

RESEARCH ARTICLE

Genomic modeling of hepatitis B virus integration frequency in the human genome

Ondrej Podlaha, George Wu, Bryan Downie, Raghuraman Ramamurthy, Anuj Gaggar, Mani Subramanian, Zhishen Ye, Zhaoshi Jiang*

Gilead Sciences Inc., Foster City, CA, United States of America

* Zhaoshi.Jiang@gilead.com



Abstract

Hepatitis B infection is a world-wide public health burden causing serious liver complications. Previous studies suggest that hepatitis B integration into the human genome plays a crucial role in triggering oncogenic process and may also constitutively produce viral antigens. Despite the progress in HBV biology and sequencing technology, our fundamental understanding of how many hepatocytes in the liver actually carry viral integrations is still lacking. Herein we provide evidence that the HBV virus integrates with a lower-bound frequency of 0.84 per diploid genome in hepatitis B positive hepatocellular cancer patients. Moreover, we calculate that integrated viral DNA generates ~80% of the HBsAg transcripts in these patients. These results underscore the need to re-evaluate the clinical end-point and treatment strategies for chronic hepatitis B patients.

OPEN ACCESS

Citation: Podlaha O, Wu G, Downie B, Ramamurthy R, Gaggar A, Subramanian M, et al. (2019) Genomic modeling of hepatitis B virus integration frequency in the human genome. PLoS ONE 14(7): e0220376. <https://doi.org/10.1371/journal.pone.0220376>

Editor: Isabelle Chemin, Centre de Recherche en Cancérologie de Lyon, FRANCE

Received: April 12, 2019

Accepted: July 14, 2019

Published: July 29, 2019

Copyright: © 2019 Podlaha et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Datasets were downloaded from the European Genome-phenome Archive <https://www.ebi.ac.uk/ega>, accession numbers EGAD00001001881, EGAD00001001880, and PRJEB2869.

Funding: This research was supported by Gilead Sciences. The funder provided support in the form of salaries for authors O.P., G.W., B.D., R.R., A.G., M.S., Z.Y., and Z.J., but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the

Introduction

Hepatitis B (HBV) infection is one of the leading causes of liver fibrosis, liver function failure, and hepatocellular carcinoma worldwide [1–6]. The virus has a compact circular genome comprised of 3.2Kb partially double-stranded DNA, which encodes four overlapping genes essential for the regulation, replication, and the formation of the nucleocapsid and the viral envelope [4–6]. During the HBV replication cycle, failure of the template-switching process results in the formation of a linear double-stranded HBV DNA [7], which can then integrate into the human genome. Although current anti-viral treatments can significantly suppress HBV replication, HBsAg seroconversion, the most widely applied clinical endpoint for evaluating HBV treatment success, is rarely achieved in chronic hepatitis B (CHB) patients [8]. A recent RNAi-based treatment study [9] demonstrated that integrated HBV DNA can contribute significantly to serum HBsAg expression. However, this conclusion was derived from a small number of chronically infected chimpanzees. It is important to confirm this observation in a large cohort of CHB patients. In order to better understand the magnitude and clinical relevance of HBV integration in CHB subjects, we set out to quantify: i) the frequency of independent HBV integration events and ii) the fraction of viral HBsAg transcription originating from integrated HBV DNA. Due to the invasive nature of liver biopsy, it is challenging to obtain genomic datasets for large CHB patient cohorts. In contrast, whole genome and transcriptome studies from surgically dissected HBV infected hepatocellular carcinoma (HCC)

manuscript. The specific roles of these authors are articulated in the 'author contributions' section.

Competing interests: O.P., G.W., B.D., R.R., A.G., M.S., Z.Y., and Z.J. are employees of Gilead Sciences and own Gilead stock. Gilead Sciences affiliation does not alter our adherence to PLOS ONE policies on sharing data and materials.

samples are widely available, including adjacent non-tumor liver tissue [10–14]. Although the adjacent non-tumor liver tissue samples from HCC patients may not exactly resemble the genetic and transcriptional profile of non-HCC CHB patients, it serves as the most relevant surrogate sample for this data analysis. We therefore collected whole genome and transcriptome sequencing data from two large genomic studies of HCC patients [10, 14] and estimated the frequency of HBV DNA integration in non-tumor liver samples and the fraction of HBsAg expression from integrated HBV DNA copies.

