



OPEN ACCESS

EDITED BY

Shihori Tanabe,
National Institute of Health Sciences (NIHS),
Japan

REVIEWED BY

Prajish Iyer,
City of Hope National Medical Center,
United States

*CORRESPONDENCE

Wenjun Deng
✉ wdeng@mgh.harvard.edu
Xinyi Zhang
✉ zhangxinyi80@163.com
Shanye Yin
✉ shanye.yin@einsteinmed.edu

†These authors have contributed equally to
this work

RECEIVED 07 November 2023

ACCEPTED 06 December 2023

PUBLISHED 21 December 2023

CITATION

Ye R, Wang A, Bu B, Luo P, Deng W, Zhang X
and Yin S (2023) Viral oncogenes, viruses, and
cancer: a third-generation sequencing
perspective on viral integration into the
human genome.
Front. Oncol. 13:1333812.
doi: 10.3389/fonc.2023.1333812

COPYRIGHT

© 2023 Ye, Wang, Bu, Luo, Deng, Zhang and
Yin. This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](#). 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.

Viral oncogenes, viruses, and cancer: a third-generation sequencing perspective on viral integration into the human genome

Ruichen Ye^{1,2,3†}, Angelina Wang^{4†}, Brady Bu^{5†},
Pengxiang Luo⁶, Wenjun Deng^{7*}, Xinyi Zhang^{8*}
and Shanye Yin^{1,2*}

¹Department of Pathology, Albert Einstein College of Medicine, Bronx, NY, United States, ²Einstein Pathology Single-cell & Bioinformatics Laboratory, Bronx, NY, United States, ³Stony Brook University, Stony Brook, NY, United States, ⁴Tufts Friedman School of Nutrition, Boston, MA, United States, ⁵Horace Mann School, Bronx, NY, United States, ⁶Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, ⁷Clinical Proteomics Research Center, Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States, ⁸Department of Respiratory Diseases, The Second Affiliated Hospital of Nanchang University, Nanchang, China

The link between viruses and cancer has intrigued scientists for decades. Certain viruses have been shown to be vital in the development of various cancers by integrating viral DNA into the host genome and activating viral oncogenes. These viruses include the Human Papillomavirus (HPV), Hepatitis B and C Viruses (HBV and HCV), Epstein-Barr Virus (EBV), and Human T-Cell Leukemia Virus (HTLV-1), which are all linked to the development of a myriad of human cancers. Third-generation sequencing technologies have revolutionized our ability to study viral integration events at unprecedented resolution in recent years. They offer long sequencing capabilities along with the ability to map viral integration sites, assess host gene expression, and track clonal evolution in cancer cells. Recently, researchers have been exploring the application of Oxford Nanopore Technologies (ONT) nanopore sequencing and Pacific BioSciences (PacBio) single-molecule real-time (SMRT) sequencing in cancer research. As viral integration is crucial to the development of cancer via viruses, third-generation sequencing would provide a novel approach to studying the relationship interlinking viral oncogenes, viruses, and cancer. This review article explores the molecular mechanisms underlying viral oncogenesis, the role of viruses in cancer development, and the impact of third-generation sequencing on our understanding of viral integration into the human genome.

KEYWORDS

viral integration, cancer sequencing, third generation sequencing, nanopore, PacBio

1 Introduction

Cancer is a complex and multifaceted disease with genetic alterations playing a central role in its initiation and progression. It is comprised of genetically diverse heterogeneous groups of diseases that are constantly evolving, stretching across populations and individuals (1–6). Throughout the years, several studies, such as the PCAWG Project (7), have investigated these variations or mutations in cancer cells to see the relationship between the genetic changes and their role in cancer (1, 8, 9). Although somatic mutations have traditionally been the focus of cancer genetics, viral infections also contribute significantly to the global cancer burden (10–12). 15.4% of all cancers are attributable to infections and 9.9% are linked to viruses, according to the World Health Organization (WHO). After *Helicobacter pylori*, the four most prominent infection-related causes of cancer are HPV, HBV, hepatitis C, and EBV. These viruses are widely known to the general population and, with their rising infection rates, are to be a significant matter to raise up. Viruses are part of a multistep pathway of oncogenesis and contribute to the development of cancer (8, 12). Characteristics that can be attributed to the development of cancer through human viruses include the presence and persistence of viral DNA in tumors, growth-promoting activity of viral genes in model systems, malignant phenotype dependence on viral oncogene expression, and epidemiological evidence that viral infection can be a major risk of cancer development (11). Recently, there have been efforts made by the worldwide scientific community to specifically analyze comprehensive genomes and whole transcriptomes from tissue samples from cancer patients. The PCAWG Project, for example, collected whole genome sequencing data from 2,658 cancers in 38 tumor types, most likely from third-generation sequencing, a new and uprising technique being used to sequence genomes (8, 11, 13–15). This review aims to provide an overview of viral oncogenes, their role in cancer, and the recent advancements in third-generation sequencing that have enhanced our ability to study viral integration into the human genome.

