Research Article

Characteristic of HPV Integration in the Genome and Transcriptome of Cervical Cancer Tissues

Weiyang Li [1,2,3] **Yanwei Qi,** [4] **Xiaofang Cui** [1] **Qing Huo,** [4]
Liangxi Zhu, [1,3] **Aiping Zhang,** [4] **Meihua Tan,** [5] **Qilan Hong,** [5] **Yan Yang,** [1,3]
Huali Zhang, [1,2] **Chuanxin Liu,** [1,2] **Qingsheng Kong,** [1,2] **Jiazheng Geng,** [1]
Yanjun Tian, [1,2] **Fancong Kong,** [1,3] **and Dongmei Man** [1,3]

[1] *Jining Medical University, Jining, Shandong 272067, China*
[2] *Collaborative Innovation Center for Birth Defect Research and Transformation of Shandong Province, Jining Medical University, Jining, Shandong 272067, China*
[3] *Affiliated Hospital of Jining Medical University, Jining, Shandong 272067, China*
[4] *BGI-Shenzhen, Shenzhen 518083, China*
[5] *University of Chinese Academy of Sciences, 19A Yuquan Road, Shijingshan District, Beijing 100049, China*

Correspondence should be addressed to Weiyang Li; 163.lwy@163.com, Xiaofang Cui; scucxf@163.com, and Dongmei Man; mandongmei@163.com

High-risk HPV is clearly associated with cervical cancer. HPV integration has been confirmed to promote carcinogenesis in the previous studies. In our study, a total of 285 DNA breakpoints and 287 RNA breakpoints were collected. We analyzed the characteristic of HPV integration in the DNA and RNA samples. The results revealed that the patterns of HPV integration in RNA and DNA samples differ significantly. *FHIT*, *KLF5*, and *LINC00392* were the hotspot genes integrated by HPV in the DNA samples. *RAD51B*, *CASC8*, *CASC21*, *ERBB2*, *TP63*, *TEX41*, *RAP2B*, and *MYC* were the hotspot genes integrated by HPV in RNA samples. Breakpoints of DNA samples were significantly prone to the region of INTRON ($P < 0.01$, Chi-squared test), whereas in the RNA samples, the breakpoints were prone to EXON. Pathway analysis had revealed that the breakpoints of RNA samples were enriched in the pathways of transcriptional misregulation in cancer, cancer pathway, and pathway of adherens junction. Breakpoints of DNA samples were enriched in the pathway of cholinergic synapse. In summary, our data helped to gain insights into the HPV integration sites in DNA and RNA samples of cervical cancer. It had provided theoretical basis for understanding the mechanism of tumorigenesis from the perspective of HPV integration in the HPV-associated cervical cancers.

1. Introduction

HPV is a DNA virus that has been widely detected in humans and animals. High-risk HPV is clearly associated with cervical intraepithelial lesions and cervical cancer. Generally, about half of HPV infections could be eliminated within one year. However, infection by high-risk HPV usually could persist for several years and these types of HPV are also associated with reduced removal efficiency [1]. Moreover, persistent HPV infection for decades is likely to induce invasive cervical cancer [2].

The microscopic HPV particle is 50-60nm in diameter, and its surface consists of 72 capsomere [3]. Wrapped inside the capsid proteins is the double-stranded HPV DNA. The HPV genome may be divided into three regions, an early (E; E1, E2, E3, E4, E5, E6, E7, and E8 genes), late (L; L1 and L2 genes), and noncoding long control region (LCR). The E region is crucial for HPV replication, transcription, translation, and transformation. The L region (~2500 bp) encoded functional regulators for HPV replication and transcription [4]. Generally, absence of HPV integration in the host genome is associated with benign lesions. Positive HPV

integrations are linked to cervical CIN grades and cervical cancer [5].

In recent years, high-throughput sequencing technology had provided robust means to investigate the characteristic and biological significance of HPV integration. Previous study had revealed that HPV integration could trigger genome instability; for instance, it results in genome structure rearrangement and copy number variation [6]. A recent study had shown that HPV integration within 8q24 region triggered a great number of rearrangement events in the study of HeLa cells haplotype and it might suggest HPV integration could directly initiate tumorigenesis [7]. In addition, a series of hotspots genes integrated by HPV had been found in the recent study [8]. Despite increased attention on HPV integration hotspots, the characteristic of HPV integration and the relationship between HPV integration and cervical cancer remained elusive.