Materials and methods

Dataset

We assembled liver cancer dataset from Fujimoto et al. [15] and Sung et al. [14], which included DNA and RNA sequencing data from tumor and adjacent non-tumor liver tissues. Our analysis focused only on a subset of HBV positive HCC samples defined by the detection of HBV DNA or RNA reads from sequencing data. The final dataset contained a total of 192 DNaseq samples (120 tumor and 72 adjacent non-tumor samples) and 60 RNAseq samples (30 tumor and 30 adjacent non-tumor samples S1 Table). Majority of these samples were paired tumor and non-tumor adjacent samples. Datasets were downloaded from the European Genome-phenome Archive <https://www.ebi.ac.uk/ega>, accession numbers EGAD00001001881, EGAD00001001880, and PRJEB2869. Clinical information related to serum HBeAg status, HBsAg levels, ALT, and other viral biomarkers were not publically available. HBV genotype was inferred via viral read mapping (see below).

Identification of HBV integrations

RNAseq and DNaseq read pairs were first aligned using bwa-mem[16] to the reference genome of HBV genotypes A, B, C, D, E, F, and H. HBV genotype for each sample was determined by assessing which HBV reference genotype produced the highest quality alignment. From here on, each sample was then processed using the best matching HBV genotype. Pairs of reads in which at least a portion of one read mapped to the HBV genome were retained and re-aligned to a hybrid GRCh38:HBV genome reference sequence in which the HBV sequence was included as a pseudo-chromosome (GenBank: GQ924620). In order to quantify viral insertion breakpoint coverage, a dual-clustering strategy was implemented. Chimeric read pairs, in which one read aligns to HBV and the second paired-read aligns to GRCh38, were extracted from the BAM alignment file and processed using bedtools bamtobed[17] in paired-end mode. Integration junction reads, in which one read was composed of both HBV and hg38 sequence, were split and treated as pseudo-read pairs and otherwise processed identically. Read pairs which aligned to more than 3 loci were discarded. This approach yielded specific breakpoint coordinates while still maintaining almost all reads deriving from chimeric nucleic acid fragments. Breakpoint read clusters were generated from processed chimeric reads using bedtools cluster with a window size of 500 bp. Clusters were defined by HBV coordinate, human coordinate, and the strand of reads supporting the cluster.

HBV genome coverage for both DNA and RNA samples was calculated using bedtools genomcov using all reads which aligned to the GRCh38:HBV hybrid genome with mapping quality > 20 and the primary read alignment flag. For visual comparisons, read depth for each sample was normalized to the maximum HBV genome depth of coverage.

Estimation of HBV integration frequency

During a typical Next Generation Sequencing (NGS) workflow, genomic DNA from a sample with a large number (typically 200,000–1,000,000) of cells is extracted and fragmented,

followed by amplification and sequencing (i.e. by Illumina technology). This sequencing workflow ends once the read depth coverage reaches a predefined level. In a typical whole genome sequencing (WGS) study, 30x coverage is often considered sufficient for calling germline variants in the diploid human genome. However, even much higher coverages than 30x would be insufficient to capture all possible HBV viral integrations across hundreds of thousands of cells, since viral integrations appear to be distributed randomly across the human genome and generate a heterogeneous integration sites across hepatocytes [12] (Fig 1A). Because the heterogeneity of the viral integration alleles at the cell-level make it impossible to recover all integrations from limited depth of coverage of sequencing, we therefore attempted to indirectly estimate the frequency of independent HBV integration events as follows: Let I be the number of genomes (content of DNA extraction) in a given sample and let J be the number of genomic locations (length of the human genome; ~3.2 billion base pairs). For a given genome i and a given genomic location j , let $x_{ij} = 1$ if an HBV integration is present and $x_{ij} = 0$ if an HBV integration is absent. By definition, the total number of HBV integrations across all genomic locations in all genomes in the human tissue sample therefore is $\sum_i \sum_j x_{ij}$, and the HBV integration frequency, i.e. the average number of HBV integrations per genome, is $\frac{\sum_i \sum_j x_{ij}}{I}$.

