



# Characteristics of HPV integration in cervical adenocarcinoma and squamous carcinoma

Yuxin Bi<sup>1,2</sup> · Junbo Hu<sup>4</sup> · Ling Zeng<sup>1,5</sup> · Gang Chen<sup>3</sup> · Hongning Cai<sup>1,2</sup> · Huang Cao<sup>1,2,5</sup> · Quanfu Ma<sup>1,2</sup> · Xufeng Wu<sup>1,2</sup>

Received: 13 August 2023 / Accepted: 25 October 2023 / Published online: 15 November 2023  
© The Author(s) 2023

## Abstract

**Purpose** HPV integration usually occurs in HPV-related cancer, and is the main cause of cancer. But the carcinogenic mechanism of HPV integration is unclear. The study aims to provide a theoretical basis for understanding the pathogenesis of cervical adenocarcinoma (AC) and cervical squamous carcinoma (SCC).

**Methods** We used HPV capture sequencing to obtain HPV integration sites in AC and SCC, and analyzed cytobands, distribution of genetic and genomic elements, identified integration hotspot genes, clinicopathological parameters, breakpoints of HPV16 and performed pathway analysis. Then we conducted immunohistochemical (IHC) assay to preliminarily verify the expression of most frequently integrated genes in AC, STARD3 and ERBB2.

**Results** The results revealed that the most frequently observed integrated cytoband was 17q12 in AC and 21p11.2 in SCC, respectively. The breakpoints in both AC and SCC were more tended to occur within gene regions, compared to intergenetic regions. Compared to SCC samples, AC samples had a higher prevalence of genomic elements. In AC, HPV integration has no significantly difference with clinicopathological parameters, but in SCC integration correlated with differentiation ( $P < 0.05$ ). Breakpoints of HPV in SCC located in LCR more frequently compared to AC, which destroyed the activation of promoter p97. Hotspot genes of HPV integration were STARD3 and ERBB2 in AC, and RNA45S rDNA and MIR3648-1 in SCC, respectively. Meanwhile, we preliminarily proved that the expression of STARD3 and ERBB2, the most frequently integrated genes, would increase after integration.

**Conclusion** These results suggested that HPV may utilize the powerful hosts' promoters to express viral oncogenes and overexpression of viral oncogenes plays a significant role in the carcinogenesis of SCC. In AC, HPV integration may affect hosts' oncogenes, and the dysregulation of oncogenes may primarily contribute to progression of AC.

**Keywords** HPV integration · HPV capture sequencing · Cervical adenocarcinoma · Cervical squamous carcinoma · Characterization of integration · Carcinogenesis of HPV integration

## Introduction

Cervical cancer is a common gynecological tumor (Li et al. 2019), with approximately 604,000 new cases and 342,000 deaths reported worldwide in 2020 (Sung et al. 2021). Squamous carcinoma (SCC) and adenocarcinoma (AC) are the two main pathological types of cervical cancer. With widespread cervical cancer screening, the mortality of SCC has significantly decreased. However, the mortality of AC remains high because of low detection sensitivity, young age of onset, low survival rate, and high rate of metastasis (Dong Hong et al. 2010; Irie et al. 2000; Shimada et al. 2006).

To our knowledge, persistence of HPV infection is the main cause of cervical cancer. HPV is a small nonenveloped double-stranded DNA virus (Hoppe-Seyler et al. 2018) that

✉ Xufeng Wu  
zwuxufeng@163.com

<sup>1</sup> Maternal and Child Health Hospital of Hubei Province, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

<sup>2</sup> Hubei Clinical Medical Research Center for Gynecologic Malignancy, Wuhan, China

<sup>3</sup> Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China

<sup>4</sup> Department of Pathology, Maternal and Child Health Hospital of Hubei Province, Wuhan, China

<sup>5</sup> Hubei Provincial Center for Medical Genetics, Wuhan, China

establishes its genome as episomes in proliferating basal cells (Bedell et al. 1991). The expression of viral oncoproteins E6 and E7 promotes the progression of cervical cancer. Although HPV integration is critical, the carcinogenic mechanism of integration is unclear. Studies have reported that HPV integration sites tend to be located in repetitive regions of the host genome, with some sites near cancer-relevant genes (Li et al. 2013), and identified integration hotspot genes, including POU5F1B, FHIT, and KLF12 (Hu et al. 2015). Kamal et al. studied the pattern of HPV integration in SCC and AC and reported that HPV integration signatures are not associated with histological subtypes (Kamal et al. 2021). Because the number of AC cases is limited, most studies on HPV integration have focused on SCC (Cancer Genome Atlas Research et al. 2017; Hu et al. 2015; Ojesina et al. 2014; Li et al. 2013); however, none have evaluated the discrepancy in integration features between SCC and AC. We systematically and comprehensively studied the differences of HPV integration characteristics between SCC and AC. The aim of this study is to understand HPV integration-mediated carcinogenesis in AC and SCC, and to provide a useful information for researches on treatment and screen of cervical cancer.

## Materials and methods

### Sample collection

A total of 66 AC samples were collected from the Maternal and Child Health Hospital of Hubei Province between 2015 and 2020. The clinicopathological parameters included age, HPV type, FIGO stage, differentiation, and lymph node metastasis.

### HPV capture sequencing

HPV capture sequencing was employed to identify breakpoints of the virus and host. The detailed method was as described in a previous study (Hu et al. 2023). Integrated sites with reads  $\geq 7$  were deemed valid integration sites. The sequencing data for 95 cases of SCC, sequenced using the same method, were provided by Tongji Hospital, which is affiliated with the Tongji Medical College of Huazhong University of Science and Technology (Fan et al. 2023). Among 95 SCC samples, 6 HPV (+) tumors were not detected HPV integration and 2 SCC were HPV negative (Fan et al. 2023).

### Analysis of integration sites

The breakpoints were annotated using the latest version of ANNOVAR in hg38 coordinates (Wang et al. 2010). Annotated genes with frequency  $\geq 3$  were defined as HPV

integration hotspots. An integration site was confirmed to be adjacent to a genomic element when the element overlapped with the flanking region (200 bp) (Hu et al. 2015). Cytobands, transcription factor binding sites (TFBSs), repetitive regions, and CpG islands were downloaded from UCSC (<http://hgdownload.cse.ucsc.edu/goldenPath/hg38/database>). DNaseI hypersensitivity sites (DHSs), methylated or acetylated histone binding sites, and chromatin state of HeLa-S3 cells were downloaded from ENCODE (Bodelon et al. 2016). Functional annotation analysis of breakpoints was performed using R packages, and a heatmap was plotted using an online platform for data analysis and visualization (<https://www.bioinformatics.com.cn>). Protein–protein interaction network analysis (PPI) was constructed using the online tool STRING (<https://www.string-db.org/>), by setting medium confidence at 0.400 and excluding irrelevant genes (Zeng et al. 2021).

### Generation of random sites

Using R package, 1000 random sites within the human chromosome were generated, serving as a random control.

### Immunohistochemical (IHC)

To verify the impact of HPV integration on expression of most frequently integrated genes, we conducted IHC assay. The detailed process of IHC staining and analytical method was as described in this paper (TAN Yao et al. 2022). We chose five samples with these genes integrated as integrated group, and five samples with no same genes integrated as control group. We observed the tissue sections under microscope and selected five images per slide, and images were taken at 400 $\times$  magnification using the same light intensity and exposure time. Image J was used to analyze the integrated optical density (IOD) value of each image.

### Statistical analysis

SPSS 25.0 was used for statistical analysis. *P*-values were calculated using the  $\chi^2$  test and corrected using Fisher's exact test.

