PLOS PATHOGENS

Cervicovaginal microbiome and natural history of HPV in a longitudinal study

Mykhaylo Usyk[1,2], **Christine P. Zolnik**[1,3], **Philip E. Castle**[4], **Carolina Porras**[5], **Rolando Herrero**[6], **Ana Gradissimo**[1], **Paula Gonzalez**[5], **Mahboobeh Safaeian**[7], **Mark Schiffman**[8], **Robert D. Burk**[1,4,9]*, **Costa Rica HPV Vaccine Trial (CVT) Group**[¶]

1 Department of Pediatrics (Genetic Medicine), Albert Einstein College of Medicine, Bronx, New York, United States of America, **2** Department of Epidemiology and Population Health, NYU School of Medicine, New York, New York, United States of America, **3** Department of Biology, Long Island University, Brooklyn, New York, United States of America, **4** Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, New York, United States of America, **5** Agencia Costarricense de Investigaciones Biomédicas (ACIB), formerly Proyecto Epidemiológico Guanacaste, Fundación INCIENSA, San José, Costa Rica, **6** Prevention and Implementation Group, International Agency for Research on Cancer, Lyon, France, **7** Roche Molecular Diagnostics, Pleasanton, California, United States of America, **8** Division of Cancer Epidemiology and Genetics (DCEG), National Cancer Institute, NIH, Bethesda, Maryland, United States of America, **9** Departments of Microbiology and Immunology, and Obstetrics and Gynecology and Women's Health, Albert Einstein College of Medicine, Bronx, New York, United States of America

¶ Members of the Costa Rica HPV Vaccine Trial (CVT) Group are listed in the Acknowledgements Section of the Manuscript

* robert.burk@einsteinmed.org

Abstract

Background

Human papillomavirus (HPV) infection is one of the most common sexually transmitted infections. However, only a small percentage of high-risk (HR) HPV infections progress to cervical precancer and cancer. In this study, we investigated the role of the cervicovaginal microbiome (CVM) in the natural history of HR-HPV.

Methods

This study was nested within the placebo arm of the Costa Rica HPV Vaccine Trial that included women aged 18–25 years of age. Cervical samples from two visits of women with an incident HR-HPV infection (n = 273 women) were used to evaluate the prospective role of the CVM on the natural history of HR-HPV. We focus specifically on infection clearance, persistence, and progression to cervical intraepithelial neoplasia grade 2 and 3 (CIN2+). The CVM was characterized by amplification and sequencing the bacterial 16S V4 rRNA gene region and the fungal ITS1 region using an Illumina MiSeq platform. OTU clustering was performed using QIIME2. Functional groups were imputed using PICRUSt and statistical analyses were performed using R.

Results

At Visit 1 (V1) abundance of *Lactobacillus iners* was associated with clearance of incident HR-HPV infections (Linear Discriminant Analysis (LDA)>4.0), whereas V1 *Gardnerella* was

the dominant biomarker for HR-HPV progression (LDA>4.0). At visit 2 (V2), increased microbial Shannon diversity was significantly associated with progression to CIN2+ (p = 0.027). Multivariate mediation analysis revealed that the positive association of V1 *Gardnerella* with CIN2+ progression was due to the increased cervicovaginal diversity at V2 (p = 0.040). A full multivariate model of key components of the CVM showed significant protective effects via V1 genus *Lactobacillus*, OR = 0.41 (0.22–0.79), V1 fungal diversity, OR = 0.90 (0.82–1.00) and V1 functional Cell Motility pathway, OR = 0.75 (0.62–0.92), whereas V2 bacterial diversity, OR = 1.19 (1.03–1.38) was shown to be predictive of progression to CIN2+.

Conclusion

This study demonstrates that features of the cervicovaginal microbiome are associated with HR-HPV progression in a prospective longitudinal cohort. The analyses indicated that the association of *Gardnerella* and progression to CIN2+ may actually be mediated by subsequent elevation of microbial diversity. Identified features of the microbiome associated with HR-HPV progression may be targets for therapeutic manipulation to prevent CIN2+.

Trial registration

ClinicalTrials.gov NCT00128661.

Author summary

Despite being the most common sexually transmitted infection and the causal agent of cervical cancer, it is still not clear why only a small proportion of high-risk HPV (HR-HPV) infections progress to cervical cancer. Our study utilizes longitudinal cervicovaginal samples from a prospective cohort, along with advanced epidemiological modeling and mediation analysis, to investigate the association between the cervicovaginal microbiome (CVM) and progression of an incident HR-HPV infection to cervical precancer. The results of our study suggest a novel association between the effect of *Gardnerella* and disruption of CVM homeostasis that can influence the pathway of HR-HPV infection progression to cervical precancer. We further show the interplay between several key components of the cervicovaginal microbiome and demonstrate that within the context of HR-HPV natural history the effect of *Gardnerella* is mediated by increased cervicovaginal bacterial diversity directly preceding the progression of a persistent infection to precancer.

Introduction

Persistent cervical infections by high-risk (HR) human papillomavirus (HPV) cause virtually all cervical cancers and their immediate precursor lesions [1]. Most sexually active women have been infected with HPV at some point in their lives and in the vast majority the infection is cleared within a few months [2]. However, a subset of women develop a persistent HPV infection that places them at high risk for cervical precancer and cancer [2–5]. **Fig 1** illustrates this paradigm canonically referred to as HPV natural history.
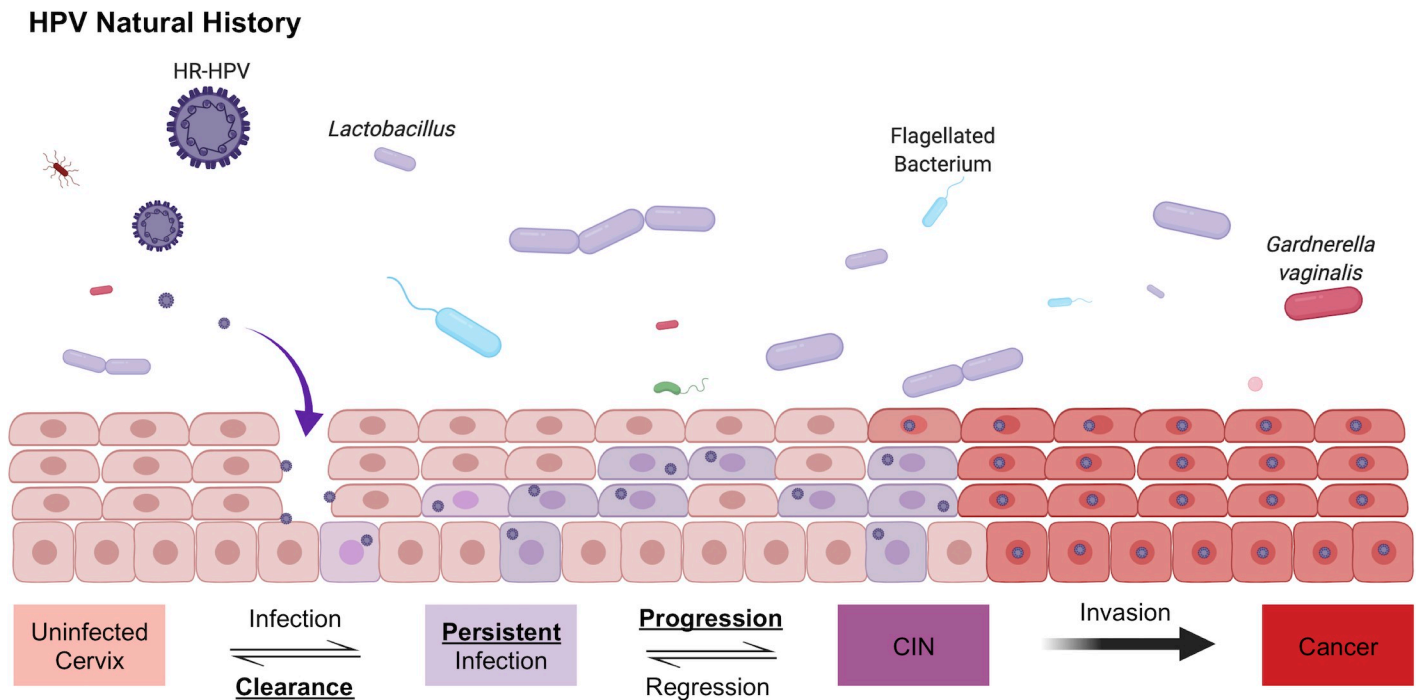
HPV Natural History

Fig 1. HPV natural history. The natural history of HR-HPV is depicted. Briefly, an incident HR-HPV infection can occur by entering the basal layer through an epithelial abrasion. Most incidence infections are cleared, however some remain persistent for years and decades. Persistence of a HR-HPV infection combined with known risk factors (e.g., smoking) may allow the persistent HR-HPV infection to progress to precancer (cervical intraepithelial neoplasia, CIN). If the lesion does not regress and the HR-HPV is able to successfully integrate into the host-cell genome, clonal expansion may occur and result in an invasive cancer.

https://doi.org/10.1371/journal.ppat.1008376.g001

Non-viral factors (HPV co-factors) associated with the outcomes of HR-HPV infections have not been fully elucidated. While smoking [6–9], hormonal contraceptive use [10, 11], and parity [12] are associated with developing precancer and cancer; systemic and local immune responses are thought to be important for clearance and control of infection (persistence vs. clearance) [13, 14]. In addition, specific host immune regulatory alleles (e.g., human leukocyte antigen) are associated with risk of cervical cancer [15, 16].

The local, cervical microenvironment, including the microbiome, may also influence the natural history of HPV infection [17]. Other studies have recently implicated the microbiome's role in the natural history of other viral infections [18] and a variety of cancers [19–21]. The cervicovaginal microbiome (CVM) is of particular interested because it has been well characterized and specific features have been associated with gynecologic disease and reproductive health [22–24]. The CVM has been categorized into community state types (CSTs) generally defined by a dominance of a specific *Lactobacillus* species (*Lactobacillus crispatus*, *Lactobacillus iners*, *Lactobacillus gasseri* or *Lactobacillus jensenii*), or a state of polymicrobialism [25, 26]. Transitions from *Lactobacillus* dominated *CSTs* have been linked to detrimental health outcomes including elevated risks for sexually transmitted infections [27], as well as higher incidences of preterm births [28].