In this study, a total of 285 DNA breakpoints and 287 RNA breakpoints were collected from previous studies [6, 8–12]. Our data revealed that the patterns of HPV integration in RNA and DNA samples differ significantly. Pathway analysis had revealed that breakpoints were enriched in the different pathways between RNA samples and DNA samples. Our study could further help to gain insights into the characteristic of HPV integration in DNA and RNA samples and provide theoretical basis for understanding the mechanism of tumorigenesis.

2. Material and Method

HPV integration sites were collected from 6 recent studies (Table S1, Table S2). Functional annotation analysis of breakpoints was performed using DAVID based on Gene Ontology and KEGG pathway databases [13, 14]. The categories of KEGG Pathways were as background databases. The breakpoints are annotated through the latest ANNOVAR in hg19 coordinates [15]. The region list of genomic elements was downloaded from the UCSC genome browser [16].

3. Gene Frequency

Because HPV integration was considered a strong cis-activator of flanking genes and cis-acting enhancers can influence their target genes over long distances [17, 18] (up to 1 Mb for upstream enhancers and 850 kb for downstream enhancers), breakpoints located <500 kb from annotated genes were included to calculate the affected gene frequency in HPV-integrated samples [8].

4. Results

4.1. HPV Integration Hotspots in DNA and RNA Samples. Based on frequency analysis, HPV integration hotspots had been identified in these samples. *FHIT*(8), *KLF5*(6), and *LINC00392*(4) were the most integrated genes in the DNA samples. In contrast, *RAD51B*(9), *CASC8*(5), *CASC21*(5), *ERBB2*(5), *TP63*(5), *TEX41*(5), *RAP2B*(4), and *MYC*(4) were the most integrated genes in RNA samples (Figures 1 and 4). Totally, we obtained 12 and 18 recurrent genes (frequency ⩾ 2)

integrated by HPV in the DNA and RNA samples, respectively (Table S3, Table S4).

4.2. Distribution of Genetic Elements. We surveyed the distribution characteristics of the HPV breakpoints in DNA and RNA samples. The results revealed that HPV breakpoints were more prone to INTRON in the DNA samples than the RNA samples ($P < 0.01$, Chi-squared test, Figure 2). However, HPV breakpoints were more prone to EXON in the RNA samples than the DNA samples ($P < 0.01$, Chi-squared test, Figure 2).

4.3. Genomic Element Distribution. The HPV integration sites (breakpoints) in our RNA and DNA samples showed similar distributions in fragile, CpG, TFBS sites. However, the HPV integration sites in the RNA samples were more prone to fragile, CpG, and TFBS than that of the DNA samples (Chi-squared test, Figure 3).

4.4. Pathway Analysis. The results revealed that the DNA pathway was enriched on the pathway of cholinergic synapse. However, the main enriched pathways of breakpoints from RNA samples were the pathways of transcriptional misregulation in cancer, cancer pathway, and pathway of adherens junction. It revealed that there was significant difference between the enrichment pathway of RNA and DNA samples (Table S5).

5. Discussion

In this study, the 285 DNA breakpoints and 287 RNA breakpoints were used to carry out the bioinformatic analysis. Among 285 integration sites of DNA breakpoints, 163 integration sites were mapped by Hu and colleagues [8]. In total, Hu et al. had identified 3,667 breakpoints in 135 samples and obtained a validation rate ~83% by PCR and Sanger. However, many of the breakpoints have low integration frequencies (NNSS value < 3). Generally, the integration events with low frequencies might have fewer impacts toward tissue functions. Additionally, the breakpoints with higher NNSS values often mean more support-reads, hence greater reliability. Owing to this matter, breakpoints of NNSS value > 3 were selected in our study. In order to get the major breakpoints, we filtered out these breakpoints surrounding the major breakpoints and obtained the 163 breakpoints (Table S6).

