Available Online: 8 September 2026 DOI: 10.54254/3049-5458/2026.36702

HPV integration landscape in precancerous lesion of cervical cancer

***Yuancheng Chen*[1], *Wei Chen*[1], *Yaping Li*[2], *Wei Wang*[2,3,*]**

[1]School of Engineering Medicine, Beihang University, Beijing, China
[2]Department of Obstetrics and Gynecology, The Fifth People's Hospital of Ningxia, Ningxia, China
[3]Department of Obstetrics and Gynecology, Peking Union Medical College Hospital; Chinese Academy of Medical Sciences and Peking Union Medical College; National Clinical Research Center for Obstetric and Gynecologic Diseases, Beijing, China

*Corresponding Author. Email: vivianwang_pumch@sina.com

Abstract. Persistent high-risk Human Papillomavirus (HPV) infection is the major cause of cervical cancer. The integration of HPV DNA into human genome is key event during cervical carcinogenesis. However, the viral integration profiles in precancerous lesion samples remain incompletely characterized. Here, this study identified HPV-human genome integration breakpoints in cervical precancerous lesion through high-risk HPV genome capture sequencing of 500 Thinprep Cytologic Test (TCT) samples. After sequencing reads alignment, breakpoints identification, contamination removal, and functional annotation, this study profiled 476 recurrent integration hotspot genes in our samples. After assigning hotspot genes to different lesion grade, patient age group, and infected HPV subtype according to the recurrent frequency of hotspot genes, this paper further compared the integration profile between above groups. In the comparison of lesion grades, high-grade lesions were more involved in mesenchymal development and membrane structure-related processes, while low-grade lesions were more involved in adhesion connections, cadherin proteins, and β-catenin-related functions. In the age group comparison, muscle cell development and protein kinase C binding were more enriched in elder group, while the genes in younger group were more involved in cell adhesion, catenin complex, and synaptic-related functions. In the HPV type comparison, HPV16/18-related genes mainly involved bicarbonate transport, one-carbon metabolism, and chromatin DNA binding, while other HPV type-related genes more involved cell adhesion, development, and membrane signal-related functions. These findings provide an overview of recurrent HPV integration-related hotspot genes and identify functional themes that may inform further studies of HPV-associated cervical precancerous lesions.

Keywords: human papillomavirus, viral integration, precancerous lesions, host genes, Gene Ontology

1. Introduction

Persistent infection with high-risk Human Papillomavirus (HPV) is the major cause of cervical precancerous lesions and cervical cancer. Although most HPV infections are transient, persistent infection can promote

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

cervical intraepithelial neoplasia and, in a subset of patients, invasive cervical cancer [1, 2]. The molecular changes accompanying this progression are complex and involve both viral and host genomic alterations.

Integration of HPV DNA into the human genome is an important event observed during cervical carcinogenesis. HPV integration may disrupt the viral genome, affect viral oncogene expression, and alter the genomic or regulatory environment surrounding an integration site [3-5]. Previous studies have shown that HPV integration is not randomly distributed throughout the human genome. Recurrent integration regions and host genes have been identified in cervical lesions and cervical cancer, suggesting that integration-associated genomic locations may provide information about virus-host interactions and lesion development [6-9].

The frequency and distribution of HPV integration may vary with lesion grade and HPV type [6, 10, 11]. However, recurrent HPV integration-related host genes have not been fully characterized in precancerous lesion samples. Previous studies have mainly examined integration status, type-specific profiles, recurrent genomic regions, or cervical cancer genomes [3, 6-11]. Large-scale genomic profiling of cervical carcinomas has identified recurrent HPV integration hotspots involving POU5F1B, FHIT, KLF12, KLF5, LRP1B, HMGA2, DLG2, and SEMA3D, with integration near MYC and HMGA2 associated with increased expression and integration within FHIT and LRP1B associated with reduced expression [12]. Further analysis of cervical cancer integration landscapes has highlighted recurrent loci involving KLF5, KLF12, MYC, TP63, RAD51B, and HMGA2, with integration hotspots frequently overlapping enhancer or super-enhancer regions and being associated with epithelial development and differentiation, cell proliferation, and DNA-repair-related functions [13]. More recent studies have identified additional integration-targeted genes, including IKZF2, IL26, AHRR, and PDCD6, and reported enrichment of cancer-related pathways such as cytokine–cytokine receptor interaction, JAK–STAT signaling, and PI3K–Akt signaling; HPV integration has also been shown to alter local CTCF binding, chromatin accessibility, transcription, and alternative splicing around integration sites [4, 9].

In this study, through high-risk HPV genome capture sequencing, this study profiled HPV-human genome integration breakpoints in 500 cervical precancerous lesion samples. Frequently observed HPV integration-related host genes were identified after breakpoint filtering and compared by lesion grade, age category, and HPV-type integration records. GO enrichment analysis and Jaccard-based GO term similarity networks were then used to summarize functional annotations in the corresponding gene groups. This paper aimed to characterize recurrent host genes, compare their observed distributions, and identify functional themes that may guide further investigation of HPV-associated cervical precancerous lesions.

