



Whole Genome Assembly of Human Papillomavirus by Nanopore Long-Read Sequencing

Shuaibing Yang^{1†}, Qianqian Zhao^{2†}, Lihua Tang³, Zejia Chen³, Zhaoting Wu³, Kaixin Li⁴, Ruoru Lin⁴, Yang Chen⁴, Danlin Ou⁴, Li Zhou^{3*}, Jianzhen Xu^{2*} and Qingsong Qin^{1,5,6*}

¹Laboratory of Human Virology and Oncology, Shantou University Medical College, Shantou, China, ²Computational Systems Biology Lab, Department of Bioinformatics, Shantou University Medical College, Shantou, China, ³Department of Gynecologic Oncology, Cancer Hospital of Shantou University Medical College, Shantou, China, ⁴Undergraduate Program of Innovation and Entrepreneurship, Shantou University Medical College, Shantou, China, ⁵Guangdong Provincial Key Laboratory of Infectious Diseases and Molecular Immunopathology, Shantou, China, ⁶Guangdong Provincial Key Laboratory for Diagnosis and Treatment of Breast Cancer, Shantou, China

OPEN ACCESS

Edited by:

Damjan Glavač,
University of Ljubljana, Slovenia

Reviewed by:

Iwao Kukimoto,
National Institute of Infectious
Diseases, Japan
Natasha Andressa Jorge,
Leipzig University, Germany

*Correspondence:

Qingsong Qin
qsqin@stu.edu.cn
Jianzhen Xu
jzxu01@stu.edu.cn
Li Zhou
zlyyzi@126.com

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Genomic Assay Technology,
a section of the journal
Frontiers in Genetics

Received: 20 October 2021

Accepted: 01 December 2021

Published: 04 January 2022

Citation:

Yang S, Zhao Q, Tang L, Chen Z, Wu Z,
Li K, Lin R, Chen Y, Ou D, Zhou L, Xu J
and Qin Q (2022) Whole Genome
Assembly of Human Papillomavirus by
Nanopore Long-Read Sequencing.
Front. Genet. 12:798608.
doi: 10.3389/fgene.2021.798608

Human papillomavirus (HPV) is a causal agent for most cervical cancers. The physical status of the HPV genome in these cancers could be episomal, integrated, or both. HPV integration could serve as a biomarker for clinical diagnosis, treatment, and prognosis. Although whole-genome sequencing by next-generation sequencing (NGS) technologies, such as the Illumina sequencing platform, have been used for detecting integrated HPV genome in cervical cancer, it faces challenges of analyzing long repeats and translocated sequences. In contrast, Oxford nanopore sequencing technology can generate ultra-long reads, which could be a very useful tool for determining HPV genome sequence and its physical status in cervical cancer. As a proof of concept, in this study, we completed whole genome sequencing from a cervical cancer tissue and a CaSki cell line with Oxford Nanopore Technologies. From the cervical cancer tissue, a 7,894 bp-long HPV35 genomic sequence was assembled from 678 reads at 97-fold coverage of HPV genome, sharing 99.96% identity with the HPV sequence obtained by Sanger sequencing. A 7904 bp-long HPV16 genomic sequence was assembled from data generated from the CaSki cell line at 3857-fold coverage, sharing 99.99% identity with the reference genome (NCBI: U89348). Intriguingly, long reads generated by nanopore sequencing directly revealed chimeric cellular-viral sequences and concatemeric genomic sequences, leading to the discovery of 448 unique integration breakpoints in the CaSki cell line and 60 breakpoints in the cervical cancer sample. Taken together, nanopore sequencing is a unique tool to identify HPV sequences and would shed light on the physical status of HPV genome in its associated cancers.

Keywords: HPV, nanopore sequencing, cervical cancer, integration, episomal genome

INTRODUCTION

Human papillomavirus (HPV), a double-stranded circular DNA virus, is a causal agent for most cervical cancers (Zur Hausen, 2002; Abu-Lubad et al., 2020) and is also associated with anal cancer (Alemany et al., 2015), oropharyngeal cancer (Meng et al., 2020), and vaginal cancer (Hellman et al., 2004). Until now, more than 200 genotypes of HPV have been identified (Bzhalava et al., 2015) and

can be classified as high-risk or low-risk genotypes based on their properties of tumorigenesis (de Villiers, 2013). Persistent infection by high-risk HPV is more likely to initially lead to precancerous cervical intraepithelial neoplasia (CIN) and finally result in invasive cervical carcinoma (ICC) (Walboomers et al., 1999; Zur Hausen, 2002; Woodman et al., 2007; Yuan et al., 2021). Integration of HPV DNA into the host genome is considered a key event in driving cervical carcinogenesis (Pett and Coleman, 2007; Zhang et al., 2016; Warburton et al., 2018). The break sites in the cellular–viral junction region are commonly found in E1 and E2 genes in the viral genome, which leads to the loss of expression of viral E2 protein and the increased expression of viral oncoproteins (E6 and E7) that inhibit cell cycle checkpoint proteins p53 and pRB (Woodman et al., 2007; Groves and Coleman, 2015) and is essential for the proliferation and survival of HPV-related cancer cells (Hoppe-Seyler et al., 2018). The frequency of HPV integration is positively correlated with cervical intraepithelial neoplasia (CIN) (Hudelist et al., 2004; Guo et al., 2007; Wang et al., 2013). In HPV-related cancers, the physical status of HPV genome has been found to be episomal, integrated, or mixed (Park et al., 1997; Nambaru et al., 2009; Niya et al., 2019). For example, the CaSki cells, a naturally derived cervical carcinoma cell line, contain a high number of concatemeric HPV genomic sequences inserted within cellular genome (Yee et al., 1985) and focal genomic variations at the integration locus (McBride and Warburton, 2017). It was previously reported that patients with episomal HPV in cancer cells have a better survival rate than those with an integrated HPV genome (Nambaru et al., 2009; Kiseleva et al., 2013). Various methods have been used to detect the physical status of the HPV genome, such as PCR-based methods (Carmody et al., 1996; Luft et al., 2001), fluorescence *in situ* hybridization (FISH) (Triglia et al., 2009; Houldsworth, 2014; Olthof et al., 2015), whole genome DNA sequencing (WGS) and high-throughput viral integration detection (HIVID) (Hu et al., 2015), RNA sequencing (RNA-seq) (Klaes et al., 1999), and single-molecule sequencing technology (SMRT-seq) (Yu et al., 2021). However, each method has its strengths and limitations. For example, whole genome sequencing by next-generation sequencing (NGS) technologies (Hu et al., 2015; Kamal et al., 2021) requires a high coverage of human genome (>30 x coverage) and a complicated algorithm to analyze HPV integrations in human genome. Short-reads (100–500 bp) generated by NGS can lead to errors and ambiguity in mapping viral integration and assembling repetitive sequences (Alkan et al., 2011; Treangen and Salzberg, 2011).

Nanopore long-read sequencing is a third-generation sequencing technology released by Oxford Nanopore Technologies in 2014 (Jain et al., 2015), which determines DNA bases by reading the current fluctuation of passing nucleotides through biological nanopores (Lu et al., 2016) and therefore has the ability of reading ultra-long sequences (Jain et al., 2018). Ultra-long reads are more likely to cover the complete viral genome, highly repetitive regions, and structural variations in human genome, allowing continuous and complete genome assembly (Jain et al., 2018; Senol Cali et al., 2018). With their long reading capacity, nanopore

sequencing technologies have shown great advantages in the rapid surveillance of Ebola (Hoenen, 2016; Quick et al., 2016) and Zika viruses (Quick et al., 2017), genotyping and genetic diversity analysis of hepatitis B virus (HBV) (Sauvage et al., 2018; McNaughton et al., 2019; Astbury et al., 2020), and monitoring emerging infectious disease outbreaks (Zhu et al., 2020). Whole genome sequencing (WGS) of SARS-CoV-2 with nanopore sequencing technologies provided a unique tool to analyze viral transmission, evolution, and genomic variation (Bull et al., 2020; Lin et al., 2021). In this study, as a proof of concept, we sequenced total DNA extracted from a cervical cancer tissue and a CaSki cell line with nanopore sequencing technologies, assembled HPV genomes with several bioinformatic tools, and analyzed the physical status of HPV.