An association between increased CVM diversity and prevalence of HR-HPV infection and/or cervical abnormalities (vs. HPV negative) has been reported in several studies [29–33]. Higher abundance of *L. crispatus* has been shown to be associated with lower HPV prevalence [34] and increased detection of normal cytology [35]. Long-term use of vaginal probiotics with *Lactobacillus* spp. has been associated with increase clearance of HPV compared to short-term use [36]. However, evidence is conflicting on the association of CVM diversity and the severity

of cervical neoplasia [37–39]. Additionally, most studies looking at the natural history of HPV and the microbiome are cross-sectional and therefore lack the ability to draw potential causal links.

For the current study, we leveraged longitudinal data and specimens from the placebo arm of a large randomized HPV vaccine trial [40] to examine the impact of the CVM on the natural history of incident HR-HPV infections to study: 1) progression to cervical precancer, 2) viral persistence, and 3) viral clearance.

Results

Subject characteristics and cervicovaginal microbiome features

A total of 273 women with an incident HR-HPV infection were included in the analyses, of whom 266 had a second sample with an average sampling interval of 1.5±0.9 years. **Table 1** presents the sample demographic information and summarizes the bacterial and fungal sequencing results after taxonomic assignment for each infection outcome at baseline. There were no significant differences between groups in age (p = 0.13), 16S rRNA gene OTU clustered read counts (p = 0.33), or ITS1 OTU clustered read counts (p = 0.53).

Fig 2 summarizes the bacterial Shannon diversity measures of the three ordered categorical HR-HPV outcomes (see **Fig 1**) within the CVT cohort at V1 and V2. Alpha diversity analysis did not reveal any significant differences at V1 in terms of bacterial Shannon alpha diversity (trend p = 0.52). At V2 there was a significant trend of rising diversity based on the Shannon diversity index (trend p = 0.024).

To evaluate the overall structure of the cervicovaginal microbiome, we performed hierarchical clustering on all available samples (n = 539) (**Fig 3A**). This analysis revealed four distinct bacterial community state types (CSTs). Two CSTs were dominated by species of the genus *Lactobacillus* (*Lactobacillus iners*, 143/539, 26.5% and *Lactobacillus crispatus*, 83/539, 15.4%), one CST by *Gardnerella vaginalis* (94/539, 17.4%), and the other CST did not contain a major group but had a highly diverse microbiome (219/539, 40.6%).

Table 1. CVT cohort characteristics.

Variable	Clearance	Persistence	Progression	p-val
Baseline Sample Count	*70*	*170*	*33*	
Age (years)	23.4 ± 2.8	22.6 ± 2.3	22.70 ± 2.9	0.13
Current/former Smoker	15.7% (11/70)	27.1% (46/170)	26.7% (8/30)	0.16
Total number of sexual partners as of visit	3 ± 3	3 ± 2	3 ± 2	0.63
Contraceptive use*	97.1% (68/70)	97.6% (166/170)	96.7% (29/30)	0.86
Condom use prior to visit	44.1% (15/34)	39.3% (33/84)	58.3% (7/12)	0.46
Contraceptive pill prior to visit	68.8% (33/48)	68.5% (76/111)	52.6% (10/19)	0.40
Injectable contraceptive prior to visit	14.8% (8/54)	21.7% (26/120)	23.8% (5/21)	0.53
Other contraceptive prior to visit**	4.6% (3/65)	0% (0/150)	0% (0/26)	0.029
Chlamydia/Gonorrhea Positive	46.2% (6/13)	30.4% (14/46)	0% (0/3)	0.32
HPV16 Positive	37.1% (26/70)	34.7% (59/170)	45.5% (15/33)	0.49
16SV4 Reads	10,642 ± 4,910	11,482 ± 4,985	11,041 ± 4,809	0.33
ITS1 Reads	1,942 ± 3,433	3,299 ± 7,055	4,660 ± 11,500	0.53

Continuous data is presented using median ± standard deviation, significance assessed using the Kruskal-Wallis test.

Count data is presented using percent with proportions shown in parenthesis, significance is assessed using Fisher's exact test.

Proportion of count data is based on only the samples which had data for a given count variable.

*Counts use of any of the following prior to the visit: condom, birth control pills, diaphragm, injectable, iud, spermicide, sponge and or other** types of contraceptives.
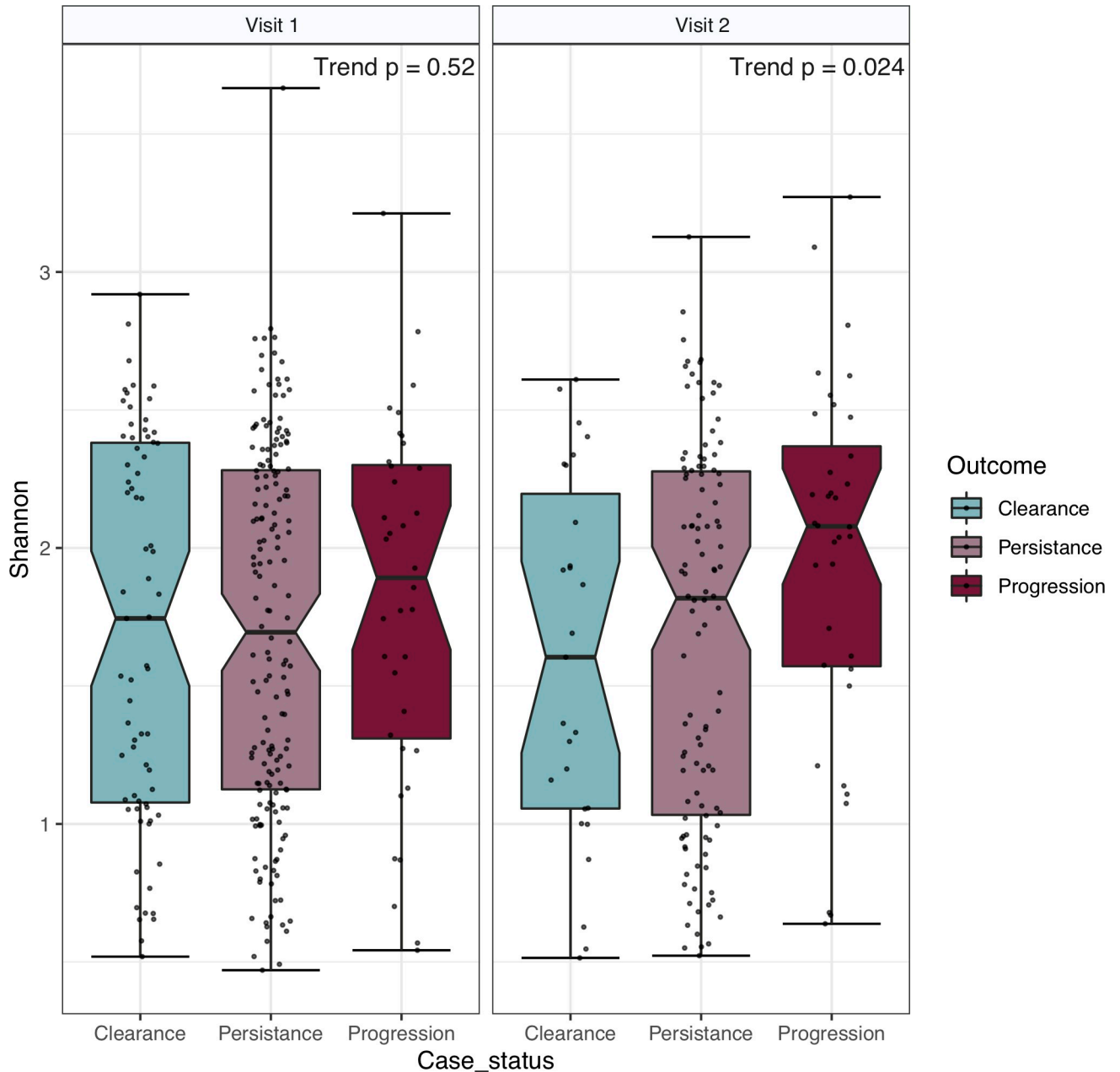
Fig 2. Bacterial Shannon diversity by visit. HR-HPV category specific microbial Shannon diversity is shown for V1 and V2. Horizontal strip labels at the top of the figure indicate visit number. V1 has an elevated diversity in the progression group, but the overall trend did not achieve statistical significance (p = 0.52). At V2 the observed trend of a rising Shannon alpha diversity from clearance to persistence to progression was statistically significant, p = 0.024.

Fig 3B shows results of hierarchical clustering based on the detected fungal species. *Candida albicans* was the dominant fungal taxa. In terms of fungal clusters, there appears to be a single clade dominated by *C. albicans* (43/208, 20.7%), one dominated by an unidentified fungal species (12/208, 5.8%) and one that is composed of a diverse fungal community (153/208, 73.6%).

