



OPEN ACCESS

EDITED BY

Tao Liu,
University of New South Wales, Australia

REVIEWED BY

Lei Bu,
New York University, United States
Yong Li,
The First Affiliated Hospital of Nanchang
University, China
Susanna Grigson,
Flinders University, Australia
Deyou Tang,
South China University of Technology, China

*CORRESPONDENCE

Wenjun Deng

✉ wdeng@mgh.harvard.edu

Shanye Yin

✉ shanye.yin@einsteinmed.edu

RECEIVED 04 December 2024

ACCEPTED 31 January 2025

PUBLISHED 19 February 2025

CITATION

Hu E, An J, Gersten AJ, Wu N, Kawachi N,
Zhu J, Rosenblatt G, Augustine S, Smith RV,
Segall JE, Ostrer H, Amelio AL, Chung CH,
Prystowsky MB, Ow TJ, Deng W and Yin S
(2025) Virusplot: a web server for viral
integration analysis and visualization.
Front. Oncol. 15:1539782.
doi: 10.3389/fonc.2025.1539782

COPYRIGHT

© 2025 Hu, An, Gersten, Wu, Kawachi, Zhu,
Rosenblatt, Augustine, Smith, Segall, Ostrer,
Amelio, Chung, Prystowsky, Ow, Deng and Yin.
This is an open-access article distributed under
the terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Virusplot: a web server for viral integration analysis and visualization

Erqiang Hu^{1,2,3}, Jianhong An^{1,2,3}, Adam J Gersten^{1,4}, Nicole Wu⁵,
Nicole Kawachi¹, Jing Zhu¹, Gregory Rosenblatt¹,
Stelby Augustine⁴, Richard V. Smith⁴, Jeffrey E Segall¹,
Harry Ostrer¹, Antonio L Amelio^{6,7}, Christine H. Chung⁷,
Michael B. Prystowsky¹, Thomas J. Ow^{1,4},
Wenjun Deng^{8*} and Shanye Yin^{1,2,3*}

¹Department of Pathology, Albert Einstein College of Medicine, Bronx, NY, United States, ²Einstein Pathology Single-cell & Bioinformatics Laboratory, Bronx, NY, United States, ³Montefiore Einstein Comprehensive Cancer Center, Albert Einstein College of Medicine, Bronx, NY, United States,

⁴Department of Otorhinolaryngology-Head and Neck Surgery, Montefiore Medical Center/Albert Einstein College of Medicine, Bronx, NY, United States, ⁵The University of Texas at Austin, Austin, TX, United States, ⁶Department of Tumor Microenvironment and Metastasis, H. Lee Moffitt Cancer Center & Research Institute, Tampa, FL, United States, ⁷Department of Head and Neck-Endocrine Oncology, H. Lee Moffitt Cancer Center & Research Institute, Tampa, FL, United States, ⁸Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States

The integration of viral DNA into the human genome is a critical event in the pathogenesis of various cancers. This process leads to genomic instability, disrupts cellular regulatory mechanisms, and activates oncogenes or inactivates tumor suppressor genes. Despite significant advancements in genome sequencing technologies, there remains a notable lack of computational tools, particularly web-based applications, specifically designed for viral integration analysis and visualization. To address this gap, we present virusPlot, a web server with the following functional modules: (i) automatic retrieval of virus genome sequences and their annotation; (ii) visualization of virus integration locations and read counts through a graphical representation that links viral and host genome integration sites, facilitating the interpretation of integration patterns; (iii) analysis of virus integration hotspots using Fisher's exact test; and (iv) integration of various functions into an interactive web platform via shinyapp. VirusPlot efficiently processes and visualizes integration data from viruses and host genomes, providing researchers with an intuitive and user-friendly analytical tool that simplifies the complexity of virus integration analysis.

KEYWORDS

viral integration, genomic visualization, hotspot analysis, computational biology, bioinformatics tool, host-virus interaction

1 Introduction

Viral insertion into the host genome is a critical event in the viral life cycle of several viruses associated with tumorigenesis (1, 2). Specifically, viruses, such as human papillomavirus (HPV) (3, 4), hepatitis B virus (HBV) (5), Epstein-Barr virus (EBV) (6), and Human T-cell leukemia virus type 1 (HTLV-1) (7), are well-known for their oncogenic potential through this mechanism. The integration of viral DNA can result in genomic alterations that drive malignant transformation, contributing to cancer development and progression (1, 8–12). Understanding the molecular mechanisms of viral integration and its impact on cellular pathways is essential for developing targeted therapies and preventive strategies against virus-associated cancers.

Recent advances in genome sequencing methods have significantly enhanced the detection of viral integration events in tumor genomes. Whole Genome Sequencing (WGS) provides a comprehensive view of the entire tumor genome and allows for the identification of viral integration sites across the genome (3, 13–15). Capture sequencing involves the enrichment of viral sequences and their adjacent host sequences before sequencing, which is highly effective for detecting viral integration sites and mapping the integration landscape with high sensitivity and specificity (14, 16, 17). The development of long-read sequencing technologies offers the advantage of reading longer DNA fragments, which can span entire integration sites and provide more accurate mapping of integration events (18–21). This method is particularly useful for resolving complex integration events and structural variations.

