Communication

Hepatitis B Virus Integration into Transcriptionally Active Loci and HBV-Associated Hepatocellular Carcinoma

Maria Bousali [1] **and Timokratis Karamitros** [1,2,*]

1   Bioinformatics and Applied Genomics Unit, Department of Microbiology, Hellenic Pasteur Institute, 11521 Athens, Greece; mbousali@med.uoa.gr
2   Laboratory of Medical Microbiology, Department of Microbiology, Hellenic Pasteur Institute, 11521 Athens, Greece
*   Correspondence: tkaram@pasteur.gr; Tel.: +30-210-6478871

Abstract: Hepatitis B Virus (HBV) DNA integrations into the human genome are considered major causative factors to HBV-associated hepatocellular carcinoma development. In the present study, we investigated whether HBV preferentially integrates parts of its genome in specific genes and evaluated the contribution of the integrations in HCC development per gene. We applied dedicated in-house developed pipelines on all of the available HBV DNA integration data and performed a statistical analysis to identify genes that could be characterized as hotspots of integrations, along with the evaluation of their association with HBV-HCC. Our results suggest that 15 genes are recurrently affected by HBV integrations and they are significantly associated with HBV-HCC. Further studies that focus on HBV integrations disrupting these genes are mandatory in order to understand the role of HBV integrations in clonal advantage gain and oncogenesis promotion, as well as to determine whether inhibition of the HBV-disrupted genes can provide a therapy strategy for HBV-HCC.

Keywords: hepatitis B virus; viral integration; VIS; pathogen–host interactions; hepatocellular carcinoma; HBV-HCC

1. Introduction

The hepatitis B virus (HBV) is a 3200 bp length hepatotropic virus of the Hepadnaviridae family and is classified into Group VII in Baltimore's system as it combines a partially double-stranded DNA (ds-DNA) and virus-encoded reverse transcriptase. The "Out of Africa" model for the origin and expansion of HBV has been supported in previous studies [1,2] and the origin of HBV in *Homo sapiens* has been estimated at 34,100 (27,600–41,300) years ago [3].

It has been found, since the early 1980s, that HBV integrates its genome, or parts of it, into the human genomic DNA [4–7], while more recent findings suggest that HBV integrations occur in the infected hepatocytes during the early stages of natural acute infections [8–11]. Multiple molecular mechanisms have been proposed to be involved in the integration process [12], of which the most prevalent are the non-homologous end joining (NHEJ) and the microhomology-mediated end joining (MMEJ) [13,14].

HBV infection causes chronic hepatitis and, subsequently, liver cirrhosis, while it has been identified as one of the most important contributors of hepatocellular carcinoma (HCC), as ~50% of the cases are HBV-associated (HBV-HCC) [15]. Moreover, the integrated HBV DNA is a continual source of HBV subviral particles, and the HBsAg they contain chronically inhibits the innate and adaptive immune responses to HBV infection [14,16–20]. This is due to the change of the organization of the HBV genome from circular to linear after the integration, preventing the expression of the HBV core, the polymerase, the E-antigen, and, frequently, the large isoform form of HBsAg and X-antigen, as the pregenomic RNA is not produced [14,20].

Retroviral and infectious viral integration events induce chromosome changes, genome instability, changes in the expression of host genes, and oncogenes activation, as well as inflammation [21]. Thus, HBV integrations are critical factors in liver disease progression and HCC development, having been detected in more than 85% of HBV-HCC cases [4,22–24].

Previous independent studies suggest that genomic areas that refer to cellular genes are favored target sites for HBV integration [10], and, specifically, that there is a preference for HBV integration into transcriptional and chromosomal regulatory regions [25]. Moreover, it has been reported that known HCC driver genes, such as the *telomerase reverse transcriptase* (*TERT*), the *lysine methyltransferase 2B* (*KMT2B*—also known as *MLL4*), and the *cyclin E1* (*CCNE1*), are affected by HBV integrations in HBV-HCC cases [26].

In this study, we investigated whether HBV preferentially integrates its genome or parts of it into specific transcriptionally active genomic loci, and evaluated the contribution of each integration in HCC development. We applied dedicated in-house developed pipelines on all of the available HBV DNA integration sites data and performed statistical analyses to identify genes that could be characterized as hotspots of HBV integrations.
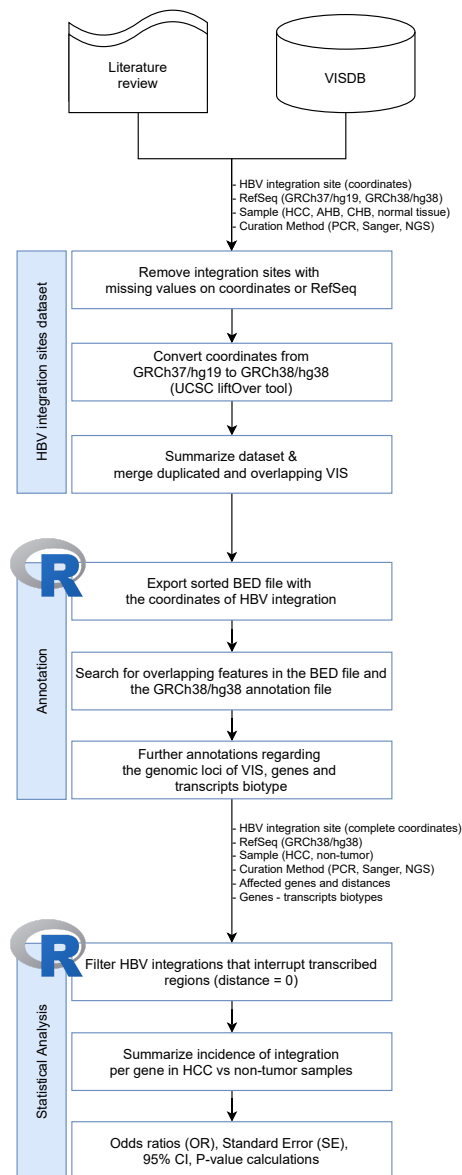
2. Methods

The coordinates of the human genome that have been reported to correspond to HBV DNA integration sites in human samples (excluding integrations reported to occur in cell lines), as well as metadata regarding the dataset characteristics, the originating samples (HCC, acute HBV infection—AHB, chronic HBV infection—CHB, or non-tumor adjacent tissue samples), and the curation method of the integration sites, were retrieved from the from 30 research articles [10,13,24,27–53] and the VISDB database [54]. In total, 20,153 HBV integration sites were evaluated (Figure 1). HBV integration sites that are registered to VISDB and have missing values in the start and the end positions of their genomic coordinates, or for which the human reference build is not reported, were removed, resulting in 19,600 integration sites.

The coordinates referring to the GRCh37/hg19 human genome reference (19,068/19,600), were converted to GRCh38/hg38 using the UCSC LiftOver tool [55]. From them, one coordinate was marked as a "lift over failure" as it has been deleted in GRCh38/hg38 reference build. After the conversion, 3267 coordinates were found to be duplicated or overlapping and were removed, resulting in 16,332 HBV integration sites, from which 16,313 are located in chromosomes 1-22, X, Y, and 19 in alternative loci scaffolds. These integrations were obtained from 4564 human samples, from which 2448 were HCC samples obtained from HCC tissues resected from patients undergoing primary hepatectomy and 2116 were non-tumor samples. The non-tumor samples category includes samples obtained either from matched adjacent tissues of the HCC samples or from liver biopsy specimens collected from patients with AHB and CHB.

