

RESEARCH

Open Access



Long-read sequencing reveals the structural complexity of genomic integration of HPV DNA in cervical cancer cell lines

Zhijie Wang^{1,2†}, Chen Liu^{1,2†}, Wanxin Liu^{1,2}, Xinyi Lv^{1,2}, Ting Hu^{1,2}, Fan Yang³, Wenhui Yang³, Liang He^{1,2*} and Xiaoyuan Huang^{1,2*}

Abstract

Background Cervical cancer (CC) causes more than 311,000 deaths annually worldwide. The integration of human papillomavirus (HPV) is a crucial genetic event that contributes to cervical carcinogenesis. Despite HPV DNA integration is known to disrupt the genomic architecture of both the host and viral genomes in CC, the complexity of this process remains largely unexplored.

Results In this study, we conducted whole-genome sequencing (WGS) at 55-65X coverage utilizing the PacBio long-read sequencing platform in SiHa and HeLa cells, followed by comprehensive analyses of the sequence data to elucidate the complexity of HPV integration. Firstly, our results demonstrated that PacBio long-read sequencing effectively identifies HPV integration breakpoints with comparable accuracy to targeted-capture Next-generation sequencing (NGS) methods. Secondly, we constructed detailed models of complex integrated genome structures that included both the HPV genome and nearby regions of the human genome by utilizing PacBio long-read WGS. Thirdly, our sequencing results revealed the occurrence of a wide variety of genome-wide structural variations (SVs) in SiHa and HeLa cells. Additionally, our analysis further revealed a potential correlation between changes in gene expression levels and SVs on chromosome 13 in the genome of SiHa cells.

Conclusions Using PacBio long-read sequencing, we have successfully constructed complex models illustrating HPV integrated genome structures in SiHa and HeLa cells. This accomplishment serves as a compelling demonstration of the valuable capabilities of long-read sequencing in detecting and characterizing HPV genomic integration structures within human cells. Furthermore, these findings offer critical insights into the complex process of HPV16 and HPV18 integration and their potential contribution to the development of cervical cancer.

Keywords Cervical cancer, HPV integration, HPV16, HPV18

[†]Zhijie Wang and Chen Liu contributed equally to this work.

*Correspondence:

Liang He
whuheliang@outlook.com
Xiaoyuan Huang
huangxy@tjh.tjmu.edu.cn

Full list of author information is available at the end of the article



Background

Cervical cancer (CC) ranks fourth for both mortality and incidence among females, with 570,000 women diagnosed with CC worldwide and approximately 311,000 died from it in 2018 [1]. Human papillomavirus (HPV) is widely recognized as the primary factor contributing to CC [2, 3]. HPV is a small DNA virus with a genome consisting of an approximately 8 kb circular, double-stranded DNA molecule, which includes the early gene region (E1-E7), the long control region (LCR), and the late gene region (L1 and L2) [4–6]. HPV is the main etiological factor in the process of CC carcinogenesis [7]. However, not all HPV infections suffered by women culminate in cervical cancer [7]. A phylogenetic tree construction based on the nucleotide sequence of the L1 gene classified HPV types into five genera – alpha, beta, gamma, mu, and nu [8–13]. The Alpha genus comprises 62 HPV types that infect the mucosal epithelium [8, 10, 14, 15]. The Alpha papillomaviruses are further classified into low-risk (LR) types and high-risk (HR) types based on their potential to cause anogenital cancer [8, 10, 11, 16]. Specifically, 15 HPV genotypes have been identified as HR types, including HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, 73, and 82 [17]. HR HPV genotype trigger the progression from normal cells to precancerous lesions and later to invasive lesions [18]. HPV16 and HPV18, the most common high-risk types, are responsible for causing cervical diseases and together contribute to approximately 70% of cervical cancers worldwide [19, 20].

The majority of HPV infections are transient and cleared by the immune system. However, 10–20% of infections persist latently [21, 22]. Persistent HPV infection is considered the main risk factor for CC [23]. More importantly, accumulated evidence indicated that HPV DNA can integrate into human genome, which is considered as one of the most important risk factors for CC development [24–27]. For instance, Hu *et al.* reported that the rate of HPV integration increased significantly from 53.8% (14 out of 26) of cervical intraepithelial neoplasia (CIN) to 81.7% (85 out of 104) of CC cases [25]. Similarly, Huang *et al.* [26] reported that HPV integration was detected in 97.8% of CC samples and 70.5% of CIN samples with HPV infection. Additionally, they found that the incidence of HPV integration was lower for low-risk HPV types compared to high-risk HPV types in both CC and CIN samples when compared to HPV-positive normal tissues [26]. Analysis of data from 169 HPV-positive cervical cancer patients from the Cancer Genome Atlas showed that HPV integration was detected in more than 80% of patients [27]. The physical status of the HPV genome in cervical cancer could be episomal, integrated, or mixed [28–31]. Initially, HPV infects host cells in a circular form, and with persistent infection, the circular

HPV genome undergoes breakage and integrates into the host genome [28–31]. During the process of HPV genome integration into the host cell genome, the E2 and/or E1 regions undergo breakage, while the long control region, and the E6 and E7 oncogenes consistently remain intact [28, 32–34]. Further studies suggest that viral integration tends to coincide with the development of high-grade cervical intraepithelial neoplasia (CIN II, III) due to the overexpression of the E6 and E7 oncogenes [35]. Measuring both the E1/E6 and/or E2/E6 ratio values is a promising prognostic tool that can offer valuable information about HPV16 integration and the physical state of HPV16 in the investigated cervical samples [36–38]. However, with the advancement of Next-Generation Sequencing (NGS) technology and targeted capture techniques, recent studies have found that breakpoints can occur anywhere in the viral genome [25–27]. Consequently, previous methods [36–38] to identify HPV16 integration status based on the E2/E6 or E1/E6 ratio may be inaccurate, as the E6 gene could be disrupted in some events.

So far, an increasing number of HPV integration sites have been identified across all chromosomes of the host cell [25, 34, 39–41]. Although These integration sites are randomly dispersed throughout the host genome, integration is not an entirely random event but also involves preferred chromosomal sites [42]. For example, HPV integration often occurring in transcriptionally active regions such as the *myc* oncogene [39, 43] or in chromosomally unstable regions known as common fragile sites [44, 45]. Moreover, several studies indicating that HPV16 DNA exhibits preferential sites for viral insertion, where chromosomal locations demonstrate significant homology to HPV16 DNA [25, 42, 46]. Additionally, HPV16 is observed to insert into repetitive elements scattered throughout the host chromosome, some of which are situated in close proximity to cancer-related genes [47]. However, the underlying mechanisms of HPV integration in tumorigenesis and tumor progression of CC are not well-understood. Among the various proposed mechanisms, the dysregulation of viral E6/E7 oncogene expression has been extensively studied. The integration of HPV DNA into the human genome can result in the dysregulation of the viral *E6* and *E7* oncogenes expression, which leads to the inactivation of the tumor suppressor proteins p53 and pRb, and consequent dysregulation of the apoptosis and cell cycle, respectively [48, 49]. In addition, HPV integration may also contribute to the progression of CC by promoting genomic instability or disrupting the expression and function of key host cellular genes which play vital roles in a wide range of biological processes including cell cycle regulation, cell proliferation, and apoptosis [25, 50, 51]. Previous studies have shown that

