Association of high-risk human papillomavirus infection duration and cervical lesions with vaginal microbiota composition

Jun Liu[1]#, Mei Luo[2]#, Yang Zhang[1], Guangming Cao[1], Shuzhen Wang[1]

[1]Department of Obstetrics and Gynecology, Beijing Chao-Yang Hospital, Capital Medical University, Beijing, China; [2]Department of Obstetrics and Gynecology, Lu-He Teaching Hospital, Capital Medical University, Beijing, China

Contributions: (I) Conception and design: J Liu, M Luo; (II) Administrative support: G Cao; (III) Provision of study materials or patients: Y Zhang; (IV) Collection and assembly of data: M Luo; (V) Data analysis and interpretation: G Cao; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

#These authors contributed equally to this work.

Correspondence to: Shuzhen Wang. Department of Obstetrics and Gynecology, Beijing Chao-Yang Hospital, Capital Medical University, no. 8 Gongti South Road, Chaoyang District, Beijing, China. Email: darrywang2003@163.com.

Background: Cervical cancer is reportedly caused by the synergistic effects of persistent high-risk human papillomavirus (HPV) infection. Cervical microbiota represent a unique and dynamically changing microecological system that is directly exposed to the vagina. The relationship between HPV and the composition of the cervical microbiome has long been a primary focus of research.

Methods: To determine the specific differential florae throughout the process of cervical cancer development, in the present study, 16S rRNA sequencing was combined with KEGG pathway enrichment analysis to analyse five groups of cervical scraping samples with increasing durations of HPV infection and cervical intraepithelial neoplasia pathological classification.

Results: The findings revealed that decreasing levels of probiotics, including *Shuttleworthia*, *Prevotella*, *Lactobacillus*, and *Sneathia*, and increasing levels of pathogenic bacteria, including *Dispar*, *Streptococcus*, and *Faecalibacterium prausnitzii*, could be the direct result of early HPV infection. Other pathogenic bacteria, such as *Bifidobacteriaceae*, might represent key factors in cancer progression. Additionally, KEGG pathway enrichment analysis indicated that HPV infection directly inhibits multiple pathways, including those of sporulation, porphyrin and chlorophyll metabolism, arginine and proline metabolism, isoquinoline alkaloid biosynthesis, and ansamycin biosynthesis, which may lead to the development of early symptoms of cervical cancer. Biomarkers were predicted based on operational taxonomic unit (OTU) abundance data, and OTU851726 and OTU715913 were undoubtedly the best potential indicators of cervical cancer.

Conclusions: The findings of the present study could assist with the development of a guideline for screening new clinical drugs for cervical cancer.

Keywords: Cervical cancer, human papillomavirus infection (HPV infection); cervical intraepithelial neoplasia pathological classification; cervical microbiota

Introduction

Cervical cancer is currently the second leading cause of cancer-related deaths among women worldwide, and is closely associated with continuous infection with human papillomavirus (HPV) (1). Microtrauma during HPV infection can destroy the biologic barrier formed by the local vaginal immune microenvironment, leading to the multiplication of a large number of abnormal florae (2). The resulting local microecological imbalance in the vagina can subsequently upregulate the expression of the HPV protein

and increase the adhesion, invasion, and colonization of abnormal flora. This process ultimately leads to an aggravation of the imbalance in the vaginal microecosystem, thereby forming a vicious cycle and eventually leading to cervical cancer (3).

HPVs are non-enveloped small DNA viruses that widely exist in nature and encompass a large family of viruses, both benign and highly carcinogenic (4). HPVs primarily replicate on the surface of squamous epithelia, such as at the junction of squamous epithelial cells and the columnar epithelium of the cervix; this viral replication can interfere with the normal replication of squamous cells, leading to abnormal proliferation of the squamous epithelia (5). Generally, as a non-lytic infection, this process does not result in a strong immune response or inflammation by activating Toll-like receptors and natural killer cells (6). Currently, more than 170 types of HPV have been discovered, and >40 of these are associated with female reproductive tract infections and cervical lesions (7). HPV types can be broadly divided into high- and low-risk groups according to their differences in pathogenicity. Low-risk types, such as HPV6, HPV11, and HPV30, can cause benign lesions, such as lower-grade squamous intraepithelial lesions (LSIL) and genital warts, whereas high-risk types, such as HPV16, HPV18, and HPV58, mainly cause higher-grade squamous intraepithelial lesions (HSIL) or lead to cervical cancer (8).

However, HPV infection is necessary but not sufficient for the development of cervical precancerous and invasive disease. The contribution of the microenvironment of the vaginal and cervical florae to the occurrence and development of cervical precancerous lesions has gained increased interest in recent years. Numerous studies have shown that bacterial vaginosis and bacterial vaginosis-associated bacteria can influence immune parameters within the vagina, including cytokine/chemokine levels (9), antimicrobial protein levels (10), and immune cell populations (11). Generally, healthy, premenopausal vaginal bacterial communities, comprising *Lactobacillus* spp., which function to ensure a low pH, provide the first-line defense against pathogens (12). Using recently developed next-generation sequencing platforms, Branch classified the vaginal bacteria into five community state types, according to the microbiome structure, mainly based on the dominance of *Lactobacillus crispatus*, *Lactobacillus gasseri*, *Lactobacillus iners*, and *Lactobacillus jensenii* (13).

As one of the most complex ecosystems among the four major ecosystems in the human body, the vaginal microecosystem has been an important focus of research (14). By resisting the invasion of external pathogens, vaginal microbial florae play an important role in human reproduction (15). Currently, there is consensus that the synergistic effects of persistent high-risk HPV infection, changes in the cervical microenvironment, and other auxiliary carcinogenic factors (i.e., prolificacy, sexual disorders, and smoking) facilitate the development of cervical precancerous lesions (16). However, the cervix, which represents a unique and dynamic microecological system, is continually affected by the microenvironment in the vagina, which makes it challenging to identify the specific differential florae, and therefore, greatly increases the difficulty of related research. In the present study, 16S rRNA sequencing and KEGG pathway enrichment analysis were performed to clarify the relationship between the vaginal flora microenvironment and the occurrence and development of HPV infection and cervical precancerous lesions. With a comprehensive analysis of the genetic information of the microbiome at the five key stages in the development of cervical cancer, the types and abundance of the differential florae, such as *Sneathia* and *Ruminococcaceae*, were determined and compared. Furthermore, using KEGG pathway enrichment analysis, the major differential pathways in the development of cervical cancer were determined, and florae, including operational taxonomic unit (OTU) 851726 and OTU715913, were identified as potential biomarkers of cervical cancer. We present the following article in accordance with the MDAR reporting checklist (available at http://dx.doi.org/10.21037/atm-20-5832).

Methods

Sample collection and study design

The samples of 122 patients (91 patients with high-risk HPV infection, including 48 persistent infections and 43 transient infections, and 31 cases as negative controls. HPV subtypes included HPV16, HPV18, HPV56, HPV52, and HPV 58) were divided into five groups based on the duration of HPV infection and cervical pathological classification: (I) high-risk HPV persistent infection with HSIL group (persistent infection with HSIL); (II) high-risk HPV persistent infection with LSIL group (persistent infection with LSIL); (III) high-risk HPV transient infection with HSIL group (transient infection with HSIL); (IV) high-risk HPV transient infection with LSIL group

(transient infection with LSIL); and (V) non-infected HPV group. Generally, persistent infection means the same HPV subtype has infected for >12 months, whereas transient infection means the same HPV subtype has regressed within 1 year.

The clinical study was approved by the Ethic Committee of Beijing Chao-Yang Hospital (No. 2019-331) and was conducted in accordance with the provisions of the Declaration of Helsinki (as revised in 2013). Written informed consent was obtained from all participants before enrolment.

Bacterial DNA extraction and quality control

DNA samples were extracted from vaginal secretions using the MOBIO PowerFecal DNA isolation kit, and then tested using quantitative agarose gel electrophoresis. Samples with a quality of at least class B were used for subsequent experiments [class B samples standards: the size of the polymerase chain reaction (PCR) product target band is correct, and the total amount meets the needs of 1, but less than 2, times the database construction, DNA concentration >5 ng/µL, and DNA mass >150 ng]. The detection parameters were as follows: for genomic DNA, the gel concentration was 1%, the voltage was 100 V, and electrophoresis time was 40 minutes; for the PCR product, the gel concentration was 2%, the voltage was 80 V, and electrophoresis time was 40 minutes.

16S rRNA amplicon pyrosequencing

A PCR-free library was constructed by sequencing using Hiseq 2500 PE250 (illumina) after the amplified PCR products passed the aforementioned test. The original offline data volume of each sample was not <50,000 tags. Specifically, the primers were first diluted and labeled according to their barcode and stored at –20 ℃. The PCR reaction mixtures were then prepared, and PCR reactions were run with the following settings: initial denaturation at 94 ℃ for 3 minutes, followed by 35 cycles of denaturation at 94 ℃ for 1 minute, annealing at 55 ℃ for 1 minute, and extension at 72 ℃ for 1 minute, with a final extension at 72 ℃ for 10 minutes. The quality of the PCR amplicons was finally assessed by running 4 µL of each combined PCR reaction on 1% agarose gel stained with ethidium bromide. Only the positive amplicons that appeared in an expected band size of 375–425 bp under an ultraviolet wavelength of 260 nm were included in the subsequent steps.

