Tumour Virus Research 20 (2025) 200325

Contents lists available at ScienceDirect

Tumour Virus Research

journal homepage: www.journals.elsevier.com/tumour-virus-research

Precise identification of viral–host integration events in HPV-positive cervical cancers by targeted long-read sequencing

Preetiparna Parida [a,1], Nivedita Mukherjee [b,1], Agastya Singh [b], Shirley Lewis [c], Krishna Sharan [c], Sandeep Mallya [d], Ashima Singh [e], Surya Sarathi Das [b], Mahadev Rao [f], Daniel S. Higginson [g], Radhakrishnan Sabarinathan [b,*], Rama Rao Damerla [a,**]

[a] Department of Medical Genetics, Kasturba Medical College Manipal, Manipal Academy of Higher Education (MAHE), Manipal, 576104, Karnataka, India
[b] National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bengaluru, India
[c] Department of Radiotherapy and Oncology, Manipal Comprehensive Cancer Care Centre, Kasturba Medical College Manipal, Manipal Academy of Higher Education (MAHE), Manipal, 576104, Karnataka, India
[d] Department of Bioinformatics, Manipal School of Life Sciences, Manipal Academy of Higher Education, Manipal, 576104, Karnataka, India
[e] Manipal School of Life Sciences, Manipal Academy of Higher Education, Manipal, 576104, Karnataka, India
[f] Department of Pharmacy Practice, Center for Translational Research, Manipal College of Pharmaceutical Sciences, Manipal Academy of Higher Education (MAHE), Manipal, 576104, Karnataka, India
[g] Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, NY, USA

ARTICLE INFO

Keywords:
HPV integration
Cervical cancer
Nanopore sequencing

ABSTRACT

Human papillomaviral (HPV) integrations into host human genome, a frequently observed event in HPV associated cervical cancer, are currently mapped through expensive Whole Genome sequencing (WGS) or RNA sequencing (RNA-seq) methodologies. This study aims to develop a targeted sequencing assay to determine HPV integrations in cervical tumors without the need for WGS or RNA-seq. We employed a library preparation strategy using tiled single primers that bind to HPV genome as a template and possibly extend HPV sequences into adjacent host human genomic sequences resulting in HPV and human chimeric sequences. Using this strategy, we sequenced known HPV integrations in HPV18 positive HeLa and HPV16 positive SiHa cell lines. We further used this method to detect HPV integration sites in four HPV-positive cervical cancer patients and confirmed these integration breakpoints by WGS and Sanger sequencing. Functional impact of HPV integrations was explored through differentially expressed genes within or near topologically associating domain (TAD) boundaries, possibly disrupted by respective integration events in these patients. We found *ZFP36L1, CPA3, CPB1* and *CXCL8* as some of the differentially expressed genes within disrupted TADs, which are known cancer associated genes. Our approach also reduced the cost of HPV integration detection by 90 % compared to WGS while also minimizing sequencing data volume. We believe that this method captures HPV integrations at significantly reduced costs and lesser sequencing data volume leading to better understanding of disease progression and monitoring cancer treatment.

1. Introduction

Human papillomavirus (HPV) infections are associated with a variety of head and neck and anogenital cancers [1]. Cervical cancer is the fourth most common cancer in women worldwide, which according to GLOBOCAN 2022, accounted for more than 662,301 new cases

* Corresponding author. National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bengaluru, 560065, Karnataka, India.
** Corresponding author. Department of Medical Genetics, Kasturba Medical College Manipal, Manipal Academy of Higher Education (MAHE), Manipal, 576104, Karnataka, India.
E-mail addresses: preetiparna.parida@learner.manipal.edu (P. Parida), niveditam@ncbs.res.in (N. Mukherjee), singhagastya00@gmail.com (A. Singh), shirley.salins@manipal.edu (S. Lewis), drkrishna.sharan@nitte.edu.in (K. Sharan), sandeep.mallya@manipal.edu (S. Mallya), ashima.singh3@learner.manipal.edu (A. Singh), suryasarathid@ncbs.res.in (S.S. Das), mahadev.rao@manipal.edu (M. Rao), higginsd@mskcc.org (D.S. Higginson), sabari@ncbs.res.in (R. Sabarinathan), rama.damerla@manipal.edu (R.R. Damerla).
[1] Contributed Equally.

https://doi.org/10.1016/j.tvr.2025.200325
Received 28 April 2025; Received in revised form 10 July 2025; Accepted 16 July 2025
Available online 17 July 2025

reported and 348,874 deaths [2,3]. Persistent HPV infection is the primary driver of more than 99 % of cervical cancers [4], 70 % of which are caused by high-risk HPV types 16 and 18 [5]. HPV is a non-enveloped double-stranded DNA virus belonging to the *Papillomaviridae* family. It infects the basal cells of the cervical epithelium and replicates as these cells divide and differentiate to form the upper epithelial layers, eventually being shed from the top layer as progeny virions that initiate reinfection [6]. HPV induces carcinogenesis by inhibiting the host tumor suppressor proteins, p53 and retinoblastoma protein (Rb) using the viral proteins- E6 and E7 respectively [7]. Integration of the HPV genome into the host chromosome is a common event observed in cervical cancers and is often associated with increased expression of viral oncogenes E6 and E7. While not essential for all cases of HPV-induced carcinogenesis, integration events frequently coincide with development of invasive cervical cancer. In the human genome, these integration sites are more likely to be enriched at common fragile sites and open chromatin regions [8–10]. Large-scale genomic analyses have identified hotspots of recurrent HPV integration with significant enrichment for sequences with microhomology between the human and HPV genomes at the breakpoints [11]. Tumors with HPV integration show upregulation of not only the viral oncogenes E6 and E7, but also dysregulation of host genes at or near the sites of integrations [12]. Recent studies have observed that genomic regions around HPV integrations are enriched with Gene Ontology (GO) terms for DNA repair, suggesting that integration events in the vicinity of DNA repair and tumor suppressor genes could accelerate genomic instability [13]. Moreover, HPV integration has been linked with changes in host chromatin structure and, consequently, in gene regulation, including long-range interactions [10,14,15].

Numerous such studies highlighting the cis and trans effects of HPV integration in inducing and promoting cervical carcinogenesis have underscored the importance of HPV integration detection in cancer research, leading to a surge in developing technologies to study HPV integrations [9]. The Cancer Genome Atlas characterized 228 primary cervical cancers and reported HPV integrations in all HPV18-positive samples and 76 % of HPV16-positive samples [10]. Campitelli et al. used these cell-viral junctions as a biomarker in circulating tumor DNA for analyzing a series of serum samples obtained from cervical cancer patients and reported that HPV integration can be used as a biomarker for the detection of minimal residual disease and subclinical relapse in HPV-associated cancers [16]. Therefore, detecting HPV integration status and identifying the integration locus are crucial steps not only for understanding the role of HPV in cervical carcinogenesis, but also for clinical diagnosis and treatment monitoring in cervical cancer patients [9].

Various techniques have been employed to determine HPV integration sites in the human genome, such as detection of integrated papillomavirus sequences by ligation-mediated PCR (DIPS-PCR) [17], Restriction Site PCR (RS-PCR) [18], amplification of papillomavirus oncogene transcripts (APOT) and next-generation sequencing (NGS). RNA-based assays for detection of HPV integration sites have substantial biological and technological constraints because of the lower stability of mRNA in biopsy samples [17]. However, the potential for discovering new integrations sites using assays like DIPS-PCR is limited [9]. NGS technologies like whole genome sequencing (WGS) or RNA sequencing (RNA-seq) have been widely used for the genome-wide characterization of HPV integrations in cervical cancer [19]. However, the method requires whole genome sequencing at high coverage (>30X) [20] and involves multiple steps [21]. Considering the complexity of the procedure and the resulting data [21], this method, though accurate, is not suitable for clinical use [9]. Further, the cost of reagents used for these assays can be a challenge for labs with less resource, space, and sample throughput [22]. With lower reagent costs and a shorter sequencing time, portable Nanopore sequencers may provide greater flexibility in this aspect [22].