2 Viral oncogenes and their role in cancer

Viral oncogenes are genes carried by certain viruses that can promote cellular transformation and lead to the development of cancer. These genes are often homologous to cellular proto-oncogenes, but their regulation and expression are altered in infected cells. In these infected cells, normal cell growth control mechanisms are disrupted, and they display uncontrolled proliferation. Tissue-specific functions cease and ultimately lead to the development of cancer. 15–20% of all human cancers have been caused via infection by oncogenic viruses (12). Key viral oncogenes include E6 and E7 from human papillomavirus (HPV), E1A and E1B from adenovirus, and Tax from human T-cell leukemia virus (HTLV-1). These oncogenes disrupt normal cellular processes, such as cell cycle regulation, apoptosis, and DNA repair, leading to uncontrolled cell growth and eventual tumor formation.

2.1 Viruses implicated in human cancer

Several viruses have been linked to the development of various cancers. Notable examples include:

2.1.1 Human papillomavirus

HPV is a DNA virus associated with cervical, anal, and oropharyngeal cancers. More than half of all malignancies related to infection around the world are caused by HPV, with an incidence of approximately 5% among all cancers worldwide (16–18). HPV16 and HPV18, with less frequency, are involved in a subset of head and neck cancers (HNCs) which are called oropharyngeal squamous cell carcinomas (OPSCCs) (19, 20). The fraction of OPSCCs arising from HPV infection has been around 20%. Integration of HPV DNA into the host genome disrupts tumor suppressor genes and promotes oncogene expression. HPV carries two critical oncogenes: E6 and E7. E6 promotes the degradation of the tumor suppressor protein p53, and E7 disrupts the function of the retinoblastoma protein (pRb), leading to uncontrolled cell division (17, 21, 22). Specifically, hrHPV E6 binds to p53 and promotes its degradation, triggering uncontrolled cell proliferation through cell cycle checkpoint evasion. HPV+ cancer cells have been shown to retain unaltered p53 and pRb genes, which is unlike most human tumors containing mutations in these tumor suppressor genes. Understanding the molecular mechanisms of these oncogenes is crucial for developing targeted therapies against HPV-associated cancers.

2.1.2 Hepatitis B and C viruses

Viral hepatitis stemming from the hepatitis B and C viruses is primarily associated with severe health complications, including liver cirrhosis, hepatocellular carcinoma, hepatic fibrosis, and steatosis. Both are considered to be a major healthcare problem and are blood-borne viruses that primarily infect the liver. Hepatitis B is a life-threatening liver disease caused by the highly contagious viral pathogen called hepatitis B virus (HBV) (23–25), which is an enveloped virus belonging to Hepadnavirida. Hepatitis C virus (HCV) is a positive-stranded RNA virus that causes severe liver disorders and is related to *Togaviridae* or *Flaviviridae*. It is a small spherical enveloped virion with an icosahedral capsid (26–28). Chronic hepatitis B and C viral infections can result in hepatocellular carcinoma (HCC) through mechanisms involving viral integration and chronic inflammation. Integration of viral DNA into the host genome can cause genomic instability and chronic inflammation, promoting the progression of liver cancer. Studying these viral integration sites and their impact on host genes is essential for identifying potential therapeutic targets. It is crucial to implement preventative strategies and therapies to eradicate chronic infection of HCV and suppress viral replication for HBV (26, 27).

2.1.3 Epstein-Barr virus

Epstein-Barr Virus, a member of the herpesvirus family, is associated with several cancers including nasopharyngeal carcinoma, Burkitt's lymphoma, and gastric carcinoma. EBV infects B cells and epithelial cells, maintaining itself in a latent

state in B cells for the lifetime of the host (29–32). The association of EBV with various human cancers has been well-documented. The virus encodes several genes that promote cell growth and inhibit apoptosis, such as LMP1 and EBNA2. LMP1 acts as a constitutively active receptor, activating several signaling pathways leading to cell proliferation and survival. EBNA2, on the other hand, controls the expression of genes promoting B cell growth (33–35). While integration of the EBV genome into the host chromosome isn't common, the expression of its viral oncogenes and their interaction with cellular pathways plays a pivotal role in the development of EBV-associated cancers (36–40).

2.1.4 Human T-cell leukemia virus

HTLV-1 is a retrovirus known to cause adult T-cell leukemia/lymphoma (ATL). This virus integrates its DNA into the host genome, leading to the clonal expansion of infected cells (41). HTLV-1 is a key viral oncogene encoding for Tax protein, which can activate a variety of cellular pathways leading to cell proliferation. While the precise factors contributing to the varied outcomes of HTLV-1 infection remain incompletely elucidated, mounting evidence indicates that a intricate interplay between the virus and the host, along with the host's immune response to HTLV-1, likely govern the emergence of HTLV-1-associated diseases (42–44). By disrupting cell cycle regulation and DNA repair mechanisms, Tax promotes genetic instability and the survival of damaged cells, paving the way for leukemogenesis. In some cases, HTLV-1-associated adult T cell leukemia/lymphoma can present with unique clinical manifestations, such as granulomatous pneumocystis jiroveci pneumonia and hypercalcemia.

2.2 Viral integration into the human genome

The contribution of viruses to carcinogenesis has been a topic of profound significance in the realm of cancer biology. Central to this paradigm is the mechanism through which viruses integrate their DNA into the human genome (45–47). Such integration events play a pivotal role in the life cycle of oncogenic viruses, often precipitating a cascade of cellular disruptions that culminate in tumorigenesis. Timely identification of these human-infecting viruses is crucial for understanding their role in carcinogenesis and for implementing appropriate interventions (48). Recent advancements have enabled rapid identification of such viruses, providing a more comprehensive understanding of their implications in human health.