Sequencing and data analysis

Each amplicon was quantified by a double-stranded DNA fluorescent nucleic acid stain to obtain an equimolar concentration pool of all PCR samples. FASTQ files were obtained by running the sequencing machine according to the manufacturer's instructions, and the overlapping paired-end FASTQ files were processed in a data curation pipeline implemented in QIIME 2, version 2017.7.017. The Greengenes 13.5 close reference OTU table was used to predict the KO table and the KEEG level 3 function table. STAMP was used to analyze differences by comparing the KEGG pathway of different group samples. Finally, KEGG-based heatmap analysis and bar plot analysis were performed.

Results

Clinical data grouping and tree map charts

To determine the influence of HPV infection duration and cervical intraepithelial neoplasia (CIN) pathological classification on the changes in the microbial population structure, the microbiome of 122 samples was analyzed. The changes in the florae in each group were observed using a maptree chart by categorizing the florae by OTU abundance, and 200 OTUs with the highest abundance were selected. As shown in *Figure 1*, the microbial composition in HPV-infected cervical pathology samples (*Figure 1A,B,C,D*) was significantly different from that of the non-infected HPV group (*Figure 1E*). In all of the HPV-infected tissues, *Firmicutes* always occupied the most central and most important position, whereas in the non-infected HPV group, the center position was covered by many other bacteria, including *Fusobacteria*, *Tenericutes*, *Proteobacteria*, and *Bacteroidetes*. Moreover, the microbiome composition of the HPV transient infection group (groups 3 and 4) was significantly different from that of the high-risk HPV persistent infection groups (groups 1 and 2), but was more similar to that of the non-infected HPV group. For instance, the high-risk HPV transient infection group (group 3) and the non-infected HPV group shared many bacteria that were not found in the persistent HPV infection groups, such as *Cyanobacteria*, TM7, WWE1, and WS3, thereby confirming the influence of the duration of HPV infection on the microbiome composition.

Microbiome differences among the five groups

To determine the specific differences of the microbiome
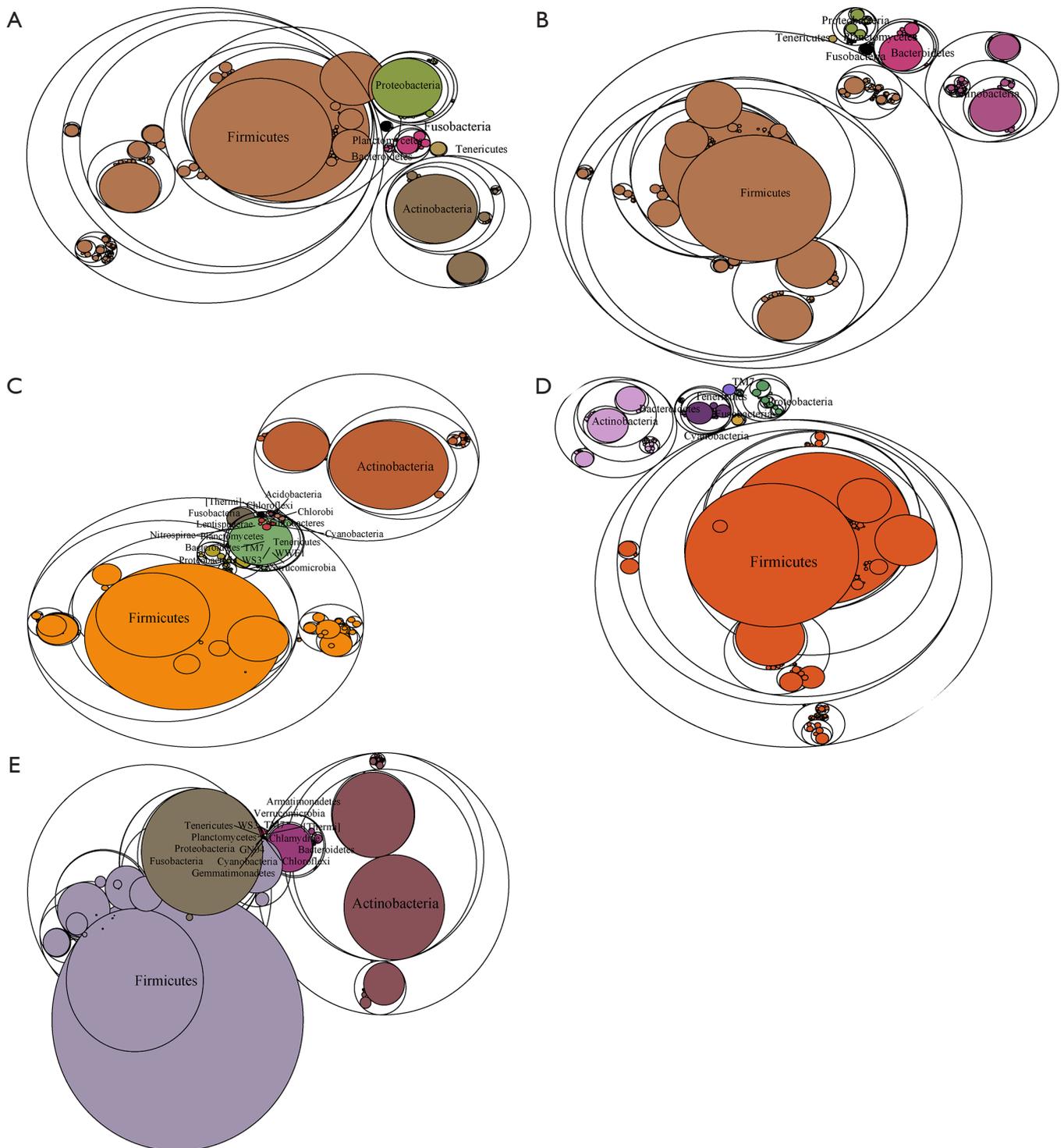
Figure 1 Maptree chart of the 200 florae with the highest operational taxonomic unit (OTU) abundance in the five groups of cervical secretion samples. The largest circle represents the kingdom level, and the gradually decreasing circles represent the phylum, class, and order, respectively. (A) Group 1: high-risk human papillomavirus (HPV) persistent infection with higher-grade squamous intraepithelial lesion (HSIL) group. (B) High-risk HPV persistent infection with LSIL group. (C) Group 3: high-risk HPV transient infection with HSIL group. (D) Group 4: high-risk HPV transient infection with LSIL group. (E) Group 5: non-infected HPV group.

composition in the non-infected HPV and HPV infection groups, a comparison of the five groups was made using a volcanic map (*Figure 2*). The changes in microorganisms throughout the deterioration process were clarified by comparing the groups with adjacent symptoms. As shown in *Figure 2A,B,C,D*, the most significant changes in the microbiome occurred in the comparison between the high-risk HPV transient infection with LSIL group (group 4) and the non-infected HPV group, indicating the influences of HPV infection on the microbiome composition. Moreover, compared with the non-infected HPV group, HPV infection decreased the level of probiotics, including *Shuttleworthia*, *Prevotella*, *Lactobacillus*, and *Sneathia*, while increasing the level of pathogenic bacteria, including *Dispar*, *Streptophyta*, and *Faecalibacterium prausnitzii*; this was more obvious in the comparison between the high-risk HPV persistent infection with HSIL group (group 1) and the non-infected HPV with no cervical lesions group (*Figure 2E*). Furthermore, the types of pathogenic bacteria, especially *Bifidobacteriaceae*, increased substantially when the duration of HPV infection and the CIN pathological grade increased, indicating that *Bifidobacteriaceae* may play a key role in cancer progression (*Figure 2B,C*). In addition to *Bifidobacteriaceae*, other families of bacteria, such as *Corynebacterium*, *Aerococcus*, and *Oscillospira*, were also greatly affected by the duration of HPV infection. These findings could assist in the development of a guideline for screening new drugs for cervical cancer from the microbiome perspective.

Statistical analysis of differential florae types

To further explore the specific differential florae in the different stages of cervical cancer, taxonomic profiles of the five groups were analyzed using STAMP, and the differences among groups were calculated using analysis of variance. The differential florae at the genus level are shown in *Table S1*. Among the differential florae, *Sneathia* showed the greatest difference among all groups, with minimum P values (corrected) of 8.37E-04 and a maximum effect size of approximately 0.15; these findings indicate the importance of *Sneathia* in the oncogenicity of the cervix. Furthermore, according to the most representative differential florae between adjacent groups, *Sneathia* and *Megasphaera* possessed the greatest difference in mean proportions with 95% confidence intervals (*Figure 3*). However, with an increase in the duration of HPV infection and an increase in grade, *Sneathia* and *Megasphaera* were

not among the differential florae, indicating that the growth and reproduction of these florae were strongly influenced by the early stages of infection, which enable them to be indicators of HPV infection. Furthermore, bacteria, such as *Comamonas*, *Gallicola*, and unidentified *Xanthomonadaceae*, contribute the most in cancer progression, as shown in *Figure 3*. Furthermore, the comparison of the differential florae across all five groups, including *Gallicola*, unidentified *Xanthomonadaceae*, *Sneathia*, and *Megasphaera*, are further shown in *Figure S1* to clarify the changes in microbiota at each stage of cervical cancer.

Abundance analysis of differential florae

To further estimate the impact of the abundance of each component (species) on the effect of the florae difference, linear discriminant analysis (LDA) using a non-parametric factorial Kruskal-Wallis rank sum test was conducted. As shown in *Figure 4*, differential florae with significantly different abundance between adjacent groups were determined. Consistent with the results, the abundance of *Sneathia* in the high-risk HPV transient infection with LSIL group and the non-infected HPV group had a significantly different LDA score of approximately $\log_{10}$ 7.3, indicating that the *Sneathia* population greatly decreased after HPV infection. Similarly, the *Pseudomonas*, *Ruminococcus gnavus*, and *Veillonellaceae* bacterium KA00182 populations also decreased sharply after HPV infection (*Figure 4D*). Interestingly, some *Ruminococcaceae* species showed fluctuations in numbers during the increase in the duration of HPV infection and an increase in the grade when comparing *Figure 4B,E*. Moreover, the evolutionary branch diagram of differential florae with different abundance values was based on the abundance analysis of differential florae to elucidate the evolutionary relationship (*Figure S2*).