Additionally, tools such as isling (22) and Vseq-Toolkit (23) have been developed for viral integration analysis, along with other tools benchmarked in these studies, providing a foundation for comparative assessments.

Viral integration events are complex and require sophisticated algorithms to accurately detect and interpret insertion sites within the host genome (9). These events can vary significantly in their genomic context, integration frequency, and impact on gene regulation, necessitating specialized tools for comprehensive analysis (1, 3, 9, 14). Currently, most available tools for viral integration analysis are either standalone software or scripts requiring a high level of bioinformatics expertise to operate (24–26). These tools often demand substantial computational resources and can be challenging for researchers without advanced programming skills. Furthermore, effective visualization tools are lacking, which are essential for interpreting the results of viral integration studies, allowing researchers to explore integration sites, their genomic contexts, and potential effects on gene expression and genome stability.

The absence of user-friendly, web-based platforms limits accessibility and hinders the widespread adoption of viral integration studies in the broader research community. To address this, we have developed virusPlot, an all-in-one analysis and visualization software. Key features of virusPlot include the automatic retrieval of virus genome sequences and annotation information, a visualization tool that represents viral integration events by connecting integration sites in the host genome with corresponding positions in the viral genome, and virus integration

hotspot analysis using Fisher's exact test to help identify possible integration hotspot regions. Importantly, virusPlot integrates all these functions into an interactive web platform via shinyapp, expanding accessibility to users without technical or computational skills. The integration of interactive, web-based visualization tools can significantly enhance the ability to communicate findings, generate hypotheses, and facilitate collaborative research.

2 Methods and functions

We have developed virusPlot, an R package that offers a comprehensive suite of tools for analyzing and visualizing virus integration into the host genome. The user interface and back-end of virusPlot are built using Shiny. Analysis results are displayed on the web page and can be downloaded in various formats, including PDF, PNG, EPS, TXT, and HTML (for more details, refer to the website help pages). The workflow and typical output schema are illustrated in Figures 1–4. Detailed functions and operations for each module are described below.

2.1 Automatic retrieval of virus genome

The virusPlot package includes two key functions for retrieving virus genome data: `get_virus_genom` and `get_virus_annotation`. The `get_virus_genom` function automates the process of obtaining viral genomic sequences from the National Center for Biotechnology Information (NCBI) database. By leveraging NCBI's extensive repository, this function provides users with accurate and up-to-date viral sequences for their research. Complementing this, the `get_virus_annotation` function retrieves gene annotation information for the obtained viral genomes, which is essential for understanding the functional roles of various viral genes and their potential impact on the host organism.

2.2 Quality control

To ensure the robustness of the input data analysis, we have incorporated a rigorous quality control mechanism. This mechanism allows users to apply filters based on two adjustable parameters: minimum read number and p-value threshold. For example, users can set a minimum read number (e.g., ≥ 10 reads) to exclude low-confidence breakpoints supported by insufficient reads, and a p-value threshold (e.g., $p < 0.05$) to ensure statistical significance. These parameters can be fine-tuned based on the specific characteristics of the dataset, such as sequencing depth or experimental design.

We have observed that adjusting these parameters has distinct effects on the dataset:

- Increasing the minimum read number threshold improves the confidence of the retained breakpoints but may reduce

