THE ROLE OF THE CERVICOVAGINAL MICROBIOME IN HUMAN PAPILLOMAVIRUS-ASSOCIATED CERVICAL CARCINOGENESIS: POTENTIAL VALUE FOR CLINICAL PRACTICE

MARGARET ANN LOGEL

DEPARTMENT OF EPIDEMIOLOGY, BIOSTATISTICS AND OCCUPATIONAL HEALTH

MCGILL UNIVERSITY

DECEMBER 2023

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science in Epidemiology

ABSTRACT

Background: Infection with high-risk human papillomavirus (hrHPV) is a necessary, but not sufficient, cause of cervical cancer and its precancerous lesion, cervical intraepithelial neoplasia (CIN). Most human papillomavirus (HPV) infections are transient; a small proportion persist and lead to cervical cancer. The paradigm for detection of cervical abnormalities is cytology, with HPV testing recently introduced for screening and risk prediction. Relevant to the latter is the role of the cervicovaginal microbiome (CVM). Evidence suggests that the CVM is implicated in HPV and carcinogenesis. However, research has been limited by small sample size studies and low taxonomic resolution.

Objectives: This thesis investigated the relationship between the CVM, HPV and CIN. The objectives were to: 1) conduct a review on the CVM in cervical cancer, 2) assess CVM composition, and 3) compare the diagnostic accuracy of the CVM, cytology, and HPV for CIN and hrHPV detection.

Methods: In manuscript 1, 3 databases were searched until July 27th, 2022. Eligible research articles discussed the CVM in HPV-associated cervical cancer, characterized the CVM via metagenomics and included a measure of association. Statistics, study design, population, and methodology were extracted and summarized. Manuscript 2 included 186 women [54 normal, 50 CIN1, 40 CIN2, 42 CIN3] referred for colposcopy following abnormal cytology. Samples were genotyped for hrHPV with the Roche cobas 4800 assay. The CVM was characterized with 16S rRNA gene sequencing of two regions (V3V4, V5V6) and bioinformatic processing via the high-resolution ANCHOR pipeline. Logistic regression models were constructed with 1) CVM species, 2) hrHPV, 3) cytology, and 4) CVM species and hrHPV as predictors and CIN2+ as the outcome. The coefficients were used to construct linear scores on CVM species, cytology, HPV, and CVM species/HPV. Species were selected via logistic regression with stepwise forward selection. Receiver operating characteristic curves were plotted, and the area under the curve (AUC) and 95% confidence intervals (CI) were compared to assess clinical performance (reported as AUC;95%CI).

Results: In manuscript 1, high CVM diversity and *Lactobacillus* depletion appear to increase and decrease the risk of adverse outcomes in HPV-associated cervical cancer, respectively. In manuscript 2, 77 species were identified; 8 unique to V3V4, 48 V5V6 and 21 shared. For CIN2+

detection, a score based on CVM species (0.64;0.57-0.71) performed the least accurately. However, a score that combined CVM species and HPV (0.80;0.74-0.86) performed similarly to cytology (0.84;0.79-0.90) and HPV (0.76;0.70-0.82).

Discussion: The CVM may be involved in cervical cancer; however, most 16S algorithms infer sequences from parametric error models, which modify sequence nucleotides resulting in low taxonomic resolution. By enhancing 16S sequencing by amplifying two regions and using ANCHOR, we found a correlation with CIN2+.

Conclusion: CVM species and HPV were correlated with CIN2+ in this cross-sectional study. A prospective cohort study could help determine a possible causal role of the CVM in cervical cancer and its diagnostic value.

RÉSUMÉ

Contexte: L'infection avec un des types haut-risque du virus du papillome humain (hrVPH) est une cause nécessaire, mais non suffisante, du cancer du col de l'utérus et de ses lésions précurseurs, les néoplasies intraépithéliales cervicales (CINs). La plupart des infections au virus du papillome humain (VPH) sont transitoires; une petite proportion persiste et mène au développement du cancer du col de l'utérus. La cytologie sert historiquement de référence absolue pour la détection des anormalités cervicaux. Plus récemment, le test VPH a été introduit comme test de dépistage et de prévision des risques. Le microbiome cervico-vaginale (CVM) pourrait jouer un rôle important au niveau de ce test en influant sur le VPH et sur la carcinogénèse. Cependant, la recherche antérieure est limitée aux études de petite taille et par une faible résolution taxonomique.

Objectifs: Ce mémoire a investigué le lien entre le CVM, le VPH et les CINs. Les objectifs étaient: 1) de mener une revue sur le CVM et le cancer du col de l'utérus, 2) d'évaluer la composition du CVM, et 3) de comparer la précision diagnostique du CVM, de la cytologie, et du test VPH pour la détection de CIN et des hrVPH.

Méthodes: Dans le manuscrit 1, nous avons mené une revue de trois bases de données jusqu'au 27 juillet, 2022. Les articles éligibles discutaient du CVM et du cancer du col de l'utérus associé au VPH, caractérisaient le CVM via des techniques métagénomiques et rapportaient un estimé de l'association. Nous avons extrait et résumé les statistiques, la conception de l'étude, la population et la méthodologie des articles inclus. Le manuscrit 2 incluait 186 femmes [54 normal, 50 CIN1, 40 CIN2, 42 CIN3] référées à la colposcopie à la suite d'un résultat de cytologie anormal. Les échantillons ont été génotypés pour les hrVPHs avec le test Roche cobas 4800. Le CVM a été caractérisé à l'aide du séquençage génétique 16S rRNA de deux régions (V3V4, V5V6) et du traitement bio-informatique via le pipeline haute-résolution ANCHOR. Des modèles de régression logistique ont été construits avec 1) les espèces du CVM, 2) les hrVPHs, 3) la cytologie, et 4) les espèces du CVM et les hrHPVs comme prédicteurs, et avec CIN2+ comme issu. Les coefficients ont été utilisés pour construire des scores linéaires sur les espèces du CVM, sur la cytologie, sur les VPHs, et sur les espèces du CVM/VPH. Les espèces ont été sélectionnées via la régression logistique avec sélection par étape ascendante. Les courbes d'efficacité de fonctionnement du récepteur ont été tracées, et l'aire sous la courbe (AUC) et les

intervalles de confiance (IC) à 95% ont été comparés afin d'évaluer leur performance clinique (rapportés en AUC;95%CI).

Résultats: Dans le manuscrit 1, une haute diversité du CVM et la déplétion du *Lactobacillus* semblent augmenter et diminuer, respectivement, le risque de résultats néfastes associés aux cancers du col de l'utérus liés au VPH. Dans le manuscrit 2, 77 espèces ont été identifiés; 8 uniques au V3V4, 48 au V5V6, et 21 partagés. Un score basé sur les espèces du CVM (0.64;0.57-0.71) était le moins précis pour la détection de CIN2+. Toutefois, la précision d'un score basé sur la combinaison des espèces du CVM et VPH (0.80;0.74-0.86) était similaire à celle de la cytologie (0.84;0.79-0.90) et du VPH (0.76;0.70-0.82).

Discussion: Le CVM pourrait être impliqué dans le cancer du col de l'utérus. Cependant, les algorithmes standards 16S en déduisent des séquences des modèles d'erreur paramétriques, modifiant les séquences nucléotidiques, et entraînant une faible résolution taxonomique. En amplifiant deux régions et en utilisant ANCHOR pour améliorer le séquençage 16S, nous avons identifié une corrélation entre le CVM et CIN2+.

Conclusion: Les espèces du CVM et du VPH étaient corrélées avec CIN2+ dans cette étude transversale. Une étude de cohorte prospective pourrait aider à déterminer si le CVM joue un rôle causal dans la carcinogénèse cervicale ainsi que la valeur diagnostique du CVM.

ACKNOWLEDGEMENTS

First and foremost, I would like to acknowledge my supervisor, Dr. Eduardo Franco. In 2020, I began at McGill University as a student in the Graduate Diploma in Oncology and Dr. Franco served as my practicum supervisor. As a student with an interest in oncology Dr. Franco introduced me to cancer epidemiology which led me to pursue a master's degree in epidemiology. A four-month internship turned into three years at the Division of Cancer Epidemiology (DCE) under Dr. Franco's supervision. Thank you, Dr. Franco for your constant guidance, encouragement and support throughout my academic journey. In addition to your mentorship, thank you for the copious opportunities you provided me, all of which have led me to pursue a career in cancer research.

Throughout my time at the DCE, the Associate Director of Research, Dr. Mariam El-Zein has helped me grow as a young researcher. Thank you, Mariam, for your daily guidance, encouragement, thorough edits and feedback, and always being available at a moment's notice for a zoom or in-person meeting. I feel privileged to have had the opportunity to work with and learn from you over the past three years.

I would also like to thank my thesis committee members. First, Dr. Emmanuel Gonzalez, thank you for teaching me about the workings of microbial characterization and helping me become confident and competent in bioinformatic analyses. I appreciate your patience, guidance, one-on-one lectures and always answering my questions no matter how simple. Thank you as well to Dr. Nicole Basta and Dr. Shirin Golchi for your support and availability throughout my degree.

I cannot forget to mention the members of the DCE whom I feel honoured to have had the opportunity to work with and learn from. Specifically thank you to Parker Tope, Andrew Arthur and Sarah Botting-Provost. Thank you, Parker, for your assistance with my narrative review, Andrew for helping me learn STATA, and both of you for your friendship and support. Sarah, I appreciate all your hard work in the administrative office of the DCE and helping translate my abstract to French. Thank you to the professors, and administrative staff in the department of EBOH. I would also like to thank Fonds de Recherche du Québec – Santé for awarding me with a master's fellowship and Dr. Franco for his financial support.

Finally, thank you to my family and all of my friends who have encouraged me throughout this journey. Specifically, to my parents and siblings for their unwavering support, guidance and encouragement throughout all my academic endeavours.

TABLE OF CONTENTS

PREFACE AND AUTHOR CONTRIBUTIONS

This thesis consists of a literature review, two manuscripts (a narrative review and an empirical research manuscript) with a preamble for each, as well as a discussion of results and final conclusions. The contributions of co-authors to the manuscripts are detailed below. The remainder of this thesis was written by MAL with guidance from ELF on the entirety of the document and EG on the bioinformatics sections.

Manuscript #1, "A narrative review of the etiological role and diagnostic utility of the cervicovaginal microbiome in human papillomavirus-associated cervical carcinogenesis"

Margaret Logel, Parker Tope, Mariam El-Zein, Emmanuel Gonzalez, Eduardo L. Franco

This manuscript is under-review at *Cancer Epidemiology, Biomarkers & Prevention*. The project was proposed by ML and conceptualized by ML, MZ and ELF. ML designed the search and screened titles and abstracts for relevance. ML and PT screened full texts for eligibility and extracted data from eligible articles. ML synthesized and analyzed the data and drafted the manuscript under the guidance of MZ, EG, and ELF. PT, MZ, EG and ELF reviewed and amended the manuscript.