HPV integration can cause amplification and rearrangement of the host cell genome, including duplications, deletions, translocations, and other events [33]. The HeLa (HPV18 integration) and SiHa (HPV16 integration) cell lines, as two classic cell line models, have been widely used in research related to HPV integration and cervical cancer. Determining the impact of HPV integration on the genome structure of these two cell lines will help further elucidate the molecular mechanisms underlying HPV integration and its role in cervical cancer. Adey *et al.* [52] conducted an exceptional and comprehensive study utilizing multiple sequencing methodologies, such as shotgun, mate-pair, and long-read sequencing, to unravel the complete genome sequence of HeLa CCL-2. The results of this study provide a high-quality reference genome that is invaluable for current and future experiments reliant on the use of HeLa cells. Akagi *et al.* [33] reported that HPV16 integration in SiHa cells is associated with structural variations (SVs) in nearby host genomic regions, as determined through WGS analysis. They also proposed a "looping" model, in which HPV integrant-mediated DNA replication and recombination can lead to the formation of viral-host DNA concatemers. Xu *et al.* [53] reported that HPV integration in SiHa cells causes genomic variation in the host cell genome. However, since the sequencing read length in these relevant reports for WGS analysis was relatively short, it may be difficult to accurately describe some relatively complex and longer SVs. Third-generation sequencing technologies, such as nanopore sequencing and PacBio sequencing, enable us to elucidate the complex structural aspects of genomes, due to its long read length and lack of PCR amplification, while shorter NGS reads cannot [54, 55]. It is reported that third-generation sequencing has the capability to generate extended contiguous sequences with an average read length exceeding 10 kb and subreads that can span over 60 kb [56].

Recently, HPV targeted capture-based long-read sequencing approaches have been utilized in several research studies to construct HPV genome sequences, determine integration status and sites, and explore

integration patterns [57–62]. For instance, Yang *et al.* provided compelling evidence and developed a tool to demonstrate the potential usefulness of nanopore technology in identifying viral status [63]. Iden *et al.* successfully identified a total of 87 integration events, comprising 267 distinct human-HPV breakpoints in 8 tumors by PacBio long read sequencing [59]. It is worth noting that although these HPV targeted capture-based long-read sequencing approaches can effectively increase the sequencing depth of targeted sequences and reduce sequencing costs, the loss of human genomic sequence information due to targeted capture prevents a comprehensive analysis of whole-genome variations caused by HPV integration. Therefore, it is necessary to perform non-targeted capture long-read sequencing on these two cervical cancer cells to accurately reveal the complexity of the impact of HPV integration on both the HPV genome and the host cell genome structure.

In this study, we aim to precisely elucidate the genomic integration structures of HPV16 and HPV18 in CC cell lines (HeLa and SiHa) by long-read sequencing via PacBio platform. These findings provide important insights into the process of HPV16 and HPV18 integration and may provide a greater understanding of the molecular mechanisms contributing to cervical cancer.

Results

HPV integration breakpoints detected by targeted-capture NGS

Previous studies have utilized targeted-capture NGS for the analysis of HPV integration [25]. Our findings indicated that HPV16 integrated into chromosome 13, specifically in the intergenic region closest to the KLF5 and LINC00392 genes, in SiHa cells (Table 1). Additionally, we observed HPV18 integration into chromosome 8, specifically in the intergenic region nearest to the CCAT1 and CASC21 genes, in HeLa cells (Table 1). These results are consistent with previous findings by others [25]. However, although HPV targeted capture NGS technology has demonstrated effectiveness and cost-efficiency in detecting HPV integration, there are limitations in

Table 1 Summary of HPV integration sites identified using targeted-capture NGS data

Sample	Breakpoint in HPV	Integrated Position	Region	Gene
SiHa	HPV16: 3134	chr13: 74,087,562	intergenic	KLF5, LINC00392
	HPV16: 3384	chr13:73,788,866	intergenic	KLF5, LINC00392
HeLa	HPV18: 5734	chr8:128,230,629	ncRNA_intronic	CCAT1
	HPV18: 2498	chr8:128,241,546	intergenic	CCAT1, CASC21
	HPV18: 7858	chr8:128,234,255	intergenic	CCAT1, CASC21
	HPV18: 3101	chr8:128,233,367	intergenic	CCAT1, CASC21

determining the structure of the integrated genome. Firstly, the short length of our NGS reads (150 bp) poses challenges in determining the complex intergration structure of the integrated genome. Secondly, the targeted-capture NGS strategy results in the loss of a significant portion of human genome reads, leading to an insufficient number of human reads for accurate identification of complex SVs associated with human genome. Specifically, our targeted-capture NGS reads specific to HPV16 and HPV18 represented 65.87% and 79.84% of the total sequencing reads, respectively. These findings suggest that HPV targeted NGS is primarily capable of identifying integration breakpoints between the human genome and the HPV genome, rather than providing a comprehensive understanding of the complex integrated genome structure.

PacBio sequencing reveals complex HPV integrated genome structures

PacBio long-read sequencing enables the generation of lengthy reads that encompass one or multiple HPV DNA fragments flanked by the human genome on both ends. This characteristic enhances the reliability and facilitates direct mapping of the reads across HPV breakpoints, thereby facilitating the elucidation of genomic integration structures. Therefore, we sequenced the genome of SiHa cells by using PacBio Sequel sequencing platform. Sequencing of SiHa cells generated a total of 194.18 Gb of bases and 8,381,208 reads after filtering out low-quality reads. The N50 read length was 34.81 kb (Table 2). Consistent with HPV targeted capture NGS results, the two integration sites: HPV16: 3134-chr13:74,087,562 and HPV16: 3384-chr13:73,788,866 were also identified by PacBio long read sequencing results in SiHa cells. These results suggesting that PacBio long-read sequencing effectively identifies HPV integration breakpoints with comparable accuracy to targeted-capture NGS methods. Furthermore, our results revealed integrated segments of

the complete HPV16 L1, L2, E1, E4, E5, E6, and E7 genes, along with partial sequences of the E2 gene, within the SiHa cell genome (Fig. 1). The integration of HPV 16 occurred twice, with a fragment of HPV16 (coordinates from 3384 to 7906/1–3134) integrating into chromosome 13 at genomic coordinates 73,788,866–74,087,562 in the human genome (Fig. 1). We also observed deletions in the HPV16 genome at positions 3460–3508 and 7757–7793. Additionally, we conducted an analysis of alterations in the human genome near the site of HPV16 integration. Our findings indicate that the HPV16 fragment integrated in a reverse orientation at chr13:73,788,866 and chr13:74,087,562 (Fig. 1). Furthermore, we directly identified the chromosomal arrangement (chr13:73,255,335–73,464,522) in close proximity to the integration sites in SiHa cells (Fig. 1). Taken together, these results imply that the integration of HPV16 may contribute to the instability of the genomic structure in the vicinity of the integration site.

Similarly, PacBio long-read sequencing was performed on HeLa cells. The sequencing runs of HeLa cells yielded a total of 211.38 Gb of bases and 7,932,747 reads after eliminating low-quality reads. The N50 read length was 40.24 kb (Table 2). Integration of the incomplete HPV18 genome (coordinates from 5735 to 7858/1–3100) was detected at chromosome 8q24 (Fig. 2). Integration sites identified through NGS were also confirmed by long read sequencing in HeLa cells. These results further suggest that PacBio long-read sequencing effectively identifies HPV integration breakpoints with comparable accuracy to targeted-capture NGS methods. Upon further analysis, we constructed two HPV integration models in HeLa cells based on whether the integration fragment contains HPV18: 2497–3100 (Fig. 2A and 2B). As shown in Fig. 2, a fragment of HPV18 (3101–5734) was deleted during the integration process of HPV18. Additionally, segments 1–2497, 1–3100, and 5736–7857 were identified as retro-inserted into the human genome, with the latter segment occurring in multiple copies (Fig. 2). Collectively, focal genome amplifications and rearrangements were observed in the human genome in the vicinity of the HPV18 integration site. These findings provide additional evidence that HPV integration disrupts genes near the integration breakpoints and may induce significant genomic instability, leading to genome rearrangements.

