Full Title: Inflammatory cytokines and a diverse cervicovaginal microbiome are associated with cervical disease in Hispanics living in Puerto Rico

Short Title: bacterial diversity and cervical cancer in Puerto Rico

Eduardo Tosado-Rodríguez¹, Loyda B. Mendez², Ana M. Espino¹, Stephanie Dorta-Estremera¹,³, Edna E. Aquino¹, Josefina Romaguera⁴ and Filipa Godoy-Vitorino¹*

¹ Department of Microbiology and Medical Zoology, School of Medicine, Medical Sciences Campus, University of Puerto Rico, USA
² University Ana G. Méndez, Carolina Campus, Carolina, Puerto Rico, USA
³ Cancer Biology, Comprehensive Cancer Center University of Puerto Rico, USA
⁴ Department of OBGYN, School of Medicine, Medical Sciences Campus, University of Puerto Rico, USA

* Corresponding author

Filipa Godoy-Vitorino, Ph.D.
filipa.godoy@upr.edu

Keywords: Cervicovaginal microbiota, Cytokines, 16S rRNA, Hispanics, inflammation

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract

Cervical cancer (CC) is women's fourth most common cancer worldwide. A worrying increase in CC rates in Hispanics suggests that besides Human Papillomaviruses infections, there may be other biological causes promoting the disease. We hypothesized that the cervical microbiome and the epithelial microenvironment favoring inflammation is conducive to disease progression. There is a lack of studies examining the part played by microbial populations in the development of precancerous lesions into cancer in Hispanic women. We aimed to investigate the relationship between the cervicovaginal microbiome and inflammation in Hispanic women living in PR while considering cervical neoplasia and HPV infection. Cervical samples collected from 91 participants coming to clinics in San Juan, underwent 16S rRNA genes (V4 region) profiling, and cytokines were measured using Luminex MAGPIX technology. Cytokines were grouped as inflammatory (IL-1β, TNFα, IFNγ, IL-6), anti-inflammatory (IL-4, IL-10, TGFβ1), and traffic-associated (IL-8, MIP1a, MCP1, IP10). They were related to microbes via an inflammation scoring index based on the quartile and tercile distribution of the cytokine's concentration. We found significant differences in the diversity and composition of the microbiota according to HPV risk, cervical disease, and cytokine abundance. The most dominant community state type (CST) was CST IV with ~90% dominance in participants with high-grade squamous intraepithelial lesions and high-risk HPV. The increasing concentration of pro-inflammatory cytokines was associated with a decrease in L. crispatus. In contrast, dysbiosis-associated bacteria such as Gardnerella, Prevotella, Atopobium concomitantly increased with pro-inflammatory cytokines. Our study highlights that the cervical microbiota of Hispanics living in PR is characteristically diverse, regardless of HPV status and that dysbiosis associated with a decrease in Lactobacillus marks inflammatory
Therefore, the joint host-microbe interaction analyses via cytokine signaling and microbiota in precancerous lesions is confirmed to have great translational potential.

**Introduction**

Cervical cancer (CC) is the 4th most common cancer among women and the 4th leading cause of cancer death in females worldwide (Cohen et al. 2019). Although CC rates have decreased in the USA (Chatterjee et al. 2016), the incidence and mortality from cervical cancer mainly occurs in low resource settings (~>85%) (Ferlay et al. 2010). According to the Pan American Health Organization for 2012, 80% of American deaths due to CC occur in Latin America and the Caribbean (1). Puerto Rico (PR) has the highest age-adjusted incidence of CC in the US (Ortiz et al. 2010), with notably higher incidence from 2001 to 2017, increasing from 9.2 to 13 per 100,000 person-years (Ortiz et al. 2021). PR also has a 79.3% CC screening rate that does not achieve CDC’s screening recommendations of 93% (Centers for Disease and Prevention 2012; Ortiz et al. 2010) posing a strain on health systems in the island (2).

Persistent infection with oncogenic Human Papillomavirus (HPV) is the most important cause of CC. HPV infections range from 90–100% prevalence in cervical abnormalities, from high-grade squamous intraepithelial lesion -HGSIL- to severe dysplasia and cancer (3), where the viral genome integration modulates the host cell cycle promoting neoplasia progression (4,5). However, other factors influence cervical disease progression, including lifestyle, genetic predisposition, immunosuppression, early onset of sexual activities, and many pregnancies and deliveries (6–8). Nevertheless, why only a fraction of women acquires a long-lasting, persistent high-risk HPV (HR-HPV) that puts them at risk for developing Cervical Intraepithelial Neoplasia (CIN) remains
unknown. One potential explanation for this variation could be the cervicovaginal microbiota and its related processes. Recent studies suggest an association between certain bacterial community types of the vaginal microbiota, HPV infection, and cervical disease (9–11). Additionally, processes like bacterial vaginosis, cervical inflammation, and increased vaginal pH are all known to affect susceptibility to cervical HPV (12).

Women of different ethnic groups differ in their vaginal microbiome signatures (13). The cervicovaginal microbiome has been characterized into five different community state types (CST) according to the dominance of *Lactobacillus* species, a concept that was introduced in 2011 by Ravel et al. to categorize vaginal microbial communities among reproductive-age women (13). CST-I refers to communities dominated by *L. crispatus* (CST-I), CST-II by *L. gasseri*, CST-III by *L. iners*, CST-V by *L. jensenii* (CST-V), or non-*Lactobacillus* dominated and highly diverse profiles such as CST-IV. Furthermore, CST-IV is often characterized by three sub-CSTs: 1) CST IV-A has a high relative abundance of Candidatus *Lachnocurva vaginae* and a moderate relative abundance of *G. vaginalis* and *Atopobium vaginae*, 2) CST IV-B also has a moderate relative abundance of *G. vaginalis* and *Atopobium vaginae*, while 3) CST IV-C has low relative abundance of *Lactobacillus* spp., as well as low levels of *G. vaginalis*, *A. vaginae*, and *Ca. L. vaginae* with high facultative and strictly anaerobic bacteria (14). While *Lactobacillus* dominates the cervicovaginal microbiome of Caucasians, Hispanic and black women have higher diversity and reduced *Lactobacillus* populations (13). A previous study demonstrated that Hispanics living in PR have a different microbial profile than Caucasians, and that patients with high-risk HPV and CIN3 cervical lesions had dominant *Atopobium vaginae* and *Gardnerella vaginalis* (11,15). However, these results could have been improved, as they identified microbiota using cervical swabs, failing to address the host response. Recent studies have highlighted the benefit of
collecting cervical lavages to understand the extent of the dynamic host-microbe symbioses via proteomic, metabolomic, or cytokine profiling analyses (16–19).