Manuscript #2, "Species-level characterization of the cervicovaginal microbiota and its role in human papillomavirus-associated cervical carcinogenesis"

Margaret Logel, Mariam El-Zein, Eduardo L. Franco, Emmanuel Gonzalez

This manuscript is under-review at the *Journal of Medical Virology*. MZ, ELF and EG formulated the research question. EG performed the microbial characterization and the bioinformatic analyses. ML assisted with the bioinformatic analyses, performed the epidemiological analyses and drafted the manuscript under the supervision of MZ, ELF and EG. MZ, ELF and EG reviewed and amended the manuscript.

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

Tables and figures prefaced with "S" refer to supplementary tables and figures.

S-Table 4-9. Stepwise logistic regression coefficients of cytology-, HPV-, and microbiome- (species raw abundance) based scores used to construct linear scores, comparing any high-risk HPV positive to negative.

S-Table 4-10. ROC curve analysis for the performance of cytology-, HPV-, and microbiome- (species raw abundance) based scores to detect CIN lesions and high-risk HPV infections.

S-Figure 4-2. Performance of cytology-, HPV-, and microbiome- (species raw abundance) based scores to detect CIN and high-risk HPV.

S-Table 4-11. Univariate logistic regression coefficients of $molBV_{V3\text{-}V4}$- and $molBV_{V5\text{-}V6}$ -based scores used to construct linear scores for the contrast groups.

S-Table 4-12. ROC curve analysis for the performance of $molBV_{V3\text{-}V4}$ - and $molBV_{V5\text{-}V6}$ - based scores to detect CIN and high-risk HPV infections.

S-Figure 4-3. Performance of $molBV_{V3\text{-}V4}$ - and $molBV_{V5\text{-}V6}$ -based scores to detect CIN and high-risk HPV.

LIST OF ACRONYMS & ABBREVIATIONS

AUC	Area under the curve
aOR	Adjusted odds ratio
AGC	Atypical glandular cells
ASC-US	Atypical squamous cells of undetermined significance
ASC-H	Atypical squamous cells-cannot exclude high-grade squamous intraepithelial lesion
aTRR	Adjusted transition rate ratio
BV	Bacterial vaginosis
CI	Confidence interval
CIN	Cervical intraepithelial neoplasia
CVM	Cervicovaginal microbiome
CST	Community state type
CCA	Supervised ordination analysis
ESV	Exact sequence variant
HIV	Human immunodeficiency virus
HPV	Human papillomavirus
HSIL	High-grade squamous intraepithelial lesion
HSV	Herpes simplex virus
hrHPV	High-risk human papillomavirus
LMIC	Low- and middle- income country
LSIL	Low-grade squamous intraepithelial lesion
lrHPV	Low-risk human papillomavirus
MARKER	Methylation Analysis Revealing Key Epigenetic Regulation
molBV	Molecular bacterial vaginosis
NILM	Negative for intraepithelial lesion or malignancy
NCBI	National Center for Biotechnology Institute
OR	Odds ratio

Pap	Papanicolaou
PCR	Polymerase chain reaction
PI	Prediction interval
PPV	Positive predictive value
pRb	Retinoblastoma protein
p53	Tumor protein p53
RA	Relative abundance
RCT	Randomized controlled trial
RR	Relative risk
RRR	Relative risk ratio
ROC	Receiver operating characteristic
STI	Sexually transmitted infection
TRR	Transition rate ratio
US	United States
WMS	Whole metagenome sequencing
16S rRNA	16S ribosomal RNA

CHAPTER 1. INTRODUCTION

1.1. Rationale

Human papillomavirus (HPV) is the most common sexually transmitted infection (STI) worldwide and infection with oncogenic high-risk HPV (hrHPV) types is a necessary cause of cervical cancer.[1,2] Cervical cancer has a significant global disease burden; it remains the fourth most common female cancer with approximately 604 000 new cases and 342 000 deaths in 2020.[3] Although hrHPV is a necessary cause of invasive cervical cancer and its precancerous lesion (cervical intraepithelial neoplasia or CIN), it is not a sufficient one, suggesting that other co-factors may contribute to carcinogenesis. Three CIN grades exist on a biological continuum of increasing severity: CIN1 (low-grade, mild dysplasia), CIN2 (high-grade, moderate dysplasia), and CIN3 (high-grade, severe dysplasia). Without treatment, high-grade CIN can progress to invasive cervical cancer. The majority of acquired hrHPV infections are transient and resolve spontaneously.[4,5] It is the persistence of hrHPV types that leads to high-grade CIN and subsequently cervical cancer.

The cervicovaginal microbiome (CVM) of a reproductive-aged women is dominated by species of the *Lactobacillus* genus corresponding to low microbial diversity.[6] Bacterial vaginosis (BV), a pathogenic state of the CVM, corresponds to high microbial diversity,[7] and has been associated with the acquisition of several STIs including chlamydia, gonorrhea, herpes simplex virus (HSV), human immunodeficiency virus (HIV), and HPV.[8–13] Systematic reviews and meta-analyses have provided evolving evidence suggesting that different communities of the CVM may contribute to HPV-associated cervical carcinogenesis.[14–20] However, a limitation of previous observational studies is inadequate characterization of the CVM; several articles that investigated the relationship between the CVM and HPV or cervical lesions/cancer characterized the CVM with 16S ribosomal RNA (16S rRNA) gene sequencing. Standard 16S algorithms fail to provide high taxonomic resolution of 16S data and subsequently identification of bacterial communities is mainly at the genus-level (or other) rather than the species-level.[21] As such, there is a critical need for a synthesis of studies with a specific focus on microbial characterization methods and assessment of the correlations between microbial communities, hrHPV infections and CIN severity at high taxonomic resolution (i.e., species-level).

1.2. Overall Aim & Objectives

The overall aim of this clinical epidemiological project was to explore the relationship between the bacterial communities that constitute the CVM and HPV-associated cervical carcinogenesis. Accordingly, two specific objectives were developed, and findings are presented in two manuscripts. The first was to compile the existing literature on the associations between the CVM and HPV prevalence, acquisition, persistence, clearance, and cytological interpretations or biopsy confirmed CIN/cervical cancer. In the same manuscript, we summarize findings from studies that assessed the diagnostic accuracy of CVM communities in cervical carcinogenesis. The second was to characterize the CVM in samples from the Methylation Analysis Revealing Key Epigenetic Regulation (MARKER) study at high taxonomic resolution and assess CVM diversity in relation to CIN lesion severity. Specifically, we explored correlations between CVM bacterial species and hrHPV infections as well as different grades of CIN lesion severity. We also compared the diagnostic accuracy of CVM bacterial species, HPV positivity, and cytology results for the detection of any hrHPV and CIN.

CHAPTER 2. LITERATURE REVIEW

2.1. Cervical Cancer Etiology

2.1.1. Human Papillomavirus (HPV) Epidemiology

Genital HPV infections are the most common STIs, and most sexually active individuals are expected to be exposed to the virus throughout their lifetime.[2] Findings from HPV-type concordance studies performed among couples have shown that the primary route for genital HPV transmission is sexual intercourse.[22] An important risk factor for HPV acquisition and prevalence is the number of recent and lifetime sexual partners.[23] HPV-associated disease burden is high among females as HPV has been established as a necessary cause of invasive cervical cancer,[1] and based on measures from multiple studies, prevalence ranges from 2 to 44% in asymptomatic women.[2] Those at highest risk for acquisition are young adults who are sexually active,[2] and historically, among asymptomatic females, prevalence has been shown to peak during young-adulthood, gradually decline overtime, and rise again around age 55.[24] Some plausible explanations for a peak at older ages include reactivation of latent infections, new intimate partners at older ages, and a birth cohort effect facilitated by differences in sexual behaviours and exposure to HPV among cohorts.[22]

Persistent infection with HPV can cause malignant diseases. Based on epidemiological evidence, the International Agency for Cancer Research (IARC) has classified infection with thirteen HPV genotypes as Group 1 carcinogens (i.e., the highest level of carcinogenicity, carcinogenic to humans): HPVs 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and 66.[25] These types are commonly referred to as hrHPV genotypes. Infection with hrHPV is well established as a necessary, but not sufficient, cause of invasive cervical cancer and its pre-cancerous lesion, CIN.[1] Two hrHPV genotypes, HPV16 and 18, cause 70% of cervical cancers.[26] Essentially all cases of cervical cancer are caused by persistent infections with hrHPV types.[27] Based on GLOBOCAN data from 2012, there were approximately 530 000 incident cases of cervical cancer and 100% were attributable to HPV infections.[28] Other HPV-associated cancers exist, however, the attributable fraction is lower: anal (88.0%), vulva (24.9%), vaginal (78.0%), penile (50.0%), oropharynx (30.8%), oral cavity (2.2%), larynx (2.4%), other pharynx (0.0%).[28]

2.1.2. HPV Virology & Pathogenesis

HPV is a small, non-enveloped, double-stranded DNA virus with an icosahedral capsid.[29] The viral diameter is approximately 55 nm.[29] HPV infects basal cells of the squamous epithelium

and can eventually lead to proliferative lesions during squamous cell maturation.[29,30] The genome is 8 kb and has three major regions; (1) the noncoding region which regulates DNA replication, (2) the early region encoding early genes (E1, E2, E4, E5, E6 and E7) for viral replication and oncogenesis, and (3) the late genes (L1 and L2) which form the capsid protein.[29,31] These different genes are expressed at different points throughout the viral life cycle. Distinct HPV genotypes have different biological properties that contribute to cancer risk and types are classified according to the nucleotide sequence in the L1 gene.[29,30]

Host cell factors facilitate HPV replication by interacting with the noncoding region which triggers the transcription of viral genes. E6 and E7 are oncogenes which increase HPV's selective growth advantage by binding to and inactivating the tumor protein p53 (p53) and the retinoblastoma protein (pRb), respectively.[32] Under normal cellular conditions, p53 and pRb are critical for controlling the cell cycle. p53 regulates apoptosis, initiates DNA repair mechanisms and cell cycle arrest at the G1 checkpoint, whereas pRB blocks the E2F-1 transcription factor which governs entry to the S phase of the cell cycle.[29] Infections with hrHPV genotypes have been demonstrated to have higher binding affinities between E6 and p53 as well as E7 and pRb, which ultimately leads to increased proliferation.[29] E5 is responsible for further stimulation of cellular growth and differentiation. E1 and E2 also facilitate HPV DNA replication.[29] Collectively, the activity of the aforementioned early genes accelerates uncontrolled cellular growth. The L1 and L2 proteins are activated by the late promoter and the capsid is formed; HPV virions are released which is facilitated by E4.[29] Overall, during the viral lifecycle, cellular proliferation of cells in the basal epithelial layer may facilitate carcinogenic changes.