KEGG pathway analysis of differential florae for biomarker prediction

PICRUSt software was used to predict the functional gene composition of the metagenome based on the OTU data by inferring the gene function profiles of their common ancestors based on the 16S rRNA full-length sequence of the tested bacterial genome. The gene function profiles were derived from the full spectrum of archaeal and bacterial domains in the Greengenes database, and the sequenced bacterial group composition was "mapped" into the database to predict the metabolic function of the
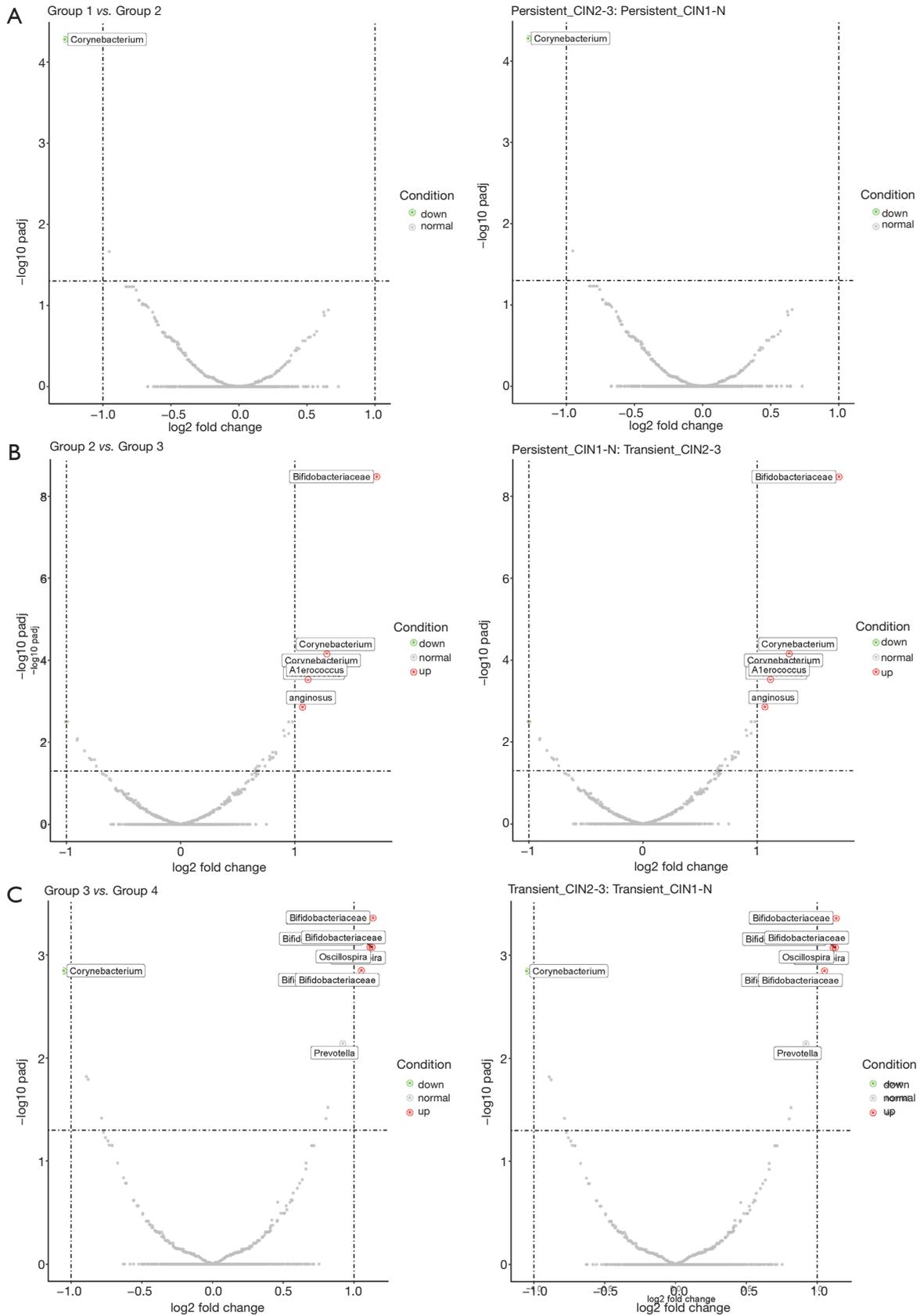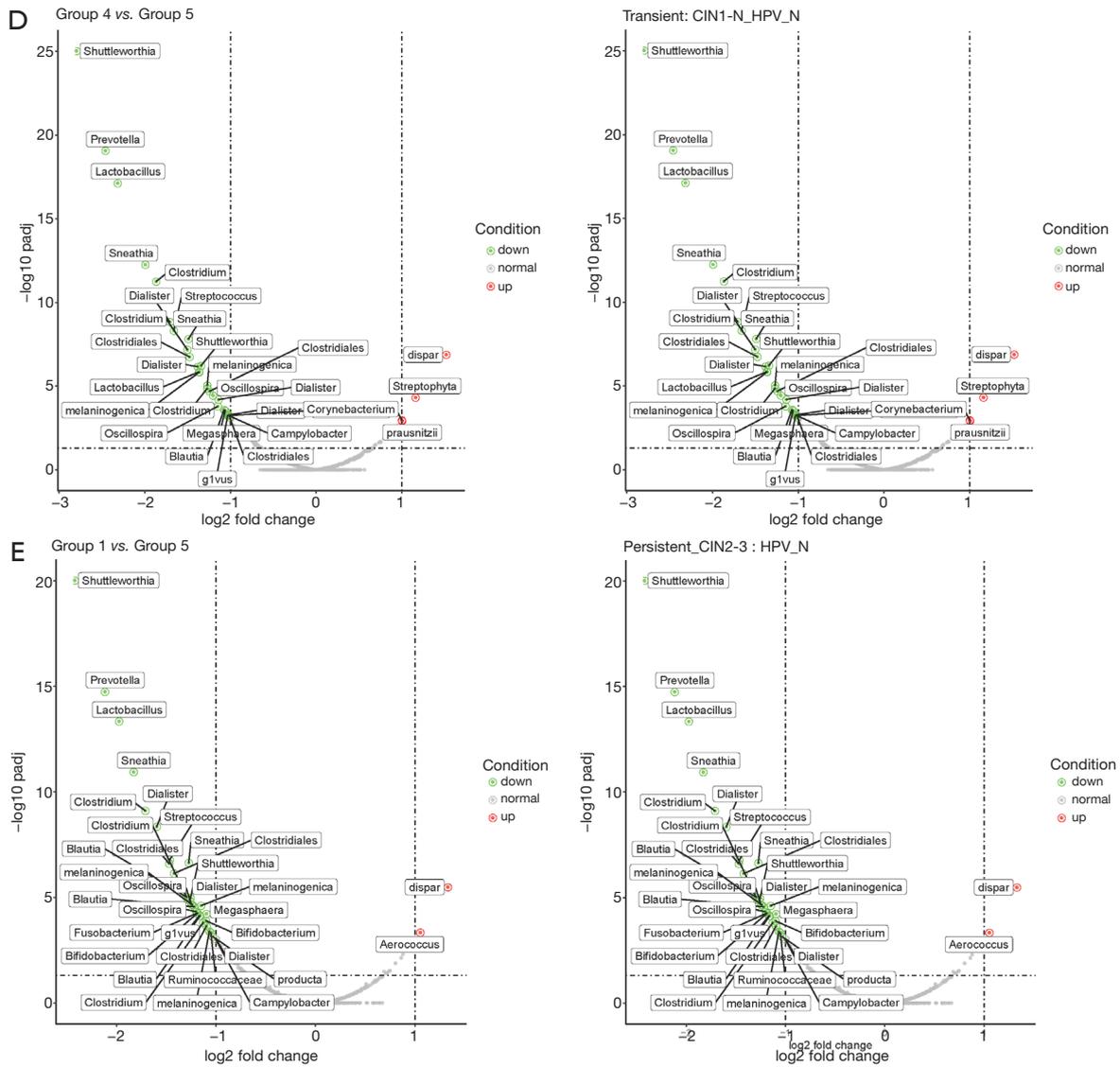
Figure 2 Volcanic map of the microbiota comparison between the different groups of cervical scraping samples. Abscissa is the log2 (fold change) and the ordinate is the padj of each species in the two groups. (A) Comparison of the microbiota between high-risk human papillomavirus (HPV) persistent infection with higher-grade squamous intraepithelial lesion (HSIL) and the high-risk HPV persistent infection with lower-grade squamous intraepithelial lesion (LSIL) group. (B) Comparison of the microbiota between the high-risk HPV persistent infection with LSIL group and high-risk HPV transient infection with HSIL group. (C) Comparison of the microbiota between high-risk HPV transient infection with HSIL group and high-risk HPV transient infection with LSIL group. (D) Comparison of the microbiota between high-risk HPV transient infection with LSIL group and the non-infected HPV group. (E) Comparison of the microbiota between high-risk HPV persistent infection with HSIL and the non-infected HPV group.
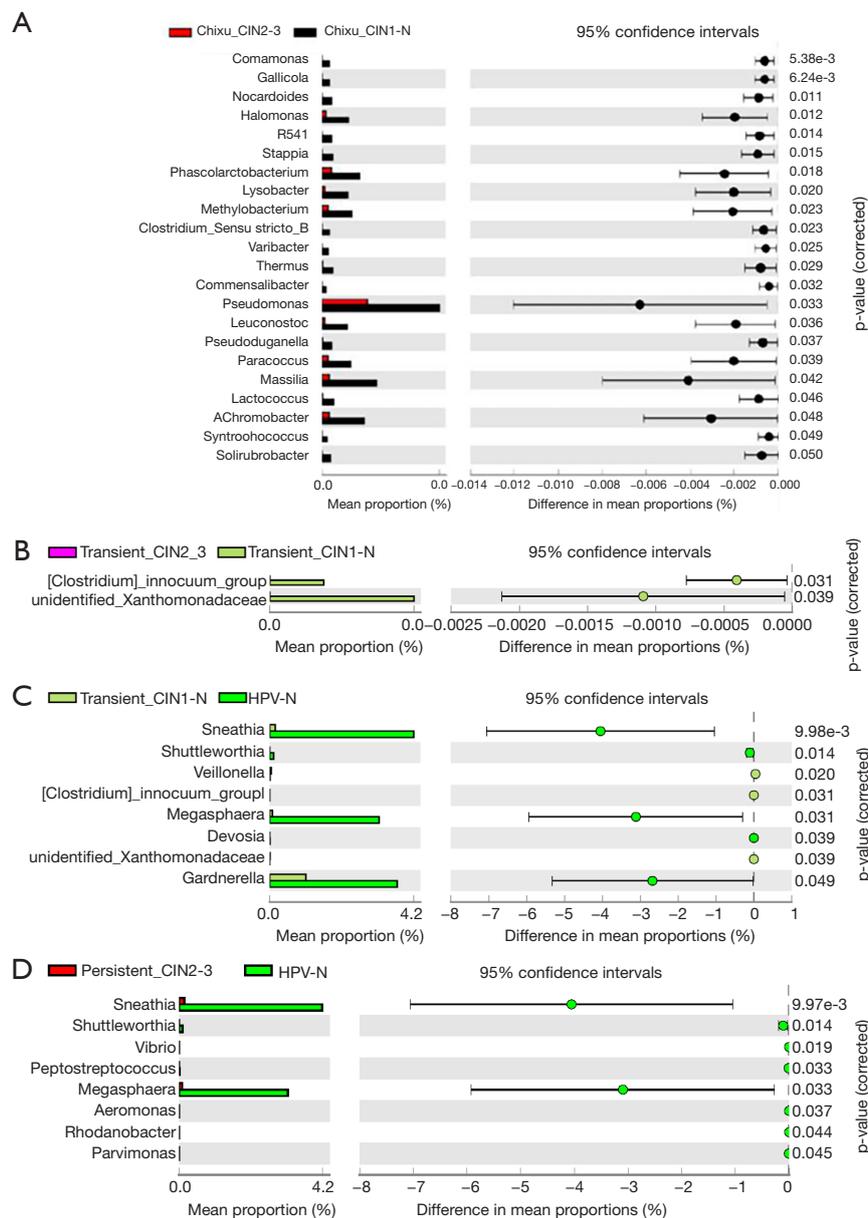