Let Y_j be the number of observed HBV integrations at a given genomic location j after sequencing the processed sample. If we assume that Y_j approximately follows a binomial distribution, $Y_j \sim \text{binomial}(n_j, p_j)$, where $p_j = \frac{\sum_i x_{ij}}{I}$ is the proportion of true HBV integrations at genomic location j and n_j is the read depth at genomic location j , then the maximum likelihood estimate (MLE) of the proportion of true HBV integrations at genomic location j is $\hat{p}_j = \frac{y_j}{n_j}$, where y_j is the observed number of HBV integrations at genomic location j . The MLE of the HBV integration frequency, i.e. the average number of HBV integrations per genome, is then $\hat{y} = \sum_j \hat{p}_j$. To validate our assumptions, we performed down-sampling of a high-coverage patient sample and found that the mean integration count matched our model well (Fig 1B).

As it is unknown whether an observed HBV integration is due to an independent viral integration event or due to clonal expansion following an integration, we conservatively estimated the number of unique HBV integration events at each genomic location j as a lower bound estimate of \hat{p}_j . This conservative estimate can be denoted as: $\hat{\theta}_j = \frac{I[y_j]}{n_j}$, where $I[y_j] = \begin{cases} 0, & \text{if } y_j = 0 \\ 1, & \text{if } y_j > 0 \end{cases}$, and serves as a lower bound because: $\hat{p}_j = \frac{y_j}{n_j} \geq \frac{I[y_j]}{n_j} = \hat{\theta}_j$. The corresponding lower bound estimate of the HBV integration frequency would then be: $\sum_j \hat{\theta}_j$. When estimating the average number of HBV integrations, we set n_j to be n , which approximates read depth of coverage at each genomic position to be the genome-wide average depth of coverage.

There are two important benefits worth noting about our approach. First, our method does not require knowing the number of cells put into the WGS workflow. We derive that sequencing a sample to, for example, 30x coverage, is equivalent to sampling 30 genomes from however many cells used as an input for the sequencing library preparation. Second, our primary goal was to identify the frequency of independent integration events, not the total number of viral integrations. Therefore, it is critical in our estimation to count for the increase of HBV integration burden due to subsequent hepatocyte clonal expansion. If a clonal expansion occurs and happens to carry a specific integration, this integration will share identical genomic locus across expanded hepatocytes. In our estimation, this event is counted once and only once because there was only a single founder integration event prior to the subsequent clonal expansion. The increase of viral integration burden due to a clonal expansion is not directly