The genomic loci referring to HBV integrations were annotated by in-house developed scripts written in R programming language [56], using the tidyverse [57] and the biomaRt [58,59] libraries.

Assuming that HBV integration is a random event and that the genomic locus of the integration depends on the size of each chromosome, the expected rate of HBV integration into each chromosome was calculated, as described by Minami et al. [60]. A binomial test was employed to analyze the frequency of integration per chromosome and the statistical significance was declared at a p-value < 0.01. Afterward, the annotated dataset was filtered in order to focus on integrations that interrupt transcriptionally active genomic regions (distance from gene = 0), and the incidence of integrations in HCC and non-tumor samples was summarized per gene. Gene ontology (GO) and other functional enrichment analyses (KEGG, REAC, CORUM) were performed with R programming language using the gprofiler2 library [61].

Figure 1. Multi-step workflow diagram of the methodology used to study the effect of HBV DNA integrations in transcribed genomic loci of the human genome (RefSeq GRCh38/hg38). Data regarding HBV DNA integration sites (coordinates) into human samples, the reference sequences used in the initial analysis, and the type of samples (HCC, AHB, CHB, normal tissue) used to export the viral integration sites (VIS), as well as wet-lab methodology information, were mined from 30 research papers and merged with additional metadata that are present in the VISDB database [54]. The dataset was curated: VIS coordinates for which both the start and end positions were absent, or the human reference build was not reported, were removed from the dataset, coordinates that referred to the GRCh37/hg19 human genome reference were converted to GRCh38/hg38 with the UCSC LiftOver tool [55], and duplicated and overlapping VIS were merged. The annotation of the VIS-disrupted genomic loci was performed with in-house developed scripts written in R language by applying a three-step approach that involves the generation of a sorted BED-formatted file with the curated VIS, the identification of overlapping regions between VIS and the GRCh38/hg38 annotation file, and the identification of the corresponding genes and transcripts biotypes. The annotated dataset was filtered in order to include only VIS that disrupt transcriptionally active regions, and the incidence of integrations in HCC and non-HCC samples was summarized per gene. The statistical analysis was based on per-gene calculations of the odds ratios (OR) between HCC and non-HCC samples; the standard error of OR, the 95% CI, and statistical significance was declared at a *p*-value < 0.01.

Odds ratios (OR) together with 95% confidence intervals (CI) were used to estimate the strength of association between the HBV DNA integrations and HCC, as previously described in [62–64]. In brief, the OR was calculated as the ratio between the odds of HBV integrations occurring in HCC samples and the odds of HBV integrations occurring in non-tumor samples. The standard error (SE) of the OR was calculated as the square root of the sum of the inverse of the four different observed states (HCC samples with integrations, non-tumor samples with integrations, HCC samples without integrations, and non-tumor samples without integrations). The 95% confidence intervals (CI) were calculated through the equation $e^{[log(OR)\pm1.96\times SE]}$, where OR is the odds ratio and SE is the standard error of the OR. For the calculation of the *p*-value from the 95% CI, the procedure described by Altman et al. [65] was followed. The statistical significance was declared at a *p*-value < 0.01.

3. Results and Discussion

The HBV DNA integrations evaluated in the present study were obtained from 4564 samples that have been analyzed in independent research studies, from which 2448 (53.46%) are HCC samples and 2116 (46.54%) are non-tumor tissue samples (Figure 2A,B).

Binomial tests were performed in order to evaluate if there is a preference or a deterrence for HBV integration to specific chromosomes (Table 1, Figure 2C). It was demonstrated that integrations in chromosomes 4, 8, 10, 12, 16, 17, and 19 were higher than anticipated and the overall frequencies were found to be statistically significant (*p* < 0.01). On the other hand, integrations in chromosomes 3, 6, 9, 13, 14, 15, and 22 were lower than anticipated and the overall frequencies were found to be statistically significant (*p* < 0.01). Previous independent studies have reported the preference for HBV integrations to chromosomes 3 [60], 11 [66], and 17 [33]. In this study, it is revealed that there is indeed a preference for integrations in chromosome 17 and a deterrence for integration for chromosome 3. These observations suggest that HBV integration is not a random event, as the genomic locus of the integration does not depend only on the size of each chromosome. As it has been previously suggested [60], selective pressure may be applied. Other factors, such as microhomology regions between the HBV and the human genome, along with the molecular mechanisms that initiate the integration process, should be further studied in order to gain more insights and explain the observed preference/deterrence of integrations into these chromosomes. Recent studies have reported that HBV frequently integrates into mobile genetic elements, such as transposons and retrotransposons, with a preference for the latter [11,67,68]. These findings form evidence for a (retro)transposon-directed integration process, which may have a potential role in the observed nonrandom HBV integration.

Figure 2D represents the biotypes of the genes that are disrupted from HBV integrations. From the 5053 transcribed genomic loci that are being disrupted from HBV integrations, 3601 (71.26%) are protein-coding genes, 1230 (24.34%) are long non-coding RNAs (lncRNA), 209 (4.14%) are pseudogenes, and 13 (0.25%) are non-coding RNAs (ncRNA).

