

LETTER TO THE EDITOR

Genome-wide virus-integration analysis reveals a common insertional mechanism of HPV, HBV and EBV

Dear Editor,

Human papillomavirus (HPV), hepatitis B virus (HBV) and Epstein–Barr virus (EBV) are the three most oncogenic DNA viruses, contributing to 15 different types of cancer.¹ Although these viruses differ in many aspects, one common key step is the integration of their DNA into the human genome, which could potentially promote carcinogenesis.^{2–4} In this study, we developed and performed a novel pipeline (Figures S1–S8, Supplementary Notes 1–3 and Table S1) named viral integration pathway analysis (VIPA) to elucidate the integration mechanism shared by HPV, HBV and EBV, thus gaining a deeper understanding towards the virus-induced carcinogenesis and the corresponding anticancer therapies.

First, we conducted HPV capture sequencing and identified 1002 HPV integration breakpoints in 24.8% (225/910) non-cancer HPV infection samples, 588 breakpoints in 38.0% (125/329) cervical precancer samples and 1597 breakpoints in 69.0% (158/227) cancer samples (Figure 1A). The total integration sample proportion was 34.7% (508/1466), and the average integration breakpoints were 6.27 per sample. We observed 24 recurrent integration hotspots (integration positions located within the 500-kb downstream/upstream of the gene, $n \geq 5$) in our dataset (Figure 1A). Among them, 10 integration hotspots were previously reported, and 14 HPV integration hotspot genes were newly identified (Table S2).

Next, we found that the distribution of HPV integration strains and status in non-cancer HPV infection, cervical precancer and cancer samples were different (Figure 1B,C). Specifically, HPV16 integration percentage was only 10% (ranked third) in non-cancer samples but increased to 33.4% (ranked first) in precancer and 55.5% (ranked first) in cancer samples. HPV18 integration percentage was only 3.1% in non-cancer samples, and 5.8% in precancer samples, and rose to 7.9% (ranked second) in cancer samples.

The average integration events for non-cancer infection were 4.4, for cervical precancer were 4.7 and for cancer samples were 10.1, indicating that HPV integration increased along with the disease progression (non-cancer vs. precancer, $p = .011$; precancer vs. cancer, $p < .0001$; Wilcoxon test, False Discovery Rate corrected) and may serve as an early warning biomarker of carcinogenesis (Figure 1C). When applying the average integration events to predict clinical outcomes, the results showed that we could distinguish high-grade squamous intraepithelial lesion (HSIL) \pm (including HSIL and Cancer) with an AUC of .722. Further, we found that HPV16 held best prediction performance towards HSIL \pm with the AUC of .859. Similarly, HPV18 shared comparable prediction performance towards HSIL \pm with the AUC of .819 (Figure 1D).

Further, motivated by the aim of finding common integration features among HPV, HBV and EBV, we collected the capture sequencing data of the three viruses. Together, we detected 4390 integration breakpoints for HPV, 4010 integration breakpoints for HBV and 174 integration breakpoints for EBV (Tables S3–S5). Intriguingly, 21 integration genes were shared by all three viruses (Table S6), indicating the potential roles of these genomic loci in oncogenic viruses-related cancers.

Next, we explored the viral integration patterns using identified human–viral junctional sequences (defined by ≥ 30 -bp human and viral sequences at the integration sites) from expanded integration datasets (Table S7 and Supplementary Notes 4 and 5). Previous studies have indicated that the integrations of three viruses were mediated by microhomology (MH)^{4–7} (Figure S9). However, it is not clear how the lateral microhomologies (defined as microhomologies with short-distance from the junction sites) mediate the integration process (Figure 2A–C). Inspired by the new understandings towards alternative end-joining,^{8,9} we speculated that synthesis-dependent

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Clinical and Translational Medicine* published by John Wiley & Sons Australia, Ltd on behalf of Shanghai Institute of Clinical Bioinformatics.