A. Abundance Plot Showing the Top 10 Bacteria in the CVT Cohort
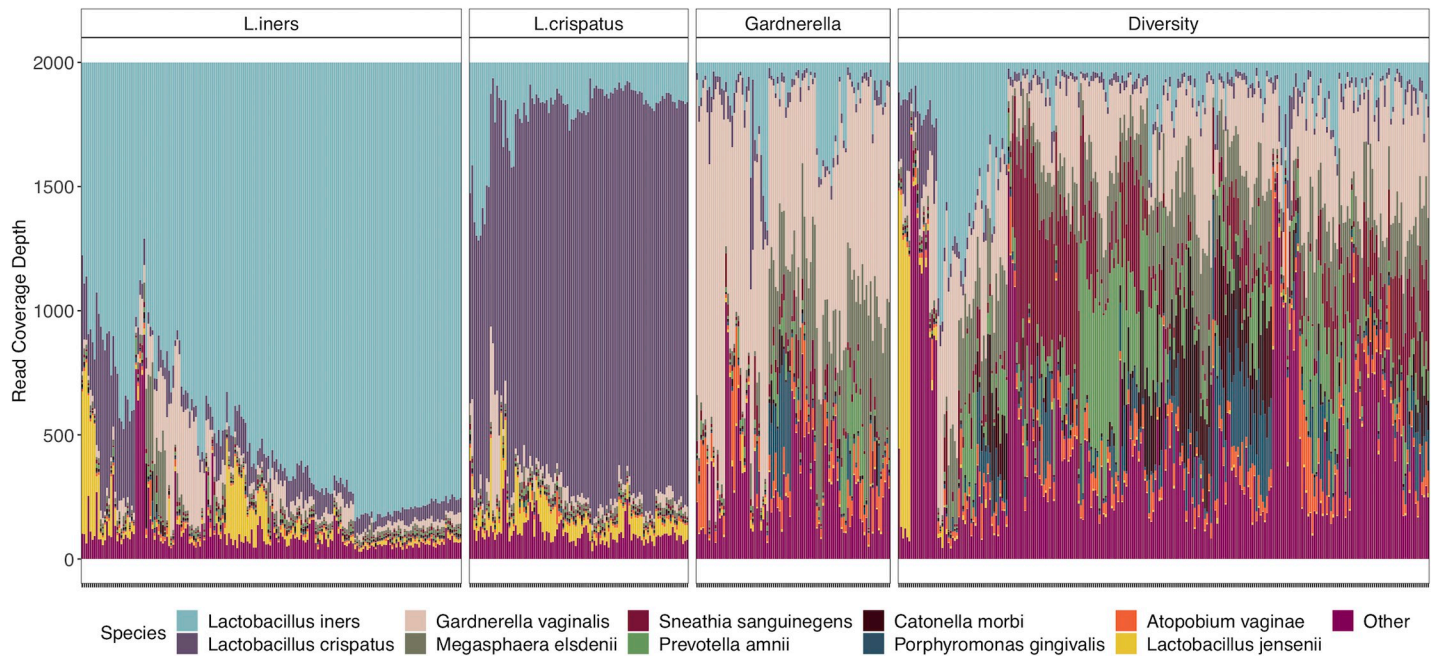
B. Abundance Plot Showing the Top 10 Fungi in the CVT Cohort
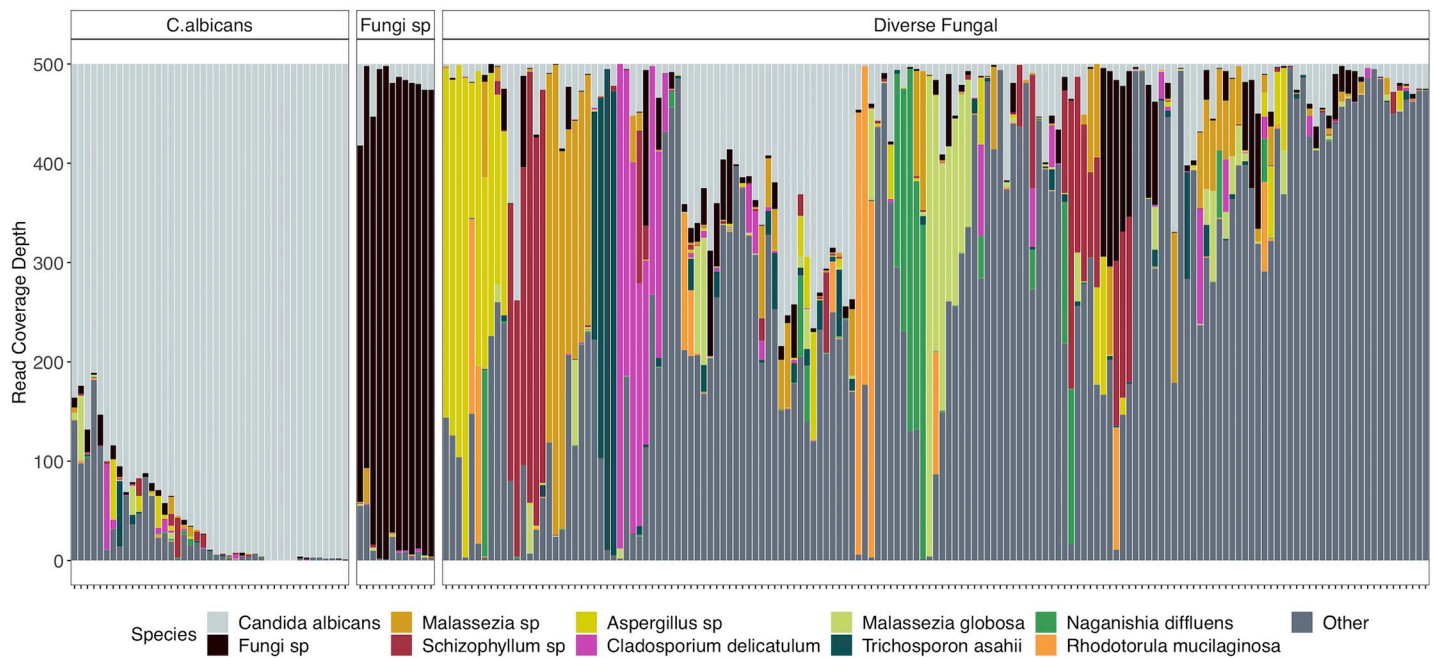
Fig 3. Bacterial and fungal communities within the study cohort. A. The abundance plot represents the bacterial community structure of the study subjects. The operational taxonomic units (OTUs) were collapsed at the species level and the top 10 species are presented. Figure boxes labeled: L. iners, L. crispatus, Gardnerella and Diversity represent the vaginal community state types (CSTs) identified using hierarchical clustering. B. Heatmap showing the top 20 fungal species identified within the study subjects. *C. albicans* has the highest mean abundance. There were three vaginal fungal community states identified using hierarchical clustering as indicated by the separate boxes.

Analysis of bacterial taxonomic categories of the microbiome associated with persistence/ progression vs. clearance revealed a total of 24 taxa that were significant (LDA>2.0) at V1. *G.*

vaginalis was the bacterial species with the highest positive correlation to progression (**Fig 4A**), while *L. iners* was the most positively associated taxon with clearance based on relative abundance. Amongst V2 samples there were a total of 13 significant taxa identified (three taxa associated with clearance and 10 with progression) (**Fig 4B**). At V2, bacteria that are commonly associated with bacterial vaginosis, such as *Prevotella amnii* and *Anaerococcus prevotii*, were significantly correlated with progression. We used a generalized linear model (GLM) to validate LEfSe biomarkers with adjustments for key covariates (age, smoking status, HPV16 and visit CST) (**S1 Table**).

Fungal biomarker discovery revealed five fungal taxa associated with HPV progression (**S1A Fig**). The average of the five combined fungal taxa were detected at rates of 0.59%, 3.55% and 3.76% for the clearance, persistence and progression outcomes, respectively (p = 0.0080, **S1B Fig**). However in the adjusted GLM analysis, none of the fungal biomarkers were determined to be significant at either visit (**S2 Table**).

To evaluate whether some common function of bacteria might be associated with HR-HPV outcomes, we used a functional analysis of gene groups imputed from the microbiome data as described in the methods. This analysis revealed 8 functional pathways at KEGG Level 2 significantly associated with progression to CIN2+ (**S3 Table**). Of the 8 identified pathways, only 2 had a mean read coverage of >1% and were considered for further analysis. The two identified pathways were "Xenobiotics Biodegradation and Metabolism" pathway, which was positively associated with progression (p = 0.0020) and the Cell Motility pathway, which was negatively associated with progression (p = 0.019). These two pathways were significantly correlated (Pearson correlation = -0.80, **S2 Fig**), and we chose to use Cell Motility in multivariate modeling since it produced a more stable GLM estimate (**S3 Table**).

Microbiome and HR-HPV natural history: GLM modeling

To investigate the contribution of components of the microbiome over time, we used a GLM in order to adjust for known covariates of CIN2+ progression that may influence the relationship of the CVM and progression to CIN2+ (e.g. age, smoking, HPV16 and CST). We utilized a GLM with a binary outcome (clearance/progression) and the significant microbial features identified in preceding sections as predictors. Specifically, we used the abundance of *Gardnerella* at V1, the abundance of *Lactobacillus* at V1, the Observed fungal species diversity at V1, the imputed Cell Motility pathway at V1 and the microbial diversity at V2. The model was adjusted for age, CST, smoking and HPV16 infection status. **Fig 5** shows the model estimates of the resulting GLM analysis. The multivariate analysis revealed a significant protective effect of V1 *Lactobacillus* (genus) abundance, OR = 0.41 (0.22–0.79), V1 fungal species diversity, OR = 0.90 (0.82–1.00) and imputed V1 Cell Motility pathway OR = 0.75 (0.62–0.92). In addition, the model showed that the V2 microbial diversity was a significant risk factor for CIN2 + progression, OR = 1.17 (1.02–1.29).

Following the multivariate analysis, we wanted to explore the reason for V1 *Gardnerella* being insignificant despite being the top microbial risk factor in differential abundance in all previous analyses (**Fig 4**, **S3 Fig** and **S1 Table**). Thus, we performed a mediation analysis to determine if V1 *Gardnerella* was acting by inducing the elevated diversity at V2 (**Fig 6A and 6B**). This analysis showed that after adjustment for V1 *Gardnerella* there was a significant association of V2 Shannon diversity with progression to CIN2+, p = 0.04. The Average Direct Effect (ADE) showed that V1 *Gardnerella* wasn't significant after adjustment for the V2 Shannon diversity, p = 0.23 supporting our mediation hypothesis (**Fig 6A**).

Power for detection of the effects of each microbiome component was performed using the *lmSupport* package. Specifically each of the microbiome components (i.e., V1 *Gardnerella*, V1