2.2. Natural History of HPV-Associated Cervical Carcinogenesis

2.2.1. Overview

The natural history of HPV-associated cervical carcinogenesis is well established and shown in **Figure 2-1**. Four major steps must occur for an HPV infection to develop into invasive cervical cancer: HPV acquisition, HPV persistence, progression to pre-cancerous lesions and invasion.[27] There is also the potential for backwards steps (except after invasion occurs); an HPV infection can clear, and lesions may regress. For each stage leading to invasive cervical cancer, only a small proportion will progress to the next; few HPV infections persist and lead to pre-cancerous lesions and even fewer progress to invasive cervical cancer.[27] Relatively little is known about additional biological co-factors that cause an HPV infection to persist and CIN

lesions to progress. Some modifiable factors shown to increase the risk of cervical pre-cancer or invasive cervical cancer include smoking, oral-contraceptive use, and multiparity.[33–35]

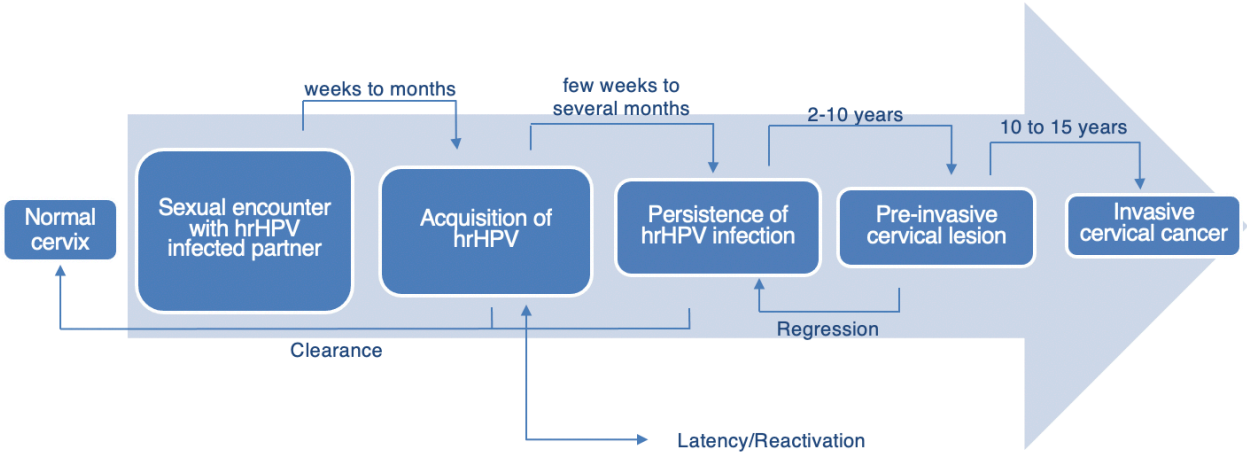


Figure 2-1. The natural history of HPV-associated cervical carcinogenesis.

2.2.2. HPV Acquisition

One must first acquire an infection with a hrHPV type through sexual intercourse with an infected partner which is transmitted via skin-to-skin or mucosa-to-mucosa contact;[22,27] those at highest risk for HPV acquisition are young adults who are sexually active.[2] Compared to other STIs (HIV and HSV), transmissibility following sexual activity is considered high.[22,27] Number of sex partners, smoking, oral contraceptive use, STIs, immunosuppressive conditions, and parity are some factors which have been shown to increase the risk of a genital HPV infection.[2]

2.2.3. HPV Clearance & Persistence

Despite high HPV prevalence in sexually active individuals,[2] most HPV infections are transient and resolve spontaneously.[4,5] They often become undetectable within 1-2 years, with a median time to clearance of approximately 6-18 months.[27] Undetectable infections may be cleared by the immune system or become latent.[27,36] Although latent infections can reactivate and become detectable, immunocompetent individuals are unlikely to develop disease.[37] The molecular mechanisms of clearance/latency are not well understood, however, they are likely affected by a cellular-mediated immune response leading to viral persistence.[38] The greatest risk factor (which is also a necessary cause) for progression to cervical pre-cancer and cervical cancer is the persistence of a hrHPV infection.[39] Persistence has been found to vary by HPV genotype. A systematic review and meta-analysis identified HPV 16, 31, 33, and 52 as the most persistent

HPV genotypes and HPV 35, 51, 66 and 68 as the least.[40] Compared to low-risk HPV (lrHPV) types, hrHPV types were found to persist longer (9.3 versus 8.4 months).[40]

2.2.4. Progression to Cervical Pre-Cancer and Invasive Cancer

Prior to the development of invasive cervical cancer, a persistent infection with a hrHPV genotype must progress to cervical lesions. Lesions can be detected via cytological screening (discussed in more detail in section 2.3.3) and diagnosed via biopsy (i.e., histology). Different terminology systems have been used for lesion classification, depending on whether they emphasize cytology or histology. The CIN system is a 3-tier system originally used for histopathological ascertainment of lesions. The Bethesda system is a 2-tiered system originally reserved for cytological reports. The Lower Anogenital Squamous Terminology (LAST) attempted to harmonize the terminology via a 2-tiered system.

CIN1 lesions are the least severe CIN grade and are generally the manifestation of an active HPV infection.[27] CIN2 lesions are ambiguously considered pre-cancerous.[27] By contrast, the most severe CIN grade is CIN3 and this lesion is classified as cervical pre-cancer.[27] CIN1 and CIN2 lesions do not necessarily progress to cervical pre-cancer or invasive cancer.[27] Cervical precancer is evident when the cervical epithelium is replaced by undifferentiated cells with genomic abnormalities and cytopathic effects.[27] A large proportion of CIN lesions will regress to a less severe state. Based on findings from a 2021 meta-analysis, the pooled-percentage of CIN1, CIN2 and CIN3 lesions that regressed to normal were 60% (95% confidence interval (CI), 55-65), 47% (95% CI, 42-51) and 18% (95% CI, 6-34), respectively.[41] By contrast, the cumulative risk of progression to CIN lesions or cervical cancer was shown to increase overtime in a recent meta-analysis of hrHPV positive females with normal cytological assessment (i.e., NILM).[42] Cumulative risk was assessed at three time points: 1-, 3-, and 5-years. With respect to the aforementioned time points, weighted risk and 95% prediction intervals (PI) for progression to CIN2+ lesions were 3.9% (95% PI, 0.0-11.2), 7.0% (95% PI, 0.0-14.0) and 9.9% (95% PI, 2.5-16.8).[42] By contrast, for progression to cervical cancer, the weighted risks were 0.8% (95% PI, 0.0-8.3), 1.2% (95% PI, 0.0-8.7) and 1.6% (95% PI, 0.0-9.1).[42] HPV type is a factor contributing to the risk of progression to high-grade lesions. In the same pooled-analysis, for a hrHPV positive female at baseline, after adjustment for HPV type, the following HPV genotypes appeared to significantly increase the risk of CIN3+ lesions: HPV 16/18, hrHPVs other than HPV 16/18, and HPV's 16, 18, 31, 33.[42]

2.2. Descriptive Epidemiology of Cervical Cancer

2.2.1. Worldwide Distribution of Cervical Cancer

Based on GLOBOCAN data from 2020, there were approximately 9.2 million new cancer cases and 4.4 million cancer deaths in females.[3] Among these cancer cases, cervical cancer is the fourth most common cancer following breast, colorectum, and lung cancer with approximately 604 000 new cases and 342 000 deaths.[3] This accounts for 6.5% and 7.7% of all incident and deaths, respectively, from any female cancer site.[3] Cervical cancer is the leading cause of cancer death in 36 countries, several of which are in the developing world (sub-Saharan Africa, Melanesia, South America, and South-Eastern Asia).[3] The high-burden of cervical cancer in low- and middle- income countries (LMICs) is largely attributable to inequities in the implementation of HPV vaccination and cervical cancer screening programs (discussed in detail in section 2.3).[3]

2.3. Prevention of Cervical Cancer

2.3.1. Overview

Cervical cancer is considered almost entirely preventable due to the development and implementation of successful primary and secondary prevention strategies.[3] Interventions can be applied at three different levels to prevent cancer prior to the development of invasive disease: primary, secondary and tertiary. These levels and some of the available prevention techniques for cervical cancer are summarized in **Table 2-1**. Common modalities for primary and secondary cervical cancer prevention are further elaborated in sections 2.3.2 and 2.3.3, respectively, as they are well established and widely utilized.

Table 2-1. Overview of the levels of cancer prevention and the existing strategies for cervical cancer.

Prevention Level	Definition	Strategies for cervical cancer
Primary	- Prevention of cancer prior to the biological onset of disease by identifying and modifying risk factors	- HPV vaccination - Abstinence - Condom use
Secondary	- Screening to detect early signs and biological markers of cancer prior to the development of signs or symptoms	- Screening modalities for the early detection of cervical lesions (i.e., cervical cytology, HPV DNA testing, etc.)

Tertiary	- Identification of prognostic factors to treat and improve the clinical outcome of cancer	- Surgery - Radiotherapy - Chemotherapy - Targeted molecular therapies

2.3.2. Primary Prevention

Six prophylactic HPV vaccines, containing L1 (i.e., one of the late genes forming the HPV capsid protein) viral like particles (VLPs), have been developed and are licenced for use to prevent against HPV-associated diseases including cervical cancer.[43] L1 VLPs are HPV genotype specific and highly immunogenic; immunization results in high levels of neutralizing antibodies against the L1 viral capsid protein.[44] The three most commonly used vaccines are Cervarix™ (Bivalent vaccine targeting HPV 16 and 18), Gardasil™ (Quadrivalent vaccine targeting HPV 6, 11, 16, 18) and Gardasil 9™ (Nonvalent vaccine targeting HPV 6, 11, 16, 18, 31, 33, 45, 52, 58).[45] All three vaccines received licensure and approval from the Food and Drug Administration in 2006 (Gardasil™), 2009 (Cervarix™), and 2014 (Gardasil 9™).[45] In randomized controlled trials (RCT), the vaccines demonstrated close to 100% efficacy against the development of cervical lesion in HPV naïve females.[46–50] Now, over 15 years post-licensure, they have proven effective by reducing the prevalence of vaccine targeted HPV genotypes and CIN.[51,52] A limitation of the currently available HPV vaccines is cost,[53] and based on data from 2020 less than 30% of LMICs had implemented an HPV vaccination program.[3]