2.2.1 Mechanisms of viral integration

Viral integration into the host genome is not a random event. Research has shown that integration sites often cluster with fragile sites in the genome, which may have implications for the activation of proto-oncogenes. It is orchestrated through a repertoire of intricate mechanisms, including non-homologous end joining (NHEJ) and microhomology-mediated end joining (MMEJ). The precise loci of these integration events can significantly differ across

instances, determining the severity of genetic disruption and subsequent activation of oncogenes. Delving into the specific mechanisms employed by different viruses helps elucidate their unique contributions to cancer.

2.2.2 Consequences of viral integration

Once viruses find their way into the human genome, their repercussions on cellular machinery are multifaceted: The integration of the viral genome into the host can be seen as a double-edged sword, with both benefits and detriments for both parties. Activation of Cellular Oncogenes: Viral DNA, when integrated, can strategically position viral enhancers or promoters in proximity to cellular proto-oncogenes. This can lead to the aberrant activation of genes like MYC in the case of HPV-associated cervical cancers, accelerating the journey to tumorigenesis. Disruption of Regulatory Elements and Pathways: Beyond the activation of proto-oncogenes, viral integration can also disrupt host genes, alter regulatory elements, or introduce novel oncogenes. Such disruptions can skew gene expression, compromise cellular function, and enhance cellular growth or survival tendencies. Inactivation of Tumor Suppressor Genes: Notably, viruses also have the ability to incapacitate tumor suppressor genes, such as TP53 and PTEN. The loss of these sentinels of cellular regulation paves the way for uncontrolled growth and genomic instability.

2.2.3 Viral-induced genomic instability and clonal evolution

The integration of viral DNA into the host genome can lead to genomic instability, fostering an environment ripe for mutations. Certain viruses can directly induce DNA damage through their replication processes, leading to genomic instability. Viruses may interfere with host DNA repair mechanisms, compromising genome stability. This interference can result in the accumulation of DNA mutations, which are critical for clonal evolution. As these mutations accumulate, as the immune system exerts selective pressure on viral variants, certain cell clones with conferred growth advantages may begin to dominate the tumor landscape. Over time, the tumor evolves, becoming a heterogeneous entity teeming with sub-clones, each potentially harboring distinct viral integration patterns. In the context of antiviral therapy, viral-induced genomic instability can contribute to the emergence of drug-resistant viral strains. Understanding the dynamics of clonal evolution is essential for designing effective therapeutic strategies. Developing strategies to interfere with viral integration or eliminate integrated viral genomes could be a promising approach to reduce viral-induced genomic instability and its associated pathologies.

3 Third-generation sequencing technologies and viral integration

The genomic landscape of cancer has been revolutionized by advancements in sequencing technologies (13, 49). A significant facet of this change has been brought about by third-generation sequencing technologies, which illuminate the dynamics of viral

integration into the human genome (50–52). Long-read sequencing allowed the sequencing of chimeric reads, incorporating both viral DNA and the host genome fragments on either side (53). This aids in detecting multiple clonal integration events and uncovering key features of viral integration in the cancer genome. The promise of these technologies lies in their capability to offer extended read lengths, real-time sequencing, and unprecedented resolution, all of which are pivotal in studying the intricate relationships between viral oncogenes, viruses, and cancer (54–57).

3.1 Pacific biosciences sequencing

PacBio sequencing has emerged as a pioneering technology in genomics, particularly for its unique capabilities in generating long reads. This groundbreaking approach enables researchers to delve deeper into the complexities of genomes, transcriptomes, and epigenomes, unlocking a wealth of biological information that was previously challenging to access. One of the standout features of PacBio sequencing is the ability to produce exceptionally long reads, often spanning thousands to tens of thousands of base pairs. This stands in stark contrast to short-read sequencing technologies, allowing for the detection of structural variants, complex genomic rearrangements, and complete characterization of repetitive regions that were previously elusive. Such extended read lengths are especially advantageous in the context of viral-related tumors, as they facilitate the precise mapping of viral integration sites within the human genome, offering a deeper understanding of the molecular events underpinning tumorigenesis. Moreover, PacBio sequencing excels in deciphering the epigenetic landscape through the analysis of DNA modifications, offering insights into the role of DNA methylation and histone modifications in viral-related tumors (58, 59). These capabilities make PacBio sequencing an invaluable tool for unraveling the intricacies of viral oncogenesis and hold significant promise for the development of advanced diagnostic and therapeutic strategies in the field of cancer research.

3.2 Nanopore sequencing

Another formidable player in the third-generation sequencing realm is nanopore sequencing, which involves the passage of DNA or RNA molecules through nanometer-sized pores, enabling real-time monitoring of nucleotide sequences (60–63). Its advantages include long-read capabilities, portability, and the potential for direct RNA sequencing. Beyond the capability to provide long reads, nanopore sequencing is distinctive due to its real-time, single-molecule sequencing of DNA strands. For studies central to viral integration, this technology holds promise by allowing direct observation of the integration process. Such insights grant researchers the capability to identify structural variants that have arisen due to viral integration, providing a more comprehensive understanding of the process. Nanopore sequencing has been instrumental in characterizing the integration patterns of the HPV genome into host DNA, providing insights into the molecular events driving cervical cancer (52, 64–66). Importantly,

the study showed that within the same sample, separate integration events often clustered closely, with partial overlap at different breakpoints (52, 65).