PacBio long read sequencing, with its longer read lengths, can detect longer segment SVs, such as duplications (DUPs), insertions (INs), deletions (DELs), translocations (TRAs) and inversions (INVs). Although the previous results showed that HPV integration can cause genomic SVs near integration sites, HPV integration events may also cause SVs in other regions of the human genome. Therefore, we further analyzed SVs

Table 2 Statistics of subreads data

	SiHa	HeLa
Subreads base(G)	194.18	211.38
Subreads number	8,381,208	7,932,747
Average subreads length	23,169	26,647
N50	34,805	40,241
Average sequencing depth	56.62	62.84
Coverage	99.04%	99.10%
≥ 4X Coverage	98.84%	98.93%
≥ 10X Coverage	98.58%	98.69%
≥ 20X Coverage	97.87%	98.15%

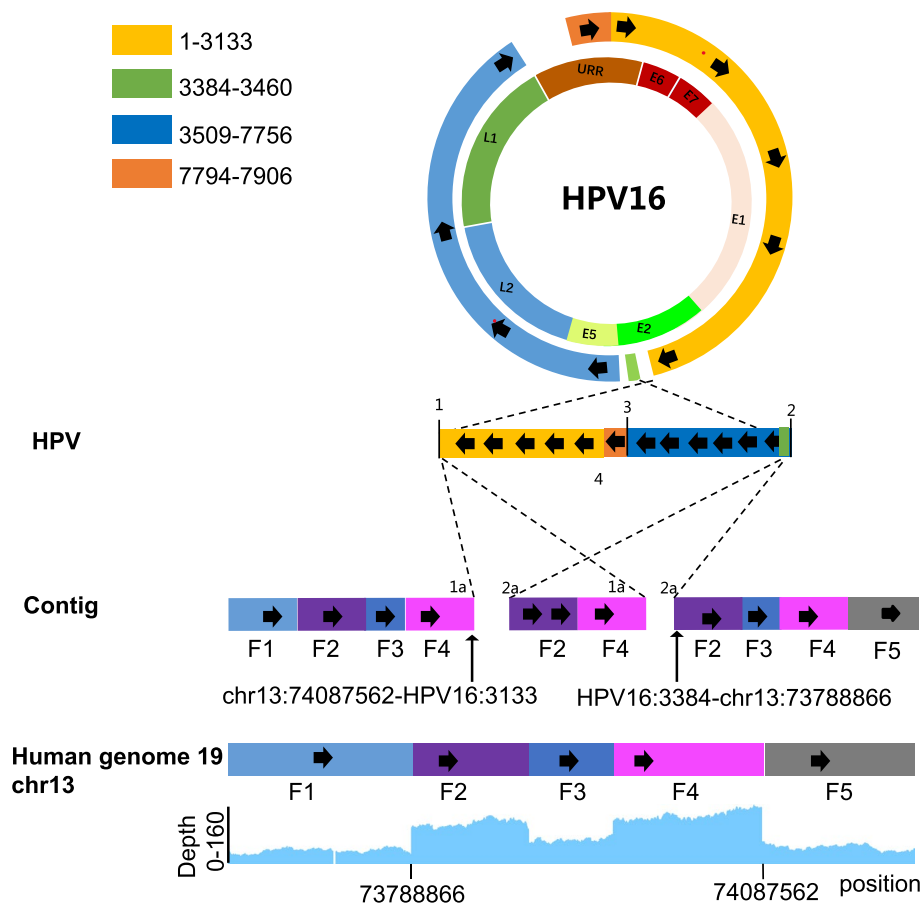


Fig. 1 HPV16 complex integrated genome structure in SiHa cells. In the top panel, the colored regions in the outer circle represent the HPV16 sequences contained within the integrated structure, while the white regions represent the sequences that were replaced or lost as a result of integration. The inner circle shows the complete HPV16 genome for reference. The bottom panel illustrates the HPV and human reference genomes, which are connected by dotted lines to a contig that covers the integration, demonstrating how it specifically mapped to each genome (1, 2, 3, and 4 represent breakpoints on the HPV16 genome; 1a and 2a represent breakpoints on human chromosome 13; F1, F2, F3, F4, F5 represent consecutive segments on human chromosome 13). Schematic annotations of the integration event were made using data from all PacBio long reads that covered the integration event

across the whole genome in SiHa and HeLa cells, respectively. Our results showed that most of SVs were located in intergenic and intronic regions, with limited representation from 3'- or 5'-UTR, splicing, ncRNA, exonic, upstream and downstream regions (Table 3). INSNs and DELs were the most common types of SVs in the human genome. Prior research has demonstrated that the integration of HPV16-LINC00393 has a significant impact on gene expression in SiHa cells [53]. Specifically, 74 genes located on chromosome 13 exhibit alterations in expression levels, with 37 genes exhibiting up-regulation and 37 genes exhibiting down-regulation [53]. SVs play a significant role in gene expression differences in humans and often affect multiple neighboring genes [64]. Therefore, to explore the potential association between changes in gene expression levels and SVs on chromosome 13 in the genome of SiHa cells, we conducted an analysis to

determine whether SVs were present in the vicinity of these 74 genes. As shown in Fig. 3, our findings demonstrate that a significant proportion of genes (50 out of 74, accounting for 67.57%) exhibited SVs, indicating a potential correlation between these genomic events and alterations in gene expression patterns.

Microhomologies (MHs) were identified between the HPV genome and the human genome in close proximity to the integration breakpoints

A previous study has indicated that microhomology-mediated recombination (MMR) could potentially serve as a mechanism for HPV integration [25]. Hence, we conducted an analysis to examine the characteristics of the HPV and human genome sequences near the integration sites in SiHa and HeLa cell lines to determine whether the integration events were associated

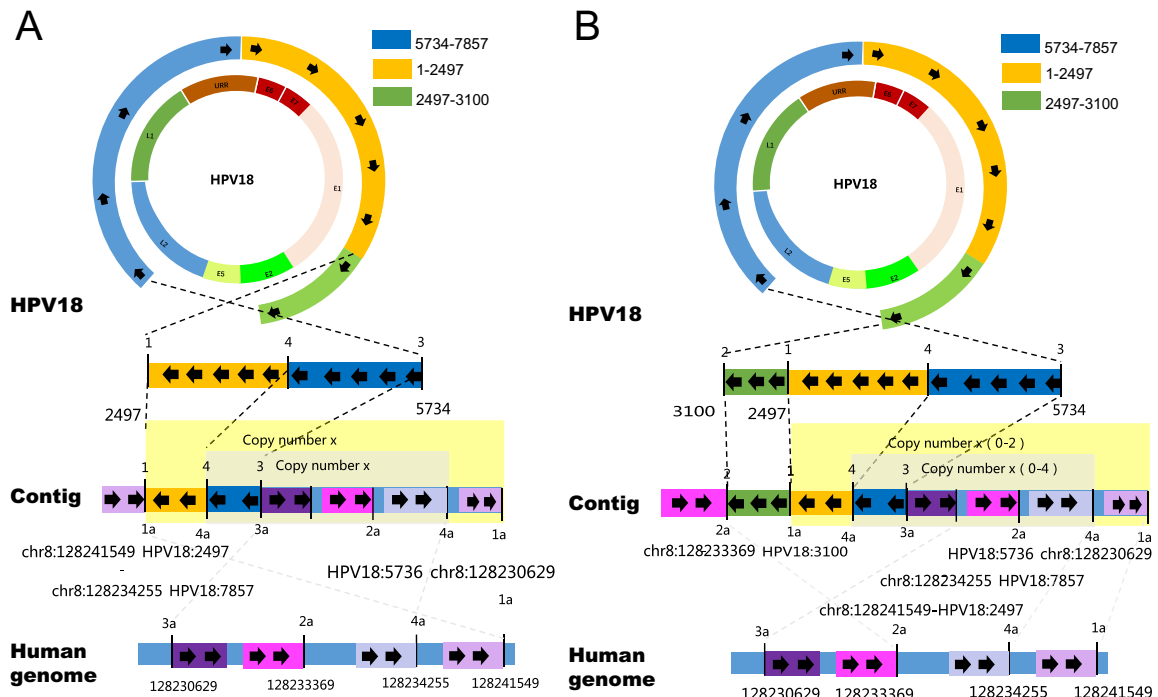


Fig. 2 HPV18 complex integrated genome structure in HeLa cells. **A, B** Two proposed models of HPV18 integration genome structure in HeLa cells. In the top panel, two proposed HPV18 integration genome structures in HeLa cells are depicted. The colored regions in the outer circle represent the HPV18 sequences contained within the integrated structure, while the white regions represent the sequences that were replaced or lost as a result of integration. The inner circle shows the complete HPV18 genome for reference. The bottom panel illustrates the HPV and human reference genomes, which are connected by dotted lines to a contig that covers the integration, demonstrating how it specifically mapped to each genome. Schematic annotations of the integration event were made using data from all PacBio long reads that covered the integration event