In theory, it would be ideal to study the characteristics of HPV integration using paired DNA and RNA samples. However, there were only 5 overlapping samples between DNA and RNA samples in our study. Those small size samples were not enough to study the relation of breakpoints in the DNA and paired RNA. In addition, we had noted that it was difficult to find sufficient breakpoints (paired RNA/DNA samples) from second-generation sequencing in the existing databases. Therefore, most of the breakpoints that we used to carry out the analysis were from unpaired RNA/DNA samples.
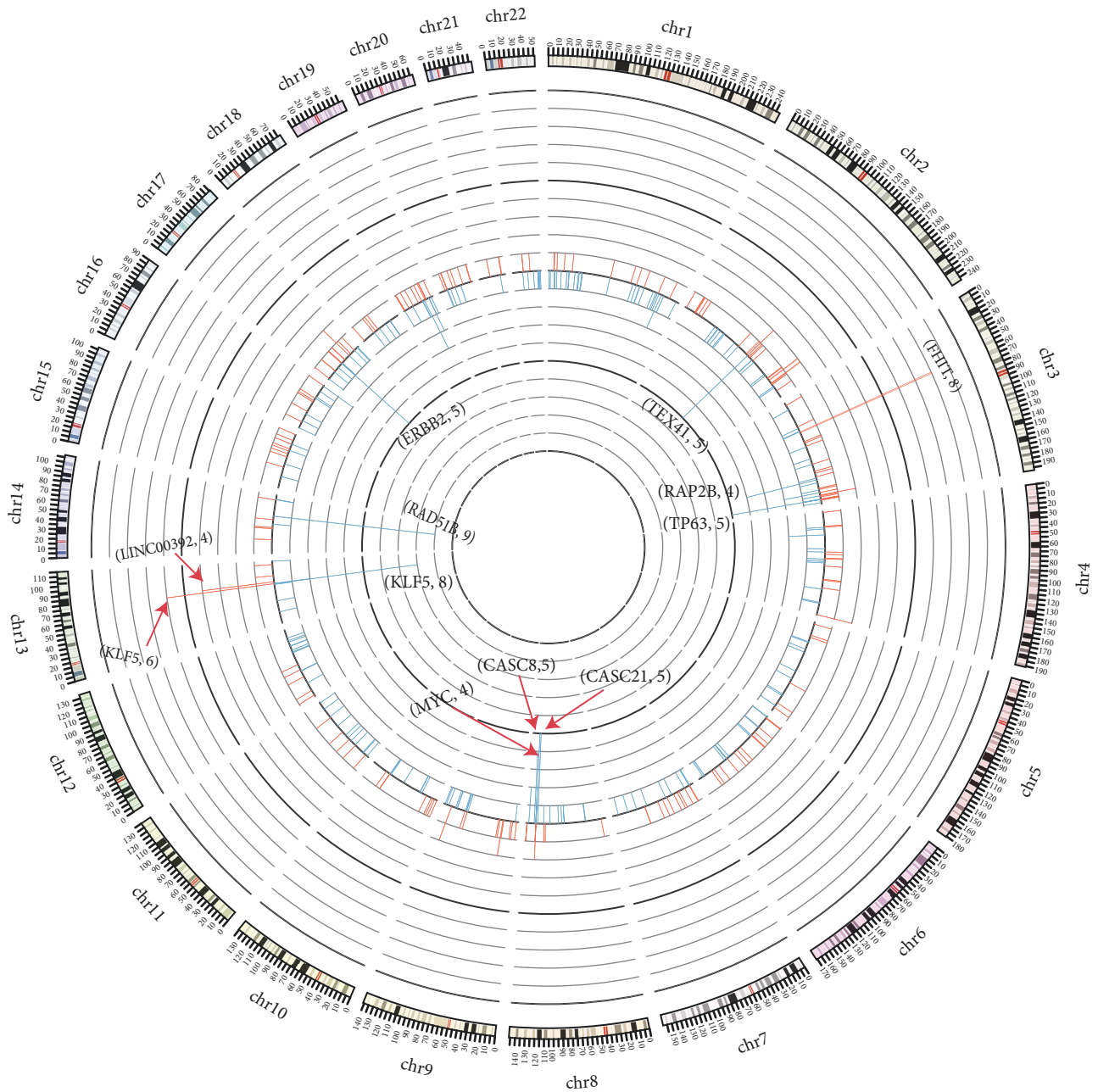
FIGURE 1: Gene frequency integrated by HPV between RNA and DNA samples. The figure showed the difference of gene frequency in DNA and RNA samples. The inner circle revealed the gene frequency (blue color) in RNA samples and the outer circle revealed the gene frequency (red color) in DNA samples. The height represented the frequency of gene integrated by HPV.