Nanopore sequencing technology, a fourth generation of sequencing technology, is portable, provides real-time data and enables rapid sequencing in clinical settings [23]. In this study, we developed a novel enrichment strategy to capture HPV-human integration breakpoints using nanopore sequencing. We validated this technology by mapping known HPV integrations from cell lines and patient samples. Furthermore, by demonstrating expression changes of cancer-associated genes near the integration sites in the patient samples, we highlight the functional significance of mapping HPV integration events.

2. Materials and methods

2.1. Cell culture

Cervical cancer cell lines SiHa (HPV-16 positive), and HeLa (HPV-18 positive), procured from the National Centre for Cell Sciences (NCCS), Pune, India were cultured in Dulbecco's modified Eagle's medium (DMEM) (HiMedia Laboratories Pvt. Ltd., Mumbai, India) supplemented with 10 % fetal bovine serum (HiMedia Laboratories Pvt. Ltd., Mumbai, India) and 1 % antibiotic/antimycotic solution (HiMedia Laboratories Pvt. Ltd., Mumbai, India). Mycoplasma PCR Detection Kit (Abcam, Cambridge, UK) was used to test extracted DNA samples from the cell cultures for mycoplasma contamination, according to the manufacturer's instructions (Cell line authentication and Mycoplasma testing report are provided in Supplementary file 1).

2.2. Patient recruitment and sample collection

Ethical approval for the study was obtained from the Institutional Ethics Committee, Kasturba Medical College Manipal, Manipal Academy of Higher Education, Manipal, India. Fresh tumor biopsies, which were HPV positive, were obtained from four cervical cancer patients admitted to Kasturba Medical College, Manipal. These samples were collected after obtaining informed consent and prior to initiating cancer therapy. Patient demographics and clinical parameters were obtained from the medical records and summarized in Table 1. Biopsy samples were snap-frozen in liquid nitrogen or dry ice ethanol bath and stored at −80 °C until further use.

2.3. DNA isolation, HPV detection and genotyping

DNA was extracted from fresh frozen tumor biopsy samples using a DNeasy Blood and Tissue kit (Qiagen, Hilden, Germany) following the manufacturer's protocol. The DNA concentration of each sample was measured using Qubit® fluorometer 3.0 (Thermo Fisher Scientific Inc., USA) using the Qubit™ dsDNA BR Assay Kit. HPV detection by PCR was performed using HPV universal primers GP5+ and GP6+. PCR reaction comprised of 1x GoTaq® Green Master Mix (Promega Corp., WI, USA), 0.5 μM of GP5+ and GP6+ primers and 50 ng of genomic DNA. Thermal cycling was performed using SimpliAmp™ Thermal Cycler (Thermo Fisher Scientific Inc., USA) with the following conditions: Initial denaturation at 95 °C for 2 min, 95 °C for 30 s, 51 °C for 30 s, 72 °C for 30s, 35 cycles and final extension at 72 °C for 5 min for 1 cycle. The PCR products obtained using the GP5+ and GP6+ general primers were purified and subjected to Sanger sequencing. The resulting sequences were analyzed using NCBI BLAST (https://blast.ncbi.nlm.nih.gov/) to identify the specific HPV genotype. To further confirm the genotype, we performed an additional PCR using genotype-specific primers targeting the E6 gene regions of the identified HPV types (adapted from Damerla et al. [24]).

2.4. Whole genome sequencing (WGS) and analysis

Genomic DNA from tumor samples was sequenced using the Illumina Novaseq/NextSeq sequencer as per the manufacturer's instructions (Illumina, San Diego, California) without any specific enrichment for HPV DNA, with 2 × 150-bp paired-end reads and a minimum coverage

Table 1
Demographic details of patients and comparison of breakpoint obtained by Illumina WGS and Nanopore sequencing.

Sl No.	Patient ID	Age	Stage of Cancer	Histology	Integration in Human	HPV type & Break-points	Illumina WGS	Nanopore Sequencing	Sanger Sequencing
1	P1	42	IB3	SCC	Chr 14: 68500410 (RAD51B gene)	HPV 18 (L1)	Yes	Yes	Yes
2	P2	42	IIIC1	SCC	Chr 3: 148344900; 148384643 LINC02046 RP11-455G2.1	HPV 16 (L2,E6)	Yes	Yes	Yes
3	P3	42	IIB	SCC	Chr 4: 73657890; 73718402 (Intergenic)	HPV 16 (L2,E1)	Yes	Yes	Yes
4	P4	43	IIB	SCC	Chr2: 145528039; 145678400 (Intergenic) RP11-707K3.1	HPV 16 (E1, E2)	Yes	Yes	Yes

Footnote: SCC: Squamous Cell Carcinoma.

of approximately 30X. WGS was performed at Medgenome laboratories based in Bangalore, India. FASTQ files from the paired-end WGS were passed through quality control and adapter trimming using the Trim Galore wrapper over Cutadapt and FastQC [25]. FASTQ files were run through the NGSCheckMate program to verify that the tumor samples are not related to each other (https://github.com/parklab/NGSCheckMate). To detect viral integrations, the trimmed FASTQ files for each sample were processed through the ViFi algorithm using hg38 as the human reference genome, along with HPV genomes [26]. ViFi employs phylogenetic methods in conjunction with reference-based read mapping to accurately identify integration events, even for novel viral strains. This analysis revealed samples with viral integration, providing details on the viral strain, the human chromosome and positional range of integration, as well as the number and identities of reads supporting the integration. All integration sites identified from the WGS data were evidenced by discordant or supporting reads, as well as junction or split reads. Discordant reads include pairs in which one read maps to the human reference genome and the other to a viral reference genome, while junction reads are a type of chimeric read that spans the integration site, mapping partially to both the human and viral reference genomes. Junction read sequences were fetched from the BAM file outputs from ViFi and verified by visualizing the alignments in the Integrative Genomics Viewer (IGV). Precise integration points were then identified from the junction read sequences through BLAST analysis against the human reference genome and the detected viral strain.

2.5. Single primer PCR extension

We developed a novel single primer extension strategy for specifically targeting HPV fragments. The schematic of the single primer extension based nanopore library preparation is depicted in Fig. 1. This method for detecting HPV-human breakpoints involves using pools of primers which cover the entire HPV genome. Two sets of primer pools were designed to amplify specific HPV genomic regions in forward and reverse directions respectively. The elongation of a single primer can capture HPV fragments and adjacent human genomic sequences during primer extension. The homopolymer tailing (C-Tailing) step followed by single primer extension produces double-stranded DNA fragments and inhibits potential PCR-generated artefacts caused by errors in polymerase incorporation. This method can also detect novel integration breakpoints, unlike the traditional HPV detection PCRs which can only amplify a certain number of integration sites based on specific genes of HPV (E1, E2 OR L1) [17]. The final round of nested amplification using nanopore adapter-specific primers makes the assay more specific by enriching the amplicons obtained in the first round of primer extension. The use of nanopore specific adapters enables barcoding, and thereby sample multiplexing, which further reduces the cost of sequencing. The use of this library preparation technique for nanopore sequencing reduces the need for whole genome sequencing for detecting HPV-human integration events while increasing the efficacy of detecting HPV-human hybrid fragments even at very low copies.