A. V1 Microbial Biomarkers
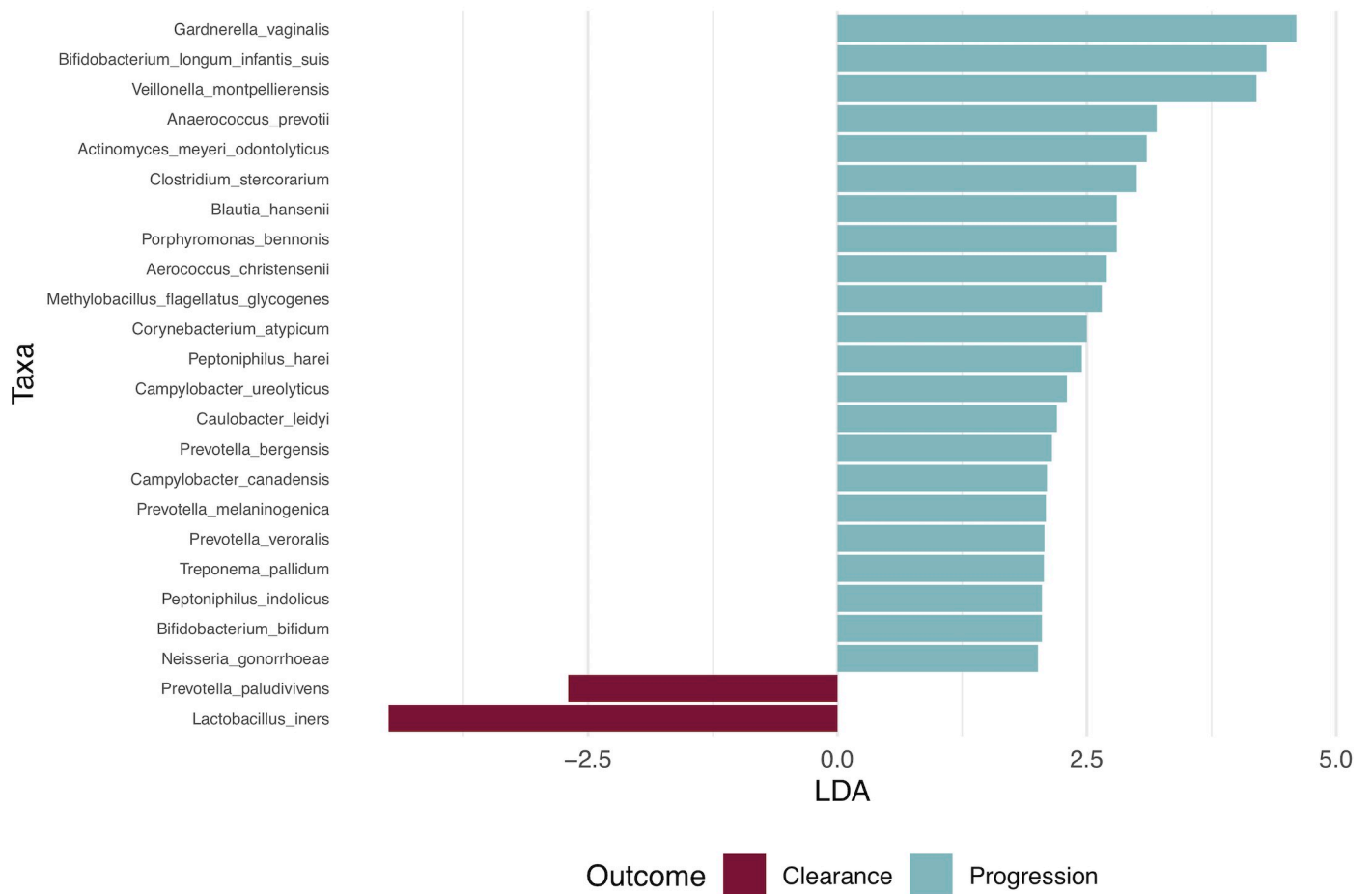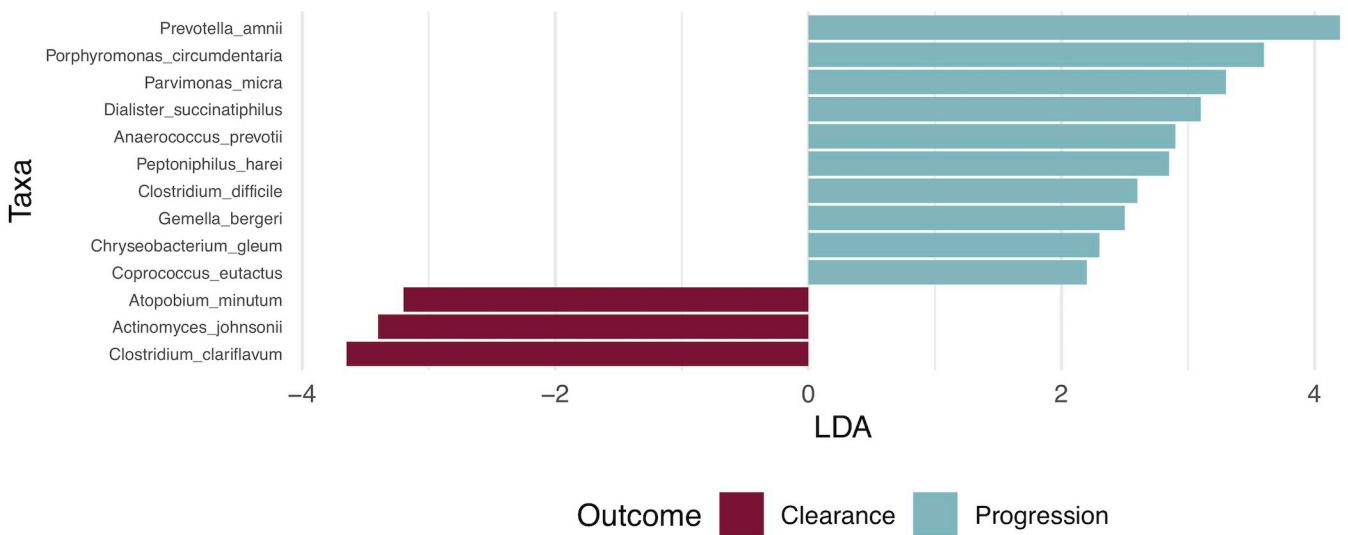

B. V2 Microbial Biomarkers

Fig 4. Bacteria associated with progression to CIN2+ identified using LEfSe. HR-HPV bacterial biomarkers for visit 1 (panel A) and visit 2 (panel B), comparing clearance vs. progression, were identified using the LEfSe tool. Only the significant bacterial taxa (LDA>2.0) are shown for both visits.
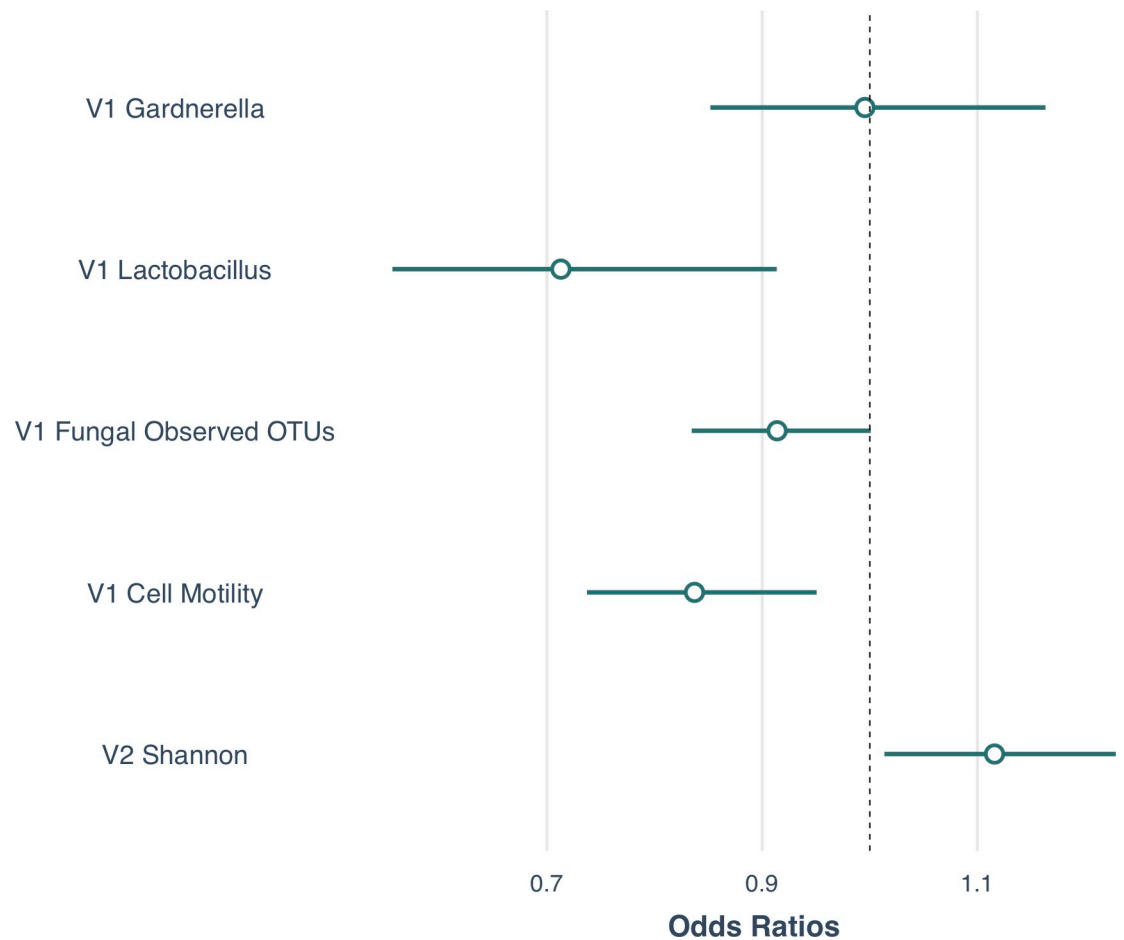
https://doi.org/10.1371/journal.ppat.1008376.g004

Fig 5. Generalized Linear Model (GLM) results showing the odds ratios of key microbial components in association with progression to CIN2+. The forest plot shows the results of variables evaluated in the univariate analysis that were then entered into a GLM. The model shows ORs (small circle) and 95% confidence interval (line extending on either side of the circle) of the microbial features associated with clearance/progression at either Visit 1 (V1) and/or Visit 2 (V2). The main variables included V1 *Gardnerella*, V1 *Lactobacillus*, V1 Fungal Observed OTUs and V1 Cell Motility and V2 Shannon diversity. The model was adjusted for age, bacterial CSTs, smoking and HPV16 infection status. 95% CIs that did not cross the Odds Ratio of 1.0 (dotted vertical line) are considered statistically significant.

https://doi.org/10.1371/journal.ppat.1008376.g005

Lactobacillus, V1 Fungal Observed OTUs, V1 Cell Motility, and V2 Shannon diversity) was tested separately to assess power with adjustment for age, CST, HPV16 status and smoking. **S4 Table** shows the results of the power calculation. Given large effect sizes for V1 *Lactobacillus*, V1 Fungal Observed OTUs, V1 Cell Motility, and V2 Shannon diversity we calculated that these could be detected with a power >98%, while the V1 *Gardnerella* had a power of 82% at the 0.05 alpha level.