As suggested by our results, the hotspots of HPV integration in the genome and transcriptome appeared to locate in different genes. Intriguingly, certain high frequency genes (i.e., ERBB2) in RNA samples appeared to have higher mutation frequencies (i.e., ERBB2, 5%) in the COSMIC database (http://cancer.sanger.ac.uk/cancergenome/projects/cosmic). Further, mutations in *ERBB2* had been known as therapeutic targets in lung and breast cancer in vitro [19, 20]. Our study had also observed that there were several important genes with high frequencies are preferential for HPV integration.

The *RAD51B* gene belongs to the *RAD51* family, which is known to play important roles in DNA repair. Frequent HPV integrated into *RAD51B* might disrupt the DNA repair mechanism, which could partially explain the HPV-rendered genomic dysfunction and chromosome instability in cancers [21].

FHIT is another gene in high frequency of HPV integration, and it is located in a fragile genomic region (*FRA3B* region). This leads to a speculation that HPV integration into such region might trigger great chromosomal instability,
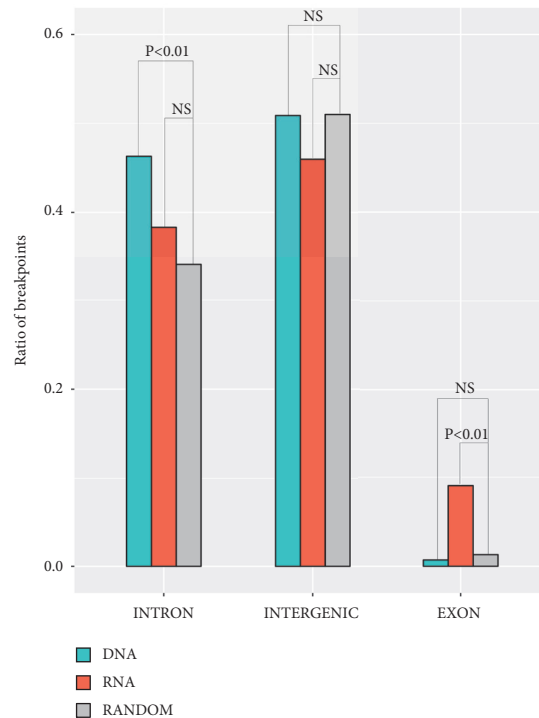
FIGURE 2: Distribution of breakpoints in genetic elements. The ratio of breakpoints in each genome element was counted. The expected ratio of each genome elements was calculated according to the random distribution of breakpoints in the whole human genome. Grey bar represented the expected ratio of breakpoints. Orange bar represented the observed ratio of breakpoints in RNA samples. Green bar represented the observed ratio of breakpoints in DNA samples. P values were calculated by Chi-squared test and were corrected by Fisher exact test.

probably via chromosomal translocation [22]. Moreover, TP63, RAP2B, KLF5, and MYC are closely related to tumorigenesis and were identified as hotspots of HPV integration [23–25]. Therefore, it is highly likely that these four genes could potentially drive tumorigenesis after genes are integrated by HPV.