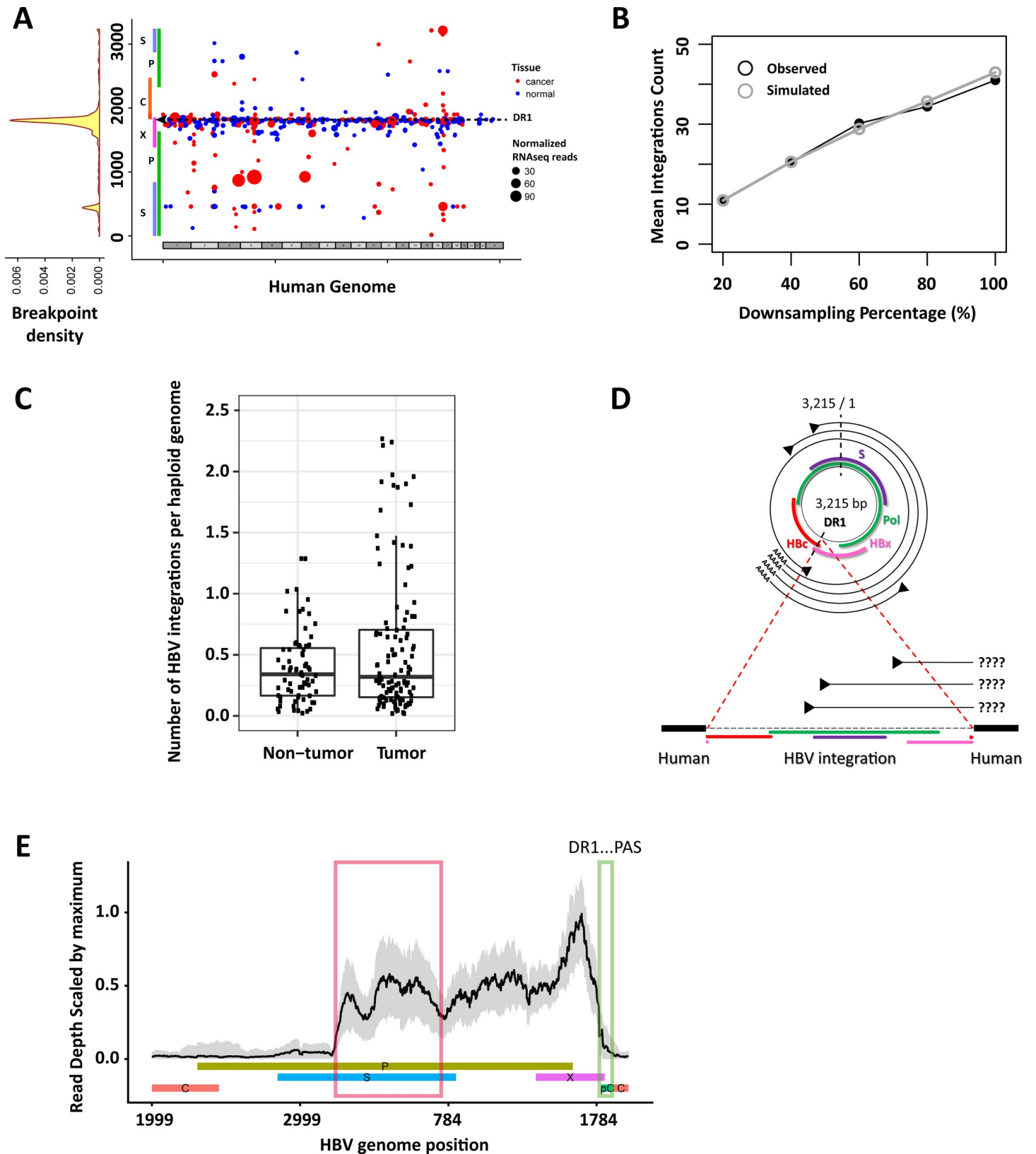


Fig 1. Estimation of hepatitis B integration frequency and its contribution to the viral S antigen expression. (A) The hot-spot of viral-human fusion junctions on the viral genome. Although viral integrations on the human genome tend to spread across every chromosome (x-axis), the DR1 region seems to be the most commonly used

fusion breakpoint in the HBV genome. Each dot represents a viral integration identified in the liver RNAseq data. The red dots represent data from tumor samples while the blue dots represent data from non-tumor samples. Size of dot is proportional to the normalized read count supporting each integration call. The read count is normalized to 1 million library reads. (B) The number of HBV DNA integrations as a function of sequencing-depth coverage: real data (black) versus result from a Binomial modeling (grey). (C) HBV integration frequency representing the number of viral integration per sequenced haploid genome in non-tumor and tumor samples. (D) Schematic of HBV genome structure, four sets of transcripts, and linearization during integration into the human genome. Most HBV integrations fused to the human genome at the DR1 region downregulate the viral expression of the DR1-PAS region and the core gene (green box and gene labeled C in panel E). (E) Aggregated RNAseq coverage across all non-tumor samples ($n = 59$) on the HBV genome (x-axis). The median value was shown as a solid line with grey bars indicating the variance. Coverage is scaled to one across all samples for comparison purposes. The depth of RNAseq coverage drops right after the DR1 locus. The expression level of DR1-PAS window (green box) represents the transcription from circular genome while the expression of the S window (red box) represents a total HBs expression from both the circular genome and integrated linear HBV genome. The core gene (labeled C) expression was depleted in the integrated HBV. PAS—viral PolyA signal site.

<https://doi.org/10.1371/journal.pone.0220376.g001>

related to the viral replication process, therefore it should not be considered as independent integration events.