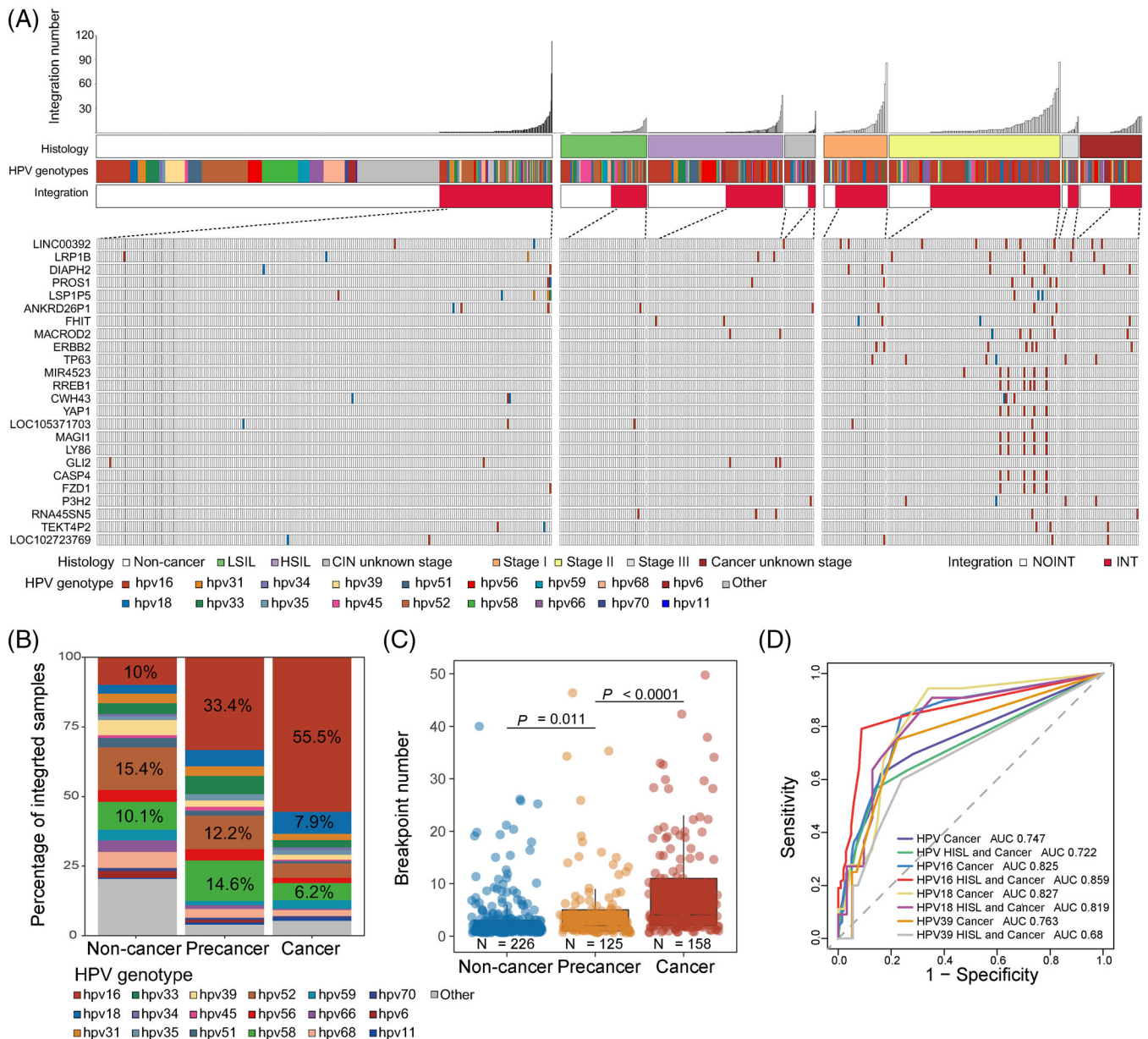


FIGURE 1 The integration landscape of new human papillomavirus (HPV) positive samples. (A) The landscape of our new HPV positive samples, including 910 HPV infection samples, 329 cervical precancer samples and 227 cervical cancer samples. The integration sample proportions were 24.8% for non-cancer HPV infection (225/910), 38.0% for cervical precancer (125/329) and 69.0% for cancer stages (158/227). Among previous HPV-integrated samples, there were 24 recurrent integration genes ($n \geq 5$ samples) were shown; (B) the distribution of integrated HPV strains in three cervical disease stages. The percentages of top three HPV strains were marked; (C) the average integration events among non-cancer infection, cervical precancer and cancer. Adjusted p values were calculated by Wilcox test; (D) the ROC of different HPV strains' average integration events to predict stages more severe than high-grade squamous intraepithelial lesion (HSIL) (HSIL and cancer) or cancer

end-joining (SD-EJ) pathway may participate in the integration process to generate multiple types of breakpoints (Figure S10), including apparent blunt joining (Figure 2A), short insertion (Figure 2B) and junctional microhomologies (Figure 2C). We validated integration structures using the nanopore sequencing of Ca Ski DNA

and Sanger sequencing of Ca Ski, HepG2.2.15 and Raji (Figures S11 and S12).