Discussion

Previous cross-sectional studies analyzing the association between the cervicovaginal microbiome and HPV infection outcomes have consistently identified *Gardnerella* as a key biomarker associated with CIN2+. This has been reported in studies that utilized both next-generation sequencing [39, 41] and other methods of microbiome analyses [52, 43]. We present data that *Gardnerella* is in fact associated with CIN2+ lesions, but rather than directly causing the CIN2+ lesion, appears to induce a higher diversity CVM over time as measured at V2,

A. Mediation Analysis: V1 *G. vaginalis* and V2 Shannon

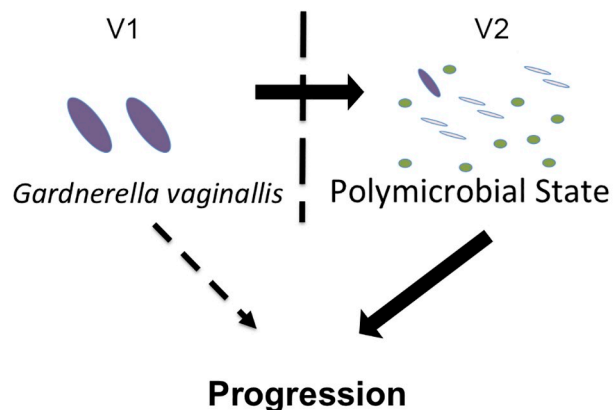|	**Estimate**	**95% CI**	**p-value**
ACME	0.19	0.0074 - 0.54	0.04
ADE	0.49	-0.27 - 1.12	0.23
Total Effect	0.68	0.018 - 1.3	0.046
Prop. Mediated	0.28	-0.10 - 2.2	0.086

B. Proposed Model of HR-HPV Progression

Fig 6. Diversity model for HPV progression with mediation analysis. Panel A shows the results of the mediation analysis that focus on V1 *Gardnerella* and V2 Shannon diversity. Top row shows Average Causal Mediation Effect (ACME) which is the full mediation effect of V2 Shannon diversity after adjusting for the direct effect of V1 *Gardnerella* on case status. The second row shows Average Direct Effect (ADE) which is the direct effect of *Gardnerella* on the clearance/progression outcome after accounting for the mediation effect of V2 Shannon diversity. The third row shows the Total Effect which is the direct, unadjusted effect, of *Gardnerella* on case outcome. The last row shows the Proportion (Prop.) Mediated, which is the proportion of the model that is mediated by V2 Shannon diversity. Based on GLM modeling, we propose the above model (Panel B) in which V1 *Gardnerella* causes an expansion of bacterial diversity at V2, which acts as a risk factor for the progression of a HR_HPV infection into a CIN2+ lesion.

which in turn mediates the observed effect of *Gardnerella* in HR-HPV disease progression. Although it is not clear how a state of polymicrobialism in the presence of a persistent HR-HPV infection leads to the development of epithelial dysplasia, recent studies on the microbiome's role in other cancers suggests that the answer lies in the establishment of a microbial microenvironment, perhaps a biofilm. For instance, it has been shown that certain cancers (e.g., colorectal cancer and prostate cancer) have distinct microbial communities at the tumor site that are associated with tumor development [44, 45]. In fact, other data from our lab using cervical biopsy tissue samples indicates that there are distinct microbial differences as cervical cancer progresses to advanced FIGO stages (manuscript in preparation). This idea is further supported by data indicating that cervical precancerous lesions that regress, compared to those that progress to cancer, harbor a different immune microenvironment [46]. The local interplay between the microbiome and the local host immune system may be important to understanding the progression of HR-HPV infection to cervical cancer.

The protective microbial biomarkers identified at V1 also suggest an association of the microbiome and host innate and acquired immunity in progression to CIN2+. Specifically, the

protective effect of bacterial Cell Motility may be due to the known phenomena of bacterial flagella activating host immunity [47–49]. This local activation may facilitate the innate immune system's ability to clear an active HPV infection. Such stimulation may be critical in HPV control since cervical lesions have been shown to be associated with local immunosuppression through the reduction of factors such as IL-17 [50]. Despite the presence of studies to support this conjecture it should be noted that this type of immune activation needs to be confirmed with rigorous experimental precision.

Gardnerella, as discussed above, continuously emerges as a risk factor for CIN2+ development and progression. Based on our findings and published data, the association may be tied to the ability of *Gardnerella* to be immunosuppressive in the cervicovaginal region [51]. Whereas, it seems that the presence of commensal bacteria (e.g. *Lactobacillus*) with the ability to stimulate a local immune response may be contributing factors to the clearance of incident HR-HPV infections. Moreover, the presence of bacteria with immunosuppressive attributes, such as *Gardnerella*, may promote viral persistence and progression.

Alternatively, there may be other explanations for the observed associations between the cervicovaginal microbiome and HPV's natural history. One possible explanation is a host developed or inherited immune deficiency that is a common cofactor for both cervical cancer progression and microbial diversity. For example, elevation of a particular inflammatory cytokine may be both necessary for successful tumor growth and be a causal factor in increasing vaginal microbial diversity. Such a factor may also explain the consistent identification of *Gardnerella*, which is commonly identified as a biomarker for increased diversity in the CVM [52] and a risk factor for CIN2+.

We have identified distinct microbial biomarkers that either protect, or promote the progression of a HR-HPV infection to CIN2+ lesions. In the context of what is known about the cervicovaginal microbiome, it may be that these factors act to suppress (in the case of progression) or activate (in the case of clearance) a localized immune response, which in turn influences the natural history of a HR-HPV infection. However, additional prospective studies are needed to establish a causal link between the cervicovaginal microbiome, the immune system and the natural history of HPV. Nevertheless, our results suggest a marker for identifying women with persistent HR-HPV infection at risk for progression by monitoring the presence of *Gardnerella* and subsequent elevation in microbial diversity. If future studies support a causal role of the cervicovaginal microbiome and disease progression, then it might be possible to manipulate the CVM in a manner to activate a local immune response. It is possible that HPV vaccination might influence the CVM and future research will be needed to evaluate such potential changes.

The strength of this study includes the prospective design and availability of a longitudinal cohort. In addition we used advanced epidemiological methods in a novel way to investigate potential causative factors in cervical intraepithelial neoplasia. Potential weaknesses in this study include the relatively small sample size, homogeneity of the population and the use of only two time points.

In summary, through the use of longitudinal samples from the CVT cohort we investigated and identified key features of the cervicovaginal microbiome potentially associated with progression of HR-HPV infection [28, 53–56] (e.g., *Gardnerella* and subsequent increase vaginal microbial diversity). Additional studies are required to validate the model proposed in this report.

Materials and methods

Clinical trial information

The study of cervicovaginal microbiome and HR-HPV natural history was a nested analysis within the previously reported CVT [57] (clinical trials registration NCT00128661). Written

informed consent was obtained from all participants in CVT. The trial protocol can be obtained from the original trial publication [57].

Ethics statement

All CVT participants were adult women between the ages of 18–25 years. All participants were shown a video describing the study design and were then required to provide written consent to continue participating in the trial. Institutional review board approval was obtained for the informed consent forms at both the NCI and in Costa Rica. Registered with Clinicaltrials.gov NCT00128661

Study population and case definitions

Subjects for this nested study were selected from the placebo arm of a community-based clinical trial of the HPV 16/18 vaccine in Costa Rica that had enrolled women 18 to 25 years of age in 2004–2005 [57]. Women with an incident HR-HPV infection (HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, or 59) were selected for analysis. Incident infections were classified based on outcomes from the well-established model of HR-HPV natural history including outcomes of clearance, persistence and progression [58]. Specifically, outcomes related to the incident HR-HPV infection included women who developed a CIN2 or CIN3 (CIN2+) lesion (progression), women with an infection for 2 or more years with the same type in the absence of a CIN2+ diagnosis (persistent), or women who cleared their incident HR-HPV infections within 1 year (clearance). This analysis included 273 women of whom all had available samples at V1 (first visit positive for the studied HPV type) and 266 who had a second sample at V2 (for persistent, visit that was positive for the same type and at least 305 days after V1; progression, closest visit before diagnosis of CIN2+; clearance, following visit that was negative for that type); all had clinical follow-up data. Seven women, one with clearance, and six with persistence either did not have an available V2 sample or the sample failed in lab testing.

Cervical microbiome characterization

DNA samples [59] were shipped to the Burk Lab on dry ice where the microbiome analysis was performed. DNA had been extracted from cervical brush samples by DDL Diagnostic Laboratory (Voorberg, The Netherlands) where they had been tested for HPV as previously described [60]. Laboratory procedures for the microbiome analyses were performed within a hood (AirClean Systems, Creedmoor, NC) in an isolated room to minimize environmental contamination and water-blank negative controls were used throughout the testing.

Bacterial DNA was amplified using barcoded-primers 16SV4_515F (GTGYCAGCMGC CGCGGTA) and 16SV4_806R (GGACTACHVGGGTWTCTAAT) that amplify the V4 variable region of the 16S rRNA gene [61]. This region has been demonstrated to accurately amplify and resolve vaginal bacteria [62]. PCR reactions were performed with 17.75 μl of nuclease-free PCR-grade water (Lonza, Rockland, ME), 2.5 μl of 10X Buffer w/ MgCl2 (Affymetrix, Santa Clara, CA), 1 μl of $MgCl_2$ (25 mM, Affymetrix, Santa Clara, California, USA), 0.5 μl of dNTPs (10 mM, Roche, Basel, Switzerland), 0.25 μl of AmpliTaq Gold DNA Polymerase (5 U/μl, Applied Biosystems, Foster City, California), 0.5 μl of HotStart-IT FideliTaq (2.5 U/μl, Affymetrix, Santa Clara, CA), 1 μl of each primer (5 μM), and 0.5 μl of sample DNA. Thermal cycling conditions consisted of initial denaturation at 95˚C for 5 min, followed by 15 cycles at 95˚C for 1 min, 55˚C for 1 min, and 68˚C for 1 min, followed by 15 cycles at 95˚C for 1 min, 60˚C for 1 min, and 68˚C for 1 min, and a final extension for 10 min at 68˚C on a GeneAmp PCR System 9700 (Applied Biosystems, Foster City, CA).

The fungal DNA ITS1 region was amplified using barcoded-primers ITS1_48F (ACACAC CGCCCGTCGCTACT) and ITS1_217R (TTTCGCTGCGTTCTTCATCG) as previously described [63]. PCR reactions were performed with 8.25 μl of nuclease-free PCR-grade water (Lonza), 2.5 μl of 10X Buffer w/ MgCl$_2$ (Affymetrix), 1 μl of MgCl$_2$ (25 mM, Affymetrix), 0.5 μl of dNTPs (10 mM, Roche), 0.25 μl of AmpliTaq Gold DNA Polymerase (5 U/μl, Applied Bio-systems), 0.5 μl of HotStart-IT FideliTaq (2.5 U/μl, Affymetrix), 1μl of each primer (5 μM), and 10 μl of sample DNA. Thermal cycling conditions consisted of initial denaturation of 95˚C for 3 min, followed by 35 cycles at 95˚C for 30 sec, 55˚C for 30 sec, and 68˚C for 2 min, followed by a final extension for 10 min at 68˚C on a GeneAmp PCR System 9700 (Applied Biosystems).