As observed in the DNA samples, the HPV integration sites were inclined to INTRON region. In contrast, the HPV integration sites found in the RNA samples were enriched in EXON, CpG, and transcription factor binding sites (TFBS). Interestingly, a large portion of HPV integration sites in RNA samples was located on the no-coding region (INTRON, INTERGENIC). It might suggest that HPV integration could directly trigger the abnormal transcription and these functions of novel transcript kept unclear. Further, we found that the ratios of HPV integration sites within ALU and LINE were significantly higher in the DNA samples than those of RNA samples. Most importantly, the overall HPV breakpoints in the RNA samples suggested specific enrichments on the pathways of transcription regulation, cancer, and adherent junction. Furthermore, we noted that Xu et al. had compared the DNA junctions with the paired RNA junctions and they found that 12 of the 20 carcinomas (60%) contained a single transcriptionally active HPV16 integrate. The other 8 tumors (40%) are featured by a transcriptionally active HPV16 integrate together with one or two probably silent HPV16 integrates [12]. The phenomenon might suggest that only part of integration sites from DNA could be transcribed efficiently. The different characteristics of HPV integration in DNA and RNA might be associated with transcriptional activity of DNA breakpoints.

Due to the significant difference observed while comparing the breakpoint profiles of the DNA and RNA samples, it raises the speculations that the genomic and transcriptomic breakpoints might play the different role in tumorigenesis.

In this study, our results had revealed characteristics of HPV integration sites in the DNA and RNA samples. Additionally, the breakpoints in the RNA samples suggested tumorigenesis might arise from disrupting transcription and interrupting DNA repair mechanism. Altogether, this study had provided theoretical basis for understanding the mechanism of tumorigenesis from the perspective of HPV integration in the HPV-associated cervical cancers.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Ethical Approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki
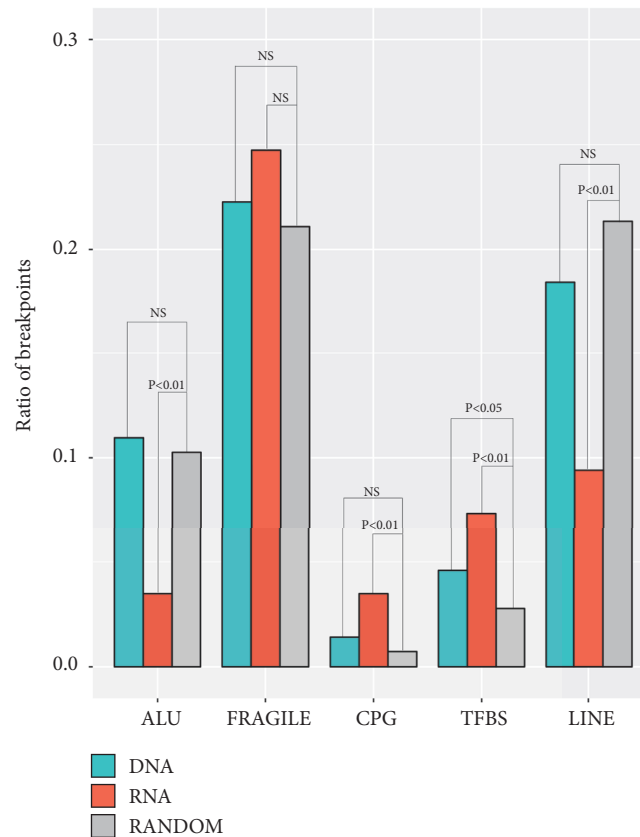
FIGURE 3: Distribution of breakpoints in the region of Fragile, CPG, TFBS, ALU, TFBS, and LINE. X axis represented different elements; Y axis represent ratio of integration breakpoints. The expected (random distribution, Grey) and the observed (actual ratio, RNA samples: orange; DNA samples: green) percentages of breakpoints are shown. P values were calculated by Chi-squared test and also were corrected by Fisher exact test.

Declaration and its later amendments or comparable ethical standards. The study had been approved by the Ethics Review Committee in the Jining Medical University.

Consent

Informed consent was obtained from all individual participants included in the study.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Authors' Contributions

Weiyang Li, Xiaofang Cui, and Dongmei Man conceived and designed the paper. Yanwei Qi, Qing Huo, Liangxi Zhu, Aiping Zhang, Yan Yang, Meihua Tan, Qilan Hong, Huali Zhang, Chuanxin Liu, and Qingsheng Kong analyzed the data. Weiyang Li, Xiaofang Cui, and Qing Huo wrote the paper. Jiazheng Geng, Yanjun Tian, and Fancong Kong provided good advice. Weiyang Li, Yanwei Qi, and Xiaofang Cui contributed equally to this work.