## Results

### Integration cytobands

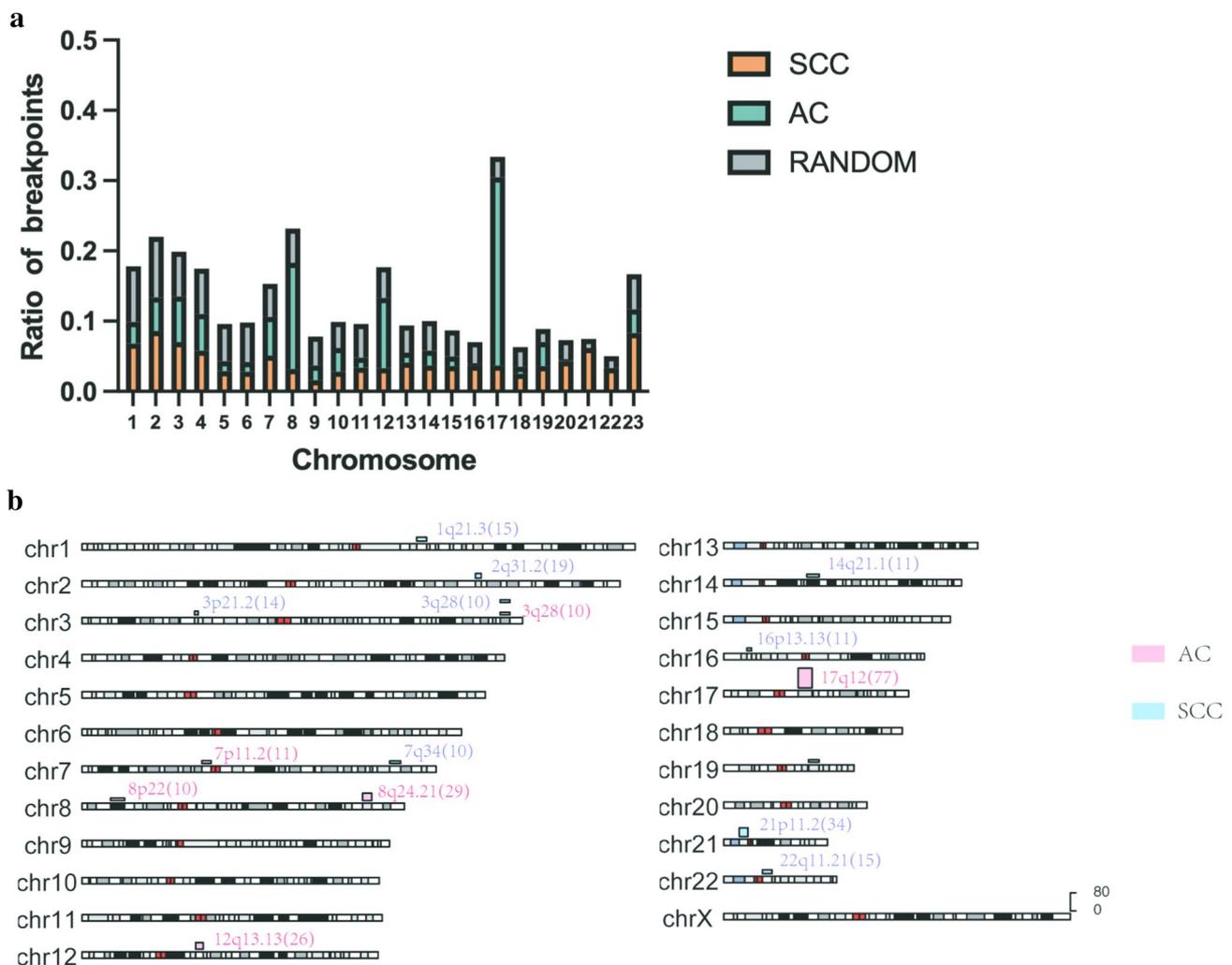
Previous studies have shown that viral integration in cervical cancer frequently occurs in some specific cytobands, such as 8q24.21 and 3q28 (Bodelon et al. 2016). Therefore, we investigated cytobands with high number (frequency  $\geq 10$ ) of observed integration events in AC and SCC to understand

the impact of HPV integration on carcinogenesis. In this study, we found integration events occur in all chromosomes (Fig. 1a). The results indicated that the integration points of AC tended to be located in chromosome 8, whereas in SCC, breakpoints tended to locate in chromosome 21. In AC, total 95 cytobands that not defined and 75 sub-bands were targeted, whereas in SCC 247 cytobands and 46 sub-bands were targeted. In AC, cytobands with a significant number of integration events (frequency  $\geq 10$ ) included 17q12 (26.5%, 77/291), 8q24.21 (10%, 29/291), 12q13.13 (9%, 26/291), 7p11.2 (3.8%, 11/291), 3q28 (3.4%, 10/291), 8p22 (3.4%, 10/291). In SCC, the most frequently integrated cytobands were 21p11.2 (4.5%, 34/760), 2q31.2 (2.5%, 19/760), 22q11.21 (2.0%, 15/760), 1q21.3 (2.0%, 15/760), 3p21.2 (1.8%, 14/760), 16p13.13 (1.4%, 11/760), 14q21.1 (1.4%, 11/760), 3q28 (1.3%, 10/760), 7q34 (1.3%, 10/760), 19q13.2

(1.3%, 10/760) (Fig. 1b). The percentage represented the ratio of breakpoints of all integration sites of all the cases. This revealed that HPV integration showed a different preference for specific cytoband in AC and SCC.

### HPV integration hotspots in AC and SCC

Many studies agreed that HPV integration showed a strong tendency to be located in some hotspot genes (Hu et al. 2015; Bodelon et al. 2016). Identifying the different HPV integration hotspot genes in AC and SCC can help clarify the pathogenesis of cervical cancer. In AC, 12 hotspot genes (frequency  $\geq 3$ ) were detected. The most commonly integrated genes were STARD3 (17%, 9/53), ERBB2 (17%, 9/53), TNS2 (7.5%, 4/53), TEX41 (7.5%, 4/53), and PVT1 (7.5%, 4/53). In SCC, 10 hotspot genes were detected,



**Fig. 1** Distribution of breakpoints in human chromosomes. **a** Distribution of breakpoints in human chromosomes. The Y-axis represents the ratio of breakpoints, whereas the X-axis denotes different chromosomes. **b** Distribution of breakpoints in cytobands with

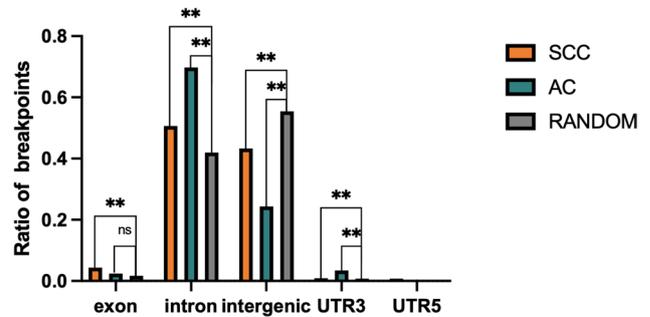
frequency  $\geq 10$ . The height of the histogram corresponds to the frequency of integration in the cytobands. Red histogram represents AC, whereas blue histogram represents SCC

including RNA45S (12.6%, 11/87), MIR3648-1 (6.9%, 6/87), and FHIT (4.6%, 4/87) (Fig. 2). We found that most hotspot genes were coding genes (8/12) in AC and noncoding genes (8/15) in SCC. The results indicated that HPV tended to integrate in different functional genes in AC and SCC.

### Distribution of breakpoints in exonic, intronic, and intergenic regions

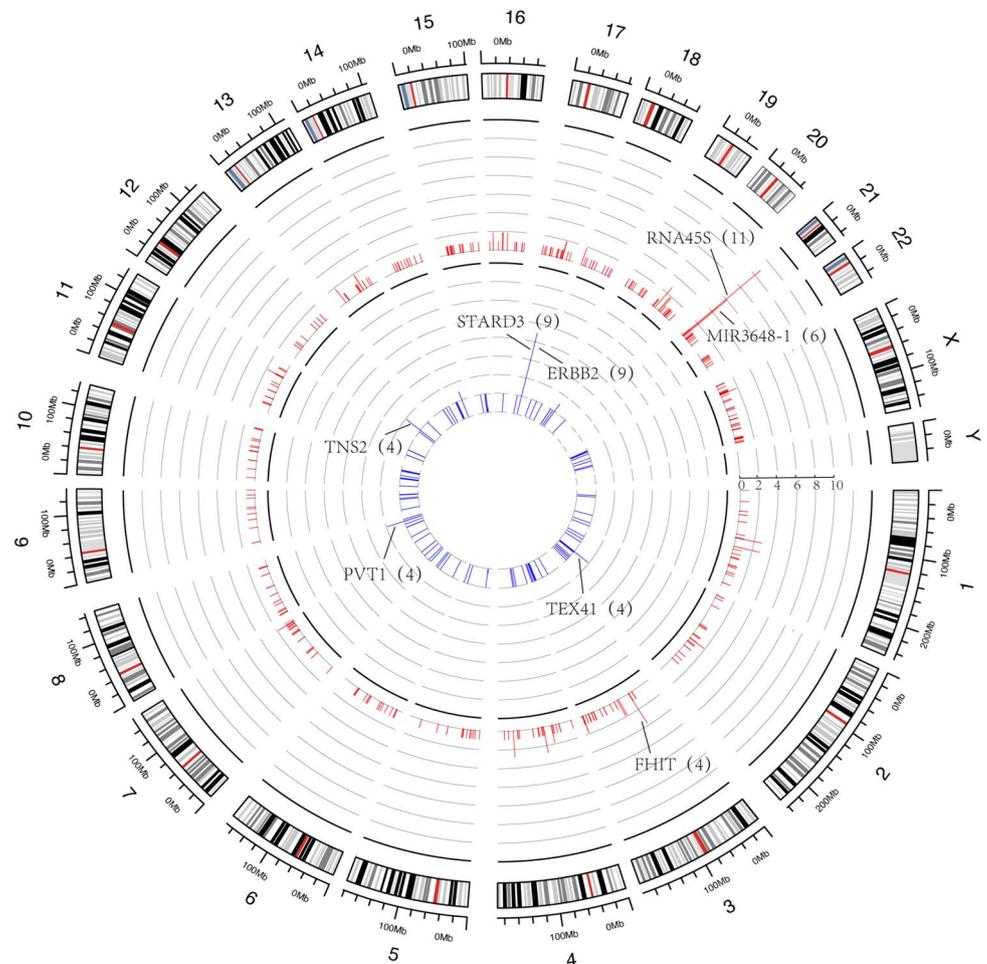
We have explored the tendency of HPV integration for genes, but the locational relationship between integration sites and known human genes is unclear. In AC, we identified 291 valid integration sites, including 7 (2.4%) in exons, 203 (69.8%) in introns, 71 (24.4%) in intergenic regions, and 10 (3.4%) in UTR'3. In SCC, we identified 760 effective integration sites, including 33 (4.3%) in exons, 385 (50.7%) in introns, 329 (43.3%) in intergenic regions, 7 (0.9%) in UTR'3 and 6 (0.9%) in UTR'5. The results indicated that compared to the randomized control, both genomes of AC, and SCC showed a tendency to

be integrated within genes, with integration sites in SCC more likely to be in exons (Fig. 3).