We hypothesized that there would be changes in the cervical microenvironment with cervical disease and HPV risk and that markers of inflammation would be associated with a decrease in *Lactobacillus* and overall dysbiosis. To test this hypothesis, we investigated the relationship between the cervicovaginal microbiome with cervical neoplasia and HPV infections and characterized the host immune profiles through profiling cytokine expression analyses from cervical lavages.

This study provides insight into specific bacterial taxonomic changes promoting a pro-inflammatory state associated with cervical dysplasia and represents the first effort to combine microbiome and cytokine profiling in Hispanics living in Puerto Rico.

**Materials and Methods**

**Participant Recruitment and Sample Collection**

Women coming for routine medical care to the University of Puerto Rico and San Juan City clinics (San Juan Metropolitan area), were recruited as part of IRB protocol 1050114 with IBC protocol 94620. Only women who did not meet the exclusion criteria were recruited and evaluated. Exclusion criteria included: 1) Active urinary infections, 2) history of common urinary incontinence, 3) sexually transmitted diseases (STDs), 4) antibiotics taken in the prior month, 5) candidiasis, and 6) treatment for or suspicion of prior toxic shock syndrome. All subjects were informed of the study's sampling procedure, risks, and benefits with written informed consent and
a verbal explanation of the document. A detailed questionnaire from which some of the metadata
variables were collected was administered. All procedures were done according to the Declaration
of Helsinki. The date range in which human subjects’ data/samples were collected was between
November 2017 and February 2020. The study and its procedures were conducted from March
2020 and October 2022.

A total of 91 women, 21-60 years old and able to provide informed consent, were enrolled in our
study. Cervical samples were collected through speculum access during the gynecological
examination by using cotton tip collection swabs. Sterile Catch-All Specimen Collection Swabs
(Epicentre Biotechnologies, Madison WI, United States) were placed in the cervix, rotated along
the lumen with a circular motion, and then placed in sterile 2 milliliters (mL) microtubes with cap.
We collected approximately 10mL of cervical lavages (CVL) by injecting pure certified nuclease-
free water (Growcells, Irvine, CA) into the vaginal canal. Collected lavages were stored in a clean
15mL collection tube and pH was measured with hydron wide range pH paper strips (Merck,
Germany). All samples were stored at −80°C and processed for nucleic acid extraction and
cytokine profiling at our UPR School of Medicine laboratory to reduce processing variation.

Cytokine Detection and Measurement with Bio-Plex Immunoassay System

A total of 91 cervical lavages were tested in duplicate with a multiplex solid-based immunoassay
(Bio-Plex Pro Human Cytokine 10-plex Assay, Bio-Rad) using the same volume of CVL across
samples (80 uL) for the Luminex MAGPIX® system. Cytokines/chemokines analyzed were IL-
1β, IL-4, IL-6, IL-8, IL-10, TNF-α, IFNγ, IP10, MIP1, and MCP1. Also, 91 samples were tested
in duplicate using Bio-Plex Pro TGF-β 3-plex immunoassay (BioRad, Hercules, CA, USA). We performed all analyses following the manufacturer’s recommendations.

Cytokine/Chemokine Profile Analyses

Data were analyzed with the Bio-Plex Manager 6.1 software (BIO-RAD, Hercules, CA) using a 5-parameter logistic curve that was verified for normality before statistical testing (20). Inflammation scores were assigned to each cytokine/chemokine for each sample using Fluorescent Intensity (FI) results. Scores were calculated by distributing samples across statistical quartiles or tertiles. If a sample was within the first quartile, a score of 1 was assigned, that is for all other three quartiles (score 1 < 0.25 < score 2 < 0.50 < score 3 < 0.75 < score 4 < 1). TGF-β scores were calculated using tertiles with scoring categories including: score 1 < 0.33, score 2 from 0.33 to 0.66 and score 3 < 1. Cytokines were grouped according to their function on host physiology, including: pro-inflammatory (IL-1β, TNF-α, IFNγ, IL-6), anti-inflammatory (IL-4, IL-10, TGFβ1) and traffic-associated cytokines (IL-8, MIP1a, MCP1, IP10) (21,22). With the scoring in either tertiles or quartiles, we also created a categorical variable, to which scores of 2 or 1 corresponded to low levels of cytokines and > 2 high levels. These ranked score assignments using the distribution of cytokines were directly related to the concentration (fluorescence intensity) of cytokines/chemokines in a sample. Therefore, we used them to combine microbial community analysis (16S rDNA profiles) with inflammation scores as a metadata variable (see methods section "Statistical Analyses of the Microbiota Profiles" for more details). This combination allowed us to detect specific taxa associated with altered cytokine profiles. For comparisons between multiple groups, we used ANOVA, and for comparisons between two groups, we used the student's t-test. All statistical analyses were done using GraphPad Prism 6.0 (San Diego, CA).
DNA Extraction and 16S rDNA Sequencing

Genomic DNA extractions from swabs were performed with the Qiagen Power Soil Kit (QIAGEN LLC, Germantown Road, Maryland, USA) following an optimized protocol for these types of samples as described in a previous study (11). DNA was quantified with a Qubit® dsDNA HS (High Sensitivity) Assay (Waltham, Massachusetts, US) (ranging from 5-100ng/µL). An average of 10-30ng/µL genomic DNA was shipped to an outsourced sequencing facility.