2.5.1. Primer design

HPV16 and 18 whole genome references were obtained from the Papilloma Virus Episteme (PaVE) database (pave.niaid.nih.gov) [27]. Primers were designed using PRIMER3 software such that they tile the whole HPV genome. These primers are listed in Supplementary Table 1. Tiled primer design was accomplished with 24 primers (10 in forward primer pool and 14 in reverse primer pool) each spanning 500bp-1kb of HPV16 genome, and with 17 primers (9 in forward pool and 8 in reverse pool) each spanning 1–1.5 kb of HPV18 genome.

2.5.2. First round PCR extension

First round of single primer extension was performed using 50 pmoles of HPV-specific primers with nanopore-specific adapters for barcoding, 500 ng of genomic DNA extracted from the respective cell line or tumor samples, 10 µl 5x PrimeSTAR GXL Buffer, 4 µl dNTP mixture (2.5 mM each), 1 µl of PrimeSTAR GXL DNA Polymerase (1.25U/50 µl) (Takara Bio, Cat# R050A). The PCR extension reaction was used to extend single-stranded long reads containing HPV sequences and the adjacent human genomic sequence. Thermal cycling was performed using SimpliAmp™ Thermal Cycler (Thermo Fisher Scientific Inc., USA) with the following conditions: 98 °C for 10 s, annealing temperature (specific to the primer) for 30 s, and 68 °C for 6 min for 50 cycles. Primer extensions were then size selected using 0.8x ratio of High Prep™ PCR Clean-up Beads, retaining DNA fragments larger than 300bp (Magbio Genomics, USA) following the manufacturer's instructions and eluted in 25 µL of nuclease and protease-free molecular biology grade water (HiMedia Ltd, Mumbai).

2.5.3. C-tailing

The purified amplification product (100 ng) was initially incubated for 5 min at 95 °C in a water bath and chilled immediately on ice for 3 min. The nucleotide tailing reaction was then set up with 5 µl of 10X TdT buffer, 5 µl of 2.5 mM CoCl2, 1 µl of 100 pmol dCTP and 0.5 µl of terminal transferase (20 units/µl) (New England Biolabs), and the mixture was incubated for 1 h at 37 °C followed by heating for 10 min at 70 °C. The reaction mixture was purified with paramagnetic beads for retaining DNA fragments larger than 300bp and eluted in 10 µl of distilled water, and 5 µl was used for the second round PCR.

2.5.4. Second round PCR extension

The second round PCR included a polyG reverse primer with nanopore-specific adapters for barcoding, 5 µl of C-tailed products of the first round of single primer extension, 10 µl 5X PrimeSTAR GXL Buffer, 4 µl dNTP mixture (2.5 mM each), 1 µl of PrimeSTAR GXL DNA Polymerase (1.25U/50 µl) (Takara Bio, Cat# R050A). This PCR was performed under the following conditions: 98 °C for 30 s, annealing temperature (specific to the primer) for 30 s and 68 °C for 6–10 min for 40 cycles followed by a 10-min incubation at 68 °C. These PCR products were purified using High Prep™ PCR Clean-up System (Magbio Genomics), retaining DNA fragments larger than 300bp and eluted in 25 µl.

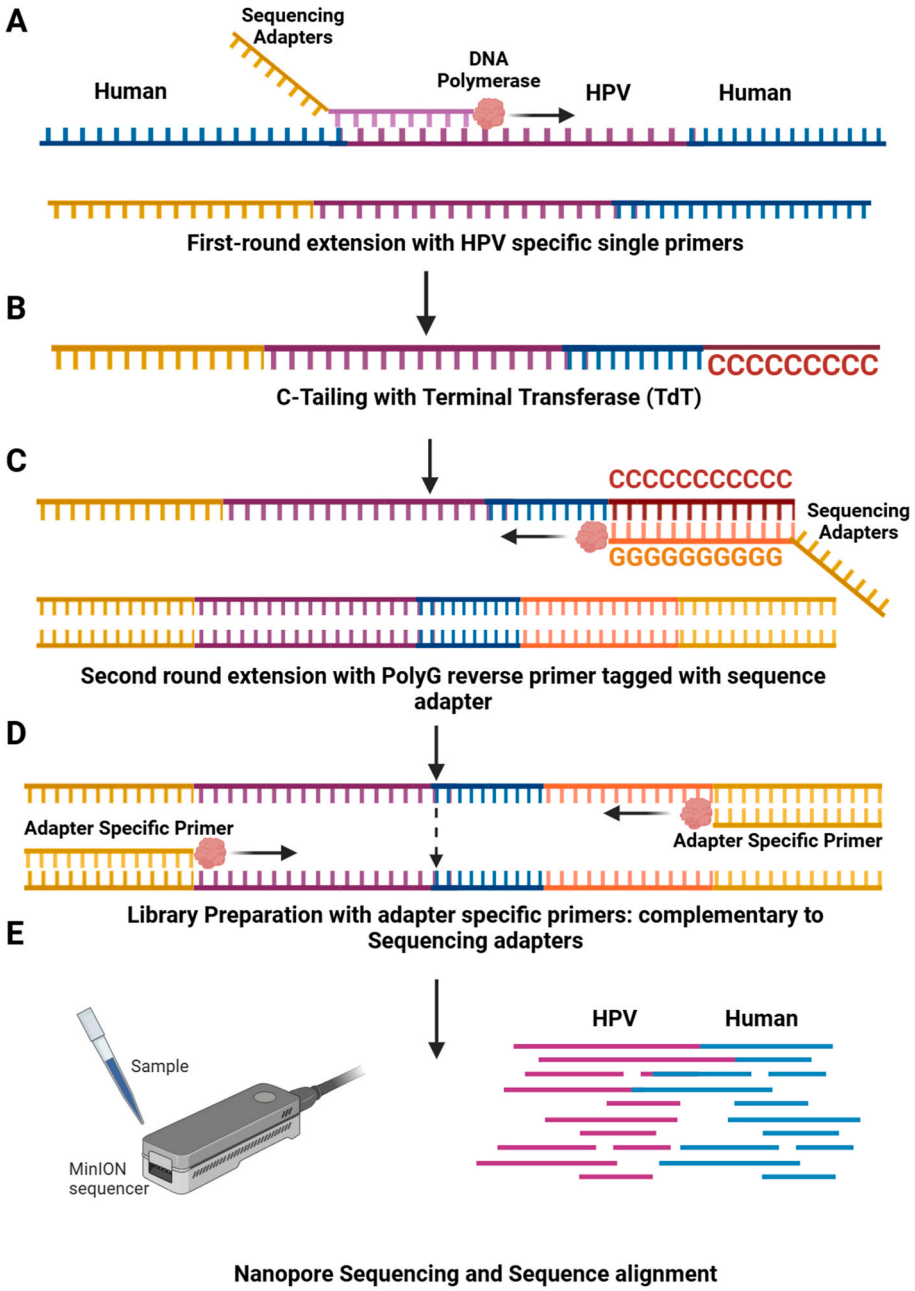

Fig. 1. Schematic representation of single primer extension method
(A) Denaturation and hybridization of a single HPV-specific primer to the HPV sequence in the integrated region was followed by single primer extension by a DNA polymerase; (B) HPV- specific single-stranded extended copies of the target DNA template molecules were then polyC-tailed using terminal transferase (TdT); (C) HPV-specific single primer-extended, polyC-tailed ssDNA molecules were then selectively amplified in second round extension with PolyG reverse primer tagged with sequencing adapter; (D) These products were used for library preparation for using adapter-specific primers: complementary to sequencing adapters. (E) Nanopore sequencing was performed and the sequences were analyzed for identifying HPV-human integration sites. Created in BioRender. Genetics, M. (2024) BioRender. com/l69i229.