MATERIALS AND METHODS

Sample Collection

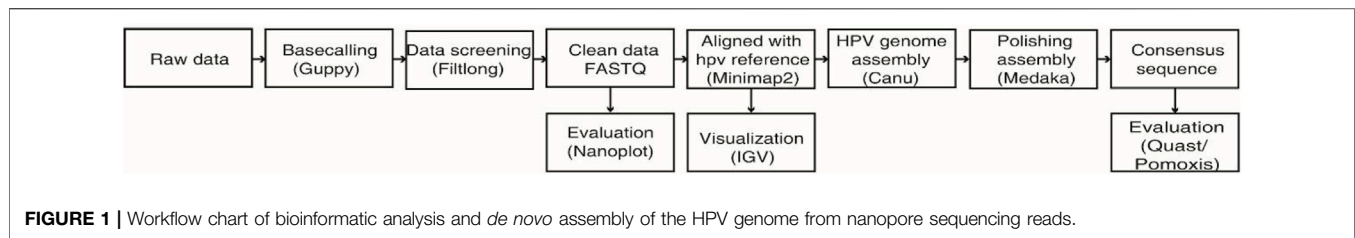
A fresh tissue sample was obtained from a 74-year-old patient with cervical carcinoma without prior chemotherapy or radiotherapy at the Cancer Hospital of Shantou University Medical College (Guangdong, China) in 2019. An informed consent form was obtained from the patient. This study was approved by the ethical review board of Shantou University Medical College (approval number: SUMC-2020-51).

Cell Culture

CaSki cells (CRL-1550, ATCC, United States) were cultured in RPMI-1640 medium (SH30809.01B, Hyclone, Logan, UT, United States) with 10% fetal bovine serum (Ausbian, Australia), 100 U/ml penicillin, and 100 mg/ml streptomycin in a humidified incubator with 5% CO₂ at 37°C. C-33A cells (TCHu176, Shanghai Cell Bank, Chinese Academy of Sciences, China) were cultured in Dulbecco's modified Eagle medium (SH30022.01, Hyclone, Logan, UT, United States) containing 10% fetal bovine serum at 37°C with 5% CO₂.

MinION DNA Library Preparation

Total DNAs were respectively extracted from a cervical cancer tissue sample and CaSki cells using a QIAamp[®] DNA Mini Kit (C51304, Qiagen, Hilden, Germany) according to the manufacturer's protocol and quantified using a Qubit dsDNA HS Assay Kit (Q33230, Thermo Fisher Scientific, Waltham, MA, United States). Two µg of DNA fragments (>10 kb) was used for library preparation and sequenced using a MinION SQK-LSK109 Oxford nanopore sequencing kit (SQK-LSKSP9, Oxford Nanopore Technologies, Oxford, United Kingdom) according to the manufacturer's instructions. Briefly, DNA was repaired with FFPE DNA Repair Mix and End repair/dA-tailing Module reagents [E7695, New England BioLabs (NEB), Ipswich, MA] and purified with AMPure XP beads (A63880, Beckman Coulter, United States), washed with 70% ethanol, and eluted with nuclease-free water. Sequencing adapters were added to the 3' ends of fragmented DNA using Adapter Mix and Quick T4 DNA Ligase with Ligation Buffer (NEB) and purified with AMPure beads. The prepared library was added into a SpotON flow cell



(FLO-MIN106D, Oxford Nanopore Technologies) and sequenced using a MinION sequencer Mk1B.

De Novo Assembly of HPV Genome With Bioinformatic Tools

As shown in **Figure 1**, FAST5 files generated by the MinION sequencing device were converted to FASTQ using GUPPY (v3.1.5), provided in the MinKNOW software package. Low-quality reads (under a Q-score = 7 threshold) were removed using Filtlong (v0.2.0). The distributions and read lengths of data were evaluated using NanoPlot (<https://github.com/wdecoster/NanoPlot>). For the assembly of HPV genome, sequences were aligned to a HPV reference sequence downloaded from the Papillomavirus Episteme database (PaVE) (<https://pave.niaid.nih.gov/>) with Minimap2 (v2.17) using the following parameters (-t 20 -ax map-ont -Y) and extracted using SAMtools (v1.9) software (Li, 2018). The coverage and sequencing depth were calculated using Bamdst (v1.0.9) (<https://github.com/shiquan/bamdst>) and the mapping of genomic data was visualized using Integrative Genomic Viewer (IGV) (Thorvaldsdóttir et al., 2013). Next, the assembly of the HPV genome was done using Canu (Koren et al., 2017) under the following parameters (genome Size = 8k -stop on Low Coverage = 5 corrected Error Rate = 0.105 -nanopore-raw). Furthermore, the assembled HPV genome was corrected by using a polishing software called medaka (v1.2.1) and with the medaka consensus option (GitHub - nanoporetech/medaka: Sequence correction provided by ONT Research). A consensus HPV genomic sequence was then obtained by removing the duplicate parts, followed by assessing the assembly quality using Quast (Gurevich et al., 2013). We evaluated the base level error rate of the assembly result using Pomoxis software developed by Oxford Nanopore Technologies (<https://github.com/nanoporetech/pomoxis>). Finally, we used the dnadiff option of mummer (4.0.0) (Marçais et al., 2018) software to analyze the differences between assembled sequences and HPV reference genomes deposited in PaVE.

Analysis of HPV Integration Sites

Reads that are both aligned to human reference genome (UCSC, hg38) and viral reference genome are defined as chimeric sequences, which are screened using Minimap2 software. Chimeric sequences were extracted using SAMtools (v1.9). The cellular-viral joint sites (also known as breakpoints) were further determined using BLAT (v35) (Kent, 2002) by repositioning chimeric sequences to human and HPV genome using BLAT (v35) with the following parameters (-stepSize = 5 -repMatch =

2,253 -minScore = 20 -minIdentity = 0). HPV integration sites were defined as gaps or overlaps between host and viral alignments resulting in less than 10 bp. Those with the highest score were selected as credible positions. Gene information of the breakpoint sites was annotated using the bedtools intersect option based on files downloaded from the website (<https://www.encodegenes.org/>).

PCR Verification

To verify the HPV35 genome obtained by nanopore sequencing, 17 pairs of primers (**Table 1**) were designed to amplify HPV genome from the cervical cancer tissue to ensure the full coverage of an intact HPV genome. Each pair of primers was designed to amplify about 400–600 bp using a PCR Mix kit (C113, Vazyme, Nanjing, China). PCR products were purified by electrophoresis and sequenced using the Sanger sequencing technologies. To verify HPV integration sites, two primers (**Table 1**) flanking integration sites were designed to amplify cellular-viral joint regions, and amplified products were subjected to gel purification and Sanger sequencing.

Plasmid Preparation

Partial E2 and E6 genes of HPV35 were amplified from DNA extracted from the cervical cancer tissue by primers 7 (F, R) and primers 1 (F, R) in **Table 1** and cloned into the pMD19-T vector (6,013, Takara, Dalian, China). Plasmids were amplified in DH5 α cells according to the manufacturer's instructions (KTSM101L, KT Life Technology, Shenzhen, China) and isolated using a Plasmid Mini Kit (P1105, GBCBIO Technologies, Guangzhou, China).

Determine Physical Status of HPV Genome in Cervical Cancer Tissue by Long-Range PCR and qPCR

To determine the existence of a circular HPV genome in the cervical cancer tissue and CaSki cells (no episomal HPV) (Yee et al., 1985), the extracted DNAs were treated with or without exonuclease V (M0345S, NEB) at 37°C for 3 h. A pair of divergent primers (primer 18 F, R) located in the L1 region of HPV35 (**Table 1**) was used to amplify a circular HPV35 genome. The circular DNA is resistant to exonuclease V. In addition, copy numbers of E2 and E6 in cervical cancer tissue and CaSki cells were quantified by qPCR using TB green qPCR kits (RR420A, Takara, Dalian, China) and calculated based on a standard curve established using 10-fold serially diluted E2 or E6 plasmids. The primers for E2 and E6 are listed in **Table 1**. The resistance of E2 or E6 to exonuclease V digestion was calculated based on qPCR results.

TABLE 1 | Primers used in this study.