2. Methods

2.1. Study samples and lesion-grade classification

The study included 500 precancerous lesion tissue samples obtained by Thinprep Cytologic Test (TCT). These samples were collected at Peking Union Medical College Hospital. This study was conducted in accordance with the principles of the Declaration of Helsinki and was approved by the ethics committees of the Peking Union Medical College Hospital. Lesion-grade information was obtained from colposcopy records, and samples were classified according to the recorded CIN or VaIN grade. Samples recorded as Cervical Intraepithelial Neoplasia Grade 1 (CIN1) or Vaginal Intraepithelial Neoplasia Grade 1 (VaIN1) were classified as low-grade lesions. Samples recorded as CIN2, CIN3, VaIN2, or VaIN3 were classified as high-grade lesions. 500 precancerous lesion tissue samples comprised 250 high-grade and 250 low-grade samples.

2.2. HPV-targeted sequencing and sequence processing

HPV-related sequences were enriched using hybridization-capture probes targeting 14 HPV types (including HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68). These capture probes were synthesized from Integrated DNA Technologies (IDT, Coralville, IA, USA). The captured sequencing libraries were subjected to paired-end sequencing on an AVITI platform (Element Bioscience San Diego, CA, USA). Raw sequencing reads were quality filtered to obtain clean paired-end reads for subsequent alignment and HPV integration analysis.

2.3. Read alignment and integration-breakpoint identification

Bwa-mem algorithm was used for alignment with default parameters. The reference genome adopted for alignment were generated through merging human hg19 with aforementioned 14 high-risk HPV reference genomes. After alignment, duplicated read-pairs were removed by MarkDuplicates function of GATK suite v4.4.0.

HPV-human integration events were defined as discordant read-pairs with one aligned to human genome and the other to HPV genome or chimeric read with part aligned to human and the remaining to HPV. Custom scripts were used to extract genomic coordinates, supporting-read types, and supporting-read counts of candidate integration breakpoints. All Candidate breakpoints were manually reviewed in Integrative Genomics Viewer v2.17.4 and annotated against hg19 using HOMER annotatePeaks.pl v5.1. Annotated breakpoint records were then summarized by host genes and originated patients for subsequent filtering and comparisons.

2.4. Cross-sample breakpoint filtering and host-gene selection

Breakpoint records were grouped by host gene and breakpoint segment. Breakpoint segments observed in seven or more samples were removed as potential cross-sample contamination. After this filtering step, host genes observed in at least five samples were retained, yielding 476 HPV integration-related hotspot genes.

2.5. Integration-associated sample-frequency comparisons

For each gene, sample frequency was defined as the number of unique samples in which at least one retained HPV integration event involving that gene was detected.

For the lesion-grade comparison, sample frequency was calculated for each gene within the high-grade and low-grade categories. Genes were classified according to whether the high-grade sample frequency was higher, the low-grade sample frequency was higher, or the two frequencies were equal.

For the age comparison, samples from patients aged at least 45 years were assigned to the older category, and samples from younger patients were assigned to the younger category. The categories contained 208 and 292 samples, respectively. Sample frequency was calculated for each gene within the older and younger categories, and genes were classified according to which category had the higher frequency or whether the two frequencies were equal.

For the HPV-type comparison, HPV-type labels were taken from gene-level integration records within each sample. Records containing HPV16 or HPV18 were assigned to the HPV16/18 category, whereas records containing other numeric HPV types were assigned to the other HPV-type category. For each gene, the number of unique samples represented in each category was calculated as the type-specific sample frequency. Mixed records allowed the same gene-sample unit to contribute to both categories. Genes were classified according to which category had the higher sample frequency or whether the two frequencies were equal.

All comparisons were based on absolute observed sample frequencies. Filtering, aggregation, and sample-frequency comparisons were performed using custom scripts.

2.6. Gene Ontology enrichment and term-similarity network analysis

For each comparison, the two higher-frequency gene lists were analyzed separately for Gene Ontology (GO) enrichment. Genes with equal category frequencies were not included in either GO input list. Enrichment analysis was performed in R v4.4.1 using clusterProfiler. Gene SYMBOLs were mapped with org.Hs.eg.db, and Biological Process, Cellular Component, and Molecular Function ontologies were analyzed.

For network visualization, up to six GO rows with the lowest nominal *p* values were selected for each comparison category and ontology. Pairwise term similarity was calculated from overlap between annotated gene sets using the Jaccard coefficient. Each node represented a GO term, node size reflected the annotated-gene count, and colored sectors denoted the relevant comparison categories. Terms represented in both categories contained sectors for both. Network edges represented annotated gene-set overlap between GO terms. Biological Process, Cellular Component, and Molecular Function prefixes identified each node's ontology in the network labels.