Table 2 and Figure 3 present the results of the comparison of HBV DNA integration in HCC between the HCC and non-tumor groups. The overall results suggest that integrations into the *TERT*, the *KMT2B*, the *CCNA2*, the *SH3RF3*, the *CCNE1*, the *GLI2*, the *FAM157A*, and the *PCMTD2* genes are more common in HCC samples than in non-tumor samples (OR > 1, *p*-value < 0.01), while the opposite is true for the *FN1*, the *ALB*, the *LINC00486*, the *CPS1*, the *KCNT2*, the *GTF2I*, and the *CHRM3* genes (OR < 1, *p*-value < 0.01). Based on this evidence, integrations disrupting the *TERT* and the *KMT2B* genes are more likely to impact the HCC development. These two genes have been found as major genomic locations of recurrent HBV DNA integrations (present in 12.75% and 7.72% of the total HCC samples, respectively) and possibly have a positive association with HCC (OR > 1, *p*-value < 0.01).
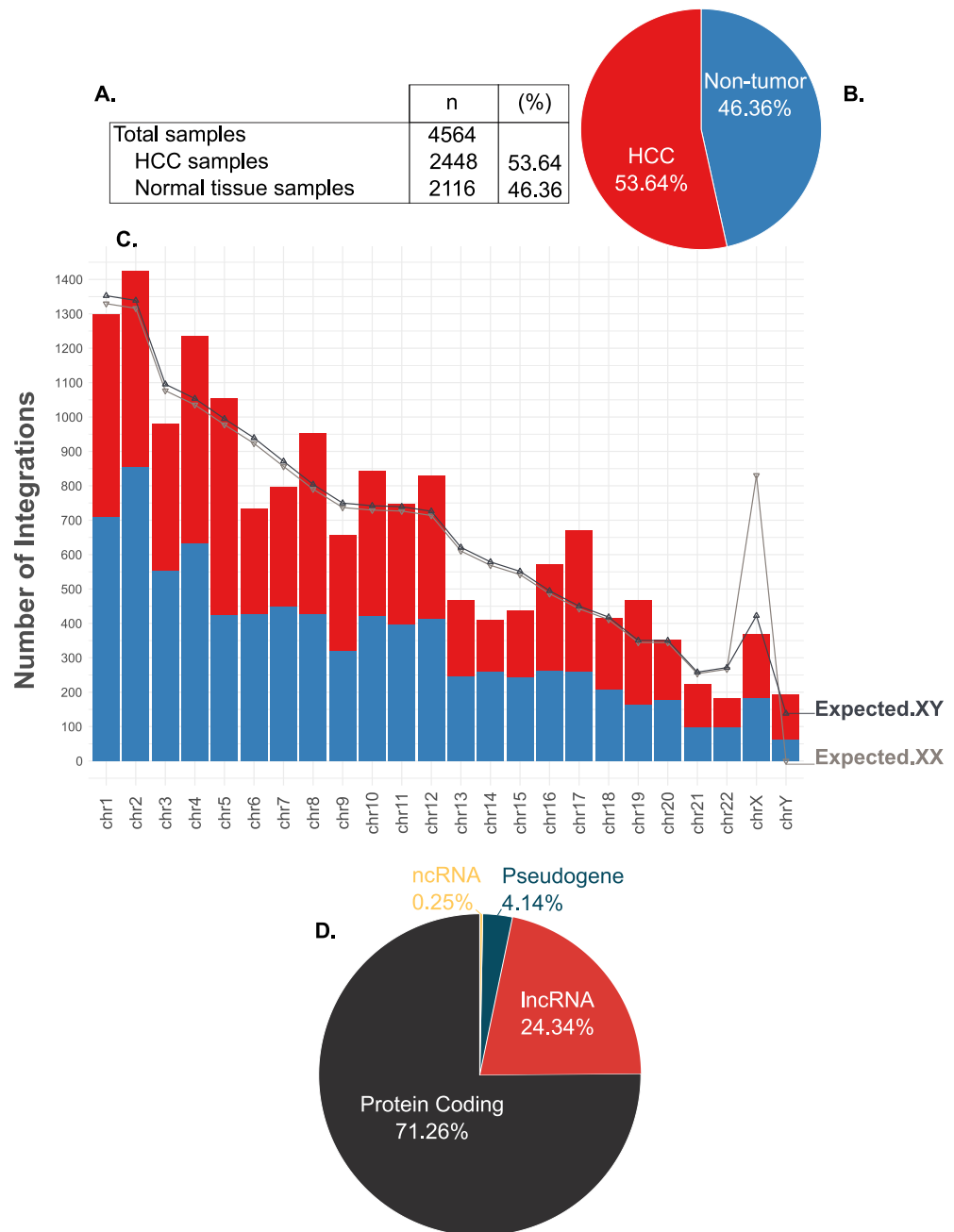
Figure 2. Graphical representation of the characteristics of the dataset used in the present study, as well as the findings obtained from the annotation of the HBV integration sites. (**A,B**) The total samples from which the HBV integrations are analyzed in the present study were obtained from previously published research papers and correspond to 2448 (53.64% of the total samples) HCC samples and 2116 (46.36% of the total samples) non-tumor tissue samples (total number of samples = 4564). (**C**) The distribution of HBV integrations in the human chromosomes (chr1–22, chrX, chrY) are depicted as bar charts. Moreover, the distribution of the HBV integrations in HCC and non-tumor samples are also depicted in the stacked bar charts, in red and blue, respectively. The expected values used in the binomial test, in which it was assumed that HBV integration is a random event and that the genomic locus of the integration depends on the size of each chromosome, are presented as lines with gray color and the (▽) symbol for the 46XX subjects and black color and the (△) symbol for the 46XY subjects. (**D**) Representation of the biotypes of the genes that are being disrupted from HBV integration. From the 5053 disrupted, transcribed genomic loci, 3601 (71.26%) are protein-coding genes, 1230 (24.34%) are long non-coding RNAs (lncRNA), 209 (4.14%) are pseudogenes, and 13 (0.25%) are non-coding RNAs (ncRNA).

Table 1. Chromosomal location of HBV integrations and evaluation of possible preference or deterrence to specific chromosomes by binomial tests ($p < 0.01$), assuming that HBV integration is a random event and that the genomic loci of the integration depend on the size of each chromosome.

chr	Size (Mb)	Frequency	Expected.XX	Expected.XY	*p*-Value	Explanation
chr1	246.1	1299	1329.52	1352.68	0.146	insignificant
chr2	243.6	1425	1316.02	1338.94	0.019	insignificant
chr3	199.3	981	1076.69	1095.45	0	deterrence
chr4	191.7	1236	1035.63	1053.68	0	preference
chr5	181.0	1054	977.83	994.86	0.061	insignificant
chr6	170.9	732	923.27	939.35	0	deterrence
chr7	158.5	797	856.28	871.19	0.012	insignificant
chr8	146.3	953	790.37	804.13	0	preference
chr9	136.4	657	736.88	749.72	0.001	deterrence
chr10	135.0	843	729.32	742.02	0	preference
chr11	134.5	748	726.62	739.28	0.741	insignificant
chr12	132.1	830	713.65	726.08	0	preference
chr13	113.0	467	610.47	621.10	0	deterrence
chr14	105.3	411	568.87	578.78	0	deterrence
chr15	100.3	437	541.86	551.30	0	deterrence
chr16	90.0	570	486.21	494.68	0.001	preference
chr17	81.9	671	442.45	450.16	0	preference
chr18	76.1	415	411.12	418.28	0.903	insignificant
chr19	63.8	467	344.67	350.68	0	preference
chr20	63.7	352	344.13	350.13	0.915	insignificant
chr21	47.0	223	253.91	258.33	0.027	insignificant
chr22	49.4	183	266.88	271.53	0	deterrence
chrX	153.7	369	830.34	422.40	0.009	deterrence
chrY	50.3	193	0.00	138.24	0	preference

The *TERT* gene contributes to the formation of the telomerase enzyme whose role is the maintenance of the telomeres—the "protectors" of the chromosomes from abnormal sticking together and degradation. In the hepatocytes, telomeres are progressively shortened because of cell division, and when they reach a critical length, they trigger the cell to stop division and undergo apoptosis. Moreover, the majority of the cancer cells (80–90%) activate the *TERT* gene in order to achieve immortality [69–71]. HBV integrations into the *TERT* gene are found in 12.75% of the HCC samples in the dataset. Thus, they should be further studied in order to determine the possible role of integrations in the gain of clonal advantage that results in HBV-HCC. Further studies will also help in clarifying whether the inhibition of the HBV-disrupted *TERT* can provide a therapy strategy for HBV-HCC, as it has been suggested for a wide spectrum of cancers [72,73]. Along with *TERT*, the *CCNA2* and the *CCNE1* genes are involved in the cell cycle G1/S transition (GO:0044843). Misregulation of this biological process has been associated with oncogenesis promotion [74]. The two cyclins, *CCNA2* and *CCNE1*, have been found as major genomic locations of recurrent HBV DNA integrations (present in 0.79% and 0.59% of the total samples, respectively), and integrations that disrupt them possibly have a positive association with HCC (OR = 3.049 and 3.829, respectively, *p*-value < 0.01). These results are in concordance with previous studies that mention that the *TERT* and *CCNE1* have oncogenetic or tumor-suppressing functions [13].

