

The integration model of hepatitis B virus genome in hepatocellular carcinoma cells based on high-throughput long-read sequencing

Weiyang Li^{a,b,*}, Wei Wei^a, Fei Hou^a, Hanshi Xu^c, Xiaofang Cui^{a,*}

^a Jining Medical University, Jining, Shandong 272067, China

^b Collaborative Innovation Center for Birth Defect Research and Transformation of Shandong Province, Jining Medical University, Jining, Shandong 272067, China

^c Department of Nasopharyngeal Carcinoma, Sun Yat-sen University Cancer Center, Guangzhou, Guangdong 510275, China

ARTICLE INFO

Keywords:

HBV Integration
HCC
Long-reads sequencing
HBV insertion sequence
Chromosome translocation
Perfect integration mode

ABSTRACT

HBV integration and function has gradually been expanding. However, the exact mode of HBV integration remains unclear. In our research, the high-throughput long-read sequencing was combined with bioinformatics to study the complete mode of HBV integration in hepatocellular carcinoma (HCC) cells. The results demonstrated that: 1) The HBV insertion sequences of HBV integration events accounted for 49.5% of the total HBV sequences. 2) Short insertion segments with the length of 0–1 kbp accounted for 50% and the long insertion segments (>3 kbp) accounted for 25% of HBV insertion events. 3) There were different HBV insertion length in the breakpoints formed within different regions. 4) The occurrence of HBV integration events was accompanied by more frequent structural variations. 5) Furthermore, multiple HBV integration patterns were confirmed based on complete HBV insertion sequences. Our research not only clarified a variety of perfect HBV integration models but also determined multiple specific features of HBV integration.

1. Introduction

In recent years, more and more attention has been paid to the role of hepatitis B virus (HBV) integration events in the occurrence and development of hepatitis and liver cancer. Previous studies have shown that HBV integration events can lead to the occurrence and development of tumor through at least four aspects: 1) HBV integration can lead to the malfunction of the transcription and changes in expression of key genes; 2) The integration of HBX (Hepatitis B virus X protein) region itself has carcinogenic function; 3) After virus integration, virus protein is continuously expressed, which can induce tumor formation; 4) Virus integration can lead to changes in genome structure and induce structural changes of adjacent or distal chromatin [1–5].

Current method for the detection of HBV integration sites is mostly based on the whole genome sequencing with next-generation sequencing and target region capture technology [6–11]. With the popularization of next-generation sequencing technology, more and more HBV integration sites are detected, and the mechanism of HBV integration events is becoming more and more comprehensive. Interestingly, the information about breakpoints (HBV integration sites) has shown certain clinical relevance [12,13]. However, due to the limitation of next-generation sequencing and Sanger sequencing in the length of

reads, it is difficult to detect two sites and complete HBV insertion sequence produced by one integration event simultaneously. If these two key problems are not effectively resolved, it will negatively affect the elucidation of virus integration characteristics and subsequent functional research. In recent years, the development of the third-generation sequencing technology has shown the advantages in resolving the problems of uncertain distance between two breakpoints as well as the uncertain direction and form of virus integration sequence inserted inside [14–17]. Therefore, in this study, we combined the third-generation sequencing with TSD software and manual check method to explore the basic HBV integration mode [18].

In this study, the pooled DNA samples of hepatocarcinoma tissues were used (Table S1), and ultra-deep sequencing was performed on the samples by Nanopore sequencing with Oxford Nanopore Technologies (ONT) systems. Through self-updated TSD calculation method combined with manual check (Fig. S1), it was found that 36 sequences could accurately identify the two ends and internal insertion sequences produced by an integration event. In this study, the characteristics of HBV insertion sequence length were described, and the loci of two ends formed were closely related to structural variation. We further explored the regularity of virus insertion length and characteristics, and systematically and exhaustively determined the integration mode of HBV

* Corresponding authors at: Jining Medical University, Jining, Shandong 272067, China.

E-mail addresses: 163.lwy@163.com (W. Li), scuexf@163.com (X. Cui).

<https://doi.org/10.1016/j.ygeno.2021.11.025>

Received 12 May 2021; Received in revised form 16 November 2021; Accepted 22 November 2021

Available online 26 November 2021

0888-7543/© 2021 The Authors.

Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

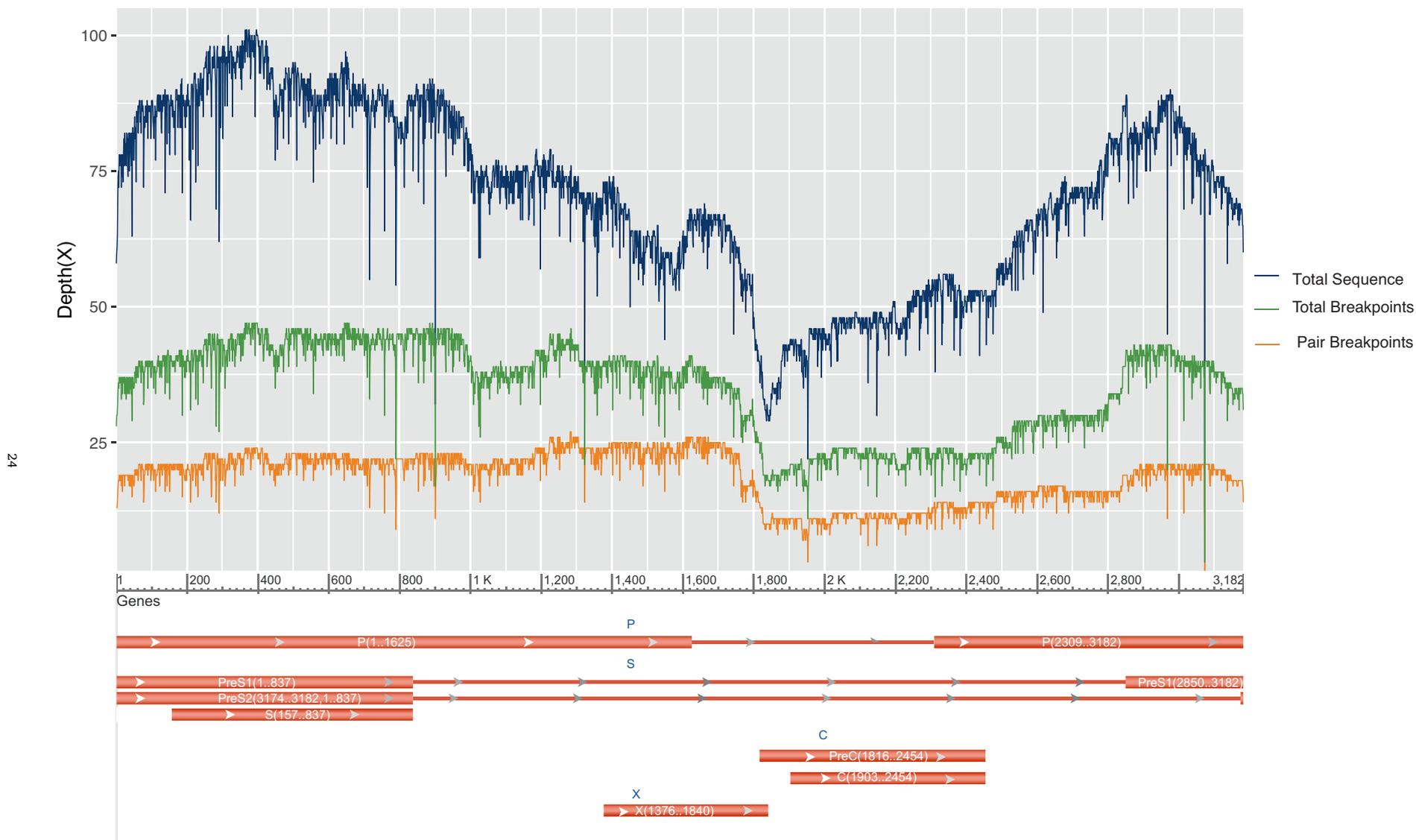


Fig. 1. HBV coverage and depth in three groups.

Total sequence (Blue Line) represented the depth of total HBV fragments; Total breakpoints (Green Line) represented the depth of HBV fragments with human genome breakpoints; Pair breakpoints (Yellow Line) represented the depth of HBV fragments with pair human genome breakpoints.

