Recurrent HBV Integration Targets as Potential Drivers in Hepatocellular Carcinoma

Selena Y. Lin [1], Adam Zhang [2], Jessica Lian [2], Jeremy Wang [1], Ting-Tsung Chang [3], Yih-Jyh Lin [4], Wei Song [1] and Ying-Hsiu Su [2,*]

1   JBS Science, Inc., Doylestown, PA 18902, USA; slin@jbs-science.com (S.Y.L.); jwang336699@gmail.com (J.W.); fsong@jbs-science.com (W.S.)
2   The Baruch S. Blumberg Research Institute, Doylestown, PA 18902, USA; adamzhang0214@gmail.com (A.Z.); jessiellian@gmail.com (J.L.)
3   Department of Internal Medicine, National Cheng Kung University Medical College, Tainan 704, Taiwan; ttchang@mail.ncku.edu.tw
4   Department of Surgery, National Cheng Kung University Medical College, Tainan 704, Taiwan; lyj007@mail.ncku.edu.tw
*   Correspondence: ying-hsiu.su@bblumberg.org; Tel.: +215-489-4907

Abstract: Chronic hepatitis B virus (HBV) infection is the major etiology of hepatocellular carcinoma (HCC), frequently with HBV integrating into the host genome. HBV integration, found in 85% of HBV-associated HCC (HBV–HCC) tissue samples, has been suggested to be oncogenic. Here, we investigated the potential of HBV–HCC driver identification via the characterization of recurrently targeted genes (RTGs). A total of 18,596 HBV integration sites from our in-house study and others were analyzed. RTGs were identified by applying three criteria: at least two HCC subjects, reported by at least two studies, and the number of reporting studies. A total of 396 RTGs were identified. Among the 28 most frequent RTGs, defined as affected in at least 10 HCC patients, 23 (82%) were associated with carcinogenesis and 5 (18%) had no known function. Available breakpoint positions from the three most frequent RTGs, *TERT*, *MLL4/KMT2B*, and *PLEKHG4B*, were analyzed. Mutual exclusivity of TERT promoter mutation and HBV integration into TERT was observed. We present an RTG consensus through comprehensive analysis to enable the potential identification and discovery of HCC drivers for drug development and disease management.

Keywords: hepatitis B virus; hepatocellular carcinoma; HBV integration; HCC driver identification

1. Introduction

Hepatocellular carcinoma (HCC) is the second leading cause of cancer deaths worldwide [1–3], and its poor prognosis is, in part, due to the lack of effective treatment options. A major etiology of this multifactorial disease is chronic hepatitis B virus (HBV) infection, which is associated with approximately 50% of HCC cases worldwide [4]. Current HBV–HCC screening guidelines vary regionally: the American Association for the Study of Liver Disease (AASLD), Canadian Association for the Study of the Liver (CASL), European Association for the Study of the Liver (EASL), Asian-Pacific Association for the Study of the Liver (APASL), and Latin-American Association for the Study of the Liver (ALEH) all have distinctive recommendations. Irrespective of the region, ultrasound with or without serum AFP screening is recommended every 6 months in HBV populations with risk factors that may include being of Asian descent, >40 years of age, cirrhosis, and PAGE-B score [5].

During infection, HBV can integrate into the host genome. It has been proposed that integration events mostly occur through non-homologous end joining (NHEJ) and micro-homologous recombination [6–9]. Although HBV DNA integration into the host genome is considered rare—with an estimate of one integration event per 10,000 HBV-infected hepatocytes [10]—integrated viral DNA has been reported in more than 85% of HBV-related

HCC (HBV–HCC) cases, suggesting a significant association between HBV integration and hepatocarcinogenesis. Additionally, mechanisms of HBV integration in HCC carcinogenesis potentially vary from patient to patient, including the insertional mutagenesis of HCC-associated genes, induction of chromosomal instability, and continuous expression of viral proteins [11,12]. Therefore, understanding the impact of integrated HBV DNA on carcinogenesis may identify potential HCC driver genes as personalized biomarkers could paving the way for precision disease management in HBV–HCC patients.

With the advent of next-generation sequencing (NGS), thousands of HBV integration sites have been identified across the human genome. Although no host sequence preference or specificity [13–18] has been identified, integration can activate known HCC driver genes and has been reported in *TERT, CCNE1,* and *MLL4* [19]. These, and other frequently affected genes, have become known as recurrently targeted genes (RTGs). Similar to the approach of identifying *BRAF* V600E driver mutations as recurrent hotspot mutations, here, we take advantage of the large number of reported integration sites from the literature and our in-house study to test the hypothesis that characterization of frequent RTGs can be a tool for the identification of HBV–HCC drivers.

In this study, we collected integration sites identified in HCC tumor tissue and defined RTGs. A total of 18,596 HBV integration sites, detected using PCR- and NGS-based approaches [5,12–35], were analyzed. By characterizing the most frequent RTGs, we demonstrate the potential of identifying HCC drivers for HCC precision medicine and drug development.

2. Materials and Methods

2.1. Data Mining/Search Strategy

We searched the PubMed (1 January 2000–1 December 2020) database using Medical Subject Heading (MeSH) terms "hepatitis B virus", "HBV integration", and "hepatitis B integration sites" to identify studies that have reported HBV integration sites by either NGS- or PCR-based approaches. Additional studies were obtained by cross-referencing from the literature. We included only English-language studies with HCC subjects. We included all such studies that identified HBV integration sites using NGS-based approaches. Among studies using PCR-based methods, we only included those with sample sizes of 10 or more HCC patients. HBV integration sites identified by RNA-seq [7,8,20] were not included, because the expression of integrated sequences can be regulated by many host cellular factors and is not within the scope of this study. We filtered out repeated integration sites to ensure that each integration site was included only once, except for two studies that utilized different methods on overlapping samples [21,22]. A total of 33 reported studies in addition to our study are included, as summarized in Table S1.

2.2. In-House HCC Specimens and HBV Integration Analysis

Archived FFPE tumor tissue DNA (Table S2) [23,24] from stage I–IIIB patients (*n* = 32) was obtained from the National Cheng-Kung University Medical Center, Taiwan, and collected following the guidelines of the Institutional Review Board. An HBV enrichment NGS assay (JBS Science, Doylestown, PA, USA) was used to capture integration sites. Briefly, NGS libraries were generated and enriched for HBV sequences, and the enriched libraries were sequenced on the Illumina MiSeq platform (Penn State Hershey Genomics Sciences Facility at Penn State College of Medicine, Hershey, PA, USA). The NGS data were analyzed using ChimericSeq [25] to identify HBV–host junction sequences. Reference genomes NC_003977.1 (HBV) and GRCh38.p2 (human) were used. The closest genes within 150 kb of the HBV integration breakpoint were identified with ChimericSeq software. Genes were defined as in NCBI's RefSeq database. Tailored junction-specific PCR-Sanger sequencing assays were designed to validate each HBV integration site of interest, identified by the HBV enrichment NGS assay.

2.3. TERT Promoter Mutation Analysis by PCR-Sanger Sequencing

HCC tissue DNA was used to amplify a 163 bp region (Chr5:1,295,151–1,295,313) of the TERT promoter by HotStart Plus Taq Polymerase (Qiagen, Valencia, CA, USA) with forward primer 5′-CAGCGCTGCCTGAAACTC-3′ and reverse primer 5′-GTCCTGCCCCTTCACCTT-3′. The PCR products were sequenced at the NAPCore Facility at the Children's Hospital of Philadelphia (Philadelphia, PA, USA) and analyzed using the ClustalW software [26].

2.4. Identification of RTGs

To identify host genes that may have been affected by HBV DNA integration across all studies, we identified the closest gene within 150 kb of the integration event, the reported distance within which host genes can be impacted by integration [27,28]. For a gene to be classified as an RTG, it must have been identified as an HBV integration locus in two or more independent studies from at least two independent laboratories, each of which included at least two HCC patients with the gene affected in the tumor tissue. The requirement that a gene must have been identified in multiple laboratories served to minimize the issue of potential cross-contamination. The full list of identified RTGs can be provided upon request.

2.5. Gene Functional Enrichment Pathway Analysis

A total of 396 RTGs were subjected to enrichment pathway analysis using Enrichr (http://amp.pharm.mssm.edu/Enrichr (accessed on 30 January 2021)), to identify significantly ($p < 0.05$) enriched pathways as determined by gene ontology.

3. Results

3.1. Overview of the Studies for RTG Identification

The studies included in RTG identification are summarized in Table 1. Twenty-three studies utilized NGS, including our in-house study (Table S1), and 10 studies utilized PCR-based approaches for HBV integration identification. For each study, the sample size and the number and percentage of HCC tumors that had detectable integration sites are listed. We compiled a total of 18,596 HBV integration sites from tumor tissues of a combined 1310 HCC patients. In all, we found that 76% of tumor tissues ($n = 1733$) contained detectable integration sites. In the 12 studies that enriched for the whole HBV genome, on average, 72% (range 54–100%) of the tumors examined were found to have integrated HBV DNA ($n = 880$). In two studies using an HBV DR1-2 enrichment NGS assay, 65% [29] and 91% (our study) of tumors examined were positive for integrated HBV DNA.