KMT2B has been described as an oncogene, as well as a tumor-suppressor gene, that helps in the prevention of uncontrolled cell growth and division [75]. HBV DNA integrations into it are present in 7.72% of the HCC samples in the dataset and there is a possible positive association with HCC (OR = 3.388, *p*-value < 0.01).

On the other hand, *FN1* and *ALB* genes that are recurrently affected by HBV integrations (6.09% and 5.26% of the total samples in the dataset) have been found to be disrupted

in 1.18% and 0.74% of HCC samples versus 11.77% and 2.46% of the non-tumor samples and they are possibly negatively associated with HCC (OR = 0.09 and 0.294, respectively, *p*-value < 0.01). The observation that FN1 is recurrently affected by HBV integration in non-tumor samples is in concordance with previous findings [31]. A reasonable explanation could be based on the fact that *FN1* has been described as an oncogene, but this has not been evaluated. HBV integrations that disrupt the *FN1* and *ALB* genes possibly do not endow the carrier hepatocytes with a selective growth advantage, but this hypothesis should, also, be further investigated. Although, due to the fact that no follow-up was performed in the cases from which non-tumor liver sample was taken for the analysis, it cannot be ruled out that these integrations may be a prognostic factor for HCC, as discussed below.
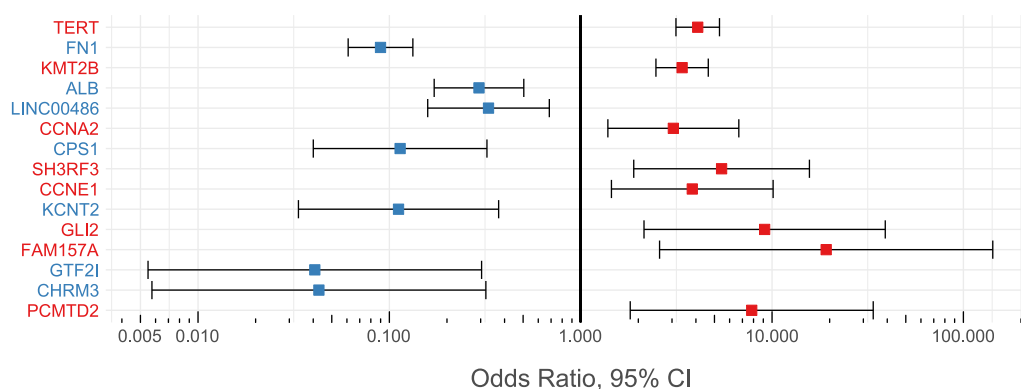
Figure 3. Odds ratios (OR) and 95% confidence intervals (95% CI) obtained by comparing the HBV integrations between the HCC and non-tumor groups in 15 human genes, represented as a $\log_{10}$-scaled plot.

At the same time, HBV integrations in four recurrently disrupted genes [12], *LINGO2*, *PRKN*, *SOX5*, and *PDE3A*, were not found to be statistically significant (*p*-value > 0.01). In brief, the *LINGO2* gene was disrupted from HBV integrations in 25 samples (0.55% of the total samples in the dataset), from which 13 were HCC samples (0.53% of the total HCC samples) and 12 were non-tumor samples (0.57% of the total non-tumor samples). The *PRKN* gene was disrupted in 23 samples (0.55%), from which 14 were HCC samples (0.57%) and 9 were non-tumor samples (0.43%). The *SOX5* gene was disrupted in 22 samples (0.48%), from which 7 were HCC samples (0.29%) and 15 were non-tumor samples (0.71%), while the *PDE3A* gene was disrupted in 22 samples (0.48%), from which 14 were HCC samples (0.57%) and 8 were non-tumor samples (0.38%).

Table 2. Association between HBV DNA integrations and HCC in 15 target human genes.

Gene Symbol	Gene Type	External Link	Genomic Location	#Samples with VIS (%Total Samples)	#HCC (%Total HCC)	#Non-Tumor (%Total Non-Tumor)	Odds Ratio	95% CI	*p*-Value
TERT	pc*	HGNC:11730	5p15.33	385 (8.44)	312 (12.75)	73 (3.45)	4.088	3.146–5.312	<0.01
FN1	pc*	HGNC:3778	2q35	278 (6.09)	29 (1.18)	249 (11.77)	0.09	0.061–0.133	<0.01
KMT2B	pc*	HGNC:15840	19q13.12	240 (5.26)	189 (7.72)	51 (2.41)	3.388	2.472–4.642	<0.01
ALB	pc*	HGNC:399	4q13.3	70 (1.53)	18 (0.74)	52 (2.46)	0.294	0.171–0.504	<0.01
CCNA2	pc*	HGNC:1578	4q27	36 (0.79)	28 (1.14)	8 (0.38)	3.049	1.387–6.704	<0.01
LINC00486	lncRNA	HGNC:42946	2p22.3	36 (0.79)	10 (0.41)	26 (1.23)	0.33	0.159–0.685	<0.01
CPS1	pc*	HGNC:2323	2q34	34 (0.74)	4 (0.16)	30 (1.42)	0.114	0.04–0.324	<0.01
SH3RF3	pc*	HGNC:24699	2q13	29 (0.64)	25 (1.02)	4 (0.19)	5.448	1.893–15.679	<0.01
CCNE1	pc*	HGNC:1589	19q12	27 (0.59)	22 (0.9)	5 (0.24)	3.829	1.447–10.128	<0.01
KCNT2	pc*	HGNC:18866	1q31.3	26 (0.57)	3 (0.12)	23 (1.09)	0.112	0.033–0.372	<0.01
GLI2	pc*	HGNC:4318	2q14.2	23 (0.5)	21 (0.86)	2 (0.09)	9.146	2.142–39.052	<0.01
FAM157A	lncRNA	HGNC:34079	3q29	23 (0.5)	22 (0.9)	1 (0.05)	19.18	2.583–142.416	<0.01
GTF2I	pc*	HGNC:4659	7q11.23	22 (0.48)	1 (0.04)	21 (0.99)	0.041	0.005–0.303	<0.01
CHRM3	pc*	HGNC:1952	1q43	21 (0.46)	1 (0.04)	20 (0.95)	0.043	0.006–0.319	<0.01
PCMTD2	pc*	HGNC:15882	20q13.33	20 (0.44)	18 (0.74)	2 (0.09)	7.83	1.815–33.784	<0.01

*pc = protein-coding.