2.5.5. Final PCR enrichment

The final round of PCR amplification included primers specific to adapter sequences, 10 µl purified product from the second round PCR, 10 µl 5x PrimeSTAR GXL Buffer, 4 µl dNTP mixture (2.5 mM each), 1 µl of PrimeSTAR GXL DNA Polymerase (1.25U/50 µl) (Takara Bio, Cat# R050A). This PCR was performed with the following conditions: 98 °C for 30 s, annealing temperature (specific to the primer) for 30 s, and 68 °C for 10 min for 40 cycles followed by a 10 min incubation at 68 °C. PCR products were purified using High Prep™ PCR Clean-up System, retaining DNA fragments larger than 300bp (Magbio Genomics) and eluted in 25 µl TE buffer. These purified products were processed for PCR barcoding for nanopore sequencing according to the manufacturer's instructions. The schematic representation of the single primer PCR extension workflow is shown in Fig. 1. The list of primers used in this assay is attached in Supplementary Table 1.

2.5.6. PCR barcoding

PCR barcoding was performed using the barcodes provided in the PCR Barcoding Expansion 1–12 kit (EXP-PBC001, Oxford Nanopore Technologies, Oxford, UK). One barcode was used per sample. The barcoding PCR reaction contained 2 µl PCR Barcode (one of BC01-BC12, at 10 µM), 500 ng of purified PCR amplification product, 10 µl of 5X

PrimeSTAR GXL Buffer, 4 µl of dNTP mixture (2.5 mM each), 1 µl of PrimeSTAR GXL DNA Polymerase (1.25U/50 µl) (Takara Bio, Cat# R050A) and nuclease-free water up to 50 µL. The cycling conditions used for barcoding PCR consisted of 35 cycles with Initial denaturation 1 cycle of 95 °C for 3 min, 35 cycles of denaturation at 95 °C for 30 s, annealing at 62 °C for 30 s, extension at 65 °C for 6–10 min, final extension 1 cycle at 65 °C for 10 min and hold at 4 °C. PCR barcoded products were purified using magnetic beads and the concentrations were measured using Qubit® fluorometer 3.0 (Thermo Fisher Scientific Inc.). These products were pooled in equimolar concentrations and 1 µg of pooled barcoded libraries was diluted in 47 µL of nuclease-free water for nanopore sequencing.

2.6. Nanopore sequencing

The barcoded library was end-prepped, ligated with adaptors, and cleaned up for sequencing using the ONT sequencing ligation kit SQK-LSK109 kit (Oxford Nanopore Technologies, UK). Qubit® fluorometer 3.0 (Thermo Fisher Scientific Inc., USA) was used to determine the concentration of the generated library. 50 fmol of the prepared library was loaded onto a ONT R9.3 MinION flow cell, after priming the flow cells with 800 µL of priming mix (30 µL Flush Tether to 1.17 mL of Flush Buffer). To prepare the library for loading, 11 µL (50 fmol) of the prepared library was mixed with 34 µL Sequencing Buffer, 25.5 µL premixed loading beads, and 4.5 µL nuclease-free water. 200 µL of priming mix was added to the priming port again while avoiding the introduction of air bubbles. Finally, 75 µL of the sample mix was added to the flow cell SpotON sample port of the R9.3 flow cells (Oxford Nanopore Technologies) on the MinION in a dropwise manner. The samples were sequenced on the MinION Sequencer for 3 h.

2.7. Viral integration detection using nanopore reads

MinKNOW software from the Oxford Nanopore Technologies was used to process the raw signals and considered only reads that passed the quality threshold ("PASS"). These reads were then aligned to a custom reference genome file using NGMLR-a long read aligner. NGMLR (https://github.com/philres/ngmlr) is designed to quickly and correctly align reads spanning complex structural variations and uses a convex gap cost model to compute precise alignments. It penalizes gap extensions for longer gaps less than for shorter ones thus accounting for large structural variants. The custom reference genome was created by concatenating hg38 and HPV reference genomes (HPV16 & HPV18). The aligned bam files (NGMLR output) were then used as input to Sniffles2 (https://github.com/fritzsedlazeck/Sniffles), a structural variation caller for long-reads. Sniffles2 outputs all types of detected structural variants (deletions, insertions, inversions and breakends). Only the breakends spanning human and viral contigs (with a minimum of five reads support) were considered viral integration points. The chimeric soft clip reads from the output.bam files were extracted for visualization. Circos plots were generated using the online tool shinyCircos (https://venyao.xyz/shinycircos/) [28].

2.8. Sanger validation

To validate the HPV integration sites in the human genome detected by nanopore sequencing, primers were designed, one against the human genome and the other against HPV near the potential site of integration. Both primers were designed 500 bp away from the integration sites detected from the nanopore sequencing results. The PCR reaction mix was prepared in a total volume of 25 µL containing 12.5 µL 2X GoTaq Green Master Mix, 0.5 µL of each forward and reverse primer (10 mM), and 10.5 µL nuclease-free water. 1 µL (30 ng) genomic DNA solution was used as a template. The PCR conditions were as follows: 5 min at 95 °C; 35 cycles of 30 s at 94 °C; 60 s at 50–60 °C for annealing; and 60 s at 72 °C; followed by 72 °C for 1 min. The PCR products were run on 1.5 % TAE (Tris base, acetic acid and EDTA) agarose gel, the bands were excised and purified using FavorPrep Gel/PCR purification mini kit (Favorgen Biotech. Corp., Taiwan, Japan). Sequencing of the purified PCR products was performed using a BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, USA). The list of primers used for validation is attached in Supplementary Table 2.

2.9. Functional validation of impact of HPV integrations on the neighboring genes

We explored the impact of HPV integrations on host gene expression in the context of topologically associating domains (TADs). For this, we selected target genes by shortlisting those with cancer-related functions in the IntOGen database [29] and existing literature. Promoter-enhancer interaction loop regions were obtained from the GeneHancer database [30]. TAD boundaries of HeLa and NHEK cell lines were obtained from existing literature [31]. All coordinates were obtained in hg38, or converted from hg19 to hg38 using UCSC LiftOver [32]. Total RNA was extracted from the four cervical cancer frozen tumor tissues and a control fibroblast cell line using TRIzol™ Reagent (Thermo Fisher, Carlsbad, CA, USA) according to the manufacturer's protocol. 1 µg of total RNA was converted into cDNA using an iScript™ gDNA Clear cDNA Synthesis Kit (BioRad, Hercules, CA, USA) according to the manufacturer's instructions. The reaction was incubated at 42 °C for 30 min, followed by 85 °C for 5 min to inactivate the reverse transcriptase. Primers were designed using primer3 software for qPCR. The list of primers used for qPCR is enlisted in Supplementary Table 3.

2.10. Reverse transcriptase-quantitative polymerase chain reaction analysis of selected genes

We performed qPCR analysis for the panel of selected genes. To assess whether gene expression changes were associated with HPV integration breakpoints and not due to general HPV infection, each patient was analyzed as a case, with the other three patients serving as controls. This intra-cohort comparative design ensured that differences in gene expression were linked to the integration event and not HPV infection. The $2^{-\Delta\Delta Ct}$ method was used for quantification and fold change for the target gene. RT-qPCR analysis was performed using TB Green Premix Ex Taq II (Tli RNase H Plus) (Takara Bio Inc., Tokyo, Japan; RR820A) on QuantStudio™ 5 (Applied Biosystems, Waltham, MA, USA). The PCR cycling conditions were as follows: 95 °C for 1 min followed by 40 cycles of 95 °C for 5 s, annealing temperature (specific to the different genes) for 30s, and 72 °C for 30s. The values were first normalized to the internal reference gene GAPDH, followed by calculating relative expression to healthy control. Statistical significance was inferred using *t*-test, with a p-value <0.05 considered statistically significant.