3. Results

3.1. Identification, genome-wide distribution, and representative breakpoint evidence

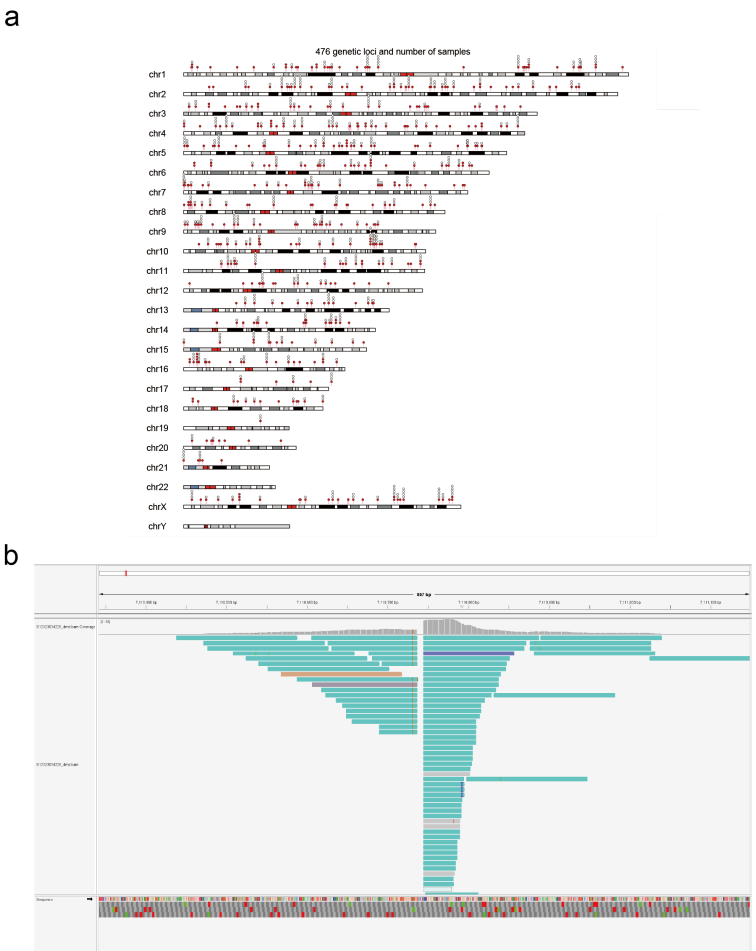

Figure 1. HPV-human genome integration profile. a) the distribution of integration breakpoint across human chromosomes. b) the example of sequencing reads distribution pattern around integration breakpoint

Note: a, The 476 retained hospot genes are plotted according to their hg19 chromosomal locations. Pink segments indicate annotated gene intervals, and recurrence symbols are positioned at the midpoint of each gene. Each red circle represents five samples, whereas each open circle represents one sample. The figure summarizes the genomic distribution and sample recurrence of the retained host genes rather than the precise coordinates of individual HPV–host integration breakpoints. b, The alignment view shows the host genomic region chr6: 7,110,342–7,111,147 (hg19). The upper panel indicates genomic coordinates and scale, the gray coverage histogram shows local sequencing depth, and the read-alignment track displays individual sequencing reads. Near the candidate junction, multiple reads terminate or begin at a concordant genomic position, accompanied by clipped or discontinuous alignment patterns and a local change in coverage. These features provide evidence supporting a candidate HPV–host integration breakpoint. Colored bases within individual reads indicate mismatches relative to the human reference sequence.

HPV integration data from 500 precancerous lesion samples were processed using the same breakpoint-filtering and host-gene selection workflow. After breakpoint segments observed in at least seven samples were removed as potential cross-sample contamination, host genes observed in at least five samples were retained, yielding 476 recurrent HPV integration-related hotspot genes. These genes were broadly distributed across the human genome rather than being restricted to a single chromosome or genomic region (Figure 1a).

Representative candidate integration breakpoints were further examined using Integrative Genomics Viewer. In sample S12123034228, reads aligned to the host genomic region chr6: 7,110,342–7,111,147 showed a local transition in read coverage and multiple alignments terminating or beginning near a concordant genomic position. The observed truncated or discontinuous alignment patterns in this region provide evidence supporting the candidate HPV-host integration sites (Figure 1b).

3.2. Lesion-grade comparison and GO enrichment

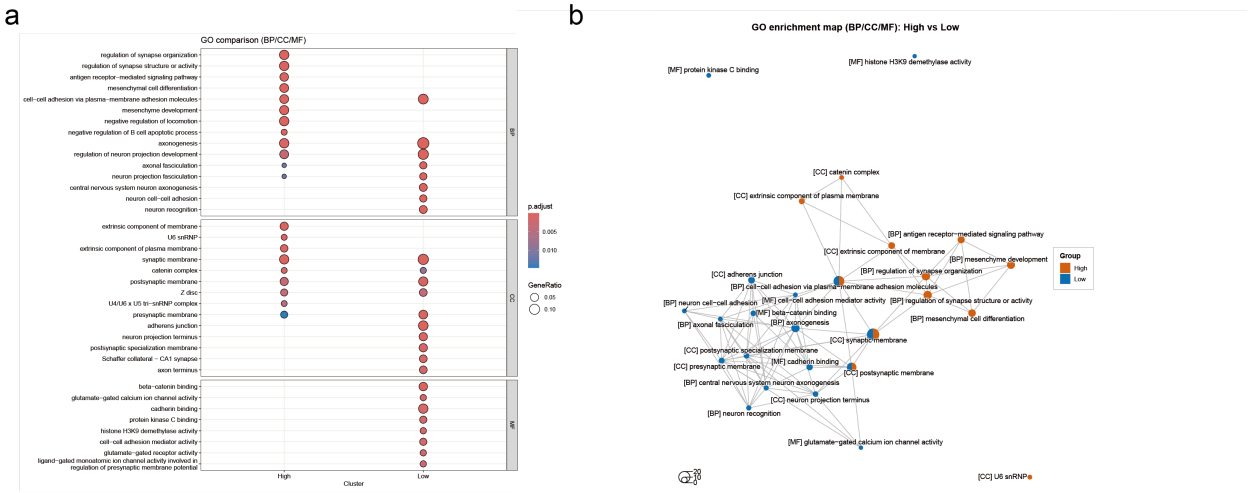

Figure 2. Functional comparison of HPV integration hotspot genes between different lesion grades. a) Dot plot of GO enrichment results of genes detected in high and low lesion grade. b) Jaccard-based GO term similarity network showed functional module differences between lesion grades