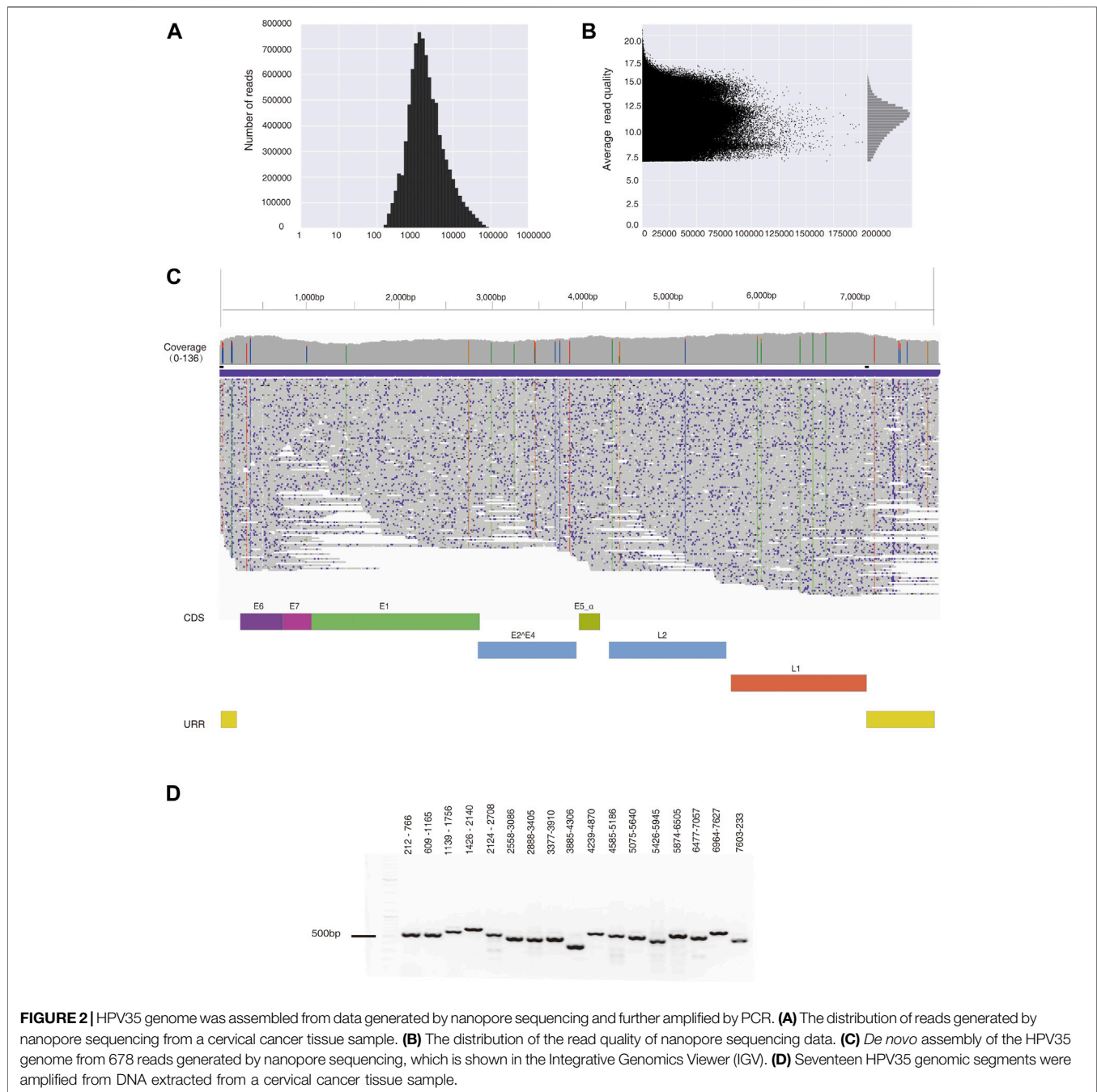
| Target | Primer designation | Sequence |
|--------------------|--------------------|----------------------------------|
| HPV35: 212–766 | Primer1(F) | 5'-CAAGAATTACAGCGGAGTGAGGT-3' |
| | Primer1(R) | 5'-ACAGACGTAGTGTGCGCTCACAT-3' |
| HPV35: 609–1,165 | Primer2(F) | 5'-ACCCGAGGCAACTGACCTATACTGT-3' |
| | Primer2(R) | 5'-AGCTCACGCTGCTAAGTGGACTA-3' |
| HPV35: 1,139–1756 | Primer3(F) | 5'-CTAGTAGTCCACTTAGCAGCGTGAG-3' |
| | Primer3(R) | 5'-TTGGTGGTTGTATTAGCATACTTGC-3' |
| HPV35: 1,426–2,140 | Primer4(F) | 5'-AGTAATGCAAAACGCAGCTATGTTG-3' |
| | Primer4(R) | 5'-AGTCACCGTCATCGTCCACCTTT-3' |
| HPV35: 2,124–2,708 | Primer5(F) | 5'-GGACGATGACGGTGACTGGAGG-3' |
| | Primer5(R) | 5'-TTGTCTCTTCTCTCGTGCAAAAT-3' |
| HPV35: 2,558–3,086 | Primer6(F) | 5'-ACTTACATAGCAGGGTAGTGGTCTT-3' |
| | Primer6(R) | 5'-CACCATCAAATGTACTTCCACTGT-3' |
| HPV35: 2,888–3,405 | Primer7(F) | 5'-GTGGTTCACACGCAGGCCATTTC-3' |
| | Primer7(R) | 5'-CCCACGGAGCAGGCTTTGGTATG-3' |
| HPV35: 3,377–3,910 | Primer8(F) | 5'-GAGACCCATACCAAAGCCTGCTC-3' |
| | Primer8(R) | 5'-GCAATAGCGAACGTACAAGCAGA-3' |
| HPV35: 3,885–4,306 | Primer9(F) | 5'-GTGTCTGCTTGTACGTTTCGCTAT-3' |
| | Primer9(R) | 5'-AACATCTGGTGGACAAGTTCCTG-3' |
| HPV35: 4,239–4,870 | Primer10(F) | 5'-TTAAACGTGCATCTGCAACACAAC-3' |
| | Primer10(R) | 5'-AGGGCGAGACCTGGAATAGGC-3' |
| HPV35: 4,585–5,186 | Primer11(F) | 5'-TGTTACACCAAGGGTCCCACCTA-3' |
| | Primer11(R) | 5'-CCTGATAATAATGTACCCGTGCC-3' |
| HPV35: 5,075–5,640 | Primer12(F) | 5'-CCTGCACTAACATCTAGGAAAGG-3' |
| | Primer12(R) | 5'-GCAGGTAGACAGTGGCTTCGTTA-3' |
| HPV35: 5,426–5,945 | Primer13(F) | 5'-CCTATAACAGCAGGGCCAGACAT-3' |
| | Primer13(R) | 5'-TACACCCAATGGCTGACCACGAC-3' |
| HPV35: 5,874–6,505 | Primer14(F) | 5'-GATCCTGCCTCCAGCGTTTGGT-3' |
| | Primer14(R) | 5'-GCATCGGAGTTACCATAGAGCCAC-3' |
| HPV35: 6,477–7,057 | Primer15(F) | 5'-CCTAGTGGCTCTATGTTAACCTCCG-3' |
| | Primer15(R) | 5'-GGAGCTGCACGCTTGCCTAATCT-3' |
| HPV35: 6,964–7,627 | Primer16(F) | 5'-CAGACTTAGATCAATTTCCGTTGG-3' |
| | Primer16(R) | 5'-GTGTGGGTGGACCACAAGTATGAA-3' |
| HPV35: 7,603–233 | Primer17(F) | 5'-TTTCATACTTGTGGTCCACCCA-3' |
| | Primer17(R) | 5'-CTCACTCCGCTGTAATTTCTTGT-3' |
| Divergent primers | Primer18(F) | 5'- ACAGCCTGTGATGTTACATAGCG -3' |
| | Primer18(R) | 5'- CAGACTTAGATCAATTTCCGTTGG -3' |
| HPV35: E2 | primer(F) | 5'-TATGGGAAGTGCATGTGGGTGGTC-3' |
| | primer(R) | 5'-GCAGTGTGCTGCGACTCGCTTGG-3' |
| HPV35: E6 | primer(F) | 5'-CTGAACGACCTTACAACTGC-3' |
| | primer(R) | 5'-TCACTCCGCTGTAATTTCTTGT-3' |
| HPV16: E2 | primer(F) | 5'-CATCAGTAACTGTGGTAGAGGGTC-3' |
| | primer(R) | 5'-GGATACTTCGTTGCTGCTAAAC-3' |
| HPV16: E6 | primer(F) | 5'-GAGCGACCCAGAAAGTTACCAC-3' |
| | primer(R) | 5'-ACGTCGCAGTAACTGTTGCTTG-3' |
| β -actin | primer(F) | 5'-TCCTCCTGAGCGCAAGTACTC-3' |
| | primer(R) | 5'-CGGACTCGTCATACTCCTGCTT-3' |
| PRR30 (BP) | primer(F) | 5'-GCTTTCCTCAACTACTGCCCTGTG-3' |
| | primer(R) | 5'-ATGCGCCAACGCCTTACATACCG-3' |

RESULTS

Assembly of HPV Genome From Data Generated by the MinION Nanopore Sequencing Platform