Results and discussion

We examined the paired-end whole genome sequencing data from 72 non-tumor samples in HBV positive HCC patients reported in recent studies [10, 14]. Of these 72 patients, 53 were infected with HBV genotype C and 19 with HBV genotype B. We estimated the average HBV integration frequency to be 0.431 per haploid genome (median = 0.347, min = 0.022, max = 2.670; Fig 1C). We did not detect any HBV genotype specific bias (mean integration rate $i_{GTB} = 0.434$ and $i_{GTC} = 0.422$, Wilcoxon rank sum test $p = 0.41$, S1 Fig). Given a typical diploid cell configuration, this translates to an average frequency of 0.844 per diploid genome. Since we counted only unique viral integration events, this represents the lower-bound estimation of the actual viral integration burden (see Methods). Although this estimate is very conservative, the order of magnitude is rather high, indicating that HBV integration in CHB patients is potentially very pervasive and higher than previously thought [1, 18]. Since this analysis is based on bulk sequencing data, this lower bounds estimate represents the average viral integration frequency throughout HBV infected liver tissue sample and is not characterizing individual cells. It is not possible to discern whether this estimate is a result of a few cells harboring many distinct integrations or large number of cells each harboring few integration events. Future single cell sequencing based study is necessary to define the spatial distribution of viral integrations in CHB patients.

This estimation method can be applied to tumor samples as well. Since only independent viral integrations with distinct genomic location are included in the frequency calculation, our method inherently controls possible over-estimation of viral integration due to clonal expansion, a phenomenon characteristic of tumor evolution. The average frequency of HBV integration across 120 tumor tissue samples was 0.674 per haploid genome (median = 0.340, min = 0.019, max = 6.005). Interestingly, integration estimates from both tumor and non-tumor tissue samples are comparable suggesting that viral integration frequency is relatively constant regardless of oncogenic progression. To confirm that none of the detected integrations in the liver tissue samples were a result of PCR or mapping artifacts, DNA sequencing data from matched blood samples was analyzed as a negative control using the same methodology. We did not detect any HBV integrations in the blood samples suggesting that our results across both tumor and non-tumor tissue samples reflect true viral integration phenomenon.

Given the prolific rate of HBV integration in non-tumor liver tissue, we further quantified the relative contribution of integrated HBV genomes to HBsAg RNA expression. Despite alternate transcription start sites, all four major families of cccDNA transcripts share a common 3' polyadenylation signal locus [19, 20] (Fig 1D). Transcripts generated from the integrated HBV DNA template, however, tend to be disrupted at the DR1 locus, a common integration junction on the HBV genome (Fig 1A and 1D). Since the majority of expressed HBV DNA

integrations are fused with the human genome across the DR1 locus (Fig 1A) and lead to a disruption of the DR1-PAS region, any transcriptional activity of the DR1-PAS region (designated as $\text{Depth}_{\text{DR1-PAS}}$) should originate primarily from circular viral template DNA (cccDNA). Conversely, transcripts covering the S gene (Depth_S) would originate from either cccDNA or integrated linear HBV DNA. Following this rationale, we estimated the proportion of transcripts originating from integrated HBV using the formula: $(\text{Depth}_S - \text{Depth}_{\text{DR1-PAS}}) / \text{Depth}_S$ (Fig 1E). Using this calculation, we determined that the median percentage of transcripts originating from integrated HBV DNA in tumor samples was 85.5% and 77.2% in non-tumor samples (median of pooled tumor and non-tumor was 82.8%). The difference between tumor and non-tumor samples was not statistically significant (Wilcoxon rank sum test, $p = 0.72$). This observation is comparable to Wooddell et al [9] findings (66.4%), which implicated a significant contribution of integrated HBV to serum HBsAg levels in CHB patients. Because DNA and RNA extraction did not sample the same cells, further investigation comparing HBV integrations from DNaseq and RNAseq directly is not adequate.