Table 1. Summary of HBV integration studies included in this analysis.

		Study	HCC Patients (n)	# of Subjects with Integrated DNA * (% of total)	# of Junctions * Identified in Subjects	Junction Sequence	Clinical Variables	Geographic Region
NGS-based	WGS	[21]	3	3 (100%)	15	Yes	Yes	NA[7]
		[30,31]	91[1]	64 (45%)	223	Yes	Yes	Japan
		[32,33]	81	76 (94%)	344	Yes	Yes	Hong Kong
		[34]	2	2 (100%)	5	Yes	Yes	China
		[22]	5[1,2]	5 (100%)	92	Yes	-	NA[7]
		[35]	5	5 (100%)	21	Yes	Yes	NA
	Whole HBV Genome	[36]	3	2 (67%)	11	Yes	Yes	China
		[37]	20	16 (80%)	424	Yes	Yes	China
		[38]	48	26 (54%)	57	-	-	Singapore
		[39]	60	51 (85%)	156	Yes	Yes	China
		[6]	426	344 (81%)	3486	Yes[3]	-	China
		[40]	49	28 (57%)	121	Yes	Yes	Japan
		[41]	40	35 (90%)	257	Yes	Yes	NA
		[42]	101	94 (93%)	510	Yes	Yes	Taiwan
		[43]	54	54 (100%)	2870	Yes	Yes	China
		[44]	1[5]	1 (100%)	2	Yes	Yes	NA
		[45]	95	85 (89%)	1563	Yes	Yes	Hong Kong
		[46]	243	72 (30%)	4347	Yes	Yes	Japan
		[47]	50	44 (88%)	289	Yes	Yes	Taiwan
	DR1-2	[29]	40	26 (65%)	42	Yes	Yes	China
		Our study	22	20 (91%)	27	Yes	Yes	Taiwan
PCR-based		[48]	13	2 (15%)	2	Yes	-	USA
		[49]	14	14 (100%)	14	Yes	Yes	China
		[50]	15	15 (100%)	15	-	Yes	NA
		[51]	60	55 (92%)	60	-	Yes	NA
		[52]	10	7 (70%)	8	Yes	Yes	Japan
		[53]	60	41 (68%)	101	Yes	Yes	China
		[54]	59[4]	45 (76%)	45	-	-	Italy
		[55]	15	9 (60%)	9	-	-	China
		[56]	18	18 (100%)	2083	-	Yes	China
		[57]	30	21 (70%)	1397[6]	-	-	China
		Total	1733	1310 (76%)	18,596			

[1] HBV (+) HCC cohorts only. [2] Three patients previously reported in Jiang 2012 [21] were removed. [3] Only human chromosome sequence position provided. [4] Cohorts of HBsAg (−)/occult (+) and HBsAg (+) HCC patients. [5] HBV capture sequencing of a single occult HBV infection (OBI) patient. [6] Five of the thirty HCC tissues were analyzed by HBV-targeted NGS. [7] Specimens were purchased commercially from vendors such as Seracare LifeSciences, ProteoGenex, or Indivumed. * HBV DNA integration sites. WGS, whole-genome next-generation sequencing; Whole HBV genome, whole HBV genome enrichment performed prior to NGS; DR1-2, HBV DR1-2 integration hotspot region enriched prior to NGS; -, not available.

3.2. Characteristics of HBV–HCC Patients with Integrated HBV DNA

To compare the general sociodemographic and clinical characteristics of the HBV–HCC population compiled in this study (age, gender, HBV genotype, and cirrhotic etiology (designated as "cirrhotic HCC")) to those previously reported in the literature [4,58,59], we divided our study population into two groups based on the detectability of integrated HBV DNA (Table 2). Analysis of each parameter was performed where available. Overall, there were no significant differences between the two groups (with integrated DNA detectable vs. not detectable), including in the sex ratio. Of the three reported HBV genotypes, genotype C was the most frequently reported in the integration-positive patient cohort (75%), while the cohort with no detectable integration had only two patients with a genotype reported (both C). Interestingly, only 47% (7/15) of the integration-negative patients had cirrhosis, which is much lower than the 70–90% range previously reported in the general HBV–HCC population [4]. In the integration-positive cohort, 62% of HCC tumors were derived from a cirrhotic liver, which is also below the previously reported range.

Table 2. Major sociodemographic and clinical features of the study subjects with and without detectable integrated HBV DNA in tumor tissue.

	Reported HBV–HCC Population [1]	Integrated HBV DNA	
		Not Detectable (*n* = 423)	Detectable (*n* = 1310)
Age (years)	NA	33–83	11–85
Range	55–65 ± NA	59.4 ± 13.0 (*n* = 40)	54.9 ± 11.6 (*n* = 359)
Avg. ± SD			
Gender (Total)		(*n* = 66)	(*n* = 635)
Male	NA	40	466
Female		15	169
Genotype (Total)		(*n* = 2)	(*n* = 84)
B		0	24 [2]
C	NA	2	76
D		0	1
Cirrhosis %	70–90% (*n* = NA)	46.7% (*n* = 7/15)	62.3% (*n* = 105/279)

[1] Characteristics of the HBV–HCC subjects obtained from previous reports [4,58,59]. [2] One patient contained a mix of HBV B and C genotypes. NA, not available; *n*, number of patients available for analysis; Avg. ± SD, average ± standard deviation.

Lastly, because it is known that HBV behavior differs in Asian and Western cohorts [60], the geographic region where patient specimens were collected is indicated for each study. The majority of studies (72.7%) collected patient specimens in Asian populations, while only 6.0% examined Western cohorts.

3.3. Recurrent Sites of HBV Integration

Next, we identified RTGs in the compiled HCC cohort using the criteria described in the Materials and Methods and then examined their associations with carcinogenesis from the published literature. Of the 18,596 integration sites examined, 2892 sites were found within 150 kb of gene-coding sequences, and 396 were RTGs. The 396 RTGs were found in 673 HCC patients, which was 51% of the integration-positive patients (*n* = 1310) or 39% of all HCC patients (*n* = 1733). To investigate the association of frequent RTGs with carcinogenesis, we analyzed the most frequently recurrent genes (n = 28), defined as found to contain HBV integrants in at least 10 HCC patients (Table 3). As expected, TERT and MLL4 were the two most recurrent genes. Interestingly, these 28 genes either have previously been associated with carcinogenesis (23/28, 82%), or have no known function (5/28, 18%). A full list of biological pathways associated with the 28 RTGs is detailed in Supplementary Table S3.

Next, the 396 RTGs were queried for significantly enriched gene ontology (GO) pathways using Enrichr [61]. The top enriched biological pathways were axon guidance ($p = 0.00007$) and positive regulation of Ras protein signal transduction ($p = 0.0001$), suggesting possible links with oncogenesis (Figure 1A). Heparin sulfate-glucosamine 3-sulfotransferase I (*HS3ST1*) activity was among the top enriched pathways from GO molecular functions (Figure 1B). Sulfotransferases have reported associations with carcinogenic activity, and *HS3ST1* in particular has been implicated in inflammation [62]. Finally, a search of the Drug Signatures Database (DSigbDB) identified trichostatin, which selectively inhibits class I and II histone deacetylases (HDACs), as the drug/compound related to most RTGs (115/396; Figure 1C).
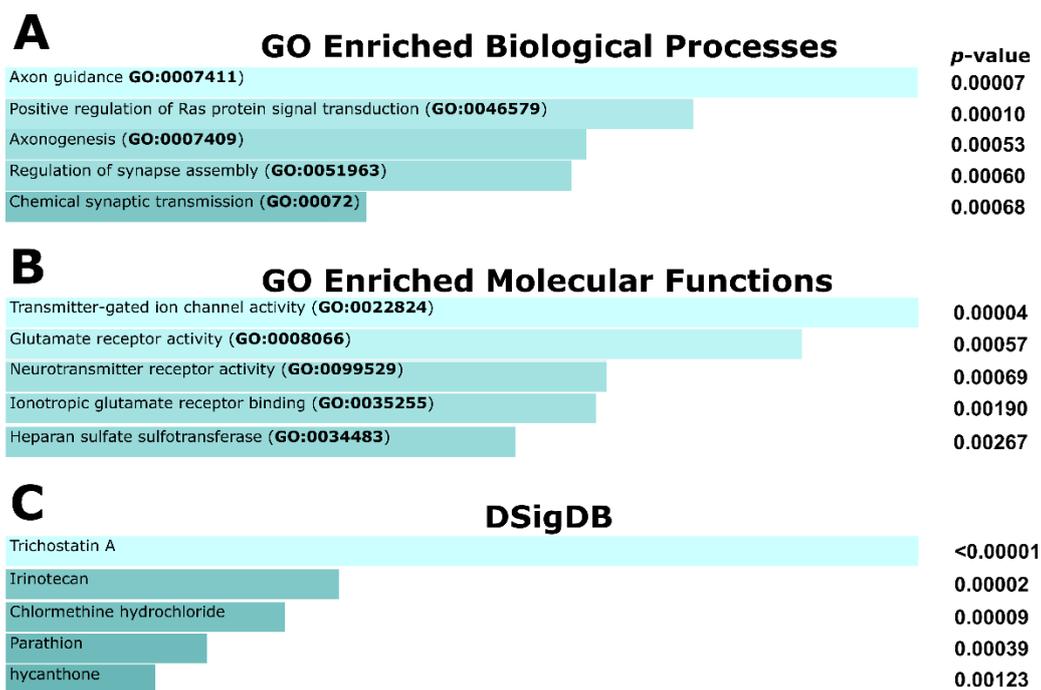
A GO Enriched Biological Processes