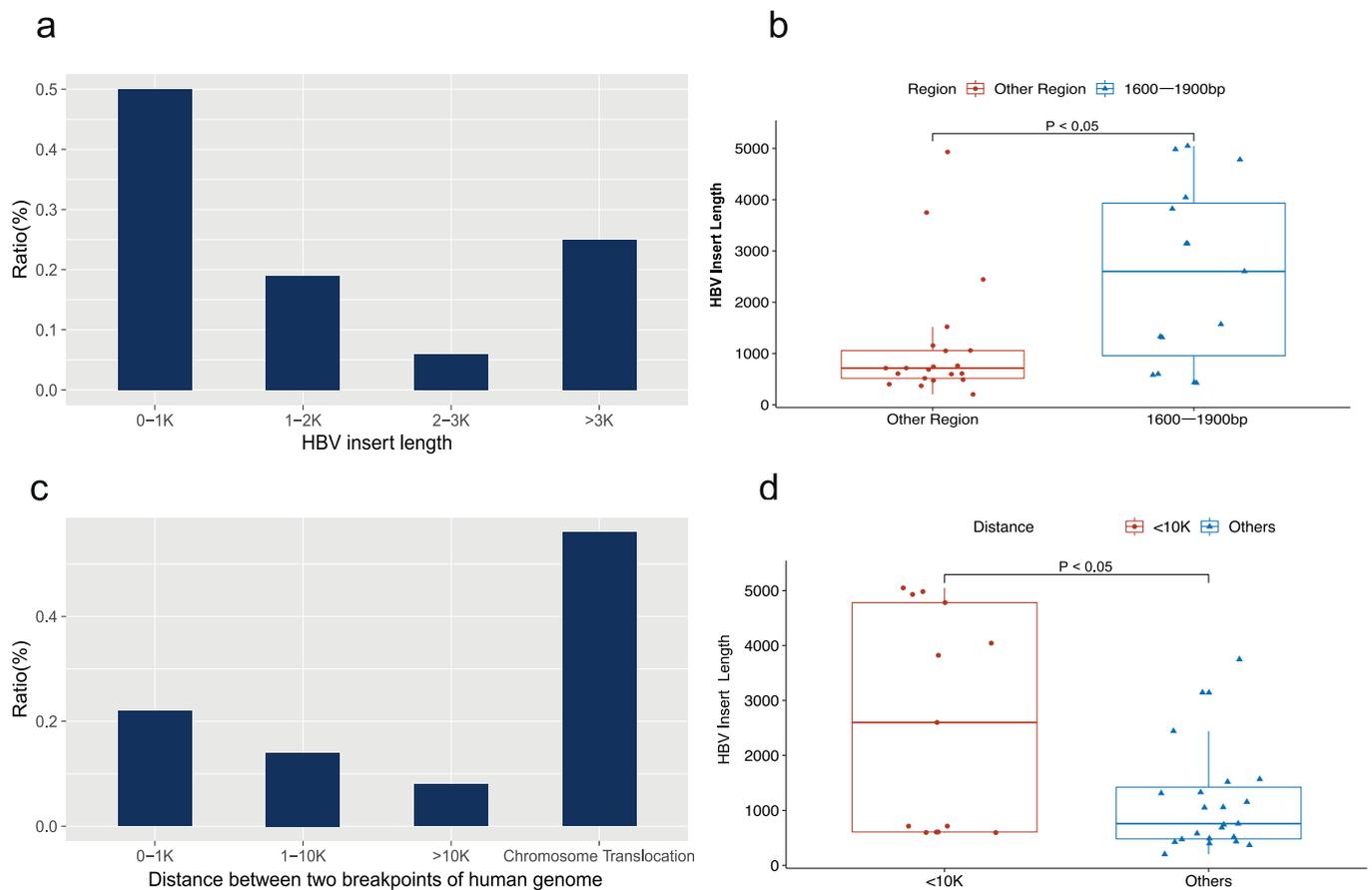


Fig. 2. HBV insert length and genome structure variation.

a) represented the ratio distribution of different HBV integration fragment length; b) represented HBV integration fragment length of breakpoints occurred in the 1600–1900 bp region and other region of HBV genome; Blue triangle represented HBV integration fragment length of breakpoints occurred in the 1600–1900 bp region, Crimson circle represented HBV integration fragment length of breakpoints occurred in the other region; c) represented the ratio distribution of different distance between the pair breakpoints in the human genome; d) represented HBV integration fragment length in the distance <10Kb and other groups; crimson circle represented HBV integration fragment length in the group with the distance <10Kb of pair breakpoints in the human genome, Blue triangle represented HBV integration fragment length with the distance >10Kb (including translocation) of pair breakpoints in the human genome.

integration sequence in liver cancer. This greatly facilitates our understanding of HBV integration modes and the functional of HBV integration, and provides a basis for effective deletion of inserted HBV sequences by gene editing technology.

2. Results

2.1. The average depth of HBV coverage

The average depth of all HBV fragments was 70.5 \times , the coverage depth of HBV sequences with breakpoints was 34.9 \times , and the depth of HBV sequences with complete virus integration mode was 18.6 \times . Furthermore, obviously higher depth was found in the PreS1 region, while lower depth was observed in the X and C region (Fig. 1).

2.2. Analysis on HBV insertion length and structural variation

The most integrated parts of HBV genome are short segments with the length of 0–1 kbp which account for 50% of insertion events. The long segments (>3 kbp) accounted for 25% of all the HBV integration events (Fig. 2a, Table 1, Table S2). Furthermore, through comparison of the breakpoints formed in the HBV genome in the region between 1.6 and 1.9 kbp to the breakpoints formed in other genomic regions, we found that the insertion sequence of HBV genome was longer in the region between 1.6 and 1.9 kbp (Fig. 2b, $P < 0.05$). The occurrence of

HBV integration events was accompanied by more frequent structural changes, of which TL (chromosome translocation) accounts for 56%, and only 22% for the two-end points with the distance below 1kbp (Fig. 2c). We discovered that the HBV integration with genome deletion <10kbp had longer HBV insertion sequence than other types (Fig. 2d, $P < 0.05$).

2.3. Analysis of HBV insertion pattern

The hot-spot regions from previous studies were selected to study the integration patterns of the integration sites found in our research. The results demonstrated that TERT region had two kinds of integration patterns at least. One was 4043 bp long and the other was 2600 bp long, with both containing complete PreS1 region (Fig. 3a, Fig. 3b, Table 1, Table S2). The insertion length on KMT2B was 713 bp, involving the complete X region (Fig. 3d). Chromosome translocation (Chr16/Chr17) had insertion with the length of 1520 bp, involving complete X and C regions (Fig. 3c). The region of DNAH9 was inserted with 5048 bp, which included the full-length sequences of complete HBV (Fig. 3f, Table 1, Table S2, Table S3).