2.11. Sequencing cost analysis

The cost per sample was calculated by considering flow cell costs, sequencing kits, wash kits, PCR reagents, and terminal transferase. It also includes all other sample preparation and purification reagents before sequencing. WGS was performed at Medgenome laboratories, India; hence, the price charged per sample for sequencing was considered.

3. Results

3.1. Single primer extensions for detection of HPV integrations in HPV positive cell lines

We developed a single primer extension method to capture HPV-human breakpoints by designing primers tiling the genomes of the most commonly occurring high-risk HPV subtypes, HPV16 and HPV18.

The single primers allowed us to randomly extend the HPV genome allowing us to capture integrations provided the DNA sequence adjacent to the primer binding site is in the vicinity of an HPV-human integration junction. The reads obtained from the nanopore sequencing were in sizes ranging from 96 to 4656 base pairs (with a median size of 228). On an average, the polyC and polyG tracts observed were 7–8 base pairs long. To establish and validate this assay, we used genomic DNA from HPV-positive cervical cancer cell lines, HeLa and SiHa, to map chimeric viral-human breakpoints. Our analysis detected all the previously reported viral-human breakpoints in these cell lines (with >100 reads supporting the integration).

We could accurately identify integration sites on chromosome 8 of the HPV18-positive HeLa cell line as reported previously using the DIPS PCR technique Luft et al. [17,33] (Fig. 2A, Supplementary Fig. 1A). Similarly, using this method, we effectively detected the integration sites on chromosome 13 in the HPV16-positive SiHa cell line as reported by Meissner [34] and Yu et al. [33]. To validate our findings, we used the Sanger sequencing approach to ensure the accuracy of detection of viral breakpoints in these cell lines (Fig. 2B, Supplementary Fig. 1B).

3.2. Characterization of HPV integrations in tumor samples of cervical cancer patients

We used the single primer extension method for analyzing HPV integration sites in four HPV-positive cervical cancer patients (details summarized in Table 1). Patient P1 had an integration of the viral genome into intron 4 of the RAD51 paralog B (*RAD51B*) gene on chromosome 14. We could find only one viral breakpoint within this gene, and the integration resulted from the disruption of L1 gene of the virus. Patient P2, had an integration of the HPV genome in a lncRNA gene LINC02046 on chromosome 3. Patients P3 and P4 had integration sites in intergenic regions of chromosomes 4 and 2, respectively. WGS was performed for the four patients using Illumina to validate the results of our analysis. WGS analysis demonstrated concordance with nanopore sequencing data (Table 1). We further validated these integration sites by Sanger sequencing (Supplementary Fig. 2). The breakpoints obtained for the four patients through Illumina WGS and Nanopore sequencing are mentioned in Table 1 and Supplementary Fig. 2.

HPV integration has been shown to disrupt host transcriptional activity by dysregulating expression of nearby genes through chromatin remodeling [35]. This transcriptional disruption is mostly confined to genes present within the same TAD (Topologically Associated Domain)–functional units of 3D genome organization that harbor genes and regulatory regions and spatially confine DNA-DNA interactions [10]. To evaluate the influence of HPV integration on transcriptional activity in the four cervical cancer patients in our cohort, we shortlisted target genes located within the same TAD as the integration sites (Supplementary Figs. 3–6). For patient P1, genes located within the same TAD were selected, including *RAD51B*, which overlaps with the integration site, and *ZFP36L1*. *ZFP36L1* was downregulated, while no significant difference in the expression of *RAD51B* was observed (Fig. 3A). In patient P2, HPV integration occurred near the 5′ TAD boundary, prompting the selection of genes near the 3′ TAD boundary which included *CPB1* (within the TAD near its 3′ boundary) and *CPA3* (outside the TAD but near its 3' boundary). Both *CPA3* and *CPB1* were significantly upregulated in P2 (Fig. 3B). For patient P3, genes *CXCL8*, *RASSF6*, *ANKRD17-DT* and *MTHFD2L* were selected since the integration site fell within their promoter-enhancer integration loops. Additionally, the cancer-associated gene *ALB*, located within the same TAD, was also selected. Only *CXCL8* among all the other genes was upregulated (Fig. 3C and D). No target genes were selected for P4, as the integration site was located outside any TAD, thus being devoid of genes or regulatory regions.

3.3. Sequencing cost

The cost per sample for WGS was $598.74 while the final cost of sequencing per sample on MinION using our novel library preparation technique was $27. The final cost was calculated assuming 10 samples per run and did not include labor charges. For calculating the per-run cost for purification, we assumed each sample volume to be 50 µL, and in each run, the magnetic bead purification was performed five times. Each flow cell was washed and reused (4 runs per flow cell). When we included the reagent costs for single primer assay along with nanopore sequencing, our cost per sample was $54.6. Detailed price analysis is included in Table 2.

4. Discussion

Nanopore has emerged as an important next generation sequencing tool since its release in 2014 [20]. In this study, we developed a novel targeted sequencing approach which combines single primer extension enrichment followed by nanopore sequencing to precisely identify HPV integration sites in HPV-positive cell lines and tumor genomic DNA from cervical cancer patient samples. Whole genome sequencing on the Nanopore sequencing platform has been previously used for studying

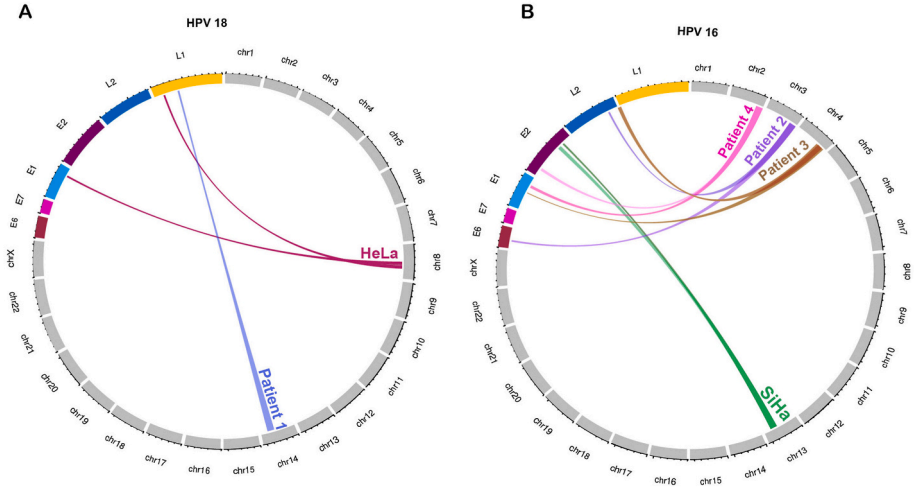

Fig. 2. Circos plot illustrating chimeras between the HPV genome (specific gene) and specific human chromosomes (A) Illustrating chimera of HPV 18 with HeLa cell line (in Chromosome 8) and Patient P1 (in Chromosome 14); (B) Illustrating chimera of HPV 16 with SiHa cell line (in Chromosome 13), Patient P2 (in Chromosome 3), Patient P3 (in Chromosome 4) and Patient P4 (in Chromosome 2).