	p-value
Axon guidance **GO:0007411**)	0.00007
Positive regulation of Ras protein signal transduction (**GO:0046579**)	0.00010
Axonogenesis (**GO:0007409**)	0.00053
Regulation of synapse assembly (**GO:0051963**)	0.00060
Chemical synaptic transmission (**GO:00072**)	0.00068

B GO Enriched Molecular Functions

Transmitter-gated ion channel activity (**GO:0022824**)	0.00004
Glutamate receptor activity (**GO:0008066**)	0.00057
Neurotransmitter receptor activity (**GO:0099529**)	0.00069
Ionotropic glutamate receptor binding (**GO:0035255**)	0.00190
Heparan sulfate sulfotransferase (**GO:0034483**)	0.00267

C DSigDB

Trichostatin A	<0.00001
Irinotecan	0.00002
Chlormethine hydrochloride	0.00009
Parathion	0.00039
hycanthone	0.00123

Figure 1. The top five significantly enriched gene ontology terms associated with RTGs as determined by the EnrichR software. (**A**) Biological processes, (**B**) molecular functions, and (**C**) Drug Signatures Database (DSigDB). Pathways are presented based on the combined EnrichR score. DSigDB relates drugs/compounds to their target genes.

3.4. Integration Breakpoints in the HBV Genome

To investigate the distribution pattern of the integration breakpoints in the HBV genome, we analyzed HBV breakpoints in tumors ($n = 4008$) where available. We omitted studies that enriched for HBV DR1-2 sequences to assess the HBV breakpoint distribution in an unbiased manner. Consistent with previous reports [6,39,43], we observed that 39.2% of all breakpoints were within the HBV nt. 1300–1900 region in tumors. This region is at the 3′ end of the *HBx* gene and contains the initiation site of viral replication/transcription. Additionally, we observed a breakpoint hotspot in the HBV DR1-2 region, representing 29.6% of all HBV breakpoints (Figure 2).

Table 3. Most frequently reported genes with HBV integrants in HCC tumors.

RTG	Full Gene Name	Subjects (*n*)	Junctions (*n*)	Cancer Associated [Ref]
TERT [1]	Telomerase reverse transcriptase	357	561	Multiple cancers [63]
MLL4 (*KMT2B*)	Lysine methyltransferase 2B	130	220	HCC [52,64], Spindle cell sarcoma [65], Gastric cancer [66]
PLEKHG4B	Pleckstrin Homology and RhoGEF Domain Containing G4B	38	115	Neuroblastoma [67]
LOC100288778	WAS protein family homolog 8 pseudogene	34	79	SCLC [68]
DDX11L1	DEAD/H-Box Helicase 11 Like 1	33	57	Function unknown
CCNE1	Cyclin E1	31	54	Multiple cancers [69]
CCNA2	Cyclin A2	26	45	Multiple cancers [70]
SNTG1	Syntrophin Gamma 1	25	27	Lung adenocarcinoma [71]
PGBD2	PiggyBac Transposable Element Derived 2	22	51	Function unknown
DUX4L4	Double homeobox 4 like 4 pseudogene	20	35	DUX4 Ewing's sarcoma [72], ALL [73]
ROCK1P1	Rho Associated Coiled-Coil Containing Protein Kinase 1 Pseudogene 1	20	35	Prostate cancer [74]
ANKRD26P1	ankyrin repeat domain 26 pseudogene 1	19	72	Breast cancer [75]
PARD6G	Par-6 Family Cell Polarity Regulator Gamma	19	42	Breast, kidney, liver, lung, ovary, and pancreatic cancers [76]
FN1	Fibronectin 1	17	18	Multiple cancers [77]
CWH43	Cell Wall Biogenesis 43 C-Terminal Homolog	14	73	CRC and TSHomas [78]
TPTE	Transmembrane Phosphatase with Tensin Homology	13	30	HCC [79], prostate cancer [80]
FAM157A	Family with Sequence Similarity 157 Member A	14	22	Function unknown
LOC728323/LINC01881	Long Intergenic Non-Protein Coding RNA 1881	13	22	Oral cancer [81]
EMBP1	Embigin pseudogene 1	12	27	Oropharyngeal carcinoma [82], multiple primary cancers [83]
OR4C6	Olfactory Receptor Family 4 Subfamily C Member 6	12	22	Pancreatic cancer [84]
PRMT2	Protein Arginine Methyltransferase 2	12	15	Glioblastoma [85]
ROCK1	Rho Associated Coiled-Coil Containing Protein Kinase 1	12	23	HCC [86–89], CRC [90]
ANHX	Anomalous Homeobox	11	16	Function unknown
CTNND2	Catenin Delta 2	11	14	HCC [91,92], prostate cancer [93], lung cancer [94]
DDX11L9	DEAD/H-Box Helicase 11 Like 9 (Pseudogene)	11	16	Function unknown
SENP5	SUMO Specific Peptidase 5	11	11	HCC [95], breast cancer [96]
ZNF595	Zinc Finger Protein 595	11	14	Lung cancer [97], Gastric cancer [98]
CDRT7	CMT1A Duplicated Region Transcript 7	10	10	Glioma [99]

[1] The number of reported *TERT* junctions is slightly skewed, because one of the references [45] only reported the number of *TERT* promoter integrations. RTG, recurrently targeted gene; HCC, hepatocellular carcinoma; NSCLC, non-small cell lung cancer; SCLC, small cell lung cancer; ALL, acute lymphocytic leukemia; CRC, colorectal cancer; TSHoma, thyrotropin-secreting pituitary adenoma; CLL, chronic lymphocytic leukemia; RCC, renal cell carcinoma.
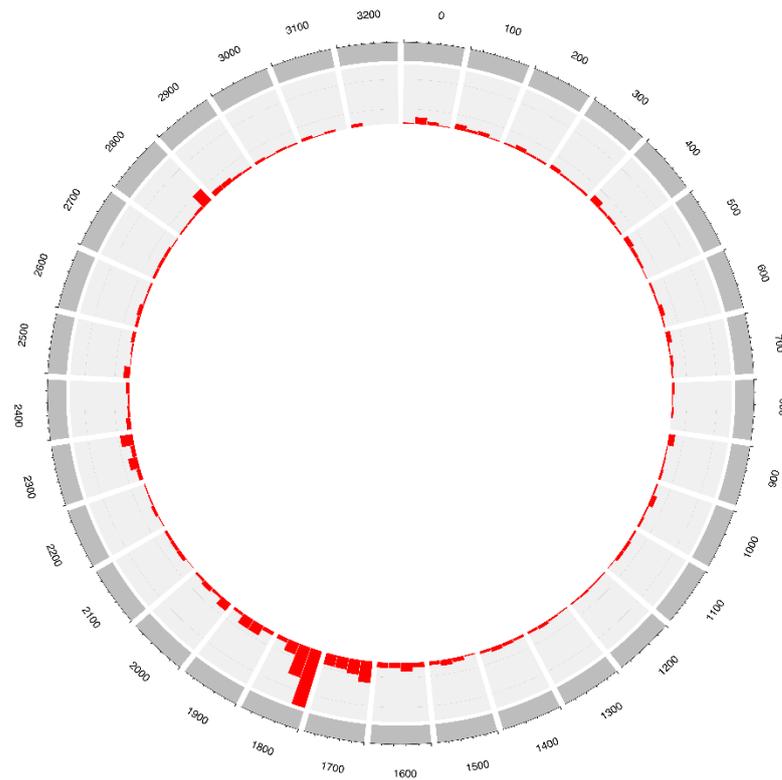
Figure 2. Distribution of integration breakpoints in the HBV genome in HCC tumor samples. A total of 3052 HBV breakpoints were plotted. The histogram represents the frequency of integration breakpoints at different loci in the HBV genome (nt. 1–3215; bin size 25) as numbered in the outer ring.