We analysed the roles of SD-EJ using computational simulation (Figure S13) in 4341 human-HPV junctional sequences (Table S3), 4010 human-HBV junctional sequences (Table S4) and 169 human-EBV junctional

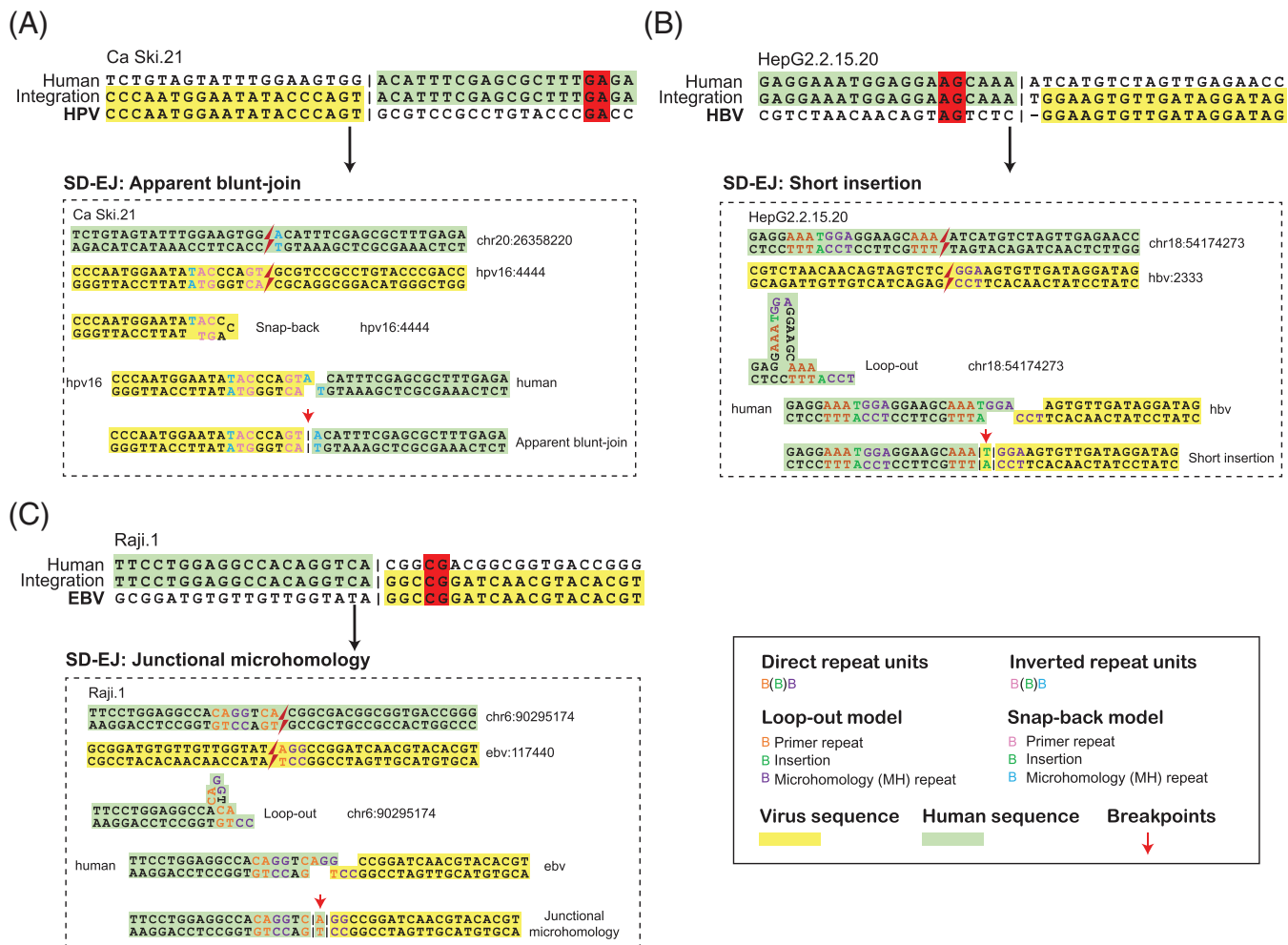


FIGURE 2 The illustration of synthesis-dependent microhomology-mediated end-joining (SD-EJ) integration pathway. Examples of SD-EJ proposal model in the integration process of Ca Ski.21 breakpoint (A), HepG2.2.15.20 (B) and Raji.1 (C), which had lateral microhomologies

sequences (Table S5). We found that SD-EJ was significantly enriched for all three viruses (Figure 3A).

Then, the repair models and products of SD-EJ were further analysed (Figure 3B). The proportions of loop-out model were 47.9%–61.4% (HPV: 61.4%; HBV: 57.7% and EBV: 47.9%), whereas those of snap-backs were 38.8%–52.1% (HPV: 38.8%; HBV: 42.3% and EBV: 52.1%). For repair products, junctional MH was the major type, accounting for 89.5% HPV, 91.3% HBV and 88.1% EBV SD-EJ integration events, followed by apparent blunt join (HPV: 8.4%; HBV: 7.9% and EBV: 10.4%) and short insertion (HPV: 2.0%; HBV: .8% and EBV: 1.5%). The occurrence of junctional MH was significantly higher in the observed group than in the expected group (Figure 3C, Supplementary Note 6). Conversely, the occurrence of apparent blunt join was significantly lower in the observed group than in the expected group. Of note, the

significant enrichment of short insertion was observed in HPV and HBV datasets, whereas there was no significant difference of short insertion between EBV's observed and expected groups ($n = 1$ vs. $n = .14$, $p = 1$, Fisher's exact test) due to relatively small dataset (Figure 3C, Supplementary Note 6).

Finally, we classified integration pathways of each dsDNA virus breakpoint into three categories: (i) SD-EJ pathway with SD-EJ structures, followed by (ii) other alt-EJ pathway with microhomologies overhangs and otherwise (iii) NHEJ pathway without the previous two signatures (Figure 3D). In 10-bp flanking length, we observed the percentages of SD-EJ pathway were 59.11% for HPV, 65.04% for HBV and 48.38% for EBV, whereas those of unclassified NHEJs were 37.15% for HPV, 28.29% for HBV and 48.55% for EBV (Figure 3E). The previous data suggested that SD-EJ repair pathway may play an important