virusPlot workflow

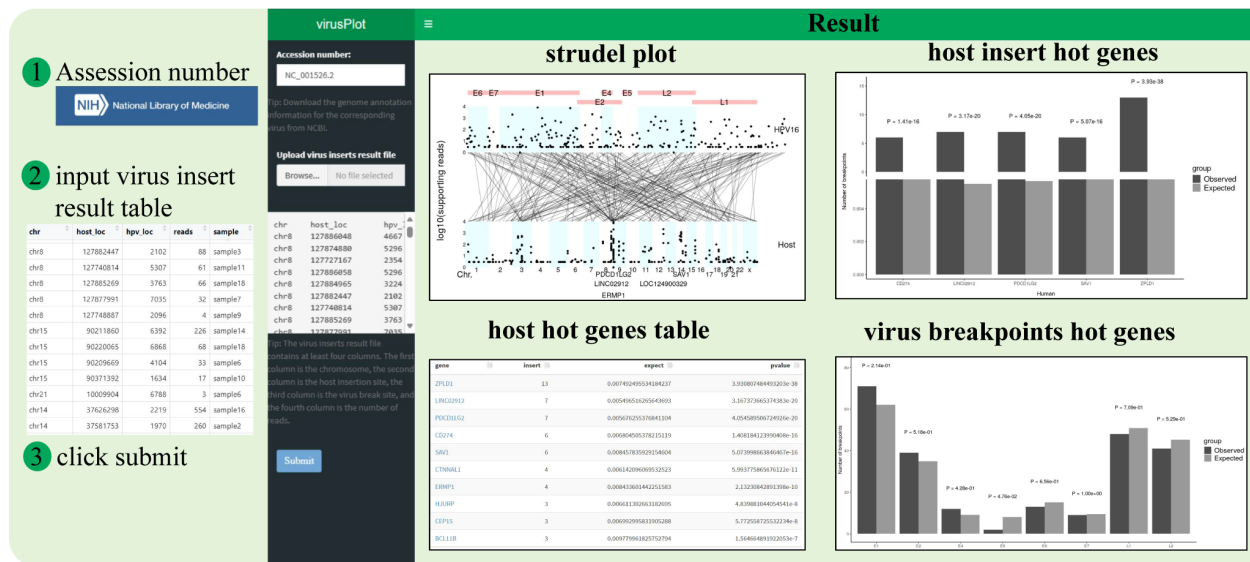


FIGURE 1

Screenshot of the virusPlot web tool. This package encompasses four main functionalities: (i) Automatic retrieval of virus genome information, (ii) Virus integration information visualization, (iii) Virus integration hot spots analysis, hot spots refer to genes with a statistically significant number of viral integration events, identified using Fisher's exact test or Chi-square test, and (iv) a user-friendly web application interface (shinyapp).

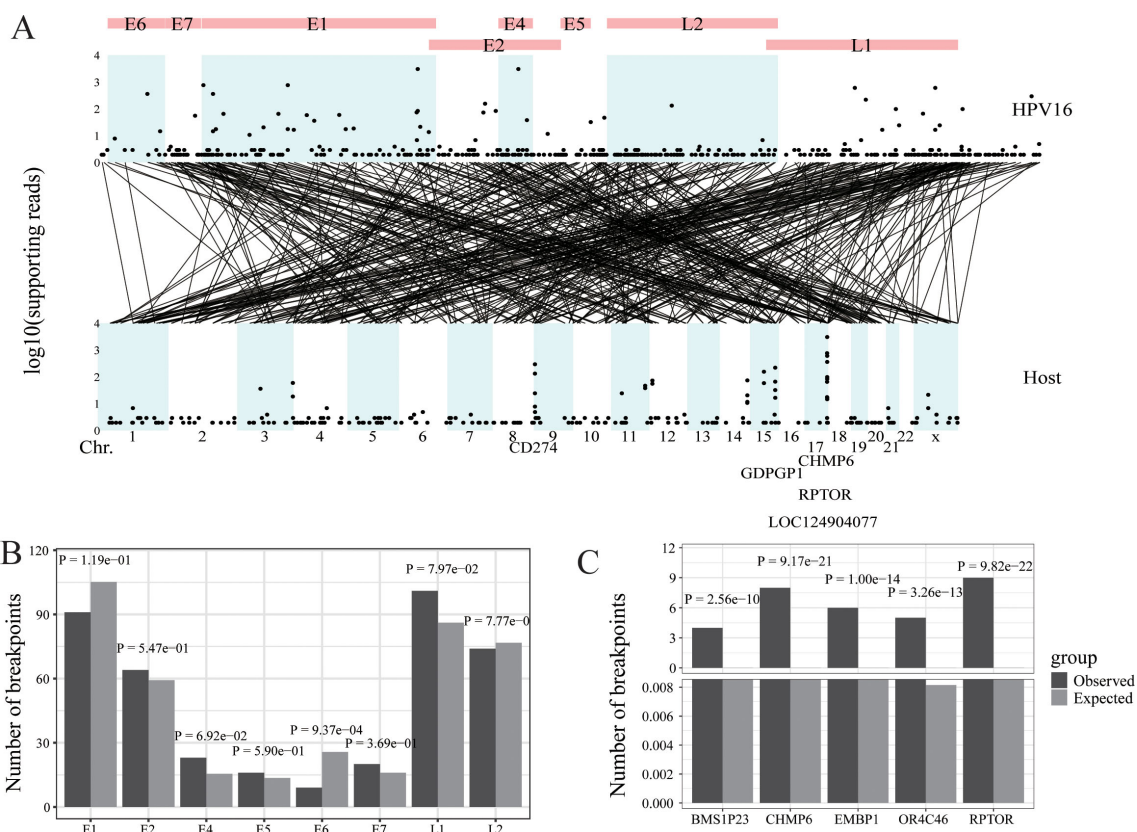
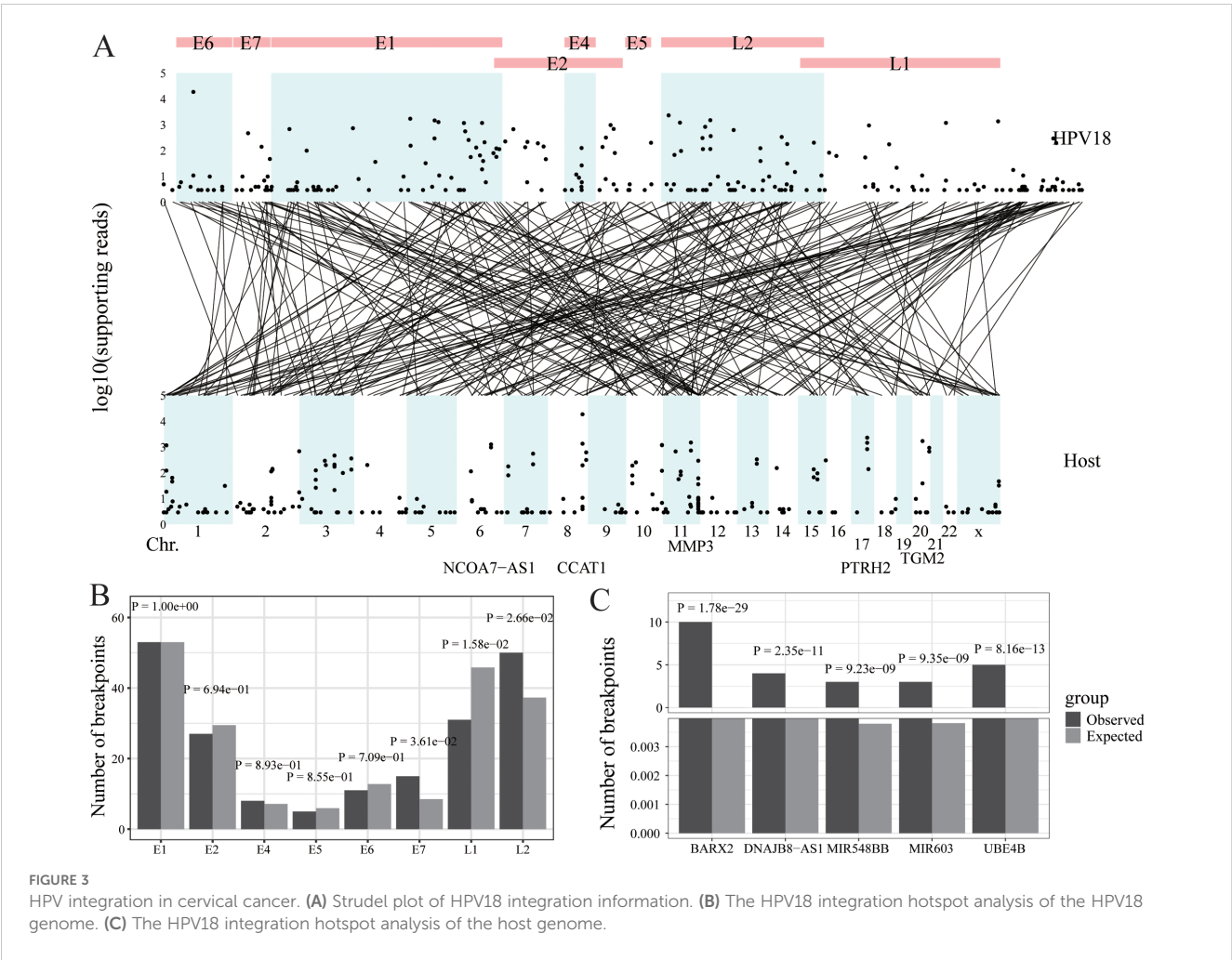