**Fig. 3** Distribution of integration sites in exonic, intronic, intergenic regions, 3'UTR and 5'UTR. The random integrated ratio of each region is counted according to the random distribution of breakpoints in the human genome (\*\* $P < 0.01$ )

**Fig. 2** Frequency and distribution of integration breakpoints in the human genome. Each line represents the location of HPV integration in human chromosomes, and the length of the line represents the frequency of integration. In the inner circles, the red bars depict the frequency of SCC samples, whereas the blue bars depict the frequency of AC samples

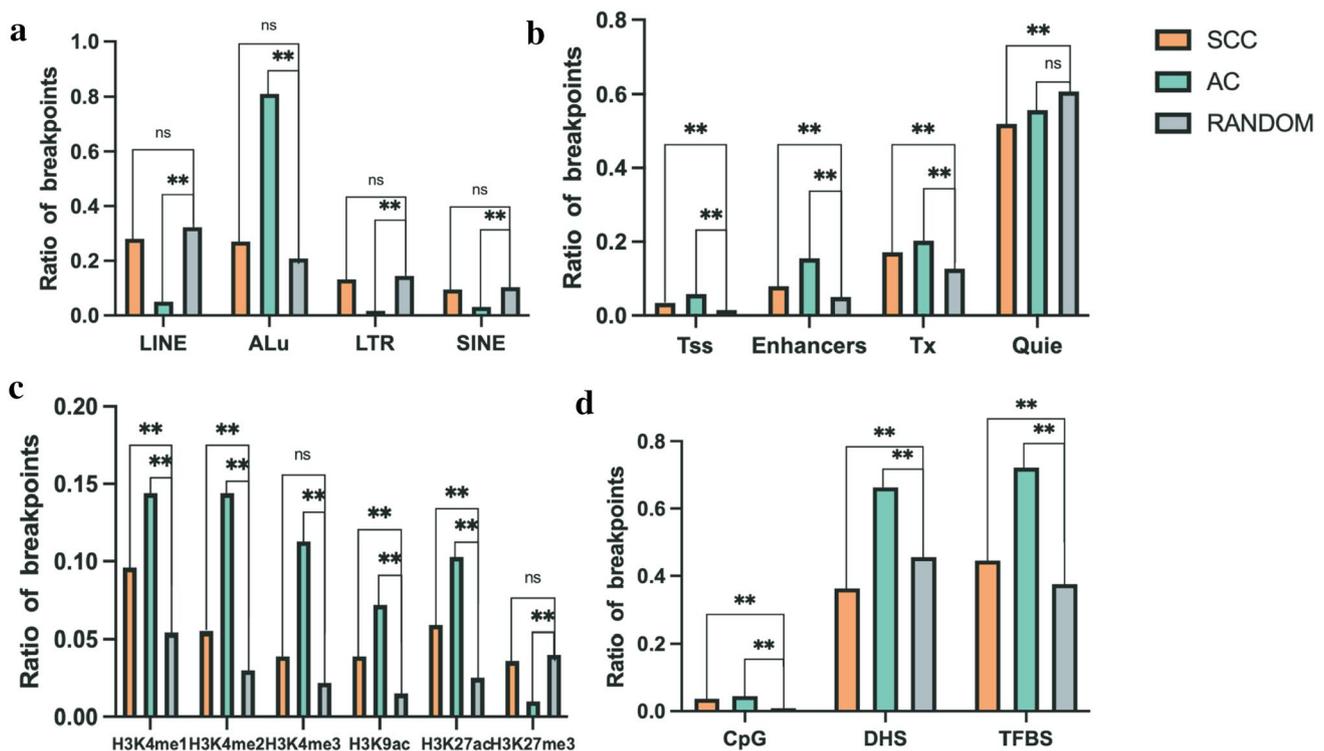


## Distribution of breakpoints in genomic elements

Next, we examined the distribution of breakpoints for AC and SCC in several genomic features, including repetitive elements, chromatin genome segmentations, methylated or acetylated histone binding sites, CpG enriched regions, and DHSs. This helps understand common molecular features that may lead to HPV integration into the human genome (Bodelon et al. 2016) and cause the discrepancy between AC and SCC. ANNOVAR was used to perform region-based annotation (Wang et al. 2010). Repetitive elements are necessary for accurate genome replication and transmission to progeny cells (Shapiro and von Sternberg 2005). Owing to the same integration sites potentially annotated by the different region-element, we calculated the number of certain elements among the amount of all annotated elements. In this study, the integration sites in AC tended to locate in Alu element (80.9%, 1184/1464). In AC, breakpoints occurred less frequently in LINE (Long interspersed nuclear elements; 5.1%, 75/1464), LTR (Long terminal repeat elements; 1.7%, 25/1464) and SINE (Short interspersed nuclear elements; 3.1%, 46/1464) compared to RANDOM (\*\* $P < 0.01$ , Fig. 4a). In SCC, HPV integration did not represent any tendency in repetitive regions.

Open chromatin regions, which are associated with transcriptional regulation, provide insights into genome function. Based on ENCODE chromatin genome segmentation information of HeLa-S3, the integration points of AC tended to located more in Enhancers (15.5%) than RANDOM (\*\* $P < 0.01$ ). Compared to the randomized control, HPV tended to integrate in transcription start sites (Tss) flanking regions (frequency rate = 3.4% and 5.8% for SCC and AC, respectively) and Tx (frequency rate = 17.2% and 20.3% for SCC and AC, respectively) (Fig. 4b).

Histone modifications, particularly acetylation and methylation on specific amino acid residues, have been implicated in gene expression regulation (Bannister and Kouzarides 2011; Greer and Shi 2012). We analyzed the association between HPV integration and histone markers binding regions. The results showed that integration events in AC occurred more frequently than RANDOM in binding sites of H3K4me1 (14.4%), H3K4me2 (14.4%), H3K4me3 (11.3%), H3K9ac (7.2%) and H3K27ac (10.3%), but not H3K27me3 (1%) (Fig. 4c). Additionally, we examined other important genomic features, such as CpG enriched regions, TFBSs, and DHSs. CpG islands help identify promoters. TFBSs play a crucial role in regulating transcription (Bartlett et al. 2017). DHSs represent the susceptibility of chromatin (Lu and Richardson 2004). In our study, integration events in AC

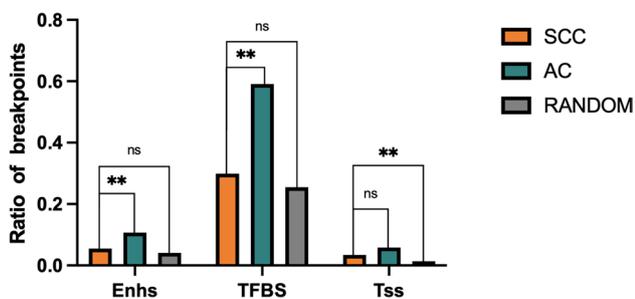


**Fig. 4** Distribution of integration sites in different genomic elements. The Y-axis represents the ratio of breakpoints, whereas the X-axis denotes different genomic elements. **a** Repetitive elements, **b** chroma-

tin genome segmentation, **c** regions binding with histone markers, **d** and CpG enriched regions, DHS, and TFBS. (\*\* $P < 0.01$ )

showed a preference for DHSs (66.3%) and TFBSs (72.2%). This indicated that in AC, HPV has a higher propensity for integration in regulatory regions compared to SCC.

To analyze the relationship between these genomics elements and HPV integration sites with great confidence, we examined the co-occurrence of integration sites in DHSs and other genomic elements in AC and SCC (Fig. 5). We found that in AC, breakpoints more frequently co-occurred in DHSs and Enhancers (10.7%) or TFBS (59.1%) compared to SCC (\*\**P* < 0.01). This suggests a strong association between breakpoints in AC and these elements. Both AC and SCC breakpoints showed a tendency to co-occurrence in Tss (Frequency rate = 3.4% and 5.8% for SCC and AC, respectively), but the difference between AC and SCC was not statistically significant.