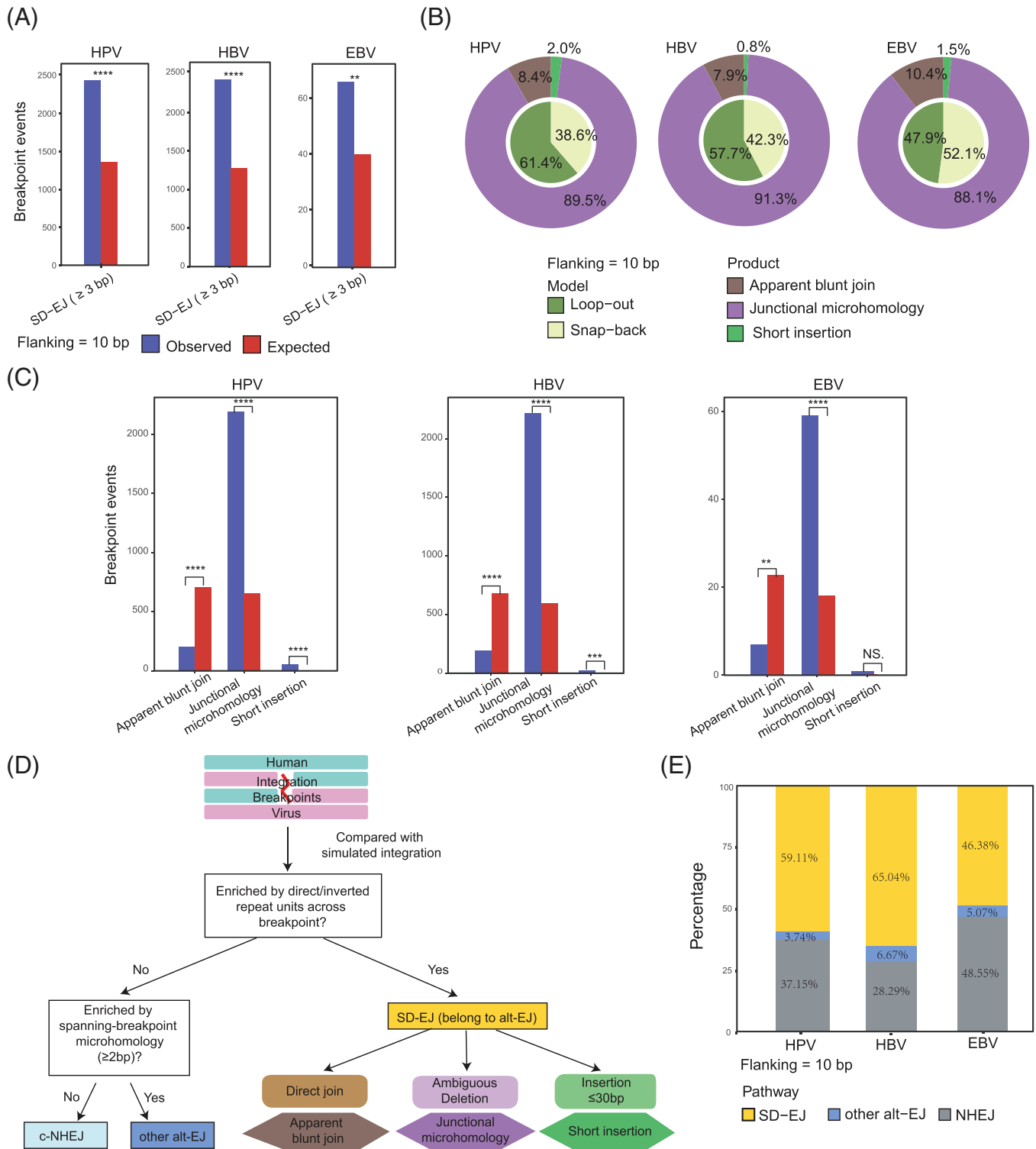


FIGURE 3 The synthesis-dependent end-joining (SD-EJ) pathways in human papillomavirus (HPV), hepatitis B virus (HBV) and Epstein-Barr virus (EBV) integration datasets. (A) The comparison of integration events with SD-EJ repeats (≥ 3 bp) between observed (actual) and expected (simulated) groups within 10-bp flanking length. The previous p values were calculated by Fisher's exact test. $*p < .05$, $**p < .01$, $***p < .001$, $****p < .0001$; (B) the composition of two models (loop-out and snap-back) and three products (apparent blunt joint, junctional microhomology and short insertion) of SD-EJ integration events in HPV, HBV and EBV datasets within 10-bp flanking length; (C) the comparison of three products (apparent blunt joint, junctional microhomology and short insertion) between observed and expected groups within 10-bp flanking length for HPV, HBV, EBV datasets. The previous p values were calculated by Fisher's exact test. $*p < .05$, $**p < .01$, $***p < .001$, $****p < .0001$. (D) The workflow details of further classification of integration pathways; (E) the proportions of SD-EJ, other alt-EJ and c-NHEJ pathways in HPV, HBV and EBV datasets within 10-bp flanking length