FIGURE 2

HPV integration in oropharyngeal cancer. (A) Strudel plot of HPV16 integration information. (B) The HPV16 integration hotspot analysis of the HPV16 genome. (C) The HPV16 integration hotspot analysis of the host genome.



the overall number of breakpoints available for downstream analysis.

- Tightening the p-value threshold enhances statistical rigor but might exclude potentially relevant breakpoints with weaker statistical signals.

2.3 Virus integration information visualization

A novel visualization approach, the strudel plot, has been developed within the virusPlot package to graphically represent the complex integration patterns of viruses within host genomes. The strudel plot is a multi-faceted structure that effectively displays integration sites, viral genome breakpoints, and the correspondence between host integration sites and viral breakpoints. The plot is organized into five distinct components:

- I. Viral Genome Retrieval: Presents the sequence and annotated information of the viral genome.

- II. Viral Integration Sites: Illustrates the positions of viral integration sites along with the number of associated reads.
- III. Integration Breakpoints: Maps the breakpoints in both the viral and host genomes.
- IV. Host Integration Sites: Shows the positions of integration sites in the host genome along with their respective read counts.
- V. Host Gene Hotspots: Identifies the host genes that are hotspots for viral integration.

VirusPlot analyzes viral integration breakpoints irrespective of the completeness of the viral genome at the site. The analysis supports both forward and reverse orientations of the virus and incorporates rearranged fragments by identifying and visualizing integration breakpoints on both the viral and host genomes.