For both amplicon experiments, 20 negative controls were randomly mixed amongst samples. Negative controls were created using nuclease-free PCR-grade water (Lonza) as described above instead of extracted DNA.

Barcoded-PCR products were combined for each amplicon type and the DNA fragments (~356 bp for 16S V4 and 400–600 for ITS1) were isolated by gel purification using a QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany). Purified PCR products were quantified using a Qubit 2.0 Fluorometric High Sensitivity dsDNA Assay (Life Technologies, Carlsbad, CA) prior to library construction using a KAPA LTP Library Preparation Kit (Kapa Biosystems, Wilmington, MA). Size integrity of the amplicons was validated with a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). High-throughput amplicon sequencing of 2x300 paired-end reads was conducted on an Illumina MiSeq (Illumina, San Diego, CA).

Bioinformatics

Illumina reads were trimmed to remove bases that had a PHRED score of <25 using prinseq-lite V0.0.4 [64]. Quality trimmed reads were then demultiplexed using Novobarcode [65]. Paired-end reads were joined using PANDAseq with default settings [66]. The merged reads were processed through the VSEARCH quality-filtering pipeline [67] to dereplicate the sequences, reduce noise and remove chimeric reads.

For bacterial 16S V4 rRNA gene reads, closed-reference OTU picking was performed using VSEARCH [67] with a custom database that contained sequences from the GreenGenes database [68] the Human Oral Microbiome Database (HOMD) [69] and cervicovaginal microbiome 16S reference sequences retrieved from NCBI [70]. Representative sequences were aligned using PyNAST [71] and taxonomy was assigned using VSEARCH [67].

PICRUSt was used to impute microbial functional gene content and to collapse identified genes into functional pathways [72]. Pathways that were associated with HR-HPV clearance were identified through the use of a generalized linear model (GLM) based on statistical significance (p<0.05) and relative abundance (1% or higher across all reads).

For fungal ITS1 reads, open-reference OTU picking was performed using VSEARCH [67] and the UNITE database [73]. Taxonomy of representative fungal sequences was assigned using BLAST [74].

Phyloseq [75] was used to import BIOM data for 16S and ITS assays into R separately, followed by the determination of Shannon and Chao1 alpha diversity. For all analyses, bacterial data was subsampled for 2,500 reads. For fungal analyses subsampling was performed at 500 reads. Biomarker discovery analysis was performed using the LEfSe tool [76]. Linear discriminant analysis (LDA) scores greater than 2.0 are considered to be significant [76]. Microbial community state types (CSTs) were assigned on the basis of hierarchical clustering of the 20 most abundant OTUs. Prior to clustering, OTUs were agglomerated at the species level or the

lowest identified taxonomic level. Clustering was performed using the wardD2 algorithm using Euclidian distances.

Statistical analysis

R v3.4.2 [77] was used for statistical analyses. The Kruskal-Wallis test was used to assess significance of continuous data. Linear regression was used to assess the significance of variables associated with the ordered outcome states of a HR-HPV infection (1). Logistic regression was performed using the GLM function and a binomial family generalized linear model in R. For categorical data, dummy variables were created and each individual factor level was tested in a univariate GLM analysis. Models were adjusted for age, smoking, HPV16 and CSTs. Smoking status was determined through a questionnaire and incorporated into the model as ordered categories: never smoked, former smoker and current smoker [40]. Power of GLM results was computed using the *lmSupport* package [78].

We performed a statistical mediation analysis to test whether V1 *Gardnerella* (an independent variable) could be acting by inducing a subsequent elevated microbiome diversity at V2 (mediator variable) that influences the outcome of HR-HPV progression using the package *mediation* [79]. The outcome model we used was binary clearance/progression. Models were adjusted for age, CST, smoking status and HPV16 infection status. In the results we present the mediation effect (average causal mediation effects (ACME)), which is the total effect of V2 Shannon diversity and V1 *Gardnerella* minus the direct effect of V1 *Gardnerella*. Additionally, we estimate the direct of effect (presented using the average direct effect (ADE)) of V1 *Gardnerella* on the binary outcome clearance/progression, minus the effect of the V2 Shannon diversity mediator; the total effect, which is the sum between the indirect effect of the V2 Shannon diversity and the direct effect of the V1 *Gardnerella*; and the proportion mediated which is the ratio of the ACME and total effect estimates.

Investigators in the International Agency for Research on Cancer/World Health Organization

Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer / World Health Organization.

Investigators in the Costa Rica HPV Vaccine Trial (CVT) group

Proyecto Epidemiológico Guanacaste, Fundación INCIENSA, San José, Costa Rica—Bernal Cortés (specimen and repository manager), Paula González (LTFU: co-principal investigator), Rolando Herrero (CVT: co-principal investigator), Silvia E. Jiménez (trial coordinator), Carolina Porras (co-investigator), Ana Cecilia Rodríguez (co-investigator).

United States National Cancer Institute, Bethesda, MD, USA—Allan Hildesheim (co-principal investigator & NCI co-project officer), Aimée R. Kreimer (LTFU: co-principal investigator & NCI co-project officer), Douglas R. Lowy (HPV virologist), Mark Schiffman (CVT: medical monitor & NCI co-project officer), John T. Schiller (HPV virologist), Mark Sherman (CVT: QC pathologist), Sholom Wacholder (statistician).

Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, Frederick, MD, USA (HPV Immunology Laboratory)—Ligia A. Pinto, Troy J. Kemp

Georgetown University, Washington, DC, USA—Mary K. Sidawy (CVT: histopathologist)

DDL Diagnostic Laboratory, Netherlands (HPV DNA Testing)—Wim Quint, Leen-Jan van Doorn, Linda Struijk.

University of California, San Francisco, CA, USA—Joel M. Palefsky (expert on anal HPV infection and disease diagnosis and management), Teresa M. Darragh (pathologist and clinical management)

University of Virginia, Charlottesville, VA, USA—Mark H. Stoler (QC pathologist)

Supporting information

S1 Table. Bacterial GLM confirmation.
(CSV)

S2 Table. Fungal GLM confirmation.
(CSV)

S3 Table. PICRUSt functional pathways.
(CSV)

S4 Table. Power calculation.
(XLSX)

S1 Fig. Fungal taxa associated with HPV natural history. Panel A shows specific fungal taxa, identified as being significant with the three HR-HPV outcomes. Values that are higher than LDA score of 2.0 are considered to be significant. Panel B shows the main fungal taxa identified in panel A with their relative abundances. The box represents the median value (with the 25–75% confidence interval as the box and the 95% confidence interval with the whiskers) for the taxa in each outcome (shown in the three separate panels and labeled at the top of the panel). There is a statistically significant increase in the sum of the five progression associated taxa when going from clearance to persistence to progression, p = 0.0080. The y-axis is the log of the relative abundance.
(PDF)

S2 Fig. Correlation between cell motility and xenobiotic metabolism. The two pathways identified after PICRUSt that were significantly different between clearance and progression were analyzed to determine their correlation to each other. Plot shows a significant negative correlation (Pearson correlation = -0.80) between the two pathways and thus they are highly correlated.
(PDF)

S3 Fig. 16S bacterial Shannon diversity index and *Gardnerella* across categorical cytology groups. Panel A shows the 16S bacterial Shannon diversity index in each cytology group. Panel B shows the relative abundance of *Gardnerella* across the cytology groups. Significance is shown above bar plots as indicated in the figure at the top right.
(PDF)

Acknowledgments

Author Contributions

Conceptualization: Mykhaylo Usyk, Carolina Porras, Paula Gonzalez, Robert D. Burk.

Data curation: Mykhaylo Usyk, Mark Schiffman, Robert D. Burk.

Formal analysis: Mykhaylo Usyk.

Funding acquisition: Robert D. Burk.

Investigation: Mykhaylo Usyk, Robert D. Burk.

Methodology: Mykhaylo Usyk, Christine P. Zolnik, Carolina Porras, Ana Gradissimo, Mahboobeh Safaeian, Mark Schiffman, Robert D. Burk.

Project administration: Robert D. Burk.

Resources: Mykhaylo Usyk, Robert D. Burk.

Software: Mykhaylo Usyk.

Supervision: Robert D. Burk.

Validation: Mykhaylo Usyk.

Visualization: Mykhaylo Usyk.

Writing – original draft: Mykhaylo Usyk, Robert D. Burk.

Writing – review & editing: Mykhaylo Usyk, Christine P. Zolnik, Philip E. Castle, Carolina Porras, Rolando Herrero, Ana Gradissimo, Paula Gonzalez, Mahboobeh Safaeian, Mark Schiffman, Robert D. Burk.

References

1. Schiffman M, Doorbar J, Wentzensen N, de Sanjose S, Fakhry C, Monk BJ, et al. Carcinogenic human papillomavirus infection. Nature reviews Disease primers. 2016; 2:16086. Epub 2016/12/03. https://doi.org/10.1038/nrdp.2016.86 PMID: 27905473.

2. Ho GY, Burk RD, Klein S, Kadish AS, Chang CJ, Palan P, et al. Persistent genital human papillomavirus infection as a risk factor for persistent cervical dysplasia. J Natl Cancer Inst. 1995; 87(18):1365–71. https://doi.org/10.1093/jnci/87.18.1365 PMID: 7658497.

3. Castle PE, Rodriguez AC, Burk RD, Herrero R, Wacholder S, Alfaro M, et al. Short term persistence of human papillomavirus and risk of cervical precancer and cancer: population based cohort study. BMJ. 2009; 339:b2569. Epub 2009/07/30. https://doi.org/10.1136/bmj.b2569 PMID: 19638649; PubMed Central PMCID: PMC2718087.

4. Kjaer SK, Frederiksen K, Munk C, Iftner T. Long-term absolute risk of cervical intraepithelial neoplasia grade 3 or worse following human papillomavirus infection: role of persistence. J Natl Cancer Inst. 2010; 102(19):1478–88. Epub 2010/09/16. https://doi.org/10.1093/jnci/djq356 PMID: 20841605; PubMed Central PMCID: PMC2950170.