Figure 3 Statistical analysis of differential florae types in the genus level between different groups of cervical scraping samples. Differential florae were filtered according to the difference in mean proportion with 95% confidence intervals. (A) Differential florae types in the comparison of the microbiota between high-risk human papillomavirus (HPV) persistent infection higher-grade squamous intraepithelial lesion (HSIL) and the high-risk HPV persistent infection with lower-grade squamous intraepithelial lesion (LSIL) group. (B) Differential florae types in the comparison of the microbiota between high-risk HPV transient infection with HSIL group and high-risk HPV transient infection with LSIL group. (C) Comparison of the microbiota between high-risk HPV transient infection with LSIL group and the non-infected HPV group. (D) Comparison of the microbiota between high-risk HPV persistent infection with HSIL and the non-infected HPV group.
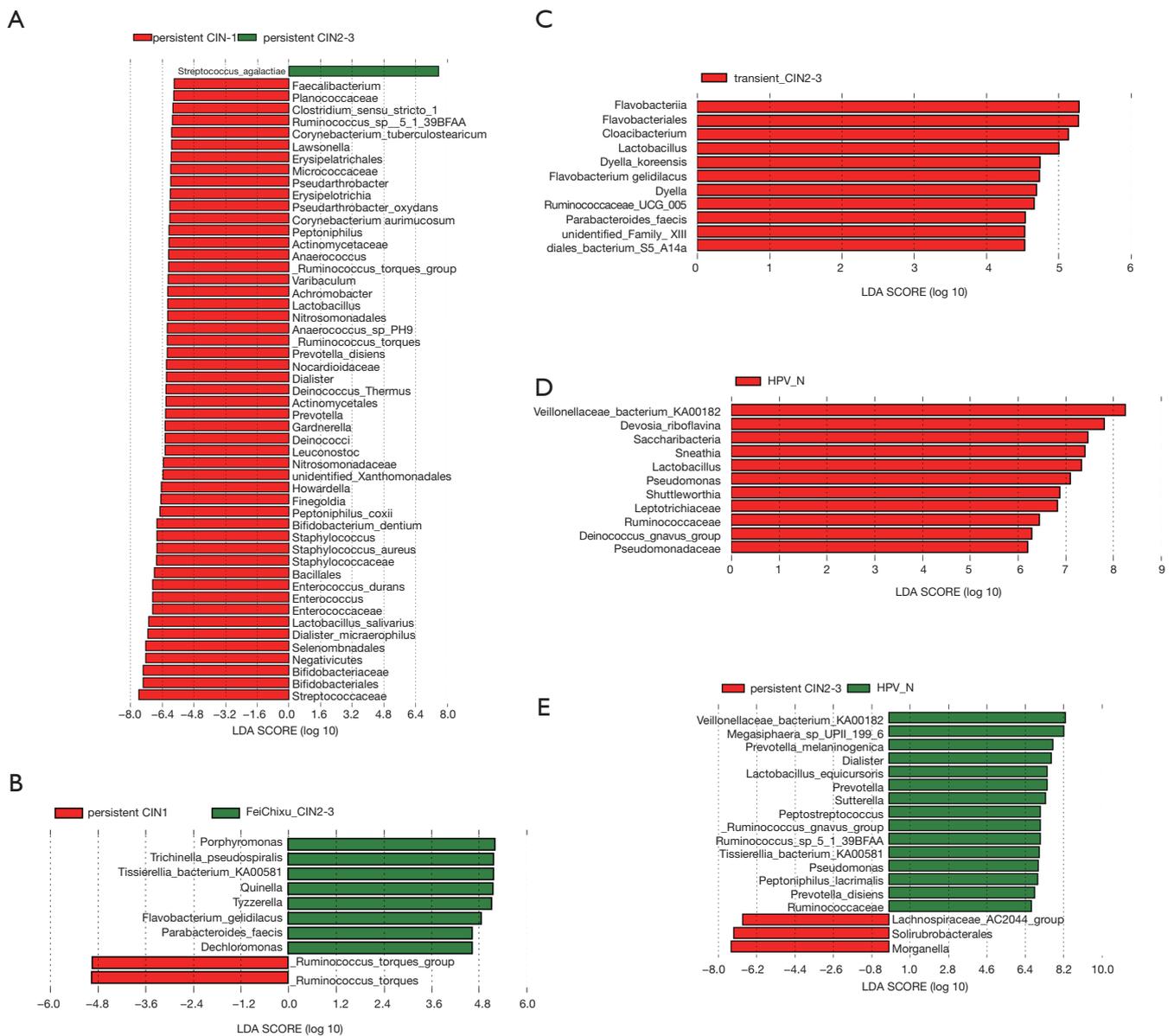
Figure 4 Abundance analysis of differential florae between the different groups of cervical scraping samples. Differential florae with significant differences in abundance were estimated by linear discriminant analysis (LDA) effect size analysis. The length of the histogram represents the magnitude of the impact of significantly different species. (A) Abundance analysis of the microbiota between high-risk human papillomavirus (HPV) persistent infection with higher-grade squamous intraepithelial lesion (HSIL) and the high-risk HPV persistent infection with lower-grade squamous intraepithelial lesion (LSIL) group. (B) Abundance analysis of the microbiota between the high-risk HPV persistent infection with LSIL group and high-risk HPV transient infection with HSIL group. (C) Abundance analysis of the microbiota between high-risk HPV transient infection with HSIL group and high-risk HPV transient infection with LSIL group. (D) Abundance analysis of the microbiota between high-risk HPV transient infection with LSIL group and the non-infected HPV group. (E) Abundance analysis of the microbiota between high-risk HPV persistent infection with HSIL and the non-infected HPV group.