2.4 Virus integration hot spots analysis

Viral integration hotspots are pivotal for comprehending virus-induced cancer mechanisms and for devising targeted therapies and preventive strategies. Identifying these hotspots holds promise for

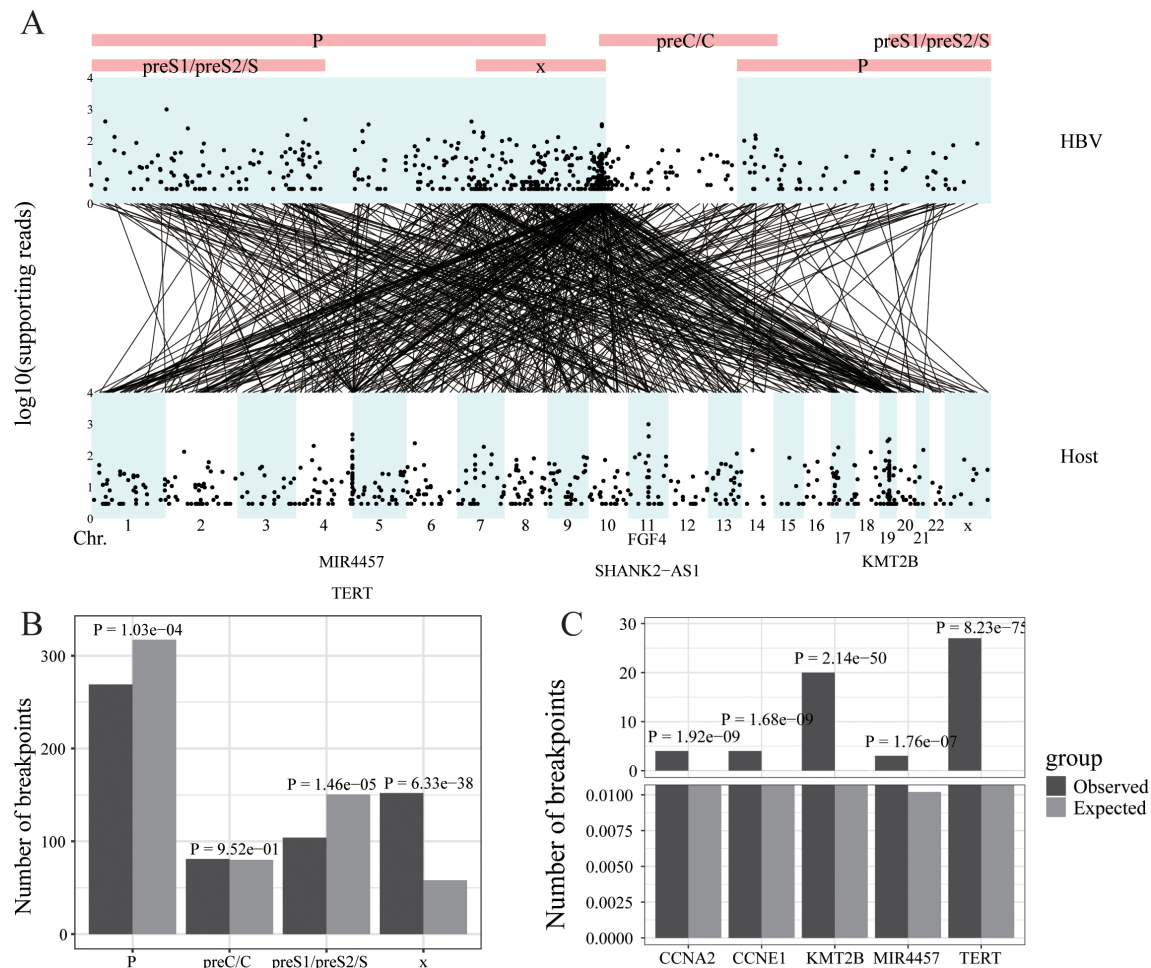


FIGURE 4
HBV integration in liver cancer. **(A)** Strudel plot of HBV integration information. **(B)** The HBV integration hotspot analysis of the HBV genome. **(C)** The HBV integration hotspot analysis of the host genome.

early detection, prognostic insights, and personalized treatment approaches. In the virusPlot package, a robust analysis tool is included for detecting viral integration hotspots within the host genome. A gene is identified as a potential hotspot if the observed number of integration sites significantly exceeds the expected count by chance. This comparison, facilitated by both Fisher's exact and chi-square tests, offers valuable insights into the preferential integration patterns of viruses, critical for deciphering viral persistence mechanisms and pathogenesis. The expected number of integration sites for each genomic region was calculated as follows:

$$\text{Expected count} = \text{Total observed integrations} \times \frac{\text{Region size (bp)}}{\text{Total genome size (bp)}}$$

Statistical significance was determined using Fisher's exact test and Chi-square test by comparing the observed and expected counts. The p-value threshold for significance was set at 0.05.