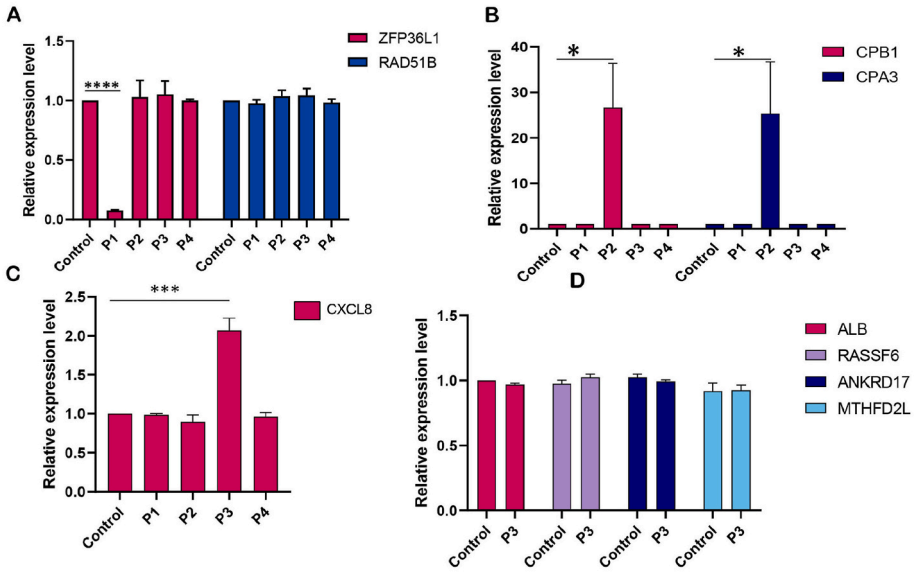

Fig. 3. Relative expression of selected target genes in all patients to evaluate the impact of HPV integration on neighboring genes (A) Relative gene expression of *ZFP36L1* and *RAD51B;* (B) Relative gene expression of *CPB1* and *CPA3;* (C) Relative gene expression of *CXCL8;* (D) Relative gene expression of *ALB, RASSF6, ANKRD17* and *MTHFD2L*.

Table 2
Cost analysis for MinION Nanopore sequencing using PCR enrichment method.

Nanopore Sequencing	Cost (in US$)	Total Units	Cost per run (in US $)
PCR	148.9	250	0.6
C-tailing (Terminal Transferase NEB)	213.65	500	0.4
Magnetic beads for purification HighPrep PCR	123.77	125 (40µl per reaction)	5.0 (5 times per run)
Minion Flowcell R9.4.1 FLO-MIN106D	1081.16	4 runs per flow cell	270.3
Ligation Sequencing and adapter ligation (SQK-LSK109 kit)	907.46	6	151.2
Barcoding (PCR Barcoding Expansion 1–12)	454.74	6	75.8
Flowcell Priming (EXP-FLP002)	155.73	6	26.0
Flowcell Wash (EXP-WSH004)	99	6	16.5
Total Cost			545.8
Cost per sample			**54.6**

viral integration due to its ability to generate long reads, enabling comprehensive genome-wide analyses [9]. Increasing the depth of sequencing by enriching the viral genome using PCR can help in attaining improved accuracy [23]. The findings of our study highlight the effectiveness of our unique targeted nanopore-based amplicon sequencing in accurately identifying integration sites within known chromosomal regions of HPV-positive cell lines (HeLa in chr8 and SiHa in chr13). Our results confirmed previously identified integration breakpoints in these cell lines [17,18].

Single Primer extension methods have been previously used for long read walking of hexamerin gene of the mosquito (*Anopheles gambiae*), *hairy* gene in *Drosophila* [36] and genotyping of ancient DNA (aDNA) [37]. We evaluated the utility of our single primer based targeted nanopore sequencing by comparing our results with Illumina WGS. Our analysis of four cervical cancer patient samples reported concordance with Illumina WGS.

Underscoring the functional and clinical relevance of detecting HPV integration sites, we found that several cancer-associated genes near the sites identified in our cervical cancer patients exhibited dysregulated expression. One patient (Patient P1) had an integration within intron 4 of the RAD51B gene located on chromosome 14. RAD51B is an important DNA double-strand break repair protein, and its loss of function is reported in uterine leiomyoma and breast cancer [38]. It has also been previously reported to be a hotspot for HPV integration [15, 38] which may lead to dysregulation of Homologous Recombination Repair (HRR), causing genomic instability—a hallmark of cancer [39]. We also observed that this integration led to the downregulation of a neighboring gene within the same TAD, Zinc finger protein 36 (*ZFP36L1*), which acts as a tumor suppressor. *ZFP36L1* is often down-regulated in several patient cohorts of bladder and breast cancers and its reduced expression is associated with worse survival in patients with breast cancer [40]. Loss of *ZFP36L1* has also been reported to promote epithelial-mesenchymal transition in hepatocellular carcinoma [41].

We also found an HPV integration site in a lncRNA gene LINC02046 in Patient P2. The sequencing data for the other two patients revealed integrations predominantly within intergenic regions, which is consistent with patterns reported in previous literature [11,33,42]. This integration shared the TAD boundaries with two genes, *CPB1* and *CPA3*, which were found to be upregulated in the patient P2. It has been reported that the overexpression of *CPB1* in ductal carcinoma in situ (DCIS), which is an early-stage breast cancer, can lead to the progression into invasive breast cancer in patients [43]. Also, gene *CPA3*, currently known as *CPA4* [44] is associated with pancreatic cancer progression and was observed to be overexpressed in pancreatic cancer patients when compared to healthy controls [45].

Integration sites in the HPV-positive cell lines HeLa and SiHa were also observed in the intergenic regions of Chr 8 (Chr8q24.21) [11,33] and Chr13 (between the KLF5 and LINC00392 genes) [42]. As hypothesized by Yang et al., intergenic HPV integrations might serve as a defense mechanism of the virus which keeps the host cell viable, as integration into human non-exonic regions would probably not cause a significant loss of function of any important gene [9]. The integration breakpoint in patient P3 fell within the promotor-enhancer integration loops of genes in that TAD including *CXCL8, RASSF6, ANKRD17* and *MTHFD2L*. *CXCL8* was overexpressed in this patient compared to the other three patients and control. None of the other genes, nor ALB, a cancer-associated gene present in the same TAD, showed significant changes in the expression. *CXCL8* is a known chemokine which plays a major role in the proliferation, invasion, and migration of cancer [46]. This gene is also reported to promote angiogenesis in breast cancer [47] and its over-expression is associated with an increased risk of cancer and

poorer prognosis in patients with colorectal cancer [48].

We believe our method has many advantages over both Illumina and Nanopore WGS. Our targeted sequencing approach significantly reduces both the cost (by over 90 %) and data volume required for the detection of HPV integrations compared to WGS. With our novel enrichment method, we were also able to analyze specific regions of interest (HPV-human Integration breakpoint) which helps avoid false positive interpretations of sequencing data. This strategic combination not only increased our sensitivity in detecting integration breakpoints but also substantially reduced background noise, enabling us to pinpoint specific integration loci within the genome.

The accurate identification and characterization of HPV integration breakpoints hold immense clinical implications, especially in understanding the molecular mechanisms underlying HPV-associated carcinogenesis [9]. Our study highlights the potential of nanopore sequencing, complemented by the single primer extension enrichment method, as a useful tool for this purpose. This method may be particularly valuable in individuals facing therapy resistance, where understanding HPV integration status and its impact on differentially regulating genes surrounding the integration sites in the host genome, could provide insights into treatment response.

Besides the reduced cost and amount of sequencing data, a simpler analysis pipeline makes HPV integration detection faster, more efficient, and easily adoptable, which is crucial for clinical applications. One of the current limitations of our study is that we have not worked on creating pipelines to reliably detect episomal and integrated forms of HPV in the same sample, and also the presence of concatenated form in HPV genomes at the sites of integration. Furthermore, we have not yet applied this approach to precancerous lesions, which limits our ability to assess its potential utility in early detection. Subsequently, once the assay is standardized and validated for use in precancerous lesions, it holds potential to identify patients at higher risk of progression to cervical cancer, thereby contributing to early detection and improved clinical management. Future studies will focus on refining the analytical framework and expanding the method's application to earlier stages of disease progression. Larger cohorts may be explored in future studies to identify new integration sites, potentially paving the way for the discovery of prognostic biomarkers or targeted therapeutic approaches for HPV-associated cancers. Our method demonstrates the superiority of targeted nanopore sequencing in identifying HPV integrations compared to Illumina WGS, emphasizing its precision, comprehensive coverage, and compatibility with a novel enrichment method.