To conclude, the present study provides new insights into the impact of HBV integrations that disrupt transcribed genomic loci on HCC development, by combining all of the available knowledge obtained from independent published research studies. The findings provided suggest that 15 genes are recurrently affected by HBV integrations and they are significantly associated with HBV-HCC, and thus may represent HCC driver genes that merit further investigation as to their possible role in therapeutical targets for HBV-HCC and molecular profiling for precision medicine. At the same time, HBV integration was found to be a non-random event, thus selective pressure may be applied, while the role of microhomologies between HBV and the human genome, as well as the molecular mechanisms initiating the integration process—including the arising evidence for (retro)transposon-mediated integration—are matters demanding further investigation. The present study included manually curated integration sites, as are those provided by the VISDB database [54], thus the biases occurring from using data from independent research studies, whose results have been exported from different methodologies, were minimized. A systematic review regarding the HBV DNA integrations has previously been conducted [12] and all available data on a topic have been summarized (reviewer selection bias minimization). It should be noted though, that cases with CHB or AHB from which samples were taken and analyzed were not followed up. Thus, integrations that are found more frequently in non-tumor samples and are not associated with tumorigenesis warrant further studies investigating their role and whether they can be used as a prognostic risk factor of tumor development. Moreover, other factors, such as the alanine aminotransferase (ALT) levels, the serum HBV DNA levels, the circulating cell-free DNA (cfDNA) [76], the level of liver fibrosis, the heterogeneity of the tumor samples, and other genomic polymorphisms [77] that could be possibly found in the samples, whether cases underwent antiviral therapy before the resection surgery [78], or the presence of anti-HBe, anti-HBs, or anti-HBc, may be confusing factors of the observed associations and should be co-evaluated in future prospective studies on HBV VIS. Further studies are mandatory to better understand the association of HBV integrations with HBV-HCC, the molecular mechanisms that mediate in the integration process, and to define the possibility to prevent or treat HCC induced by HBV DNA integrations.

Abbreviations

The following abbreviations are used in this manuscript:

HBV	Hepatitis B virus
AHB	Acute HBV infection
CHB	Chronic HBV infection
HCC	Hepatocellular carcinoma
HBV-HCC	HBV-associated hepatocellular carcinoma
HBeAg	Hepatitis B e antigen
HBcAg	Hepatitis B core antigen

HBsAg	Hepatitis B surface antigen
anti-HBs	Hepatitis B surface antibody
anti-HBc	Total hepatitis B core antibody
IgM anti-HBc	IgM antibody to hepatitis B core antigen

References

1. Paraskevis, D.; Magiorkinis, G.; Magiorkinis, E.; Ho, S.Y.; Belshaw, R.; Allain, J.P.; Hatzakis, A. Dating the origin and dispersal of hepatitis B virus infection in humans and primates. *Hepatology* **2013**, *57*, 908–916. [CrossRef] [PubMed]
2. Kostaki, E.G.; Karamitros, T.; Stefanou, G.; Mamais, I.; Angelis, K.; Hatzakis, A.; Kramvis, A.; Paraskevis, D. Unravelling the history of hepatitis B virus genotypes A and D infection using a full-genome phylogenetic and phylogeographic approach. *Elife* **2018**, *7*, e36709. [CrossRef] [PubMed]
3. Paraskevis, D.; Angelis, K.; Magiorkinis, G.; Kostaki, E.; Ho, S.Y.; Hatzakis, A. Dating the origin of hepatitis B virus reveals higher substitution rate and adaptation on the branch leading to F/H genotypes. *Mol. Phylogenet Evol.* **2015**, *93*, 44–54. [CrossRef] [PubMed]
4. Shafritz, D.A.; Shouval, D.; Sherman, H.I.; Hadziyannis, S.J.; Kew, M.C. Integration of Hepatitis B Virus DNA into the Genome of Liver Cells in Chronic Liver Disease and Hepatocellular Carcinoma. *N. Engl. J. Med.* **1981**, *305*, 1067–1073. [CrossRef]
5. Koshy, R.; Koch, S.; Freytag von Loringhoven, A.; Kahmann, R.; Murray, K.; Hofschneider, P.H. Integration of hepatitis B virus DNA: Evidence for integration in the single-stranded gap. *Cell* **1983**, *34*, 215–223. [CrossRef]
6. Brechot, C.; Pourcel, C.; Louise, A.; Rain, B.; Tiollais, P. Presence of integrated hepatitis B virus DNA sequences in cellular DNA of human hepatocellular carcinoma. *Nature* **1980**, *286*, 533–535. [CrossRef]
7. Chakraborty, P.R.; Ruiz-Opazo, N.; Shouval, D.; Shafritz, D.A. Identification of integrated hepatitis B virus DNA and expression of viral RNA in an HBsAg-producing human hepatocellular carcinoma cell line. *Nature* **1980**, *286*, 531–533. [CrossRef]
8. Tarocchi, M.; Polvani, S.; Marroncini, G.; Galli, A. Molecular mechanism of hepatitis B virus-induced hepatocarcinogenesis. *World J. Gastroenterol.* **2014**, *20*, 11630–11640. [CrossRef]
9. Murakami, Y.; Saigo, K.; Takashima, H.; Minami, M.; Okanoue, T.; Bréchet, C.; Paterlini-Bréchot, P. Large scaled analysis of hepatitis B virus (HBV) DNA integration in HBV related hepatocellular carcinomas. *Gut* **2005**, *54*, 1162–1168. [CrossRef]
10. Jiang, S.; Yang, Z.; Li, W.; Li, X.; Wang, Y.; Zhang, J.; Xu, C.; Chen, P.J.; Hou, J.; McCrae, M.A.; et al. Re-evaluation of the Carcinogenic Significance of Hepatitis B Virus Integration in Hepatocarcinogenesis. *PLoS ONE* **2012**, *7*, e40363. [CrossRef]
11. Chauhan, R.; Shimizu, Y.; Watashi, K.; Wakita, T.; Fukasawa, M.; Michalak, T.I. Retrotransposon elements among initial sites of hepatitis B virus integration into human genome in the HepG2-NTCP cell infection model. *Cancer Genet.* **2019**, *235–236*, 39–56. [CrossRef] [PubMed]
12. Bousali, M.; Papatheodoridis, G.; Paraskevis, D.; Karamitros, T. Hepatitis B Virus DNA Integration, Chronic Infections and Hepatocellular Carcinoma. *Microorganisms* **2021**, *9*, 1787. [CrossRef] [PubMed]
13. Zhao, L.H.; Liu, X.; Yan, H.X.; Li, W.Y.; Zeng, X.; Yang, Y.; Zhao, J.; Liu, S.P.; Zhuang, X.H.; Lin, C.; et al. Genomic and oncogenic preference of HBV integration in hepatocellular carcinoma. *Nat. Commun.* **2016**, *7*, 12992. [CrossRef] [PubMed]
14. Tu, T.; Budzinska, M.A.; Shackel, N.A.; Urban, S. HBV DNA Integration: Molecular Mechanisms and Clinical Implications. *Viruses* **2017**, *9*, 75. [CrossRef]
15. Akinyemiju, T.; Abera, S.; Ahmed, M.; Alam, N.; Alemayohu, M.A.; Allen, C.; Al-Raddadi, R.; Alvis-Guzman, N.; Amoako, Y.; Artaman, A.; et al. The Burden of Primary Liver Cancer and Underlying Etiologies From 1990 to 2015 at the Global, Regional, and National Level: Results From the Global Burden of Disease Study 2015. *JAMA Oncol.* **2017**, *3*, 1683–1691. [CrossRef] [PubMed]
16. Lindh, M.; Rydell, G.E.; Larsson, S.B. Impact of integrated viral DNA on the goal to clear hepatitis B surface antigen with different therapeutic strategies. *Curr. Opin. Virol.* **2018**, *30*, 24–31. [CrossRef]
17. Saitta, C.; Tripodi, G.; Barbera, A.; Bertuccio, A.; Smedile, A.; Ciancio, A.; Raffa, G.; Sangiovanni, A.; Navarra, G.; Raimondo, G.; et al. Hepatitis B virus (HBV) DNA integration in patients with occult HBV infection and hepatocellular carcinoma. *Liver Int.* **2015**, *35*, 2311–2317. [CrossRef] [PubMed]
18. Peiffer, K.H.; Kuhnhenn, L.; Jiang, B.; Mondorf, A.; Vermehren, J.; Knop, V.; Susser, S.; Walter, D.; Dietz, J.; Carra, G.; et al. Divergent preS Sequences in Virion-Associated Hepatitis B Virus Genomes and Subviral HBV Surface Antigen Particles From HBV e Antigen-Negative Patients. *J. Infect. Dis.* **2018**, *218*, 114–123. [CrossRef]
19. Ishii, T.; Tamura, A.; Shibata, T.; Kuroda, K.; Kanda, T.; Sugiyama, M.; Mizokami, M.; Moriyama, M. Analysis of HBV Genomes Integrated into the Genomes of Human Hepatoma PLC/PRF/5 Cells by HBV Sequence Capture-Based Next-Generation Sequencing. *Genes* **2020**, *11*, 661. [CrossRef]
20. Vaillant, A. HBsAg, Subviral Particles, and Their Clearance in Establishing a Functional Cure of Chronic Hepatitis B Virus Infection. *ACS Infect. Dis.* **2021**, *7*, 1351–1368. [CrossRef]
21. Karamitros, T.; Hurst, T.; Marchi, E.; Karamichali, E.; Georgopoulou, U.; Mentis, A.; Riepsaame, J.; Lin, A.; Paraskevis, D.; Hatzakis, A.; et al. is associated with intravenous drug abuse and modulates transcription in a cell-line model. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 10434–10439. [CrossRef] [PubMed]
22. Matsubara, K.; Tokino, T. Integration of hepatitis B virus DNA and its implications for hepatocarcinogenesis. *Mol. Biol. Med.* **1990**, *7*, 243–260. [PubMed]