The web-based version of VirusPlot processes viral integration data using the hg38 human genome assembly by default. For users of the R package version, the genome assembly can be customized by specifying the TxDb parameter in functions such as `get_hot_gene` and `strudel_plot`. This flexibility allows users to work with data aligned to different genome versions, ensuring compatibility with their specific datasets and annotations.

2.5 A user-friendly web application interface

To make virusPlot widely accessible and easy to use, we've integrated a shinyapp, providing a user-friendly web interface for accessing all features without the need for coding. This web app simplifies operations, requiring no local setup or specific hardware, as all computations are cloud-based, ensuring compatibility across different systems. Through practical case studies using our own data and published HPV and HBV integration data in various cancers

(i.e., oropharyngeal cancer, cervical cancer and hepatocellular cancer) (Figures 2–4), virusPlot has proven effective in displaying integration sites and identifying hotspots, underscoring virusPlot's utility in simplifying the complex analysis of viral integration and providing valuable insights into the genomic alternations and regulatory disruptions associated with virus-related cancers.

2.6 A data security and privacy

VirusPlot employs multiple measures to ensure the security of user data. All data transfers between the user and the server are encrypted using HTTPS. Uploaded data is processed in-memory without being stored permanently on the server, and all data is automatically deleted upon completion of the analysis. The server is hosted on a secure platform with firewalls and regular updates to protect against vulnerabilities.

3 Case study

3.1 HPV integration in oropharyngeal cancer (Data in this study)

Whole Exome & HPV capture sequencing (WEHS) was performed on 20 head and neck cancer samples. We used the Survirus software (24) to detect HPV16 integration information. This analysis revealed 471 integration sites and 13,007 reads encompassing these sites. The viral integration sites were evenly distributed throughout the HPV16 genome but concentrated in specific chromosomal regions of the host genome, such as chr17, chr9, and chr15. The five genes with the highest number of viral read insertions were CHMP6, RPTOR, LOC124904077, CD274, and GDPGP1 (Figure 2A). Regarding the HPV16 genome, fewer integration sites were observed in the E6 gene compared to random occurrences, while no significant differences were noted for other genes, and no hotspot integration genes were identified (Figure 2B). In the host (human) genome, several hotspots for HPV16 insertion were identified, including BMS1P23, CHMP6, EBMP1, OR4C46, and RPTOR, compared to random occurrences (Figure 2C).

3.2 HPV integration in cervical cancer

Ma's Lab conducted DNA sequencing on 39 cervical cancer samples collected from Tongji Hospital in Wuhan and Jingmen No. 2 People's Hospital in Hubei Province, China, between 2007 and 2014 (14). Exfoliated cervical epithelial cells were collected using cervical brushes. The high-throughput Viral Integration Detection (HIVID) (27) was then applied to these samples to detect HPV18 integration information. This analysis identified 241 integration sites and 48,603 reads including these sites. The viral integration sites were evenly distributed across the HPV18 genome and concentrated in specific chromosomal regions of the host genome, such as chr8, chr11, and chr17. The five genes with the

highest number of viral read insertions were CCAT1, PTRH2, TGM2, MMP3, and NCOA7-AS1 (Figure 3A). In the HPV18 genome, the actual number of integration sites in the E7 and L2 genes was significantly higher than random, indicating hotspots for integration, while the L1 gene had significantly fewer integration sites than random (Figure 3B). In the host (human) genome, multiple hotspots for HPV18 insertion were identified, including BARX2, DNAJB8-AS1, MIR548BB, MIR603, and UBE4B, compared to random occurrences (Figure 3C).

3.3 HBV integration in liver cancer

Wang's Lab performed DNA sequencing on 138 hepatocellular carcinomas (HCCs) collected at the Eastern Hepatobiliary Surgery Hospital in Shanghai from 2009 to 2010 using high-throughput viral integration detection (HIVID) method (28). This analysis identified 546 integration sites and 13,242 reads encompassing these sites. The reads containing viral integration sites were primarily concentrated in the X gene of the HBV genome and on chromosomes chr5, chr11, and chr19 of the host genome. The five genes with the highest number of viral read insertions were MIR4457, TERT, FGF4, SHANK2-AS1, and KM12B (Figure 4A). Regarding the HBV genome, the actual number of integration sites in the X gene was significantly higher than random, indicating hotspots for integration, whereas the PreS1/PreS2/S and P genes had significantly fewer integration sites than random (Figure 4B). In the host (human) genome, multiple hotspots for HBV insertion were identified, including TEAR, KMT2B, CCNE1, CCNA2, and MIR4457, compared to random occurrences (Figure 4C).