3. Discussions

The depth of HBV sequence with integration site accounted for about 49.5% of the total average depth, indicating that the HBV integration

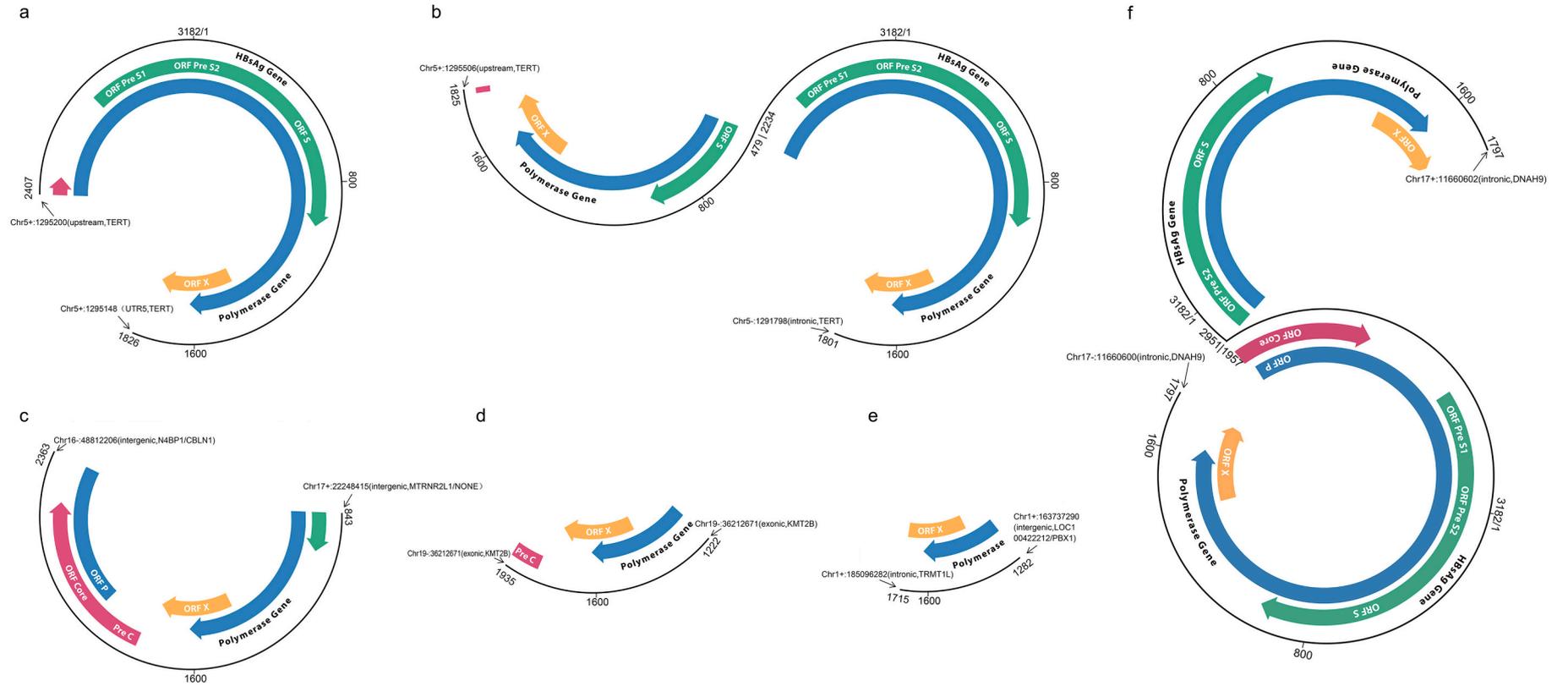


Fig. 3. HBV integration mode in the hotspot regions

a) HBV insertion sequence and two breakpoints in TERT gene; b) HBV insertion sequence and two breakpoints in TERT gene; c) HBV insertion sequence and two breakpoints in the chromosome translocation region; d) HBV insertion sequence and two breakpoints in the KMT2B gene; e) HBV insertion sequence and two breakpoints in the intron and intergenic region; f) HBV insertion sequence and two breakpoints in the DNAH9 gene. Black arrow represented the chromosome position of two breakpoints, HBV genes with different functions are shown in the inner circles.

Table 1
Summary of HBV integration mode.

Left-Chr	Left-breakpoint	HBV genome Start	HBV insert length	HBV genome End	Right-Chr	Right-Breakpoint
chr1+	185,096,282	1715	433	1282	chr1+	163,737,290
chr6-	116,347,205	1797	4781	1852	chr6+	116,347,215
chr17-	11,660,600	1797	5048	1797	chr17+	11,660,602
chr6-	116,347,205	1797	3144	1760	chr4-	62,916,487
chr1+	249,240,419	1763	3143	1797	chr6+	116,347,205
chr19+	36,212,670	1221	714	1935	chr19+	36,212,674
chr19-	36,212,671	1935	713	1222	chr19-	36,212,671
chr18+	64,512,544	3086	3748	2603	chr3+	166,536,190
chr17-	11,660,600	1797	4982	1797	chr17+	11,660,600
chr5+	2,843,075	1285	2443	2023	chr17-	11,747,600
chr17-	22,250,071	2873	596	291	chr17+	22,256,723
chr16-	48,812,206	2363	1520	843	chr17+	22,248,415
chr12+	5,874,726	2310	4930	2616	chr12+	5,874,718
chr16+	48,857,924	319	3822	1833	chr16-	48,854,612
chr1-	163,737,290	1282	424	1706	chr1-	185,096,281
chr8-	60,470,354	2890	605	332	chr8+	60,464,493
chr10-	4,653,495	2284	399	2683	chr7-	2,735,921
chr7+	149,038,603	2205	607	1598	chr7-	149,040,564
chr5-	24,285,266	1180	579	1759	chr2+	141,553,668
chr5-	24,285,245	1170	598	1768	chr5+	24,285,398
chr5+	1,295,506	1825	4043	1801	chr5-	1,291,798
chr5+	1,295,148	1826	2600	2407	chr5+	1,295,200
chr4+	5,564,310	1305	1057	248	chr10-	42,376,002
chr10-	42,597,048	2848	741	414	chr3+	75,691,193
chr4+	5,564,279	1302	1051	251	chr4+	49,635,295
chr16+	84,289,588	1056	201	1257	chr3-	17,719,387
chr3-	75,691,186	430	759	2848	chr10+	42,387,516
chr7+	2,570,694	1675	1311	364	chr3-	35,766,748
chr7-	109,591,657	3079	1154	1392	chr3-	75,692,765
chr3-	75,691,274	379	687	2873	chr10+	42,387,516
chr1+	249,240,538	1849	1568	3130	chr19+	245,596
chr13-	65,967,656	851	473	378	chr14-	79,918,007
chr5-	11,505	2486	368	2854	chr16+	79,840,919
chr14+	104,521,988	2939	488	2018	chr3+	5,469,100
chr7-	71,179,695	1831	1328	503	chr13-	105,879,661
chr12+	95,672	450	515	2927	chr10+	3,104,853