3.3 Implications on epigenetics and functional genomics

Third-generation sequencing technologies, when juxtaposed with functional genomics methodologies like RNA-seq and chromatin immunoprecipitation sequencing (ChIP-seq), forge a powerful toolset. In the context of investigating CpG methylation in novel and existing transposable element (TE) insertions, both Oxford Nanopore Technologies and PacBio long-read sequencing techniques offer valuable capabilities (60, 67, 68). These sequencing platforms enable the analysis of CpG methylation patterns within TE insertions in a comprehensive and detailed manner. Specifically, researchers can employ ONT and PacBio long-read sequencing to capture long sequences and comprehensively cover TE insertions, many of which can be extensive and complex. Moreover, these long sequencing technologies can resolve methylation patterns and allow for precise mapping of CpG methylation, providing important insights into the epigenetic alterations associated with tumorigenesis and other biological processes (69, 70). This combination enables researchers to probe deeper, investigating the ramifications of viral integration on host gene expression, epigenetic shifts, and chromatin configurations. It is this intersection of technologies that holds the potential to unravel the mechanisms steering viral oncogenesis. Moreover, as sequencing technologies evolve, so do the methods for virus detection. Current and emerging molecular and immunological methods are expanding the toolkit for researchers, allowing for more comprehensive and accurate identification of viruses in various samples.

3.4 Analysis of HPV integration

Recent progress in sequencing technology has led to the capability of producing exceptionally long reads, averaging around 100 kb. This advancement includes the generation of high-throughput, full-length mRNA or cDNA reads and the construction of genomic contigs exceeding 100 Mb. However, this technological leap presents a challenge: existing alignment programs are either incapable or inefficient in handling such extensive data, highlighting the urgent need for new, more capable alignment algorithms. The field of bioinformatics has been evolving rapidly to meet these demands, with significant strides made in the development of alignment methods for long sequencing reads. These methods are particularly adept at chimeric sequence-aware alignment, crucial for identifying reads that align with both human and viral reference genomes – a key process in understanding pathogen-host interactions and disease mechanisms.

One standout tool in this domain is Minimap2 (71), a versatile alignment program designed to map DNA or long mRNA sequences against extensive reference databases. Alongside Minimap2, other notable aligners have emerged, including desalt

(72), which is tailored for aligning long DNA or mRNA sequences. Additionally, tools originally developed for short reads are being adapted and enhanced for the unique challenges posed by error-prone long reads. This includes TAGET (73) and a newly adapted mode of the STAR aligner (73). These improvements reflect the ongoing evolution of bioinformatics tools, ensuring they remain at the forefront of managing and interpreting the burgeoning data from next-generation sequencing technologies.

4 Discussion

Viral-induced genomic instability and clonal evolution represent an intricate and multifaceted aspect of viral infections. Understanding the mechanisms and consequences of this phenomenon is essential for the development of effective therapeutic strategies. As our knowledge continues to expand, further research is needed to unveil new insights into this complex relationship and its potential implications for human health. Ultimately, deciphering the interplay between viruses and host genomes will be critical in advancing our ability to prevent and treat viral-associated diseases.

Third-generation sequencing technologies have ushered in a new era in the study of viral-related tumors, offering unprecedented insights into the molecular intricacies of viral oncogenesis. The intricate interplay between viruses, oncogenes, and human cancers remains a vital research frontier, with oncogenic viruses playing an undeniable role in global cancer epidemiology. Our understanding of this dynamic has been immensely enriched by third-generation sequencing technologies, which have cast light on the hitherto elusive mechanisms underlying viral integration into the human genome.

The advent of third generation sequencing has revolutionized our capacity to pinpoint viral integration sites with exceptional resolution, enabling a deeper comprehension of their implications on functional genomics. As the genomic landscape of cancer continually evolves, these technologies provide a dynamic lens, capturing shifts in viral integration sites over time and documenting tumor progression and heterogeneity. The ability to map viral integration sites, study epigenetic modifications, and track clonal evolution provides a multifaceted view of these complex diseases. The practical implications are vast, from innovative diagnostic markers to personalized treatment strategies and the exploration of novel therapeutic targets.

However, the significance of these insights extends beyond the domain of academic research. The translational potential of understanding viral oncogenesis is vast. Recognizing unique viral integration patterns can provide innovative diagnostic markers, heralding early detection and improved patient outcomes. The meticulous mapping of the relationship between viral integration and cellular response can pave the way for novel therapeutic strategies. The insights gained from third-generation sequencing in viral-related tumors hold significant potential for clinical applications. Here are some practical implications: 1) Innovative Diagnostic Markers: Third-generation sequencing can identify unique patterns of viral integration and epigenetic alterations in viral-related tumors. These patterns can serve as innovative

diagnostic markers, allowing for the early detection and stratification of cancer subtypes. 2) Personalized Treatment Strategies: Understanding the clonal evolution and heterogeneity of viral-related tumors through third-generation sequencing helps in tailoring personalized treatment approaches. This can lead to more effective therapies and improved patient outcomes. 3) Exploring Novel Therapeutic Targets: The precise characterization of viral oncogenes and their interactions with the host genome can uncover novel therapeutic targets. Inhibiting viral-specific factors can be a promising avenue for targeted cancer therapies. 4) Monitoring Treatment Response: Third-generation sequencing can be used to monitor the response of viral-related tumors to treatments over time. This real-time monitoring can guide adjustments in therapeutic strategies and help predict the course of the disease.