**Fig. 5** The relationship between these genomics elements and HPV integration sites. Breakpoints represent integration sites located both in DHSs and corresponding chromatin states. (\*\**P* < 0.01)

### Clinical characterization

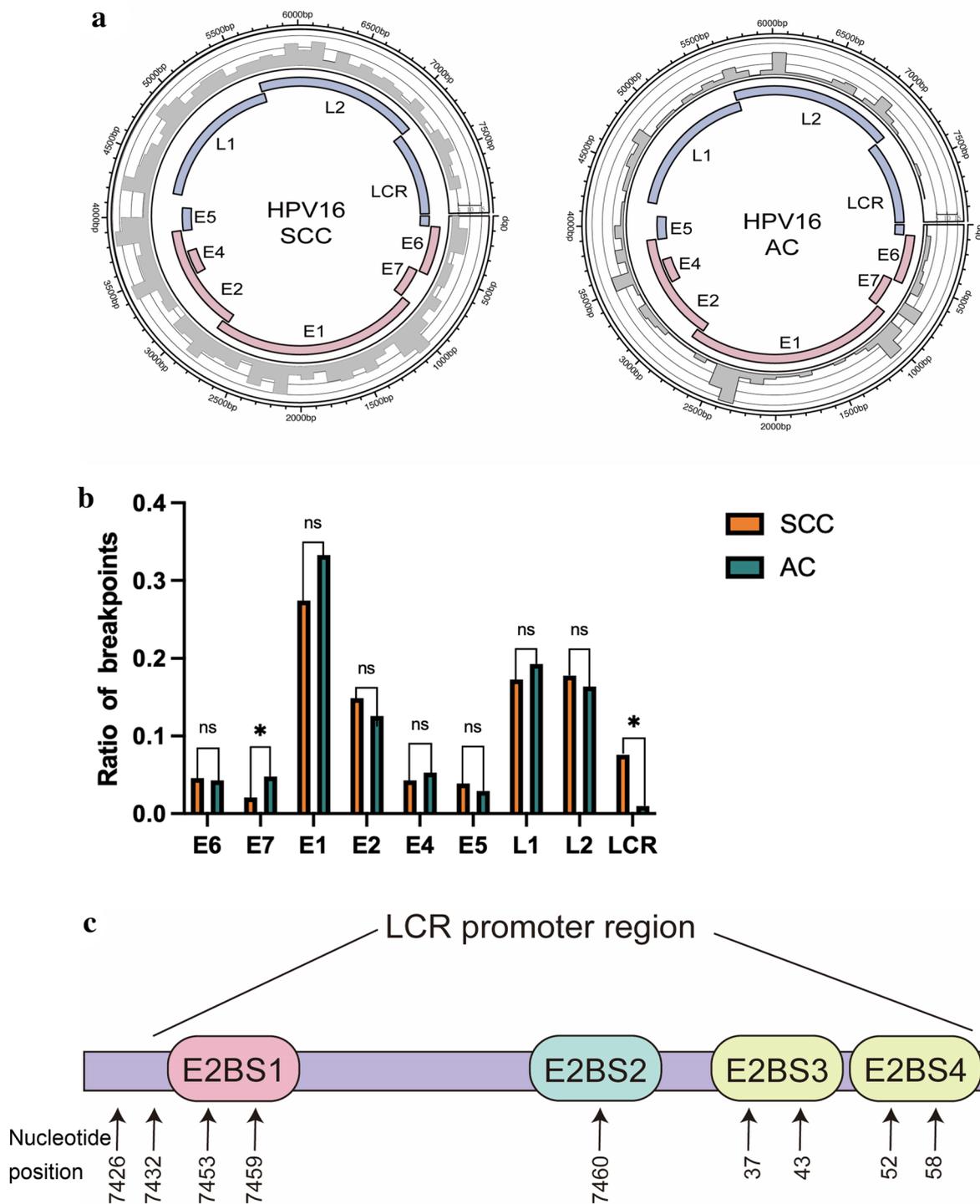
A total of 66 AC samples were collected, of which 11 had invalid integration sites (frequency ≤ 3). HPV integration didn't occur in two samples; therefore, 53 AC samples had valid breakpoints. A total of 95 SCC samples were collected, of which 87 were detected valid integration. 6 HPV (+) tumors were not detected HPV integration and 2 SCC were HPV negative. The results revealed that in AC HPV integration had no significant difference with common clinical parameters, including age, differentiation, stage, HPV type and lymph metastasis (Table 1). In SCC, HPV integration only correlated with differentiation and the corresponding *p*-value was denoted by bold (*P* = 0.026).

### Breakpoints of HPV16 in SCC and AC

On the other hand, we also studied the relationship between different genomic elements features of integration sites and HPV type. We found that there was no significant relation (supplementary Table S1). Then we analyzed the HPV breakpoints of AC and SCC. The results revealed that HPV16 breakpoints of AC intended to be located in E7 compared to that of SCC. Meanwhile, breakpoints of SCC were prone to be located in LCR compared to that of AC (Fig. 6a, b). Further we found that in SCC breakpoints of HPV16 tended to be located in EBS1 and promoter regions (nucleotide position: 7432–116) compared to AC (Fig. 6c). The percentage of breakpoints in this region among all breakpoints

**Table 1** Relationship between clinicopathological parameters of AC and SCC, and HPV integration

Variable	Valid HPV integration of AC				<i>P</i> -value	Valid HPV integration of SCC				
	Negative		Positive			Negative		Positive		<i>P</i> -value
	No	%	No	%		No	%	No	%	
<i>Age</i>					0.115					0.522
< 40	8	32.0%	17	68.0%		1	7.7%	12	92.3%	
40–49	2	8.3%	22	91.7%		1	3.4%	28	96.6%	
> = 50	3	17.6%	14	82.4%		6	11.3%	47	88.7%	
<i>Differentiation</i>					0.144					<b>0.026</b>
Well	2	9.1%	20	90.9%		3	33.3%	6	66.7%	
Moderately or poorly	2	40%	3	60%		5	5.8%	81	94.2%	
<i>Stage</i>					1.000					0.273
CIS or stage 1	9	20%	36	20%		6	12.8%	41	87.2%	
Stage 2 +	1	80%	4	80%		2	4.8%	40	95.2%	
<i>Type of HPV</i>					0.881					1.000
HPV16	6	16.7%	30	83.3%		5	7.4%	63	92.6%	
HPV18	4	16.7%	20	83.3%		0	0.0%	6	100%	
Non-HPV16/18	1	25.0%	3	75.0%		1	5.3%	18	94.7%	
<i>Lymph metastasis</i>					0.171					0.676
Positive	1	100%	0	0%		1	4.5%	21	95.5%	
Negative	5	14.7%	29	85.3%		7	9.6%	66	90.4%	



**Fig. 6 a** Distribution of integration breakpoints in the HPV16 genome. Histograms (grey) of frequency of breakpoints in the samples are constructed for 100-bp intervals. Histogram axis units represent numbers of breakpoints, and outer DNA numbering is given in bases. HPV genes with different functions are colored. **b** Comparison of SCC and AC breakpoints ratio in the HPV16 genome. The Y-axis

represents the ratio of HPV breakpoints, while the X-axis represents HPV16 genome. P-values are calculated by  $\chi^2$  test. **c** We represent the E2 binding sites in the HPV16 long control region (LCR). E2BS1 and promoter region covered from 7432 to 116 (Chaiwongkot et al. 2013)

that were located in LCR in SCC and AC samples were respectively 78.4% (40/51) and 2.9% (1/34) (\*\* $P < 0.01$ ). The breakpoints of HPV16 were shown in the supplementary material.

### Pathway analysis

To clarify the function of genes affected by HPV integration in AC and SCC, we used GO Pathway and PPI network analysis. GO analysis of hotspot genes revealed significant differences in enriched pathways between AC and SCC (Fig. 7a). Transmembrane receptor protein kinase pathway was enriched for AC; cell junction and cell–cell adhesion pathway were enriched for SCC. These results revealed that there was a significant difference between the enrichment pathway of AC and SCC. To understand the interaction of corresponding genes, the PPI network was constructed. Because we found no significant interaction between hotspot genes (frequency  $\geq 3$ ) of SCC, the picture is not shown here. 150 genes affected by HPV integration in AC samples were analyzed, and 115 genes were included in the network after the elimination of disconnected genes. The network is shown in Fig. 7b. The results indicated that in AC, genes ERBB2 and MYC were closely related to other affected genes by integration.

### Immunohistochemical (IHC) staining

To preliminarily verify the expression of most integrated genes in AC, we conducted IHC assay. The samples with ERBB2 integrated were the same as those with STARD3 integrated. ERBB2 is mainly expressed in cytomembrane and STARD3 is mainly existed in cytoplasm. Among five integrated samples, histocytes of one sample were most strongly stained with ERBB2 and STARD3 simultaneously compared to non-integrated samples (Fig. 8a). We used the integrated optical density (IOD) to quantize the expression of ERBB2 and STARD3. The average IOD of ERBB2, respectively, was  $7.2 \times 10^5$  and  $7.6 \times 10^6$  in integrated group and non-integrated group. The average IOD of STARD3, respectively, was  $7.6 \times 10^5$  and  $2.1 \times 10^5$  respectively. The results revealed that the expression of integrated genes was higher compared to non-integrated group ( $P < 0.01$ ) (Fig. 8b).

### Discussion

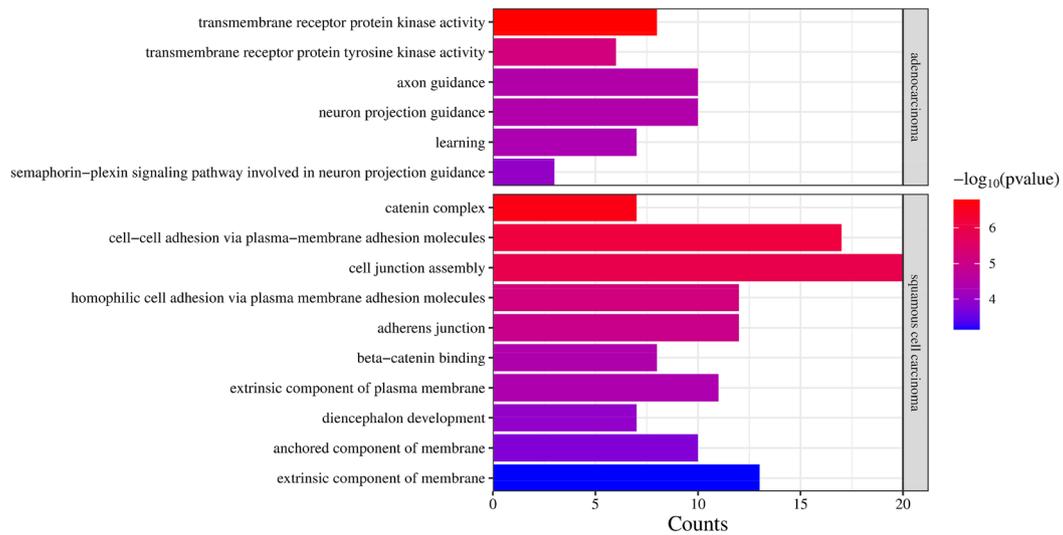
To our knowledge, HPV is the main driver of cervical cancer and plays a significant role in the progression of cancers through various mechanisms. First, expression of the E6 and E7 oncogenes, that target p53 and Rb family proteins respectively, and lead to their degradation, results in uncontrolled

proliferation (Vande Pol and Klingelhut [2013](#); Roman and Munger [2013](#)). In addition, HPV integration induces structural alterations and instability in the host genome (Porter and Marra [2022](#)). These alterations may include gene amplification, deletion, extrachromosomal circular DNA formation, and focal rearrangement (Porter and Marra [2022](#)). This study provides insights into HPV integration-mediated carcinogenesis in cervical cancer by analyzing the characteristics of HPV integration in SCC and AC from different genetic perspectives.