form was an important existence form of HBV in HCC cells. Considering that ONT sequencing is a PCR-free unbiased sequencing, according to the average depth of HBV ($70.5\times$) and human genome ($44\times$), there were 3.2 copies of HBV genome in one cell ($70.5\times 2/44$), and there were 1.6 copies of integrated HBV in one cell ($34.9\times 2/44$). The probability of integration event was much higher than $1/10^5-10^6$ previously reported [19,20]. This finding indicated that a large number of cells with HBV integration were obviously enriched in HCC cells. Besides, through the comparison of HBV depth, it was found that HBV with integration sites accounted for about 49.5% of all HBV fragments, indicating that HBV was a very stable and key existence form in HCC cells. This may suggest that cells that frequently express virus antigens can be efficiently killed, while those cells with increased growth signal response, lost virus expression and failed antigen immune process can gain growth advantages [21]. Interestingly, the cells with HBV integration are consistent with the above characteristics [22]. Thus, the integrated hepatocytes show a more stable existence form, which also indicates that part of HCC cells are likely to originate from hepatocytes with HBV integration. Meanwhile, we observed that the interval of HBV integration shows certain patterns, and we found that the coverage depth of the region encoding PRE S1 protein is higher than in the other regions, indicating that this region tends to be completely preserved. However, why this happens remains to be further studied.

There is a larger distance between the two ends of HBV virus integration, with a large number of structural variations, which indicates that HBV virus integration is accompanied by frequent variations in genome structure. According to the results of our research, 56% of the two ends of HBV integration events show chromosome translocation. Besides, virus integration also leads to frequent deletion of genome sequences, which provides more direct evidence for the instability of genome caused by HBV integration. Since the normally generated HBV

double stranded linear DNA (dsDNA) are in the range of 1816–3182/1–1832 [4], and our research results show that the 32/36 sequence is in this range, it is speculated that the HBV integrated sequences are mainly derived from normal HBV dsDNA. After these HBV dsDNA are integrated into the human genome through various methods such as Classical non-homologous end joining (C-NHEJ), microhomology-mediated end joining (MMEJ), Single-strand annealing (SSA) etc. [23], they evolve into the characteristics of existing HBV integrated fragments through rearrangement, deletion, secondary integration and other methods, as well as the interaction with cell selection. However, we also observed that some inserted HBV fragments could cover HBV genome linear end, that is, the HBV insertion fragment was in the range (843–2263 bp) of HBV genome, indicating that there were other mechanisms of virus integration. In general, it is difficult for normal HBV dsDNA to cross the 1600–1900 bp region, so we speculated that the possible mechanism for the HBV integration fragments that cross this region is as follows: 1) The reverse transcription process leads to the formation of excessively redundant HBV dsDNA sequences, which are integrated into the human genome through the C-NHEJ and MMEJ mechanism (Fig. 4a, Fig. 4b) [4]. 2) The abnormally redundant HBV ssDNA is integrated into the human genome through the SSA mechanism (Fig. 4c). 3) The closed circular HBV genome is integrated into the human genome through the MMEJ mechanism [24] (Fig. 4d). The HBV sequences integrated into the human genome through the above methods and evolved into the existing integrated model through rearrangement, deletion, and cell selection and evolution. Due to limited research on the characteristics and mechanisms of HBV integration, the integration mechanism of HBV sequence in the normal integration form remains unclear, and research on the virus integration mechanism of these exceptional cases is lacking. Therefore, the clarification of HBV integration mechanism and characteristics still requires numerous in-depth studies. In addition, we