4 Discussion

The virusPlot platform offers a comprehensive suite of tools and workflows designed to simplify the analysis and visualization of viral integration events within host genomes. Serving as an visualization solution, virusPlot streamlines the process of identifying integration hotspots and understanding the impact of viral integration on cancer development. With its intuitive interface, virusPlot enables researchers, including experimental biologists without computational programming skills, to explore complex viral integration patterns and gain valuable insights into virus-associated cancers. By integrating multiple analytical approaches and statistical tests, virusPlot enhances the analysis of viral integration data, complementing traditional methods and enabling more comprehensive investigations. As sequencing methods continue to evolve and become more cost-effective, an increasing number of viral-associated cancers and viral integration events are expected to be uncovered (1, 20). This expanding dataset of viral-related cancer genomes and integration sites will provide valuable insights into the role of viruses in oncogenesis and facilitate the development of targeted therapies and preventive strategies against virus-associated cancers. We are committed to maintaining the virusPlot platform and continuously updating it with new data

and methods, ensuring its relevance and utility for the research community over the coming years. Through its user-friendly interface and powerful analytical capabilities, virusPlot aims to accelerate discoveries in viral oncology and facilitate the identification of novel cancer pathways and therapeutic targets.

To evaluate the usability of VirusPlot, we compared it with other visualization tools commonly used for integration data, such as Circos (29). Users reported that VirusPlot's interactive interface significantly reduced the complexity of generating visualizations, compared to the manual configurations required by Circos. Furthermore, VirusPlot provides integration-specific features, such as the strudel plot and hotspot analysis, which are not available in these general-purpose tools. Additionally, the web-based accessibility of VirusPlot eliminates the need for local installations, making it a more user-friendly and efficient tool for analyzing viral integration data.

Future updates to virusPlot aim to expand its capabilities by including support for additional genomic data types, such as RNA-seq and long-read sequencing data. Planned features include new visualization tools, such as circular diagrams and VCF diagrams, which are currently under development. We also aim to integrate more advanced statistical modules to accommodate complex integration scenarios, such as those involving rearranged viral genomes.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <https://github.com/huerqiang/virusPlot>.

Author contributions

SY: Methodology, Project administration, Software, Writing – original draft, Writing – review & editing. EH: Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. JA: Writing – review & editing. AG: Writing – review & editing. NW: Writing – review & editing. NK: Writing – review & editing. JZ: Writing – review & editing. GR: Writing – review & editing. SA: Writing – review & editing. RS: Writing – review & editing. JS: Writing – review & editing. HO: Writing – review & editing.

References

1. Ye R, Wang A, Bu B, Luo P, Deng W, Zhang X, et al. Viral oncogenes, viruses, and cancer: A third-generation sequencing perspective on viral integration into the human genome. *Front Oncol.* (2023) 13:1333812. doi: 10.3389/fonc.2023.1333812
2. Zapatka M, Borozaan I, Brewer DS, Iskar M, Grundhoff A, Alawi M, et al. The landscape of viral associations in human cancers. *Nat Genet.* (2020) 52:320–30. doi: 10.1038/s41588-019-0558-9
3. Akagi K, Li J, Broutian TR, Padilla-Nash H, Xiao W, Jiang B, et al. Genome-wide analysis of hpv integration in human cancers reveals recurrent, focal genomic instability. *Genome Res.* (2014) 24:185–99. doi: 10.1101/gr.164806.113
4. Lechner M, Liu J, Masterson L, Fenton TR. Hpv-associated oropharyngeal cancer: epidemiology, molecular biology and clinical management. *Nat Rev Clin Oncol.* (2022) 19:306–27. doi: 10.1038/s41571-022-00603-7
5. Arzumanyan A, Reis HM, Feitelson MA. Pathogenic mechanisms in hbv- and hcv-associated hepatocellular carcinoma. *Nat Rev Cancer.* (2013) 13:123–35. doi: 10.1038/nrc3449
6. Young LS, Yap LF, Murray PG. Epstein-barr virus: more than 50 years old and still providing surprises. *Nat Rev Cancer.* (2016) 16:789–802. doi: 10.1038/nrc.2016.92
7. Matsuoka M, Jeang KT. Human T-cell leukaemia virus type 1 (Htlv-1) infectivity and cellular transformation. *Nat Rev Cancer.* (2007) 7:270–80. doi: 10.1038/nrc2111
8. Williams VM, Filippova M, Soto U, Duerksen-Hughes PJ. Hpv-DNA integration and carcinogenesis: putative roles for inflammation and oxidative stress. *Future Virol.* (2011) 6:45–57. doi: 10.2217/fvl.10.73
9. Symer DE, Akagi K, Geiger HM, Song Y, Li G, Emde AK, et al. Diverse tumorigenic consequences of human papillomavirus integration in primary oropharyngeal cancers. *Genome Res.* (2022) 32:55–70. doi: 10.1101/gr.275911.121

AA: Writing – review & editing. CC: Writing – review & editing. MP: Writing – review & editing. TO: Writing – review & editing. WD: Project administration, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was in part supported by NIH R01CA291607 (SY), R21 CA267527-01 (SY), the Feldstein Medical Foundation grant (SY), R01INS134819 (WD), Fundings from the Moffitt Cancer Center and State of Florida (AA).