As we move forward, addressing the challenges of data analysis, integration with other omics data, validation in clinical settings, and cost considerations will be essential (74, 75). It is also important to integrate these findings with other functional and clinical data. This holistic approach will be essential in realizing the full potential of these insights, from bench to bedside. While third-generation sequencing has brought about remarkable advancements in the study of viral-related tumors, several challenges and future directions warrant attention. 1) Data Analysis and Interpretation: Managing the large volumes of data generated by third-generation sequencing is a considerable challenge. The development of robust bioinformatics tools and analytical pipelines is crucial to extract meaningful insights. 2) Integration with Other Omics Data: A holistic understanding of viral-related tumors often requires the integration of genomic, transcriptomic, proteomic, and epigenomic data. Future research should focus on methods for seamless integration of these multidimensional datasets. 3) Validation in Clinical Settings: To translate the findings from third-generation sequencing into clinical practice, rigorous validation studies are essential. These studies should assess the diagnostic and prognostic utility of the identified markers and therapeutic targets. 4) Cost and Accessibility: While third-generation sequencing technologies have advanced, their cost and accessibility remain barriers to widespread adoption. Efforts to reduce costs and improve accessibility will be instrumental in realizing the full potential of these technologies in clinical settings.

Author contributions

RY: Conceptualization, Investigation, Writing – original draft, Writing – review & editing. AW: Data curation, Writing – original draft, Writing – review & editing. BB: Investigation, Writing – original draft, Writing – review & editing. PL: Data curation, Writing – review & editing. WD: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Supervision, Writing – original draft, Writing – review & editing. XZ: Funding acquisition, Investigation, Methodology, Resources, Supervision, Writing – original draft, Writing – review & editing. SY: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Resources, Supervision, Writing – original draft, Writing – review & editing, Methodology, Project administration, Software, Validation, Visualization.

Acknowledgments

This work was in part supported by NIH R21 CA267527-01 (SY) and the Feldstein Medical Foundation grant (SY), Domestic Cooperation Project of the Science and Technology Department of Jiangxi province No. 20212BDH81021 (XZ), Project of the Science and Technology Department of Jiangxi province No.20203BBGL73148 (XZ) and Project of Health Commission of Jiangxi Province No.202210038 (XZ).

Conflict of interest

WD and SY hold equity in Yihui Bio, Inc.