Note: a) Terms are separated into Biological Process (BP), Cellular Component (CC), and Molecular Function (MF). Dot size indicates GeneRatio, and dot color indicates the nominal *p* value displayed as p.adjust; *q* value Cutoff = 0.2 was retained. b), Nodes represent GO terms, node size indicates annotated-gene count, and orange and blue sectors denote high-grade and low-grade categories. Edges represent Jaccard similarity based on shared annotated genes between GO terms at a threshold of at least 0.1.

The lesion-grade comparison included 250 high-grade and 250 low-grade samples. Among the 476 genes, 266 showed higher integration-associated sample frequency in the high-grade group, 165 showed higher frequency in the low-grade group, and 45 showed equal sample frequency in the two groups. Because the two groups contained the same number of samples, the larger number of genes with higher sample frequency in the high-grade group indicates a broader representation of recurrent integration-related host genes in high-grade lesions.

GO analysis showed differences in the functional composition of the two gene sets. Genes with higher sample frequency in the high-grade group were represented by mesenchymal cell differentiation, mesenchyme development, extrinsic component of membrane, and catenin complex, whereas genes with higher sample frequency in the low-grade group were represented by axonogenesis, neuron recognition, adherens junction, cadherin binding, and beta-catenin binding. Cell-cell adhesion via plasma-membrane adhesion molecules was represented in both groups (Figure 2a and b). Overall, the high-grade group showed a relative shift toward mesenchymal-development and membrane-organization annotations, whereas the low-grade group retained stronger representation of neuronal-development and adherens-junction-related annotations. These differences may be relevant to changes in tissue organization during lesion advancement.

3.3. Age-category comparison and GO enrichment

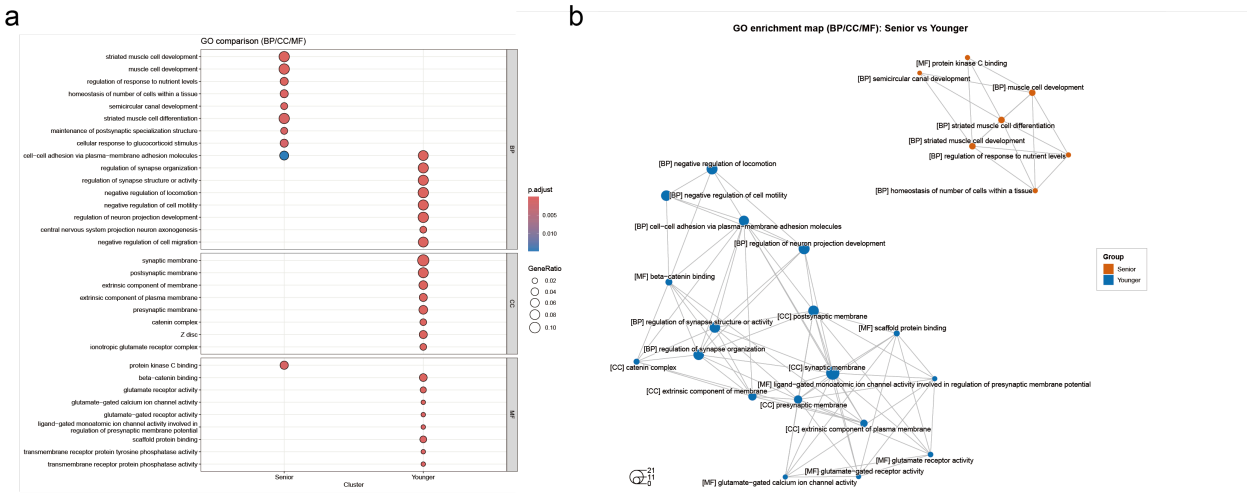

Figure 3. Functional comparison of HPV integration hotspot genes between age groups. a) Dot plot of GO enrichment results of genes detected in senior and younger groups. b) Jaccard-based GO term similarity network showed functional module differences between age groups

Note: patients with age higher or less than 45 were grouped into Senior and Younger group, respectively.

The age comparison included 208 samples from patients aged at least 45 years and 292 samples from younger patients. Among the 476 genes, 125 showed higher integration-associated sample frequency among patients aged at least 45 years, 303 showed higher sample frequency among younger patients, and 48 had equal sample frequency.

The two age categories showed different functional compositions. Genes with higher sample frequency among patients aged at least 45 years were represented by striated muscle cell development, muscle cell development, and protein kinase C binding. Genes with higher sample frequency among younger patients were more strongly represented by cell adhesion, catenin complex, beta-catenin binding, synaptic membrane,

postsynaptic membrane, and glutamate receptor activity (Figure 3a and b). Thus, the two age groups differed in the functional distribution of recurrent integration-related host genes, with older patients showing greater representation of muscle development and protein kinase C–related functions, whereas younger patients showed greater representation of adhesion-, catenin-, and synaptic signaling–related functions.