DNA from cervical samples was normalized to 4nM during 16S library prep. The V4 hypervariable region of the 16S ribosomal RNA marker gene (~291bp) was amplified using universal bacterial primers: 515F (5’GTGCCAGCMGCCGCGGTAA3’) and 806R (5’GGACTACHVGGGTWTCTAAT3’) as in the Earth Microbiome Project (http://www.earthmicrobiome.org/emp-standard-protocols/16s/) (23). The 16S-rRNA reads and corresponding sample metadata was deposited in QIITA study ID 12871 (24), and the raw sequences are available in the European Nucleotide Archive ENA Project under accession number EBI: ERP136546.

HPV Genotyping process

We used a kit based on a highly sensitive short-polymerase chain reaction-fragment assay (Labo Biomedical Products, Rijswijk, The Netherlands, licensed Innogenetics technology). It uses SPF10 primers to amplify a 65-bp fragment of the L1 open reading frame of HPV genotypes, followed by a Reverse-Hybridization step to determine which HPV type is present by comparing to kit-provided controls. This assay allows for the identification of the following common mucosal HPV genotypes: 14 High-risk types including (HPV 16, 18, 31, 33, 35, 39,45, 51, 52, 56, 58, 59, 66...

Quality Control and Sequence Processing

Raw read pre-processing of demultiplexed files was done with a Phred offset of 33 and default parameters using split libraries FASTQ (QIIME2 1.9.1) in QIITA (24). Sequences were trimmed to 250 bp, followed by the deblurring workflow (deblur 1.1.0) (24,25). For taxonomy assignment we used a modified Greengenes reference database (Greengenes_13.8) (26,27) with a minimum similarity threshold of 97%. The species table was downloaded for downstream analyses using a locally run version of QIIME (28). Singletons, amplicon sequence variants (ASVs) with less than three reads, and sequences matching chloroplasts, mitochondria, and taxonomically unassigned sequences were removed from downstream analyses. All samples were rarefied to 7400 reads per sample.

Vaginal Community State Type Nearest Centroid Classifier (VALENCIA)

The taxonomy of cervical microbial communities was sorted into categories termed community state types (CSTs). This categorization collapses a taxonomic profile into a single categorical variable, allowing data exploration and statistical modeling (14). In addition, categorization allows unbiased analysis of small and sizable vaginal microbiota datasets, comparisons between datasets, and meta-analyses that combine multiple datasets. We used the data package at github.com/ravel-
lab/VALENCIA to analyze the following standard parameters and categorize our cervicovaginal communities into different CST and sub-CST categories (14).

Analyses of Microbial Community Structure and Diversity

For community-level analyses, we computed the pairwise Bray-Curtis distances (Beta diversity) which quantify the compositional dissimilarity between samples (29). We visualized bacterial community composition and structure differences using non-metric multidimensional scaling (NMDS). Alpha diversity measure of Chao1 (richness), and Shannon index (30) (diversity) were computed and plotted using R with packages: “phyloseq”, “vegan” and “ggplot2” (31–33). Bar Plots showing genus and species-level taxa were computed using a locally run version of QIIME2 (28).

The metadata variables included in our study were: 1) cervical disease, defined as the cytological phenotype of negative for intraepithelial lesion or malignancy (NILM), low-grade squamous intraepithelial lesion (LGSIL) and high-grade squamous intraepithelial lesion (HGSIL); 2) HPV risk, defined as HPV negative, HPV positive with co-infection (low and high-risk HPV simultaneously), and HPV positive high-risk (exclusively high-risk types); 3) cytokine ranks (estimated through quartile and tertile distribution of the fluorescence scores as detected by Luminex MAGPIX® system); and 4) community state types (CSTs).

Statistical Analyses of the Microbiota Profiles

For beta diversity analyses, we tested if distances between sample groups were statistically significant using ANOSIM, a non-parametric statistical test (34). Using a Bray-Curtis distance
table, we used this test to compare ranked beta diversity distances between the different variables. These tests were done using the script “qiime diversity beta-group-significance” for each specific test in QIIME2 (28) with the distance matrix as the input file and 999 permutations (34). Statistical analyses for alpha diversity were done using the script “qiime diversity alpha-group-significance” in QIIME2 (28), which uses a non-parametric t-test with Monte Carlo permutations. Additionally, to ascertain if there are statistically significant differences between the ASV abundance in the different sample groups, we used linear discriminant analysis (LDA) effect size (LEfSe) (35) through R, using MicrobiomeAnalyst (36). Significant ASVs were determined considering a p-value of < 0.05 and LDA scores of 1.5 for all cytokine groups. Cytokine fluorescence values (as logarithms) were associated to differentially abundant levels of *L. crispatus, L. iners, Sneathia, Gardnerella, Atopobium* and *Prevotella* using a linear mixed model (LMM) in R (ggplot and reshape2 libraries), correcting for age and body mass index (BMI). Differential abundance analysis was performed using amplicon sequence variants corresponding to *L. crispatus, L. iners, Sneathia, Atopobium, Gardnerella* and *Prevotella*. They were evaluated for their differential presence across the cytokine groups. Initially, a summary of each one of the ASVs of interest was used to perform groupings due to abundance. The groupings were separated as follows ASV 0-50 counts = very low abundance; ASV count 51-500 = low abundance; and >500 = high abundance.

### Results

#### Population description and Community State Types (CSTs) prevalence

A total of 91 women were included in the study. The mean age was 39 years old. Women had an average pH value of 5.49. The distribution of the groups of samples according to cervical neoplasia
severity included Negative for Intraepithelial Lesion or Malignancy (NILM) (N=56) Low-Grade Squamous Intraepithelial Lesion (LGSIL) (N=18), and High Grade Squamous Intraepithelial Lesion (HGSIL) (N=17), with differential HPV status and cytokine ranks (Table 1, S1 Table). HPV infections had a prevalence of 67.0% (including infections with any HPV type) with a higher prevalence of HPV+ among women with LGSIL (94.4%) and HSIL lesions (88.2%) (Table 1). As expected, HR-HPV prevalence was higher in HGSIL (82.4%) compared to in LGSIL (50.0%) and negative cervical lesions (39.3%) (p=0.015). Participants with HPV co-infection were primarily observed in participants with low-grade lesions (44.4%) (p=0.040), and less abundant in the group with high-grade lesions (5.9%). There were no significant differences in the distribution of the BMI groups across cervical disease, with equal proportions of normal weight, overweight and obese women (Table1). Across all women, cervical lesion prevalence was 38.5% (including LGSIL and HGSIL). Individually, LGSIL and HGSIL comprised 19.8% and 18.7% of the samples, respectively. CST1 mainly was present in 16% of the NILM cases and 22% LGSIL while not being present in HGSIL. In contrast, CST IV-B was only identified on participants with high-grade lesions (HGSIL) (p=0.025) (Table 1). Prevalence analysis of CSTs per levels (high and low) of cytokine groups (pro-inflammatory, traffic-associated, and anti-inflammatory) revealed that CST IV-C is the more prevalent in all cytokines (S1 Fig., S2 Table).