3.5. Genomic Breakpoints of TERT, MLL4 and PLEKHG4B RTGs

HBV integration is believed to be non-sequence-specific; therefore, it was of interest to examine all RTG breakpoint coordinates for similarity to each other. To do so, we plotted the available human and HBV breakpoint coordinates of the three most frequent RTGs identified, *TERT*, *MLL4*, and *PLEKHG4B* (Figure 3).

For *TERT*, the most frequent RTG, 291 of 561 junctions from 205 HCC patients, had both human and HBV breakpoint coordinates available. As expected, most of these breakpoints were centered between DR2 and DR1 of the viral genome and were highly concentrated at the promoter region of the *TERT* gene (Figure 3A). Of note, 142 (48.7%) *TERT* integration junction sequences were unique, i.e., reported from 142 different subjects, whereas the rest (57 sequences) were each found in two or more HCC patients. Of the 477 available breakpoint coordinates in the *TERT* gene, 361 (75.6%) junctions were located upstream of exon 1 and, of these upstream breakpoints, 242 (50.7%) were located within the *TERT* promoter region (Chr5:1,295,162–1,296,162).

MLL4 was the second most frequently reported RTG, with 220 junctions identified from 130 HCC patients. Among them, 150 breakpoints from 83 HCC patients had both human and viral coordinates available and are plotted in Figure 3B. As with *TERT*, most of the breakpoints were clustered between DR2 and DR1 of the viral genome and concentrated within exon 3 of the *MLL4* gene. Eleven different recurring breakpoints were found in two or more HCC patients, accounting for 37 of the 150 junctions examined.
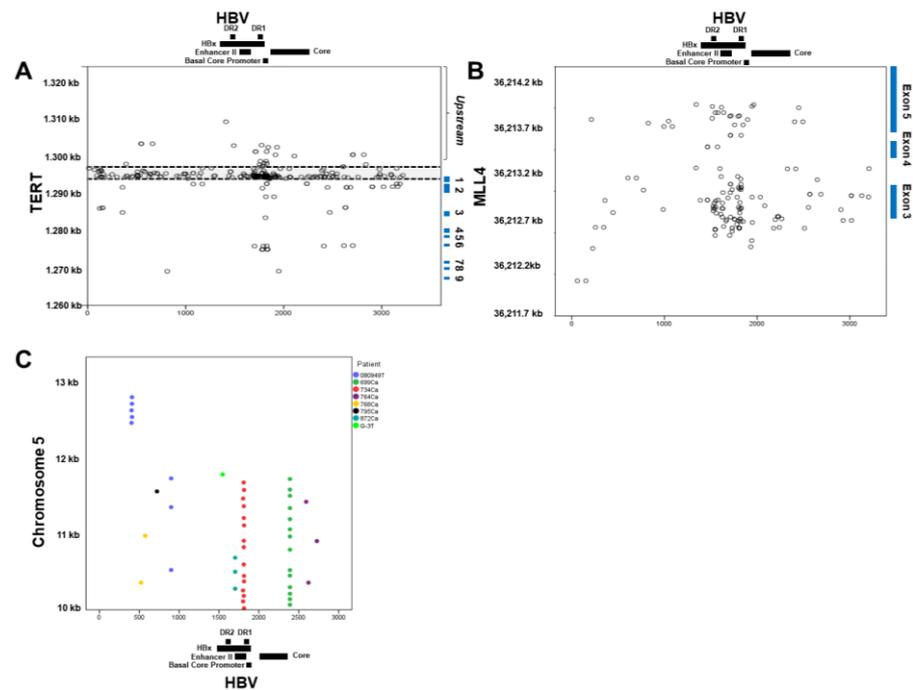
Figure 3. Mapping of *TERT, MLL4,* and *PLEKH4G4B* HBV integration breakpoints along the human and HBV genomes. (**A**) *TERT* breakpoints. A total of 219 *TERT* integration breakpoints derived from 161 patients are plotted. The *y*-axis coordinates decrease from 1320 kb to 1260 kb to represent the direction of the transcriptional start site from a 5′–3′ orientation. The expanded view of the region with the most integration sites is shown for the human genome position 1297 kb to 1294 kb and the HBV nt. 1500–2000. (**B**) *MLL4* breakpoints. A total of 115 *MLL4* integration breakpoints, derived from 64 patients, are plotted. Blue squares denoting exon regions are shown. (**C**) *PLEKH4G4B* breakpoints. A total of 47 of the 116 reported *PLEKHG4B* breakpoints plotted are derived from 8 unique HCC patients. Colored dots correspond to each unique patient. Each dot represents the mapped locations of the integration sites where the human gene breakpoints (GRCh37) are located on the *y*-axis, and HBV breakpoints are located on the *x*-axis, in accordance with the reported locations.

The third most reported RTG was PLEKHG4B, with 116 integrations reported in this study. Interestingly, all the 116 breakpoints were centered within a 3 kb region which was 131 kb upstream from the *PLEKHG4B* coding region. A total of 47 breakpoints from eight HCC patients had both viral and human coordinates available, as shown in Figure 3C. All the breakpoints were found upstream of the transcription start site (Chr5:140,373). Unlike in *TERT* and *MLL4*, the viral breakpoints were centered in two HBV regions (nt. 1802–1814 and 2390; 15 and 14 breakpoints, respectively) but at various human coordinates. Further analysis of the human sequences (Chr5:10,000–13,000) in the upstream integration region revealed a 1877 bp segment containing simple repeat sequences and a 1057 bp segment containing satellite sequences. This finding is interesting because HBV has been suggested to have a higher propensity to integrate into repeat regions/retrotransposons, as recently shown by Chauhan et al. [100]. In addition, a motif, TAAACCCTAAC, was discovered, appearing four times in the Chr5:10,000–13,000 region and once in the HBV genome, each with $p < 0.0001$ by the Student's *t*-test. However, querying the GenomeNet database for this motif produced no matches, and analysis of the region for known motifs produced no results. No recurrent breakpoints were identified among the integration sites in the *PLEKHG4B* gene. Notably, seven of the eight HCC patients with this unique junction pattern, where the same HBV breakpoint was joined to different host chromosome breakpoints, were reported from one study by Yang et al. [43].

TERT hotspot promoter mutations (−124, −146) are the most frequently reported mutations in HCC, found in about 50% of cases [101–106]. In HBV–HCC, up-regulation of

TERT expression could also be caused by HBV integration at or near the *TERT* promoter region [28–30,32,38,48]. Therefore, we next compared the frequencies of *TERT* promoter mutation and HBV integration. For our in-house cohort (*n* = 22), shown in Figure 4A, promoter mutations were found in 6 of 22 samples, and integrations in the *TERT* gene were found in 5 of 22 samples in a mutually exclusive fashion. Together, *TERT* alterations were detected in 50% (11/22) of this cohort. To determine if this mutual exclusivity applied to a larger sample size, we examined *TERT* alterations identified by us and others [40,42,45] together, as summarized in Figure 4B. Of the 347 HBV–HCC patients, 174 (50%) were found to have detectable *TERT* alterations. Promoter mutations and HBV integrations were mutually exclusive and comprised 40% (70/174) and 60% (104/174) of all alterations, respectively.
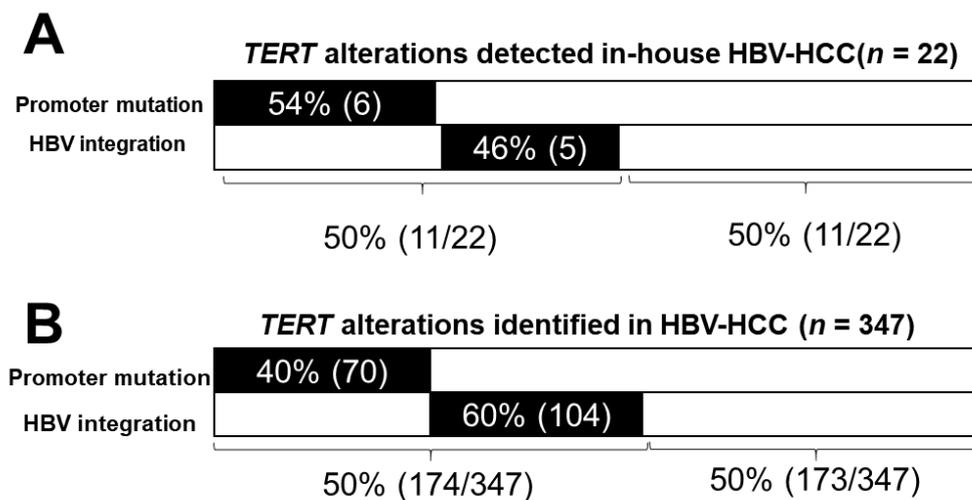
Figure 4. *TERT* gene alterations identified in HBV–HCC tissues. (**A**) In-house cohort (*n* = 22). (**B**) Compiled HBV–HCC cohort. Patients are derived from our in-house study (*n* = 22) and from the literature (*n* = 129) [40,42,45]. The number of HCC patients is indicated in parentheses.