3.4. HPV16/18 and other HPV-type comparison and GO enrichment

Compare the integration situations of different HPV types. Records containing HPV16 or HPV18 were assigned to the HPV16/18 category, whereas records containing other numeric HPV types were assigned to the pooled other HPV-type category. If the same gene is integrated by both the 16/18 type HPV and other types of HPV, it will be recorded in both sets of categories simultaneously. Among the 476 genes, 45 genes showed high-frequency integration of HPV16/18, 405 genes were highly integrated by other HPV types, and the frequencies of 26 genes were equal in the samples.

Genes with higher HPV16/18-associated sample frequency were characterized by bicarbonate transport, bicarbonate transmembrane transporter activity, digestive system process, one-carbon metabolism, and chromatin DNA binding. In contrast, genes with higher sample frequency in the other HPV-type category were represented by cell-cell adhesion, axonogenesis, developmental pigmentation, adherens junction, beta-catenin binding, synaptic membrane, and transmembrane-ephrin receptor activity (Figure 4a and b). Thus, host genes more frequently associated with HPV16/18 and other HPV types differed in their functional distribution, with transport-, metabolism-, and chromatin-related functions more prominent in the HPV16/18 category and adhesion-, development-, and membrane signaling–related functions more prominent in the other HPV-type category.

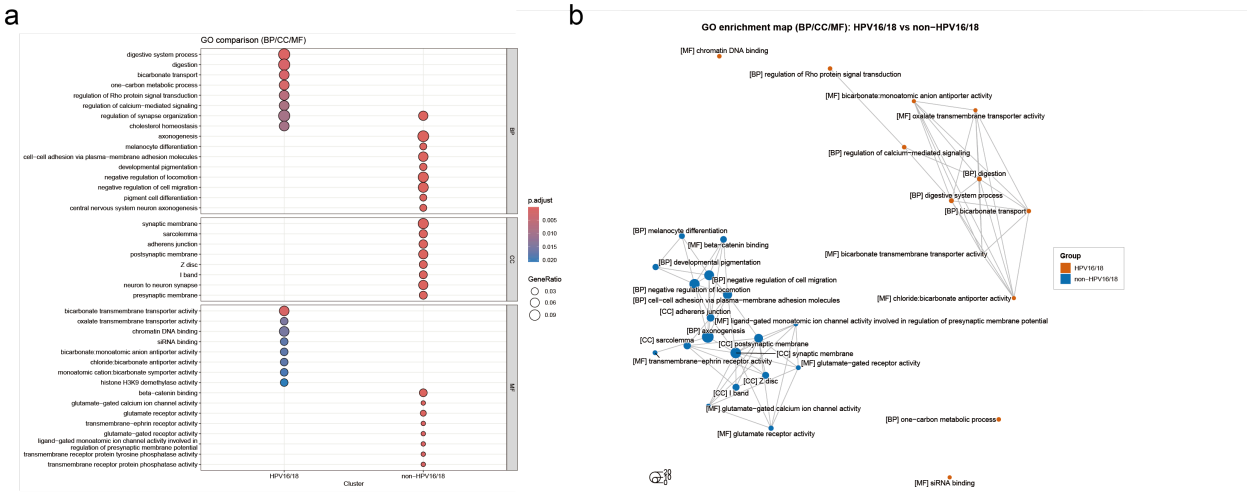

Figure 4. Functional comparison of HPV integration hotspot genes between infected HPV types. a) Dot plot of GO enrichment results of genes detected in HPV16/18 infection and other high-risk type infection groups. b) Jaccard-based GO term similarity network showed functional module differences between HPV groups

4. Discussion

In this study, the authors identified 476 recurrent HPV integration-related hotspot genes across 500 cervical precancerous lesion samples and found that these genes were distributed across human genome. Comparisons by lesion grade, age, and HPV type further revealed distinct patterns in both the sample frequencies and

functional annotations of the integration-related host genes. Adhesion-, membrane-, and junction-related functions appeared repeatedly across the three comparisons, whereas mesenchymal development, protein kinase C binding, chromatin DNA binding, bicarbonate transport, and neuronal or synaptic functions showed more group-specific patterns. Together, these findings indicate that the host-gene landscape associated with recurrent HPV integration varies with both clinical and viral characteristics.

In the comparison of lesion grades, there were 250 samples for both the high-grade and low-grade lesions. Among the 476 host genes, 266 genes showed a higher frequency of integration in samples from high-grade lesions, while 165 genes had a higher frequency in samples from low-grade lesions, and another 45 genes had the same frequency in both groups. This result indicates that the frequency distribution of HPV-integrated host genes is overall more skewed towards higher-grade lesions. Previous studies have observed the relationship between HPV integration and the severity of cervical lesions. Tian et al. in a prospective cohort study found that the positive rate of HPV integration increased as the cervical pathological grade increased [6, 14]. This study further demonstrated at the host gene level that as the lesion severity changed, the host gene profile repeatedly affected by HPV integration also underwent corresponding changes, which was consistent with the overall cumulative trend of HPV integration during the development of cervical lesions.