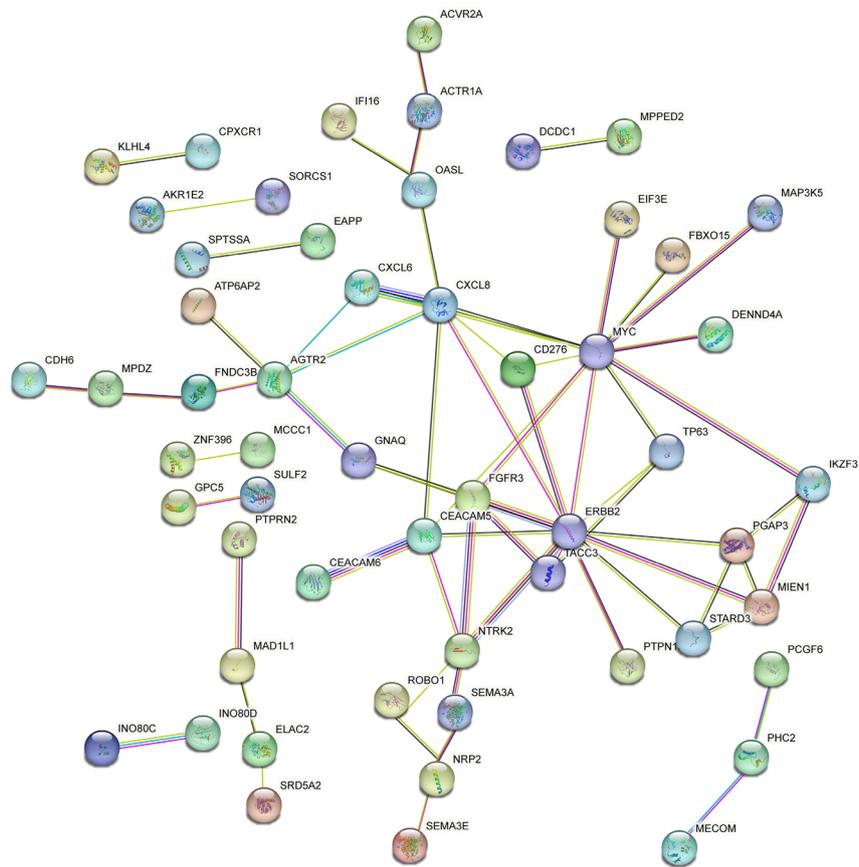
We examined the cytobands of HPV integration in AC and SCC samples. In AC, 17q12 (26.5%) was the most frequently integrated cytoband in AC in this study. Consistently, Hu supported the association between mutations in cytoband 17q12 and cervical cancer (Shi et al. [2013](#)). Within cytoband 17q12, ERBB2 has been recognized as a potential target for treating cervical cancer (Oh et al. [2015](#)). In this study, we found some sites were integrated frequently, such as chr17:39,695,542 (ERBB2) and chr17:39,647,045 (STARD3). Furthermore, we found that integration mainly occurs in intron of ERBB2, which is consistent with the results that integration sites tended to be located within introns. Martina Schmitz found that HPV integration sites were located within the intron sequences of known genes, and insertional mutagenesis could influence the function of genes (Schmitz et al. [2012](#)).

In addition to 17q12 (26.5%), 8q24.21 (10%) was the second most frequently integrated bands. This result agrees with previous studies (Hu et al. [2015](#); Bodelon et al. [2016](#); Shi et al. [2013](#)). Among 28 integration sites on band 8q24.21, 8 were located within a distance of 2 kb to 460 kb upstream of the MYC gene and 20 were within a distance of 13 kb to 366 kb downstream of the MYC gene. These distances represent exceedingly short regions in genomic terms. M Peter had proposed that viral and MYC sequences were co-amplified in an amplicon between less than 50 and 800 kb in size (Peter et al. [2006](#)). The amplification of the MYC oncogene was shown to significantly associate with heavily-rearranged extrachromosomal circular DNA (ecDNA) amplicons, which contained distal ( $> 1$  Mb) amplified segments (Kim et al. [2020](#)). So, we believed that the ecDNA amplicon could simultaneously promote viral oncogene and MYC gene co-amplifying. This may play an important role in carcinogenesis of HPV.

21p11.2 (4.5%) was the most frequently integrated cytoband in SCC. In this cytoband, genes encoding 45S ribosomal RNA (RNA45S rDNA), which harbor strong promoters, are abundant. Viral sequences could utilize hosts' promoters to express its proteins. Inagaki proposed that two distinct cDNA clones of the HPV type 18 transcript are present in HeLa cells (Inagaki et al. [1988](#)). These clones contain the viral genes E6 and E7, as well as a shared human sequence located at chr8:127,228,560 -127,229,294 (Inagaki et al.



**a**

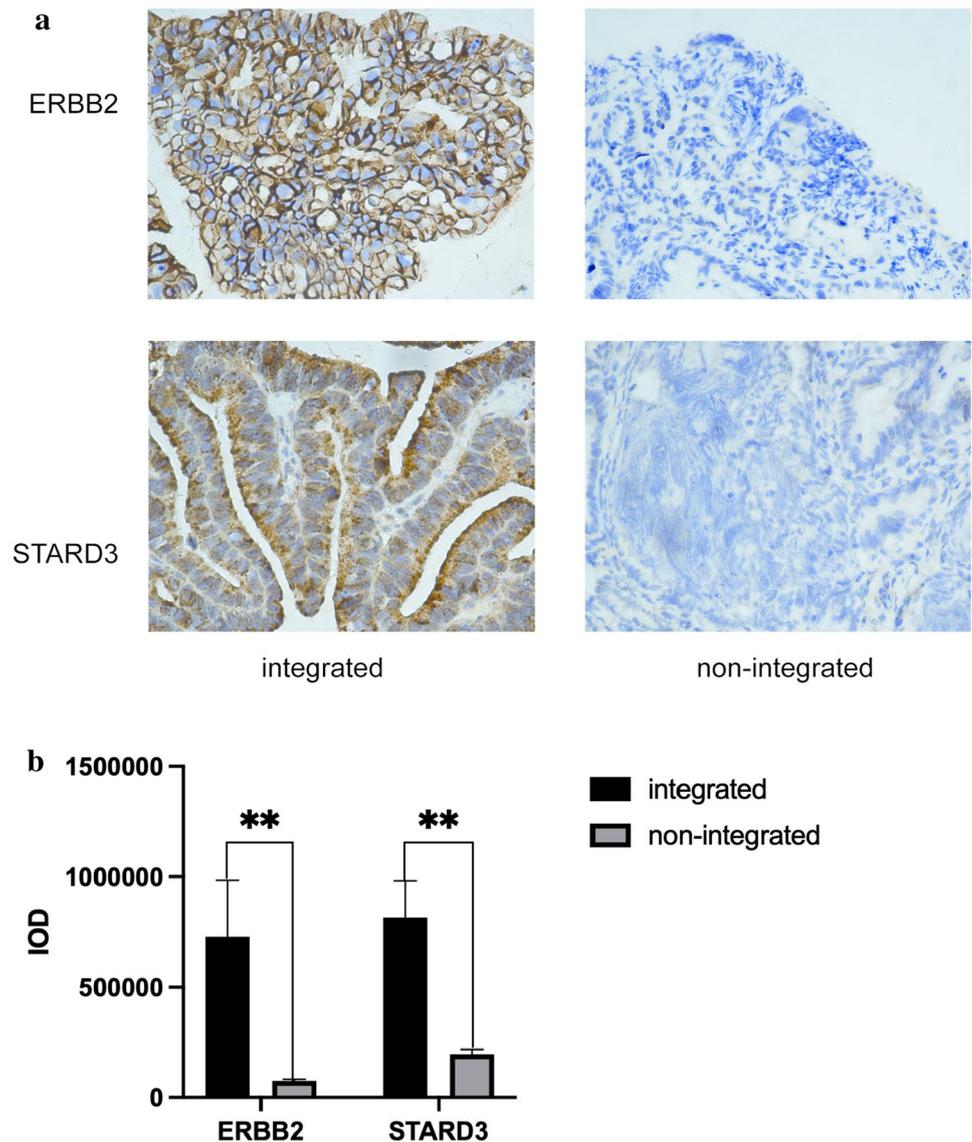


**b**

**Fig. 7 a** GO enrichment pathway. The Y-axis represents enrichment pathways, whereas the X-axis represents counts of genes enriched for the corresponding GO terms. The color of the bars indicates *P*-values, with “red” indicating a significant enrichment of genes in

the respective GO pathway. **b** PPI network for AC. The network was exported and visualized using online tool STRING. (<https://www.string-db.org/>)

**Fig. 8** **a** From left to right, the IHC images of integrated group and negative group in sequence are represented. **b** The histogram represents the integrated optical density (IOD) of integrated group and non-integrated group. The bars in the picture represents the standard error (SE)



1988). They proposed that HPV 18 integrates into the lncRNA and utilizes its promoter to express the viral oncoprotein E6 and E7 in HeLa cells (Inagaki et al. 1988). In our study, viral sequence integrated in 21p11.2 cytoband could likewise utilize RNA45S rDNA' promoters to express viral proteins in SCC.