Two additional techniques to prevent the acquisition of an HPV infection include complete abstinence and condom use. However, these methods are less effective at preventing an HPV infection, as complete abstinence is unrealistic in the real-world context.[27] Although a reduction in the risk of genital HPV infections has been evident with consistent male condom use,[54] it is not entirely effective due to the exposure of male anogenital skin.[27]

2.3.3. Secondary Prevention

Cervical cytology and HPV DNA testing are the two most common screening modalities for the early detection of cervical lesions; both techniques detect cervical abnormalities prior to the development of invasive cancer. Cytological assessment is conducted utilizing the Papanicolaou (Pap) smear, a technique which consists of assessing cervical cells microscopically to detect cellular abnormalities. For the detection of high-grade cervical lesions, this technique

has a high specificity (ranging from 96-98%) and poor sensitivity (ranging from 51-53%).[55–57] HPV DNA testing relies on the detection of hrHPV genotypes to indicate the presence of low- or high-grade cervical lesions. Findings from one of the first RCTs comparing HPV DNA testing to cytology for the detection of CIN2+ lesions demonstrated that HPV DNA testing has a much higher sensitivity (94.6% vs. 55.4%) and somewhat lower specificity (94.1% vs. 96.8%) to cytology for the detection of CIN2+ lesions.[58] Since the implementation of HPV vaccination, there has been a reduction in HPV and CIN prevalence,[51,52] and a corresponding decrease in the positive predictive value (PPV) of cytological assessment and HPV DNA testing is expected. With a decreasing PPV of these common screening modalities, a larger proportion of females will receive a false positive test result. Thus, as this decline in prevalence continues there may be a need for reformulating screening algorithms and practices as the PPV continues to decrease.[59,60]

Screening guidelines vary according to region and are highly dependent on cost and resource availability.[27] CIN3 lesions that progress to cervical cancer can be largely attributed to the absence of and/or ineffective screening.[27] As of 2020, only 44% of women residing in LMICs had been screened for cervical cancer,[3] indicating a possible need for interventions to increase screening rates and additional low-cost testing options.

2.4. Cervicovaginal Microbiome (CVM)

2.4.1. Overview of the Human Microbiome

The human microbiome is composed of genomic content from several micro-organisms residing in various regions of the body and organs; there are between 10-100 trillion microbial cells and micro-organisms can be communalistic, mutualistic, or pathogenic.[61,62] Some of the micro-organisms that colonize different anatomical sites of the human body include archaea, bacteria, fungi, plasmids, and viruses. The prokaryotes that inhabit the mucosal and epithelial surfaces of the body constitute one component of the microbiome, the bacteriome.[63] Age, sex, body-mass index, mental health, physical health, diet, antibiotic usage, socioeconomic status, and immune response are some of many environmental factors that can alter the composition of the bacteriome.[63,64] Dysbiosis (i.e., harmful changes in the bacteriome) has been shown to lead to the proliferation of oncogenic bacteria, and subsequently linked to carcinogenesis in some human cancers including hepatocellular carcinoma, lung, colorectal, and breast cancers.[63,65]

2.4.2. Metagenomics for Microbial Characterization

Metagenomics is broadly defined as studying nucleotide sequences that are isolated and present in a sample.[66] This field is evolving rapidly and consists of various techniques to identify microbial communities including, but not limited to, shotgun metagenomics, metabolomics, metaproteomics, metatranscriptomics and gene marker analysis. Each technique relies on different methods, detection ability, and has various strengths and limitations for microbial characterization. Shotgun metagenomics is utilized for identification of the complete microbiome as all microbial genomes in a sample can potentially be sequenced.[66] Similarly, metatranscriptomics also identifies the complete microbiome, but instead identifies microbial gene expression patterns by sequencing the messenger RNA transcribed from genes that encode protein. Metabolomics and metaproteomics target metabolites and proteins produced by microbes within the microbiome, respectively.[66] Gene marker analysis is the most common method for microbial characterization and consists of capturing specific regions of the genome of microbial species.[66] One example of this is 16S rRNA gene sequencing which is utilized for bacterial and archaeal identification (see footnote 1).[1] Briefly, 16S rRNA gene sequencing relies on amplifying variable regions of the taxonomic 16S rRNA gene by targeting and amplifying via polymerase chain reaction (PCR) nine highly conserved regions (V1 to V9) common to bacteria and archaea.[67] The resulting amplicons are analyzed using bioinformatic pipelines and their nucleotide sequence is compared to existing databases to assign taxonomy. Due to technical difficulties (both at the sequencing and bioinformatics steps) the first generation of bioinformatics pipelines including, QIIME1, and MOTHUR generally failed to annotate the majority of 16S data at high taxonomic resolution. Taxonomic resolution refers to the ability to detect specific bacterial taxa,[68] where a high resolution corresponds to the detection of species or strains (i.e., the lowest taxonomic rank according to the taxonomic hierarchy). A new generation of pipelines (DADA2, QIIME2, UPARSE) obtain higher resolution but this capacity decreases with sample complexity.[21] By contrast the lowest taxonomic resolution corresponds to the identification of domains (i.e., the highest taxonomic rank according to the taxonomic hierarchy).

[1] Further extrapolated as this is the characterization technique utilized in the present work.

In 2019, Gonzalez and colleagues developed the high-resolution ANCHOR pipeline which has been shown to surpass the resolution of standard bioinformatics pipelines for replicated samples.[21] As shown in **Figure 2-4**, ANCHOR can annotate the majority of counts at the species-level, whereas DADA2, UPARSE, QIIME1 and MOTHUR are mainly at the genus level or higher.[21]

Figure 2-2. The taxonomic resolution of the ANCHOR pipeline compared to standard bioinformatics pipelines for gene marker analysis.[21] *Reprinted with permission from Dr. Emmanuel Gonzalez.*

2.4.3. Composition of the CVM

The CVM consists of microorganisms that colonize the cervix and vagina. With respect to the microbiome, these anatomical regions can be considered interchangeable (see footnote 2);[2] the overall concordance rate for detection of specific microbes from cervical and vaginal samples from 96 study participants was 92%.[69] After sequencing the CVM of 396 healthy females with 16S rRNA gene sequencing, five different community state types (CST) were identified based on the relative abundance (RA) and diversity of the micro-organisms. CST I, II, III and V were dominated by *Lactobacillus crispatus*, *Lactobacillus gasseri, Lactobacillus iners*, and *Lactobacillus jensenii*, respectively.[6] Of the study participants, the CVM composition of 26.2%, 6.3%, 34.1% and 5.3% fell into the respective CSTs.[6] By contrast, CST IV was characterized by

[2] In the present work, the cervical and vaginal microbiome are considered interchangeable and referred to as the cervicovaginal microbiome.

a loss of *Lactobacillus* dominance and increase in anaerobic bacteria of the following genera: *Prevotella, Dialister, Atopobium, Gardnerella, Megasphaera, Peptoniphilus, Sneathia, Eggerthella, Aerococcus, Finegoldia,* and *Mobiluncus*.[6] This group accounted for approximately 27% of the study population.[6]

2.4.4. Dysbiosis of the CVM & STIs

Lactobacillus species may be protective in the CVM and defend against disease-associated states by lowering the vaginal pH through the production of lactic acid.[6,70] BV is a clinical condition of the CVM which is often referred to as dysbiosis. This state is characterized by a loss of *Lactobacillus* dominance, and increased diversity due to the presence of several anaerobic microbes (i.e, *Gardnerella, Prevotella, Atopobium, Mobiluncus, Bifidobacterium, Sneathia, Leptotrichia, Clostridiales*).[7] BV is a highly prevalent condition among females; based on data collected between 2001-2004 in the United States (US), 29% of women aged 14 to 49 were positive for BV.[71] Some important risk factors for BV include menses, a new intimate partner, vaginal douching, receiving oral sex and not using condoms during sexual intercourse.[72] Evidence suggests that BV increases the risk of several common STIs including chlamydia, gonorrhea, HSV and HIV.[8,9,13]

The association between BV and HPV has also been investigated, yet few articles have used metagenomic sequencing for BV detection. Clinically, BV can be diagnosed via the Nugent score (gram-staining), Amsel criteria or the presence of clue cells on a Pap smear.[73–75] However, in these cases, there is subjectivity in interpretation and an overall measurement of bacterial diversity rather than the identification of specific bacterial communities. Among 1136 HIV-negative and HIV-positive women from the US who were followed for up to 15 visits, BV diagnosed via the Nugent Score, was significantly associated with an increased odds of HPV prevalence (adjusted odds ratio (aOR), 1.14; 95% CI, 1.04-1.26), HPV incidence (aOR, 1.24; 95% CI, 0.72-0.97) and decreased the likelihood of HPV clearance (adjusted hazard ratio, 0.84; 95% CI, 0.72-0.97).[10] Additionally, a longitudinal analysis among 1125 females in the US, found a significant association between BV, diagnosed using wet mounts, and an increased risk of HPV persistence (hazard ratio, 1.60; 95% CI, 1.07-2.39).[12] As demonstrated by the statistically significant associations with BV, research suggests that microbial communities are involved in the natural history of an HPV infection.

Several factors have led to an abundant amount of literature assessing the relationship between the composition of the CVM, with respect to the identification of specific bacterial communities via metagenomic methods, in HPV-associated cervical carcinogenesis (comprehensively summarized and analyzed in Chapter 3). Three key factors include that: (1) the microbiome has been implicated in human cancers,[63,65] (2) *Lactobacillus* species dominate the CVM and may protect against diseases,[6,70] and (3) BV has been associated with several STIs including HPV.[8–13] Generally, findings are aligned with the proposed mechanism: *Lactobacillus* species seem to play a protective role, whereas high diversity may be related to HPV and subsequently cervical carcinogenesis. A meta-analysis conducted in 2019 found that a *Lactobacillus* depleted microbiome was significantly associated with an increase in the odds of HPV prevalence (pooled odds ratio (OR), 1.53; 95% CI, 1.23-1.82; I^2, 66%, n = 20 studies).[19] Similarly, increased diversity of the CVM, identified using 16S rRNA gene sequencing, has been associated with the development of CIN2+ lesions among women with an incident hrHPV infection.[76] As the field of metagenomics for microbial characterization is evolving, there is a need for a compilation of the epidemiological associations between CVM communities in HPV-associated cervical carcinogenesis. Additionally, few studies have utilized high-resolution microbial characterization methods to assess correlations between CVM species, hrHPV infections, and CIN severity. Both of the aforementioned gaps in the literature will be addressed in this thesis.