The functional differences between high- and low-grade lesions further support this trend. Genes with higher sample frequency in high-grade lesions were enriched for mesenchymal cell differentiation, mesenchyme development, extrinsic component of membrane, and catenin complex, whereas genes with higher frequency in low-grade lesions were more strongly associated with adherens junction, cadherin binding, beta-catenin binding, and some functions related to neural development. Cell-cell adhesion via plasma-membrane adhesion molecules was represented in both groups. Previous studies have shown that during the progression from low-grade to high-grade cervical epithelial lesions, there is a continuous remodeling of cell adhesion and epithelial structure. The decrease in E-cadherin expression is associated with an increase in the severity of CIN. Another study that examined the progression from normal cervix, CIN to cervical squamous cell carcinoma found that the expressions of E-cadherin and β-catenin gradually decreased, while mesenchymal markers such as N-cadherin, vimentin, and fibronectin gradually increased [15, 16]. Therefore, the functions related to adherens junction, cadherin and β-catenin shown in the lower grade lesions in this study, and the more prominent functions related to mesenchymal development and membrane structure in the higher-grade lesions. This is consistent with the biological changes of epithelial junction remodeling and mesenchymal feature enhancement during cervical lesions. The functional composition changes of the host genes related to HPV integration may thereby reflect the alterations in the cellular tissue structure and the environment of cell-to-cell interactions during the progression of lesion severity.

The age comparison revealed a different pattern of functional organization. The group with an age of at least 45 years included 208 samples, while the group with an age less than 45 years included 292 samples. There were 125 and 303 genes respectively that showed a higher frequency of integrated-related samples in the corresponding age groups. Genes with higher sample frequency among patients aged at least 45 years were enriched for striated muscle cell development, muscle cell development, and protein kinase C binding, whereas genes with higher frequency among younger patients were more strongly associated with cell adhesion, catenin complex, beta-catenin binding, synaptic membrane, postsynaptic membrane, and glutamate receptor activity. Age is closely related to the natural history of HPV infection. The duration of HPV infection is an important factor in the formation of high-grade cervical lesions. Age is also associated with the characteristics of persistent high-risk HPV infection and viral clearance. Moreover, it has been observed that in older women with persistent HPV infection, there is a decline in overall immune response capacity [17-19]. These age-related viral infections and changes in the host immune environment provide a biological

background for the emergence of different HPV integration-related host gene functional profiles in patients of different ages. The different long-term interactions between the virus and the host may further influence the range and functional composition of host genes involved in HPV integration events.

The occurrence of protein kinase C binding in the at least 45-year-old group is particularly noteworthy. There is a known connection between PKC and the life cycle of the papillomavirus. Experimental studies have shown that PKC can participate in the activation of the late promoter of HPV31 and the amplification of the viral genome, and is related to the regulation of the viral life cycle and host cell differentiation-dependent regulation [20]. Therefore, the occurrence of PKC-binding-related functions in the older group forms a connection worthy of further attention between this group's integrated host genes and the papillomavirus life cycle and host cell differentiation regulation. In contrast, the re-occurrence of cell adhesion, catenin complex and beta-catenin binding in the younger group corresponds to the repeatedly observed themes of cell adhesion and cell connection in the lesion grade comparison, indicating that these host functions may be an important biological background shared by HPV integrated-related genes in different clinical groups. The muscle cell development and nerve, synaptic-related functions that appear simultaneously in the age comparison constitute the unique functional characteristics of this group, and the specific contributing genes and their roles in HPV-related lesions still need further analysis.

The comparison of HPV types also reveals significantly different host gene functional profiles. Among 476 genes, 45 genes have a higher frequency in samples related to HPV16/18, 405 genes have a higher frequency in records related to other HPV types, and another 26 genes have the same frequency in both types of records. The other HPV types consist of multiple genotypes, so the core information reflected by this comparison is that the host gene functional compositions corresponding to HPV16/18 and other HPV types are different. Existing studies have shown that different high-risk HPV genotypes have different integration characteristics. There are different integration patterns between HPV16 and HPV18, and comparisons of HPV16, HPV18, HPV31, HPV33, and HPV45 also show that there are significant differences in integration frequencies and APOBEC3 profiles among different viral phylogenetic branches [10, 11, 14]. Therefore, the HPV type-related host gene differences observed in this study are consistent with the previously discovered genotype-dependent integration characteristics, and further indicate that different HPV types may correspond to different host genomic functional environments.

In terms of specific functions, the genes related to HPV16/18 with higher frequencies are mainly enriched in bicarbonate transport, bicarbonate transmembrane transporter activity, one-carbon metabolism, and chromatin DNA binding. While the genes related to other HPV types are more involved in cell-cell adhesion, adherens junction, beta-catenin binding, axonogenesis, and membrane and synapse-related functions. Among them, chromatin DNA binding has a relatively direct biological connection with HPV integration itself. Existing studies have found that HPV integration can be accompanied by changes in chromatin state near the integration site, changes in CTCF binding patterns, and increased chromatin accessibility, and further affects the transcription and splicing of adjacent host genes [4]. Therefore, the enrichment of chromatin DNA binding in HPV16/18-related host genes makes chromatin regulation an important direction for understanding this group of integration-related genes. At the same time, one-carbon metabolism is associated with DNA methylation and the occurrence of cervical cancer, and the ion transport system is also involved in various biological processes of cervical cancer cells [21, 22]. These existing studies provide corresponding biological backgrounds for the metabolic and transmembrane transport-related enrichments in the HPV16/18 group.