5. Conclusion

We developed a quick and efficient targeted sequencing method on nanopore to characterize HPV integrations in cervical cancer patient samples. We believe that this method will contribute significantly to the understanding of the role of HPV integrations in cervical carcinogenesis by accurately identifying integration breakpoints. Our novel technique could be easily adapted to studying other viral integrations like retroviral, Adenovirus-associated and Human Herpes viral integration events. In addition to mapping viral integrations, this method could also be a crucial and cost-effective solution for detecting gene fusions in several cancers. Larger cohort studies are further required to explore the clinical significance of this method and translate it to diagnostics or companion diagnostic solutions.

CRediT authorship contribution statement

Preetiparna Parida: Writing – review & editing, Writing – original draft, Methodology, Investigation, Conceptualization. **Nivedita Mukherjee:** Writing – review & editing, Writing – original draft, Visualization, Investigation, Data curation. **Agastya Singh:** Writing – review & editing, Writing – original draft, Visualization, Investigation, Data curation. **Shirley Lewis:** Resources. **Krishna Sharan:** Resources. **Sandeep Mallya:** Data curation. **Ashima Singh:** Investigation. **Surya Sarathi Das:** Writing – review & editing, Formal analysis. **Mahadev Rao:** Writing – review & editing. **Daniel S. Higginson:** Writing – review & editing, Conceptualization. **Radhakrishnan Sabarinathan:** Writing – review & editing, Writing – original draft, Supervision, Funding acquisition, Conceptualization. **Rama Rao Damerla:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Funding acquisition, Conceptualization.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request. Raw sequencing files for cell lines are available in the supplementary file 1.

Ethics approval and consent to participate

Ethical approval for the study was obtained from the Institutional Ethical Committee, Manipal Academy of Higher Education (IEC No: 774/2019). This study was also registered as an observational study with Clinical Trials Registry – India with the number CTRI/2020/01/022862.

Consent for publication

All authors have consented to submit this article for publication.

Funding

This study was supported by the Department of Biotechnology, Government of India, Ramalingaswami Fellowship (BT/RLF/Re-entry/21/2018) and Lady Tata Memorial Trusts Institutional Research Grant (application number: 3680) given to RD. RS would like to acknowledge funding support from the NCBS-TIFR and the DBT/Wellcome Trust India Alliance Fellowship [grant number IA/I/20/1/504928].

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Rama Rao Damerla has patent #Method for sequencing integration and transposition events in a genome. Indian Patent Office. Application number 202341083889 pending to Manipal Academy of Higher Education. Preetiparna Parida has patent #Method for sequencing integration and transposition events in a genome. Indian Patent Office. Application number 202341083889 pending to Manipal Academy of Higher Education. Sandeep Mallya has patent #Method for sequencing integration and transposition events in a genome. Indian Patent Office. Application number 202341083889 pending to Manipal Academy of Higher Education. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

PP would like to acknowledge the support from Senior Research Fellowship (SRF-Direct) (CSIRAWARD/SRF-DIRECT2024/15079) awarded by Council for Scientific and Industrial Research, Ministry of Science & Technology, Government of India. NM would like to acknowledge support from NCBS-TIFR and the Shyama Prasad Mukherjee Fellowship (SPMF) awarded by Council for Scientific and Industrial Research, Ministry of Science & Technology, Government of India. RD would like to acknowledge faculty seed money support from Manipal Academy of Higher Education, Manipal, Karnataka, India.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tvr.2025.200325.

Data availability

Data will be made available on request.

References

[1] M.-N. Didelot, N. Boulle, A. Damay, V. Costes, M. Segondy, Comparison of the PapilloCheck® assay with the digene HC2 HPV DNA assay for the detection of 13 high-risk human papillomaviruses in cervical and anal scrapes, J. Med. Virol. 83 (8) (2011) 1377–1382, https://doi.org/10.1002/jmv.22148.

[2] Cancer today. https://gco.iarc.who.int/today/. (Accessed 25 September 2024).

[3] F. Bray, et al., Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J. Clin. 74 (3) (2024) 229–263, https://doi.org/10.3322/caac.21834.

[4] Cervical cancer. https://www.who.int/health-topics/cervical-cancer. (Accessed 25 September 2024).

[5] Cervical cancer causes, risk factors, and prevention - NCI. https://www.cancer.gov/types/cervical/causes-risk-prevention. (Accessed 25 September 2024).

[6] C.B.J. Woodman, S.I. Collins, L.S. Young, The natural history of cervical HPV infection: unresolved issues, Nat. Rev. Cancer 7 (1) (Jan. 2007) 11–22, https://doi.org/10.1038/nrc2050.

[7] Y. Zhang, K. Qiu, J. Ren, Y. Zhao, P. Cheng, Roles of human papillomavirus in cancers: oncogenic mechanisms and clinical use, Signal Transduct. Targeted Ther. 10 (1) (Jan. 2025) 44, https://doi.org/10.1038/s41392-024-02083-w.

[8] Y. Liu, Z. Lu, R. Xu, Y. Ke, Comprehensive mapping of the human papillomavirus (HPV) DNA integration sites in cervical carcinomas by HPV capture technology, Oncotarget 7 (5) (Dec. 2015) 5852–5864, https://doi.org/10.18632/oncotarget.6809.

[9] W. Yang, et al., Accurate detection of HPV integration sites in cervical cancer samples using the nanopore MinION sequencer without error correction, Front. Genet. 11 (2020) [Online]. Available: https://www.frontiersin.org/articles/10.3389/fgene.2020.00660. (Accessed 10 May 2023).

[10] A.K. Singh, K. Walavalkar, D. Tavernari, G. Ciriello, D. Notani, R. Sabarinathan, Cis-regulatory effect of HPV integration is constrained by host chromatin architecture in cervical cancers, Mol. Oncol. 18 (5) (May 2024) 1189–1208, https://doi.org/10.1002/1878-0261.13559.

[11] Z. Hu, et al., Genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism, Nat. Genet. 47 (2) (Feb. 2015) 158–163, https://doi.org/10.1038/ng.3178.

[12] C.A. Moody, L.A. Laimins, Human papillomavirus oncoproteins: pathways to transformation, Nat. Rev. Cancer 10 (8) (Aug. 2010) 550–560, https://doi.org/10.1038/nrc2886.

[13] J.-W. Zhao, et al., HPV16 integration probably contributes to cervical oncogenesis through interrupting tumor suppressor genes and inducing chromosome instability, J. Exp. Clin. Cancer Res. 35 (1) (Nov. 2016) 180, https://doi.org/10.1186/s13046-016-0454-4.

[14] A.A. McBride, A. Warburton, The role of integration in oncogenic progression of HPV-associated cancers, PLoS Pathog. 13 (4) (Apr. 2017) e1006211, https://doi.org/10.1371/journal.ppat.1006211.

[15] A. Warburton, T.E. Markowitz, J.P. Katz, J.M. Pipas, A.A. McBride, Recurrent integration of human papillomavirus genomes at transcriptional regulatory hubs, npj Genom. Med 6 (1) (Nov. 2021) 1–15, https://doi.org/10.1038/s41525-021-00264-y.

[16] M. Campitelli, et al., Human papillomavirus mutational insertion: specific marker of circulating tumor DNA in cervical cancer patients, PLoS One 7 (8) (2012) e43393, https://doi.org/10.1371/journal.pone.0043393.

[17] F. Luft, et al., Detection of integrated papillomavirus sequences by ligation-mediated PCR (DIPS-PCR) and molecular characterization in cervical cancer cells, Int. J. Cancer 92 (1) (2001) 9–17, https://doi.org/10.1002/1097-0215(200102)9999:9999<::AID-IJC1144>3.0.CO;2-L.