Together, these results highlight the impact of integrated HBV DNA in CHB patients when developing HBV therapeutic endpoints. With an average HBV DNA integration rate in non-tumor tissue samples of 0.844 integrations per diploid genome, integrated HBV would be expected to exert a dominant HBsAg transcriptional burden in patients—a hypothesis supported by the finding that ~80% of HBs transcripts are derived from integrated HBV DNA. Although we have restricted our analyses to conservative estimates of integration frequency and transcript proportion, there are additional caveats which should be highlighted. First, due to limitations in sample availability, we do not know the extent to which the integration profile of non-tumor tissue samples from HCC patients diverge from that of liver tissue of CHB patients. Second, sequenced bulk liver tissue samples will be composed of a heterogeneous mix of cell types. While we expect that only hepatocytes would be infected by HBV, DNA from all tissue resident cell types was sequenced, which would lead to an under-estimation of viral integration rate using our method. Third, HBV transcripts could undergo degradation of the 3' end near the polyadenylation tail, leading to fewer reads spanning the DR1-PAS region and possible overestimation of HBsAg transcription from integrated HBV DNA template. This, however, doesn't seem to be the case as analysis of human gene RNA-seq coverage did not reveal 3'-end degradation or 5'-end bias.

Additionally, we do not account for integration sites outside the DR1-PAS region when estimating relative contribution of integrated DNA to transcriptional activity. Given that the DR1 site has been described as the major junction for viral integrations, our expectation would be that RNA originating from the fraction of non-DR1-PAS integration sites would not substantially impact our findings.

We consider this work to have two main implications for the long-term development of HBV-directed therapies. First, our work supports the observation that integrated HBV DNA serves as a template for HBs antigen transcription and implies that quantitative serum level of HBsAg is a confounded biomarker when evaluating disease progression in CHB patients. Based on the findings of our work and others[21], HBc gene expression would be expected to reflect HBV viral replication activity more accurately. Combination of therapeutics targeting infected cells with either cccDNA or integrated viral DNA will be needed to eliminate HBsAg levels and increase HBsAg seroconversion rates. Our results emphasize the significance of various strategies that are currently being tested including RNAi, targeted DNA editing technologies [9, 22, 23], and the modulation of the host immune system via therapeutic vaccine or immune checkpoint blockades[24, 25]. Second, CHB patients with normal alanine aminotransferase (ALT) levels are generally not treated with interventional therapies[26]. We would like to raise the awareness that patients with normal ALT levels may still be exposed to the risk associated with

constitutive viral integrations. The accumulation of HBV integrations in the human genome imposes a mutation burden by disrupting important regulatory genes, drives aberrant gene expression by fusion of human-viral sequences, and induces genomic rearrangements[5], all of which pose the risk of oncogenic transformation. Indeed, a recent study on non-cirrhotic CHB patients with normal to minimally elevated ALT demonstrated that patients receiving antiviral therapy had reduced risk of liver cancer compared to an untreated cohort [27].

In summary, we observed that genomic integration of HBV DNA occurs frequently in non-tumor liver tissue of HBV infected HCC human subjects, and that integrated HBV DNA is the primary contributor to the HBsAg transcriptional burden. These findings highlight the need to re-visit the clinical interpretation of HBsAg levels of CHB patients and to determine the increased oncogenic risk to untreated CHB patients caused by viral integrations.

Supporting information

S1 Fig. HBV integration frequency across HBV genotypes. The median HBV integration frequencies across HBV genotypes B and C are 0.434 and 0.422, respectively. This difference is not statistically significant (Wilcoxon rank sum test, $p = 0.41$).

(TIF)

S1 Table. Origin of samples used in this study.

(DOCX)

Acknowledgments

This research was supported by Gilead Sciences. We further want to acknowledge and thank respective research groups Sung et al. and Fujimoto et al. for providing this valuable dataset to the public.

Author Contributions

Conceptualization: Ondrej Podlaha, Bryan Downie, Anuj Gaggar, Mani Subramanian, Zhishen Ye, Zhaoshi Jiang.

Data curation: Raghuraman Ramamurthy.

Formal analysis: Ondrej Podlaha, George Wu, Bryan Downie, Raghuraman Ramamurthy.

Funding acquisition: Zhaoshi Jiang.

Investigation: Zhaoshi Jiang.

Project administration: Zhaoshi Jiang.

Resources: Zhaoshi Jiang.

Supervision: Zhaoshi Jiang.

Visualization: Raghuraman Ramamurthy.

Writing – original draft: Ondrej Podlaha, George Wu, Zhaoshi Jiang.

Writing – review & editing: Ondrej Podlaha.

References

1. Zeisel MB, Lucifora J, Mason WS, Sureau C, Beck J, Levrero M, et al. Towards an HBV cure: state-of-the-art and unresolved questions—report of the ANRS workshop on HBV cure. *Gut*. 2015; 64(8):1314–26. <https://doi.org/10.1136/gutjnl-2014-308943> PMID: 25670809.