After filtering data with low-quality scores ($Q < 7$), 8,683,345 reads were obtained from a cervical cancer tissue sample. The average read length (N50) equals 9,875 bp (**Figure 2A**), and the mean read quality score is 11.2 (**Figure 2B**). A total of 36 Gb of data ensured 11.6-fold coverage of the human genome. Out of these, 678 reads are aligned to HPV35

genome (PaVE, GI:396997), accounting for 0.01% of total reads. The average sequencing depth for the assembled HPV35 genomic DNA is 97 x. A final consensus HPV sequence is 7,894 bp-long (**Figure 2C**, **Supplementary Table S3**), which shares 99.93% identity with a reference sequence (HPV35 16B, NCBI: KX514416.1). To further validate the assembled HPV35 from nanopore sequencing, 17 segments that cover the whole HPV genome were amplified from the cervical cancer tissue sample and sequenced with Sanger sequencing (**Figure 2D**). The final assembled genome from Sanger sequencing is 7,894 bp (**Supplementary Table S3**), which is 99.96% identical to the sequence obtained by



nanopore sequencing. In contrast, 4,211,267 reads were obtained from CaSki cells after filtering low-quality reads (Q score <7). The average read length (N50) is 11,862 bp, and the mean read quality score is 10.6. A total of 30 Gb data ensured about 10-fold coverage of the human genome. 15,947 reads were aligned to HPV16 genome (NCBI: U89348), accounting for 0.38% of total reads, and the average sequencing depth for HPV genome was 3,857 x. The final consensus HPV sequence is 7904 bp-long (**Supplementary Table S3**), which shares 99.99% identity with the reference sequence (NCBI: U89348). Taken together, from the whole

genome sequencing data generated by nanopore sequence technology, the assembly of HPV genome with high accuracy can be achieved with our bioinformatic strategy.

Unique Features of Ultra-Long Reads Generated by Nanopore Sequencing Reveal Different Forms of HPV Genome in Cervical Cancer Tissue Sample and CaSki Cells

In the two datasets generated from cervical cancer tissue and CaSki cells, there are two types of ultra-long reads containing viral

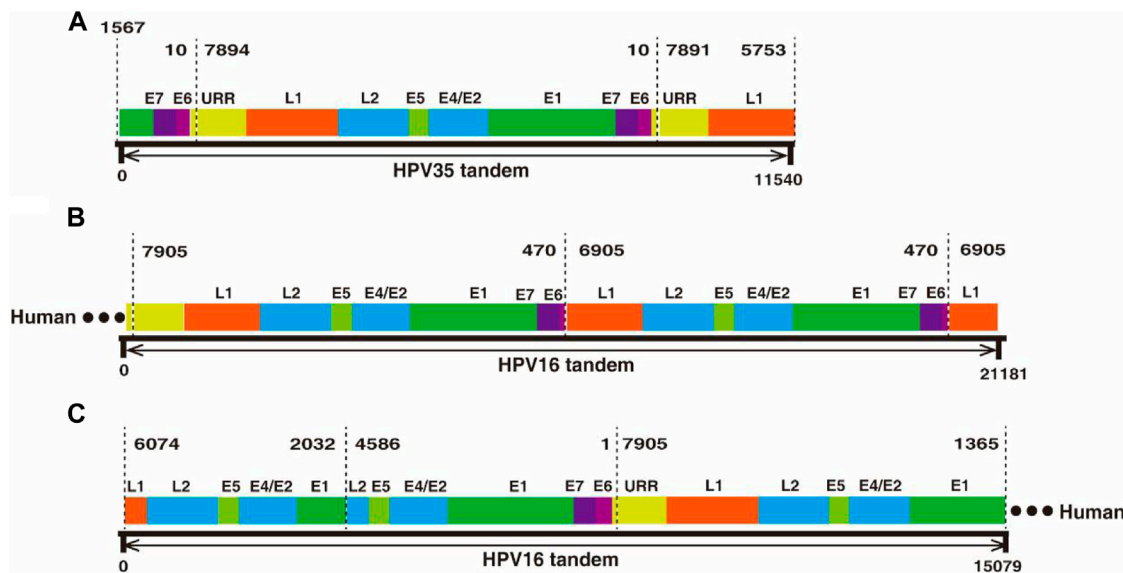


FIGURE 3 | Nanopore sequencing found tandem HPV genomes in cervical cancer tissues and CaSki cells, respectively. **(A)** An 11.54 kb-long HPV 35 tandem genomic sequence was obtained by nanopore sequencing. The 11.54 kb-long sequence was aligned to HPV35 16B (NCBI: KX514416.1), which consists of three genomic segments, E1-E7-E6 (1567–10 nt), a whole genome frame (7894–10 nt), and URR-L1 (1561–10 nt). **(B)** A 21.18 kb-long HPV tandem sequence flanked with human genomic sequence at one end was obtained by nanopore sequencing from CaSki cells. The sequence was aligned to HPV16 (NCBI: U89348), in which a truncated genome (470–7905 nt) was connected to another truncated one (470–6905 nt) in a head-to-tail manner. **(C)** Another 15.07 kb-long HPV tandem sequence was obtained by nanopore sequencing from CaSki cells and was flanked at one end by the human gene; concatemers are formed by joining of incomplete HPV genomes with two spliced sites at 2032 nt and 4,586 nt.

sequences, partial concatemeric amplicons, and chimeric cellular–viral sequences as shown in **Figure 3**. For example, an 11.54 kb-long HPV nucleotide sequence (**Supplementary Table S3**) obtained from the cervical cancer sample consists of a nearly intact HPV genome (7,894 bp) flanked by two adjacent ends of HPV genome (**Figure 3A**), which could be derived from the breakage of concatemeric amplicons during the rolling circle replication of HPV genome or from integrated tandem viral genomic sequences as reported in CaSki cells (Yee et al., 1985; Mincheva et al., 1987). Interestingly, in this study, 300 tandem viral genomic sequences flanked with cellular genes were found in CaSki cells. Representative features of the sequences are drawn as shown in **Figures 3B,C** and sequences are attached in **Supplementary Table S3**. Most of the tandem viral genomic sequences are composed of incomplete genomic units joined at 4 hot spliced sites (position 470 nt (E6), 6,905 nt (L1), 2032 nt (E1), and 4,586 nt (L2)), which is consistent with previous studies (Meissner, 1999; Xu et al., 2015). In total, 60 and 448 chimeric cellular-viral reads (**Supplementary Tables S1, S2**) were found in the cervical cancer tissue and CaSki cells, respectively, which would shed light on hot spots of HPV integration in human genome.

Determination of Physical Status of HPV Genome in a Cervical Cancer Tissue Sample and CaSki Cells

The discovery of concatemeric genomic sequences raises a question on whether HPV genome in cancer tissue exists in the episomal or integrated form. Exonuclease V digestion and

long-range PCR (Forsslund et al., 2019) were applied to address this question. Because exonuclease V specifically digests linear DNA but not circular DNA, circular HPV genome can be amplified with a pair of divergent primers within L1 gene (**Table 1**) from DNA treated with exonuclease V. The circular form of HPV genome was detected in cervical cancer tissue as shown in **Figure 4**, suggesting that the 11.54 kb-long concatemeric HPV sequence revealed by nanopore sequencing could be an intermediate product during the rolling circle replication of HPV genome. To further determine the physical status of the HPV genome in the cancer tissue, we tested the resistance percentage of E2 and E6 to exonuclease V digestion, followed by qPCR (Myers et al., 2019; Myers et al., 2020). Results showed that the average resistance of E2 and E6 to exonuclease V was in the range of 0.2–0.4 (**Figure 4; Table 2**), while the average resistance of β -actin to exonuclease V was almost 0, which suggests that integrated and episomal HPV genome coexist in the cervical cancer tissue. In contrast, in CaSki cells, the resistance percentages of E2, E6, and β -actin to exonuclease V digestion were almost 0 (**Table 2**), suggesting that only integrated forms of HPV genome exist. This result is consistent with that of a previous study (Myers et al., 2019).