bacterial group (*Table S2*). According to the heatmap of the KEGG pathway analysis, the five groups had gene functions that were obviously different from each other (*Figure 5*). Furthermore, compared with the non-infected HPV group, the main differences in pathways in the HPV infection groups were sporulation, porphyrin and chlorophyll metabolism, arginine and proline metabolism, isoquinoline alkaloid biosynthesis, and ansamycin biosynthesis. These findings indicate that HPV infection directly inhibits these biologic process, which may be the cause of the early symptoms of cervical neoplasia. Interestingly, according to the clustering result, the high-risk HPV transient infection with HSIL group was closer to the non-infected HPV group than the other HPV infection groups; this may be due to the observed fluctuations in the populations of some florae, such as *Ruminococcaceae* species. Additionally, the comparison of differential pathways, including flavonoid, stilbenoid, diarylheptanoid, and gingerol biosynthesis; base excision repair; and arginine and proline metabolism, among the five groups were drafted using box plots, from which the specific differences in these pathways for each group are shown (*Figure 6*).

Furthermore, the random forest machine learning method and 10-fold cross-validation were adopted to predict biomarkers from the OTU abundance data. By selecting HPV persistent infection with the HSIL group and non-infected HPV group as the training group (otu_train) and HPV transient infection with HSIL, HPV persistent infection with LSIL, HPV transient infection with HSIL, and HPV transient infection with LSIL as the testing data (otu_test), 30 potential biomarker combinations were screened using the R Package randomforest 4.6–14 (*Figure 7*). By repeating the 10-fold cross-validation five times, 20 potential biomarkers were finally confirmed according to their mean decrease accuracy, and among all the biomarkers, OTU851726 and OTU715913 were undoubtedly the best potential indicators of cervical cancer, owing to their high mean decrease accuracy and mean decrease Gini coefficient (*Table S3*, *Figure 7*).

Discussion

The vagina is a complex microecological system with a unique anatomical structure, microecological flora, local immunity, and endocrine regulation. Together, these components constitute three lines of defense: the first line is the closed labia and the natural fit of the anterior and posterior vaginal walls, the second line is the "self-cleaning

factory" that consists of more than 200 microbes living in the vagina, and the third line is the mucus plug at the mouth of the cervix that prevents vaginal inflammation from spreading to the uterus (17,18). Because of the physiological structure of the cervix and vagina, imbalance in the vaginal microecosystem directly increases the chance of HPV infection and accelerates the process of cervical precancerous lesions (19).

Cervical cancer associated with HPV infection occurs, it is well known, but the majority of HPV infection is the infection, a small number of HPV in persistent infection state, natural clear HPV infection mechanism is not clear at present, the cause of persistent infection, persistent infection status of prediction and adjustment is cervical cancer early prediction and prevention from earlier stages has more important significance. The status of HPV infection should be related to the local vaginal environment. Some small sample cross-sectional studies showed that hr-HPV infection with VMB dominated by Lactobacillus gallinaceus had the fastest regression of the virus, while low lactobacillus gallinaceus content and high content of Atoporum showed the slowest regression of the virus. In the HPV infection state, THE VMB is mostly in the dominant state of lactobacillus or anaerobic state. The ratio of VMB dominated by anaerobes increased in step with the degree of cervical lesions. The VMB of HSIL women had higher abundance of Cilia, Prussians, digestive Streptococcus, and lower abundance of Jannella than that of LSIL women. Studies have shown that characteristic VMB may be related to HPV infection status, and certain bacterial metabolites may cause certain consequences. It is speculated that the specific vaginal microecology may affect the HPV infection in the cervical epithelium and the outcome after infection through a certain way. But by what mechanism does THE VMB dominate HPV infection and infection status? There are no predictable experimental results. Basic research shows: the VMB advantage for anaerobic bacteria or inert lactobacillus bacteria, lactic acid bacteria are lacking, vaginal epithelial cells producing L-lactic acid/D-the ratio of lactic acid increased, which can lead to the extracellular matrix metalloproteinases inducer expression increasing and metal protease activity-8, the metalloproteinases-8 can damage the integrity of cervical epithelium and promoting the transformation of HPV enters the cervical area, home in basement membrane cells, viral replication and spread, and lead to the happening of the CIN, HPV entered into a state of persistent infection.

Among the microorganisms inhabiting the vagina
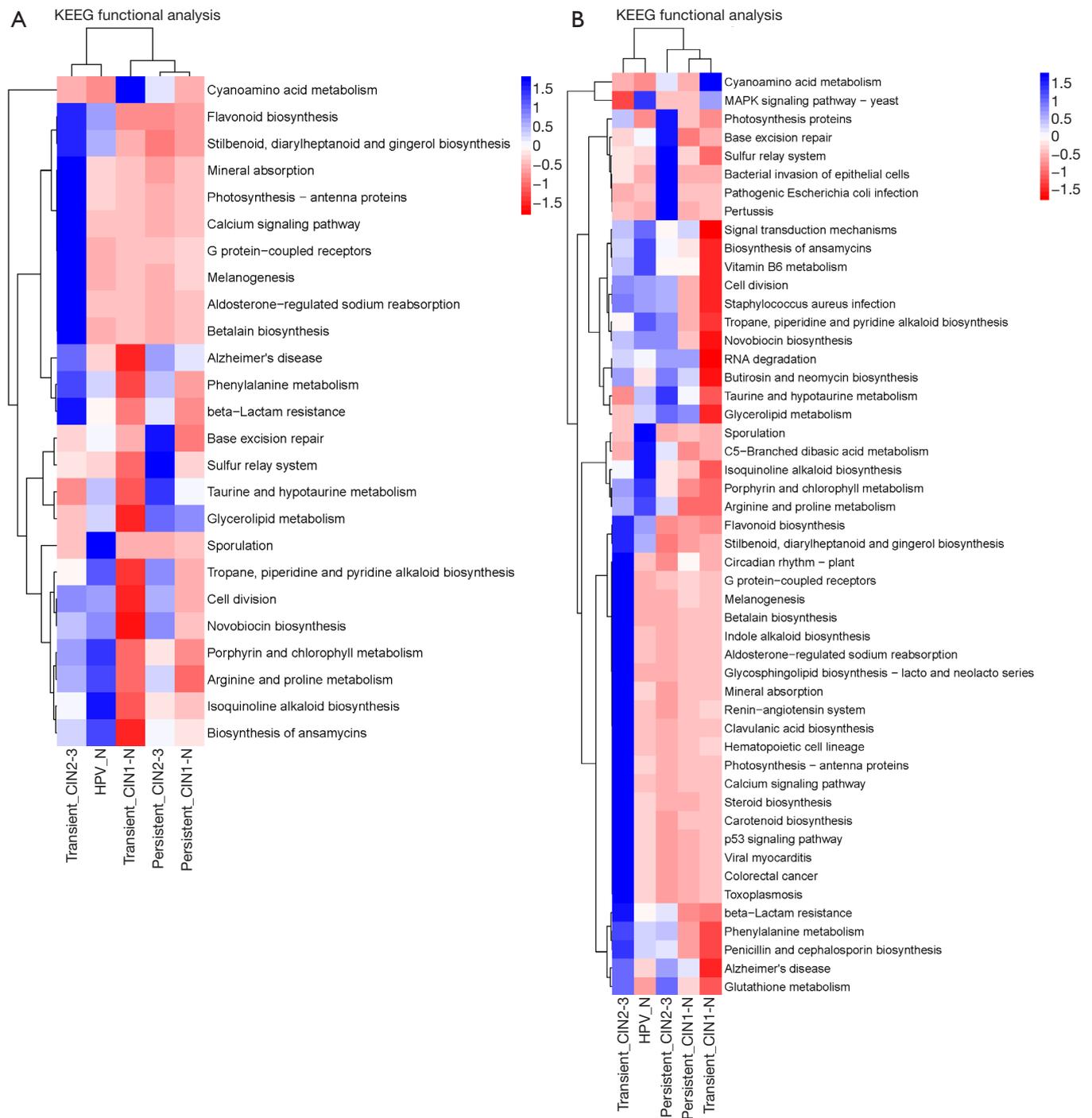
Figure 5 Heatmap of the KEGG enriched pathways among the five groups of cervical scraping samples. PICRUSt software was used to predict the functional gene composition of the metagenome based on operational taxonomic unit (OTU) data. Greengenes 13.5 close reference OTU table was used to predict the KO table and the KEEG level 3 function table. STAMP was used to analyze differences by comparing the KEGG pathway of different group samples.
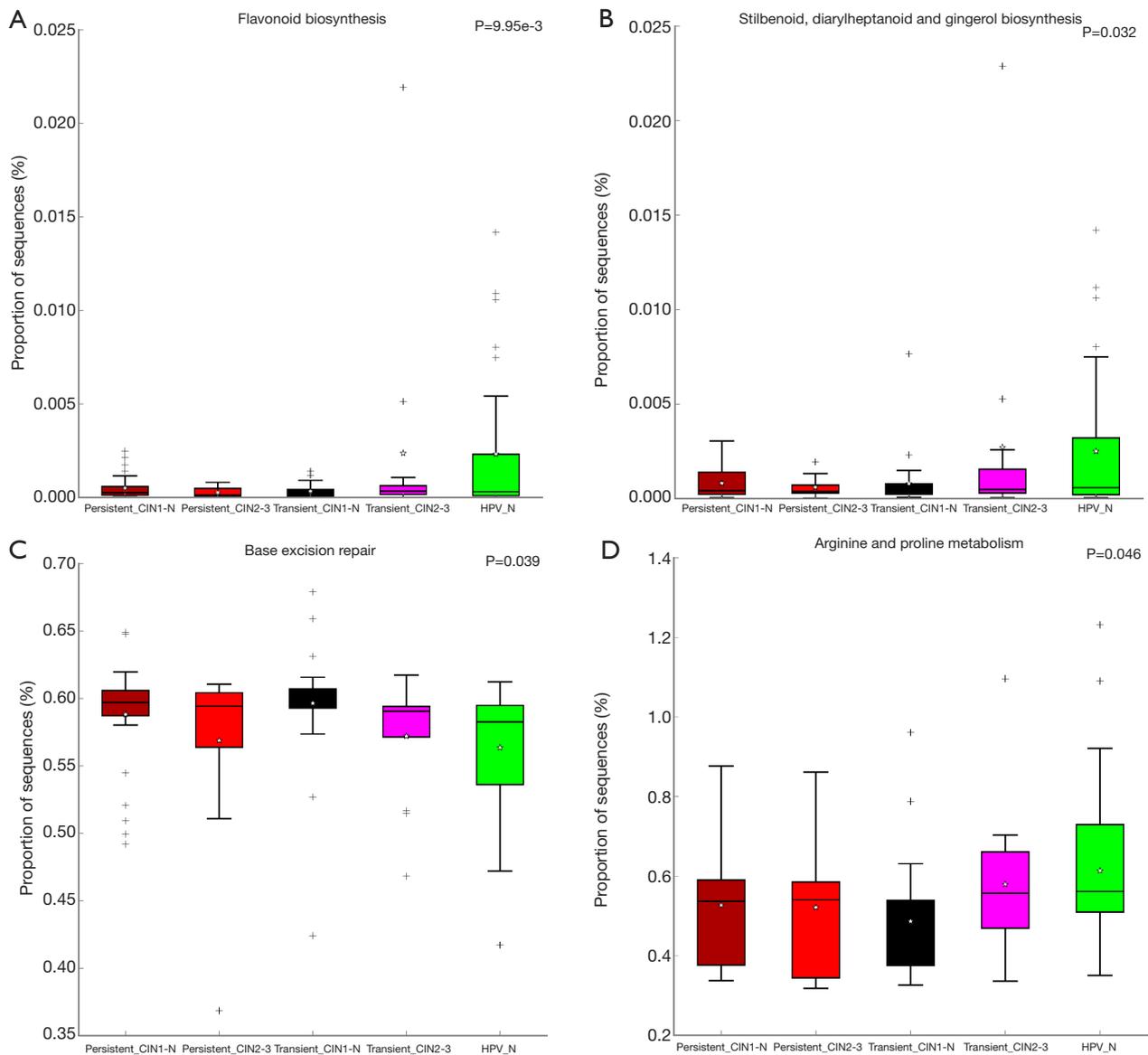
Figure 6 Box plot analysis of the KEGG enriched pathways among the five groups of cervical scraping samples. Based on the gene function profiles predicted with the PICRUSt software, the relative expression of the KEGG pathways in the five groups were compared. (A) Comparison of flavonoid biosynthesis in the five groups. (B) Comparison of stilbenoid, diarylheptanoid, and gingerol biosynthesis in the five groups. (C) Comparison of base excision repair in the five groups. (D) Comparison of arginine and proline metabolism in the five groups. *, infinitesimal deviation; +, mild diverge.