Table 3 Statistical analysis of SVs in SiHa and HeLa cell lines

Sample	type	exonic	splicing	ncRNA	UTR5	UTR3	intronic	Up-stream	Down-stream	intergenic
SiHa	DUP	51	1	5	0	0	13	0	1	87
	INV	87	0	21	0	1	20	1	1	117
	TRA	7	1	74	2	7	357	7	13	1460
	DEL	145	6	655	10	53	3281	75	86	8186
	INS	361	4	2273	87	281	11,823	333	256	19,282
HeLa	DUP	19	0	2	0	0	11	1	1	59
	INV	78	1	12	0	0	14	0	1	112
	TRA	9	2	71	2	6	322	6	10	1481
	DEL	135	8	814	14	64	4071	95	116	9869
	INS	349	11	2735	83	290	13,232	327	265	22,220

with MMR. Two integration sites were identified in SiHa cells: HPV16: 3134-chr13:74,087,562 and HPV16: 3384-chr13:73,788,866. A microhomologous "ATGC" fragment was observed at the integration site HPV16: 3134-chr13:74,087,562. The integration site HPV16: 3384-chr13:73,788,866 exhibited a microhomologous "TATT" fragment (Fig. 4A). Four integration sites were identified in HeLa cells: HPV18: 2498-chr8: 128,241,546,

HPV18: 3101-chr8: 128,233,367, HPV18: 5735-chr8: 128,230,629, and HPV18: 7857-chr8: 128,234,255. A microhomologous "TAAC/TACA" fragment was observed at the integration site HPV18: 2498-chr8: 128,241,546. The integration site HPV18: 5735-chr8: 128,230,629 exhibited a microhomologous "ATAA" fragment. At the HPV18: 7857-chr8: 128,234,255 integration site, a "TACT/TACA" fragment was observed, while no

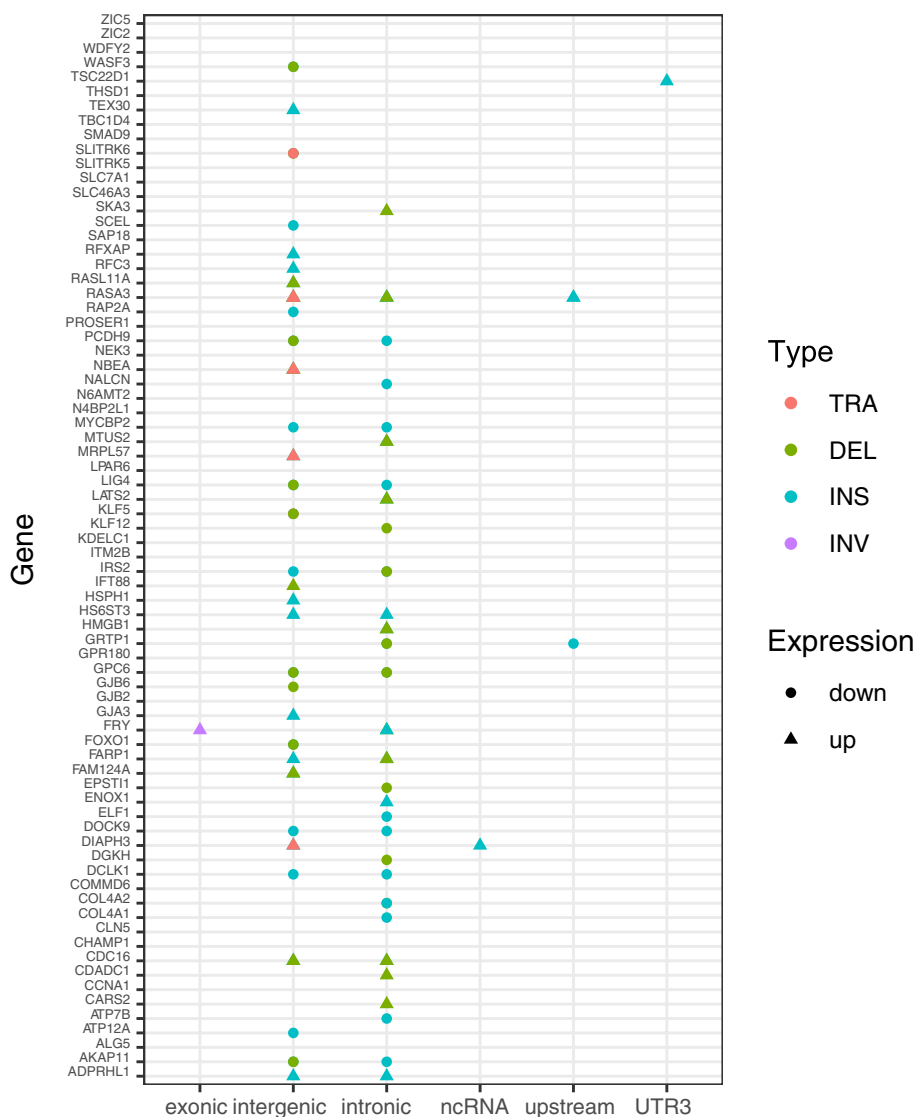


Fig. 3 Distribution of SVs among Differentially Expressed Genes on Chromosome 13 in SiHa Cells. The x-axis represents the different regions of the genome, including exonic, splicing, ncRNA, UTR5, UTR3, intronic, upstream, downstream, and intergenic regions. The y-axis shows the differentially expressed genes. Different types of SVs are represented by different colors: TRAs (rose red), DELs (dark green), INSs (dark cyan), and INVs (pale purple). The upregulated genes are denoted by triangles, while the downregulated genes are represented by circles

microhomologous fragment was found at the HPV18: 3101-chr8: 128,233,367 integration site (Fig. 4B). These results provide further evidence supporting the notion that MMR may indeed be the mechanism underlying HPV integration.