**Table 1.** Characteristics of the population.

<table>
<thead>
<tr>
<th></th>
<th>NILM (N=56)</th>
<th>LGSIL (N=18)</th>
<th>HGSIL (N=17)</th>
<th>TRUE (N=91)</th>
<th>P-value *</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>40.4 (12.0)</td>
<td>39.6 (9.55)</td>
<td>33.9 (9.47)</td>
<td>39.1 (11.3)</td>
<td>0.213</td>
</tr>
</tbody>
</table>
Median [Min, Max]  
38.5 [21.0, 60.0]  
38.5 [25.0, 56.0]  
33.0 [22.0, 56.0]  
37.0 [21.0, 60.0]  

<table>
<thead>
<tr>
<th>BMI_Analysis</th>
<th>Normal</th>
<th>Obese</th>
<th>Overweight</th>
<th>Not recorded</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20 (35.7%)</td>
<td>18 (32.1%)</td>
<td>18 (32.1%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td></td>
<td>5 (27.8%)</td>
<td>8 (44.4%)</td>
<td>4 (22.2%)</td>
<td>1 (5.6%)</td>
</tr>
<tr>
<td></td>
<td>7 (41.2%)</td>
<td>5 (29.4%)</td>
<td>5 (29.4%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td></td>
<td>32 (35.2%)</td>
<td>31 (34.1%)</td>
<td>27 (29.7%)</td>
<td>1 (1.1%)</td>
</tr>
</tbody>
</table>

pH  
Mean (SD)  
5.50 (0.500)  
5.36 (0.413)  
5.60 (0.573)  
5.49 (0.500)  
0.486  

Median [Min, Max]  
5.50 [5.00, 7.00]  
5.25 [5.00, 6.00]  
5.50 [5.00, 6.50]  
5.50 [5.00, 7.00]  

<table>
<thead>
<tr>
<th>HPV_Status</th>
<th>HPV-</th>
<th>HPV+</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>27 (48.2%)</td>
<td>29 (51.8%)</td>
</tr>
<tr>
<td></td>
<td>1 (5.6%)</td>
<td>17 (94.4%)</td>
</tr>
<tr>
<td></td>
<td>2 (11.8%)</td>
<td>15 (88.2%)</td>
</tr>
<tr>
<td></td>
<td>30 (33.0%)</td>
<td>61 (67.0%)</td>
</tr>
<tr>
<td>HPV_Risk</td>
<td>HPV-</td>
<td>Co-infection</td>
</tr>
<tr>
<td></td>
<td>27 (48.2%)</td>
<td>7 (12.5%)</td>
</tr>
<tr>
<td></td>
<td>1 (5.6%)</td>
<td>8 (44.4%)</td>
</tr>
<tr>
<td></td>
<td>2 (11.8%)</td>
<td>1 (5.9%)</td>
</tr>
<tr>
<td></td>
<td>30 (33.0%)</td>
<td>16 (17.6%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CST</th>
<th>I</th>
<th>IV-A</th>
<th>IV-C</th>
<th>IV-B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>9 (16.1%)</td>
<td>12 (21.4%)</td>
<td>35 (62.5%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td></td>
<td>4 (22.2%)</td>
<td>5 (27.8%)</td>
<td>9 (50.0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td></td>
<td>0 (0%)</td>
<td>5 (29.4%)</td>
<td>8 (47.1%)</td>
<td>4 (23.5%)</td>
</tr>
<tr>
<td></td>
<td>13 (14.3%)</td>
<td>22 (24.2%)</td>
<td>52 (57.1%)</td>
<td>4 (4.4%)</td>
</tr>
</tbody>
</table>

*Kruskal-Wallis test for continuous and Fisher's test for categorical variables

Cytokine changes according to high-grade lesions and community state types (CSTs)
There were significant differences in fluorescence between pro-inflammatory cytokines IL-1\(\beta\) and IFN\(\gamma\) among women with high grade disease, compared to healthy participants. There was significantly higher fluorescence values between HGSIL and (NILM / HPV- p=0.008) or with participants with low-grade disease (p=0.015) (Fig. 1 A). Although there were no significant differences among groups when evaluating traffic-associated cytokines (Fig. 1B) or anti-inflammatory cytokines (Fig. 1C), there was a tendency for higher fluorescence levels in participants with any cervical disease (LGSIL and HGSIL). Although an analysis evaluating the cytokines according to HPV risk was done, no significant differences were identified (S2 Fig.).

Evaluation of cytokines according to CSTs revealed significant differences in IL-1\(\beta\), IL-6, MIP1a and IP10 concentrations. CST IV-B was found to be associated with a significantly higher concentration of IL-1\(\beta\) (p=1.0EE-4), IL-6 (p=0.050) and MIP1a (p=0.008) when compared to *Lactobacillus*-dominated CST I. (Fig. 2). In contrast, IP10 was significantly higher in CST I compared to CST IV-A (p=0.031) (Fig. 2).

**Fig. 1.** Cytokine Profiles grouped as pro inflammatory (A) (IL-1\(\beta\), TNF\(\alpha\), IFN\(\gamma\), IL-6), anti-inflammatory (B) (IL-4, IL-10, TGF\(\beta\)1) and traffic-associated cytokines(C) (IL-8, MIP1a, MCP1, IP10) cytokines evaluated through LUMINEX and plotted according to cervical disease. Fluorescence Intensity values were used to compute multiple comparison analysis using ordinary one-way ANOVA with Tukey’s multiple comparisons test. Results were depicted in boxplots for cervical disease. Significant differences are highlighted by brackets and corresponding p-values.