of women of childbearing age, approximately 80% are *Lactobacillus*, which maintain a pH between 3.8 and 4.5 (20). Previous studies have demonstrated that a decreased level of cervicovaginal lactobacilli is associated with increased detection of HPV infection and cervical cancer development (21). In the present study, the *Lactobacillus* population decreased in the comparison between the non-

infected HPV group and HPV infection groups, which is consistent with the majority of previous reports. However, Brusselaers *et al.* argued that dominant florae would be the opposite if different technology was used for detection. In their microscopic analyses, lactobacilli-dominated microbiota was correlated with an increased detection of HPV infection and persistence, whereas non-*Lactobacillus*

Top 30-Variable Biomark
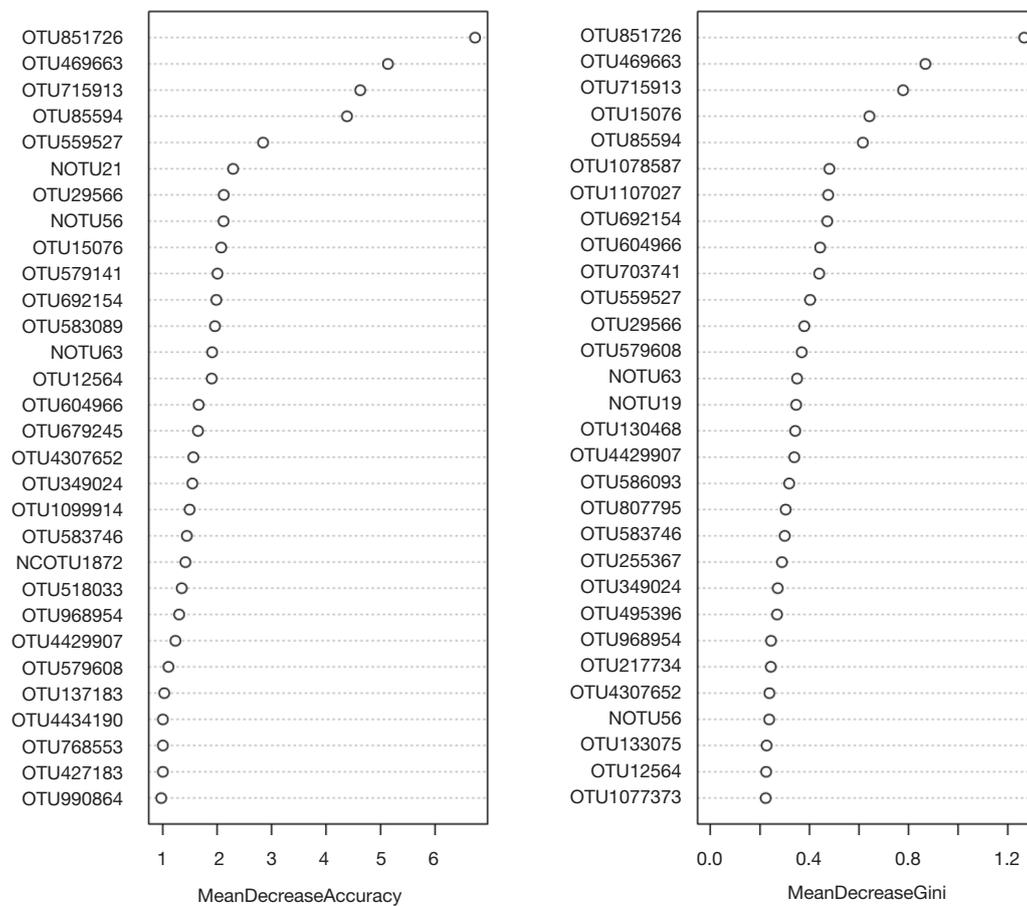
Figure 7 Thirty potential biomarker combinations screened based on operational taxonomic unit (OTU) abundance data. The mean decrease accuracy and the mean decrease Gini coefficient are two important indicators in the random forest model. Mean decrease accuracy indicates the degree of decrease in the accuracy of the random forest prediction; a larger value indicates a greater importance of the variable. Mean decrease Gini coefficient calculates the difference between each variable and the observed value on each node of the classification tree. Qualitative effects, used to compare the importance of variables, indicate that, the greater the value, the greater the importance of the variable.

crispatus-dominated vaginal microbiota, associated with an increased detection of HPV infection and persistence, was observed (22). Similarly, the *Sneathia* population decreased in the HPV infection groups in the present study, which was in contrast to the results of previous studies that identified it as a risk factor (23-25). This difference may be due to the difference in the detection and analysis methods used. However, *Bifidobacteriaceae* populations were also observed to increase sharply when the duration of HPV infection increased, which is consistent with the findings of a previous study (26).

In addition to determining changes in the taxonomic

composition in the development of cervical cancer, the functional composition of the microbiome was further assessed using KEGG pathway enrichment analysis (27). Of the influenced pathways following HPV infection, porphyrin and chlorophyll metabolism is one of the main sources of heme (28); this may partly explain why some patients with cervical cancer have low hemoglobin levels (29). Additionally, in the important nitrogen cycle pathway, arginine and proline metabolism plays a key role in providing organic material for bacterial microorganisms (30). As reported previously, the dominant vaginal florae in healthy women has the ability to improve neutrophil functionality

and the innate immune system as a whole. This may be due to the enhanced pathways observed in non-HPV infected patients in the present study, such as flavonoid, stilbenoid, diarylheptanoid, gingerol biosynthesis, and arginine and proline metabolism (31-33). In addition to these pathways, several studies have identified other signaling pathways that are associated with cervical cancer progression. Kwon *et al.* suggested that cervical metagenomes could be enriched during the course of peptidoglycan synthesis and depleted by dioxin degradation and 4-oxalocrotonate tautomerase due to the essential role of peptidoglycan synthesis in maintaining cell integrity against osmotic pressure for cell survival (34).

Understanding of the importance of vaginal microecology is increasing (35,36), and the combination of microbiologic basic research and clinical gynecological and obstetric infectious disease research seeks to establish a more comprehensive and standardized vaginal microecological clinical evaluation system. However, the relationship between the microenvironment of the vaginal flora, HPV infection, and the occurrence and development of cervical precancerous lesions, as well as the associated mechanisms, remain unclear. Therefore, in the present study, we conducted in-depth research on the relationship between the vaginal flora microenvironment and the occurrence and development of HPV infection and cervical precancerous lesions. Moreover, new reference indicators for judging the treatment, prognosis, and HPV infection and outcome of CIN and cancer have been proposed to assist with the elucidation of the mechanisms of cancerous changes.

Conclusions

The present study provides a comprehensive profile of the cervical microbiome and its relationship with the duration of HPV infection and CIN pathological classification. The findings of the present study could assist with the development of a guideline for screening new clinical drugs for cervical cancer.

Acknowledgments

Footnote

Reporting Checklist: The authors have completed the MDAR reporting checklist. Available at http://dx.doi.org/10.21037/atm-20-5832

Data Sharing Statement: Available at http://dx.doi.org/10.21037/atm-20-5832

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/atm-20-5832). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The clinical study was approved by the Ethic Committee of Beijing Chao-Yang Hospital (No. 2019-331) and was conducted in accordance with the provisions of the Declaration of Helsinki (as revised in 2013). Written informed consent was obtained from all participants before enrolment.