4. Discussion

In this study, we compiled and studied a total of 18,596 HBV integration sites from 1733 HCC patients reported in 32 previous studies and our in-house study, to test our hypothesis that frequent HBV RTGs are associated with HCC carcinogenesis and thus may represent HCC driver genes that can reveal potential therapeutic targets and molecular profiling for precision medicine. Using three criteria for RTG selection, we identified 396 RTGs. Encouragingly, the most frequent RTGs (*n* = 28), reported in at least 10 HCC patients each, either have known involvement in carcinogenesis (23/28; 82%) or have no known function (5/28; 18%). By gene ontology analysis, RTGs were mapped to functions related to carcinogenesis. We describe a potential tool to identify or discover HCC drivers by the characterization of frequent RTGs. More studies are needed to demonstrate the association of carcinogenesis with the frequency of RTGs that have unknown functions.

Three criteria were applied to candidate genes to identify 396 HBV RTGs in this study: (1) location within 150 kb of the breakpoint, the reported distance within which host genes can be impacted by integration [27,28]; (2) the presence of HBV integrants in ≥2 HCC patients; and (3) identification in ≥2 independent laboratories to avoid the possibility of contamination within a laboratory. We are aware that the identification of RTGs across multiple studies is complex due to multi-faceted underlying variables such as integration detection methodologies and patient populations. For instance, some studies do not contain any of the 396 RTGs that we identified [54,55], while others report a high detection rate of a particular RTG, such as *MLL4* [52] or *cMYC* [39]. We are also aware that different methodologies for identifying integrations may have different sensitivities that can result in different integration site profiles. Furthermore, how DNA samples were prepared varied across the studies. Despite these limitations, the detection of highly

frequent RTGs as potential HBV–HCC drivers could be clinically useful for HCC patients. Notably, our approach can only identify driver genes involved in HBV–HCC (rather than HCC in general) and is limited to insertional mutagenesis. In addition, HBV integration could contribute to carcinogenesis by causing chromosomal instability, expression of the oncogenic HBV protein X, and inflammation induced by other HBV proteins expressed from the integrated DNA. Although the most frequent RTG, *TERT*, is also the most frequently mutated gene in HCC in general, other frequently mutated HCC genes, such as *TP53* and *CTNNB1*, were not among the top 28 RTGs. These findings suggest that using RTG analysis as a tool to identify driver genes is limited to HBV–HCC.

By detailed analysis of integrations in the three most frequent RTGs (*TERT*, *MLL4*, and *PLEKHG4B*), we revealed three important features. First, as expected, most of the junction coordinates are different, supporting a non-sequence-specific mechanism of integration in the host genome. Among the junction sequences identified in multiple HCC patients, such as the 57 junctions described in this study, frequent integration in the promoter region of the *TERT* gene highlights the potential importance of the site in hepatocarcinogenesis. Secondly, an interesting pattern of integration into repetitive sequences was observed in the *PLEKH4G4B* junctions shown in Figure 3C. Furthermore, a highly repetitive sequence, satellite sequences, and the TAAACCCTAAC motif were identified in these regions, consistent with previous observations that integration occurred frequently in repetitive sequences [100]. However, because these unique integration sequence patterns were reported in only one study and validation in original tissue DNA was not reported, the possibility of an artifact has not been excluded. Lastly, the mutually exclusive detection of *TERT* promoter mutations and *TERT* integrations was shown in our small cohort of 22 HCC patients as well as in a larger compiled cohort of 347 HCC patients [40,42,45]. *TERT* promoter mutations account for 50% of all alterations in this gene, suggesting the importance of identifying *TERT* integration. This observation also further emphasizes the need for analysis of frequent RTGs to better characterize HCC.

In summary, we identified 396 RTGs from an analysis of 18,596 HBV integration sites. Twenty-eight of the RTGs were identified in 10 or more HCC patients. The RTGs were found to be significantly enriched in several GO pathways. The 28 most frequent RTGs have previously been implicated in carcinogenesis (23/28) or have no known function (5/28). Taken together, our findings demonstrate the potential of RTG identification as a tool for HBV–HCC driver discovery and tumor characterization, to be used in precision medicine and drug development. More studies are needed to further refine the criteria for RTG identification for its applications.

References

1. Howlader, N.N.A.; Krapcho, M.; Miller, D.; Bishop, K.; Altekruse, S.F.; Kosary, C.L.; Yu, M.; Ruhl, J.; Tatalovich, Z.; Mariotto, A. (Eds.) *SEER Cancer Statistics Review, 1975–2013*; Based on November 2015 SEER Data Submission, Posted to the SEER Web Site; National Cancer Institute: Bethesda, MD, USA, April 2016. Available online: http://seer.cancer.gov/csr/1975_2013/ (accessed on 5 December 2020).

2. Shinozaki, M.; Hoon, D.S.; Giuliano, A.E.; Hansen, N.M.; Wang, H.-J.; Turner, R.; Taback, B. Distinct Hypermethylation Profile of Primary Breast Cancer Is Associated with Sentinel Lymph Node Metastasis. *Clin. Cancer Res.* **2005**, *11*, 2156–2162. [CrossRef]

3. Ferlay, J.S.I.; Ervik, M.; Dikshit, R.; Eser, S.; Mathers, C.; Rebelo, M.; Parkin, D.M.; Forman, D.; Bray, F. *GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11*; International Agency for Research on Cancer: Lyon, France, 2013; Available online: http://globocan.iarc.fr (accessed on 14 March 2017).

4. El-Serag, H.B. Epidemiology of Viral Hepatitis and Hepatocellular Carcinoma. *Gastroenterology* **2012**, *142*, 1264–1273. [CrossRef] [PubMed]

5. Anugwom, C.M.; Allaire, M.; Akbar, S.M.F.; Sultan, A.; Bollipo, S.; Mattos, A.Z.; Debes, J.D. Hepatitis B-related hepatocellular carcinoma: Surveillance strategy directed by immune-epidemiology. *Hepatoma Res.* **2021**, *7*. [CrossRef]

6. 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]

7. 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]

8. 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]

9. Chauhan, R.; Churchill, N.D.; Mulrooney-Cousins, P.M.; I Michalak, T. 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] [PubMed]

10. Bill, C.A.; Summers, J. Genomic DNA double-strand breaks are targets for hepadnaviral DNA integration. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 11135–11140. [CrossRef] [PubMed]

11. Budzinska, M.A.; Shackel, N.A.; Urban, S.; Tu, T. Sequence analysis of integrated hepatitis B virus DNA during HBeAg-seroconversion. *Emerg. Microbes Infect.* **2018**, *7*, 1–12. [CrossRef]

12. 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]

13. Scotto, J.; Hadchouel, M.; Hery, C.; Alvarez, F.; Yvart, J.; Tiollais, P.; Bernard, O.; Brechot, C. Hepatitis B virus DNA in children's liver diseases: Detection by blot hybridisation in liver and serum. *Gut* **1983**, *24*, 618–624. [CrossRef]

14. Huang, H.-P.; Tsuei, D.-J.; Wang, K.-J.; Chen, Y.-L.; Ni, Y.-H.; Jeng, Y.-M.; Chen, H.-L.; Hsu, H.-Y.; Chang, M.-H. Differential integration rates of hepatitis B virus DNA in the liver of children with chronic hepatitis B virus infection and hepatocellular carcinoma. *J. Gastroenterol. Hepatol.* **2005**, *20*, 1206–1214. [CrossRef]

15. 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]

16. Koshy, R.; Maupas, P.; Muller, R.; Hofschneider, P.H. Detection of Hepatitis B Virus-specific DNA in the Genomes of Human Hepatocellular Carcinoma and Liver Cirrhosis Tissues. *J. Gen. Virol.* **1981**, *57*, 95–102. [CrossRef]

17. Takada, S.; Gotoh, Y.; Hayashi, S.; Yoshida, M.; Koike, K. Structural rearrangement of integrated hepatitis B virus DNA as well as cellular flanking DNA is present in chronically infected hepatic tissues. *J. Virol.* **1990**, *64*, 822–828. [CrossRef] [PubMed]

18. Tu, T.; Budzinska, M.A.; Vondran, F.W.R.; Shackel, N.A.; Urban, S. Hepatitis B Virus DNA Integration Occurs Early in the Viral Life Cycle in anIn VitroInfection Model via Sodium Taurocholate Cotransporting Polypeptide-Dependent Uptake of Enveloped Virus Particles. *J. Virol.* **2018**, *92*, e02007–e02017. [CrossRef]