Another cytoband 3q28 in SCC has been reported in previous study (Bodelon et al. 2016). In 3q28, mutations in TP63, a member of the p53 family, are associated with various cancers: notably, lung and bladder cancers (Lu et al. 2018; Wang et al. 2016). Min Liu hypothesized that highly expressed HPV-human fusion transcripts could promote overexpression of E6\**I* and E7 and inhibit the transcription of tumor suppressor genes TP63 and P3H2 (Liu et al. 2023). In our study, 3 breakpoints were located in intergenic regions between TP63 and P3H2. Therefore, this indicated that HPV-human fusion transcripts may play an important

role in carcinogenesis of SCC, by promoting overexpression of viral oncoproteins and inhibiting the transcription of tumor suppressor genes. In conclusion, we proposed that HPV integration in different cytobands promotes the progression of cancers through different mechanisms and shows a tendency for integration into specific cytobands.

We identified integration sites within regions with various elements, such as repetitive elements, open chromatin (DHSs), and chromatin genome segmentation (chromatin 18-state of HeLa-S3). The findings revealed that breakpoints in AC tended to locate in Alu elements, Enhancers, DHSs, TFBSs and H3K4 methylated histone binding sites compared with that in SCC. Among these elements, DHSs serve as a global measure of chromatin accessibility (Dao et al. 2021), providing insights into the functional elements within the genome. We identified Enh, TFBS, and Tss with great confidence by intersecting DHSs with chromatin state

regions identified in the ENCODE project (Fig. 5). This suggested that in AC, HPV has a higher propensity for integration in regulatory regions, potentially influencing host gene expression and cellular processes.

HPV integration at recurrent loci may provide selection advantage to host cells (Schmitz et al. 2012), leading to the recurrence of hotspot genes. We found that hotspots were coding genes (8/12) in AC and noncoding genes (8/15) in SCC. In AC samples, ERBB2 (9) and STARD3 (9) were the most integrated genes. The mutation rate of ERBB2 was reported to be significantly higher in AC than SCC (Xiang et al. 2018). Moreover, the positive rate of ERBB2 protein in AC was more than one-half (Shi et al. 2021). Further through GO analysis, we found that integrated genes in AC enriched in transmembrane receptor protein kinase pathway, and that of SCC enriched in cell junction and cell–cell adhesion pathway. Through the PPI network for AC, we found ERBB2 was closely related to other genes affected by HPV integration. STARD3 (9) is a member of the lipid transporter subfamily and plays a critical role in the synthesis and maintenance of cholesterol balance (Wilhelm et al. 2017). Studies on STARD3 have focused on gastric carcinoma and breast cancer (Wilhelm et al. 2017; Lodi et al. 2023); however, studies on cervical cancer are lacking. Previous studies have shown that STARD3 co-amplified and co-expressed with HER2 in breast cancer (Bieche et al. 1996; Tomasetto et al. 1995). The reduction of STARD3 expression in HER2-positive cancer cell lines inhibit their growth (Kao and Pollack 2006; Sahlberg et al. 2013). In this study, STARD3 was related to ERBB2 closely in the PPI network. Through IHC assay, we found that ERBB2 and STARD3 were most strongly expressed in the same sample compared to no integrated samples. And the rest integrated samples have also difference with non-integrated group ( $P < 0.01$ ). We further analyzed the integrated sites. All samples with ERBB2 integrated have the same breakpoints in chr17: 39,695,542/39695556/39703981 (ERBB2) and chr17: 39,647,045 (STARD3). The stained most strongly sample not only have these sites, but also other different sites. This sample was also detected mostly integrated sites of STARD3 and the highest expression of STARD3. It was not clear why the expression of integrated genes was mostly increased in the same sample. The amounts of integrated sites or certain specific site maybe reasons and more researches need to be done. In a word, the samples with integration in ERBB2 was same as that with integration in STARD3 and they both overexpressed. We strongly believed that STARD3 co-amplified and co-expressed with ERBB2 in AC. Meanwhile, the FIGO stage of the most strongly stained sample was more advanced. Of particular note is that owing to small set of samples, more researches need to be done. STARD3 and ERBB2 potentially were recognized as novel candidate biomarkers of AC. RAD51B (3) is known as a hotspot gene

(Li et al. 2018). It plays a critical role in homologous recombinational repair of DNA double-strand breaks (important for genomic stability and a hallmark of cancer) (Hang et al. 2016). In this study, HPV integration sites were found in intron 7 of RAD51B-204 (the main transcript) and its breakage may be promoting progression of cancer. Therefore, we hypothesized that HPV integration may lead to the dysregulation of oncogenes, that plays a more important role in HPV-related AC.

RNA45S rDNA (11) serves as the most frequently integrated hotspot gene of SCC. One-half of the integration sites were located in intergenic regions between RNA45S (distance = minimum 2 kb and maximum 400 kb in chr21) and MIR3648-1 (distance = minimum 510 kb and maximum 530 kb in chr21). rDNA is located in fragile sites in the chromosomes (Sidler 2016), making it susceptible to breakage and HPV integration. Genes encoding 45S ribosomal RNA are abundant in eukaryotic genomes (Havlova and Fajkus 2020). The change of rDNA expression due to integration maybe negligible, but as mentioned above, rDNA possesses powerful promoters that can promote the overexpression of viral oncogenes. In SCC, viral oncogenes may utilize the powerful hosts' promoters to express viral oncogenes and the overexpression of viral oncogenes may play a more important role in the development of carcinogenesis in SCC than that of AC.

We studied the relationship between clinicopathological parameters and HPV integration and found that HPV integration correlated with FIGO stage and pathology (\*\* $P < 0.01$ ). However, the amount of AC samples was a little and some data of AC were missing, this result need more study to verify. Then, we investigated the association between the breakpoints of HPV16, AC and SCC. The results revealed that HPV16 breakpoints of AC intended to locate in E7 compared to that of SCC. Meanwhile, breakpoints of SCC prone to occur in LCR compared to that of AC (Fig. 6a, b). Previous study has revealed that E2 bond with high-affinity E2BS1 and the p97 promoter was activated (Reuschenbach et al. 2015). Then the level of E2 enhanced. E2 then also bonded to the low-affinity E2BS3 and E2BS4, and this resulted in inhibition of the p97 promoter (Reuschenbach et al. 2015). Arkom Chaiwongkot proved that E2BS1 and promoter region covered from 7,432 to 116 (Chaiwongkot et al. 2013). We found breakpoints of HPV16 in SCC tended to locate in the E2BS1 and promoter region, which destroyed the binding of E2 and E2BS1. The activation of p97 promoter that promoted the expression of E6/7 was unable. Therefore, this result was corresponding to our hypothesis that viral oncogenes may utilize the powerful hosts' promoters to express viral oncogenes.

In our study, we focused on the characterization of HPV integration sites in AC and SCC. Notably, we only included cancer samples and did not enroll non-cancer samples with

HPV integration. Thus, the characteristics of the observed regions probably contribute to carcinogenesis, instead of conferring susceptibility to HPV integration. We analyzed the results and speculated the carcinogenic mechanisms of HPV integration in SCC and AC focusing on highly frequent integration cyto bands, characterization of genomic elements, integration of hotspot genes, enrichment pathways, and clinicopathological parameters. Meanwhile, we analyzed the HPV16 breakpoints in AC and SCC. In conclusion, the HPV-mediated carcinogenic mechanisms in SCC and AC were different. We proposed that the overexpression of viral oncogenes plays a more important role and it may utilize the powerful hosts' promoters to express viral oncogenes in carcinogenesis in SCC compared to AC. In AC, HPV integration may affect hosts' oncogenes, such as ERBB2 and STARD3. And the dysregulation of oncogenes may primarily contribute to progression of AC.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00432-023-05494-4>.

**Author contributions** Y.B., H.C., Q.M. and X.W. contributed to the study conception and design. Y.B., L.Z., G.C., H.C. and J.H. participated in material preparation, data collection and analysis. Y.B. wrote the first draft of the manuscript and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

**Funding** This research was supported by Hubei Provincial Natural Science Foundation of China (2023AFB1064, 2021CFB240).

**Data availability** The datasets generated and analyzed during the current study are available in the Supplementary Information.