References

1. Wu Y, Zhu B, Han J, et al. A comparative dosimetric study of cervical cancer patients with para-aortic lymph node metastasis treated with volumetric modulated arc therapy vs. 9-field intensity-modulated radiation therapy. Ann Transl Med 2019;7:675.

2. Kero K, Rautava J, Syrjanen K, et al. Association of asymptomatic bacterial vaginosis with persistence of female genital human papillomavirus infection. Eur J Clin Microbiol Infect Dis 2017;36:2215-9.

3. Leach CR, Schoenberg NE. The vicious cycle of inadequate early detection: a complementary study on barriers to cervical cancer screening among middle-aged and older women. Prev Chronic Dis 2007;4:A95.

4. Van Doorslaer K, Chen Z, Bernard HU, et al. ICTV Virus Taxonomy Profile: Papillomaviridae. J Gen Virol

2018;99:989-90.

5. Leeman A, Jenkins D, Del Pino M, et al. Expression of p16 and HPV E4 on biopsy samples and methylation of FAM19A4 and miR124-2 on cervical cytology samples in the classification of cervical squamous intraepithelial lesions. Cancer Med 2020;9:2454-61.

6. Brendle SA, Bywaters SM, Christensen ND. Pathogenesis of infection by human papillomavirus. Curr Probl Dermatol 2014;45:47-57.

7. Layman H, Rickert KW, Wilson S, et al. Development and validation of a multiplex immunoassay for the simultaneous quantification of type-specific IgG antibodies to E6/E7 oncoproteins of HPV16 and HPV18. PLoS One 2020;15:e0229672.

8. Yang J, Wang W, Wang Z, et al. Prevalence, genotype distribution and risk factors of cervical HPV infection in Yangqu, China: a population-based survey of 10086 women. Hum Vaccin Immunother 2020;16:1645-52.

9. Cherpes TL, Marrazzo JM, Cosentino LA, et al. Hormonal contraceptive use modulates the local inflammatory response to bacterial vaginosis. Sex Transm Infect 2008;84:57-61.

10. Knupp de Souza DM, Diniz CG, Filho DS, et al. Antimicrobial susceptibility and vaginolysin in Gardnerella vaginalis from healthy and bacterial vaginosis diagnosed women. J Infect Dev Ctries 2016;10:913-9.

11. Joag V, Obila O, Gajer P, et al. Impact of Standard Bacterial Vaginosis Treatment on the Genital Microbiota, Immune Milieu, and Ex Vivo Human Immunodeficiency Virus Susceptibility. Clin Infect Dis 2019;68:1675-83.

12. Mandras N, Tullio V, Furneri PM, et al. Key Roles of Human Polymorphonuclear Cells and Ciprofloxacin in Lactobacillus Species Infection Control. Antimicrob Agents Chemother 2015;60:1638-41.

13. Branch JD. Effect of creatine supplementation on body composition and performance: a meta-analysis. Int J Sport Nutr Exerc Metab 2003;13:198-226.

14. Hu ZQ, Zhou W, Yue XA, et al. Dynamical evaluation of vaginal micro-ecosystem in a Chinese woman with recurrent vulvovaginal candidiasis. Genet Mol Res 2015;14:2626-30.

15. Berger RE. Rates of Torsades de Pointes Associated With Ciprofloxacin, Ofloxacin, Levofloxacin, Gatifloxacin, and Moxifloxacin. J Urology 2005;174:165.

16. Feng RM, Hu SY, Zhao FH, et al. Role of active and passive smoking in high-risk human papillomavirus infection and cervical intraepithelial neoplasia grade 2 or worse. J Gynecol Oncol 2017;28:e47.

17. Fichorova RN, Anderson DJ. Differential expression of immunobiological mediators by immortalized human cervical and vaginal epithelial cells. Biol Reprod 1999;60:508-14.

18. Ostrzenski A. The G-Spot Anatomical Structure Spotted. J Sex Med 2018;15:803.

19. Kovachev SM. Cervical cancer and vaginal microbiota changes. Arch Microbiol 2020;202:323-7.

20. Amin M, Moradi Choghakabodi P, et al. In vitro antimicrobial activities of metabolites from vaginal Lactobacillus strains against Clostridium perfringens isolated from a woman's vagina. J Chin Med Assoc 2017;80:29-33.

21. Wang H, Ma Y, Li R, et al. Associations of Cervicovaginal Lactobacilli With High-Risk Human Papillomavirus Infection, Cervical Intraepithelial Neoplasia, and Cancer: A Systematic Review and Meta-Analysis. J Infect Dis 2019;220:1243-54.

22. Brusselaers N, Shrestha S, van de Wijgert J, et al. Vaginal dysbiosis and the risk of human papillomavirus and cervical cancer: systematic review and meta-analysis. Am J Obstet Gynecol 2019;221:9-18.e8.

23. Gondwe T, Ness R, Totten PA, et al. Novel bacterial vaginosis-associated organisms mediate the relationship between vaginal douching and pelvic inflammatory disease. Sex Transm Infect 2020;96:439-44.

24. Stout M, LaRosa P, Shannon W, et al. 392: The vaginal microbiome of preterm birth. Am J Obstet Gynecol 2014;210:S199.

25. Zhou Y, Wang L, Pei F, et al. Patients With LR-HPV Infection Have a Distinct Vaginal Microbiota in Comparison With Healthy Controls. Front Cell Infect Microbiol 2019;9:294.

26. Lugli GA, Milani C, Turroni F, et al. Comparative genomic and phylogenomic analyses of the Bifidobacteriaceae family. BMC Genomics 2017;18:568.

27. Peng W, Yang J, Xu P, et al. Association of gut microbiota composition and function with a senescence-accelerated mouse model of Alzheimer's Disease using 16S rRNA gene and metagenomic sequencing analysis. Aging 2018;10:4054-65.

28. Bonkovsky HL, Guo JT, Hou W, et al. Porphyrin and heme metabolism and the porphyrias. Compr Physiol 2013;3:365-401.

29. Bishop AJ, Allen PK, Klopp AH, et al. Relationship between low hemoglobin levels and outcomes after treatment with radiation or chemoradiation in patients with cervical cancer: has the impact of anemia been

overstated? Int J Radiat Oncol Biol Phys 2015;91:196-205.

30. Rath M, Muller I, Kropf P, et al. Metabolism via Arginase or Nitric Oxide Synthase: Two Competing Arginine Pathways in Macrophages. Front Immunol 2014;5:532.

31. Petersen B, Busch CJ, Schleifer G, et al. Arginase impairs hypoxic pulmonary vasoconstriction in murine endotoxemia. Respir Res 2019;20:109.

32. Petersen KS, Marnewick JL, Smith C. Grape polyphenols corrects ageing-related detriments in neutrophil functionality via modulation of specific molecular targets. Inflammopharmacology 2018;26:1349-58.

33. Smith NC, Christian SL, Taylor RG, et al. Immune modulatory properties of 6-gingerol and resveratrol in Atlantic salmon macrophages. Mol Immunol 2018;95:10-9.

34. Kwon M, Seo SS, Kim MK, Lee DO, Lim MC. Compositional and Functional Differences between Microbiota and Cervical Carcinogenesis as Identified by Shotgun Metagenomic Sequencing. Cancers (Basel) 2019;11:309.

35. Liang Y, Chen M, Qin L, et al. A meta-analysis of the relationship between vaginal microecology, human papillomavirus infection and cervical intraepithelial neoplasia. Infect Agent Cancer 2019;14:29.

36. Zheng JJ, Song JH, Yu CX, et al. Difference in vaginal microecology, local immunity and HPV infection among childbearing-age women with different degrees of cervical lesions in Inner Mongolia. BMC Womens Health 2019;19:109.