2. Lavanchy D. Hepatitis B virus epidemiology, disease burden, treatment, and current and emerging prevention and control measures. *J Viral Hepat.* 2004; 11(2):97–107. PMID: [14996343](#).
3. Custer B, Sullivan SD, Hazlet TK, Iloeje U, Veenstra DL, Kowdley KV. Global epidemiology of hepatitis B virus. *J Clin Gastroenterol.* 2004; 38(10 Suppl 3):S158–68. PMID: [15602165](#).
4. Rehermann B, Nascimbeni M. Immunology of hepatitis B virus and hepatitis C virus infection. *Nat Rev Immunol.* 2005; 5(3):215–29. <https://doi.org/10.1038/nri1573> PMID: [15738952](#).
5. Arzumanyan A, Reis HM, Feitelson MA. Pathogenic mechanisms in HBV- and HCV-associated hepatocellular carcinoma. *Nat Rev Cancer.* 2013; 13(2):123–35. <https://doi.org/10.1038/nrc3449> PMID: [23344543](#).
6. Samal J, Kandpal M, Vivekanandan P. Molecular mechanisms underlying occult hepatitis B virus infection. *Clin Microbiol Rev.* 2012; 25(1):142–63. <https://doi.org/10.1128/CMR.00018-11> PMID: [22232374](#); PubMed Central PMCID: [PMC3255968](#).
7. Habig JW, Loeb DD. The conformation of the 3' end of the minus-strand DNA makes multiple contributions to template switches during plus-strand DNA synthesis of duck hepatitis B virus. *J Virol.* 2003; 77(23):12401–11. Epub 2003/11/12. <https://doi.org/10.1128/JVI.77.23.12401-12411.2003> PMID: [14610164](#); PubMed Central PMCID: [PMC262595](#).
8. Kwon H, Lok AS. Hepatitis B therapy. *Nat Rev Gastroenterol Hepatol.* 2011; 8(5):275–84. <https://doi.org/10.1038/nrgastro.2011.33> PMID: [21423260](#).
9. Wooddell CI, Yuen MF, Chan HL, Gish RG, Locarnini SA, Chavez D, et al. RNAi-based treatment of chronically infected patients and chimpanzees reveals that integrated hepatitis B virus DNA is a source of HBsAg. *Sci Transl Med.* 2017; 9(409). <https://doi.org/10.1126/scitranslmed.aan0241> PMID: [28954926](#).
10. Fujimoto A, Totoki Y, Abe T, Boroevich KA, Hosoda F, Nguyen HH, et al. Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat Genet.* 2012; 44(7):760–4. <https://doi.org/10.1038/ng.2291> PMID: [22634756](#).
11. Guichard C, Amadeo G, Imbeaud S, Ladeiro Y, Pelletier L, Maad IB, et al. Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. *Nat Genet.* 2012; 44(6):694–8. Epub 2012/05/09. <https://doi.org/10.1038/ng.2256> PMID: [22561517](#); PubMed Central PMCID: [PMC3819251](#).
12. Rao HY, Sun DG, Jiang D, Yang RF, Guo F, Wang JH, et al. IL28B genetic variants and gender are associated with spontaneous clearance of hepatitis C virus infection. *J Viral Hepat.* 2012; 19(3):173–81. <https://doi.org/10.1111/j.1365-2893.2011.01497.x> PMID: [22329371](#).
13. Kan Z, Zheng H, Liu X, Li S, Barber TD, Gong Z, et al. Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. *Genome Res.* 2013; 23(9):1422–33. Epub 2013/06/22. <https://doi.org/10.1101/gr.154492.113> PMID: [23788652](#); PubMed Central PMCID: [PMC3759719](#).
14. Sung WK, Zheng H, Li S, Chen R, Liu X, Li Y, et al. Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nat Genet.* 2012; 44(7):765–9. <https://doi.org/10.1038/ng.2295> PMID: [22634754](#).
15. Fujimoto A, Furuta M, Totoki Y, Tsunoda T, Kato M, Shiraishi Y, et al. Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. *Nat Genet.* 2016; 48(5):500–9. <https://doi.org/10.1038/ng.3547> PMID: [27064257](#).
16. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009; 25(14):1754–60. <https://doi.org/10.1093/bioinformatics/btp324> PMID: [19451168](#); PubMed Central PMCID: [PMC2705234](#).
17. Quinlan AR. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics.* 2014; 47:11.2.1–34. <https://doi.org/10.1002/0471250953.bi1112s47> PubMed Central PMCID: [PMC4213956](#). PMID: [25199790](#)
18. Mason WS, Gill US, Litwin S, Zhou Y, Peri S, Pop O, et al. HBV DNA Integration and Clonal Hepatocyte Expansion in Chronic Hepatitis B Patients Considered Immune Tolerant. *Gastroenterology.* 2016; 151(5):986–98 e4. Epub 2016/07/28. <https://doi.org/10.1053/j.gastro.2016.07.012> PMID: [27453547](#).
19. Seeger C, Mason WS. Molecular biology of hepatitis B virus infection. *Virology.* 2015; 479–480:672–86. <https://doi.org/10.1016/j.virol.2015.02.031> PubMed Central PMCID: [PMC4424072](#). PMID: [25759099](#)
20. Gish RG, Yuen MF, Chan HL, Given BD, Lai CL, Locarnini SA, et al. Synthetic RNAi triggers and their use in chronic hepatitis B therapies with curative intent. *Antiviral Res.* 2015; 121:97–108. <https://doi.org/10.1016/j.antiviral.2015.06.019> PMID: [26129970](#).
21. Matsui T, Kang JH, Kimura Y, Tanaka K, Tsuji K, Maguchi H, et al. Hepatitis B Core Related Antigen in Sera Is the Surrogate Marker for Hbv Covalently Closed Circular DNA. *Journal of Hepatology.* 2016; 64:S527–S8. WOS:000398715800116.