5. Ho, G. Y., Burk, R. D., Klein, S., Kadish, A. S., Chang, C. J., Palan, P., ... & Romney, S. (1995). Persistent genital human papillomavirus infection as a risk factor for persistent cervical dysplasia. JNCI: Journal of the National Cancer Institute, 87(18), 1365–1371.

6. Castle PE, Wacholder S, Lorincz AT, Scott DR, Sherman ME, Glass AG, et al. A prospective study of high-grade cervical neoplasia risk among human papillomavirus-infected women. J Natl Cancer Inst. 2002; 94(18):1406–14. Epub 2002/09/19. https://doi.org/10.1093/jnci/94.18.1406 PMID: 12237286.

7. Roura E, Castellsague X, Pawlita M, Travier N, Waterboer T, Margall N, et al. Smoking as a major risk factor for cervical cancer and pre-cancer: results from the EPIC cohort. Int J Cancer. 2014; 135(2):453–66. https://doi.org/10.1002/ijc.28666 PMID: 24338632.

8. Plummer M, Herrero R, Franceschi S, Meijer CJ, Snijders P, Bosch FX, et al. Smoking and cervical cancer: pooled analysis of the IARC multi-centric case—control study. Cancer Causes Control. 2003; 14 (9):805–14. https://doi.org/10.1023/b:caco.0000003811.98261.3e PMID: 14682438.

9. Eldridge RC, Pawlita M, Wilson L, Castle PE, Waterboer T, Gravitt PE, et al. Smoking and subsequent human papillomavirus infection: a mediation analysis. Ann Epidemiol. 2017. Epub 2017/11/07. https://doi.org/10.1016/j.annepidem.2017.10.004 PMID: 29107447.

10. Smith JS, Green J, Berrington de Gonzalez A, Appleby P, Peto J, Plummer M, et al. Cervical cancer and use of hormonal contraceptives: a systematic review. Lancet. 2003; 361(9364):1159–67. Epub 2003/04/11. https://doi.org/10.1016/s0140-6736(03)12949-2 PMID: 12686037.

11. Moreno V, Bosch FX, Munoz N, Meijer CJ, Shah KV, Walboomers JM, et al. Effect of oral contraceptives on risk of cervical cancer in women with human papillomavirus infection: the IARC multicentric case-control study. Lancet. 2002; 359(9312):1085–92. https://doi.org/10.1016/S0140-6736(02)08150-3 PMID: 11943255.

12. Munoz N, Franceschi S, Bosetti C, Moreno V, Herrero R, Smith JS, et al. Role of parity and human papillomavirus in cervical cancer: the IARC multicentric case-control study. Lancet. 2002; 359(9312):1093–101. https://doi.org/10.1016/S0140-6736(02)08151-5 PMID: 11943256.

13. Stanley MA, Sterling JC. Host Responses to Infection with Human Papillomavirus. Curr Probl Dermatol. 2014; 45:58–74. Epub 2014/03/20. https://doi.org/10.1159/000355964 PMID: 24643178.

14. Doorbar J. Host control of human papillomavirus infection and disease. Best practice & research Clinical obstetrics & gynaecology. 2018; 47:27–41. Epub 2017/09/19. https://doi.org/10.1016/j.bpobgyn.2017.08.001 PMID: 28919159.

15. Leo PJ, Madeleine MM, Wang S, Schwartz SM, Newell F, Pettersson-Kymmer U, et al. Defining the genetic susceptibility to cervical neoplasia-A genome-wide association study. PLoS genetics. 2017; 13 (8):e1006866. https://doi.org/10.1371/journal.pgen.1006866 PMID: 28806749; PubMed Central PMCID: PMC5570502.

16. Habbous S, Pang V, Xu W, Amir E, Liu G. Human papillomavirus and host genetic polymorphisms in carcinogenesis: A systematic review and meta-analysis. J Clin Virol. 2014. https://doi.org/10.1016/j.jcv.2014.07.019 PMID: 25174543.

17. Castle PE, Giuliano AR. Chapter 4: Genital tract infections, cervical inflammation, and antioxidant nutrients —assessing their roles as human papillomavirus cofactors. J Natl Cancer Inst Monogr. 2003;(31):29–34. Epub 2003/06/17. https://doi.org/10.1093/oxfordjournals.jncimonographs.a003478 PMID: 12807942.

18. Pfeiffer JK. Host response: Microbiota prime antiviral response. Nature microbiology. 2016; 1(2):15029.

19. Vallianou NG, Tzortzatou-Stathopoulou F. Microbiota and cancer: an update. J Chemother. 2019; 31 (2):59–63. https://doi.org/10.1080/1120009X.2018.1541046 PMID: 30646834

20. Rajagopala SV, Vashee S, Oldfield LM, Suzuki Y, Venter JC, Telenti A, et al. The human microbiome and cancer. Cancer Prevention Research. 2017; 10(4):226–34. https://doi.org/10.1158/1940-6207.CAPR-16-0249 PMID: 28096237

21. Vogtmann E, Hua X, Zeller G, Sunagawa S, Voigt AY, Hercog R, et al. Colorectal cancer and the human gut microbiome: reproducibility with whole-genome shotgun sequencing. PloS one. 2016; 11(5): e0155362. https://doi.org/10.1371/journal.pone.0155362 PMID: 27171425

22. Champer M, Wong AM, Champer J, Brito IL, Messer PW, Hou JY, et al. The role of the vaginal microbiome in gynecological cancer: a review. BJOG: an international journal of obstetrics and gynaecology. 2017. Epub 2017/03/10. https://doi.org/10.1111/1471-0528.14631 PMID: 28278350.

23. Kroon SJ, Ravel J, Huston WM. Cervicovaginal microbiota, women's health, and reproductive outcomes. Fertil Steril. 2018; 110(3):327–36. https://doi.org/10.1016/j.fertnstert.2018.06.036 PMID: 30098679

24. Kyrgiou M, Mitra A, Moscicki A-B. Does the vaginal microbiota play a role in the development of cervical cancer? Translational Research. 2017; 179:168–82. https://doi.org/10.1016/j.trsl.2016.07.004 PMID: 27477083

25. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, McCulle SL, et al. Vaginal microbiome of reproductive-age women. Proc Natl Acad Sci U S A. 2011; 108 Suppl 1:4680–7. Epub 2010/06/11. https://doi.org/10.1073/pnas.1002611107 PMID: 20534435; PubMed Central PMCID: PMC3063603.

26. Smith BC, Zolnik CP, Usyk M, Chen Z, Kaiser K, Nucci-Sack A, et al. Focus: Microbiome: Distinct Ecological Niche of Anal, Oral, and Cervical Mucosal Microbiomes in Adolescent Women. The Yale journal of biology and medicine. 2016; 89(3):277. PMID: 27698612

27. Bik EM, Bird SW, Bustamante JP, Leon LE, Nieto PA, Addae K, et al. A novel sequencing-based vaginal health assay combining self-sampling, HPV detection and genotyping, STI detection, and vaginal microbiome analysis. PloS one. 2019; 14(5):e0215945. https://doi.org/10.1371/journal.pone.0215945 PMID: 31042762

28. Freitas AC, Bocking A, Hill JE, Money DM. Increased richness and diversity of the vaginal microbiota and spontaneous preterm birth. Microbiome. 2018; 6(1):117. https://doi.org/10.1186/s40168-018-0502-8 PMID: 29954448

29. Di C, Jin F. Value of combined detection of claudin 4 and high-risk human papilloma virus in high-grade squamous intraepithelial lesion and cervix squamous cell carcinoma. Zhejiang da xue xue bao Yi xue ban = Journal of Zhejiang University Medical sciences. 2018; 47(4):344–50. PMID: 30511519

30. Audirac-Chalifour A, Torres-Poveda K, Bahena-Román M, Téllez-Sosa J, Martínez-Barnetche J, Cortina-Ceballos B, et al. Cervical microbiome and cytokine profile at various stages of cervical cancer: a pilot study. PloS one. 2016; 11(4):e0153274. https://doi.org/10.1371/journal.pone.0153274 PMID: 27115350

31. Dareng E, Ma B, Famooto A, Akarolo-Anthony S, Offiong R, Olaniyan O, et al. Prevalent high-risk HPV infection and vaginal microbiota in Nigerian women. Epidemiol Infect. 2016; 144(1):123–37. https://doi.org/10.1017/S0950268815000965 PMID: 26062721

32. Gao W, Weng J, Gao Y, Chen X. Comparison of the vaginal microbiota diversity of women with and without human papillomavirus infection: a cross-sectional study. BMC infectious diseases. 2013; 13(1):271.

33. Lee JE, Lee S, Lee H, Song Y-M, Lee K, Han MJ, et al. Association of the vaginal microbiota with human papillomavirus infection in a Korean twin cohort. PloS one. 2013; 8(5):e63514. https://doi.org/10.1371/journal.pone.0063514 PMID: 23717441

34. Reimers LL, Mehta SD, Massad LS, Burk RD, Xie X, Ravel J, et al. The cervicovaginal microbiota and its associations with human papillomavirus detection in HIV-infected and HIV-uninfected women. The Journal of infectious diseases. 2016; 214(9):1361–9. https://doi.org/10.1093/infdis/jiw374 PMID: 27521363

35. Audirac-Chalifour A, Torres-Poveda K, Bahena-Roman M, Tellez-Sosa J, Martinez-Barnetche J, Cortina-Ceballos B, et al. Cervical Microbiome and Cytokine Profile at Various Stages of Cervical Cancer: A Pilot Study. PLoS One. 2016; 11(4):e0153274. https://doi.org/10.1371/journal.pone.0153274 PMID: 27115350

36. Palma E, Recine N, Domenici L, Giorgini M, Pierangeli A, Panici PB. Long-term Lactobacillus rhamnosus BMX 54 application to restore a balanced vaginal ecosystem: a promising solution against HPV-infection. BMC infectious diseases. 2018; 18(1):13. https://doi.org/10.1186/s12879-017-2938-z PMID: 29304768

37. Piyathilake CJ, Ollberding NJ, Kumar R, Macaluso M, Alvarez RD, Morrow CD. Cervical Microbiota Associated with Risk of Higher Grade Cervical Intraepithelial Neoplasia in Women Infected with High-Risk Human Papillomaviruses. Cancer Prev Res (Phila). 2016. Epub 2016/03/05. https://doi.org/10.1158/1940-6207.capr-15-0350 PMID: 26935422.

38. Mitra A, MacIntyre DA, Lee YS, Smith A, Marchesi JR, Lehne B, et al. Cervical intraepithelial neoplasia disease progression is associated with increased vaginal microbiome diversity. Scientific reports. 2015; 5:16865. https://doi.org/10.1038/srep16865 PMID: 26574055; PubMed Central PMCID: PMC4648063.