IDENTIFICATION AND VERIFICATION OF HPV INTEGRATION SITES IN HUMAN GENOME

To further identify HPV integration sites, a large number of closely located integration sites with distances less than 200 bp

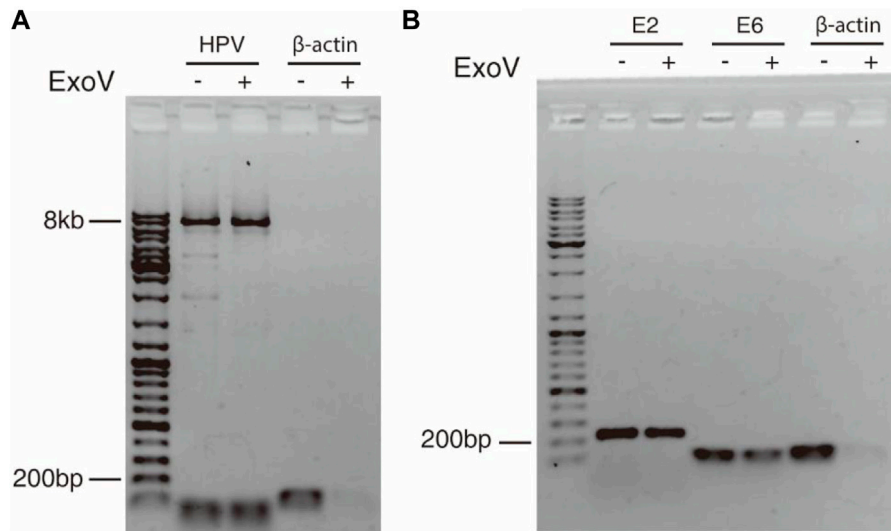


FIGURE 4 | Circular episomal HPV DNA was detected by exonuclease V digestion and PCR. **(A)** DNA extracted from cervical cancer tissue was digested with exonuclease V and the HPV genome was amplified by long-range PCR using a pair of divergent primers. **(B)** E2 and E6 were amplified with E2 and E6 primers from DNA treated with or without exonuclease V. β -actin served as a control.

TABLE 2 | Physical state determination of the HPV genome by exonuclease V (ExoV)-qPCR-based assay.

| Items | HPV35 in cervical cancer | | | HPV16 in CaSki | | |
|------------------------------|--------------------------|------|----------------|----------------|--------|----------------|
| Genes | E2 | E6 | β -actin | E2 | E6 | β -actin |
| Resistance to ExoV digestion | 0.25 | 0.34 | 0.003 | 0.0005 | 0.0006 | 0.001 |
| Physical state of HPV | Integrated and episomal | | | Integrated | | |

Note: a resistance rate of the detected genes equal to 1 indicates the HPV genome is exclusively episomal; a rate close to 0 indicates exclusively integrated; between 0–1, mixed.

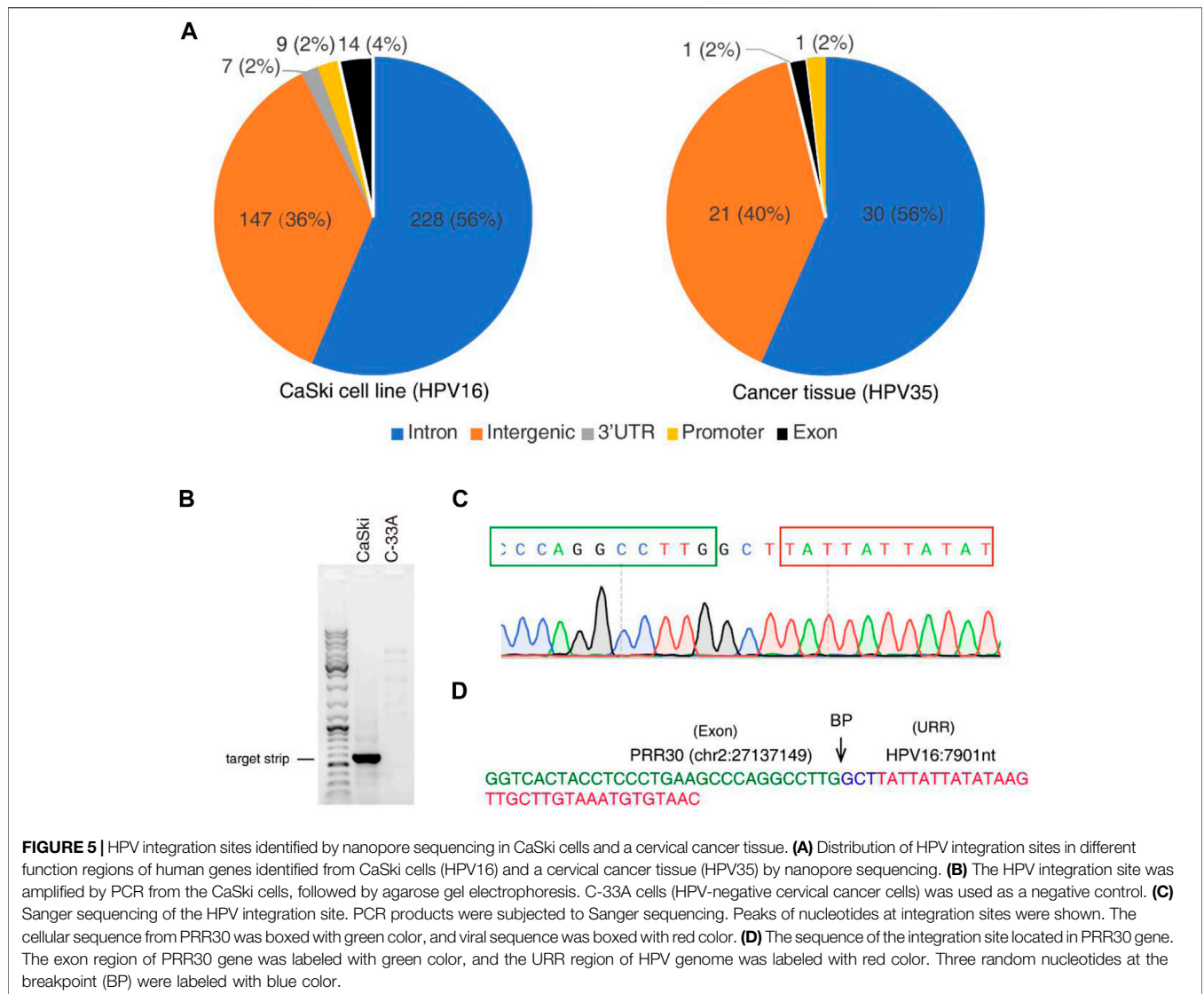
were considered the same, and 448 and 60 unique breakpoints were identified from the CaSki cell line and the cervical cancer tissue sample (Supplementary Tables S1, S2). HPV integration sites are mainly distributed in introns and intergenic regions as shown in Figure 5A, which is consistent with the results of previous studies (Koneva et al., 2018; Pinatti et al., 2018). Very few integration sites are located in exon, promoter, and 3'-UTR regions. For example, an integrated site at the exonic region of PRR30 in CaSki cells was revealed by 4 reads in nanopore sequencing data. To further verify this integration site, the integration region was amplified with PCR primers listed in Table 1, and PCR products (Figure 5B) were further sequenced by Sanger sequencing (Figures 5C,D). The HPV integration located in this exon is likely to affect the transcription and translation of PRR30, and further investigation is warranted.

DISCUSSION

Nanopore sequencing technology has attracted enormous attention since its release by Oxford Nanopore Technologies in 2014 due to its unique capability of generating ultra-long reads compared with secondary generation sequencing platforms. Nanopore sequencing has been applied for whole genome sequencing for humans (Jain et al., 2018), bacteria (Moss

et al., 2020), and viruses (Beaulaurier et al., 2020). However, the error rate of nanopore sequencing remains a big concern. In this study, by sequencing cervical cancer tissue with nanopore sequencing, 10 x coverage of human genome and 97-fold coverage of the HPV genome were achieved; a *de novo* assembled HPV 35 genome is 99.96% identical to HPV genome assembled by Sanger sequencing. In contrast, for CaSki cells, with 10 x coverage of human genome and 3,857 x coverage of HPV genome, the assembled HPV 16 genome is 99.99% identical to previously published reference genome. Although the low abundance of HPV positive cells in the cervical cancer tissue sample resulted in much lower coverage of HPV genome compared to CaSki cells, whole genome sequencing by nanopore sequencing can still generate HPV genome with high accuracy using our improved bioinformatic strategy.

Nanopore sequencing also demonstrated its advantage in detecting the physical status of the HPV genome in cervical cancer tissues. The level of HPV integration is positively correlated with the cervical intraepithelial neoplasia (CIN) grade and progression stages (Andersson et al., 2005; Briolat et al., 2007). Thus, identifying the episomal and integrated status of the HPV genome in cervical cancer will yield insight into HPV-induced cervical carcinogenesis. During viral replication, episomal circular genomic DNA co-exists with concatemeric



amplicons as HPV adopts a rolling circle mode of replication (Flores and Lambert, 1997). Integration events presumably occur following the breakup of circular genomes or concatemeric amplicons under certain circumstances. Integration of a single genome or concatemeric amplicon is referred to as type I or type II integration, respectively (Baker et al., 1987; McBride and Warburton, 2017). The integration mechanism is poorly understood. Through nanopore sequencing, we not only found integrated sites with high efficacy compared to the NGS platform (Hu et al., 2015) but also found different integrated patterns of HPV genome exhibited in long reads as shown in **Figure 3**. This unique feature will yield insight into the integration mechanism of HPV in cervical cancer cells and facilitate our understanding of the role of HPV in carcinogenesis.