22. Karimova M, Beschorner N, Dammermann W, Chemnitz J, Indenbirken D, Bockmann JH, et al. CRISPR/Cas9 nickase-mediated disruption of hepatitis B virus open reading frame S and X. *Sci Rep.* 2015; 5:13734. <https://doi.org/10.1038/srep13734> PMID: 26334116; PubMed Central PMCID: PMC4558539.
23. Ramanan V, Shlomai A, Cox DB, Schwartz RE, Michailidis E, Bhatta A, et al. CRISPR/Cas9 cleavage of viral DNA efficiently suppresses hepatitis B virus. *Sci Rep.* 2015; 5:10833. <https://doi.org/10.1038/srep10833> PMID: 26035283; PubMed Central PMCID: PMC4649911.
24. Grupp SA, Kalos M, Barrett D, Aplenc R, Porter DL, Rheingold SR, et al. Chimeric antigen receptor-modified T cells for acute lymphoid leukemia. *N Engl J Med.* 2013; 368(16):1509–18. <https://doi.org/10.1056/NEJMoa1215134> PMID: 23527958; PubMed Central PMCID: PMC4058440.
25. Robbins PF, Morgan RA, Feldman SA, Yang JC, Sherry RM, Dudley ME, et al. Tumor regression in patients with metastatic synovial cell sarcoma and melanoma using genetically engineered lymphocytes reactive with NY-ESO-1. *J Clin Oncol.* 2011; 29(7):917–24. <https://doi.org/10.1200/JCO.2010.32.2537> PMID: 21282551; PubMed Central PMCID: PMC3068063.
26. European Association for the Study of the Liver. Electronic address eee, European Association for the Study of the L. EASL 2017 Clinical Practice Guidelines on the management of hepatitis B virus infection. *J Hepatol.* 2017; 67(2):370–98. <https://doi.org/10.1016/j.jhep.2017.03.021> PMID: 28427875.
27. Hoang JK, Yang HI, Le A, Nguyen NH, Lin D, Vu VD, et al. Lower liver cancer risk with antiviral therapy in chronic hepatitis B patients with normal to minimally elevated ALT and no cirrhosis. *Medicine (Baltimore).* 2016; 95(31):e4433. Epub 2016/08/09. <https://doi.org/10.1097/MD.0000000000004433> PMID: 27495067; PubMed Central PMCID: PMC4979821.