References

- Li YL, Roberts ND, Wala JA, Shapira O, Schumacher SE, Kumar K, et al. Patterns of somatic structural variation in human cancer genomes. *Nature* (2020) 578:112. doi: 10.1038/s41586-019-1913-9
- Sakamoto Y, Sereewattanawoot S, Suzuki A. A new era of long-read sequencing for cancer genomics. *J Hum Genet* (2020) 65:3–10. doi: 10.1038/s10038-019-0658-5
- Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. *Nat Rev Genet* (2006) 7:85–97. doi: 10.1038/nrg1767
- Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, et al. Fine-scale structural variation of the human genome. *Nat Genet* (2005) 37:727–32. doi: 10.1038/ng1562
- Lin YL, Gokcumen O. Fine-scale characterization of genomic structural variation in the human genome reveals adaptive and biomedically relevant hotspots. *Genome Biol Evol* (2019) 11:1136–51. doi: 10.1093/gbe/evz058
- Yadav S, Hu CL, Nathanson KL, Weitzel JN, Goldgar DE, Kraft P, et al. Germline pathogenic variants in cancer predisposition genes among women with invasive lobular carcinoma of the breast. *J Clin Oncol* (2021) 39:3918. doi: 10.1200/JCO.21.00640
- Giunta S. Decoding human cancer with whole genome sequencing: a review of PCAWG Project studies published in February 2020. *Cancer Metastasis Rev* (2021) 40:909–24. doi: 10.1007/s10555-021-09969-z
- Shendure J, Balasubramanian S, Church GM, Gilbert W, Rogers J, Schloss JA, et al. DNA sequencing at 40: past, present and future. *Nature* (2017) 550:345–53. doi: 10.1038/nature24286
- Manier S, Park J, Capelletti M, Bustoros M, Freeman SS, Ha G, et al. Whole-exome sequencing of cell-free DNA and circulating tumor cells in multiple myeloma. *Nat Commun* 9 (2018). doi: 10.1038/s41467-018-04001-5
- Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* (2013) 499:214–8. doi: 10.1038/nature12213
- Zapatka M, Borozan 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
- Krump NA, You J. Molecular mechanisms of viral oncogenesis in humans. *Nat Rev Microbiol* (2018) 16:684–98. doi: 10.1038/s41579-018-0064-6
- van Dijk EL, Jaszczyszyn Y, Naquin D, Thermes C. The third revolution in sequencing technology. *Trends Genet* (2018) 34:666–81. doi: 10.1016/j.tig.2018.05.008
- Wang KC, Chang HY. Epigenomics technologies and applications. *Circ Res* (2018) 122:1191–9. doi: 10.1161/CIRCRESAHA.118.310998
- Flaherty P, Natsoulis G, Muralidharan O, Winters M, Buenrostro J, Bell J, et al. Ultrasensitive detection of rare mutations using next-generation targeted resequencing. *Nucleic Acids Res* 40 (2012). doi: 10.1093/nar/gkr861
- de Villiers EM, Fauquet C, Broker TR, Bernard HU, zur Hausen HU. Classification of papillomaviruses. *Virology* (2004) 324:17–27. doi: 10.1016/j.virol.2004.03.033
- zur Hausen H. Papillomaviruses and cancer: From basic studies to clinical application. *Nat Rev Cancer* (2002) 2:342–50. doi: 10.1038/nrc798
- Gillison ML, D'Souza G, Westra W, Sugar E, Xiao WH, Begum S, et al. Distinct risk factor profiles for human papillomavirus type 16-positive and human papillomavirus type 16-negative head and neck cancers. *Jnci-Journal Natl Cancer Institute* (2008) 100:407–20. doi: 10.1093/jnci/djn025
- Wakeham K, Kavanagh K, Cuschieri K, Millan D, Pollock KG, Bell S, et al. HPV status and favourable outcome in vulvar squamous cancer. *Int J Cancer* (2017) 140:1134–46. doi: 10.1002/ijc.30523
- Tuna M, Amos CI. Next generation sequencing and its applications in HPV-associated cancers. *Oncotarget* (2017) 8:8877–89. doi: 10.18632/oncotarget.12830
- Shen CL, Liu YZ, Shi S, Zhang RY, Zhang T, Xu Q, et al. Long-distance interaction of the integrated HPV fragment with MYC gene and 8q24.22 region upregulating the allele-specific MYC expression in HeLa cells. *Int J Cancer* (2017) 141:540–8. doi: 10.1002/ijc.30763
- Mirabello L, Yeager M, Yu K, Clifford GM, Xiao YZ, Zhu B, et al. HPV16 E7 genetic conservation is critical to carcinogenesis. *Cell* (2017) 170:1164–74. doi: 10.1016/j.cell.2017.08.001
- Bouchard MJ, Navas-Martin S, Hepatitis B and C virus hepatocarcinogenesis: lessons learned and future challenges. *Cancer Lett* (2011) 305:123–43. doi: 10.1016/j.canlet.2010.11.014
- Guidotti LG, Chisari FV. Immunobiology and pathogenesis of viral hepatitis. *Annu Rev Pathol* (2006) 1:23–61. doi: 10.1146/annurev.pathol.1.110304.100230
- Chisari FV. Hepatitis B virus transgenic mice: models of viral immunobiology and pathogenesis. *Curr Top Microbiol Immunol* (1996) 206:149–73. doi: 10.1007/978-3-642-85208-4_9
- Das A, Rivera-Serrano EE, Yin X, Walker CM, Feng Z, Lemon SM. Cell entry and release of quasi-enveloped human hepatitis viruses. *Nat Rev Microbiol* (2023) 21:573–89. doi: 10.1038/s41579-023-00889-z
- Biswas A, Datta S. Editorial: Origin and evolution of hepatitis viruses, volume II. *Front Microbiol* (2023) 14:1241705. doi: 10.3389/fmicb.2023.1241705
- Anzola M. Hepatocellular carcinoma: role of hepatitis B and hepatitis C viruses proteins in hepatocarcinogenesis. *J Viral Hepat* (2004) 11:383–93. doi: 10.1111/j.1365-2893.2004.00521.x
- Heawchaiyaphum C, Ekalaksananan T, Patarapadungkit N, Vatanasapt P, Pientong C. Association of human papillomavirus and epstein-barr virus infection with tonsil cancer in Northeastern Thailand. *Asian Pac J Cancer Prev* (2022) 23:781–7. doi: 10.31557/APJCP.2022.23.3.781
- Mashaly M, Ghorab D, Hegazy M, Abdelkhalek M, Gaballah K, Elzehery R. Association between epstein-barr virus gene polymorphism and breast cancer risk among Egyptian females. *Asian Pac J Cancer Prev* (2022) 23:641–50. doi: 10.31557/APJCP.2022.23.2.641
- Tavakoli A, Monavari SH, Solaymani Mohammadi F, Kiani SJ, Armat S, Farahmand M. Association between Epstein-Barr virus infection and gastric cancer: a systematic review and meta-analysis. *BMC Cancer* (2020) 20:493. doi: 10.1186/s12885-020-07013-x
- Wu SS, Chen B, Fleming CW, Shah AA, Griffith CC, Domb C, et al. Nasopharyngeal cancer: Incidence and prognosis of human papillomavirus and Epstein-Barr virus association at a single North American institution. *Head Neck* (2022) 44:851–61. doi: 10.1002/hed.26976
- Arias-Calvachi C, Blanco R, Calaf GM, Aguayo F. Epstein-barr virus association with breast cancer: evidence and perspectives. *Biol (Basel)* (2022) 11. doi: 10.3390/biology11060799
- Carvalho M, de Melo YLF. Association between human papillomavirus and epstein-barr virus infections and cancer of the uterine cervix. *Crit Rev Oncog* (2019) 24:379–83. doi: 10.1615/CritRevOncog.2019031545