[18] R. Raybould, A. Fiander, G.W.G. Wilkinson, S. Hibbitts, HPV integration detection in CaSki and SiHa using detection of integrated papillomavirus sequences and restriction-site PCR, J. Virol. Methods 206 (Sep. 2014) 51–54, https://doi.org/10.1016/j.jviromet.2014.05.017.

[19] L. Zhou, et al., Long-read sequencing unveils high-resolution HPV integration and its oncogenic progression in cervical cancer, Nat. Commun. 13 (1) (May 2022) 1, https://doi.org/10.1038/s41467-022-30190-1.

[20] S. Yang, et al., Whole genome assembly of human papillomavirus by nanopore long-read sequencing, Front. Genet. 12 (2022) [Online]. Available: https://www.frontiersin.org/articles/10.3389/fgene.2021.798608. (Accessed 15 May 2023).

[21] L.S. Arroyo Mühr, D. Guerendiain, K. Cuschieri, K. Sundström, Human papillomavirus detection by whole-genome next-generation sequencing: importance of validation and quality assurance procedures, Viruses 13 (7) (Jul. 2021) 1323, https://doi.org/10.3390/v13071323.

[22] W.S. Chan, et al., An economical nanopore sequencing assay for human papillomavirus (HPV) genotyping, Diagn. Pathol. 15 (1) (May 2020) 45, https://doi.org/10.1186/s13000-020-00964-6.

[23] R.N. Brancaccio, A. Robitaille, S. Dutta, D.E. Rollison, M. Tommasino, T. Gheit, MinION nanopore sequencing and assembly of a complete human papillomavirus genome, J. Virol. Methods 294 (Aug. 2021) 114180, https://doi.org/10.1016/j.jviromet.2021.114180.

[24] R.R. Damerla, et al., Detection of early human papillomavirus–associated cancers by liquid biopsy, JCO Precis. Oncol. 3 (Apr. 2019) PO–18.00276, https://doi.org/10.1200/PO.18.00276.

[25] F. Krueger, FelixKrueger/TrimGalore, Perl, Sep. 24, 2024 [Online]. Available: https://github.com/FelixKrueger/TrimGalore. (Accessed 25 September 2024).

[26] N.-P.D. Nguyen, V. Deshpande, J. Luebeck, P.S. Mischel, V. Bafna, ViFi: accurate detection of viral integration and mRNA fusion reveals indiscriminate and unregulated transcription in proximal genomic regions in cervical cancer, Nucleic Acids Res. 46 (7) (Apr. 2018) 3309–3325, https://doi.org/10.1093/nar/gky180.

[27] K. Van Doorslaer, et al., The papillomavirus episteme: a major update to the papillomavirus sequence database, Nucleic Acids Res. 45 (Database issue) (Jan. 2017) D499–D506, https://doi.org/10.1093/nar/gkw879.

[28] Y. Wang, et al., shinyCircos-V2.0: leveraging the creation of circos plot with enhanced usability and advanced features, Imeta 2 (2) (May 2023) e109, https://doi.org/10.1002/imt2.109.

[29] A. Gonzalez-Perez, et al., IntOGen-mutations identifies cancer drivers across tumor types, Nat. Methods 10 (11) (2013) 1081–1082, https://doi.org/10.1038/nmeth.2642.

[30] S. Fishilevich, et al., GeneHancer: genome-wide integration of enhancers and target genes in GeneCards, Database 2017 (Apr. 2017) bax028, https://doi.org/10.1093/database/bax028.

[31] R.B. McCole, J. Erceg, W. Saylor, C.-T. Wu, Ultraconserved elements occupy specific arenas of three-dimensional mammalian genome organization, Cell Rep. 24 (2) (Jul. 2018) 479–488, https://doi.org/10.1016/j.celrep.2018.06.031.

[32] A.S. Hinrichs, et al., The UCSC genome browser database: update 2006, Nucleic Acids Res. 34 (Database issue) (Jan. 2006) 590–598, https://doi.org/10.1093/nar/gkj144.

[33] L. Yu, et al., HPV oncogenes expressed from only one of multiple integrated HPV DNA copies drive clonal cell expansion in cervical cancer, mBio 15 (5) (Apr. 2024) e00729, https://doi.org/10.1128/mbio.00729-24, 24.

[34] J.D. Meissner, 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) (1999) 1725–1733, https://doi.org/10.1099/0022-1317-80-7-1725.

[35] A.I. Ojesina, et al., Landscape of genomic alterations in cervical carcinomas, Nature 506 (7488) (Feb. 2014) 371–375, https://doi.org/10.1038/nature12881.

[36] G.-S. Min, J.R. Powell, Long-distance genome walking using the long and accurate polymerase chain reaction, Biotechniques 24 (3) (Mar. 1998) 398–400, https://doi.org/10.2144/98243bm13.

[37] P. Brotherton, et al., Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions, Nucleic Acids Res. 35 (17) (2007) 5717–5728, https://doi.org/10.1093/nar/gkm588.

[38] M. Parfenov, et al., Characterization of HPV and host genome interactions in primary head and neck cancers, Proc. Natl. Acad. Sci. 111 (43) (Oct. 2014) 15544–15549, https://doi.org/10.1073/pnas.1416074111.

[39] W. Li, et al., Characteristic of HPV integration in the genome and transcriptome of cervical cancer tissues, BioMed Res. Int. 2018 (Jun. 2018) 6242173, https://doi.org/10.1155/2018/6242173.

[40] X.-Y. Loh, et al., RNA-binding protein ZFP36L1 suppresses hypoxia and cell-cycle signaling, Cancer Res. 80 (2) (Jan. 2020) 219–233, https://doi.org/10.1158/0008-5472.CAN-18-2796.

[41] J. Chen, S. Patial, Loss of RNA binding protein, ZFP36L1, promotes EMT in hepatocellular cancer cells by regulating EMT-inducing transcription factor ZEB2, FASEB J. 35 (S1) (2021), https://doi.org/10.1096/fasebj.2021.35.S1.04347.

[42] K. Akagi, et al., Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability, Genome Res. 24 (2) (Feb. 2014) 185–199, https://doi.org/10.1101/gr.164806.113.

[43] C. Kothari, et al., Is carboxypeptidase B1 a prognostic marker for ductal carcinoma in situ? Cancers 13 (7) (Apr. 2021) 1726, https://doi.org/10.3390/cancers13071726.

[44] K. He, et al., CPA4 overexpression correlates with poor prognosis and tumor progression in endometrial cancer, Eur. J. Med. Res. 30 (Mar. 2025) 174, https://doi.org/10.1186/s40001-025-02293-0.

[45] L. Sun, et al., CPA4 is a promising diagnostic serum biomarker for pancreatic cancer, Am. J. Cancer Res. 6 (1) (Dec. 2015) 91–96.

[46] Q. Liu, et al., The CXCL8-CXCR1/2 pathways in cancer, Cytokine Growth Factor Rev. 31 (Oct. 2016) 61–71, https://doi.org/10.1016/j.cytogfr.2016.08.002.

[47] E. Azenshtein, T. Meshel, S. Shina, N. Barak, I. Keydar, A. Ben-Baruch, The angiogenic factors CXCL8 and VEGF in breast cancer: regulation by an array of pro-malignancy factors, Cancer Lett. 217 (1) (Jan. 2005) 73–86, https://doi.org/10.1016/j.canlet.2004.05.024.

[48] M. Bălăşoiu, et al., Serum and tumor microenvironment IL-8 values in different stages of colorectal cancer, Rom. J. Morphol. Embryol. 55 (2 Suppl) (2014) 575–578.