## Declarations

**Conflict of interest** The authors declare no competing interests.

**Ethical approval** The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of Maternal and Child Health Hospital of Hubei Province.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Bannister AJ, Kouzarides T (2011) Regulation of chromatin by histone modifications. *Cell Res* 21(3):381–395. <https://doi.org/10.1038/cr.2011.22>
- Bartlett A, O'Malley RC, Huang SC, Galli M, Nery JR, Gallavotti A, Ecker JR (2017) Mapping genome-wide transcription-factor binding sites using DAP-seq. *Nat Protoc* 12(8):1659–1672. <https://doi.org/10.1038/nprot.2017.055>
- Bedell MA, Hudson JB, Golub TR, Turyk ME, Hosken M, Wilbanks GD, Laimins LA (1991) Amplification of human papillomavirus genomes in vitro is dependent on epithelial differentiation. *J Virol* 65(5):2254–2260. <https://doi.org/10.1128/JVI.65.5.2254-2260.1991>
- Bieche I, Tomasetto C, Regnier CH, Moog-Lutz C, Rio MC, Lidereau R (1996) Two distinct amplified regions at 17q11-q21 involved in human primary breast cancer. *Cancer Res* 56(17):3886–3890
- Bodelon C, Untereiner ME, Machiela MJ, Vinokurova S, Wentzensen N (2016) Genomic characterization of viral integration sites in HPV-related cancers. *Int J Cancer* 139(9):2001–2011. <https://doi.org/10.1002/ijc.30243>
- 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, Research Institute at Christiana Care Health S, HudsonAlpha Institute for B, Ilsbio LLC, Indiana University School of M, Institute of Human V, Institute for Systems B, International Genomics C, Leidos B, Massachusetts General H, McDonnell Genome Institute at Washington U, Medical College of W, Medical University of South C, Memorial Sloan Kettering Cancer C, Montefiore Medical C, NantOmics, National Cancer I, National Hospital AN, National Human Genome Research I, National Institute of Environmental Health S, National Institute on D, Other Communication D, Ontario Tumour Bank LHSC, Ontario Tumour Bank OIICR, Ontario Tumour Bank TOH, Oregon H, Science U, Samuel Oschin Comprehensive Cancer Institute C-SMC, International SRA, St Joseph's Candler Health S, Eli, Edythe LBIoMIoT, Harvard U, Research Institute at Nationwide Children's H, Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins U, University of B, University of Texas MDACC, University of Abuja Teaching H, University of Alabama at B, University of California I, University of California Santa C, University of Kansas Medical C, University of L, University of New Mexico Health Sciences C, University of North Carolina at Chapel H, University of Oklahoma Health Sciences C, University of P, University of Sao Paulo RaPMS, University of Southern C, University of W, University of Wisconsin School of M, Public H, Van Andel Research I, Washington University in St L (2017) Integrated genomic and molecular characterization of cervical cancer. *Nature* 543(7645):378–384. <https://doi.org/10.1038/nature21386>
- Chaiwongkot A, Vinokurova S, Pientong C, Ekakaksananan T, Kongyingoes B, Kleebkaow P, Chumworathayi B, Patarapadungkit N, Reuschenbach M, von Knebel DM (2013) Differential methylation of E2 binding sites in episomal and integrated HPV 16 genomes in preinvasive and invasive cervical lesions. *Int J Cancer* 132(9):2087–2094. <https://doi.org/10.1002/ijc.27906>
- Dao FY, Lv H, Su W, Sun ZJ, Huang QL, Lin H (2021) iDHS-Deep: an integrated tool for predicting DNase I hypersensitive sites by deep neural network. *Brief Bioinform*. <https://doi.org/10.1093/bib/bbab047>
- Fan J, Fu Y, Peng W, Li X, Shen Y, Guo E, Lu F, Zhou S, Liu S, Yang B, Qin X, Hu D, Xiao R, Li X, Yang S, Yuan C, Shu Y, Huang H, Wan T, Pi Y, Wang S, Chen W, Wang H, Zhong L, Yuan L, Wen B, Kong B, Mills GB, Zou D, Xia B, Song K, Chen G, Ma D, Sun

- C (2023) Multi-omics characterization of silent and productive HPV integration in cervical cancer. *Cell Genom* 3(1):100211. <https://doi.org/10.1016/j.xgen.2022.100211>
- Greer EL, Shi Y (2012) Histone methylation: a dynamic mark in health, disease and inheritance. *Nat Rev Genet* 13(5):343–357. <https://doi.org/10.1038/nrg3173>
- Hang D, Zhou W, Jia M, Wang L, Zhou J, Yin Y, Ma H, Hu Z, Li N, Shen H (2016) Genetic variants within microRNA-binding site of RAD51B are associated with risk of cervical cancer in Chinese women. *Cancer Med* 5(9):2596–2601. <https://doi.org/10.1002/cam4.797>
- Havlova K, Fajkus J (2020) G4 structures in control of replication and transcription of rRNA genes. *Front Plant Sci* 11:593692. <https://doi.org/10.3389/fpls.2020.593692>
- Hong Dong, Shuang Li, Wang Dan Hu (2010) Comparison of the biological behaviors and prognostic factors in cervical adenocarcinoma and squamous carcinoma. *J Huazhong Univer Sci Technol Health Sci* 02:254–257
- Hoppe-Seyler K, Bossler F, Braun JA, Herrmann AL, Hoppe-Seyler F (2018) The HPV E6/E7 oncogenes: key factors for viral carcinogenesis and therapeutic targets. *Trends Microbiol* 26(2):158–168. <https://doi.org/10.1016/j.tim.2017.07.007>
- Hu Z, Zhu D, Wang W, Li W, Jia W, Zeng X, Ding W, Yu L, Wang X, Wang L, Shen H, Zhang C, Liu H, Liu X, Zhao Y, Fang X, Li S, Chen W, Tang T, Fu A, Wang Z, Chen G, Gao Q, Li S, Xi L, Wang C, Liao S, Ma X, Wu P, Li K, Wang S, Zhou J, Wang J, Xu X, Wang H, Ma D (2015) Genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism. *Nat Genet* 47(2):158–163. <https://doi.org/10.1038/ng.3178>
- Hu T, Li K, He L, Huang F, Yang F, Chen S, Wang H, Ma D, Huang X, Wu P (2023) Testing for viral DNA integration among HPV-positive women to detect cervical precancer: an observational cohort study. *BJOG*. <https://doi.org/10.1111/1471-0528.17597>
- Inagaki Y, Tsunokawa Y, Takebe N, Nawa H, Nakanishi S, Terada M, Sugimura T (1988) Nucleotide sequences of cDNAs for human papillomavirus type 18 transcripts in HeLa cells. *J Virol* 62(5):1640–1646. <https://doi.org/10.1128/JVI.62.5.1640-1646.1988>
- Irie T, Kigawa J, Minagawa Y, Itamochi H, Sato S, Akeshima R, Terakawa N (2000) Prognosis and clinicopathological characteristics of Ib-IIb adenocarcinoma of the uterine cervix in patients who have had radical hysterectomy. *Eur J Surg Oncol* 26(5):464–467. <https://doi.org/10.1053/ejso.1999.0923>
- Kamal M, Lameiras S, Deloger M, Morel A, Vacher S, Lecerf C, Dupain C, Jeannot E, Girard E, Baulande S, Dubot C, Kenter G, Jordanova ES, Berns E, Bataillon G, Popovic M, Rouzier R, Cacheux W, Le Tourneau C, Nicolas A, Servant N, Scholl SM, Bieche I, Consortium RA (2021) Human papilloma virus (HPV) integration signature in Cervical Cancer: identification of MACROD2 gene as HPV hot spot integration site. *Br J Cancer* 124(4):777–785. <https://doi.org/10.1038/s41416-020-01153-4>
- Kao J, Pollack JR (2006) RNA interference-based functional dissection of the 17q12 amplicon in breast cancer reveals contribution of coamplified genes. *Genes Chromosomes Cancer* 45(8):761–769. <https://doi.org/10.1002/gcc.20339>
- Kim H, Nguyen NP, Turner K, Wu S, Gujar AD, Luebeck J, Liu J, Deshpande V, Rajkumar U, Namburi S, Amin SB, Yi E, Menghi F, Schulte JH, Henssen AG, Chang HY, Beck CR, Mischel PS, Bafna V, Verhaak RGW (2020) Extrachromosomal DNA is associated with oncogene amplification and poor outcome across multiple cancers. *Nat Genet* 52(9):891–897. <https://doi.org/10.1038/s41588-020-0678-2>
- Li H, Yang Y, Zhang R, Cai Y, Yang X, Wang Z, Li Y, Cheng X, Ye X, Xiang Y, Zhu B (2013) Preferential sites for the integration and disruption of human papillomavirus 16 in cervical lesions. *J Clin Virol* 56(4):342–347. <https://doi.org/10.1016/j.jcv.2012.12.014>
- Li W, Qi Y, Cui X, Huo Q, Zhu L, Zhang A, Tan M, Hong Q, Yang Y, Zhang H, Liu C, Kong Q, Geng J, Tian Y, Kong F, Man D (2018) Characteristic of HPV integration in the genome and transcriptome of cervical cancer tissues. *Biomed Res Int* 2018:6242173. <https://doi.org/10.1155/2018/6242173>
- Li W, Tian S, Wang P, Zang Y, Chen X, Yao Y, Li W (2019) The characteristics of HPV integration in cervical intraepithelial cells. *J Cancer* 10(12):2783–2787. <https://doi.org/10.7150/jca.31450>
- Liu M, Han Z, Zhi Y, Ruan Y, Cao G, Wang G, Xu X, Mu J, Kang J, Dai F, Wen X, Zhang Q, Li F (2023) Long-read sequencing reveals oncogenic mechanism of HPV-human fusion transcripts in cervical cancer. *Transl Res* 253:80–94. <https://doi.org/10.1016/j.trsl.2022.09.004>
- Lodi M, Voilquin L, Alpy F, Moliere S, Reix N, Mathelin C, Chenard MP, Tomasetto CL (2023) STARD3: a new biomarker in HER2-positive breast cancer. *Cancers (Basel)*. <https://doi.org/10.3390/cancers15020362>
- Lu Q, Richardson B (2004) DNaseI hypersensitivity analysis of chromatin structure. *Methods Mol Biol* 287:77–86. <https://doi.org/10.1385/1-59259-828-5-077>
- Lu C, Yang Y, Ma S (2018) A functional variant (Rs35592567) in TP63 at 3q28 is associated with gastric cancer risk via modifying its regulation by MicroRNA-140. *Cell Physiol Biochem* 47(1):235–244. <https://doi.org/10.1159/000489802>
- Oh DY, Kim S, Choi YL, Cho YJ, Oh E, Choi JJ, Jung K, Song JY, Ahn SE, Kim BG, Bae DS, Park WY, Lee JW, Song S (2015) HER2 as a novel therapeutic target for cervical cancer. *Oncotarget* 6(34):36219–36230. <https://doi.org/10.18632/oncotarget.5283>
- Ojesina AI, Lichtenstein L, Freeman SS, Pedamallu CS, Imaz-Rosshandler I, Pugh TJ, Cherniack AD, Ambrogio L, Cibulskis K, Bertelsen B, Romero-Cordoba S, Trevino V, Vazquez-Santillan K, Guadarrama AS, Wright AA, Rosenberg MW, Duke F, Kaplan B, Wang R, Nickerson E, Walline HM, Lawrence MS, Stewart C, Carter SL, McKenna A, Rodriguez-Sanchez IP, Espinosa-Castilla M, Woie K, Bjorge L, Wik E, Halle MK, Hoivik EA, Krakstad C, Gabino NB, Gomez-Macias GS, Valdez-Chapa LD, Garza-Rodriguez ML, Maytorena G, Vazquez J, Rodea C, Cravioto A, Cortes ML, Greulich H, Crum CP, Neuberger DS, Hidalgo-Miranda A, Escareno CR, Akslen LA, Carey TE, Vintermyr OK, Gabriel SB, Barrera-Saldana HA, Melendez-Zajgla J, Getz G, Salvesen HB, Meyerson M (2014) Landscape of genomic alterations in cervical carcinomas. *Nature* 506(7488):371–375. <https://doi.org/10.1038/nature12881>
- Peter M, Rosty C, Couturier J, Radvanyi F, Teshima H, Sastre-Garau X (2006) MYC activation associated with the integration of HPV DNA at the MYC locus in genital tumors. *Oncogene* 25(44):5985–5993. <https://doi.org/10.1038/sj.onc.1209625>
- Porter VL, Marra MA (2022) The drivers, mechanisms, and consequences of genome instability in HPV-driven cancers. *Cancers (Basel)*. <https://doi.org/10.3390/cancers14194623>
- Reuschenbach M, Huebbers CU, Prigge ES, Bermejo JL, Kalteis MS, Preuss SF, Seuthe IM, Kolligs J, Speel EJ, Olthof N, Kremer B, Wagner S, Klussmann JP, Vinokurova S, von Knebel DM (2015) Methylation status of HPV16 E2-binding sites classifies subtypes of HPV-associated oropharyngeal cancers. *Cancer* 121(12):1966–1976. <https://doi.org/10.1002/cncr.29315>
- Roman A, Munger K (2013) The papillomavirus E7 proteins. *Virology* 445(1–2):138–168. <https://doi.org/10.1016/j.viro.2013.04.013>
- Sahlberg KK, Hongisto V, Edgren H, Makela R, Hellstrom K, Due EU, Moen Volland HK, Sahlberg N, Wolf M, Borresen-Dale AL, Perala M, Kallioniemi O (2013) The HER2 amplicon includes several genes required for the growth and survival of HER2 positive