CHAPTER 3. NARRATIVE REVIEW MANUSCRIPT

3.1. Preface

Numerous systematic reviews and meta-analyses have explored whether the CVM is involved in HPV-associated cervical carcinogenesis. Yang et al.,[15] Mortaki et al.,[14] and Sims et al.,[16] performed systematic reviews in 2018, 2020 and 2021, respectively. Similarly, three meta-analyses were published by Brusselaers et al.,[18] Tamarelle et al.,[19] and Wang et al.[20] in 2019. Generally, *Lactobacillus* species appear to protect against poor health outcomes as a hrHPV infection is acquired, and develops into CIN or cervical cancer, whereas high diversity or a loss of *Lactobacillus* dominance may increase the risk of unfavourable outcomes throughout this process. Specific findings from these systematic reviews and meta-analyses are detailed in Chapter 3. Importantly, most of these reviews did not restrict study inclusion based on characterization techniques to assess microbial variability. Wang et al. excluded studies using microscopic techniques; however, the authors only explored the role of *Lactobacilli* in cervical carcinogenesis, rather than the entire composition of the CVM.[20] As several methods for microbial characterization have evolved overtime (i.e., microscopic identification of BV to the more novel utilization of metagenomic sequencing for the identification of specific microbial communities) this is an important consideration as methodological differences may impact both the heterogeneity and comparability of findings across the literature.

In Chapter 3 we present a manuscript titled, *A narrative review of the etiological role and diagnostic utility of the cervicovaginal microbiome in human papillomavirus-associated cervical carcinogenesis.* Our study builds upon findings from previous systematic reviews and meta-analyses by restricting to observational studies that characterized the CVM using metagenomic techniques. Specifically, we summarize all observational studies, systematic reviews and meta-analyses assessing epidemiological associations between the CVM and HPV prevalence, acquisition, persistence, clearance and cervical lesions/cancer. We also identify and compare the literature on the clinical role of the CVM in these processes; no review article has summarized studies assessing whether CVM components can detect HPV, CIN lesions or cervical cancer. Following the initial submission of this thesis, this manuscript was submitted to *Cancer Epidemiology Biomarkers and Prevention*.

3.2. MANUSCRIPT 1: A NARRATIVE REVIEW OF THE ETIOLOGICAL ROLE AND DIAGNOSTIC UTILITY OF THE CERVICOVAGINAL MICROBIOME IN HUMAN-PAPILLOMAVIRUS ASSOCIATED CERVICAL CARCINOGENESIS

Authors:

Margaret Logel,[1] Parker Tope,[1] Mariam El-Zein,[1] Emmanuel Gonzalez,[2,3] Eduardo L. Franco[1]

Affiliations:

[1]Division of Cancer Epidemiology, McGill University, Montréal, Québec, Canada

[2] Microbiome Unit, Canadian Centre for Computational Genomics (C3G), Department of Human Genetics, McGill University, Montreal, Canada

[3] Centre for Microbiome Research, McGill University, Montreal, Canada

Correspondence:

Margaret Logel

Division of Cancer Epidemiology, McGill University

5100 Maisonneuve Blvd West, Suite 720

Montreal, QC, H4A 3T2, Canada

E-mail address: margaret.logel@mail.mcgill.ca

3.2.1. Abstract

Increasing evidence suggests that communities of the cervicovaginal microbiome (CVM) contribute to human papillomavirus (HPV)-associated cervical carcinogenesis. However, few reviews have focused on metagenomic sequencing. We summarized the literature on the etiologic role of the CVM in cervical carcinogenesis by systematically searching Medline, Web of Science and Embase until July 27th, 2022 (and monitoring subsequent publications in PubMed). Additionally, we identified studies assessing the diagnostic role of the CVM in HPV-associated cervical carcinogenesis by searching PubMed until June 23rd, 2023. We searched Google and Google Scholar to compare common CVM characterization techniques. Twenty-eight records (21 observational studies, four systematic reviews, and three meta-analyses) presented or summarized associations between the CVM and HPV acquisition, prevalence, persistence, clearance, and cervical lesions or cancer. Three observational studies conducted a receiver operating characteristic curve analysis to identify bacterial taxa detecting high-risk HPV prevalence or cervical lesions. The area under the curve from the identified studies ranged from 0.802 – 0.952. 16S ribosomal RNA gene sequencing and whole metagenome sequencing have sufficient resolution to study the CVM bacteriome. Bacterial communities may have important causal and clinical implications in cervical cancer; however, there is a need for methodological standards for CVM characterization.

3.2.2. Introduction

Human papillomavirus (HPV), the most common sexually transmitted infection (STI),[1] is a necessary cause of invasive cervical cancer and its precancerous lesion, cervical intraepithelial neoplasia (CIN).[2] Globally, cervical cancer is the fourth most common female cancer despite being preventable with HPV vaccination and cervical cancer screening.[3] As of 2020, there were approximately 604 000 new cases and 342 000 attributable deaths.[3] Vaccination coverage is suboptimal across Canada (uptake rates <80% in most provinces)[4] and in the United States (US) (uptake rates <60%).[5] Moreover, since the primary target age for HPV vaccination is pre-adolescence and adolescence, most adults remain unvaccinated and at high risk of acquiring an HPV infection. Vaccination and screening uptake are inferior in low- and middle-income countries; less than 30% have implemented vaccination programs, and only 44% of women have been screened for cervical cancer.[3] Although vaccination and screening will continue to be the main prevention strategies for cervical cancer, there is a need for continued clinical and epidemiological research to reduce the global disease burden.

An area of growing research interest is the relationship between the communities constituting the cervicovaginal microbiome (CVM) and HPV infections. The microbiome consists of several microorganisms, including archaea, bacteria, fungi, plasmids, and viruses. In particular, the bacteriome (i.e. bacterial community) has been shown to vary over time based on genetic, hormonal, and environmental factors.[6,7] High concordance rates between microbes identified in cervical and vaginal samples from the same subjects suggest that the microbial composition at these anatomical sites is generally comparable.[8] Evidence suggests that the vaginal microbiome of a healthy female is dominated by bacterial species of the *Lactobacillus* genus corresponding to low microbial diversity.[9] Bacterial vaginosis (BV), also known as vaginal dysbiosis, is characterized by a loss of *Lactobacillus* dominance and an increase in anaerobic bacteria contributing to high species diversity.[10] This state may be pathogenic and has been associated with the acquisition of STIs, including gonorrhea and Chlamydia.[11]

There are thirteen high-risk HPV (hrHPV) types that are considered carcinogenic to the uterine cervix.[2,12] Particularly, HPV16 and HPV18 cause 70% of cervical cancers.[13] Although necessary, hrHPV infections are not a sufficient cause, indicating the involvement of additional co-factors along the causal pathway. The majority of HPV infections are cleared or kept latent by the immune system; it is the persistence of hrHPV that can lead to CIN lesions (which can

spontaneously regress) and cervical cancer.[14–17] Little is known about the biological factors and mechanisms that influence the prevalence, acquisition, persistence, and clearance of hrHPV infections and the progression and regression of CIN lesions. Meta-analyses have provided evidence that different CVM communities may play a role in different stages of HPV-associated cervical carcinogenesis.[18–20] However, none of these meta-analyses considered all of the stages of HPV-associated cervical carcinogenesis: HPV acquisition, prevalence, persistence, clearance and CIN lesions or cervical cancer.

Numerous microbial characterization techniques and bioinformatic pipelines of microbial community data exist and have evolved in technical sophistication, which enhanced our understanding of the CVM. Microscopic evaluation is utilized to identify BV; some techniques include the Nugent score (a gram-staining method),[21] Amsel criteria,[22] and the Papanicolaou smear for detecting clue cells.[23] The above methods are morphological and thus lack resolution. Molecular testing approaches are necessary to study CVM microbial diversity. In this regard, metagenomics permits studying the community structure, activity, and functional potential of microorganisms in a biological sample. Several different metagenomic techniques exist, including, but not limited to, gene marker analysis, shotgun metagenomics, -transcriptomics, -proteomics, and metabolomics, all of which have different detection abilities, strengths, and limitations for identifying microbial communities.[24] Gene marker analysis is a targeted sequencing method. Specifically, for CVM characterization, 16S ribosomal RNA (16S rRNA) gene sequencing is a common technique that targets bacterial taxa. Ravel and colleagues were the first group to identify community state types (CST) of the CVM in reproductive-aged women using 16S rRNA gene sequencing.[9]

In light of the critical need for continued research to reduce cervical cancer burden and the variation in CVM characterization methods, the current narrative review aimed to summarize the etiologic role and diagnostic potential of the CVM in HPV-associated cervical carcinogenesis, with a particular focus on studies that characterized the CVM using metagenomic techniques. We also compared 16S rRNA gene sequencing to whole metagenome sequencing (WMS), considered the gold standard technique for microbial characterization, regarding methodological characteristics and CVM identification ability.

3.2.3. Methods

The current narrative review consisted of two components: (1) a systematic search of the literature to summarize the CVM's role in HPV-associated cervical carcinogenesis, and (2) a search of Google and Google Scholar to summarize metagenomic techniques for CVM characterization. For simplicity, and due to the high concordance in bacterial diversity between the cervix and vagina as sampling sites,[8] we considered the cervical and vaginal microbiome interchangeably (referred to as the CVM) without distinguishing between the two.

First, we identified original research articles, systematic reviews, and meta-analyses on the etiologic role of the CVM in HPV-associated cervical carcinogenesis by searching OVID Medline, Embase, and Web of Science from inception to July 27th, 2022, using MeSH headings and keywords related to "HPV", "cervical cancer", and "CVM". Records later published were monitored in PubMed for relevance until October 13th, 2023. The search strategies, reviewed by a McGill University librarian, are presented in **S-Table 3-1**. Records were managed in EndNote version X9, and duplicates were removed automatically and manually. Titles and abstracts were screened for relevance in Rayyan, and the full texts of relevant articles were independently assessed for eligibility by two reviewers (ML and PT). Disagreements were reviewed and resolved by consensus. To be eligible, original research articles had to characterize the CVM using a metagenomic technique and include a relative or absolute measure of association between the CVM and one or more of the outcomes of interest (HPV acquisition, prevalence, persistence, clearance, and/or cytology interpretations or biopsy-confirmed CIN and cervical cancer). Meta-analyses and systematic reviews had to discuss the relationship between the CVM and one or more of the outcomes of interest. We applied a language restriction, including only English publications. For research articles, we extracted the overall findings, exposures, outcomes, effect estimates and 95% confidence intervals (CI), and details regarding the study design, population, methods, microbial characterization and HPV genotyping techniques. For systematic reviews and meta-analyses, we extracted the general findings relating to the outcomes of interest, number of included studies, microbial characterization techniques of the included studies and limitations. Additionally, we extracted the inclusion/exclusion criteria, exposures, outcomes, pooled effect estimates, and their corresponding heterogeneity statistics for meta-analyses. Effect estimates were considered statistically significant when the 95% CI did not include the null value of the outcome measure.