**Fig. 2.** Cytokine Profiles of IL-1\(\beta\), IL-6, MIP1a and IP10 cytokines evaluated through LUMINEX, plotted according to CST classification. Fluorescence Intensity values were used
to compute multiple comparison analyses using ordinary one-way ANOVA with Tukey's multiple comparisons tests. Results are depicted in boxplots for cervical disease, and brackets and corresponding p-values highlight significant differences.

HPV infection and cervical disease are associated with altered bacterial community composition.

A total of 2,067,074 good-quality 16S rRNA sequence reads and 8,167 OTUs were analyzed, with a rarefaction level of 7,400 reads per sample. Our samples mainly included reproductive-age women (n = 73) and women in menopause (n = 18). To evaluate whether changes in cervicovaginal microbiota are associated according to the women's physiological state, we analyzed beta and alpha diversity between these categories. We found no significant differences neither in beta (Bray-Curtis) (S3 Fig. A) nor in alpha diversity (Shannon) (S3 Fig. B). The microbiota composition did not reveal any unique taxa associated with any of the evaluated groups (S3 Fig. C). Additionally, we evaluated the cytokine profiles associated to menopause state (S3 Fig. D, E and F). We found no significant differences among menopause and non-menopause women according to the cytokine profiles. All the associations of cytokines with specific taxa were corrected for age.

Fig. 3. Diversity analyses according to cervical disease and HPV risk categories. Bray-Curtis analysis represented by Non-metric multidimensional scaling (NMDS) using cervical disease and HPV risk as metadata categories (A, D). Alpha diversity (Shannon) was calculated and depicted.
in Figs B and E. Bar Plots depicting relative abundance of bacteria at the species level for cervical disease (C) and HPV risk (F). Red asterisks mean significant differences among groups (p<0.05).

We evaluated the differences in the microbiota according to cervical disease and HPV risk. Beta diversity analyses considering cervical disease, revealed that bacterial community structure was significantly different according to cervical disease severity (LGSIL p=0.026) (HGSIL p=0.003) when compared to healthy controls (NILM/HPV-) (Fig. 3A). As for HPV risk, community structure was found to be significantly altered due to HR-HPV infection (p=0.042) and HPV co-infections (p=0.004) compared to HPV negative samples (Fig. 3D, S3 Table). Alpha diversity (Shannon) increased concomitantly to the severity of the epithelial lesions, including significant differences between NILM and HGSIL (p=0.003), regardless of HPV infection (Fig. 3B, S4 Table). In contrast, alpha diversity increased in participants with HPV co-infections as compared to HPV negatives (p=0.032) (Fig. 3E, S4 Table).

The microbiota composition between NILM and LGSIL participants was not different and was dominated by *L. crispatus* and *L. inners*. Participants with HGSIL had a change in *Lactobacillus* distribution, being dominated by *L. inners* and other taxa associated with dysbiosis, such as *Lachnospiraceae sp.*, *P. bivia*, *A. vaginae* and *G. vaginalis* (Fig. 3C). Bacterial composition changed dramatically according to HPV Risk, favoring a predominance of *L. crispatus* in HPV negative women, and *L. iners* in high-risk participants (Fig. 3F). Participants with HPV co-infections had a highly diverse taxonomic profile with dysbiosis associated taxa including, *Prevotella sp.*, *Megasphaera sp.*, *Sneathia*, *Aerococcus christinsenii* and *Streptococcus anginosus* (Fig. 3F).
As identified by linear discriminant analysis (LDA) effect size (LEfSe), putative microbial biomarkers detected *Parabacteroides* and *Akkermansia* in participants with HGSIL. As for HPV risk, we found that *Sneathia* was more abundant in HPV-negative participants with cervical disease. In addition, for HPV co-infections, we identified *Megasphaera, Parabacteroides, Akkermansia* and *Bacteroides* as significantly differentiated taxa (Fig. 4).

**Fig. 4. Linear discriminant analysis (LDA) effect size (LEfSe) for cervical disease and HPV risk at the Genus level.** Cutoff was 1.5 LDA score and a p value of 0.05. Log-level changes in LDA score are displayed on the x axis.

**Association of bacterial taxa with cytokine levels**

In a balanced vaginal tract, *Lactobacillus* species impede the colonization of other bacteria that would otherwise cause infections. *L. Crispatus* produce lactic acid, bacteriocins and H$_2$O$_2$ acting as a chemical barrier (Valenti et al. 2018). Very low levels of *L. crispatus*, a key vaginal bacterium in eubiosis, were significantly associated to high concentrations of pro-inflammatory cytokines IL-1β (p=2.16E-6) and IFNγ (p=0.013) (Fig. 5). High levels of *L. iners* were significantly associated with high levels of IL-6, also a pro-inflammatory cytokine (p=0.019), suggesting that this species of *Lactobacillus* might stimulate IL6 production (Fig. 5). On the contrary, low levels of *L. iners* correlate with lower expression of IFNγ (p=0.039) (Fig. 5). Very low levels of *Atopobium*, a mostly pathogenic bacteria were significantly associated with low levels of IL-1β (p=0.002). As for *Sneathia*, another dysbiosis associated taxa, detected at low levels, had correlation with higher levels of IFNγ (p=7.00E-6), TNFα (p=7.00E-4) TNFα (p=0.010) (Fig. 5). Absence, or very low levels anaerobic *Gardnerella*, were associated with lower fluorescence...
values of IL-1β (p=1.00E-6) (Fig. 5). Very low abundance of *Prevotella* correlated with lower concentrations of IL-1β in our samples (p=3.00E-6) (Fig. 5).