23. Guerrero, R.B.; Roberts, L.R. The role of hepatitis B virus integrations in the pathogenesis of human hepatocellular carcinoma. *J. Hepatol.* **2005**, *42*, 760–777. [CrossRef]

24. Yan, H.; Yang, Y.; Zhang, L.; Tang, G.; Wang, Y.; Xue, G.; Zhou, W.; Sun, S. Characterization of the genotype and integration patterns of hepatitis B virus in early- and late-onset hepatocellular carcinoma. *Hepatology* **2015**, *61*, 1821–1831. [CrossRef] [PubMed]

25. Horikawa, I.; Barrett, J.C. Transcriptional regulation of the telomerase hTERT gene as a target for cellular and viral oncogenic mechanisms. *Carcinogenesis* **2003**, *24*, 1167–1176. [CrossRef] [PubMed]

26. Li, X.; Zhang, J.; Yang, Z.; Kang, J.; Jiang, S.; Zhang, T.; Chen, T.; Li, M.; Lv, Q.; Chen, X.; et al. The function of targeted host genes determines the oncogenicity of HBV integration in hepatocellular carcinoma. *J. Hepatol.* **2014**, *60*, 975–984. [CrossRef]

27. Gozuacik, D.; Murakami, Y.; Saigo, K.; Chami, M.; Mugnier, C.; Lagorce, D.; Okanoue, T.; Urashima, T.; Bréchot, C.; Paterlini-Bréchot, P. Identification of human cancer-related genes by naturally occurring Hepatitis B Virus DNA tagging. *Oncogene* **2001**, *20*, 6233–6240. [CrossRef]

28. Paterlini-Bréchot, P.; Saigo, K.; Murakami, Y.; Chami, M.; Gozuacik, D.; Mugnier, C.; Lagorce, D.; Bréchot, C. Hepatitis B virus-related insertional mutagenesis occurs frequently in human liver cancers and recurrently targets human telomerase gene. *Oncogene* **2003**, *22*, 3911–3916. [CrossRef]

29. Mason, W.S.; Liu, C.; Aldrich, C.E.; Litwin, S.; Yeh, M.M. Clonal expansion of normal-appearing human hepatocytes during chronic hepatitis B virus infection. *J. Virol.* **2010**, *84*, 8308–8315. [CrossRef]

30. Jiang, Z.; Jhunjhunwala, S.; Liu, J.; Haverty, P.M.; Kennemer, M.I.; Guan, Y.; Lee, W.; Carnevali, P.; Stinson, J.; Johnson, S.; et al. The effects of hepatitis B virus integration into the genomes of hepatocellular carcinoma patients. *Genome Res.* **2012**, *22*, 593–601. [CrossRef]

31. Sung, W.K.; Zheng, H.; Li, S.; Chen, R.; Liu, X.; Li, Y.; Lee, N.P.; Lee, W.H.; Ariyaratne, P.N.; Tennakoon, C.; et al. Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nat. Genet.* **2012**, *44*, 765–769. [CrossRef] [PubMed]

32. Fujimoto, A.; Totoki, Y.; Abe, T.; Boroevich, K.A.; Hosoda, F.; Nguyen, H.H.; Aoki, M.; Hosono, N.; Kubo, M.; Miya, F.; et al. Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat. Genet.* **2012**, *44*, 760–764. [CrossRef] [PubMed]

33. Ding, D.; Lou, X.; Hua, D.; Yu, W.; Li, L.; Wang, J.; Gao, F.; Zhao, N.; Ren, G.; Li, L.; et al. Recurrent targeted genes of hepatitis B virus in the liver cancer genomes identified by a next-generation sequencing-based approach. *PLoS Genet.* **2012**, *8*, e1003065. [CrossRef] [PubMed]

34. Khoury, J.D.; Tannir, N.M.; Williams, M.D.; Chen, Y.; Yao, H.; Zhang, J.; Thompson, E.J.; Meric-Bernstam, F.; Medeiros, L.J.; Weinstein, J.N.; et al. Landscape of DNA virus associations across human malignant cancers: Analysis of 3775 cases using RNA-Seq. *J. Virol.* **2013**, *87*, 8916–8926. [CrossRef] [PubMed]

