Lesbats, P., Mauro, E., Lapaillerie, D., Dupuy, J. W., Lopez, A. P., . . . , & Parissi, V. (2017). Modulation of chromatin structure by the FACT histone chaperone complex regulates HIV-1 integration. *Retrovirology*, *14*(1), 39. https://doi.org/10.1186/s12977-017-0363-4

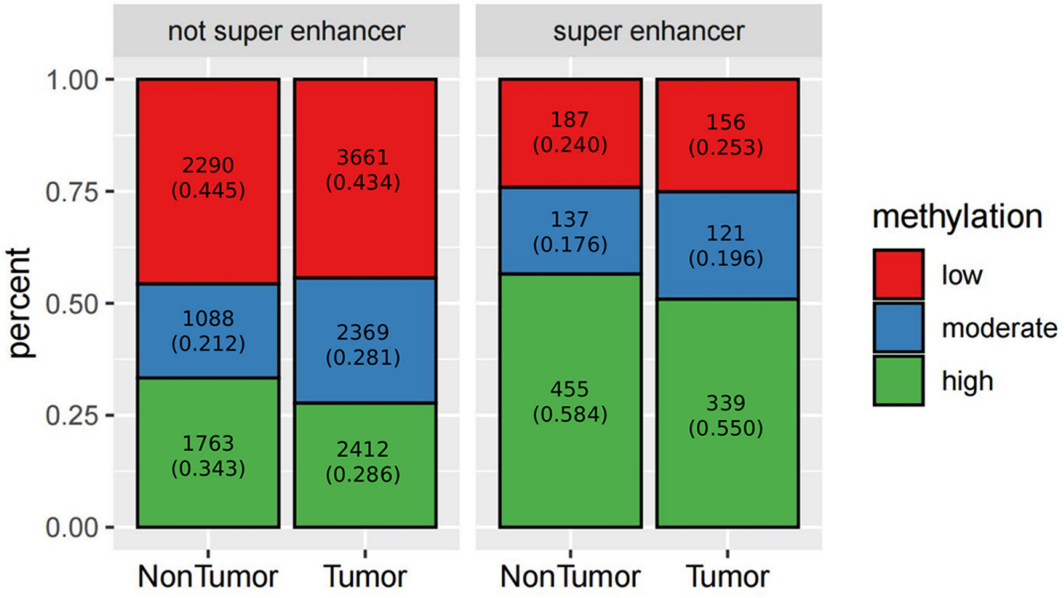
[41] Monteagudo-Sánchez, A., Noordermeer, D., & Greenberg, M. V. C. (2024). The impact of DNA methylation on CTCF-mediated 3D genome organization. *Nature Structural & Molecular Biology*, *31*(3), 404–412. https://doi.org/10.1038/s41594-024-01241-6

How to Cite: Zhang, M., Ming, Y., Du, Y., Guo, Z., Xin, Z., Li, Y., . . . , & Jiang, Z. (2025). HBV Integration into Host Superenhancers Might Not Be a Risk Factor for Hepatocellular Carcinoma. *Medinformatics*, *2*(4), 309–317. https://doi.org/10.47852/bonviewMEDIN52025080

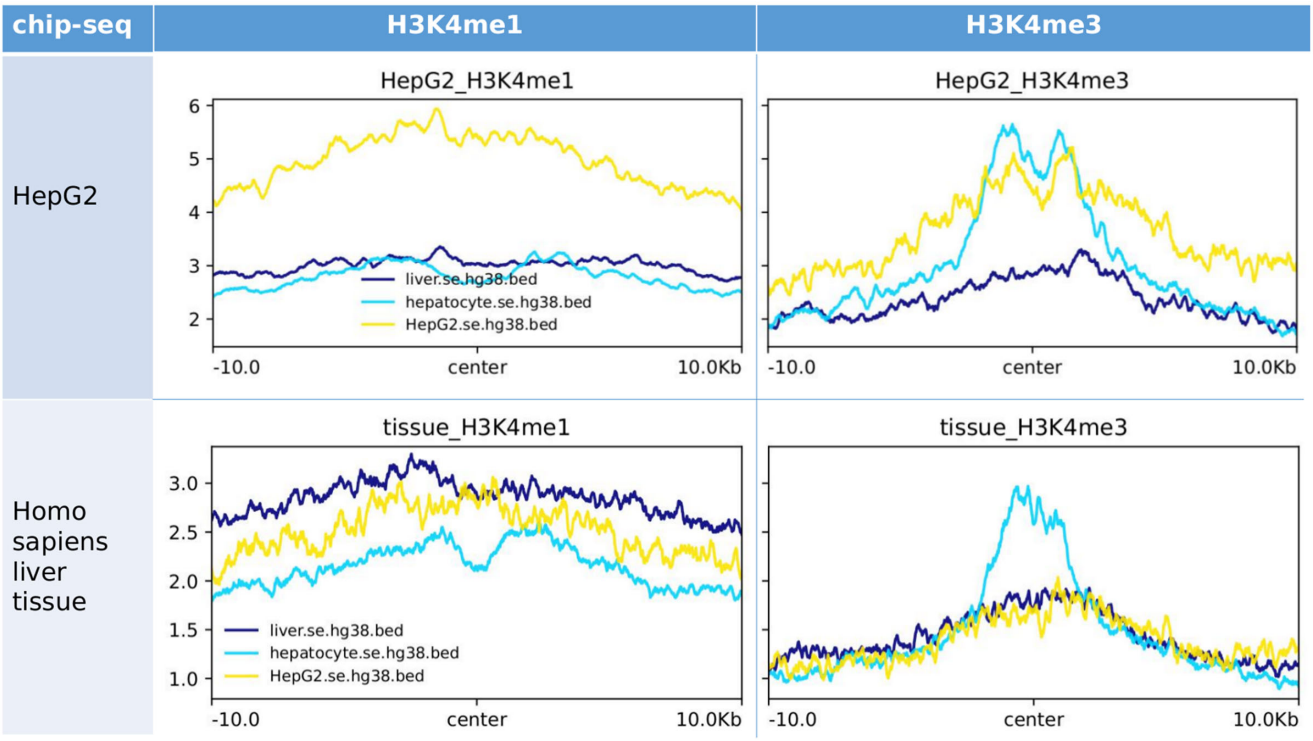
Supplementary Information



Supplementary Figure S1. Methylation rate calculated as described in Methods. (A) A validation of our methylation rate calculation method with methylation panel. In methylation chip results, 0 means not methylated and 1000 means hypermethylated. (B) The distribution of methylation rate of CpG islands in the whole genome calculated as our method.



Supplementary Figure S2. In superenhancers, both tumor and nontumor integration sites (with 100 bp flanking regions) showed a higher rate to be hypermethylated.



Supplementary Figure S3. Histone modification levels of SE regions from SEdb. We additionally collected Chip-seq data of homo sapiens liver tissue of a female adult (25 years) from ENCODE database. se.hg38.bed were SE locations obtained from SEdb. Though HepG2 cell line showed higher histone modification levels than hepatocyte and liver tissue, the genomic regions that were annotated as SEs in liver tissue and HepG2 showed similar H3K4me1 and H3K4me3 levels when aligned to tissue Chip-seq data.

Abbreviations

CTCF	CCCTC-binding factor
HBV	hepatitis B virus
HCC	hepatocellular carcinoma
OR	odds ratio
SE	superenhancer
TFBS	transcription factor-binding site
TSS	transcription start site