Conflict of interest

WD and SY hold equity in Yihui Bio, Inc. AA is a Global Advisory Board member and paid consultant for LG Chem Life Sciences Innovation Center.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

10. Porter VL, Marra MA. The drivers, mechanisms, and consequences of genome instability in hpv-driven cancers. *Cancers (Basel)*. (2022) 14. doi: 10.3390/cancers14194623
11. Kang JJ, Yu Y, Chen L, Zakeri K, Gelblum DY, McBride SM, et al. Consensuses, controversies, and future directions in treatment deintensification for human papillomavirus-associated oropharyngeal cancer. *CA Cancer J Clin*. (2022). doi: 10.3322/caac.21758
12. Sabatini ME, Chiocca S. Human papillomavirus as a driver of head and neck cancers. *Br J Cancer*. (2020) 122:306–14. doi: 10.1038/s41416-019-0602-7
13. Cancer Genome Atlas N. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature*. (2015) 517:576–82. doi: 10.1038/nature14129
14. Hu Z, Zhu D, Wang W, Li W, Jia W, Zeng X, 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) 47:158–63. doi: 10.1038/ng.3178
15. Muhr LSA, Guerendian D, Cuschieri K, Sundstrom K. Human papillomavirus detection by whole-genome next-generation sequencing: importance of validation and quality assurance procedures. *Viruses*. (2021) 13. doi: 10.3390/v13071323
16. Pinatti LM, Gu W, Wang Y, Elhossiny A, Bhargale AD, Brummel CV, et al. Searchpv: A novel approach to identify and assemble human papillomavirus-host genomic integration events in cancer. *Cancer*. (2021) 127:3531–40. doi: 10.1002/cncr.33691
17. Liu Y, Lu Z, Xu R, Ke Y. Comprehensive mapping of the human papillomavirus (Hpv) DNA integration sites in cervical carcinomas by hpv capture technology. *Oncotarget*. (2016) 7:5852–64. doi: 10.18632/oncotarget.6809
18. Yang S, Zhao Q, Tang L, Chen Z, Wu Z, Li K, et al. Whole genome assembly of human papillomavirus by nanopore long-read sequencing. *Front Genet*. (2021) 12:798608. doi: 10.3389/fgene.2021.798608
19. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol*. (2020) 21:30. doi: 10.1186/s13059-020-1935-5
20. Zhou L, Qiu Q, Zhou Q, Li J, Yu M, Li K, et al. Long-read sequencing unveils high-resolution hpv integration and its oncogenic progression in cervical cancer. *Nat Commun*. (2022) 13:2563. doi: 10.1038/s41467-022-30190-1
21. Yang W, Liu Y, Dong R, Liu J, Lang J, Yang J, et al. Accurate detection of hpv integration sites in cervical cancer samples using the nanopore minion sequencer without error correction. *Front Genet*. (2020) 11:660. doi: 10.3389/fgene.2020.00660
22. Scott S, Hallwirth CV, Hartkopf F, Grigson S, Jain Y, Alexander IE, et al. Isling: A tool for detecting integration of wild-type viruses and clinical vectors. *J Mol Biol*. (2022) 434:167408. doi: 10.1016/j.jmb.2021.167408
23. Afzal S, Fronza R, Schmidt M. Vseq-toolkit: comprehensive computational analysis of viral vectors in gene therapy. *Mol Ther Methods Clin Dev*. (2020) 17:752–7. doi: 10.1016/j.omtm.2020.03.024
24. Rajaby R, Zhou Y, Meng Y, Zeng X, Li G, Wu P, et al. Survirus: A repeat-aware virus integration caller. *Nucleic Acids Res*. (2021) 49:e33. doi: 10.1093/nar/gkaa1237
25. Baheti S, Tang X, O'Brien DR, Chia N, Roberts LR, Nelson H, et al. Hgt-id: an efficient and sensitive workflow to detect human-viral insertion sites using next-generation sequencing data. *BMC Bioinf*. (2018) 19:271. doi: 10.1186/s12859-018-2260-9
26. Zeng X, Zhao L, Shen C, Zhou Y, Li G, Sung WK. Hivid2: an accurate tool to detect virus integrations in the host genome. *Bioinformatics*. (2021) 37:1821–7. doi: 10.1093/bioinformatics/btab031
27. Li W, Zeng X, Lee NP, Liu X, Chen S, Guo B, et al. Hivid: an efficient method to detect hbv integration using low coverage sequencing. *Genomics*. (2013) 102:338–44. doi: 10.1016/j.ygeno.2013.07.002
28. Zhao LH, Liu X, Yan HX, Li WY, Zeng X, Yang Y, et al. Genomic and oncogenic preference of hbv integration in hepatocellular carcinoma. *Nat Commun*. (2016) 7:12992. doi: 10.1038/ncomms12992
29. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. *Genome Res*. (2009) 19:1639–45. doi: 10.1101/gr.092759.109