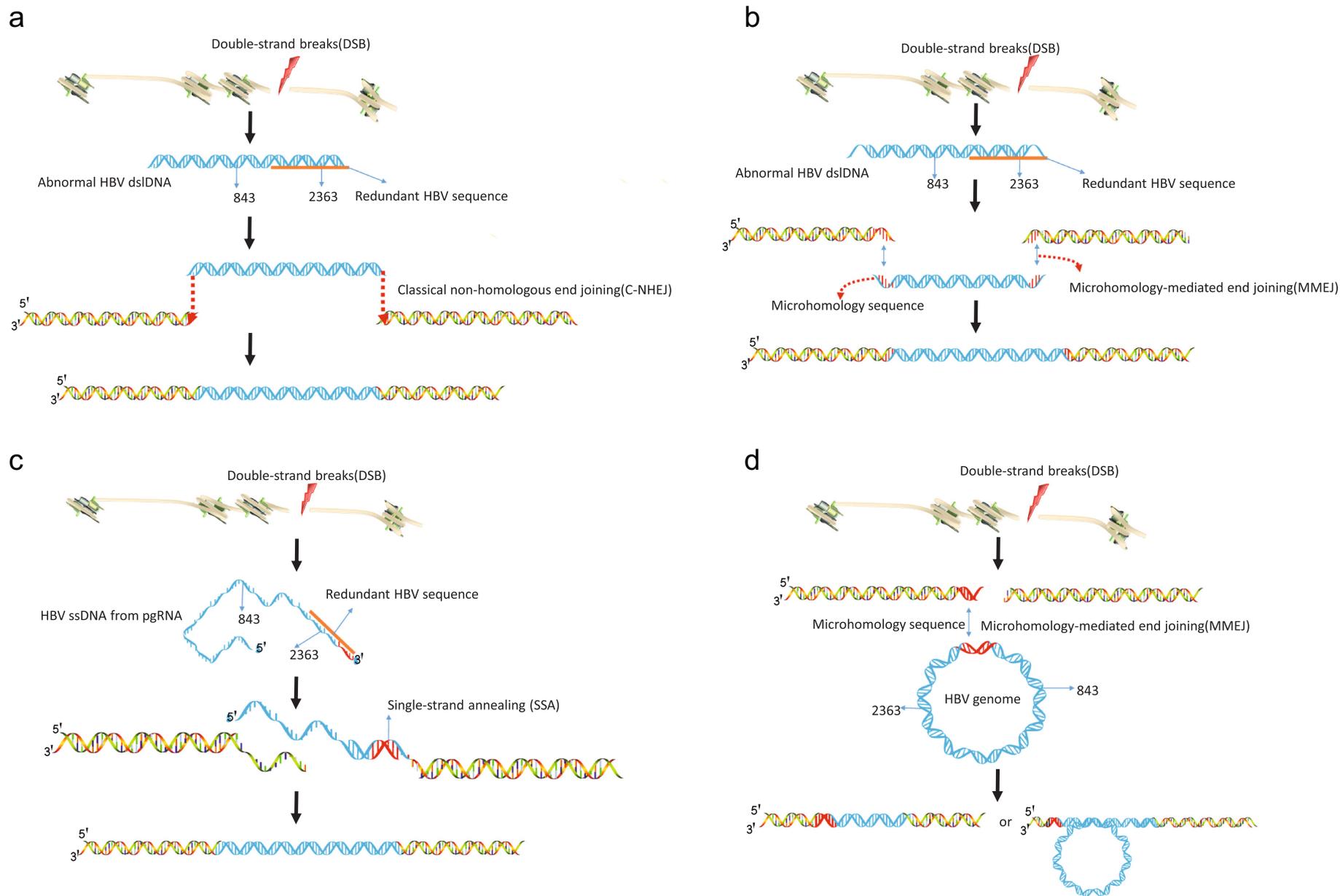


Fig. 4. Potential mechanism of HBV integration fragments spanning the 1600-1900 bp region.

a) HBV double stranded linear DNA (dsDNA) integration by Classical non-homologous end joining(C-NHEJ); b) HBV double stranded linear DNA (dsDNA) integration by microhomology-mediated end joining (MMEJ); c) HBV single-stranded DNA (ssDNA) integration by single-stranded annealing (SSA); d) HBV circular DNA integration by microhomology-mediated end joining (MMEJ).

selected data on HBV and HPV from previous studies [12,24] to compare the integrated genes of the viruses. Interestingly, 364 genes were shared in the integrated genes of previous HPV and HBV studies (Fig. S2, Table S4). We adopted the 364 shared genes to conduct pathway analysis. These genes were significantly enriched in the pathway of axon guidance, glutamatergic synapse, GABAergic synapse, retrograde endocannabinoid signaling, morphine addiction and ovarian steroidogenesis (Fig. S3, Table S5). Most of the pathways are associated with the nervous system, so it is speculated that both viral integration events are closely related to the nerve signal conduction status of patients.

TERT and KMT2B have been identified as the hot spots of HBV virus integration. In this study, the complete HBV integration pattern in the region of these genes was also found. Two integration patterns were found in TERT region, one with an insertion length of 4043 bp, forming a special integration form (as shown in Fig. 3), and the other with an insertion length of 2600 bp. They both retained the complete PreS1 region, while the KMT2B insertion was a short segment of 713 bp, which retained the complete X region. The chromosome translocation (Chr16/Chr17) insertion retained the complete X and C regions. Hence, it could be seen that HBV insertion had diversity, showing different insertion modes in different genomic positions, but the exact mechanism remains to be studied.

In this study, a comprehensive integration model of HBV was found through third-generation sequencing with ONT, combined with the analysis of biological information. This integration model provides strong support for the development of HBV integration algorithm, and also provides a theoretical basis for the research of HBV integration carcinogenesis and function.