5. Conclusion

Overall, this study reveals the different functional characteristics between lesion grade, age, and HPV type from the perspective of integration-related host genes of recurrent HPV. The lesion grade comparison shows that more host genes have a higher frequency of integration-related samples in high-grade lesions, and there is a change from cell adhesion and connection-related functions to mesenchymal development and membrane structure-related functions; the age comparison shows that the integration-related host genes of different age patients involve different cell differentiation, PKC binding, adhesion, and membrane-related functions; the HPV type comparison further shows the differences in chromatin, metabolism, transport, and adhesion-related functions between HPV16/18 and other HPV types. The three comparisons jointly indicate that the host genes involved in HPV integration do not present a single functional pattern, but form different functional combinations depending on clinical and viral characteristics. The recurrent cell adhesion, cell connection, and membrane structure-related functions provide a key direction for further understanding the relationship between HPV integration and the tissue changes in precancerous lesions.

Author contribution

Y.C. was responsible for data analysis and manuscript writing. W.C. provided guidance on data analysis. Y.L. was responsible for sample collection. W.C. and W.W. designed the experiments and supervised the entire research process. All authors critically reviewed the draft, provided important feedback, and collectively approved the final version of the manuscript.

Funding

The Key Research and Development Program of Ningxia Hui Autonomous Region (2022BEG02055)

Ethics approval and consent to participate

The studies involving human participants were reviewed and approved by the Ethics Committee of Peking Union Medical College Hospital (approval No. I-25PJ1294). All patients or participants provided written informed consent to participate in the study.

References

[1] Koshiol, J., Lindsay, L., Pimenta, J. M., Poole, C., Jenkins, D., & Smith, J. S. (2008). Persistent human papillomavirus infection and cervical neoplasia: A systematic review and meta-analysis. *American Journal of Epidemiology, 168*(2), 123–137. https://doi.org/10.1093/aje/kwn036

[2] Schiffman, M., & Wentzensen, N. (2013). Human papillomavirus infection and the multistage carcinogenesis of cervical cancer. *Cancer Epidemiology, Biomarkers & Prevention, 22*(4), 553–560. https://doi.org/10.1158/1055-9965.EPI-12-1406

[3] Burk, R. D., Chen, Z., Saller, C., Tarvin, K., Carvalho, A. L., Scapulatempo-Neto, C., Silveira, H. C., Fregnani, J. H., Creighton, C. J., Anderson, M. L., Castro, P., Wang, S. S., Yau, C., Benz, C., Robertson, A. G., Mungall, K., Lim, L., Bowlby, R., Sadeghi, S., … University of North Carolina at Chapel Hill. (2017). Integrated genomic and molecular characterization of cervical cancer. *Nature, 543*(7645), 378–384. https://doi.org/10.1038/nature21386

[4] Karimzadeh, M., Arlidge, C., Rostami, A., Lupien, M., Bratman, S. V., & Hoffman, M. M. (2023). Human papillomavirus integration transforms chromatin to drive oncogenesis. *Genome Biology, 24*(1), 142. https://doi.org/10.1186/s13059-023-02926-9

[5] Jeon, S., & Lambert, P. F. (1995). Integration of human papillomavirus type 16 DNA into the human genome leads to increased stability of E6 and E7 mRNAs: Implications for cervical carcinogenesis. *Proceedings of the National Academy of Sciences, 92*(5), 1654–1658. https://doi.org/10.1073/pnas.92.5.1654

[6] Tian, X., Weng, D., Chen, Y., Wang, Y., Li, X., Wang, X., Cao, C., Gong, D., Zeng, Z., Wu, Q., Wang, X., Wu, P., Fan, L., Zhang, Q., Wang, H., Hu, Z., Cheng, X., & Ma, D. (2024). Risk assessment and triage strategy of cervical cancer primary screening on HPV integration status: 5-year follow-up of a prospective cohort study. *Journal of the National Cancer Center, 4*(4), 311–317. https://doi.org/10.1016/j.jncc.2024.08.001

[7] Zhao, J., Zheng, W., Wang, L., Jiang, H., Wang, X., Hou, J., Xu, A., & Cong, J. (2023). Human papillomavirus (HPV) integration signature in cervical lesions: Identification of MACROD2 gene as HPV hot spot integration site. *Archives of Gynecology and Obstetrics, 307*(4), 1115–1123. https://doi.org/10.1007/s00404-022-06748-1

[8] Schmitz, M., Driesch, C., Jansen, L., Runnebaum, I. B., & Dürst, M. (2012). Non-random integration of the HPV genome in cervical cancer. *PLOS ONE, 7*(6), e39632. https://doi.org/10.1371/journal.pone.0039632

[9] Xu, X.-S., Ma, Y.-S., Dai, R.-H., Zhang, H.-L., Yang, Q.-X., Fan, Q.-Y., Liu, X.-Y., Liu, J.-B., Feng, W.-W., Meng, H., Fu, D., Yu, H., & Shen, J. (2024). Identification of novel genomic hotspots and tumor-relevant genes via comprehensive analysis of HPV integration in Chinese patients of cervical cancer. *American Journal of Cancer Research, 14*(9), 4665–4682. https://doi.org/10.62347/KKLE8602