Previous works have found that HPV integration is not a random event, preferring to specific chromosomal regions, including oncogenes or tumor suppressor genes (e.g., BCL2,

FANCC, HDAC2, RAD51B, and CSMD1) (Rusan et al., 2015; Groves and Coleman, 2018). Our study also revealed that one integration site (chr8:4915392-HPV35:6168) in the cervical cancer tissue sample is located in a previously identified tumor suppressor gene CSMD1 (Escudero-Esparza et al., 2016). Besides, when compared with known cancer genes in the COSMIC database (Tate et al., 2019), twenty-nine integration sites overlapped with records in COSMIC. For example, CSMD3 and ZFH3 are involved in ovarian cancer and endometrial carcinoma, respectively, which indicated that integration of HPV may lead to malfunctions of host oncogenes and tumor suppressors.

Currently, nanopore sequencing technology is widely used for sequencing DNA and RNA and determining the methylation status of DNA (Ni et al., 2019) and RNA (Lorenz et al., 2020). Our study demonstrated that the nanopore sequencing platform is a unique tool to assemble HPV genome and study the integration mechanism. Long reads simplify the genomic assembly algorithm

and further improve the accuracy of genomic assembly and integration analysis.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://ngdc.cncb.ac.cn/search/>, PRJCA006562; HRA001318, HRA001319.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the ethical review board of Shantou University Medical College (approval number: SUMC-2020-51). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SY and QZ contributed equally to experiments, bioinformatic analysis, and writing. LT, ZC, and ZW partially contributed to experiments. KL, RL, YC, and DO partially contributed to bioinformatic analysis. QQ, JX, and LZ are responsible for funding acquisition, experimental design, and writing. All

authors approved the final version of the manuscript before submission.

FUNDING

QQ was supported by grants from the Department of Education of Guangdong province (2020KZDZX1084), the Science and Technology Bureau of Shantou (200114115876753), and the National Natural Science Foundation of China (82072292). LZ was supported by a Cross-Disciplinary Research grant (2020LKSFG13B) from the Li Ka Shing Foundation and from the Science and Technology Bureau of Shantou, Guangdong, China (200110115871710).

ACKNOWLEDGMENTS

We would like to thank professor Stanley Lin at Shantou University Medical College for proofreading our manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.798608/full#supplementary-material>

REFERENCES

- Abu-Lubad, M. A., Jarajreh, D. A. A., Helaly, G. F., Alzoubi, H. M., Haddadin, W. J., Dabobash, M. D., et al. (2020). Human Papillomavirus as an Independent Risk Factor of Invasive Cervical and Endometrial Carcinomas in Jordan. *J. Infect. Public Health* 13 (4), 613–618. doi:10.1016/j.jiph.2019.08.017
- Aleman, L., Saunier, M., Alvarado-Cabrero, I., Quirós, B., Salmeron, J., Shin, H.-R., et al. (2015). Human Papillomavirus DNA Prevalence and Type Distribution in Anal Carcinomas Worldwide. *Int. J. Cancer* 136 (1), 98–107. doi:10.1002/ijc.28963
- Alkan, C., Sajjadian, S., and Eichler, E. E. (2011). Limitations of Next-Generation Genome Sequence Assembly. *Nat. Methods* 8 (1), 61–65. doi:10.1038/nmeth.1527
- Andersson, S., Safari, H., Mints, M., Lewensohn-Fuchs, I., Gyllensten, U., and Johansson, B. (2005). Type Distribution, Viral Load and Integration Status of High-Risk Human Papillomaviruses in Pre-stages of Cervical Cancer (CIN). *Br. J. Cancer* 92 (12), 2195–2200. doi:10.1038/sj.bjc.6602648
- Astbury, S., Costa Nunes Soares, M. M., Peprah, E., King, B., Jardim, A. C. G., Shimizu, J. F., et al. (2020). Nanopore Sequencing from Extraction-free Direct PCR of Dried Serum Spots for Portable Hepatitis B Virus Drug-Resistance Typing. *J. Clin. Virol.* 129, 104483. doi:10.1016/j.jcv.2020.104483
- Baker, C. C., Phelps, W. C., Lindgren, V., Braun, M. J., Gonda, M. A., and Howley, P. M. (1987). Structural and Transcriptional Analysis of Human Papillomavirus Type 16 Sequences in Cervical Carcinoma Cell Lines. *J. Virol.* 61 (4), 962–971. doi:10.1128/JVI.61.4.962-971.1987
- Beaulaurier, J., Luo, E., Eppley, J. M., Uyl, P. D., Dai, X., Burger, A., et al. (2020). Assembly-free Single-Molecule Sequencing Recovers Complete Virus Genomes from Natural Microbial Communities. *Genome Res.* 30 (3), 437–446. doi:10.1101/gr.251686.119
- Briolat, J., Dalstein, V., Saunier, M., Joseph, K., Caudroy, S., Prétet, J. L., et al. (2007). HPV Prevalence, Viral Load and Physical State of HPV-16 in Cervical Smears of Patients with Different Grades of CIN. *Int. J. Cancer* 121 (10), 2198–2204. doi:10.1002/ijc.22959
- Bull, R. A., Adikari, T. N., Ferguson, J. M., Hammond, J. M., Stevanovski, I., Beukers, A. G., et al. (2020). Analytical Validity of Nanopore Sequencing for Rapid SARS-CoV-2 Genome Analysis. *Nat. Commun.* 11 (1), 6272. doi:10.1038/s41467-020-20075-6
- Bzhalava, D., Eklund, C., and Dillner, J. (2015). International Standardization and Classification of Human Papillomavirus Types. *Virology* 476, 341–344. doi:10.1016/j.virol.2014.12.028
- Carmody, M. W., Jones, M., Tarraza, H., and Vary, C. P. H. (1996). Use of the Polymerase Chain Reaction to Specifically Amplify Integrated HPV-16 DNA by Virtue of its Linkage to Interspersed Repetitive DNA. *Mol. Cell Probes* 10 (2), 107–116. doi:10.1006/mcpr.1996.0015
- de Villiers, E. M. (2013). Cross-roads in the Classification of Papillomaviruses. *Virology* 445 (1–2), 2–10. doi:10.1016/j.virol.2013.04.023
- Escudero-Esparza, A., Bartoschek, M., Gialeli, C., Okroj, M., Owen, S., Jirström, K., et al. (2016). Complement Inhibitor CSMD1 Acts as Tumor Suppressor in Human Breast Cancer. *Oncotarget* 7 (47), 76920–76933. doi:10.18632/oncotarget.12729
- Flores, E. R., and Lambert, P. F. (1997). Evidence for a Switch in the Mode of Human Papillomavirus Type 16 DNA Replication during the Viral Life Cycle. *J. Virol.* 71 (10), 7167–7179. doi:10.1128/JVI.71.10.7167-7179.1997
- Forslund, O., Sugiyama, N., Wu, C., Ravi, N., Jin, Y., Swoboda, S., et al. (2019). A Novel Human *In Vitro* Papillomavirus Type 16 Positive Tonsil Cancer Cell Line with High Sensitivity to Radiation and Cisplatin. *BMC Cancer* 19 (1), 265. doi:10.1186/s12885-019-5469-8
- Groves, I. J., and Coleman, N. (2018). Human Papillomavirus Genome Integration in Squamous Carcinogenesis: what Have Next-Generation Sequencing Studies Taught Us? *J. Pathol.* 245 (1), 9–18. doi:10.1002/path.5058