4. Materials and methods

We adopted 20 liver cancer DNA samples with pooled DNA (20 µg). After DNA was purified by magnetic beads, the ends of DNA fragments were repaired. The purified product was connected by sequencing linker in SQK-LSK109 kit. The constructed DNA library was quantified by Qubit. After the construction of the library was completed, the DNA library was added to the flow cell, and the flow cell was transferred to the PromethION sequencer (Oxford Nanopore Technologies, Oxford, UK) for real-time single molecule sequencing [25] (Wuhan Benagen Tech Solutions Company Limited). Low-quality sequences (sequences with average mass value less than or equal to 7 were excepted) and adaptor sequences were removed from the original sequencing data, thus obtaining 132.99GB of clean data [18]. Then, these sequences containing partial or complete HBV genome are retained after clean data is mapped to the HBV genome (NC_003977.2) by using local BLAST (NCBI-Blast-2.11.0). Afterwards, the virus integration site was detected by TSD analytical method [18], with HG19 and NC_003977.2 as reference genomes (Fig. S1). Finally, HBV and human genome end points were determined by manual check.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ygeno.2021.11.025>.

Data availability

All the data that support the present study have been deposited in the Supplementary Table and NCBI (PRJNA736806).

Funding

This work was supported by the Natural Science Foundation of Shandong Province in China [grant number ZR2019MC024].

Author contribution

W.Y.L. conceived and designed the research. W.W., F.H., H.S.X analyzed the data. X.F.C. and W.Y.L wrote the main manuscript. All

authors reviewed the manuscript.

Ethics approval

All procedures performed in the present study involving human participants were carried out in accordance with the ethical standards of the institutional research committee and the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The present study had been approved by the Ethics Review Committee of the Jining Medical University and informed consent was obtained from each individual.

Declaration of Competing Interest

The authors declare that there are no competing interests associated with the manuscript.

Acknowledgements

We sincerely thank Guofeng Meng and Yuhui Sun for excellent advice, as well as all the study participants, research staff, and students who participated in this work.

References

- [1] T.C. Tseng, L.R. Huang, Immunopathogenesis of Hepatitis B Virus, *J. Infect. Dis.* 216 (2017) S765–S770.
- [2] H.E. Blum, D. Moradpour, Viral pathogenesis of hepatocellular carcinoma, *J. Gastroenterol. Hepatol.* 17 (Suppl. 3) (2002) S413–S420.
- [3] L. Zong, H. Peng, C. Sun, F. Li, M. Zheng, Y. Chen, H. Wei, R. Sun, Z. Tian, Breakdown of adaptive immunotolerance induces hepatocellular carcinoma in HBsAg-tg mice, *Nat. Commun.* 10 (2019) 221.
- [4] T. Tu, M.A. Budzinska, N.A. Shackel, S. Urban, HBV DNA Integration: Molecular Mechanisms and Clinical Implications, *Viruses*, 9, 2017.
- [5] T. Tu, M.A. Budzinska, N.A. Shackel, A.R. Jilbert, Conceptual models for the initiation of hepatitis B virus-associated hepatocellular carcinoma, *Liver Int. : Off. J. Int. Assoc. Study Liver* 35 (2015) 1786–1800.
- [6] R. Rajaby, Y. Zhou, Y. Meng, X. Zeng, G. Li, P. Wu, W.K. Sung, SurVirus: a repeat-aware virus integration caller, *Nucleic Acids Res.* 49 (2021), e33.
- [7] X. Zeng, L. Zhao, C. Shen, Y. Zhou, G. Li, W.K. Sung, HIVID2: an accurate tool to detect virus integrations in the host genome, *Bioinformatics* 37 (2021) 1821–1827.
- [8] W. Li, X. Zeng, N.P. Lee, X. Liu, S. Chen, B. Guo, S. Yi, X. Zhuang, F. Chen, G. Wang, R.T. Poon, S.T. Fan, M. Mao, Y. Li, S. Li, J. Wang, X. Xu Jianwang, H. Jiang, X. Zhang, HIVID: an efficient method to detect HBV integration using low coverage sequencing, *Genomics* 102 (2013) 338–344.
- [9] X. Cui, W. Wei, C. Wang, Y. Qi, X. Qin, L. Huang, W. Li, Studies on the correlation between mutation and integration of HBV in hepatocellular carcinoma, *Biosci. Rep.* 40 (2020).
- [10] W. Li, Y. Qi, H. Xu, W. Wei, X. Cui, Profile of different Hepatitis B virus integration frequency in hepatocellular carcinoma patients, *Biochem. Biophys. Res. Commun.* 553 (2021) 160–164.
- [11] N.D. Nguyen, V. Deshpande, J. Luebeck, P.S. Mischel, V. Bafna, ViFi: accurate detection of viral integration and mRNA fusion reveals indiscriminate and unregulated transcription in proximal genomic regions in cervical cancer, *Nucleic Acids Res.* 46 (2018) 3309–3325.
- [12] L.H. Zhao, X. Liu, H.X. Yan, W.Y. Li, X. Zeng, Y. Yang, J. Zhao, S.P. Liu, X. H. Zhuang, C. Lin, C.J. Qin, Y. Zhao, Z.Y. Pan, G. Huang, H. Liu, J. Zhang, R. Y. Wang, Y. Yang, W. Wen, G.S. Lv, H.L. Zhang, H. Wu, S. Huang, M.D. Wang, L. Tang, H.Z. Cao, L. Wang, T.L. Lee, H. Jiang, Y.X. Tan, S.X. Yuan, G.J. Hou, Q. F. Tao, Q.G. Xu, X.Q. Zhang, M.C. Wu, X. Xu, J. Wang, H.M. Yang, W.P. Zhou, H. Y. Wang, Genomic and oncogenic preference of HBV integration in hepatocellular carcinoma, *Nat. Commun.* 7 (2016) 12992.
- [13] W.K. Sung, H. Zheng, S. Li, R. Chen, X. Liu, Y. Li, N.P. Lee, W.H. Lee, P. N. Ariyaratne, C. Tennakoon, F.H. Mulawadi, K.F. Wong, A.M. Liu, R.T. Poon, S. T. Fan, K.L. Chan, Z. Gong, Y. Hu, Z. Lin, G. Wang, Q. Zhang, T.D. Barber, W. C. Chou, A. Aggarwal, K. Hao, W. Zhou, C. Zhang, J. Hardwick, C. Buser, J. Xu, Z. Kan, H. Dai, M. Mao, C. Reinhard, J. Wang, J.M. Luk, Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma, *Nat. Genet.* 44 (2012) 765–769.
- [14] C.J. Westbrook, J.A. Karl, R.W. Wiseman, S. Mate, G. Koroleva, K. Garcia, M. Sanchez-Lockhart, D.H. O'Connor, G. Palacios, No assembly required: Full-length MHC class I allele discovery by PacBio circular consensus sequencing, *Hum. Immunol.* 76 (2015) 891–896.
- [15] B.D. Betz-Stablein, A. Topfer, M. Littlejohn, L. Yuen, D. Colledge, V. Sozzi, P. Angus, A. Thompson, P. Revill, N. Beerwinkel, N. Warner, F. Luciani, Single-molecule sequencing reveals complex genome variation of hepatitis B virus during 15 years of chronic infection following liver transplantation, *J. Virol.* 90 (2016) 7171–7183.