To identify studies that assessed the diagnostic performance of CVM components in HPV-associated cervical carcinogenesis, we performed a separate PubMed search (by title and abstract) from inception until June 23rd, 2023, using keywords relating to "area under the curve (AUC)", "receiver operating characteristic (ROC)", "HPV", "cervical cancer" and "CVM" (**S-Table 3-1**). Eligible records included those that characterized the CVM with metagenomic techniques, assessed the diagnostic performance of the CVM to detect an outcome of interest using a ROC curve analysis, and reported an AUC and 95% CI. Record screening and data extraction (study design, population, microbial characterization methods, microbial components, and AUC with corresponding 95% CI) were performed by one reviewer (ML).

For the second component of the narrative review, we identified 16S rRNA gene sequencing as the most common metagenomic technique for CVM characterization among the studies included in the narrative synthesis. We then used Google and Google Scholar to compile information on 16S rRNA gene sequencing and WMS to provide an overview of metagenomic approaches in microbial research.

3.2.4. Results

The overall methodology is shown in **Figure 3-1.** Our systematic search yielded 1272 articles following duplicate removal. After primary screening, 276 full texts were assessed for eligibility, and 26 were deemed eligible; two additional articles were identified in PubMed following the systematic search for a total of 28 included articles. Seven articles on the diagnostic performance of the CVM in HPV-associated cervical carcinogenesis were identified in PubMed, and three records were eligible following primary and secondary screening. Additionally, information on 16S rRNA gene sequencing and WMS was acquired from 11 web pages/scholarly articles identified in Google (web pages) or Google Scholar (scholarly articles).

Etiologic role of the CVM in HPV-associated cervical carcinogenesis

Detailed information on the 21 included observational studies presenting epidemiological associations are shown in **S-Table 3-2** whereas **Table 3-1** lists the main findings. All records were published between 2014 and 2023. Observational studies were conducted in populations from Brazil,[25] Canada,[26] China,[27,28] Costa Rica,[29,30] Italy,[31] Korea[32–34], Mexico,[35] Nigeria,[36–38] Portugal,[39] South Africa,[40] and the US.[41–45] Sample sizes of included studies ranged from 12 to 807 women. Measures of association were reported between the CVM and HPV prevalence (n =

8), acquisition (n = 4), persistence (n = 4), and clearance (n = 4), and cytological interpretations or biopsy-confirmed CIN/cervical cancer (n = 13). Most studies were cross-sectional (n = 10) and longitudinal (n = 9) with two nested-case-control studies. The majority of studies (n = 18) utilized 16S rRNA gene sequencing for CVM characterization, differing by the hypervariable regions amplified (V1-V2, V1-V3, V3-V4, V3-V5, V3-V6, or V4) and sequencing platforms utilized (Genome Sequencer Titanium Roche-454, Illumina HiSeq, Illumina MiSeq, NovaSeq). The remaining three studies characterized the CVM using either PCR amplification, *Allplex*™ Bacterian Vaginosis Assay, or 16S ribosomal DNA gene sequencing. Specific study findings grouped by the outcomes of interest are detailed below.

HPV Prevalence

Collectively, findings from the eight studies that assessed the relationship between the CVM and HPV prevalence generally suggest that *Lactobacilli* are protective, whereas high diversity may be pathogenic.[25,27,33,38–41,44] At the community level, CST type was identified as a significant predictor of HPV prevalence (adjusted odds ratio (aOR), 0.74; 95% CI, 0.59 – 0.93).[27] Specifically, CSTs dominated by *Lactobacillus iners* (aOR, 0.67; 95% CI, 0.29 – 1.57 and aOR, 0.70; 95% CI 0.30 – 1.30), and *Lactobacillus crispatus* (aOR, 0.40; 95% CI, 0.10 – 1.70) appeared to reduce the odds of hrHPV.[38,41] Likewise, a community dominated by *Lactobacillus crispatus* or *Lactobacillus jensenii* appeared to non-significantly decrease the risk of any hrHPV among women with human immunodeficiency virus (HIV).[40] Three studies identified an increasing relative abundance (RA) of *Lactobacillus* species or *Lactobacillus crispatus* as protective against hrHPV.[33,40,44] A reduction in the odds of several hrHPVs was associated with an increasing prevalence of *Lactobacillus* species (relative risk (RR) range, 0.50 – 0.69),[39] and the presence of *Lactobacillus gasseri* (aOR, 0.50; 95% CI 0.25 – 1.02).[41] Several studies assessed the relationship between high microbial diversity and HPV prevalence. A high diversity CST and one dominated by anaerobic bacteria were positively and negatively associated with a decrease in the odds of hrHPV and HPV16 (aOR, 1.53; 95% CI 0.62 – 3.76, relative risk (RR), 0.75; 95% CI, 0.44 – 1.26, respectively).[25,41] The latter estimate was among HIV/HPV co-infected pregnant women. Other measures of bacterial diversity – the Simpson index and RA of BV anaerobes – suggested a non-significant increase in the risk of hrHPV among HIV-positive women (relative risk ratio (RRR) range, 2.066 – 2.395).[40] The presence of *Mobiluncus* species

was significantly associated with an increased risk of HPV types targeted by the nine-valent vaccine (RR, 1.85; 95% CI, 1.07 – 3.20).[39] By contrast, diversity measured by prevalent BV bacteria and *Gardnerella vaginalis* appeared to significantly decrease the risk of multiple hrHPVs, HPV 16 or 18 and 9-valent HPV (RR range, 0.50 – 0.71).[39]

HPV Acquisition

Temporal associations between the CVM and HPV acquisition were assessed by three longitudinal studies and one nested-case-control study; generally *Lactobacillus* communities were negatively associated with HPV incidence (effect estimate range, 0.125 – 0.910).[26,40,44,45] The aforementioned range excludes findings for a community dominated by *Lactobacillus iners*, which appeared to increase the risk of an incident HPV infection (adjusted transition rate ratio (aTRR), 1.79; 95% CI, 0.71 – 4.51).[45] A decrease in the risk of hrHPV was found among HIV-positive populations with a VMB dominated by *Lactobacillus crispatus* or *Lactobacillus jensenii* (RRR, 0.125; p = 0.019),[40] and an increasing RA of *Lactobacillus crispatus* (odds ratio (OR), 0.91; 0.84 – 1.01).[26] Similarly, a transition to a CST dominated by *Lactobacillus crispatus* appeared to reduce the risk of an incident HPV infection (aTRR, 0.20; 95% CI, 0.03 – 1.14) among HIV-positive and HIV-negative subjects,[44] whereas among the general population, a CST dominated by *Lactobacillus gasseri* was negatively associated with a transition from an HPV negative to a positive state (aTRR, 0.34; 95% CI, 0.06 – 1.85).[45]

HPV Persistence

Longitudinal and nested-case-control studies that assessed the relation between CVM and HPV persistence found associations between *Lactobacillus* species as well as microbial diversity and hrHPV persistence.[31,33,37,40] The directionality of effect estimates for *Lactobacillus* species varied across studies and populations. A significant negative relationship was found between *Lactobacillus* species (≥ 70%) and hrHPV persistence (aOR, 0.35; 95% CI, 0.14 – 0.89) in HIV-negative women, whereas in HIV-positive women, a non-significant positive association with the odds of hrHPV persistence was evident (aOR; 1.25; 95% CI, 0.73 – 2.14).[37] Conversely, in HIV-positive women, a community dominated by *Lactobacillus crispatus* or *Lactobacillus jensenii* was associated with a decreased risk of type-specific hrHPV persistence (RRR, 0.315, P = 0.074).[40] At the species level, among women with normal or ASCUS cytology, an increasing RA

of *Lactobacillus johnsonii* appeared to significantly increase the odds of hrHPV persistence (aOR, 16.4; 95% CI, 1.77 – 152.2).[33] Microbial diversity was generally indicative of an increased risk of persistence. CST IV-BV (*Lactobacillus* species depleted with aerobic and anaerobic bacteria) was significantly associated with an increase in the odds of hrHPV persistence (aOR, 9.38; 95% CI, 1.85 – 47.52).[31] Moreover, in women living with HIV, an increasing RA of BV anaerobes and the Simpson index were associated with an increased risk of type-specific hrHPV persistence.[40] Conversely, CST IV-AV (*Lactobacillus* species depleted with strictly anaerobic bacteria) appeared to decrease the odds of hrHPV persistence (aOR, 0.11; 95% CI, 0.01 – 0.93).[31]

HPV Clearance

Four studies using longitudinal data (cohort or nested-case-control studies) found significant associations between CVM components and HPV clearance.[29,33,40,45] Two studies found lower diversity or *Lactobacillus* dominance to be significantly associated with HPV clearance.[29,45] Usyk et al. assessed this association using a molecular BV (*molBV*) score, where higher scores correspond to increased dysbiosis. Individuals who transitioned from a high to low *molBV* score appeared to have a lower likelihood of hrHPV clearance relative to those with a consistently low score (adjusted hazard ratio; 0.55, 95% CI 0.30 – 0.97).[29] Similarly, a CST dominated by *Lactobacillus gasseri* was significantly associated with faster HPV clearance than one dominated by *Lactobacillus crispatus* (aTRR, 4.43; 95% CI 1.11 – 17.7).[45] Conversely, among HIV-positive women, microbial diversity appeared to significantly increase the likelihood of hrHPV clearance by almost four times (RRR, 3.856; P = 0.034).[40] One study identified individual species to be significantly associated with increased odds of hrHPV clearance: *Eubacterium eligen* (aOR, 11.5; 95% CI, 1.31 – 101.4)*, Gardnerella vaginalis* (aOR, 17.0; 95% CI, 2.18 – 131.8), and *Ureaplasma urealyticum* (aOR, 7.42; 1.30 – 42.5).[33]

Associations with CIN and Cervical Cancer

Four of five cross-sectional studies assessing associations between the CVM and biopsy-confirmed lesions identified microbial components that significantly increased the risk of CIN or CIN2+ lesions.[27,28,32,34,43] CST type was not a significant predictor of CIN lesions (aOR, 1.09; 95% CI 0.85 – 1.41),[27] and the odds of CIN were approximately six times greater in women with

a risky microbial pattern (defined by dominance of *Atopobium vaginae, Lactobacillus iners, Gardnerella vaginalis* and depletion of *Lactobacillus crispatus*) (aOR, 5.80; 95% CI 1.73 – 19.4) or a high RA of *Atopoium vaginae* (aOR, 6.63; 95% CI 1.61 – 27.2).[34] Dominance of *Lactobacillus iners* and unclassified *Lactobacilli* appeared to significantly increase the odds of CIN2+ (aOR, 3.48; 95% CI 1.27 – 9.55).[43] One study identified *Atopobium vaginae, Dialister invisus, Finegoldia magna, Gardnerella vaginalis, Prevotella buccalis* and *Prevotella timonensis* as bacterial species significantly associated with an increase in the odds of CIN2+,[32] whereas another suggested that *Atopobium vaginae* and *Pseudomonas stutzeri* were associated with a significant decrease in the odds.[28]