Acknowledgments

Supplementary Materials

Supplementary 1. Table S1: the breakpoints from DNA samples.

Supplementary 2. Table S2: the breakpoints from RNA samples.

Supplementary 3. Table S3: gene frequency of breakpoints from DNA samples.

Supplementary 4. Table S4: gene frequency of breakpoints from RNA samples.

Supplementary 5. Table S5: pathway analysis of breakpoints.

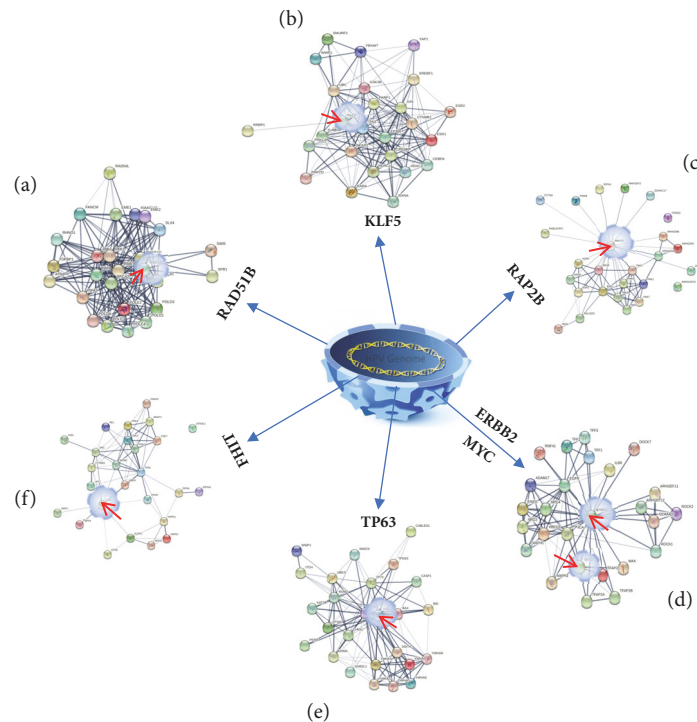*Supplementary 6.* Table S6: the 163 breakpoints from Hu et al.

FIGURE 4: The interaction network of interacting protein for hotspot genes. The figure revealed the interaction network of hotspot genes. Line thickness indicates the strength of data support.

References

[1] A. C. Rodríguez, M. Schiffman, R. Herrero et al., "Rapid clearance of human papillomavirus and implications for clinical focus on persistent infections," *Journal of the National Cancer Institute*, vol. 100, no. 7, pp. 513–517, 2008.

[2] A. P. Pinto and C. P. Crum, "Natural history of cervical neoplasia: Defining progression and its consequence," *Clinical Obstetrics and Gynecology*, vol. 43, no. 2, pp. 352–362, 2000.

[3] K.-C. Chao, P. H. Wang, M. S. Yen, and C. C. Chang, "Detection of HPV infection by analyzing the changes in structure of peripheral blood lymphocytes specifically induced by HPV E7 antigen," *European Journal of Gynaecological Oncology*, vol. 24, no. 1, pp. 30–32, 2003.

[4] K. Morshed, D. Polz-Gruszka, M. Szymański, and M. Polz-Dacewicz, "Human Papillomavirus (HPV)—structure, epidemiology and pathogenesis," *Otolaryngologia Polska*, vol. 68, no. 5, pp. 213–219, 2014.

[5] N. Wentzensen, S. Vinokurova, and M. Von Knebel Doeberitz, "Systematic review of genomic integration sites of human papillomavirus genomes in epithelial dysplasia and invasive cancer of the female lower genital tract," *Cancer Research*, vol. 64, no. 11, pp. 3878–3884, 2004.

[6] K. Akagi, J. Li, T. R. Broutian et al., "Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability," *Genome Research*, vol. 24, no. 2, pp. 185–199, 2014.

[7] A. Adey, J. N. Burton, J. O. Kitzman et al., "The haplotype-resolved genome and epigenome of the aneuploid HeLa cancer cell line," *Nature*, vol. 500, no. 7461, pp. 207–211, 2013.