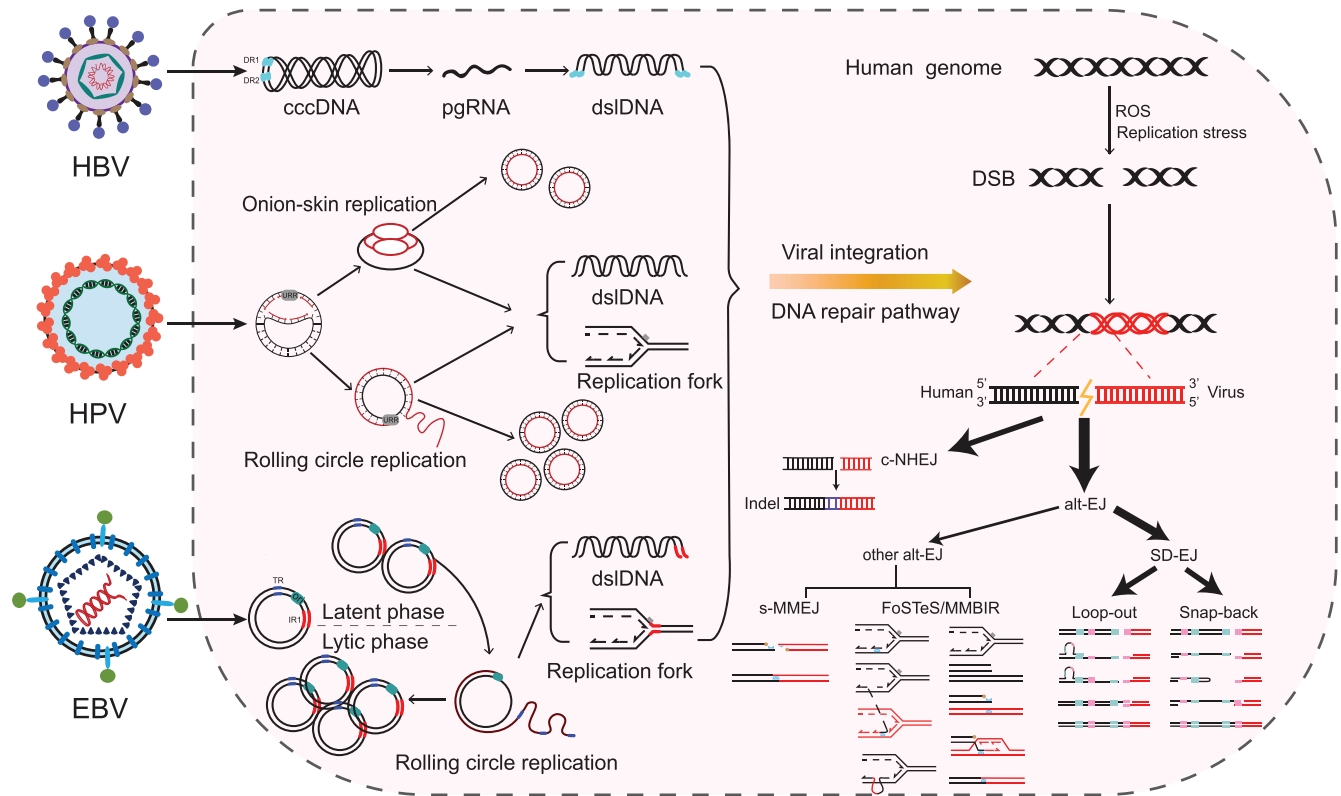


FIGURE 4 Model of DNA repair pathways involved in the integration of human papillomavirus (HPV), hepatitis B virus (HBV) and Epstein–Barr virus (EBV). Although viruses are replicated in different ways, their common feature is the production of large amounts of double-stranded linear DNA (dsIDNA) and replication forks. When the host cells encounter replication stresses or genetic insults (e.g. ROS), these replication products could serve as substrates of DNA repair pathway for fusion with double stranded DNA breaks (DSBs) generated from human genome, thereby promoting virus integration. Our data demonstrate that viral insertional events of HPV, HBV and EBV are mainly mediated via synthesis-dependent end-joining (SD-EJ) DNA repair mechanism, followed by c-NHEJ and other alt-EJ (s-MMEJ and FoSTeS) DNA repair mechanisms.

role in the integrations of three viruses into human genome.

Together, we report the largest genome-wide landscape of HPV, HBV and EBV insertional mutageneses. We uncovered HPV, HBV and EBV to share the same common SD-EJ integration mechanism. Based on our identified integration patterns and the biology features of three viruses, we proposed a new model of the integration process of HPV, HBV and EBV (Figure 4), providing insights into virus-induced cancer.

ACKNOWLEDGEMENTS

We thank the Tianhe Supercomputer Center for computational support and GeneRuler for probe design and partial experiment.