**Fig. 5. Association of key vaginal bacteria with levels of pro-inflammatory cytokines.** Cytokine logarithm fluorescence values were associated to differentially abundant levels of *L. crispatus*, *L. iners*, *Sneathia*, *Gardnerella*, *Atopobium* and *Prevotella* using a linear mixed model (LMM), correcting for age and BMI. Differential abundance groupings were separated as follows ASV 0-50 counts = very low abundance; ASV count 51-500 = low abundance; and >500 = high abundance.

In beta diversity, axis 1 of the NMDS1 plot separated samples according to their cytokine levels (S4 Fig. A, D, G). When compared to healthy controls (NILM/HPV-) we found changes in microbial community structure significantly associated with high levels of pro-inflammatory cytokines regardless of HPV infection (NILM/HPV-, p=0.006; NILM/HPV+, p=0.001) and cervical disease severity (LGSIL, p=0.002; HGSIL, p=0.001) (S4 Fig. A, S3 Table). High levels of anti-inflammatory cytokines associated with changes in bacterial composition in similar patterns as pro-inflammatory cytokines, with significant differences regardless of HPV infection (NILM/HPV-, p=0.001; NILM/HPV+, p=0.001) and cervical disease severity (LGSIL, p=0.012; HGSIL, p=0.001) (S4 Fig. D, S3 Table). High levels of traffic-associated cytokines significantly altered bacterial composition regardless of cervical disease severity (LGSIL, p=0.012; High_HGSIL, p=0.001) (S4 Fig. G, S3 Table). Alpha diversity significantly increased with cytokine levels for both pro-inflammatory (NILM/HPV-, p=0.035; NILM/HPV+, p=0.002) (LGSIL, p=0.002; HGSIL p=0.001) and anti-inflammatory (NILM/HPV-, p=0.039; NILM/HPV+,
p=0.005) (LGSIL, p=0.001; HGSIL, p=3.55E-4) cytokines regardless of HPV infection and cervical disease (S4 Fig. B, 4E). Traffic-associated cytokines did not have significant alpha diversity differences associated with cytokine concentrations (S4 Fig. H, S4 Table). Taxonomic abundance at species level showed that in the 3 cytokine groups, populations of *L. crispatus* decreases as cytokine levels and cervical disease severity increased. Participants with high levels of cytokines and high-grade cervical disease were dominated by *L. iners*, as *L. crispatus* nearly disappears. Additionally, this group of participants were dominated by *Prevotella sp.*, *A. vaginae* and *G. vaginalis*. Simultaneously, *Lachnospiraceae sp.* was detected mainly on participants with High cytokine levels and HGSIL, confirming the dysbiosis typical of *Lactobacillus* depleted communities (S4 Fig. C, 4F and 4I). LefSe analysis (p=0.05 and 1.5 LDA score cutoffs) revealed an increase in *Lactobacillus* when pro-inflammatory and anti-inflammatory cytokines were low (Fig. 6). Low levels of traffic-associated cytokines presented anaerobic bacteria like *Prevotella* and *Atopobium* with higher abundance (Fig. 6). In contrast, with high cytokine levels in pro-inflammatory and anti-inflammatory cytokines, *Atopobium* and *Prevotella* populations decreased. *Pseudomonas* communities also were reduced in association with pro-inflammatory and traffic-associated cytokines (Fig. 6).

**Fig. 6.** Linear discriminant analysis (LDA) effect size (LEfSe) for (A) pro-inflammatory (IL-1β, TNFα, IFNγ, IL-6) (LDA score 1.5), (B) anti-inflammatory (IL-4, IL-10, TGFβ1) (LDA score 2.25) and (C) traffic (IL-8, MIP1a, MCP1, IP10) (1.5 LDA score) cytokines evaluated considering low and high cytokine levels and a p-value cutoff of 0.05.
Cervical cancer (CC) is a complex disease where the cervical environment is modified, including cytokine expression that leads to a local immunosuppression state (18). Approximately 15 high-risk HPV strains cause most cervical cancers; however, the detection of HPV has a low positive predictive value for invasive cervical cancer, as few HPV+ women progress to High-Grade Squamous Intraepithelial Lesions (HGSIL) (6,37). Inflammation and the microbiome could be the missing link to understanding the actual cellular mechanisms leading to dysplasia and cancer.

Indeed, to provide an accurate diagnosis and prognosis of cervical disease progression, screening tests have limited sensitivity for histopathology and cytology (38). Our sensitive HPV genotyping method demonstrated that HPV prevalence in this cohort was 67.0% (encompassing infections with any HPV type) and that co-infections with any HPV serotype were primarily present in low-grade lesions (44.4%) and exclusively high-risk types were more prevalent in HGSIL (82.3%), a prevalence previously reported by our team (11) and Hispanic and Amerindian peoples (39). This higher prevalence exceeds those reported for Costa Ricans (40), or women in Surinam (41), Nigeria (42), Turkey (43) or Italy (44). This can be attributed to the fact that these women are recruited in colposcopy clinics and are therefore not representative of the Puerto Rican population.

Modifying the host innate immune system is a well-known feature of numerous malignancies, including HPV-driven cancers (45). HPV interferes with the cytokines, Toll-like receptors (TLRs), and the overall inflammatory process, occurring in the early stages of the immune response (46,47). Together, cytokines and TLRs identify HPV and encourage immunological signalings such as cell development and differentiation (45,47). In cervical HPV infections, an impaired immune response to the virus may lead to cervical disease progression (48). Overall, we found that
any HPV infection elevated all cytokine levels but not all cytokines were significant. This suggests that HPV infections alone cannot drive cervical dysplasia progression to cancer (49). Most HR-HPV-infected women do not develop CC because an effective immune response controls the infection and therefore, malignant transformation is halted (50). Additionally, HPV uses several strategies to evade the immune system (Miura et al., 2010), such as reducing the amount of IFNγ, altering the activation of antiviral and antitumor immune cells such as NK cells and T cells with antiviral and antitumor activity (51). In addition to this, an imbalance of the mucosal microbiome, such as a loss in protective *Lactobacillus* and physicochemical changes, causes alterations of the vaginal mucosa and the cervical epithelium, exercising a pattern of selective pressure on the microbiota (52).