Table S1 Differential florae types in genus level among the five groups

Taxonomy	P value	P value (corrected)	Effect size
Sneathia	8.37E-04	8.37E-04	0.147736608
Alloscardovia	1.51E-03	1.51E-03	0.138179981
Shuttleworthia	1.75E-03	1.75E-03	0.13585771
Gallicola	1.98E-03	1.98E-03	0.133821274
Unidentified Anaerolineaceae	4.84E-03	4.84E-03	0.11888726
Megasphaera	7.46E-03	7.46E-03	0.111497558
Unidentified Xanthomonadaceae	9.15E-03	9.15E-03	0.1079653
Peptoclostridium	4.17E-02	4.17E-02	0.080698959
Alloprevotella	4.23E-02	4.23E-02	0.080431788
Propionimicrobium	4.23E-02	4.23E-02	0.080412983
Leptotrichia	4.39E-02	4.39E-02	0.079710818
Buchnera	4.52E-02	4.52E-02	0.079176561
Flavobacterium	4.56E-02	4.56E-02	0.078984607
Subdoligranulum	4.90E-02	4.90E-02	0.077643485
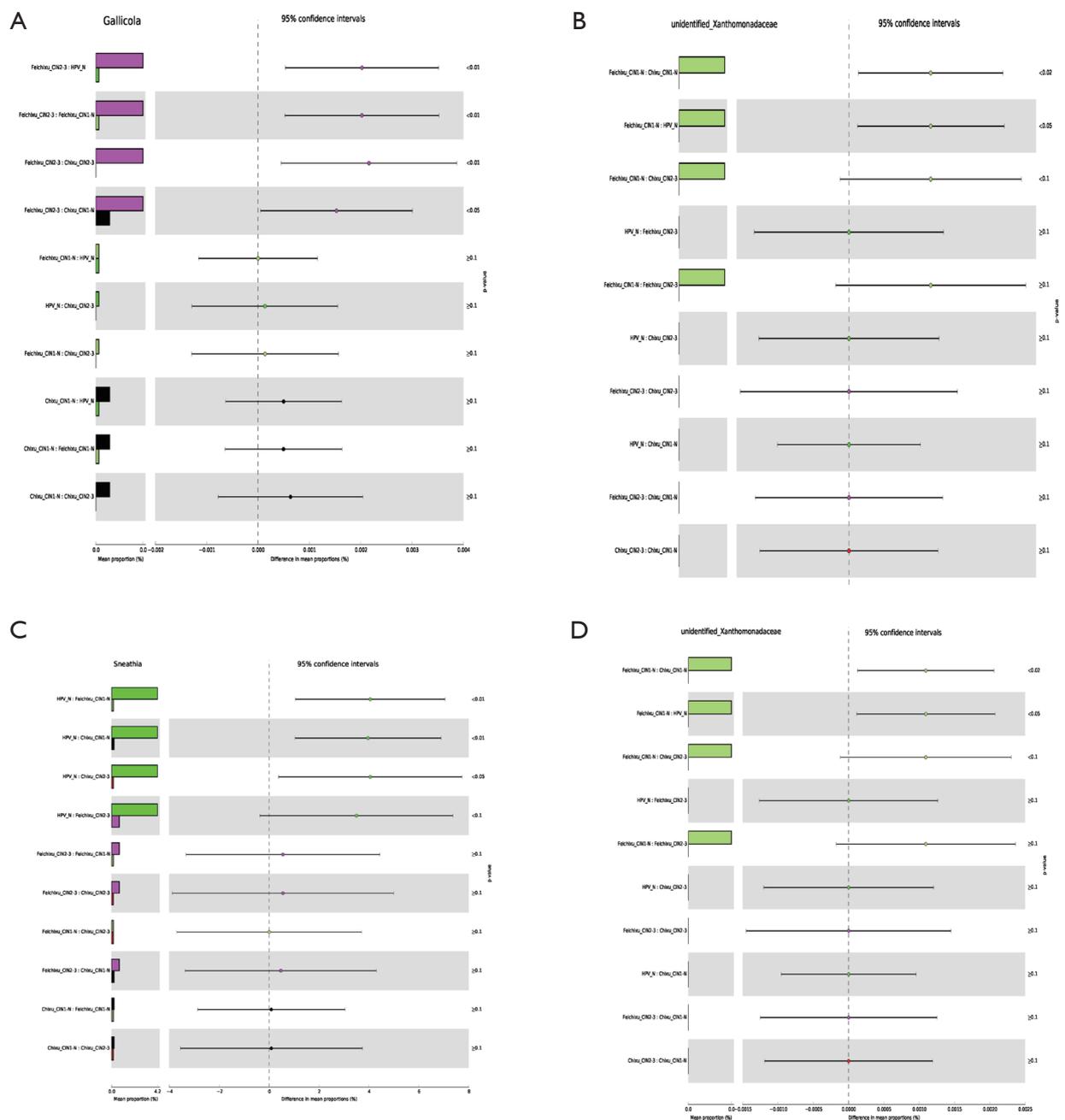
Figure S1 Comparison of the differential florae in the five groups using *Gallicola*, unidentified *Xanthomonadaceae*, *Sneathia*, and *Megasphaera* as examples. Differential florae were filtered according to the differences in mean proportions with 95% confidence intervals. Ten comparisons for differential florae were made [non-infected human papillomavirus (HPV) group *vs.* transient infection with lower-grade squamous intraepithelial lesion (LSIL), non-infected HPV group *vs.* persistent infection with LSIL, non-infected HPV group *vs.* persistent infection with higher-grade squamous intraepithelial lesion (HSIL), non-infected HPV group *vs.* transient infection with HSIL, transient infection with HSIL *vs.* transient infection with LSIL, transient infection with HSIL *vs.* persistent infection with HSIL, transient infection with LSIL *vs.* persistent infection with HSIL, transient infection with HSIL *vs.* persistent infection with LSIL, persistent infection with LSIL *vs.* transient infection with LSIL, and persistent infection with LSIL *vs.* persistent infection with HSIL]. (A) *Gallicola* expression in the 10 comparisons. (B) Unidentified *Xanthomonadaceae* expression in the 10 comparisons. (C) *Sneathia* expression in the 10 comparisons. (D) *Megasphaera* expression in the 10 comparisons.

A

Chixu_CIN1-N

Deinococcus_Thermus
Deinococci

- a: Varibaculum
- b: Actinomycetaceae
- c: Actinomycetales
- d: Gardnerella
- e: Bifidobacteriaceae
- f: Bifidobacteriales
- g: Lawsonella
- h: Pseudarthrobacter
- i: Micrococcaceae
- j: Nocardioidaceae
- k: Prevotella
- l: Planococcaceae
- m: Staphylococcus
- n: Staphylococcaceae
- o: Bacillales
- p: Enterococcus
- q: Enterococcaceae
- r: Lactobacillus
- s: Leuconostoc
- t: Streptococcaceae
- u: Clostridium_sensu_stricto_
- v: Anaerococcus
- w: Finegoldia
- x: Peptoniphilus
- y: Howardella
- z: _Ruminococcus__torques_
- a0: Faecalibacterium
- a1: Erysipelotrichales
- a2: Dialister
- a3: Selenomonadales
- a4: Achromobacter
- a5: Nitrosomonadaceae
- a6: Nitrosomonadales
- a7: unidentified_Xanthomon

B

Chixu_CIN1-N
Feichixu_CIN2-3

- a: Porphyromonas
- b: Tyzzerella
- c: _Ruminococcus__torques_
- d: Quinella
- e: Dechloromonas

C

Feichixu_CIN2-3

Flavobacteriia

- a: Cloacibacterium
- b: Flavobacteriales
- c: Lactobacillus
- d: unidentified_Family_XIII
- e: Ruminococcaceae_UCG_0(
- f: Dyella

D

HPV_N

Spachaibacteria

- a: Lactobacillus
- b: Shuttleworthia
- c: _Ruminococcus__gnavus_
- d: Ruminococcaceae
- e: Sneathia
- f: Leptotrichiaceae
- g: Pseudomonas
- h: Pseudomonadaceae

E

Chixu_CIN2-3
HPV_N

- a: Solirubrobacterales
- b: Prevotella
- c: Lachnospiraceae_AC2044_
- d: _Ruminococcus__gnavus_
- e: Peptostreptococcus
- f: Ruminococcaceae
- g: Dialister
- h: Sutterella
- i: Morganella
- j: Pseudomonas

Figure S2 Evolutionary branch diagram of differential florae with different abundance between groups. Circle radiating from the inside to the outside represents the classification level from the door to the genus (or species). Each small circle at a different classification level represents a classification at that level, and the diameter of the small circle is proportional to the relative abundance. (A) Evolutionary branch diagram of the microbiota with different abundance between high-risk human papillomavirus (HPV) persistent infection with higher-grade squamous intraepithelial lesion (HSIL) and the high-risk HPV persistent infection with lower-grade squamous intraepithelial lesion (LSIL) group. (B) Evolutionary branch diagram of the microbiota with different abundance between the high-risk HPV persistent infection with LSIL group and high-risk HPV transient infection with HSIL group. (C) Evolutionary branch diagram of the microbiota with different abundance between high-risk HPV transient infection with HSIL group and high-risk HPV transient infection with LSIL group. (D) Evolutionary branch diagram of the microbiota with different abundance between high-risk HPV transient infection with LSIL group and the non-infected HPV group. (E) Evolutionary branch diagram of the microbiota with different abundance between high-risk HPV persistent infection with HSIL and the non-infected HPV group.

Table S2 Table of differential KEGG pathways

OUT ID	P value	P value (corrected)	Effect size
Porphyrin and chlorophyll metabolism	6.89E-03	6.89E-03	1.13E-01
Flavonoid biosynthesis	9.95E-03	9.95E-03	1.07E-01
Stilbenoid, diarylheptanoid, and gingerol biosynthesis	3.20E-02	3.20E-02	8.56E-02
Base excision repair	3.90E-02	3.90E-02	8.19E-02
Arginine and proline metabolism	4.62E-02	4.62E-02	7.87E-02
Isoquinoline alkaloid biosynthesis	5.14E-02	5.14E-02	7.67E-02
Alzheimer's disease	5.63E-02	5.63E-02	7.50E-02
Sulfur relay system	5.63E-02	5.63E-02	7.50E-02
Cyanoamino acid metabolism	6.69E-02	6.69E-02	7.17E-02
Biosynthesis of ansamycins	6.71E-02	6.71E-02	7.16E-02
Taurine and hypotaurine metabolism	7.59E-02	7.59E-02	6.93E-02
Photosynthesis—antenna proteins	7.81E-02	7.81E-02	6.87E-02
Glycerolipid metabolism	7.94E-02	7.94E-02	6.84E-02
Calcium signaling pathway	7.96E-02	7.96E-02	6.83E-02

OUT, operational taxonomic unit.

Table S3 Top 20 potential biomarkers screened by repeating the 10-fold cross-validation five times

OUT ID
OTU851726
OTU715913
OTU469663
OTU85594
OTU586093
OTU495396
OTU217734
OTU703741
OTU15076
OTU1111294
OTU692154
OTU1083194
OTU604966
OTU583089
OTU4429907
NOTU63
OTU484304
OTU4307652
OTU1078587
OTU956654

OUT, operational taxonomic unit.