- Groves, I. J., and Coleman, N. (2015). Pathogenesis of Human Papillomavirus-Associated Mucosal Disease. *J. Pathol.* 235 (4), 527–538. doi:10.1002/path.4496
- Guo, M., Sneege, N., Silva, E. G., Jan, Y. J., Cogdell, D. E., Lin, E., et al. (2007). Distribution and Viral Load of Eight Oncogenic Types of Human Papillomavirus (HPV) and HPV 16 Integration Status in Cervical Intraepithelial Neoplasia and Carcinoma. *Mod. Pathol.* 20 (2), 256–266. doi:10.1038/modpathol.3800737
- Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: Quality Assessment Tool for Genome Assemblies. *Bioinformatics* 29 (8), 1072–1075. doi:10.1093/bioinformatics/btt086
- Hellman, K., Silfversward, C., Nilsson, B., Hellstrom, A. C., Frankendal, B., and Pettersson, F. (2004). Primary Carcinoma of the Vagina: Factors Influencing the Age at Diagnosis. The Radiumhemmet Series 1956–96. *Int. J. Gynecol. Cancer* 14 (3), 491–501. doi:10.1111/j.1048-891x.2004.014310.x
- Hoenen, T. (2016). Sequencing of Ebola Virus Genomes Using Nanopore Technology. *Bio Protoc.* 6 (21). doi:10.21769/BioProtoc.1998
- Hoppe-Seyler, K., Bossler, F., Braun, J. A., Herrmann, A. L., and Hoppe-Seyler, F. (2018). The HPV E6/E7 Oncogenes: Key Factors for Viral Carcinogenesis and Therapeutic Targets. *Trends Microbiology* 26 (2), 158–168. doi:10.1016/j.tim.2017.07.007
- Houldsworth, J. (2014). FFACT: the FISH-Based HPV-Associated Cancer Test that Detects Nonrandom Gain at Four Genomic Loci as Biomarkers of Disease Progression. *Expert Rev. Mol. Diagn.* 14 (8), 921–934. doi:10.1586/14737159.2014.965685
- Hu, Z., Zhu, D., Wang, W., Li, W., Jia, W., Zeng, X., et al. (2015). Genome-wide Profiling of HPV Integration in Cervical Cancer Identifies Clustered Genomic Hot Spots and a Potential Microhomology-Mediated Integration Mechanism. *Nat. Genet.* 47 (2), 158–163. doi:10.1038/ng.3178
- Hudelist, G., Manavi, M., Pischinger, K. L., Watkins-Riedel, T., Singer, C. F., Kubista, E., et al. (2004). Physical State and Expression of HPV DNA in Benign and Dysplastic Cervical Tissue: Different Levels of Viral Integration Are Correlated with Lesion Grade. *Gynecol. Oncol.* 92 (3), 873–880. doi:10.1016/j.ygyno.2003.11.035
- Jain, M., Fiddes, I. T., Miga, K. H., Olsen, H. E., Paten, B., and Akeson, M. (2015). Improved Data Analysis for the MinION Nanopore Sequencer. *Nat. Methods* 12 (4), 351–356. doi:10.1038/nmeth.3290
- Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., et al. (2018). Nanopore Sequencing and Assembly of a Human Genome with Ultra-long Reads. *Nat. Biotechnol.* 36 (4), 338–345. doi:10.1038/nbt.4060
- Kamal, M., Lameiras, S., Deloger, M., Morel, A., Vacher, S., Lecerf, C., et al. (2021). Human Papilloma Virus (HPV) Integration Signature in Cervical Cancer: Identification of MACROD2 Gene as HPV Hot Spot Integration Site. *Br. J. Cancer* 124 (4), 777–785. doi:10.1038/s41416-020-01153-4
- Kent, W. J. (2002). BLAT—the BLAST-like Alignment Tool. *Genome Res.* 12 (4), 656–664. doi:10.1101/gr.229202
- Kiseleva, V. I., Krikunova, L. I., Mkrtchian, L. S., Liubina, L. V., Beziaeva, G. P., Panarina, L. V., et al. (2013). The Significance of Physical Status of Human Papillomavirus Type 16 for Predicting the Effectiveness of Invasive Cervical Cancer Treatment. *Vopr Onkol* 59 (6), 756–760.
- Klaes, R., Woerner, S. M., Ridder, R., Wentzensen, N., Duerst, M., Schneider, A., et al. (1999). Detection of High-Risk Cervical Intraepithelial Neoplasia and Cervical Cancer by Amplification of Transcripts Derived from Integrated Papillomavirus Oncogenes. *Cancer Res.* 59 (24), 6132–6136.
- Koneva, L. A., Zhang, Y., Virani, S., Hall, P. B., McHugh, J. B., Chepeha, D. B., et al. (2018). HPV Integration in HNSCC Correlates with Survival Outcomes, Immune Response Signatures, and Candidate Drivers. *Mol. Cancer Res.* 16 (1), 90–102. doi:10.1158/1541-7786.mcr-17-0153
- Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., and Phillippy, A. M. (2017). Canu: Scalable and Accurate Long-Read Assembly via Adaptive K-mer Weighting and Repeat Separation. *Genome Res.* 27 (5), 722–736. doi:10.1101/gr.215087.116
- Li, H. (2018). Minimap2: Pairwise Alignment for Nucleotide Sequences. *Bioinformatics* 34 (18), 3094–3100. doi:10.1093/bioinformatics/bty191
- Lin, J., Tang, C., Wei, H.-C., Du, B., Chen, C., Wang, M., et al. (2021). Genomic Monitoring of SARS-CoV-2 Uncovers an Nsp1 Deletion Variant that Modulates Type I Interferon Response. *Cell host & microbe* 29 (3), 489–502. e8. doi:10.1016/j.chom.2021.01.015
- Lorenz, D. A., Sathe, S., Einstein, J. M., and Yeo, G. W. (2020). Direct RNA Sequencing Enables m6A Detection in Endogenous Transcript Isoforms at Base-specific Resolution. *Rna* 26 (1), 19–28. doi:10.1261/rna.072785.119
- Lu, H., Giordano, F., and Ning, Z. (2016). Oxford Nanopore MinION Sequencing and Genome Assembly. *Genomics, Proteomics & Bioinformatics* 14, 265–279. (Electronic). doi:10.1016/j.gpb.2016.05.004
- Luft, F., Klaes, R., Nees, M., Dürst, M., Heilmann, V., Melsheimer, P., et al. (2001). Detection of Integrated Papillomavirus Sequences by Ligation-Mediated PCR (DIPS-PCR) and Molecular Characterization in Cervical Cancer Cells. *Int. J. Cancer* 92 (1), 9–17. doi:10.1002/1097-0215(200102)9999:9999<aid-ijc1144>3.0.co;2-1
- Marçais, G., Delcher, A. L., Phillippy, A. M., Coston, R., Salzberg, S. L., and Zimin, A. (2018). MUMmer4: A Fast and Versatile Genome Alignment System. *PLoS Comput. Biol.* 14 (1), e1005944. doi:10.1371/journal.pcbi.1005944
- McBride, A. A., and Warburton, A. (2017). The Role of Integration in Oncogenic Progression of HPV-Associated Cancers. *Plos Pathog.* 13 (4), e1006211. doi:10.1371/journal.ppat.1006211
- McNaughton, A. L., Roberts, H. E., Bonsall, D., de Cesare, M., Mokaya, J., Lumley, S. F., et al. (2019). Illumina and Nanopore Methods for Whole Genome Sequencing of Hepatitis B Virus (HBV). *Sci. Rep.* 9 (3), 7081. doi:10.1038/s41598-019-43524-9
- Meissner, J. D. (1999). Nucleotide Sequences and Further Characterization of Human Papillomavirus DNA Present in the CaSki, SiHa and HeLa Cervical Carcinoma Cell Lines. *J. Gen. Virol.* 80 (7), 1725–1733. doi:10.1099/0022-1317-80-7-1725
- Meng, H.-X., Yang, X.-X., Liu, R.-Q., Bao, J.-J., Hou, Y.-J., Sun, J., et al. (2020). The Relationship between Human Papillomavirus, OFD1 and Primary Ciliogenesis in the Progression of Oropharyngeal Cancer: A Retrospective Cohort Study. *Pgpm* Vol. 13, 633–644. doi:10.2147/pgpm.s271735
- Mincheva, A., Gissmann, L., and zur Hausen, H. (1987). Chromosomal Integration Sites of Human Papillomavirus DNA in Three Cervical Cancer Cell Lines Mapped by *In Situ* Hybridization. *Med. Microbiol. Immunol.* 176 (5), 245–256. doi:10.1007/BF00190531
- Moss, E. L., Maghini, D. G., and Bhatt, A. S. (2020). Complete, Closed Bacterial Genomes from Microbiomes Using Nanopore Sequencing. *Nat. Biotechnol.* 38 (6), 701–707. doi:10.1038/s41587-020-0422-6
- Myers, J. E., Guidry, J. T., Scott, M. L., Zwolinska, K., Raikhy, G., Prasai, K., et al. (2019). Detecting Episomal or Integrated Human Papillomavirus 16 DNA Using an Exonuclease V-qPCR-Based Assay. *Virology* 537, 149–156. doi:10.1016/j.virol.2019.08.021
- Myers, J. E., Zwolinska, K., Sapp, M. J., and Scott, R. S. (2020). An Exonuclease V-qPCR Assay to Analyze the State of the Human Papillomavirus 16 Genome in Cell Lines and Tissues. *Curr. Protoc. Microbiol.* 59 (1), e119. doi:10.1002/cpmc.119
- Nambaru, L., Meenakumari, B., Swaminathan, R., and Rajkumar, T. (2009). Prognostic Significance of HPV Physical Status and Integration Sites in Cervical Cancer. *Asian Pac. J. Cancer Prev. : APJCP* 10 (3), 355–360.
- Ni, P., Huang, N., Zhang, Z., Wang, D. P., Liang, F., Miao, Y., et al. (2019). DeepSignal: Detecting DNA Methylation State from Nanopore Sequencing Reads Using Deep-Learning. *Bioinformatics* 35. (Oxford, England), 4586–4595. doi:10.1093/bioinformatics/btz276
- Niya, M. H. K., Kesheh, M. M., Keshtmand, G., Basi, A., Rezvani, H., Imanzade, F., et al. (2019). Integration Rates of Human Papilloma Virus Genome in a Molecular Survey on Cervical Specimens Among Iranian Patients. *Eur. J. Cancer Prev. : official J. Eur. Cancer Prev. Organisation (Ecp)* 28 (6), 537–543. doi:10.1097/cej.0000000000000498
- Olthof, N. C., Huebbers, C. U., Kolligs, J., Henfling, M., Ramaekers, F. C. S., Cornet, I., et al. (2015). Viral Load, Gene Expression and Mapping of Viral Integration Sites in HPV16-Associated HNSCC Cell Lines. *Int. J. Cancer* 136 (5), E207–E218. doi:10.1002/ijc.29112
- Park, J. S., Hwang, E. S., Park, S. N., Ahn, H. K., Um, S. J., Kim, C. J., et al. (1997). Physical Status and Expression of HPV Genes in Cervical Cancers. *Gynecol. Oncol.* 65 (1), 121–129. doi:10.1006/gyno.1996.4596
- Pett, M., and Coleman, N. (2007). Integration of High-Risk Human Papillomavirus: a Key Event in Cervical Carcinogenesis? *J. Pathol.* 212 (4), 356–367. doi:10.1002/path.2192
- Pinatti, L. M., Walline, H. M., and Carey, T. E. (2018). Human Papillomavirus Genome Integration and Head and Neck Cancer. *J. Dent Res.* 97 (6), 691–700. doi:10.1177/0022034517744213