Several studies have already established background data on how the microbiome is modified by HPV infections, stating that HPV infections alone are not enough to drive significant changes that translate into epithelial modifications to cancer (11,53). Nonetheless, other studies suggested changes in the composition of the vaginal microbiota associated with human papillomavirus infection (54–56). Our results show greater microbiota diversity and a lower abundance of lactobacilli among HPV-positive women, which is congruent with previous studies (J. E. Lee et al. 2013; Oh et al. 2015; Reimers et al. 2016, Godoy-Vitorino et al. 2018). It is still unclear why some HR-HPV infections clear while others persist, causing dysplasia and even cervical cancer. Therefore, since only persistent HR-HPV infection has been associated with immune response and vaginal dysbiosis, molecular differences between transient and persistent HR-HPV infection may hold information to explain this phenomenon (37).

Concerning cervical lesions, we found a strong microbiome association with the cervical phenotype (lesion severity), with a significantly higher diversity value (Shannon) in HGSIL
compared to NILM/HPV- and NILM HPV+ participants, but not substantially different from LGSIL. Our findings coincide with most studies suggesting an association between *L. iners* and HGSIL (60,61). *Lactobacillus* produces have several mechanisms that function as a defensive barrier for invasive bacteria, such as, the production of significant amounts of bacteriocins that contributes to colonization the use of available glycogen to produce lactic acid, lowering vaginal pH, responsible of lysing bacteria other than *Lactobacillus* (62–65)(Amir et al. 2020; Cruickshank 1934; O’Hanlon et al. 2013; Zhou et al. 2019). The loss of this defensive strategy due to *Lactobacillus* community reduction, explains the increase of *Atopobium vaginae*, *Prevotella* spp., and *Gardnerella* spp.

Our study confirms cervicovaginal microbial dysbiosis in high-risk HPV and cervical disease severity associated with pro-inflammatory cytokines, validating other studies in other cohorts (18,66). IL-1β and IFNγ were significantly more abundant in high-grade disease participants, associated with a loss in *Lactobacillus* validating previous studies where high concentrations of pro-inflammatory cytokines were detected in women with cervical neoplasia (67–69). The rise in IL-1β can be attributed to cellular mechanisms such as tumorigenesis, angiogenesis (70) or metastasis (71); as well as increased risk of CC in women with high grade cervical lesions (72) associated with either polymorphism (73), reactive oxygen species (ROS), COX2 or NADPH oxidase (74). We found higher levels of expression of IL-6 associated with high levels of *L. iners*. IL-6 is essential for the growth, metastasis, and initiation of cancer; affecting tumor growth, differentiation, and the host immune defense mechanism (75). Many immune and nonimmune cells, such as T cells, monocytes, macrophages, dendritic cells, and some cancer cells -such as CC cells- produce IL-6 (76,77). IL-6 induces CC growth by vascular endothelial growth factor (VEGF)-dependent angiogenesis (78). In addition, higher expression of IL-6 was shown in
cancerous tissues than in adjacent non-cancer tissues in early-stage CC patients, confirming its association to cancer development (79). The less protective Lactobacilli (*L. iners*) may be stimulating the production of these mediators, this could be due in part to their heterogeneous genome, that unlike the protective counterparts, do not produce as much lactic acid or hydrogen peroxide (80). Although the slight increase in IL-10 among LGSIL is not significant, we hypothesize that IL-10 may increase in parallel with the development of intraepithelial lesion as it does in cancer, promoting the expression of HPV oncoproteins (Berti et al. 2017).

Taxonomic profiles of cervical microbial communities were sorted into community state types (CSTs), allowing a valuable categorization for statistical modeling and comparisons between datasets (13,14). The vaginal microbiota in this Hispanic women cohort of women living in Puerto Rico was dominated by CST IV-C (57.1%, non-*Lactobacillus* profile) and CST IV-A (24.1%, *G. vaginalis* moderate), which confirms previous work of our lab (11), and suggests similarities to other US Hispanics (13) and Venezuelans (81). Even when CST IV-C dominated all three categories of cervical cytology (NILM, LGSIL, HGSIL), but was more predominant in healthy patients (NILM) (62.5%). CST IV-C is characterized by a low relative abundance of *Lactobacillus spp.*, *G. vaginalis*, and *A. vaginae*, while instead is characterized by the abundance of a diverse array of facultative and strictly anaerobic bacteria, thus revealing that Hispanic women living in PR are in constant dysbiosis with a microbiome dominated by anaerobic non-lactobacillus communities. Overall, the high prevalence of CST-IV across all samples reflects a cervical microbial environment with low stability and less likelihood to return to its equilibrium state after a perturbation, which increases the risk of developing adverse health outcomes (82,83). In Hispanics living in Puerto Rico, the typical healthy vaginal microbial profile seems to be already diverse. It does not correspond to that of US Caucasians, mostly dominated by CST-I, being more
like that of other US Hispanics, which are more prone to vaginal dysbiosis regulation and inflammatory conditions (Ravel et al. 2011). CST-IV represents a highly diverse bacterial profile and has been associated with high pro-inflammatory cytokine levels, which damage the endocervix’s columnar epithelial barrier, exposing it to colonization to other microbial or pathogenic agents (12,84).

Microbiome analysis revealed that high levels of all three groups of cytokines (anti-inflammatory, pro-inflammatory, and traffic-associated) were related to significant bacterial community composition changes and high diversity levels (Shannon). Low levels of Sneathia spp. are associated with higher concentrations of IL-1β and TNFα, this taxa that is part of the normal microbiota of the genitourinary tracts of men and women is now considered a possible pathogen of the female reproductive tract (85), Sneathia spp. are also regarded as good biomarkers for bacterial vaginosis (86), a classical dysbiosis associated with abnormalities in cervical smears (87).