- [16] C. Peneau, S. Imbeaud, T. La Bella, T.Z. Hirsch, S. Caruso, J. Calderaro, V. Paradis, J.F. Blanc, E. Letouze, J.C. Nault, G. Amaddeo, J. Zucman-Rossi, Hepatitis B virus integrations promote local and distant oncogenic driver alterations in hepatocellular carcinoma, *Gut* 0 (2021) 1–11.
- [17] A.L. McNaughton, H.E. Roberts, D. Bonsall, M. de Cesare, J. Mokaya, S.F. Lumley, T. Golubchik, P. Piazza, J.B. Martin, C. de Lara, A. Brown, M.A. Ansari, R. Bowden, E. Barnes, P.C. Matthews, Illumina and Nanopore methods for whole genome sequencing of hepatitis B virus (HBV), *Sci. Rep.* 9 (2019) 7081.
- [18] G. Meng, Y. Tan, Y. Fan, Y. Wang, G. Yang, G. Fanning, Y. Qiu, TSD: a computational tool to study the complex structural variants using PacBio targeted sequencing data, *G3* 9 (2019) 1371–1376.
- [19] J. Summers, A.R. Jilbert, W. Yang, C.E. Aldrich, J. Saputelli, S. Litwin, E. Toll, W. S. Mason, Hepatocyte turnover during resolution of a transient hepadnaviral infection, *Proc. Natl. Acad. Sci. U. S. A.* 100 (2003) 11652–11659.
- [20] W. Yang, J. Summers, Integration of hepadnavirus DNA in infected liver: evidence for a linear precursor, *J. Virol.* 73 (1999) 9710–9717.
- [21] W.S. Mason, A.R. Jilbert, J. Summers, Clonal expansion of hepatocytes during chronic woodchuck hepatitis virus infection, *Proc. Natl. Acad. Sci. U. S. A.* 102 (2005) 1139–1144.
- [22] W.S. Mason, H.C. Low, C. Xu, C.E. Aldrich, C.A. Scougall, A. Grosse, A. Clouston, D. Chavez, S. Litwin, S. Peri, A.R. Jilbert, R.E. Lanford, Detection of clonally expanded hepatocytes in chimpanzees with chronic hepatitis B virus infection, *J. Virol.* 83 (2009) 8396–8408.
- [23] A. Sfeir, L.S. Symington, Microhomology-mediated end joining: a back-up survival mechanism or dedicated pathway? *Trends Biochem. Sci.* 40 (2015) 701–714.
- [24] Z. Hu, D. Zhu, W. Wang, W. Li, W. Jia, X. Zeng, W. Ding, L. Yu, X. Wang, L. Wang, H. Shen, C. Zhang, H. Liu, X. Liu, Y. Zhao, X. Fang, S. Li, W. Chen, T. Tang, A. Fu, Z. Wang, G. Chen, Q. Gao, S. Li, L. Xi, C. Wang, S. Liao, X. Ma, P. Wu, K. Li, S. Wang, J. Zhou, J. Wang, X. Xu, H. Wang, D. Ma, Genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism, *Nat. Genet.* 47 (2015) 158–163.
- [25] D. Deamer, M. Akeson, D. Branton, Three decades of nanopore sequencing, *Nat. Biotechnol.* 34 (2016) 518–524.