Findings varied across the four cross-sectional studies that assessed the relation between the CVM and cytological interpretations.[35,36,39,41] A diverse CST appeared to increase the odds of abnormal cytology results (aOR, 1.63; 95% CI 0.66 – 4.03),[41] and high-grade squamous intra-epithelial lesions (SIL) or invasive cervical cancer (ICC) (aOR, 1.31; 95% CI 0.39 – 4.41).[36] Similarly, microbial diversity (measured by an increasing Shannon index and PD whole tree) appeared to non-significantly increase the odds of SIL or cervical cancer by over three times.[35] One study identified a CST dominated by *Lactobacillus iners* as protective against abnormal cytology (aOR, 0.67; 95% CI 0.28 – 1.59),[41] whereas another suggested it was pathogenic and increased the odds of high-grade SIL or ICC by approximately 13%.[36] Only one study's findings – with assessment at the genus and species level – were statistically significant; *Lactobacillus* species, *Gardnerella vaginalis, Atopbium vaginae* and *Mobiluncus* species were significantly associated with a decreased risk of cytological abnormalities.[39]

Four longitudinal studies identified a significant relationship between high bacterial diversity or depletion of *Lactobacillus* species and progression or regression of CIN2+ lesions.[29,30,40,42] High diversity - represented by an increasing *molBV* score, *Lactobacillus* depletion, the RA of BV-anaerobes, the Shannon index, and the Simpson index – appeared to significantly increase the odds of CIN2+ prevalence, progression to CIN2+, non-regression of CIN2, or incident CIN2+ (effect estimate range, 1.17 - 7.352).[29,30,40,42] Accordingly, increasing baseline abundance of *Lactobacillus* was associated with a 59% decrease in the odds of CIN2+ lesion progression (aOR, 0.41; 95% CI 0.22 – 0.79).[30] In contrast to findings suggesting that an increase in diversity and reduction of *Lactobacilli* increase the risk of CIN2+, high-diversity BV

anaerobes were significantly associated with a higher chance of CIN2+ clearance among women living with HIV.[40]

Systematic Reviews

Table 3-2 summarizes the main findings, which were consistent, from the four included systematic reviews published between 2018 and 2022.[46–49] A low abundance of *Lactobacillus* was related to HPV prevalence, and low levels of *Lactobacillus jensenii* and *crispatus* as well as high levels of *Lactobacillus iners* and *Lactobacillus* species, have been found in CIN and cervical cancer.[49]Another review that investigated the relationship between the vaginal microbiome and HPV infections found that high microbial diversity was related to CIN lesions and HPV acquisition, prevalence, and persistence but slower clearance of HPV.[46] Sims et al. and Gardella et al. evaluated the role of the CVM in CIN and cervical cancer.[47,48] Generally, both reviews suggested that dysbiosis and a reduction in *Lactobacillus* species were potentially pathogenic via their association with either HPV acquisition, prevalence, persistence, clearance, and cytology or biopsy-confirmed CIN or cervical cancer.

Meta-Analyses

Table 3-3 details the main components from all three meta-analyses published in 2019.[18–20] One meta-analysis provided separate effect estimates from studies that used microscopic and molecular characterization techniques; based on findings from the two molecular studies (with no heterogeneity; $I^2 = 0.0\%$), low *Lactobacilli* anaerobes appeared to significantly increase the odds of a persistent HPV infection (relative risk (RR), 2.00; 95% CI, 1.05 – 3.81).[18] Two meta-analyses assessed the relationship between CVM and HPV prevalence. The first meta-analysis found that low *Lactobacillus* species appeared to significantly increase the odds of an HPV infection based on 20 included studies (with substantial heterogeneity; $I^2 = 66.1\%$) (OR, 1.53; 95% CI 1.23 – 1.82).[19] Similarly, the second, based on nine included studies (with low heterogeneity; $I^2 = 6.0\%$), found a CST dominated by *Lactobacillus* species to be significantly protective against hrHPV (OR, 0.64; 95% CI 0.48 – 0.87).[20] Moreover, based on 8 included studies (with low heterogeneity; $I^2 = 10.0\%$), the same meta-analysis found a *Lactobacillus crispatus* dominant CST to be associated with lower odds of hrHPV prevalence. With respect to CIN lesions, a *Lactobacillus* dominant CST (n = 6 studies, with no heterogeneity; $I^2 = 0.0\%$) and

a CST dominated by *Lactobacillus crispatus* (n = 5 studies, with no heterogeneity; $I^2 = 0.0\%$) was significantly associated with a decreased odds of CIN lesions by approximately 47% and 50%, respectively.[20] In the same meta-analysis, a *Lactobacillus* dominant CST was significantly protective against cervical cancer based on findings from 3 studies (with no heterogeneity; $I^2 = 0.0\%$). None of the identified meta-analyses assessed the association between CVM and HPV clearance. Brusselaers et al. showed that non-*Lactobacillus crispatus* dominance may increase the risk of HPV acquisition by almost 2-fold based on two molecular studies (with moderate heterogeneity; $I^2 = 55.7\%$); however, the estimate was not statistically significant.[18]

Diagnostic ability of the CVM in HPV-associated cervical carcinogenesis

Three cross-sectional studies, published in 2020 or 2021, assessed the diagnostic performance of CVM components to detect HPV (n = 2)[50,51] or high-grade cervical lesions (n = 1),[52] the details of which are presented in **Table 3-4**. All studies detected the outcomes of interest with high accuracy, as demonstrated by their respective AUCs. Among 546 women of reproductive age in Brazil, 30 bacterial taxa were strongly correlated with hrHPV (AUC = 0.802; 95% CI, 0.752 – 0.853).[50] Similarly, among 52 women attending a colposcopy clinic following screening for cervical cancer, 17 genera (AUC = 0.819; 95% CI, 0.684 – 0.954) and seven species (AUC = 0.918; 95% CI, 0.839-0.997) had good clinical performance to detect HPV16 positivity.[51] The abundance of 33 bacterial species was extremely able to discriminate between CIN2+ and CIN1- lesions among 66 women in Korea (AUC = 0.952; 95% CI, 0.820-1.000).[52]

16S rRNA Gene Sequencing vs. WMS

Most research articles in the narrative synthesis utilized 16S rRNA gene sequencing (n = 20) for CVM characterization. **Table 3-5** compares the identification ability, processes, advantages, and disadvantages of 16S rRNA gene sequencing with those of WMS. For CVM characterization, both techniques begin with the collection of a cervical or vaginal sample and extraction of genomic DNA. 16S rRNA gene sequencing relies on amplifying different hypervariable region(s) (i.e. amplicon) of the 16S rRNA bacterial gene (a subunit of a ribosome found in all bacteria and archaea) using polymerase chain reaction, whereas WMS relies on all genomic DNA. High-throughput sequencing analyzed with a bioinformatics pipeline allows for assessing a sample's bacterial (16S rRNA) and microbial (WMS) diversity. The major difference

between the two techniques is that WMS is more comprehensive, as it has the ability to identify the complete microbiome, whereas 16S rRNA gene sequencing can only detect archaea and bacteria that contain the primer set specific sequences in their genome. Nevertheless, after sequencing the CVM with both WMS and 16S rRNA, both techniques identify the bacteriome.[53] These methods have numerous advantages and disadvantages. Although more comprehensive with respect to the microbiome, WMS is expensive and generates large amounts of meaningless data. By contrast, 16S rRNA gene sequencing, despite limited taxonomic resolution, is cost-effective and well-established as a starting point for microbial identification.

3.2.5. Discussion

Findings from the current review may aid in understanding associations between the CVM and HPV prevalence, acquisition, persistence, clearance, and pre-cancerous lesions and cervical cancer. To the best of our knowledge, this is the first review on the topic that only included studies that used metagenomic methods without restricting to specific microbes and summarized the diagnostic performance of CVM communities. Our findings from the assessment of individual studies were aligned with previous systematic reviews assessing similar relationships.[46–49]Although associations were not consistent across included studies, we identified a general trend of *Lactobacillus* communities protecting against adverse outcomes, whereas high diversity appeared pathogenic in HPV-associated cervical carcinogenesis. Additionally, distinct combinations of microbial components were strongly correlated with and able to detect HPV16,[51] hrHPV[50] and CIN lesions.[52]

In our review, several species appear to be pathogenic or protective against HPV-associated cervical carcinogenesis, some of which overlapped. At the level of individual species, *Lactobacillus crispatus,*[33,44] *Eubacterium eligen,*[33] *Gardnerella vaginalis,*[33,39] *Ureaplasma urealyticum,*[33] *Pseudomonas stutzeri,*[28] and *Atopbium vaginae*[28,39] were significantly protective across the outcomes of interest suggesting a clear association. On the other hand, *Atopbium vaginae*,[32] *Dialister invisus*,[32] *Finegoldia magna*,[32] *Gardnerella vaginalis*,[32] *Prevotella buccalis*,[32] *Prevotella timonensis*,[32] and *Lactobacillus johnsonii*[33] were significantly pathogenic. Although the aforementioned list is not comprehensive and estimates were presented for additional species, their role remains inconclusive as estimates were not statistically significant. Conflicting findings at the species level suggest that a microbial community rather than a single

species may contribute to cervical carcinogenesis. Alternatively, this could highlight a need for a comprehensive assessment of bacterial species using high-resolution characterization methods. Nevertheless, a communal role is supported by the observation that a CVM dominated by *Lactobacillus* species is often significantly protective[20,30,37,39,40] and high diversity (or a loss of *Lactobacillus* dominance) significantly pathogenic in HPV prevalence, acquisition, persistence, clearance, and cervical lesions and cervical cancer.[18,19,29–31,34,40,42] Furthermore, *Lactobacillus* species sustain a low vaginal pH due to the production of lactic acid,[54] and some *Lactobacilli* produce lactocin (a bacteriocin).[55] Both of these substances may help to maintain a healthy vaginal environment, further contributing to the plausibility of a communal role in cervical carcinogenesis.

The CVM of reproductive-aged women is thought to be dominated by *Lactobacillus* species and generally represents a healthy state, whereas diversity indicates a diseased one.[9,10] However, CVM composition has been shown to vary by ethnicity. Among women in Amsterdam, a high diversity CVM consisting of *Gardnerella vaginalis* was more common in those of sub-Saharan African ancestry than Dutch.[56] Hence, a unique genetic or environmental factor pertaining to country of origin may impact the predisposition of the presence and favoured harbouring of certain CVM microbes. The identified studies looking at the association between the CVM and HPV-associated cervical carcinogenesis were conducted in various countries and, in some cases, specific ethnicities; it is plausible to assume that ethnicity varied greatly between studies. The inconsistency in effect estimates and the lack of inter-study reproducibility may result from incomparable study populations concerning country of origin and differing microbial compositions.