35. Gupta I, Al Farsi H, Jabeen A, Skenderi F, Al-Thawadi H, AlAhmad YM, et al. High-risk human papillomaviruses and Epstein-Barr virus in colorectal cancer and their association with clinicopathological status. *Pathogens* (2020) 9. doi: 10.3390/pathogens9060452
36. Rodriguez H, Zenklusen JC, Staudt LM, Doroshow JH, Lowy DR. The next horizon in precision oncology: Proteogenomics to inform cancer diagnosis and treatment. *Cell* (2021) 184:1661–70. doi: 10.1016/j.cell.2021.02.055
37. Hou YM, Dong J, Liu MY, Yu S. Expression of Epstein-Barr virus-induced gene 3 in cervical cancer: Association with clinicopathological parameters and prognosis. *Oncol Lett* (2016) 11:330–4. doi: 10.3892/ol.2015.3849
38. Patel PD, Alghareeb R, Hussain A, Maheshwari MV, Khalid N. The association of Epstein-Barr virus with cancer. *Cureus* (2022) 14:e26314. doi: 10.7759/cureus.26314
39. Santos E, Silva M, Canedo NHS, Gauí MFD, Brito A, Silva W, et al. Epstein-Barr virus in gastric cancer and association with 30 bp del-latent membrane protein 1 polymorphism. *Rev Assoc Med Bras* (1992) (2023) 69:e20221571. doi: 10.1590/1806-9282.20221571
40. Zhang WT, Zhu GL, Xu WQ, Zhang W, Wang HZ, Wang YB, et al. Association of PD-1/PD-L1 expression and Epstein-Barr virus infection in patients with invasive breast cancer. *Diagn Pathol* (2022) 17:61. doi: 10.1186/s13000-022-01234-3
41. Miura M, Naito T, Saito M. Current perspectives in human T-cell leukemia virus type 1 infection and its associated diseases. *Front Med (Lausanne)* (2022) 9:867478. doi: 10.3389/fmed.2022.867478
42. Ghobadi MZ, Afsaneh E, Emamzadeh R, Soroush M. Potential miRNA-gene interactions determining progression of various ATLL cancer subtypes after infection by HTLV-1 oncovirus. *BMC Med Genomics* (2023) 16:62. doi: 10.1186/s12920-023-01492-0
43. Valcarcel B, Enriquez-Vera D, De-la-Cruz-Ku G, Chambergo-Michilot D, Calderon-Huaycochea H, Malpica L. Epidemiological features and outcomes of HTLV-1 carriers diagnosed with cancer: A retrospective cohort study in an endemic country. *JCO Glob Oncol* (2023) 9:e2200369. doi: 10.1200/GO.22.00369
44. Zarei Ghobadi M, Afsaneh E, Emamzadeh R. Gene biomarkers and classifiers for various subtypes of HTLV-1-caused ATLL cancer identified by a combination of differential gene co-expression and support vector machine algorithms. *Med Microbiol Immunol* (2023) 212:263–70. doi: 10.1007/s00430-023-00767-8
45. Gao G, Wang JT, Kasperbauer JL, Tombers NM, Teng F, Gou HL, et al. Whole genome sequencing reveals complexity in both HPV sequences present and HPV integrations in HPV-positive oropharyngeal squamous cell carcinomas. *BMC Cancer* (2019). doi: 10.1186/s12885-019-5536-1
46. Koneva LA, Zhang YX, Virani S, Hall PB, McHugh JB, Chepeha DB, et al. HPV integration in HNSCC correlates with survival outcomes, immune response signatures, and candidate drivers. *Mol Cancer Res* (2018) 16:90–102. doi: 10.1158/1541-7786.MCR-17-0153
47. Ibragimova M, Tsyganov M, Shpileva O, Churuksaeva O, Bychkov V, Kolomiets L, et al. HPV status and its genomic integration affect survival of patients with cervical cancer. *Neoplasia* (2018) 65:441–8. doi: 10.4149/neo_2018_170414N277
48. Kumar S, Razzaq SK, Vo AD, Gautam M, Li H. Identifying fusion transcripts using next generation sequencing. *Wiley Interdiscip Reviews-Rna* (2016) 7:811–23. doi: 10.1002/wrna.1382
49. Sedlazeck FJ, Lee H, Darby CA, Schatz MC. Piercing the dark matter: bioinformatics of long-range sequencing and mapping. *Nat Rev Genet* (2018) 19:329–46. doi: 10.1038/s41576-018-0003-4
50. Lee JY, Cutts RJ, White I, Augustin Y, Garcia-Murillas I, Fenwick K, et al. Next generation sequencing assay for detection of circulating HPV DNA (cHPV-DNA) in patients undergoing radical (Chemo)Radiotherapy in anal squamous cell carcinoma (ASCC). *Front Oncol* (2020) 10. doi: 10.3389/fonc.2020.00505
51. Mühr LSA, Lagheden C, Eklund C, Lei JY, Nordqvist-Kleppe S, Sparén P, et al. Sequencing detects human papillomavirus in some apparently HPV-negative invasive cervical cancers. *J Gen Virol* (2020) 101:265–70. doi: 10.1099/jgv.0.001374
52. 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