Previous studies have described that Lactobacillus can prevent the colonization of exogenous pathogens by producing lactic acid, bacteriocins and reactive oxygen species (ROS) (13,88,89). Lactobacilli, promote homeostasis in the cervicovaginal environment and are associated with low levels of pro-inflammatory and anti-inflammatory cytokines. In contrast, low concentrations of traffic-associated-related cytokines are associated with an increase in Prevotella and Atopobium, suggesting that there is little induction of immune cell tissue migration to combat these bacteria. Previous studies reported that Atopobium vaginae could trigger an innate immune response involving IL-6 and IL-8 using an in vitro model of bacterial vaginosis (90), and this could be happening in vivo. Also, a study suggested that Prevotella may provide nutrients such as ammonia and amino acids to other microbial community members such as Gardnerella and Peptostreptococcus (91), assuming a metabolic hub for vaginal microbiota and
being a key player in vaginal dysbiosis (92). *Prevotella* is therefore critical for maintaining a dysbiotic ecosystem in the cervix.

Cervical cancer is a complex multifactorial disease, and further research, including time-series analyses of longitudinal sampling, is needed to improve our understanding of its etiology. Our study evidence that high concentrations of all three cytokine groups (inflammatory, anti-inflammatory, and traffic-associated) were associated with cervical dysbiosis, which occurs through dramatic losses of protective *Lactobacilli* and the presence of diverse anaerobic taxa. The cervical microbiota in Hispanic women living in Puerto Rico is highly diverse and dominated by CSTs III and IV. Further changes of the microbiota with HR-HPV and inflammation reveal that the joint host-microbe interaction analyses via cytokine signaling and microbiota in precancerous lesions have great translational potential.

## Acknowledgments

We would like to acknowledge Dr. Robert D. Burk, Dr. Mykhaylo Usyk for kindly providing the curated extended GreenGenes database we cite in the methods section, which integrates the GreenGenes database, the Human Oral Microbiome Database (HOMD) and cervicovaginal microbiome 16S reference sequences retrieved from NCBI. Also, we want to thank Dr. Michael T. France for his help in adapting our taxonomy for the VALENCIA classification. We thank all the medical personnel from the gynecology clinics who participated in the screening program and sample collection and the students who helped with metadata curation. We offer special thanks to the participants for contributing to our study.
References


70. Filippi I, Carraro F, Naldini A. Interleukin-1β Affects MDAMB231 Breast Cancer Cell Migration under Hypoxia: Role of HIF-1α and NFκB Transcription Factors. Mediators Inflammm. 2015;2015:789414.


Supporting information

S1 Fig. CST prevalence distribution among cytokine levels and groups. CSTs according to category groups were used to compute Fisher’s test using pairwise analysis to compare between groups. Results were depicted in boxplots for cervical disease. Significant differences are highlighted by brackets and corresponding p-values.

S2 Fig. Cytokine Profiles of (A) pro inflammatory (IL-1β, TNFα, IFNγ, IL-6), (B) anti-inflammatory (IL-4, IL-10, TGFβ1) and (C) traffic (IL-8, MIP1α, MCP1, IP10) cytokines evaluated through LUMINEX, plotted according to HPV risk. Fluorescence Intensity values were used to compute multiple comparison analysis using ordinary one-way ANOVA with Tukey’s multiple comparisons test. Results were depicted in boxplots for cervical disease. Significant differences are highlighted by brackets and corresponding p-values.

S3 Fig. Diversity and cytokine profile analyses comparing samples, according to menopause status. Beta and alpha diversity analyses are represented by non-metric multidimensional scaling
(NMDS) (A) and Shannon index boxplots (B). Relative abundance of bacteria at the species level is shown in a relative abundance bar plot (C). Cytokine Profiles of pro-inflammatory (D), anti-inflammatory (E) and traffic (F) cytokines, plotted according to HPV risk. Fluorescence Intensity values were used to compute multiple comparison analysis using ordinary one-way ANOVA with Tukey’s multiple comparisons test. Significant differences are highlighted by brackets and corresponding p-values.

S4 Fig. Diversity analyses comparing samples, according to inflammatory (IL-1β, TNFa, IFNg, IL-6), anti-inflammatory (IL-4, IL-10, TGFβ1), and trafficking (IL-8, MIP1a, MCP1, IP10) cytokines. Bray-Curtis analysis represented by Non-metric multidimensional scaling (NMDS) using pro-inflammatory (A), anti-inflammatory (D) and trafficking cytokines (G) as metadata categories. Alpha diversity (Shannon) was calculated and depicted in Figures B, E and H. Bar Plots showing relative abundance of bacteria at the species level for cytokine groups (C, F, I).

S1 Table. Study population characteristics (n=91) described by cervical disease status, including HPV risk, community state types, and cytokine ranks.

S2 Table. Pairwise comparisons using Pairwise comparison of proportions (Fisher) of the cytokine levels. P values were adjusted using Bonferroni corrections.
S3 Table. Non-parametric analyses using ranked beta diversity distances between different group depths found in the mapping file, based on the Bray-Curtis table.

S4 Table. Non-parametric t-test comparing observed richness between the different metadata categories.
**Figure 5**

- **L. crispatus**
  - *IFNγ*
  - IL-6
  - TNFα
  - *IL-1β*
  - Fluorescent Intensity (FI)

- **L. iners**
  - *IFNγ*
  - IL-6
  - TNFα
  - IL-1β
  - Fluorescent Intensity (FI)

- **Sneathia**
  - *IFNγ*
  - IL-6
  - TNFα
  - IL-1β
  - Fluorescent Intensity (FI)

- **Gardnerella**
  - IFNγ
  - IL-6
  - TNFα
  - *IL-1β*
  - Fluorescent Intensity (FI)

- **Atopobium**
  - IFNγ
  - IL-6
  - TNFα
  - *IL-1β*
  - Fluorescent Intensity (FI)

- **Prevotella**
  - IFNγ
  - IL-6
  - TNFα
  - *IL-1β*
  - Fluorescent Intensity (FI)

Legend:
- **Very Low**
- **Low**
- **High**
Figure 1
Figure 4
Figure 6

Pro Inflammatory

Lactobacillus
Pseudomonas
Atopobium
Not_Assigned
Prevotella

Anti Inflammatory

Lactobacillus
Pseudomonas
Not_Assigned
Atopobium
Prevotella

Traffic

Prevotella
Pseudomonas

Low Cytokine Levels
High Cytokine Levels

LDA score