FUNDING INFORMATION

This work was supported by the National Science and Technology Major Project of the Ministry of science and technology of China (Grant no. 2018ZX10301402); The National Natural Science Foundation of China (Grant

no. 32171465 and 82102392); General Program of Natural Science Foundation of Guangdong Province of China (Grant no. 2021A1515012438); the National Postdoctoral Program for Innovative Talent (Grant no. BX20200398); the China Postdoctoral Science Foundation (Grant no. 2020M672995); Guangdong Basic and Applied Basic Research Foundation (Grant no. 2020A1515110170); the Major projects of Wuhan Municipal Health Commission (Grant no. WX19M02); the National Ten Thousand Plan-Young Top Talents of China.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

Rui Tian^{1,2}
 Yuyan Wang³
 Weiping Li⁴
 Zifeng Cui³
 Ting Pan⁵
 Zhuang Jin³
 Zhaoyue Huang³

Lifang Li³
 Bin Lang⁶
 Jian Wu⁷
 Hongxian Xie⁸
 Yiqin Lu⁹
 Xun Tian¹
 Zheng Hu^{2,3} 

Xun Tian, Department of Obstetrics and Gynecology, Academician expert workstation, the Central Hospital of Wuhan, Tongji Medical College, Huazhong University of Science and Technology, Shengli Street, Jang'an, Wuhan 430030, Hubei, China.
 Email: tianxun@zxhospital.com

Rui Tian, Yuyan Wang, Weiping Li, Zifeng Cui and Ting Pan contributed equally to this work.

¹Department of Obstetrics and Gynecology, Academician expert workstation, the Central Hospital of Wuhan, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei 430014, China

²Center of Cervical Cancer Precision Prevention and Treatment, Sun Yat-sen University Nanchang Research Institution, Nanchang, Jiangxi 330200, China

³Department of Obstetrics and Gynecology, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou, Guangdong 510080, China

⁴Department of Obstetrics and Gynecology, General Hospital of People's Liberation Army, Beijing 100039, China

⁵Institute of Human Virology, Key Laboratory of Tropical Disease Control of Ministry of Education, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, Guangdong 510080, China

⁶School of Health Sciences and Sports, Macao Polytechnic Institute, Macao 999078, China

⁷MyGenostics Inc, Beijing 101300, China

⁸Generulor Company Bio-X Lab, Zhuhai, Guangdong 519000, China

⁹Department of Gynecology, Dongzhimen Hospital, Beijing University of Chinese Medicine, Beijing 100700, China

Correspondence

Zheng Hu, Department of Obstetrics and Gynecology, The First Affiliated Hospital, Sun Yat-sen University, Zhongshan 2nd Road, Yuexiu, Guangzhou 510080, Guangdong, China; Center of Cervical Cancer Precision Prevention and Treatment, Sun Yat-sen University Nanchang Research Institution, Nanchang 330200, Jiangxi, China.

Email: huzheng1998@163.com

Yiqin Lu, Department of Gynecology, Dongzhimen Hospital, Beijing University of Chinese Medicine, Beijing 100700, China.

Email: k_lyq@sina.com

ORCID

Zheng Hu  <https://orcid.org/0000-0001-9306-9442>

REFERENCES

- Oh JK, Weiderpass E. Infection and cancer: global distribution and burden of diseases. *Ann Glob Health*. 2014;80:384-392.
- Akagi K, Li J, Broutian TR, et al. Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability. *Genome Res*. 2014;24:185-199.
- Sung WK, Zheng H, Li S, et al. Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nat Genet*. 2012;44:765-769.
- Xu M, Zhang WL, Zhu Q, et al. Genome-wide profiling of Epstein-Barr virus integration by targeted sequencing in Epstein-Barr virus associated malignancies. *Theranostics*. 2019;9:1115-1124.
- Hu Z, Zhu D, Wang W, et al. Genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism. *Nat Genet*. 2015;47:158-163.
- Zhao LH, Liu X, Yan HX, et al. Genomic and oncogenic preference of HBV integration in hepatocellular carcinoma. *Nat Commun*. 2016;7:12992.
- Leeman JE, Li Y, Bell A, et al. Human papillomavirus 16 promotes microhomology-mediated end-joining. *Proc Natl Acad Sci USA*. 2019;116:21573-21579.
- Ramsden DA, Carvajal-Garcia J, Gupta GP. Mechanism, cellular functions and cancer roles of polymerase-theta-mediated DNA end joining. *Nat Rev Mol Cell Biol*. 2022;23:125-140.
- Yu AM, McVey M. Synthesis-dependent microhomology-mediated end joining accounts for multiple types of repair junctions. *Nucleic Acids Res*. 2010;38:5706-5717.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.