19. Hai, H. Role of hepatitis B virus DNA integration in human hepatocarcinogenesis. *World J. Gastroenterol.* **2014**, *20*, 6236–6243. [CrossRef] [PubMed]

20. 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]

21. 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] [PubMed]

22. Jhunjhunwala, S.; Jiang, Z.; Stawiski, E.W.; Gnad, F.; Liu, J.; Mayba, O.; Du, P.; Diao, J.; Johnson, S.; Wong, K.-F. Diverse modes of genomic alterations in hepatocellular carcinoma. *Genome Biol.* **2014**, *15*, 436. [PubMed]

23. Lin, S.Y.; Dhillon, V.; Jain, S.; Chang, T.-T.; Hu, C.-T.; Lin, Y.-J.; Chen, S.-H.; Chang, K.-C.; Song, W.; Yu, L.; et al. A Locked Nucleic Acid Clamp-Mediated PCR Assay for Detection of a p53 Codon 249 Hotspot Mutation in Urine. *J. Mol. Diagn.* **2011**, *13*, 474–484. [CrossRef]

24. Jain, S.; Xie, L.; Boldbaatar, B.; Lin, S.Y.; Hamilton, J.P.; Meltzer, S.J.; Chen, S.-H.; Hu, C.-T.; Block, T.M.; Song, W.; et al. Differential methylation of the promoter and first exon of the RASSF1A gene in hepatocarcinogenesis. *Hepatol. Res.* **2014**, *45*, 1110–1123. [CrossRef] [PubMed]

25. Shieh, F.-S.; Jongeneel, P.; Steffen, J.D.; Lin, S.; Jain, S.; Song, W.; Su, Y.-H. ChimericSeq: An open-source, user-friendly interface for analyzing NGS data to identify and characterize viral-host chimeric sequences. *PLoS ONE* **2017**, *12*, e0182843. [CrossRef]

26. Madeira, F.; Park, Y.M.; Lee, J.; Buso, N.; Gur, T.; Madhusoodanan, N.; Basutkar, P.; Tivey, A.R.N.; Potter, S.C.; Finn, R.D.; et al. The EMBL-EBI search and sequence analysis tools APIs. *Nucleic Acids Res.* **2019**, *47*, W636–W641. [CrossRef]

27. Shamay, M.; Agami, R.; Shaul, Y. HBV integrants of hepatocellular carcinoma cell lines contain an active enhancer. *Oncogene* **2001**, *20*, 6811–6819. [CrossRef] [PubMed]

28. Horikawa, I.; Barrett, J.C. cis-Activation of the Human Telomerase Gene (hTERT) by the Hepatitis B Virus Genome. *J. Natl. Cancer Inst.* **2001**, *93*, 1171–1173. [CrossRef] [PubMed]

29. 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]

30. 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]

31. 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]

32. 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]

33. 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]

34. Miao, R.; Luo, H.; Zhou, H.; Li, G.; Bu, D.; Yang, X.; Zhao, X.; Zhang, H.; Liu, S.; Zhong, Y.; et al. Identification of prognostic biomarkers in hepatitis B virus-related hepatocellular carcinoma and stratification by integrative multi-omics analysis. *J. Hepatol.* **2014**, *61*, 840–849. [CrossRef] [PubMed]

35. Hama, N.; Totoki, Y.; Miura, F.; Tatsuno, K.; Saito-Adachi, M.; Nakamura, H.; Arai, Y.; Hosoda, F.; Urushidate, T.; Ohashi, S.; et al. Epigenetic landscape influences the liver cancer genome architecture. *Nat. Commun.* **2018**, *9*, 1643. [CrossRef]

36. Duan, M.; Hao, J.; Cui, S.; Worthley, D.L.; Zhang, S.; Wang, Z.; Shi, J.; Liu, L.; Wang, X.; Ke, A.; et al. Diverse modes of clonal evolution in HBV-related hepatocellular carcinoma revealed by single-cell genome sequencing. *Cell Res.* **2018**, *28*, 359–373. [CrossRef] [PubMed]

37. Chen, W.; Zhang, K.; Dong, P.; Fanning, G.; Tao, C.; Zhang, H.; Guo, S.; Wang, Z.; Hong, Y.; Yang, X.; et al. Noninvasive chimeric DNA profiling identifies tumor-originated HBV integrants contributing to viral antigen expression in liver cancer. *Hepatol. Int.* **2020**, *14*, 326–337. [CrossRef]

38. Toh, S.T.; Jin, Y.; Liu, L.; Wang, J.; Babrzadeh, F.; Gharizadeh, B.; Ronaghi, M.; Toh, H.C.; Chow, P.K.-H.; Chung, A.Y. Deep sequencing of the hepatitis B virus in hepatocellular carcinoma patients reveals enriched integration events, structural alterations and sequence variations. *Carcinogenesis* **2013**, *34*, 787–798. [CrossRef] [PubMed]

39. 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]

40. 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.* **2015**, *51*, 473–486. [CrossRef]

41. Furuta, M.; Tanaka, H.; Shiraishi, Y.; Uchida, 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]

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

43. 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]

44. Chen, X.-P.; Long, X.; Jia, W.-L.; Wu, H.-J.; Zhao, J.; Liang, H.-F.; Laurence, A.; Zhu, J.; Dong, D.; Chen, Y. Viral integration drives multifocal HCC during the occult HBV infection. *J. Exp. Clin. Cancer Res.* **2019**, *38*, 1–11.

45. Sze, K.M.; Ho, D.W.; Chiu, Y.; Tsui, Y.; Chan, L.; Lee, J.M.; Chok, K.S.; Chan, A.C.; Tang, C.; Tang, V.W.; et al. Hepatitis B Virus–Telomerase Reverse Transcriptase Promoter Integration Harnesses Host ELF4, Resulting in Telomerase Reverse Transcriptase Gene Transcription in Hepatocellular Carcinoma. *Hepatology* **2021**, *73*, 23–40. [CrossRef]

46. Tatsuno, K.; Midorikawa, Y.; Takayama, T.; Yamamoto, S.; Nagae, G.; Moriyama, M.; Nakagawa, H.; Koike, K.; Moriya, K.; Aburatani, H. Impact of AAV2 and Hepatitis B Virus Integration into Genome on Development of Hepatocellular Carcinoma in Patients with Prior Hepatitis B Virus Infection. *Clin. Cancer Res.* **2019**, *25*, 6217–6227. [CrossRef]

47. Li, C.; Ho, M.; Lin, Y.; Tzeng, S.; Chen, Y.; Pai, H.; Wang, Y.; Chen, C.; Lee, Y.; Chen, D.; et al. Cell-Free Virus-Host Chimera DNA From Hepatitis B Virus Integration Sites as a Circulating Biomarker of Hepatocellular Cancer. *Hepatology* **2020**, *72*, 2063–2076. [CrossRef]

48. Ferber, M.J.; Montoya, D.P.; Yu, C.; Aderca, I.; McGee, A.; Thorland, E.C.; Nagorney, D.M.; Gostout, B.S.; Burgart, L.J.; Boix, L.; et al. Integrations of the hepatitis B virus (HBV) and human papillomavirus (HPV) into the human telomerase reverse transcriptase (hTERT) gene in liver and cervical cancers. *Oncogene* **2003**, *22*, 3813–3820. [CrossRef]

49. Wang, Y.; Lau, S.H.; Sham, J.S.-T.; Wu, M.-C.; Wang, T.; Guan, X.-Y. Characterization of HBV integrants in 14 hepatocellular carcinomas: Association of truncated X gene and hepatocellular carcinogenesis. *Oncogene* **2004**, *23*, 142–148. [CrossRef] [PubMed]

50. Tamori, A.; Yamanishi, Y.; Kawashima, S.; Kanehisa, M.; Enomoto, M.; Tanaka, H.; Kubo, S.; Shiomi, S.; Nishiguchi, S. Alteration of Gene Expression in Human Hepatocellular Carcinoma with Integrated Hepatitis B Virus DNA. *Clin. Cancer Res.* **2005**, *11*, 5821–5826. [CrossRef]

51. Murakami, Y.; Saigo, K.; Takashima, H.; Minami, M.; Okanoue, T.; Bréchot, 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] [PubMed]

52. Saigo, K.; Yoshida, K.; Ikeda, R.; Sakamoto, Y.; Murakami, Y.; Urashima, T.; Asano, T.; Kenmochi, T.; Inoue, I. Integration of hepatitis B virus DNA into the myeloid/lymphoid or mixed-lineage leukemia (MLL4) gene and rearrangements of MLL4 in human hepatocellular carcinoma. *Hum. Mutat.* **2008**, *29*, 703–708. [CrossRef]

53. 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] [PubMed]

54. 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]

55. Fang, X.; Wu, H.-H.; Ren, J.-J.; Liu, H.-Z.; Li, K.-Z.; Li, J.-L.; Tang, Y.-P.; Xiao, C.-C.; Huang, T.-R.; Deng, W. Associations between serum HBX quasispecies and their integration in hepatocellular carcinoma. *Int. J. Clin. Exp. Pathol.* **2017**, *10*, 11857–11866.