Discussion

Among all cancer types, CC ranks fourth in incidence and mortality in females worldwide, with 85% of cases occurring in developing countries [1, 65]. Accumulated evidence has indicated that integration of HPV genome into the human genome is considered as one of the most

important risk factors for CC development [24–27]. Although the HPV targeted-capture NGS is a highly sensitive method for detecting HPV infection and integration status, it remains challenging to accurately identify complex integrated genome structure upon HPV integration using targeted-capture NGS. Third-generation sequencing technologies, such as nanopore sequencing and PacBio sequencing, enable us to elucidate the complex structural aspects of genomes, due to their long-read length and lack of PCR amplification, whereas shorter NGS reads cannot [54, 55]. In this study, PacBio long-read sequencing was employed to accurately identify

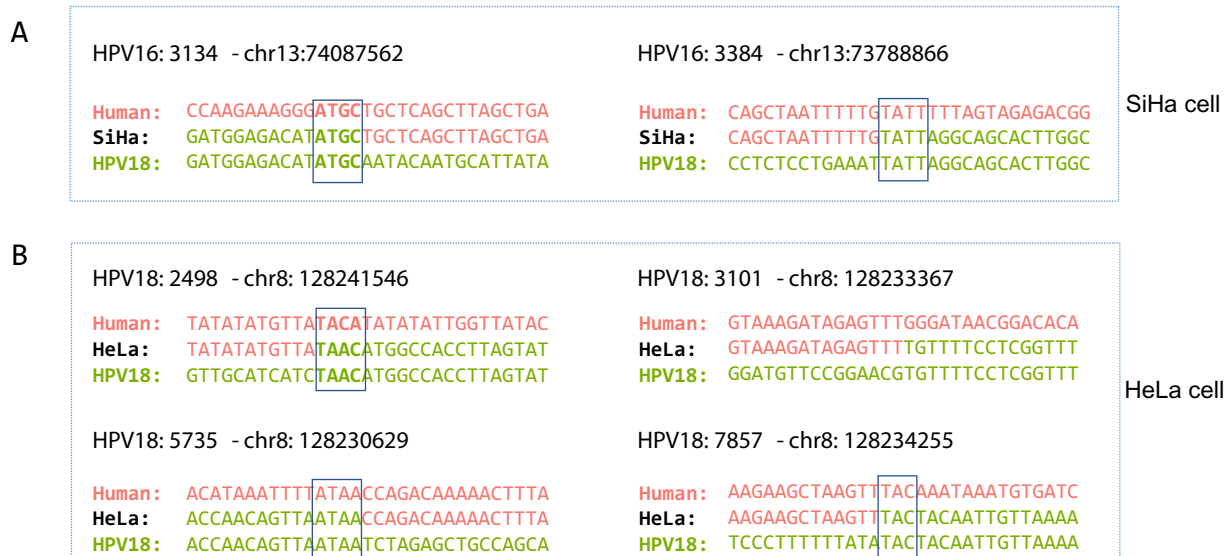


Fig. 4 HPV integration mechanisms in SiHa and HeLa cells. **A, B** The figure shows the alignment of the sequence around the integration site between the human genome (rose red) and the HPV16/HPV18 genome in SiHa cells or HeLa cells (dark green), respectively. The black box represents the aligned nucleotides in the microhomology (MH) region of the two reference sequences at the HPV integration site

complex integrated genome structure upon HPV integration in SiHa and HeLa cell lines. We directly identified the chromosomal arrangements associated with integrated HPV16 DNA in SiHa cells (Fig. 1). Furthermore, HPV18 integration was found to induce microduplication in both HPV and human genomes (Fig. 2). HPV integration not only induces rearrangements in the chromosomes near the integration sites, but also leads to alterations within the HPV genome itself (Fig. 1–2). Specifically, as illustrated in Fig. 1, deletions were observed in the HPV16 genome at positions 3460–3508 and 7757–7793 in SiHa cells. HPV16 manifested in a concatenated repeat form. Similar outcomes were noted in HeLa cells, a fragment of HPV18 (3101–5734) was deleted during the integration process of HPV18. Additionally, segments 1–2497, 1–3100, and 5736–7857 were identified as retro-inserted into the human genome, with the latter segment occurring in multiple copies (Fig. 2). These findings are consistent with earlier reports [33, 36, 66], highlighting that HPV16 integration can trigger rearrangements within both the human and HPV genomes in cell lines and clinical samples. For example, in a study [36], extensive mapping analysis of HPV-16 E1 and E2 genes in 37 selected tumors revealed deletions in both E1 and E2 genes, with the maximum number of losses (78.4%) observed within the HPV-16 E2 hinge region. Akagi *et al.* [33] noted that HPV integrant-mediated DNA replication and recombination may lead to the formation of viral–host DNA concatemers, often disrupting genes implicated in oncogenesis and amplifying HPV oncogenes E6 and E7. Furthermore, Tsakogiannis *et al.* using

the Restriction Site-PCR (RS-PCR) method, discovered two distinct rearranged HPV16 intra-viral sequences [66]. These rearrangements involve the conjunction of E2 and L1 genes, as well as the conjunction of E1 and L1 genes with inverted orientation [66]. Together, complex integrated genome structures including the HPV genome and the nearby human genome were depicted through PacBio long-read whole genome sequencing in HeLa and SiHa cells. Therefore, this study advances our understanding of the impact of HPV16 and HPV18 integration on alterations in chromosome architecture and cervical tumorigenesis.

Our results from HPV targeted-capture NGS (Table 1), along with previous studies, strongly indicate that SiHa cells have only two virus-human junctions [67, 68]. However, Diao *et al.* reported that SiHa cells contain two copies of HPV16 DNA, suggesting a possibility of four virus-human junctions [69]. These findings suggest that the two integrated HPV fragments may have the same junction and partially overlap with each other at the integration site [33]. Our PacBio long-read sequencing results have confirmed that two copies of HPV16 DNA were integrated into the human genome, which is consistent with recent findings reported by Akagi *et al.* [33]. Meanwhile, we identified multiple integration events in HeLa cells, and these integration events share one or two breakpoints. Additionally, some studies have shown that HPV integration can increase genomic instability [33, 70]. Therefore, we speculated that the initial integration of a single HPV fragment into a specific location in the human genome triggers increased genomic instability

near the integration site, leading to subsequent focal genome amplifications and rearrangements in both the HPV and human genome sequences in the vicinity of the integration site. As a result, multiple integration events with shared breakpoints occur.

Despite the close association between HPV integration and cervical cancer development, the molecular mechanisms underlying HPV integration remain poorly understood. Previous studies have suggested that MMR is a potential mechanism for HPV integration [25]. Our findings demonstrate that two integration sites in SiHa cells and three out of the four common integration sites in HeLa cells are consistent with the mechanism of MMR, while one site deviates from this pattern. These findings indicate that the molecular mechanisms involved in HPV integration are intricate. Although MMR is a plausible molecular mechanism for HPV integration, further studies are needed to explore additional mechanisms.

Additionally, Xu *et al.* reported a significant impact of HPV16-LINC00393 integration on gene expression in SiHa cells [53]. Specifically, 74 genes located on chromosome 13 show changes in expression levels, with 37 genes up-regulated and 37 genes down-regulated [53]. Our findings revealed that the majority of these genes (50/74, 67.57%) displayed SVs in SiHa cells, indicating that integration of HPV16-LINC00393 may result in genomic SVs, thereby influencing the expression levels of associated genes (Fig. 3). Importantly, these structural variations manifest in various forms, such as DELs, INs, DUPs, INVs, and TRAs. Furthermore, these SVs manifest in diverse regions of the human genome, encompassing UTR3, UTR5, upstream, downstream, non-coding RNA, splicing regions, exonic regions, intronic regions, and intergenic regions. Further research is required to ascertain the precise impact of these structural variations on gene expression. Meanwhile, some genes and their surrounding regions (24/74, 32.43%) may undergo expression changes in the absence of SVs, suggesting involvement of alternative regulatory mechanisms in gene expression. Elucidating these molecular mechanisms will provide additional insights into the molecular mechanisms underlying HPV integration and its association with CC.