- Quick, J., Grubaugh, N. D., Pullan, S. T., Claro, I. M., Smith, A. D., Gangavarapu, K., et al. (2017). Multiplex PCR Method for MinION and Illumina Sequencing of Zika and Other Virus Genomes Directly from Clinical Samples. *Nat. Protoc.* 12 (6), 1261–1276. doi:10.1038/nprot.2017.066
- Quick, J., Loman, N. J., Duraffour, S., Simpson, J. T., Severi, E., Cowley, L., et al. (2016). Real-time, Portable Genome Sequencing for Ebola Surveillance. *Nature* 530 (7589), 228–232. doi:10.1038/nature16996
- Rusan, M., Li, Y. Y., and Hammerman, P. S. (2015). Genomic Landscape of Human Papillomavirus-Associated Cancers. *Clin. Cancer Res.* 21 (9), 2009–2019. doi:10.1158/1078-0432.ccr-14-1101
- Sauvage, V., Boizeau, L., Candotti, D., Vandenberghe, M., Servant-Delmas, A., Caro, V., et al. (2018). Early MinION™ Nanopore Single-Molecule Sequencing Technology Enables the Characterization of Hepatitis B Virus Genetic Complexity in Clinical Samples. *PLoS one* 13 (3), e0194366. doi:10.1371/journal.pone.0194366
- Senol Cali, D., Kim, J. S., Ghose, S., Alkan, C., and Mutlu, O. (2018). Nanopore Sequencing Technology and Tools for Genome Assembly: Computational Analysis of the Current State, Bottlenecks and Future Directions. *Brief. Bioinform.* 20 (4), 1542–1559. doi:10.1093/bib/bby017
- Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., Bindal, N., et al. (2019). COSMIC: the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* 47 (D1), D941–D947. doi:10.1093/nar/gky1015
- Thorvaldsdóttir, T., Robinson, J. T., and Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): High-Performance Genomics Data Visualization and Exploration. *Brief. Bioinform.* 14 (2), 178–192. doi:10.1093/bib/bbs01
- Treangen, T. J., and Salzberg, S. L. (2011). Repetitive DNA and Next-Generation Sequencing: Computational Challenges and Solutions. *Nat. Rev. Genet.* 13 (1), 36–46. doi:10.1038/nrg3117
- Triglia, R. D. M., Metzke, K., Zeferino, L. C., and Andrade, L. A. L. D. A. (2009). HPV *In Situ* Hybridization Signal Patterns as a Marker for Cervical Intraepithelial Neoplasia Progression. *Gynecol. Oncol.* 112 (1), 114–118. doi:10.1016/j.ygyno.2008.09.047
- Walboomers, J. M., Jacobs, M. V., Manos, M. M., Bosch, F. X., Kummer, J. A., Shah, K. V., et al. (1999). Human Papillomavirus Is a Necessary Cause of Invasive Cervical Cancer Worldwide. *J. Pathol.* 189 (1), 12–19. doi:10.1002/(SICI)1096-9896(199909)189:1<12:AID-PATH431>3.0.CO;2-F
- Wang, L., Dai, S.-Z., Chu, H.-J., Cui, H.-F., and Xu, X.-Y. (2013). Integration Sites and Genotype Distributions of Human Papillomavirus in Cervical Intraepithelial Neoplasia. *Asian Pac. J. Cancer Prev.* 14 (6), 3837–3841. doi:10.7314/apjcp.2013.14.6.3837
- Warburton, A., Redmond, C. J., Dooley, K. E., Fu, H., Gillison, M. L., Akagi, K., et al. (2018). HPV Integration Hijacks and Multimerizes a Cellular Enhancer to Generate a Viral-Cellular Super-enhancer that Drives High Viral Oncogene Expression. *Plos Genet.* 14 (1), e1007179. doi:10.1371/journal.pgen.1007179
- Woodman, C. B., Collins, S. I., and Young, L. S. (2007). The Natural History of Cervical HPV Infection: Unresolved Issues. *Nat. Rev. Cancer* 7 (Suppl. 3), 11–22. doi:10.1038/nrc2050
- Xu, F., Cao, M., Shi, Q., Chen, H., Wang, Y., and Li, X. (2015). Integration of the Full-Length HPV16 Genome in Cervical Cancer and Caski and SiHa Cell Lines and the Possible Ways of HPV Integration. *Virus Genes* 50 (2), 210–220. doi:10.1007/s11262-014-1164-7
- Yee, C., Krishnan-Hewlett, I., Baker, C. C., Schlegel, R., and Howley, P. M. (1985). Presence and Expression of Human Papillomavirus Sequences in Human Cervical Carcinoma Cell Lines. *Am. J. Pathol.* 119 (3), 361–366.
- Yu, L., Majerciak, V., Xue, X.-Y., Uberoi, A., Lobanov, A., Chen, X., et al. (2021). Mouse Papillomavirus Type 1 (MmuPV1) DNA Is Frequently Integrated in Benign Tumors by Microhomology-Mediated End-Joining. *Plos Pathog.* 17 (8), e1009812. doi:10.1371/journal.ppat.1009812
- Yuan, Y., Cai, X., Shen, F., and Ma, F. (2021). HPV post-infection Microenvironment and Cervical Cancer. *Cancer Lett.* 497, 243–254. doi:10.1016/j.canlet.2020.10.034
- Zhang, R., Shen, C., Zhao, L., Wang, J., McCrae, M., Chen, X., et al. (2016). Dysregulation of Host Cellular Genes Targeted by Human Papillomavirus (HPV) Integration Contributes to HPV-Related Cervical Carcinogenesis. *Int. J. Cancer* 138 (5), 1163–1174. doi:10.1002/ijc.29872
- Zhu, X., Yan, S., Yuan, F., and Wan, S. (2020). The Applications of Nanopore Sequencing Technology in Pathogenic Microorganism Detection. *Can. J. Infect. Dis. Med. Microbiol.* 2020, 6675206, 2020. Journal canadien des maladies infectieuses et de la microbiologie medicale. doi:10.1155/2020/6675206
- Zur Hausen, H. (2002). Papillomaviruses and Cancer: from Basic Studies to Clinical Application. *Nat. Rev. Cancer* 2 (Suppl. 5), 342–350. doi:10.1038/nrc798

Conflict of Interest: The 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.

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.

Copyright © 2022 Yang, Zhao, Tang, Chen, Wu, Li, Lin, Chen, Ou, Zhou, Xu and Qin. 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.