56. 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.* **2018**, *19*, 1410–1416. [CrossRef] [PubMed]

57. He, Z.; Zhu, J.; Mo, J.; Zhao, H.; Chen, Q. HBV DNA integrates into upregulated ZBTB20 in patients with hepatocellular carcinoma. *Mol. Med. Rep.* **2020**, *22*, 380–386. [CrossRef] [PubMed]

58. Yang, J.D.; Kim, W.R.; Coelho, R.; Mettler, T.A.; Benson, J.T.; Sanderson, S.O.; Therneau, T.M.; Kim, B.; Roberts, L.R. Cirrhosis Is Present in Most Patients with Hepatitis B and Hepatocellular Carcinoma. *Clin. Gastroenterol. Hepatol.* **2011**, *9*, 64–70. [CrossRef]

59. El–Serag, H.B.; Rudolph, K.L. Hepatocellular Carcinoma: Epidemiology and Molecular Carcinogenesis. *Gastroenterology* **2007**, *132*, 2557–2576. [CrossRef]

60. Mueller-Breckenridge, A.J.; Garcia-Alcalde, F.; Wildum, S.; Smits, S.L.; De Man, R.A.; Van Campenhout, M.J.H.; Brouwer, W.P.; Niu, J.; Young, J.A.T.; Najera, I.; et al. Machine-learning based patient classification using Hepatitis B virus full-length genome quasispecies from Asian and European cohorts. *Sci. Rep.* **2019**, *9*, 1–12. [CrossRef]

61. Kuleshov, M.V.; Jones, M.R.; Rouillard, A.D.; Fernandez, N.F.; Duan, Q.; Wang, Z.; Koplev, S.; Jenkins, S.L.; Jagodnik, K.M.; Lachmann, A.; et al. Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **2016**, *44*, W90–W97. [CrossRef]

62. Smits, N.C.; Kobayashi, T.; Srivastava, P.K.; Skopelja, S.; Ivy, J.A.; Elwood, D.J.; Stan, R.V.; Tsongalis, G.J.; Sellke, F.W.; Gross, P.L. HS3ST1 genotype regulates antithrombin's inflammomodulatory tone and associates with atherosclerosis. *Matrix Biol.* **2017**, *63*, 69–90. [CrossRef]

63. Heidenreich, B.; Rachakonda, P.S.; Hemminki, K.; Kumar, R. TERT promoter mutations in cancer development. *Curr. Opin. Genet. Dev.* **2014**, *24*, 30–37. [CrossRef] [PubMed]

64. Tamori, A.; Nishiguchi, S.; Shiomi, S.; Hayashi, T.; Kobayashi, S.; Habu, D.; Takeda, T.; Seki, S.; Hirohashi, K.; Tanaka, H.; et al. Hepatitis B Virus DNA Integration in Hepatocellular Carcinoma After Interferon-Induced Disappearance of Hepatitis C Virus. *Am. J. Gastroenterol.* **2005**, *100*, 1748–1753. [CrossRef]

65. O'Meara, E.; Stack, D.; Phelan, S.; McDonagh, N.; Kelly, L.; Sciot, R.; Debiec-Rychter, M.; Morris, T.; Cochrane, D.; Sorensen, P.; et al. Identification of anMLL4-GPS2fusion as an oncogenic driver of undifferentiated spindle cell sarcoma in a child. *Genes Chromosom. Cancer* **2014**, *53*, 991–998. [CrossRef]

66. Pan, X.; Ji, X.; Zhang, R.; Zhou, Z.; Zhong, Y.; Peng, W.; Sun, N.; Xu, X.; Xia, L.; Li, P.; et al. Landscape of somatic mutations in gastric cancer assessed using next-generation sequencing analysis. *Oncol. Lett.* **2018**, *16*, 4863–4870. [CrossRef] [PubMed]

67. Chicard, M.; Boyault, S.; Daage, L.C.; Richer, W.; Gentien, D.; Pierron, G.; Lapouble, E.; Bellini, A.; Clement, N.; Iacono, I.; et al. Genomic Copy Number Profiling Using Circulating Free Tumor DNA Highlights Heterogeneity in Neuroblastoma. *Clin. Cancer Res.* **2016**, *22*, 5564–5573. [CrossRef]

68. Li, J.; Wang, J.; Chen, Y.; Yang, L.; Chen, S. A prognostic 4-gene expression signature for squamous cell lung carcinoma. *J. Cell. Physiol.* **2017**, *232*, 3702–3713. [CrossRef] [PubMed]