However, this study has certain limitations. Firstly, the sample size is small, and secondly, cell lines were used instead of tissue samples. Further investigations with larger tissue sample sizes are required to explore this topic in more depth. Although tissue samples were not the primary focus of this study, Zhou *et al.* utilized another third-generation long read sequencing technology—Oxford Nanopore Technology (ONT) sequencing—to investigate HPV integration in clinical samples

[57]. Notably, ONT sequencing revealed that HPV16 integration tends to induce genomic instability, rearrangements and SVs near integration sites [57]. This aligns with the conclusions drawn from our cell line samples, reinforcing the findings obtained through PacBio sequencing in cell lines. While ONT sequencing was employed in the investigation of HPV integration, our research utilized PacBio sequencing. Both methods belong to third-generation sequencing technologies, leveraging their extended read lengths to detect intricate genomic structures. It is crucial to acknowledge that each technology, including ONT and PacBio, has its own set of advantages and disadvantages. On the one hand, ONT sequencing generates longer contiguous sequence reads, with higher throughput, portability, and lower cost, making ONT promising for various applications, especially for genome-wide and transcriptome-wide studies requiring large amounts of data [71]. On the other hand, PacBio can produce higher quality raw data with a lower error rate and higher mappability compared to ONT raw data, proving invaluable for dissecting intricate genomic features, uncovering SVs, and elucidating epigenetic modifications [71, 72]. The selection of sequencing technology should be tailored to meet the specific needs of our study. Thus, each of these two technologies has its own strengths and weaknesses and should not be used interchangeably in different studies.

Conclusions

In this study, we have successfully depicted high-resolution complex models of HPV integrated genome structures, specifically HPV16 and HPV18, in SiHa and HeLa cell lines using PacBio long-read sequencing. This achievement serves as a compelling demonstration of the invaluable capabilities of long-read sequencing in the detection and characterization of viral genomic integration structures within host cells. Moreover, these findings offer significant insights into the intricate process of HPV16 and HPV18 integration, thereby contributing to a better understanding of the molecular mechanisms underlying cervical cancer development.

Methods and materials

Cell culture

HeLa and SiHa cell lines were obtained from the American Type Culture Collection (Manassas, VA, USA). Two cell lines were cultured in DMEM medium (Gibco) supplemented with 1% streptomycin / penicillin solution (Gibco) and 10% fetal bovine serum (Gibco). The cell lines were maintained in an incubator with 5% CO₂ at 37 °C.

Targeted-hybridization short-read sequencing

The Genomic DNA Extraction Kit (Vazyme, Nanjing, China) was used to extract genomic DNA according to the instructions of the manufacturer. The DNA concentration was quantified using the Qubit dsDNA HS Assay Kit (Thermo Fisher). 500 ng of DNA was divided into fragments (200 bp to 300 bp) by Enzymes (Vazyme, Nanjing, China), then these fragments were end-repaired, dA-tailed, and adaptor ligated. ligation products were PCR-amplified and obtained a DNA library. The concentration of the DNA library was quantified using a Qubit dsDNA HS Assay Kit. Hybridization was performed using the HPV probes for the full-length HPV16/18 genomes and a GenCap enrichment kit according to the Target Enrichment Protocol (iGeneTech, Beijing, China). In brief, 500 ng of DNA library were hybridized with HPV probes at 65 °C for 16 h, and the washing buffer was used to remove un-captured fragments. These enriched fragments were amplified by PCR. The hybridization procedure was conducted for twice. Lastly, the enriched libraries were sequenced on an Illumina Sequencer (HiSeq 2000) with 150 bp paired-end reads.

PacBio long-read sequencing

Genomic DNA of HeLa and SiHa cancer cell lines were extracted using the Genomic DNA Extraction Kit (Vazyme, Nanjing, China). DNA integrity was checked with the Agilent 4200 Bioanalyzer (Agilent Technologies). A total of 8 µg genomic DNA was sheared using Covaris g-TUBEs (Covaris) and then purified using AMPure PB magnetic beads. SMRT bell libraries were prepared with the SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences). Each library was size-selected on a BluePippin system for 25 kb molecules, followed by annealing the sequencing primer and the binding polymerase to SMRTbell templates. Libraries were then sequenced on Pacific Bioscience Sequel II platform (Pacific Biosciences) at Frasiergen Bioinformatics Co., Ltd (Wuhan, China).

Short-read sequence alignment

The targeted-capture NGS data was used to determine the HPV integration status present in samples. Firstly, clean reads were obtained by removing adaptor-contaminated reads, duplicated reads and low-quality reads. Clean reads were aligned to the human genome (GRCh38) and HPV genome (HPV16: NC_001526.4, https://www.ncbi.nlm.nih.gov/nuccore/NC_001526.4/; HPV18: AY262282.1, <https://www.ncbi.nlm.nih.gov/nuccore/AY262282.1>) using a Burrows–Wheeler Aligner (BWA) [73]. Subsequently, the paired-end reads that aligned perfectly to HPV or human reference

genome were removed. Lastly, the human-HPV chimeric paired-end reads were used to define the position of breakpoints in the human genome and HPV genome.

Long-read sequence alignment

The subreads sequences were obtained by processing the raw sequence data on SMRTlink v9.0 software. The clean reads were mapped to a merged reference genome that comprises human genome (GRCh38) and HPV genome (HPV16: NC_001526.4, HPV18: AY262282.1) with a long-read mapper: Minimap2 (v2.24; <https://github.com/lh3/minimap2>) [74]. Output sequence alignment/map (SAM) files were transformed to binary format (BAM) files, and then were sorted and indexed with samtools (v1.13; <http://samtools.sourceforge.net/>) [75]. Then Sniffles software was used to call SVs from the bam file. ANNOVAR was used to do the annotation for the integrated breakpoints [76].

Acknowledgements

We thank Yanbing Cheng, Wei Li and Rong Huang from Frasiergen Bioinformatics Co., Ltd (Wuhan, China). for assistance, help, and advice during bioinformatics analysis.

Authors' contributions

HX, HL and WZ contributed to the study design. YF and YW contributed to data collection. WZ, LC, HT, HL and HX performed data analysis and interpretation. LW and LX performed experiments. WZ and LC drafted the manuscript. HL and HX revised the manuscript. All authors read and approved the final manuscript.

Funding

This study was supported by Department of Science and Technology of Hubei Province (2021BCA108), the China National Natural Science Foundation grants (82100344) and Wuhan Science and Technology Bureau (2019030703011518).

Availability of data and materials

The sequencing raw data have been deposited in the Genome Sequence Archive of the National Genomics Data Center, China National Center for Bioinformatics, under accession number HRA007085 (<https://ngdc.cncb.ac.cn>).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, Hubei, China. ²Cancer Biology Research Center (Key Laboratory of the Ministry of Education), Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, Hubei, China. ³Wuhan Kandwise Biotechnology, Inc. Wuhan, Hubei, China.