39. Seo SS, Oh HY, Lee JK, Kong JS, Lee DO, Kim MK. Combined effect of diet and cervical microbiome on the risk of cervical intraepithelial neoplasia. Clin Nutr. 2016. https://doi.org/10.1016/j.clnu.2016.03.019 PMID: 27075319.

40. Hildesheim A, Herrero R, Wacholder S, Rodriguez AC, Solomon D, Bratti MC, et al. Effect of human papillomavirus 16/18 L1 viruslike particle vaccine among young women with preexisting infection: a randomized trial. JAMA. 2007; 298(7):743–53. Epub 2007/08/21. https://doi.org/10.1001/jama.298.7.743 PMID: 17699008.

41. Zhang C, Liu Y, Gao W, Pan Y, Gao Y, Shen J, et al. The direct and indirect association of cervical microbiota with the risk of cervical intraepithelial neoplasia. Cancer medicine. 2018. Epub 2018/04/03. https://doi.org/10.1002/cam4.1471 PMID: 29608253.

42. Takač I, Marin J, Gorišek B. Human papillomavirus 16 and 18 infection of the uterine cervix in women with different grades of cervical intraepithelial neoplasia (CIN). International Journal of Gynecology & Obstetrics. 1998; 61(3):269–73.

43. Di Paola M, Sani C, Clemente AM, Iossa A, Perissi E, Castronovo G, et al. Characterization of cervico-vaginal microbiota in women developing persistent high-risk Human Papillomavirus infection. Scientific reports. 2017; 7(1):10200. https://doi.org/10.1038/s41598-017-09842-6 PMID: 28860468

44. Louis P, Hold GL, Flint HJ. The gut microbiota, bacterial metabolites and colorectal cancer. Nature Reviews Microbiology. 2014; 12(10):661. https://doi.org/10.1038/nrmicro3344 PMID: 25198138

45. Sfanos KS, Yegnasubramanian S, Nelson WG, De Marzo AM. The inflammatory microenvironment and microbiome in prostate cancer development. Nature Reviews Urology. 2018; 15(1):11. https://doi.org/10.1038/nrurol.2017.167 PMID: 29089606

46. Øvestad IT, Gudlaugsson E, Skaland I, Malpica A, Kruse A-J, Janssen EA, et al. Local immune response in the microenvironment of CIN2–3 with and without spontaneous regression. Mod Pathol. 2010; 23(9):1231. https://doi.org/10.1038/modpathol.2010.109 PMID: 20512116

47. Ramos HC, Rumbo M, Sirard J-C. Bacterial flagellins: mediators of pathogenicity and host immune responses in mucosa. Trends Microbiol. 2004; 12(11):509–17. https://doi.org/10.1016/j.tim.2004.09.002 PMID: 15488392

48. Vijay-Kumar M, Gewirtz A. Flagellin: key target of mucosal innate immunity. Mucosal immunology. 2009; 2(3):197. https://doi.org/10.1038/mi.2009.9 PMID: 19242410

49. Cullender TC, Chassaing B, Janzon A, Kumar K, Muller CE, Werner JJ, et al. Innate and adaptive immunity interact to quench microbiome flagellar motility in the gut. Cell host & microbe. 2013; 14(5):571–81.

50. Gosmann C, Mattarollo SR, Bridge JA, Frazer IH, Blumenthal A. IL-17 Suppresses Immune Effector Functions in Human Papillomavirus-Associated Epithelial Hyperplasia. J Immunol. 2014. https://doi.org/10.4049/jimmunol.1400216 PMID: 25063870.

51. Murphy K, Mitchell CM. The interplay of host immunity, environment and the risk of bacterial vaginosis and associated reproductive health outcomes. The Journal of infectious diseases. 2016; 214(suppl_1): S29–S35.

52. Schellenberg JJ, Patterson MH, Hill JE. Gardnerella vaginalis diversity and ecology in relation to vaginal symptoms. Res Microbiol. 2017; 168(9–10):837–44. https://doi.org/10.1016/j.resmic.2017.02.011 PMID: 28341009

53. Arokiyaraj S, Seo SS, Kwon M, Lee JK, Kim MK. Association of cervical microbial community with persistence, clearance and negativity of Human Papillomavirus in Korean women: a longitudinal study. Scientific reports. 2018; 8(1):15479. Epub 2018/10/21. https://doi.org/10.1038/s41598-018-33750-y PMID: 30341386.

54. Godoy-Vitorino F, Romaguera J, Zhao C, Vargas-Robles D, Ortiz-Morales G, Vazquez-Sanchez F, et al. Cervicovaginal Fungi and Bacteria Associated With Cervical Intraepithelial Neoplasia and High-Risk Human Papillomavirus Infections in a Hispanic Population. Front Microbiol. 2018; 9:2533. Epub 2018/11/09. https://doi.org/10.3389/fmicb.2018.02533 PMID: 30405584; PubMed Central PMCID: PMC6208322.

55. Kwasniewski W, Wolun-Cholewa M, Kotarski J, Warchol W, Kuzma D, Kwasniewska A, et al. Microbiota dysbiosis is associated with HPV-induced cervical carcinogenesis. Oncology letters. 2018; 16(6):7035–47. Epub 2018/12/14. https://doi.org/10.3892/ol.2018.9509 PMID: 30546437; PubMed Central PMCID: PMC6256731.

56. Suehiro TT, Malaguti N, Damke E, Uchimura NS, Gimenes F, Souza RP, et al. Association of human papillomavirus and bacterial vaginosis with increased risk of high-grade squamous intraepithelial cervical lesions. International journal of gynecological cancer: official journal of the International Gynecological Cancer Society. 2019. Epub 2019/01/12. https://doi.org/10.1136/ijgc-2018-000076 PMID: 30630884.

57. Herrero R, Hildesheim A, Rodríguez AC, Wacholder S, Bratti C, Solomon D, et al. Rationale and design of a community-based double-blind randomized clinical trial of an HPV 16 and 18 vaccine in Guanacaste, Costa Rica. Vaccine. 2008; 26(37):4795–808. https://doi.org/10.1016/j.vaccine.2008.07.002 PMID: 18640170

58. Schiffman M, Wentzensen N. Human papillomavirus infection and the multistage carcinogenesis of cervical cancer. Cancer Epidemiol Biomarkers Prev. 2013; 22(4):553–60. https://doi.org/10.1158/1055-9965.EPI-12-1406 PMID: 23549399.

59. Hildesheim A, Wacholder S, Catteau G, Struyf F, Dubin G, Herrero R, et al. Efficacy of the HPV-16/18 vaccine: final according to protocol results from the blinded phase of the randomized Costa Rica HPV-16/18 vaccine trial. Vaccine. 2014; 32(39):5087–97. https://doi.org/10.1016/j.vaccine.2014.06.038 PMID: 25018097

60. Herrero R, Wacholder S, Rodríguez AC, Solomon D, González P, Kreimer AR, et al. Prevention of persistent human papillomavirus infection by an HPV16/18 vaccine: a community-based randomized clinical trial in Guanacaste, Costa Rica. Cancer discovery. 2011; 1(5):408–19. https://doi.org/10.1158/2159-8290.CD-11-0131 PMID: 22586631

61. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. The ISME journal. 2012; 6 (8):1621. https://doi.org/10.1038/ismej.2012.8 PMID: 22402401

62. Van Der Pol WJ, Kumar R, Morrow CD, Blanchard EE, Taylor CM, Martin DH, et al. In silico and experimental evaluation of primer sets for species-level resolution of the vaginal microbiota using 16s ribosomal rna gene sequencing. The Journal of infectious diseases. 2018; 219(2):305–14.

63. Usyk M, Zolnik CP, Patel H, Levi MH, Burk RD. Novel ITS1 fungal primers for characterization of the mycobiome. mSphere. 2017; 2(6):e00488–17. https://doi.org/10.1128/mSphere.00488-17 PMID: 29242834

64. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 2011; 27(6):863–4. Epub 2011/02/01. https://doi.org/10.1093/bioinformatics/btr026 PMID: 21278185; PubMed Central PMCID: PMC3051327.

65. Hercus C. Novocraft short read alignment package. Website http://www.novocraft.com. 2009.

66. Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD. PANDAseq: paired-end assembler for illumina sequences. BMC Bioinformatics. 2012; 13:31. Epub 2012/02/16. https://doi.org/10.1186/1471-2105-13-31 PMID: 22333067; PubMed Central PMCID: PMC3471323.

67. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. PeerJ. 2016; 4:e2584. https://doi.org/10.7717/peerj.2584 PMID: 27781170

68. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol. 2006; 72(7):5069–72. https://doi.org/10.1128/AEM.03006-05 PMID: 16820507

69. Chen HS, Bromberg-White J, Conway MJ, Alam S, Meyers C. Study of infectious virus production from HPV18/16 capsid chimeras. Virology. 2010; 405(2):289–99. https://doi.org/10.1016/j.virol.2010.05.019 PMID: 20598725; PubMed Central PMCID: PMC2923236.

70. Coordinators NR. Database resources of the national center for biotechnology information. Nucleic acids research. 2017; 45(Database issue):D12. https://doi.org/10.1093/nar/gkw1071 PMID: 27899561

71. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nature methods. 2010; 7(5):335–6. https://doi.org/10.1038/nmeth.f.303 PMID: 20383131; PubMed Central PMCID: PMC3156573.

72. Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat Biotechnol. 2013; 31 (9):814. https://doi.org/10.1038/nbt.2676 PMID: 23975157

73. Kõljalg U, Larsson KH, Abarenkov K, Nilsson RH, Alexander IJ, Eberhardt U, et al. UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. New Phytologist. 2005; 166(3):1063–8. https://doi.org/10.1111/j.1469-8137.2005.01376.x PMID: 15869663

74. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of molecular biology. 1990; 215(3):403–10. https://doi.org/10.1016/S0022-2836(05)80360-2 PMID: 2231712

75. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PloS one. 2013; 8(4):e61217. https://doi.org/10.1371/journal.pone.0061217 PMID: 23630581

76. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. Genome biology. 2011; 12(6):R60. https://doi.org/10.1186/gb-2011-12-6-r60 PMID: 21702898

77. Team RC. R: A language and environment for statistical computing. 2013.

78. Curtin J. lmSupport: support for linear models. R package version. 2015; 2(2).

79. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. Mediation: R package for causal mediation analysis. 2014.