[8] Z. Hu, D. Zhu, W. Wang et al., "Genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism," *Nature Genetics*, vol. 47, no. 2, pp. 158–163, 2015.

[9] C. Bodelon, M. E. Untereiner, M. J. Machiela, S. Vinokurova, and N. Wentzensen, "Genomic characterization of viral integration sites in HPV-related cancers," *International Journal of Cancer*, vol. 139, no. 9, pp. 2001–2011, 2016.

[10] A. I. Ojesina, L. Lichtenstein, and S. S. Freeman, "Landscape of genomic alterations in cervical carcinomas," *Nature*, vol. 506, no. 7488, pp. 371–375, 2013.

[11] C. Bodelon, S. Vinokurova, J. N. Sampson et al., "Chromosomal copy number alterations and HPV integration in cervical precancer and invasive cancer," *Carcinogenesis*, vol. 37, no. 2, pp. 188–196, 2015.

[12] B. Xu, S. Chotewutmontri, S. Wolf et al., "Multiplex Identification of Human Papillomavirus 16 DNA Integration Sites in Cervical Carcinomas," *PLoS ONE*, vol. 8, no. 6, Article ID e66693, 2013.

[13] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists," *Nucleic Acids Research*, vol. 37, no. 1, pp. 1–13, 2009.

[14] D. W. Huang, B. T. Sherman, and R. A. Lempicki, "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources," *Nature Protocols*, vol. 4, no. 1, pp. 44–57, 2009.

[15] K. Wang, M. Li, and H. Hakonarson, "ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data," *Nucleic Acids Research*, vol. 38, no. 16, article e164, 2010.

[16] W. J. Kent, C. W. Sugnet, T. S. Furey et al., "The human genome browser at UCSC," *Genome Research*, vol. 12, no. 6, pp. 996–1006, 2002.

[17] L. A. Lettice, S. J. H. Heaney, L. A. Purdie et al., "A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly," *Human Molecular Genetics*, vol. 12, no. 14, pp. 1725–1735, 2003.

[18] L. Li, J. A. Zhang, M. Dose et al., "A far downstream enhancer for murine Bcl11b controls its T-cell specific expression," *Blood*, vol. 122, no. 6, pp. 902–911, 2013.

[19] H. Greulich, B. Kaplan, P. Mertins et al., "Functional analysis of receptor tyrosine kinase mutations in lung cancer identifies oncogenic extracellular domain mutations of ERBB2," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 109, no. 36, pp. 14476–14481, 2012.

[20] R. Bose, S. M. Kavuri, A. C. Searleman et al., "Activating HER2 mutations in HER2 gene amplification negative breast cancer," *Cancer Discovery*, vol. 3, no. 2, pp. 224–237, 2013.

[21] N. Suwaki, K. Klare, and M. Tarsounas, "RAD51 paralogs: roles in DNA damage signalling, recombinational repair and tumorigenesis," *Seminars in Cell & Developmental Biology*, vol. 22, no. 8, pp. 898–905, 2011.

[22] K. Yoshino, T. Enomoto, T. Nakamura et al., "Aberrant FHIT transcripts in squamous cell carcinoma of the uterine cervix," *International Journal of Cancer*, vol. 76, no. 2, pp. 176–181, 1998.

[23] G. Wu, S. Nomoto, M. O. Hoque et al., "DeltaNp63alpha and TAp63alpha regulate transcription of genes with distinct biological functions in cancer and development," *Cancer research*, vol. 63, no. 10, pp. 2351–2357, 2003.

[24] S. Miyamoto, T. Suzuki, S. Muto et al., "Positive and Negative Regulation of the Cardiovascular Transcription Factor KLF5 by p300 and the Oncogenic Regulator SET through Interaction and Acetylation on the DNA-Binding Domain," *Molecular and Cellular Biology*, vol. 23, no. 23, pp. 8528–8541, 2003.

[25] S. N. Finver, K. Nishikura, L. R. Finger et al., "Sequence analysis of the MYC oncogene involved in the t(8;14)(q24;q11) chromosome translocation in a human leukemia T-cell line indicates that putative regulatory regions are not altered," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 85, no. 9, pp. 3052–3056, 1988.