53. Mitchell K, Brito JJ, Mandric I, Wu QZ, Knyazev S, Chang S, et al. Benchmarking of computational error-correction methods for next-generation sequencing data. *Genome Biol* (2020). doi: 10.1186/s13059-020-01988-3
54. Pareek CS, Smoczynski R, Tretyn A. Sequencing technologies and genome sequencing. *J Appl Genet* (2011) 52:413–35. doi: 10.1007/s13353-011-0057-x
55. Mardis ER. The impact of next-generation sequencing technology on genetics. *Trends Genet* (2008) 24:133–41. doi: 10.1016/j.tig.2007.12.007
56. Branton D, Deamer DW, Marziali A, Bayley H, Benner SA, Butler T, et al. The potential and challenges of nanopore sequencing. *Nat Biotechnol* (2008) 26:1146–53. doi: 10.1038/nbt.1495
57. Chan WS, Chan TL, Au CH, Leung CP, To MY, Ng MK, et al. An economical Nanopore sequencing assay for human papillomavirus (HPV) genotyping. *Diagn Pathol* (2020) 15. doi: 10.1186/s13000-020-00964-6
58. Sarkar A, Gasperi W, Begley U, Nevins S, Huber SM, Dedon PC, et al. Detecting the epitranscriptome. *Wiley Interdiscip Reviews-Rna* (2021) 12. doi: 10.1002/wrna.1663
59. Li YY. Modern epigenetics methods in biological research. *Methods* (2021) 187:104–13. doi: 10.1016/j.ymeth.2020.06.022
60. Wang Y, Zheng DL, Tan QL, Wang MX, Gu LQ. Nanopore-based detection of circulating microRNAs in lung cancer patients. *Nat Nanotechnology* (2011) 6:668–74. doi: 10.1038/nnano.2011.147
61. Gu LQ, Wanunu M, Wang MX, McReynolds L, Wang Y. Detection of miRNAs with a nanopore single-molecule counter. *Expert Rev Mol Diagnostics* (2012) 12:573–84. doi: 10.1586/erm.12.58
62. Wylie TN, Wylie KM, Herter BN, Storchl GA. Enhanced virome sequencing using targeted sequence capture. *Genome Res* (2015) 25:1910–20. doi: 10.1101/gr.191049.115
63. Chandrani P, Kulkarni V, Iyer P, Upadhyay P, Chaubal R, Das P, et al. NGS-based approach to determine the presence of HPV and their sites of integration in human cancer genome. *Br J Cancer* (2015) 112:1958–65. doi: 10.1038/bjc.2015.121
64. Li X, Wei X, Liu X, Wang N, Xu F, Liu X, et al. The analysis of HPV integration sites based on nanopore sequencing and the profiling changes along the course of photodynamic therapy. *BMC Cancer* (2023) 23:1052. doi: 10.1186/s12885-023-11538-2
65. Yang WJ, Liu Y, Dong RY, Liu J, Lang JD, Yang JL, et al. Accurate detection of HPV integration sites in cervical cancer samples using the nanopore minION sequencer without error correction. *Front Genet* (2020) 11. doi: 10.3389/fgenet.2020.00660
66. Mühr LSA, Lagheden C, Lei JY, Eklund C, Kleppe SN, Sparén P, et al. Deep sequencing detects human papillomavirus (HPV) in cervical cancers negative for HPV by PCR. *Br J Cancer* (2020) 123:1790–5. doi: 10.1038/s41416-020-01111-0
67. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods* (2010) 7:461–U72. doi: 10.1038/nmeth.1459
68. Wallace EVB, Stoddard D, Heron AJ, Mikhailova E, Maglia G, Donohoe TJ, et al. Identification of epigenetic DNA modifications with a protein nanopore. *Chem Commun* (2010) 46:8195–7. doi: 10.1039/c0cc02864a
69. Nebbioso A, Tambaro FP, Dell'Aversana C, Altucci L. Cancer epigenetics: Moving forward. *PLoS Genet* (2018). doi: 10.1371/journal.pgen.1007362
70. Rivera CM, Ren B. Mapping human epigenomes. *Cell* (2013) 155:39–55. doi: 10.1016/j.cell.2013.09.011
71. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* (2018) 34:3094–100. doi: 10.1093/bioinformatics/bty191
72. Liu B, Liu Y, Li J, Guo H, Zang T, Wang Y. deSALT: fast and accurate long transcriptomic read alignment with de Bruijn graph-based index. *Genome Biol* (2019) 20:274. doi: 10.1186/s13059-019-1895-9
73. Xia Y, Jin Z, Zhang C, Ouyang L, Dong Y, Li J, et al. TAGET: a toolkit for analyzing full-length transcripts from long-read sequencing. *Nat Commun* (2023) 14:5935. doi: 10.1038/s41467-023-41649-0
74. Olivier M, Asmis R, Hawkins GA, Howard TD, Cox LA. The need for multi-omics biomarker signatures in precision medicine. *Int J Mol Sci* (2019). doi: 10.3390/ijms20194781
75. Nagahashi M, Shimada Y, Ichikawa H, Kameyama H, Takabe K, Okuda S, et al. Next generation sequencing-based gene panel tests for the management of solid tumors. *Cancer Sci* (2019) 110:6–15. doi: 10.1111/cas.13837