[10] Lagström, S., Løvestad, A. H., Umu, S. U., Ambur, O. H., Nygård, M., Rounge, T. B., & Christiansen, I. K. (2021). HPV16 and HPV18 type-specific APOBEC3 and integration profiles in different diagnostic categories of cervical samples. *Tumour Virus Research, 12*, 200221. https://doi.org/10.1016/j.tvr.2021.200221

[11] Løvestad, A. H., Repesa, A., Costanzi, J.-M., Lagström, S., Christiansen, I. K., Rounge, T. B., & Ambur, O. H. (2022). Differences in integration frequencies and APOBEC3 profiles of five high-risk HPV types adheres to phylogeny. *Tumour Virus Research, 14*, 200247. https://doi.org/10.1016/j.tvr.2022.200247

[12] 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., … Ma, D. (2015). Genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism. *Nature Genetics, 47*(2), 158–163. https://doi.org/10.1038/ng.3178

[13] Warburton, A., Markowitz, T. E., Katz, J. P., Pipas, J. M., & McBride, A. A. (2021). Recurrent integration of human papillomavirus genomes at transcriptional regulatory hubs. *npj Genomic Medicine, 6*(1), 101. https://doi.org/10.1038/s41525-021-00264-y

[14] Vinokurova, S., Wentzensen, N., Kraus, I., Klaes, R., Driesch, C., Melsheimer, P., Kisseljov, F., Dürst, M., Schneider, A., & von Knebel Doeberitz, M. (2008). Type-dependent integration frequency of human papillomavirus genomes in cervical lesions. *Cancer Research, 68*(1), 307–313. https://doi.org/10.1158/0008-5472.CAN-07-2754

[15] Branca, M., Giorgi, C., Ciotti, M., Santini, D., Di Bonito, L., Costa, S., Benedetto, A., Bonifacio, D., Di Bonito, P., Paba, P., Accardi, L., Mariani, L., Syrjänen, S., Favalli, C., Syrjänen, K., & HPV-PathogenISS Study Group. (2006). Down-regulation of E-cadherin is closely associated with progression of cervical intraepithelial neoplasia (CIN), but not with high-risk human papillomavirus (HPV) or disease outcome in cervical cancer. *European Journal of Gynaecological Oncology, 27*(3), 215–223

[16] Jiang, J., Li, X., Yin, X., Zhang, J., & Shi, B. (2019). Association of low expression of E-cadherin and β-catenin with the progression of early stage human squamous cervical cancer. *Oncology Letters, 17*(6), 5729–5739. https://doi.org/10.3892/ol.2019.10266

[17] Rodríguez, A. C., Schiffman, M., Herrero, R., Hildesheim, A., Bratti, C., Sherman, M. E., Solomon, D., Guillén, D., Alfaro, M., Morales, J., Hutchinson, M., Katki, H., Cheung, L., Wacholder, S., & Burk, R. D. (2010). Longitudinal study of human papillomavirus persistence and cervical intraepithelial neoplasia grade

2/3: Critical role of duration of infection. *Journal of the National Cancer Institute, 102*(5), 315–324. https://doi.org/10.1093/jnci/djq001

[18] Li, W., Meng, Y., Wang, Y., Cheng, X., Wang, C., Xiao, S., Zhang, X., Deng, Z., Hu, M., Shen, P., Xu, S., Fu, C., Jiang, W., Wu, B., Li, K., Chen, G., Wei, J., Xi, L., Hu, J., … Wu, P. (2019). Association of age and viral factors with high-risk HPV persistence: A retrospective follow-up study. *Gynecologic Oncology, 154*(2), 345–353. https://doi.org/10.1016/j.ygyno.2019.05.026

[19] García-Piñeres, A. J., Hildesheim, A., Herrero, R., Trivett, M., Williams, M., Atmetlla, I., Ramírez, M., Villegas, M., Schiffman, M., Rodríguez, A. C., Burk, R. D., Hildesheim, M., Freer, E., Bonilla, J., Bratti, C., Berzofsky, J. A., & Pinto, L. A. (2006). Persistent human papillomavirus infection is associated with a generalized decrease in immune responsiveness in older women. *Cancer Research, 66*(22), 11070–11076. https://doi.org/10.1158/0008-5472.CAN-06-2034

[20] Bodily, J. M., Alam, S., & Meyers, C. (2006). Regulation of human papillomavirus type 31 late promoter activation and genome amplification by protein kinase C. *Virology, 348*(2), 328–340. https://doi.org/10.1016/j.virol.2006.01.036

[21] Pathak, S., Bhatla, N., & Singh, N. (2012). Cervical cancer pathogenesis is associated with one-carbon metabolism. *Molecular and Cellular Biochemistry, 369*(1–2), 1–7. https://doi.org/10.1007/s11010-012-1362-3

[22] Chen, Y.-F., & Shen, M.-R. (2022). The important role of ion transport system in cervical cancer. *International Journal of Molecular Sciences, 23*(1), 333. https://doi.org/10.3390/ijms23010333