Received: 16 June 2023 Accepted: 8 February 2024

Published online: 20 February 2024

References

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394–424.
- Walboomers JM, Jacobs MV, Manos MM, Bosch FX, Kummer JA, Shah KV, Snijders PJ, Peto J, Meijer CJ, Munoz N. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J Pathol*. 1999;189(1):12–9.
- Bosch FX, Manos MM, Munoz N, Sherman M, Jansen AM, Peto J, Schiffman MH, Moreno V, Kurman R, Shah KV. Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. International biological study on cervical cancer (IBSCC) Study Group. *J Natl Cancer Inst*. 1995;87(11):796–802.
- Kadaja M, Silla T, Ustav E, Ustav M. Papillomavirus DNA replication - from initiation to genomic instability. *Virology*. 2009;384(2):360–8.
- Kajitani N, Schwartz S. Role of Viral Ribonucleoproteins in Human Papillomavirus Type 16 Gene Expression. *Viruses*. 2020;12(10):1110.
- McBride AA. Mechanisms and strategies of papillomavirus replication. *Biol Chem*. 2017;398(8):919–27.
- Bosch FX, Lorincz A, Munoz N, Meijer CJ, Shah KV. The causal relation between human papillomavirus and cervical cancer. *J Clin Pathol*. 2002;55(4):244–65.
- Bletsa G, Zagouri F, Amoutzias GD, Nikolaidis M, Zografos E, Markoulatos P, Tsakogiannis D. Genetic variability of the HPV16 early genes and LCR. Present and future perspectives. *Expert Rev Mol Med*. 2021;23:e19.
- Horvath CA, Boulet GA, Renoux VM, Delvenne PO, Bogers JP. Mechanisms of cell entry by human papillomaviruses: an overview. *Virology*. 2010;7:11.
- Bernard HU, Burk RD, Chen Z, van Doorslaer K, zur Hausen H, de Villiers EM: Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. *Virology*. 2010;401(1):70–9.
- Bernard HU, Calleja-Macias IE, Dunn ST. Genome variation of human papillomavirus types: phylogenetic and medical implications. *Int J Cancer*. 2006;118(5):1071–6.
- Doorbar J. Molecular biology of human papillomavirus infection and cervical cancer. *Clin Sci (Lond)*. 2006;110(5):525–41.
- Gheit T. Mucosal and Cutaneous Human Papillomavirus Infections and Cancer Biology. *Front Oncol*. 2019;9:355.
- Chen Z, de Freitas LB, Burk RD. Evolution and classification of oncogenic human papillomavirus types and variants associated with cervical cancer. *Methods Mol Biol*. 2015;1249:3–26.
- de Villiers EM, Fauquet C, Broker TR, Bernard HU, zur Hausen H: Classification of papillomaviruses. *Virology*. 2004;324(1):17–27.
- zur Hausen H. Papillomavirus infections—a major cause of human cancers. *Biochim Biophys Acta*. 1996;1288(2):F55–78.
- Munoz N, Bosch FX, de Sanjose S, Herrero R, Castellsague X, Shah KV, Snijders PJ, Meijer CJ. International Agency for Research on Cancer Multicenter Cervical Cancer Study G: Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N Engl J Med*. 2003;348(6):518–27.
- Schiffman M, Castle PE, Jeronimo J, Rodriguez AC, Wacholder S. Human papillomavirus and cervical cancer. *Lancet*. 2007;370(9590):890–907.
- Li N, Franceschi S, Howell-Jones R, Snijders PJ, Clifford GM. Human papillomavirus type distribution in 30,848 invasive cervical cancers worldwide: Variation by geographical region, histological type and year of publication. *Int J Cancer*. 2011;128(4):927–35.
- Guan P, Howell-Jones R, Li N, Bruni L, de Sanjose S, Franceschi S, Clifford GM. Human papillomavirus types in 115,789 HPV-positive women: a meta-analysis from cervical infection to cancer. *Int J Cancer*. 2012;131(10):2349–59.
- Shanmugasundaram S, You J. Targeting Persistent Human Papillomavirus Infection. *Viruses*. 2017;9(8):229.
- Crosbie EJ, Einstein MH, Franceschi S, Kitchener HC. Human papillomavirus and cervical cancer. *Lancet*. 2013;382(9895):889–99.
- zur Hausen H. Papillomaviruses and cancer: from basic studies to clinical application. *Nat Rev Cancer*. 2002;2(5):342–50.
- Pett M, Coleman N. Integration of high-risk human papillomavirus: a key event in cervical carcinogenesis? *J Pathol*. 2007;212(4):356–67.
- Hu Z, Zhu D, Wang W, Li W, Jia W, Zeng X, Ding W, Yu L, Wang X, Wang L, et al. Genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism. *Nat Genet*. 2015;47(2):158–63.
- Huang J, Qian Z, Gong Y, Wang Y, Guan Y, Han Y, Yi X, Huang W, Ji L, Xu J, et al. Comprehensive genomic variation profiling of cervical intraepithelial neoplasia and cervical cancer identifies potential targets for cervical cancer early warning. *J Med Genet*. 2019;56(3):186–94.
- Cancer Genome Atlas Research N, Albert Einstein College of M, Analytical Biological S, Barretos Cancer H, Baylor College of M, Beckman Research Institute of City of H, Buck Institute for Research on A, Canada's Michael Smith Genome Sciences C, Harvard Medical S, Helen FGCC et al. Integrated genomic and molecular characterization of cervical cancer. *Nature*. 2017;543(7645):378–384.
- Tsakogiannis D, Kyriakopoulou Z, Ruether IGA, Amoutzias GD, Dimitriou TG, Diamantidou V, Kotsovassilis C, Markoulatos P. Determination of human papillomavirus 16 physical status through E1/E6 and E2/E6 ratio analysis. *J Med Microbiol*. 2014;63(Pt 12):1716–23.
- Park JS, Hwang ES, Park SN, Ahn HK, Um SJ, Kim CJ, Kim SJ, Namkoong SE. Physical status and expression of HPV genes in cervical cancers. *Gynecol Oncol*. 1997;65(1):121–9.
- Nambaru L, Meenakumari B, Swaminathan R, Rajkumar T. Prognostic significance of HPV physical status and integration sites in cervical cancer. *Asian Pac J Cancer Prev*. 2009;10(3):355–60.
- Karbalaei Niya MH, Mobini Keshah M, Keshtmand G, Basi A, Rezvani H, Imanzade F, Panahi M, Rakhshani N. Integration rates of human papilloma virus genome in a molecular survey on cervical specimens among Iranian patients. *Eur J Cancer Prev*. 2019;28(6):537–43.
- Xu B, Chotewutmontri S, Wolf S, Klos U, Schmitz M, Durst M, Schwarz E. Multiplex Identification of Human Papillomavirus 16 DNA Integration Sites in Cervical Carcinomas. *PLoS ONE*. 2013;8(6): e66693.
- Akagi K, Li J, Broutian TR, Padilla-Nash H, Xiao W, Jiang B, Rocco JW, Teknos TN, Kumar B, Wangsa D, et al. Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability. *Genome Res*. 2014;24(2):185–99.
- Wentzensen N, Vinokurova S, von Knebel DM. Systematic review of genomic integration sites of human papillomavirus genomes in epithelial dysplasia and invasive cancer of the female lower genital tract. *Cancer Res*. 2004;64(11):3878–84.
- Doorbar J, Quint W, Banks L, Bravo IG, Stoler M, Broker TR, Stanley MA. The biology and life-cycle of human papillomaviruses. *Vaccine*. 2012;30(Suppl 5):F55–70.
- Arias-Pulido H, Peyton CL, Joste NE, Vargas H, Wheeler CM. Human papillomavirus type 16 integration in cervical carcinoma in situ and in invasive cervical cancer. *J Clin Microbiol*. 2006;44(5):1755–62.
- Boulet GA, Benoy IH, Depuydt CE, Horvath CA, Aerts M, Hens N, Ver-eecken AJ, Bogers JJ. Human papillomavirus 16 load and E2/E6 ratio in HPV16-positive women: biomarkers for cervical intraepithelial neoplasia >or=2 in a liquid-based cytology setting? *Cancer Epidemiol Biomarkers Prev*. 2009;18(11):2992–9.
- Gradissimo Oliveira A, Delgado C, Verdasca N, Pista A. Prognostic value of human papillomavirus types 16 and 18 DNA physical status in cervical intraepithelial neoplasia. *Clin Microbiol Infect*. 2013;19(10):E447–450.
- Tsakogiannis D, Gartzonika C, Levidiotou-Stefanou S, Markoulatos P. Molecular approaches for HPV genotyping and HPV-DNA physical status. *Expert Rev Mol Med*. 2017;19: e1.
- Matovina M, Sabol I, Grubisic G, Gasperov NM, Grce M. Identification of human papillomavirus type 16 integration sites in high-grade precancerous cervical lesions. *Gynecol Oncol*. 2009;113(1):120–7.
- Chandrani P, Kulkarni V, Iyer P, Upadhyay P, Chaubal R, Das P, Mulherkar R, Singh R, Dutt A. NGS-based approach to determine the presence of HPV and their sites of integration in human cancer genome. *Br J Cancer*. 2015;112(12):1958–65.
- Schmitz M, Driesch C, Jansen L, Runnebaum IB, Durst M. Non-random integration of the HPV genome in cervical cancer. *PLoS ONE*. 2012;7(6): e39632.
- Ferber MJ, Thorland EC, Brink AA, Rapp AK, Phillips LA, McGovern R, Gostout BS, Cheung TH, Chung TK, Fu WY, et al. Preferential integration of human papillomavirus type 18 near the c-myc locus in cervical carcinoma. *Oncogene*. 2003;22(46):7233–42.
- Thorland EC, Myers SL, Ding W, Smith DI. Common fragile sites are preferential targets for HPV16 integrations in cervical tumors. *Oncogene*. 2003;22(8):1225–37.