- breast cancer cells. *Mol Oncol* 7(3):392–401. <https://doi.org/10.1016/j.molonc.2012.10.012>
- Schmitz M, Driesch C, Beer-Grondke K, Jansen L, Runnebaum IB, Durst M (2012) Loss of gene function as a consequence of human papillomavirus DNA integration. *Int J Cancer* 131(5):E593–602. <https://doi.org/10.1002/ijc.27433>
- Shapiro JA, von Sternberg R (2005) Why repetitive DNA is essential to genome function. *Biol Rev Camb Philos Soc* 80(2):227–250. <https://doi.org/10.1017/s1464793104006657>
- Shi Y, Li L, Hu Z, Li S, Wang S, Liu J, Wu C, He L, Zhou J, Li Z, Hu T, Chen Y, Jia Y, Wang S, Wu L, Cheng X, Yang Z, Yang R, Li X, Huang K, Zhang Q, Zhou H, Tang F, Chen Z, Shen J, Jiang J, Ding H, Xing H, Zhang S, Qu P, Song X, Lin Z, Deng D, Xi L, Lv W, Han X, Tao G, Yan L, Han Z, Li Z, Miao X, Pan S, Shen Y, Wang H, Liu D, Gong E, Li Z, Zhou L, Luan X, Wang C, Song Q, Wu S, Xu H, Shen J, Qiang F, Ma G, Liu L, Chen X, Liu J, Wu J, Shen Y, Wen Y, Chu M, Yu J, Hu X, Fan Y, He H, Jiang Y, Lei Z, Liu C, Chen J, Zhang Y, Yi C, Chen S, Li W, Wang D, Wang Z, Di W, Shen K, Lin D, Shen H, Feng Y, Xie X, Ma D (2013) A genome-wide association study identifies two new cervical cancer susceptibility loci at 4q12 and 17q12. *Nat Genet* 45(8):918–922. <https://doi.org/10.1038/ng.2687>
- Shi H, Shao Y, Lu W, Lu B (2021) An analysis of HER2 amplification in cervical adenocarcinoma: correlation with clinical outcomes and the international endocervical adenocarcinoma criteria and classification. *J Pathol Clin Res* 7(1):86–95. <https://doi.org/10.1002/cjp2.184>
- Shimada M, Kigawa J, Nishimura R, Yamaguchi S, Kuzuya K, Nakaniishi T, Suzuki M, Kita T, Iwasaka T, Terakawa N (2006) Ovarian metastasis in carcinoma of the uterine cervix. *Gynecol Oncol* 101(2):234–237. <https://doi.org/10.1016/j.ygyno.2005.10.004>
- Sidler C (2016) Chapter 29 - Genomic Instability and Aging: Causes and Consequences. In: Kovalchuk I, Kovalchuk O (eds) *Genome Stability*. Academic Press, Boston, pp 511–525. <https://doi.org/10.1016/B978-0-12-803309-8.00029-X>
- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F (2021) Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 Countries. *CA Cancer J Clin* 71(3):209–249. <https://doi.org/10.3322/caac.21660>
- Tomasetto C, Regnier C, Moog-Lutz C, Mattei MG, Chenard MP, Lidereau R, Basset P, Rio MC (1995) Identification of four novel human genes amplified and overexpressed in breast carcinoma and localized to the q11–q21.3 region of chromosome 17. *Genomics* 28(3):367–376. <https://doi.org/10.1006/geno.1995.1163>
- Vande Pol SB, Klingelutz AJ (2013) Papillomavirus E6 oncoproteins. *Virology* 445(1–2):115–137. <https://doi.org/10.1016/j.virol.2013.04.026>
- Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 38(16):e164. <https://doi.org/10.1093/nar/gkq603>
- Wang M, Du M, Ma L, Chu H, Lv Q, Ye D, Guo J, Gu C, Xia G, Zhu Y, Ding Q, Yuan L, Fu G, Tong N, Qin C, Yin C, Xu J, Zhang Z (2016) A functional variant in TP63 at 3q28 associated with bladder cancer risk by creating an miR-140-5p binding site. *Int J Cancer* 139(1):65–74. <https://doi.org/10.1002/ijc.29978>
- Wilhelm LP, Wendling C, Védie B, Kobayashi T, Chenard MP, Tomasetto C, Drin G, Alpy F (2017) STARD3 mediates endoplasmic reticulum-to-endosome cholesterol transport at membrane contact sites. *EMBO J* 36(10):1412–1433. <https://doi.org/10.15252/embj.201695917>
- Xiang L, Jiang W, Ye S, He T, Pei X, Li J, Chan DW, Ngan HYS, Li F, Tao P, Shen X, Zhou X, Wu X, Yang G, Yang H (2018) ERBB2 mutation: A promising target in non-squamous cervical cancer. *Gynecol Oncol* 148(2):311–316. <https://doi.org/10.1016/j.ygyno.2017.12.023>
- Yao TAN, Yujing XIONG, Meng XU, Bin YAN, Huang CAO, Junbo HU, Xufeng WU, Hanming W (2022) Differential expression level of trefoil factor 3 in different histological pathological types of cervical cancer and its clinical significance. *Medical Journal of Wuhan University* 02:238–242
- Zeng X, Shi G, He Q, Zhu P (2021) Screening and predicted value of potential biomarkers for breast cancer using bioinformatics analysis. *Sci Rep* 11(1):20799. <https://doi.org/10.1038/s41598-021-00268-9>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.