35. Li, W.; Zeng, X.; Lee, N.P.; Liu, X.; Chen, S.; Guo, B.; Yi, S.; Zhuang, X.; Chen, F.; Wang, G.; et al. HIVID: An efficient method to detect HBV integration using low coverage sequencing. *Genomics* **2013**, *102*, 338–344. [CrossRef] [PubMed]

36. Tang, K.W.; Alaei-Mahabadi, B.; Samuelsson, T.; Lindh, M.; Larsson, E. The landscape of viral expression and host gene fusion and adaptation in human cancer. *Nat. Commun.* **2013**, *4*, 2513. [CrossRef] [PubMed]

37. Lau, C.C.; Sun, T.; Ching, A.K.; He, M.; Li, J.W.; Wong, A.M.; Co, N.N.; Chan, A.W.; Li, P.S.; Lung, R.W.; et al. Viral-human chimeric transcript predisposes risk to liver cancer development and progression. *Cancer Cell* **2014**, *25*, 335–349. [CrossRef] [PubMed]

38. Fujimoto, A.; Furuta, M.; Shiraishi, Y.; Gotoh, K.; Kawakami, Y.; Arihiro, K.; Nakamura, T.; Ueno, M.; Ariizumi, S.; Nguyen, H.H.; et al. Whole-genome mutational landscape of liver cancers displaying biliary phenotype reveals hepatitis impact and molecular diversity. *Nat. Commun.* **2015**, *6*, 6120. [CrossRef]

39. Zhang, X.; Liu, S.; Shen, C.; Wu, Y.; Zhang, L.; Chen, X.; Lu, F. DNA methylation consistency implicates the primary tumor cell origin of recurrent hepatocellular carcinoma. *Epigenomics* **2015**, *7*, 581–592. [CrossRef]

40. Dong, H.; Zhang, L.; Qian, Z.; Zhu, X.; Zhu, G.; Chen, Y.; Xie, X.; Ye, Q.; Zang, J.; Ren, Z.; et al. Identification of HBV-MLL4 Integration and Its Molecular Basis in Chinese Hepatocellular Carcinoma. *PLoS ONE* **2015**, *10*, e0123175. [CrossRef]

41. Kawai-Kitahata, F.; Asahina, Y.; Tanaka, S.; Kakinuma, S.; Murakawa, M.; Nitta, S.; Watanabe, T.; Otani, S.; Taniguchi, M.; Goto, F.; et al. Comprehensive analyses of mutations and hepatitis B virus integration in hepatocellular carcinoma with clinicopathological features. *J. Gastroenterol.* **2016**, *51*, 473–486. [CrossRef] [PubMed]

42. Chiu, Y.T.; Wong, J.K.; Choi, S.W.; Sze, K.M.; Ho, D.W.; Chan, L.K.; Lee, J.M.; Man, K.; Cherny, S.; Yang, W.L.; et al. Novel pre-mRNA splicing of intronically integrated HBV generates oncogenic chimera in hepatocellular carcinoma. *J. Hepatol.* **2016**, *64*, 1256–1264. [CrossRef] [PubMed]

43. Fujimoto, A.; Furuta, M.; Totoki, Y.; Tsunoda, T.; Kato, M.; Shiraishi, Y.; Tanaka, H.; Taniguchi, H.; Kawakami, Y.; Ueno, M.; et al. Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. *Nat. Genet.* **2016**, *48*, 500–509. [CrossRef] [PubMed]

44. Cao, S.; Wendl, M.C.; Wyczalkowski, M.A.; Wylie, K.; Ye, K.; Jayasinghe, R.; Xie, M.; Wu, S.; Niu, B.; Grubb, R.; et al. Divergent viral presentation among human tumors and adjacent normal tissues. *Sci. Rep.* **2016**, *6*, 28294. [CrossRef] [PubMed]

45. Mason, W.S.; Gill, U.S.; Litwin, S.; Zhou, Y.; Peri, S.; Pop, O.; Hong, M.L.; Naik, S.; Quaglia, A.; Bertoletti, A.; et al. HBV DNA Integration and Clonal Hepatocyte Expansion in Chronic Hepatitis B Patients Considered Immune Tolerant. *Gastroenterology* **2016**, *151*, 986–998. [CrossRef] [PubMed]

46. Gao, Q.; Wang, Z.C.; Duan, M.; Lin, Y.H.; Zhou, X.Y.; Worthley, D.L.; Wang, X.Y.; Niu, G.; Xia, Y.; Deng, M.; et al. Cell Culture System for Analysis of Genetic Heterogeneity Within Hepatocellular Carcinomas and Response to Pharmacologic Agents. *Gastroenterology* **2017**, *152*, 232–242. [CrossRef] [PubMed]

47. Ally, A.; Balasundaram, M.; Carlsen, R.; Chuah, E.; Clarke, A.; Dhalla, N.; Holt, R.A.; Jones, S.J.M.; Lee, D.; Ma, Y.; et al. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. *Cell* **2017**, *169*, 1327–1341. [CrossRef]

48. Yoo, S.; Wang, W.; Wang, Q.; Fiel, M.I.; Lee, E.; Hiotis, S.P.; Zhu, J. A pilot systematic genomic comparison of recurrence risks of hepatitis B virus-associated hepatocellular carcinoma with low- and high-degree liver fibrosis. *BMC Med.* **2017**, *15*, 214. [CrossRef]

49. Furuta, M.; Tanaka, H.; Shiraishi, Y.; Unida, T.; Imamura, M.; Fujimoto, A.; Fujita, M.; Sasaki-Oku, A.; Maejima, K.; Nakano, K.; et al. Characterization of HBV integration patterns and timing in liver cancer and HBV-infected livers. *Oncotarget* **2018**, *9*, 25075–25088. [CrossRef]

50. Li, C.L.; Li, C.Y.; Lin, Y.Y.; Ho, M.C.; Chen, D.S.; Chen, P.J.; Yeh, S.H. Androgen Receptor Enhances Hepatic Telomerase Reverse Transcriptase Gene Transcription After Hepatitis B Virus Integration or Point Mutation in Promoter Region. *Hepatology* **2019**, *69*, 498–512. [CrossRef]

51. Yang, L.; Ye, S.; Zhao, X.; Ji, L.; Zhang, Y.; Zhou, P.; Sun, J.; Guan, Y.; Han, Y.; Ni, C.; et al. Molecular Characterization of HBV DNA Integration in Patients with Hepatitis and Hepatocellular Carcinoma. *J. Cancer* **2018**, *9*, 3225–3235. [CrossRef] [PubMed]

52. Bayard, Q.; Meunier, L.; Peneau, C.; Renault, V.; Shinde, J.; Nault, J.C.; Mami, I.; Couchy, G.; Amaddeo, G.; Tubacher, E.; et al. Cyclin A2/E1 activation defines a hepatocellular carcinoma subclass with a rearrangement signature of replication stress. *Nat. Commun.* **2018**, *9*, 5235. [CrossRef] [PubMed]

53. Ruan, P.; Dai, X.; Sun, J.; He, C.; Huang, C.; Zhou, R.; Cao, Z.; Ye, L. Different types of viral-host junction found in HBV integration breakpoints in HBV-infected patients. *Mol. Med. Rep.* **2019**, *19*, 1410–1416. [CrossRef] [PubMed]

54. Tang, D.; Li, B.; Xu, T.; Hu, R.; Tan, D.; Song, X.; Jia, P.; Zhao, Z. VISDB: A manually curated database of viral integration sites in the human genome. *Nucleic Acids Res.* **2019**, *48*, D633–D641. [CrossRef] [PubMed]

55. Hinrichs, A.S.; Karolchik, D.; Baertsch, R.; Barber, G.P.; Bejerano, G.; Clawson, H.; Diekhans, M.; Furey, T.S.; Harte, R.A.; Hsu, F.; et al. The UCSC Genome Browser Database: Update 2006. *Nucleic Acids Res.* **2006**, *34*, D590–D598. [CrossRef] [PubMed]

56. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2021.