69. Donnellan, R.; Chetty, R. Cyclin E in human cancers. *FASEB J.* **1999**, *13*, 773–780. [CrossRef] [PubMed]
70. Otto, T.; Sicinski, T.O.P. Cell cycle proteins as promising targets in cancer therapy. *Nat. Rev. Cancer* **2017**, *17*, 93–115. [CrossRef]
71. Meng, F.; Zhang, L.; Ren, Y.; Ma, Q. The genomic alterations of lung adenocarcinoma and lung squamous cell carcinoma can explain the differences of their overall survival rates. *J. Cell. Physiol.* **2019**, *234*, 10918–10925. [CrossRef]
72. Yoshimoto, T.; Tanaka, M.; Homme, M.; Yamazaki, Y.; Takazawa, Y.; Antonescu, C.R.; Nakamura, T. CIC-DUX4 Induces Small Round Cell Sarcomas Distinct from Ewing Sarcoma. *Cancer Res.* **2017**, *77*, 2927–2937. [CrossRef]
73. Yasuda, T.; Tsuzuki, S.; Kawazu, M.; Hayakawa, F.; Kojima, S.; Ueno, T.; Imoto, N.; Kohsaka, S.; Kunita, A.; Doi, K.; et al. Recurrent DUX4 fusions in B cell acute lymphoblastic leukemia of adolescents and young adults. *Nat. Genet.* **2016**, *48*, 569–574. [CrossRef] [PubMed]
74. Luo, Y.; Jiang, Q.-W.; Wu, J.-Y.; Qiu, J.-G.; Zhang, W.-J.; Mei, X.-L.; Shi, Z.; Di, J.-M. Regulation of migration and invasion by Toll-like receptor-9 signaling network in prostate cancer. *Oncotarget* **2015**, *6*, 22564–22574. [CrossRef] [PubMed]
75. Krepischi, A.C.V.; Achatz, M.I.W.; Santos, E.M.M.; Costa, S.S.; Lisboa, B.C.G.; Brentani, H.; Santos, T.M.; Goncalves, A.; Nobrega, A.F.; Pearson, P.L.; et al. Germline DNA copy number variation in familial and early-onset breast cancer. *Breast Cancer Res.* **2012**, *14*, R24. [CrossRef]
76. Marqués, E.; Englund, J.I.; A Tervonen, T.; Virkunen, E.; Laakso, M.; Myllynen, M.; Mäkelä, A.; Ahvenainen, M.; Lepikhova, T.; Monni, O.; et al. Par6G suppresses cell proliferation and is targeted by loss-of-function mutations in multiple cancers. *Oncogene* **2015**, *35*, 1386–1398. [CrossRef] [PubMed]
77. Hynes, R.O. Fibronectins. *Sci. Am.* **1986**, *254*, 42–51. [CrossRef]
78. Chu, C.-M.; Yao, C.-T.; Chang, Y.-T.; Chou, H.-L.; Chou, Y.-C.; Chen, K.-H.; Terng, H.-J.; Huang, C.-S.; Lee, C.-C.; Su, S.-L.; et al. Gene Expression Profiling of Colorectal Tumors and Normal Mucosa by Microarrays Meta-Analysis Using Prediction Analysis of Microarray, Artificial Neural Network, Classification, and Regression Trees. *Dis. Markers* **2014**, *2014*, 1–11. [CrossRef]
79. Dong, X.-Y.; Su, Y.-R.; Qian, X.-P.; Yang, X.-A.; Pang, X.-W.; Wu, H.-Y.; Chen, W.-F. Identification of two novel CT antigens and their capacity to elicit antibody response in hepatocellular carcinoma patients. *Br. J. Cancer* **2003**, *89*, 291–297. [CrossRef]
80. Singh, A.P.; Bafna, S.; Chaudhary, K.; Venkatraman, G.; Smith, L.; Eudy, J.D.; Johansson, S.L.; Lin, M.-F.; Batra, S.K. Genome-wide expression profiling reveals transcriptomic variation and perturbed gene networks in androgen-dependent and androgen-independent prostate cancer cells. *Cancer Lett.* **2008**, *259*, 28–38. [CrossRef]
81. Ambatipudi, S.; Gerstung, M.; Gowda, R.; Pai, P.; Borges, A.M.; Schäffer, A.A.; Beerenwinkel, N.; Mahimkar, M.B. Genomic Profiling of Advanced-Stage Oral Cancers Reveals Chromosome 11q Alterations as Markers of Poor Clinical Outcome. *PLoS ONE* **2011**, *6*, e17250. [CrossRef]
82. Ren, S.; Gaykalova, D.; Wang, J.; Guo, T.; Danilova, L.; Favorov, A.; Fertig, E.; Bishop, J.; Khan, Z.; Flam, E.; et al. Discovery and development of differentially methylated regions in human papillomavirus-related oropharyngeal squamous cell carcinoma. *Int. J. Cancer* **2018**, *143*, 2425–2436. [CrossRef]
83. Park, S.L.; Caberto, C.P.; Lin, Y.; Goodloe, R.J.; Dumitrescu, L.; Love, S.-A.; Matise, T.C.; Hindorff, L.A.; Fowke, J.H.; Schumacher, F.R.; et al. Association of Cancer Susceptibility Variants with Risk of Multiple Primary Cancers: The Population Architecture using Genomics and Epidemiology Study. *Cancer Epidemiol. Biomark. Prev.* **2014**, *23*, 2568–2578. [CrossRef] [PubMed]
84. Weber, L.; Maßberg, D.; Becker, C.; Altmüller, J.; Ubrig, B.; Bonatz, G.; Wölk, G.; Philippou, S.; Tannapfel, A.; Hatt, H.; et al. Olfactory Receptors as Biomarkers in Human Breast Carcinoma Tissues. *Front. Oncol.* **2018**, *8*, 33. [CrossRef] [PubMed]
85. Dong, F.; Li, Q.; Yang, C.; Huo, D.; Wang, X.; Ai, C.; Kong, Y.; Sun, X.; Wang, W.; Zhou, Y.; et al. PRMT2 links histone H3R8 asymmetric dimethylation to oncogenic activation and tumorigenesis of glioblastoma. *Nat. Commun.* **2018**, *9*, 4552. [CrossRef]
86. Tang, J.; Liu, C.; Xu, B.; Wang, D.; Ma, Z.; Chang, X. ARHGEF10L contributes to liver tumorigenesis through RhoA-ROCK1 signaling and the epithelial-mesenchymal transition. *Exp. Cell Res.* **2019**, *374*, 46–68. [CrossRef] [PubMed]
87. Liu, W.; Zhang, Q.; Tang, Q.; Hu, C.; Huang, J.; Liu, Y.; Lu, Y.; Wang, Q.; Li, G.; Zhang, R. Lycorine inhibits cell proliferation and migration by inhibiting ROCK1/cofilin-induced actin dynamics in HepG2 hepatoblastoma cells. *Oncol. Rep.* **2018**, *40*, 2298–2306. [CrossRef]
88. Ding, W.; Tan, H.; Zhao, C.; Li, X.; Li, Z.; Jiang, C.; Zhang, Y.; Wang, L. MiR-145 suppresses cell proliferation and motility by inhibiting ROCK1 in hepatocellular carcinoma. *Tumor Biol.* **2016**, *37*, 6255–6260. [CrossRef] [PubMed]
89. Deng, Q.; Xie, L.; Li, H. MiR-506 suppresses cell proliferation and tumor growth by targeting Rho-associated protein kinase 1 in hepatocellular carcinoma. *Biochem. Biophys. Res. Commun.* **2015**, *467*, 921–927. [CrossRef]
90. Song, G.-L.; Jin, C.-C.; Zhao, W.; Tang, Y.; Wang, Y.-L.; Li, M.; Xiao, M.; Li, X.; Li, Q.-S.; Lin, X.; et al. Regulation of the RhoA/ROCK/AKT/β-catenin pathway by arginine-specific ADP-ribosytransferases 1 promotes migration and epithelial-mesenchymal transition in colon carcinoma. *Int. J. Oncol.* **2016**, *49*, 646–656. [CrossRef]
91. Chen, R.; Dong, Y.; Xie, X.; Chen, J.; Gao, D.; Liu, Y.; Ren, Z.; Cui, J. Screening candidate metastasis-associated genes in three-dimensional HCC spheroids with different metastasis potential. *Int. J. Clin. Exp. Pathol.* **2014**, *7*, 2527–2535.
92. Huang, F.; Chen, J.; Lan, R.; Wang, Z.; Chen, R.; Lin, J.; Fu, L. Hypoxia induced δ-Catenin to enhance mice hepatocellular carcinoma progression via Wnt signaling. *Exp. Cell Res.* **2019**, *374*, 94–103. [CrossRef]
93. Zhang, P.; Schaefer-Klein, J.; Cheville, J.C.; Vasmatzis, G.; Kovtun, I.V. Frequently rearranged and overexpressed δ-catenin is responsible for low sensitivity of prostate cancer cells to androgen receptor and β-catenin antagonists. *Oncotarget* **2018**, *9*, 24428–24442. [CrossRef] [PubMed]

94. Huang, F.; Chen, J.; Wang, Z.; Lan, R.; Fu, L.; Zhang, L. δ-Catenin promotes tumorigenesis and metastasis of lung adenocarcinoma. *Oncol. Rep.* **2017**, *39*, 809–817. [CrossRef]

95. Jin, Z.-L.; Pei, H.; Xu, Y.-H.; Yu, J.; Deng, T. The SUMO-specific protease SENP5 controls DNA damage response and promotes tumorigenesis in hepatocellular carcinoma. *Eur. Rev. Med. Pharmacol. Sci.* **2016**, *20*, 3566–3573.

96. Cashman, R.; Cohen, H.; Ben-Hamo, R.; Zilberberg, A.; Efroni, S. SENP5 mediates breast cancer invasion via a TGFβRI SUMOylation cascade. *Oncotarget* **2014**, *5*, 1071–1082. [CrossRef] [PubMed]

97. Kanwal, M.; Ding, X.-J.; Ma, Z.-H.; Li, L.-W.; Wang, P.; Chen, Y.; Huang, Y.-C.; Cao, Y. Characterization of germline mutations in familial lung cancer from the Chinese population. *Gene* **2018**, *641*, 94–104. [CrossRef]

98. Cui, J.; Yin, Y.; Ma, Q.; Wang, G.; Olman, V.; Zhang, Y.; Chou, W.-C.; Hong, C.S.; Zhang, C.; Cao, S.; et al. Comprehensive characterization of the genomic alterations in human gastric cancer. *Int. J. Cancer* **2014**, *137*, 86–95. [CrossRef]

99. Hu, C.; Zhou, Y.; Liu, C.; Kang, Y. Risk assessment model constructed by differentially expressed lncRNAs for the prognosis of glioma. *Oncol. Rep.* **2018**, *40*, 2467–2476. [CrossRef] [PubMed]

100. 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]

101. Nault, J.C.; Calderaro, J.; Di Tommaso, L.; Balabaud, C.; Zafrani, E.S.; Bioulac-Sage, P.; Roncalli, M.; Zucman-Rossi, J. Telomerase reverse transcriptase promoter mutation is an early somatic genetic alteration in the transformation of premalignant nodules in hepatocellular carcinoma on cirrhosis. *Hepatology* **2014**, *60*, 1983–1992. [CrossRef]

102. Nault, J.C.; Mallet, M.; Pilati, C.; Calderaro, J.; Bioulac-Sage, P.; Laurent, C.; Laurent, A.; Cherqui, D.; Balabaud, C.; Zucman-Rossi, J. High frequency of telomerase reverse-transcriptase promoter somatic mutations in hepatocellular carcinoma and preneoplastic lesions. *Nat. Commun.* **2013**, *4*, 2218. [CrossRef]

103. Nault, J.-C.; Zucman-Rossi, J. Genetics of hepatocellular carcinoma: The next generation. *J. Hepatol.* **2014**, *60*, 224–226. [CrossRef] [PubMed]

104. Pinyol, R.; Tovar, V.; Llovet, J.M. TERT promoter mutations: Gatekeeper and driver of hepatocellular carcinoma. *J. Hepatol.* **2014**, *61*, 685–687. [CrossRef]

105. Quaas, A.; Oldopp, T.; Tharun, L.; Klingenfeld, C.; Krech, T.; Sauter, G.; Grob, T.J. Frequency of TERT promoter mutations in primary tumors of the liver. *Virchows Arch.* **2014**, *465*, 673–677. [CrossRef] [PubMed]

106. Totoki, Y.; Tatsuno, K.; Covington, K.R.; Ueda, H.; Creighton, C.J.; Kato, M.; Tsuji, S.; A Donehower, L.; Slagle, B.L.; Nakamura, H.; et al. Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. *Nat. Genet.* **2014**, *46*, 1267–1273. [CrossRef] [PubMed]