45. Thorland EC, Myers SL, Persing DH, Sarkar G, McGovern RM, Gostout BS, Smith DL. Human papillomavirus type 16 integrations in cervical tumors frequently occur in common fragile sites. *Cancer Res.* 2000;60(21):5916–21.
46. Tsakogiannis D, Gortsilas P, Kyriakopoulou Z, Ruether IG, Dimitriou TG, Orfanoudakis G, Markoulatos P. Sites of disruption within E1 and E2 genes of HPV16 and association with cervical dysplasia. *J Med Virol.* 2015;87(11):1973–80.
47. Li H, Yang Y, Zhang R, Cai Y, Yang X, Wang Z, Li Y, Cheng X, Ye X, Xiang Y, et al. Preferential sites for the integration and disruption of human papillomavirus 16 in cervical lesions. *J Clin Virol.* 2013;56(4):342–7.
48. Harden ME, Munger K. Human papillomavirus molecular biology. *Mutat Res Rev Mutat Res.* 2017;772:3–12.
49. Bose P, Brockton NT, Dort JC. Head and neck cancer: from anatomy to biology. *Int J Cancer.* 2013;133(9):2013–23.
50. Jeon S, Allen-Hoffmann BL, Lambert PF. Integration of human papillomavirus type 16 into the human genome correlates with a selective growth advantage of cells. *J Virol.* 1995;69(5):2989–97.
51. Balasubramaniam SD, Balakrishnan V, Oon CE, Kaur G. Key Molecular Events in Cervical Cancer Development. *Medicina (Kaunas).* 2019;55(7):384.
52. Adey A, Burton JN, Kitzman JO, Hiatt JB, Lewis AP, Martin BK, Qiu R, Lee C, Shendure J. The haplotype-resolved genome and epigenome of the aneuploid HeLa cancer cell line. *Nature.* 2013;500(7461):207–11.
53. Xu X, Han Z, Ruan Y, Liu M, Cao G, Li C, Li F. HPV16-LINC00393 Integration Alters Local 3D Genome Architecture in Cervical Cancer Cells. *Front Cell Infect Microbiol.* 2021;11: 785169.
54. Pennisi E. New technologies boost genome quality. *Science.* 2017;357(6346):10–1.
55. Zhao G, Zou C, Li K, Wang K, Li T, Gao L, Zhang X, Wang H, Yang Z, Liu X, et al. The *Aegilops tauschii* genome reveals multiple impacts of transposons. *Nat Plants.* 2017;3(12):946–55.
56. Gordon D, Huddleston J, Chaisson MJ, Hill CM, Kronenberg ZN, Munson KM, Malig M, Raja A, Fiddes I, Hillier LW, et al. Long-read sequence assembly of the gorilla genome. *Science.* 2016;352(6281):aae0344.
57. Zhou L, Qiu Q, Zhou Q, Li J, Yu M, Li K, Xu L, Ke X, Xu H, Lu B, et al. Long-read sequencing unveils high-resolution HPV integration and its oncogenic progression in cervical cancer. *Nat Commun.* 2022;13(1):2563.
58. Brancaccio RN, Robitaille A, Dutta S, Rollison DE, Tommasino M, Gheit T. MinION nanopore sequencing and assembly of a complete human papillomavirus genome. *J Virol Methods.* 2021;294: 114180.
59. Iden M, Tsai SW, Huang YW, Liu P, Xiao M, Flister MJ, Rader JS. Multi-omics mapping of human papillomavirus integration sites illuminates novel cervical cancer target genes. *Br J Cancer.* 2021;125(10):1408–19.
60. Yang S, Zhao Q, Tang L, Chen Z, Wu Z, Li K, Lin R, Chen Y, Ou D, Zhou L, et al. Whole Genome Assembly of Human Papillomavirus by Nanopore Long-Read Sequencing. *Front Genet.* 2021;12: 798608.
61. Zhuo Z, Rong W, Li H, Li Y, Luo X, Liu Y, Tang X, Zhang L, Su F, Cui H, et al. Long-read sequencing reveals the structural complexity of genomic integration of HBV DNA in hepatocellular carcinoma. *NPJ Genom Med.* 2021;6(1):84.
62. Liu M, Han Z, Zhi Y, Ruan Y, Cao G, Wang G, Xu X, Mu J, Kang J, Dai F, et al. Long-read sequencing reveals oncogenic mechanism of HPV-human fusion transcripts in cervical cancer. *Transl Res.* 2023;253:80–94.
63. Yang W, Liu Y, Dong R, Liu J, Lang J, Yang J, Wang W, Li J, Meng B, Tian G. Accurate Detection of HPV Integration Sites in Cervical Cancer Samples Using the Nanopore MinION Sequencer Without Error Correction. *Front Genet.* 2020;11:660.
64. Scott AJ, Chiang C, Hall IM. Structural variants are a major source of gene expression differences in humans and often affect multiple nearby genes. *Genome Res.* 2021;31(12):2249–57.
65. Arbyn M, Weiderpass E, Bruni L, de Sanjose S, Saraiya M, Ferlay J, Bray F. Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis. *Lancet Glob Health.* 2020;8(2):e191–203.
66. Tsakogiannis D, Bletsas M, Kyriakopoulou Z, Dimitriou TG, Kotsovassilis C, Panotopoulou E, Markoulatos P. Identification of rearranged sequences of HPV16 DNA in precancerous and cervical cancer cases. *Mol Cell Probes.* 2016;30(1):6–12.
67. el Awady MK, Kaplan JB, O'Brien SJ, Burk RD. Molecular analysis of integrated human papillomavirus 16 sequences in the cervical cancer cell line SiHa. *Virology.* 1987;159(2):389–98.
68. Baker CC, Phelps WC, Lindgren V, Braun MJ, Gonda MA, Howley PM. Structural and transcriptional analysis of human papillomavirus type 16 sequences in cervical carcinoma cell lines. *J Virol.* 1987;61(4):962–71.
69. Diao MK, Liu CY, Liu HW, Li JT, Li F, Mehryar MM, Wang YJ, Zhan SB, Zhou YB, Zhong RG, et al. Integrated HPV genomes tend to integrate in gene desert areas in the CaSki, HeLa, and SiHa cervical cancer cell lines. *Life Sci.* 2015;127:46–52.
70. Waggoner SE. Cervical cancer. *Lancet.* 2003;361(9376):2217–25.
71. Weirather JL, de Cesare M, Wang Y, Piazza P, Sebastiano V, Wang XJ, Buck D, Au KF. Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Res.* 2017;6:100.
72. Logsdon GA, Vollger MR, Eichler EE. Long-read human genome sequencing and its applications. *Nat Rev Genet.* 2020;21(10):597–614.
73. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754–60.
74. Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, Schatz MC. Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods.* 2018;15(6):461–8.
75. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. Genome Project Data Processing S: The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078–9.
76. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010;38(16):e164.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.