57. Wickham, H.; Averick, M.; Bryan, J.; Chang, W.; McGowan, L.D.; François, R.; Grolemund, G.; Hayes, A.; Henry, L.; Hester, J.; et al. Welcome to the tidyverse. *J. Open Source Softw.* **2019**, *4*, 1686. [CrossRef]

58. Durinck, S.; Spellman, P.T.; Birney, E.; Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **2009**, *4*, 1184–1191. [CrossRef]

59. Durinck, S.; Moreau, Y.; Kasprzyk, A.; Davis, S.; De Moor, B.; Brazma, A.; Huber, W. BioMart and Bioconductor: A powerful link between biological databases and microarray data analysis. *Bioinformatics* **2005**, *21*, 3439–3440. [CrossRef]

60. Minami, M.; Daimon, Y.; Mori, K.; Takashima, H.; Nakajima, T.; Itoh, Y.; Okanoue, T. Hepatitis B virus-related insertional mutagenesis in chronic hepatitis B patients as an early drastic genetic change leading to hepatocarcinogenesis. *Oncogene* **2005**, *24*, 4340–4348. [CrossRef]

61. Kolberg, L.; Raudvere, U.; Kuzmin, I.; Vilo, J.; Peterson, H. gprofiler2—An R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Research* **2020**, *9*, ELIXIR-709. [CrossRef]

62. Szumilas, M. Explaining odds ratios. *J. Can. Acad. Child. Adolesc. Psychiatry* **2010**, *19*, 227–229. [PubMed]

63. Andrade, C. Understanding relative risk, odds ratio, and related terms: As simple as it can get. *J. Clin. Psychiatry* **2015**, *76*, e857–e861. [CrossRef] [PubMed]

64. Cummings, P. The relative merits of risk ratios and odds ratios. *Arch. Pediatr. Adolesc. Med.* **2009**, *163*, 438–445. [CrossRef] [PubMed]

65. Altman, D.G.; Bland, J.M. How to obtain the P value from a confidence interval. *BMJ* **2011**, *343*, d2304. [CrossRef] [PubMed]

66. Tokino, T.; Matsubara, K. Chromosomal sites for hepatitis B virus integration in human hepatocellular carcinoma. *J. Virol.* **1991**, *65*, 6761–6764. [CrossRef] [PubMed]

67. Chauhan, R.; Michalak, T.I. Earliest hepatitis B virus-hepatocyte genome integration: Sites, mechanism, and significance in carcinogenesis. *Hepatoma Res.* **2021**, *7*, 20. [CrossRef]

68. Chauhan, R.; Churchill, N.D.; Mulrooney-Cousins, P.M.; Michalak, T.I. Initial sites of hepadnavirus integration into host genome in human hepatocytes and in the woodchuck model of hepatitis B-associated hepatocellular carcinoma. *Oncogenesis* **2017**, *6*, e317. [CrossRef]

69. Shay, J.W.; Bacchetti, S. A survey of telomerase activity in human cancer. *Eur. J. Cancer* **1997**, *33*, 787–791. [CrossRef]

70. Kim, N.W.; Piatyszek, M.A.; Prowse, K.R.; Harley, C.B.; West, M.D.; Ho, P.L.; Coviello, G.M.; Wright, W.E.; Weinrich, S.L.; Shay, J.W. Specific association of human telomerase activity with immortal cells and cancer. *Science* **1994**, *266*, 2011–2015. [CrossRef]

71. Barthel, F.P.; Wei, W.; Tang, M.; Martinez-Ledesma, E.; Hu, X.; Amin, S.B.; Akdemir, K.C.; Seth, S.; Song, X.; Wang, Q.; et al. Systematic analysis of telomere length and somatic alterations in 31 cancer types. *Nat. Genet.* **2017**, *49*, 349–357. [CrossRef]

72. Man, R.J.; Chen, L.W.; Zhu, H.L. Telomerase inhibitors: A patent review (2010–2015). *Expert Opin. Ther. Pat.* **2016**, *26*, 679–688. [CrossRef] [PubMed]

73. Wen, L.; Zhao, C.; Song, J.; Ma, L.; Ruan, J.; Xia, X.; Chen, Y.E.; Zhang, J.; Ma, P.X.; Xu, J. CRISPR/Cas9-Mediated TERT Disruption in Cancer Cells. *Int. J. Mol. Sci.* **2020**, *21*, 653. [CrossRef] [PubMed]

74. Bertoli, C.; Skotheim, J.M.; De Bruin, R.A. Control of cell cycle transcription during G1 and S phases. *Nat. Rev. Mol. Cell Biol.* **2013**, *14*, 518–528. [CrossRef] [PubMed]

75. Guo, C.; Chen, L.H.; Huang, Y.; Chang, C.C.; Wang, P.; Pirozzi, C.J.; Qin, X.; Bao, X.; Greer, P.K.; McLendon, R.E.; et al. KMT2D maintains neoplastic cell proliferation and global histone H3 lysine 4 monomethylation. *Oncotarget* **2013**, *4*, 2144–2153. [CrossRef]

76. Papatheodoridi, A.; Chatzigeorgiou, A.; Chrysavgis, L.; Lembessis, P.; Loglio, A.; Facchetti, F.; Cholongitas, E.; Koutsilieris, M.; Lampertico, P.; Papatheodoridis, G. Circulating cell-free DNA species affect the risk of hepatocellular carcinoma in treated chronic hepatitis B patients. *J. Viral Hepat.* **2021**, *28*, 464–474. [CrossRef] [PubMed]

77. Karamitros, T.; Papatheodoridis, G.; Paraskevis, D.; Hatzakis, A.; Mbisa, J.L.; Georgopoulou, U.; Klenerman, P.; Magiorkinis, G. Impact of Interferon-α Receptor-1 Promoter Polymorphisms on the Transcriptome of the Hepatitis B Virus-Associated Hepatocellular Carcinoma. *Front. Immunol.* **2018**, *9*, 777. [CrossRef]

78. Papatheodoridi, M.; Papatheodoridis, G. Chronic Hepatitis B Virus Infection: Interferon Therapy and Long-Term Outcomes. In *Hepatitis B Virus and Liver Disease*; Kao, J.H., Ed.; Springer: Singapore, 2021; pp. 267–285. [CrossRef]