Several screening modalities have been implemented for cervical cancer. Historically, cervical cytology was the paradigm for screening. Despite its high-test specificity for the detection of pre-cancerous lesions (96-98%), its sensitivity is poor (51-53%),[57–59] highlighting a need for repeat testing to minimize false negatives. Comparatively, HPV DNA testing has been shown to have a higher sensitivity (94.6%) and slightly lower specificity (94.1%) for detecting high-grade lesions.[60] By combining cytology and HPV primary testing, the specificity or sensitivity are maximized. We identified three studies suggesting that the presence and/or RA of bacterial taxa can robustly detect HPV or CIN lesions. Thus, microbial testing could be a feasible additional and cost-effective option for cervical cancer screening. Nevertheless, further research

is required to identify specific CVM components, including reproducibility in longitudinal data to establish causality.

Most observational studies we identified utilized 16S rRNA gene sequencing to identify microbial communities and explore associations with the outcomes of interest. Variation exists within 16S rRNA gene sequencing, and the identified observational studies used different sequencing platforms and bioinformatics annotation pipelines and amplified at various hypervariable regions. It is reasonable to hypothesize that both the diversity in the identified bacterial communities and variation in CVM exposure assignment across investigations could be attributable to these methodological differences. Few studies were able to examine the same relationships between the CVM and outcomes of interest due to varying exposures. Thus, to enable scientific replicability and confirm associations across studies, it may be important to establish a harmonious CVM characterization technique to identify similar bacterial communities. Standardization of the methodological procedures within 16S rRNA gene sequencing may be a feasible option. Although this method is not the most comprehensive for microbial identification as it targets bacterial communities,[24] we identified literature showing that it detects similar CVM microbial species as WMS.[53] The latter technique can identify the complete microbiome.[24] This detection comparability suggests that bacteria may be the main constituent of the CVM. As such, the cost-effectiveness and efficiency of 16S rRNA gene sequencing render it a potential option for CVM characterization.

Our review comprehensively summarized the current literature. However, there are several considerations to be taken into account. Among the included observational studies, there was a large variation in sample size, HPV genotyping methods, microbial characterization methods, and covariate adjustment. Exposure misclassification may have arisen in studies using 16S rRNA gene sequencing due to limited taxonomic resolution and errors across 16S reference databases. Thus, some participants may have been mistakenly categorized with respect to their microbiome exposure, which may bias the directionality of effect estimates. Our findings may also be limited by the inclusion of several cross-sectional studies. In these studies, temporal changes in CVM composition were not observed; thus, causation cannot be established. Moreover, two of the longitudinal studies only characterized the CVM in baseline samples; changes in microbial variability were not examined over time concerning hrHPV persistence and CIN lesion regression.[31,42] Consequently, there is a need for more longitudinal studies to infer

causality or explore whether the CVM could mediate the relationship between HPV infections and cervical cancer. We may have missed identifying important literature by excluding abstracts and articles published in languages other than English. Nevertheless, our search was developed with the assistance of a librarian, and we comprehensively searched three databases.

To conclude, evidence suggests that the CVM may be involved in HPV-associated cervical carcinogenesis among studies that characterized the CVM using metagenomics. However, conflicting findings across the literature highlight a need for further research focusing on standardizing techniques for CVM characterization and sampling approaches to control for intra-individual CVM variation over time.

3.2.6. Declaration of Interests and Sources of Funding

ELF reports grants and/or consulting fees from Merck and the Canadian Institutes of Health Research (CIHR), outside of the submitted work. ELF and MZ hold a patent related to the discovery of "DNA methylation markers for early detection of cervical cancer", registered at the Office of Innovation and Partnerships, McGill University, Montreal, Quebec, Canada (October 2018). ELF is an Advisory Board Member of Cancer Research Organizations (Camargo Cancer Center, Charbonneau Cancer Institute, Universidade do Minho). ELF receives an honorarium as an editor for various journals (Oxford University Press, Elsevier and Elifesciences Ltd). MZ receives an honorarium from Elsevier as an Associate Editor for Preventive Medicine. ML and PRT receive an honorarium from Elsevier as an Assistant Editor for Preventive Medicine and Preventive Medicine Reports. ML was awarded a master's fellowship by Fonds de Recheré du Québec – Santé (FRQS). EG has nothing to disclose. This work was supported by a grant from the Canadian Institutes of Health Research (grant FDN-143347 to ELF).

3.2.7. Acknowledgements

We would like to acknowledge Genevieve Gore (Liason Librarian, Schulich Library of Physical Sciences, Life Sciences, and Engineering) for helping us with the development of the search strategies. We also thank Mykhaylo Usyk for consulting and providing insights on this project.

3.2.8. Author Contributions

ML, MZ and ELF formulated the research question. ML created the searches and performed primary screening for the epidemiological associations. ML and PT performed secondary screening and extracted data from relevant records. ML performed all screening and extraction for the diagnostic performance. ML drafted the manuscript under the supervision of EG, MZ and ELF. PT, MZ, EG and ELF reviewed and amended the manuscript. All authors read and approved the final version of the manuscript.

3.2.9. Manuscript 1 References

1. WHO. Comprehensive Cervical Cancer Control. Geneva. 2014;366–78.

2. Walboomers JMM, Jacobs M V., Manos MM, Bosch FX, Kummer JA, Shah K V., et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. J Pathol. 1999;189(1):12–9.

3. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021 May 4;71(3):209–49.

4. Canadian Partnership Against Cancer. HPV immunization for the prevention of cervical cancer [Internet]. 2021. Available from: https://www.partnershipagainstcancer.ca/topics/hpv-immunization-policies/

5. Kaiser Family Foundation. The HPV Vaccine: Access and Use in the U.S. | KFF [Internet]. 2021. Available from: https://www.kff.org/womens-health-policy/fact-sheet/the-hpv-vaccine-access-and-use-in-the-u-s/

6. Luqman A. The orchestra of human bacteriome by hormones. Microb Pathog. 2023 Jul;180:106125.

7. Stern J, Miller G, Li X, Saxena D. Virome and bacteriome: two sides of the same coin. Curr Opin Virol. 2019 Aug;37:37–43.

8. Smith WL, Hedges SR, Mordechai E, Adelson ME, Trama JP, Gygax SE, et al. Cervical and Vaginal Flora Specimens Are Highly Concordant with Respect to Bacterial Vaginosis-Associated Organisms and Commensal Lactobacillus Species in Women of Reproductive Age. Forbes BA, editor. J Clin Microbiol. 2014 Aug;52(8):3078–81.

9. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SSK, McCulle SL, et al. Vaginal microbiome of reproductive-age women. Proc Natl Acad Sci U S A. 2011 Mar 15;108(SUPPL. 1):4680–7.

10. Chen X, Lu Y, Chen T, Li R. The Female Vaginal Microbiome in Health and Bacterial Vaginosis. Front Cell Infect Microbiol. 2021 Apr 7;11:631972.

11. Gallo MF, Macaluso M, Warner L, Fleenor ME, Hook EW, Brill I, et al. Bacterial Vaginosis, Gonorrhea, and Chlamydial Infection Among Women Attending a Sexually Transmitted Disease Clinic: A Longitudinal Analysis of Possible Causal Links. Ann Epidemiol. 2012 Mar;22(3):213–20.

12. Cogliano V, Baan R, Straif K, Grosse Y, Secretan B, Ghissassi F El. Carcinogenicity of human papillomaviruses. Lancet Oncol. 2005 Apr;6(4):204.

13. Muñoz N, Bosch FX, Castellsagué X, Díaz M, de Sanjose S, Hammouda D, et al. Against

which human papillomavirus types shall we vaccinate and screen? the international perspective. Int J Cancer. 2004 Aug 20;111(2):278–85.

14. Castellsagué X. Natural history and epidemiology of HPV infection and cervical cancer. Gynecol Oncol. 2008 Sep;110(3):S4–7.

15. Richardson H, Kelsall G, Tellier P, Voyer H, Abrahamowicz M, Ferenczy A, et al. The natural history of type-specific human papillomavirus infections in female university students. Cancer Epidemiol Biomarkers Prev. 2003 Jun;12(6):485–90.

16. Schiffman M, Castle PE, Jeronimo J, Rodriguez AC, Wacholder S. Human papillomavirus and cervical cancer. Lancet. 2007 Sep;370(9590):890–907.

17. Gravitt P, Winer R. Natural History of HPV Infection across the Lifespan: Role of Viral Latency. Viruses. 2017 Sep 21;9(10):267.

18. Brusselaers N, Shrestha S, van de Wijgert J, Verstraelen H. Vaginal dysbiosis and the risk of human papillomavirus and cervical cancer: systematic review and meta-analysis. Am J Obstet Gynecol. 2019 Jul;221(1):9-18.e8.

19. Tamarelle J, Thiébaut ACM, de Barbeyrac B, Bébéar C, Ravel J, Delarocque-Astagneau E. The vaginal microbiota and its association with human papillomavirus, Chlamydia trachomatis, Neisseria gonorrhoeae and Mycoplasma genitalium infections: a systematic review and meta-analysis. Clin Microbiol Infect. 2019 Jan;25(1):35–47.

20. Wang H, Ma Y, Li R, Chen X, Wan L, Zhao W. Associations of Cervicovaginal Lactobacilli With High-Risk Human Papillomavirus Infection, Cervical Intraepithelial Neoplasia, and Cancer: A Systematic Review and Meta-Analysis. J Infect Dis. 2019 Sep 13;220(8):1243–54.

21. Nugent RP, Krohn MA, Hillier SL. Reliability of diagnosing bacterial vaginosis is improved by a standardized method of gram stain interpretation. J Clin Microbiol. 1991 Feb;29(2):297–301.

22. Amsel R, Totten PA, Spiegel CA, Chen KCS, Eschenbach D, Holmes KK. Nonspecific vaginitis. Diagnostic criteria and microbial and epidemiological associations. Am J Med. 1983 Jan;74(1):14–22.

23. Platz-Christensen JJ, Larsson P-G, Sundström E, Wlqvlst N. Detection of bacterial vaginosis in wet mount, Papanicolaou stained vaginal smears and in Gram stained smears. Acta Obstet Gynecol Scand. 1995 Jan;74(1):67–70.

24. Galloway-Peña J, Hanson B. Tools for Analysis of the Microbiome. Dig Dis Sci. 2020 Mar 31;65(3):674–85.

25. Siqueira J, Curty G, Xutao D, Hofer C, Machado E, Seuánez H, et al. Composite Analysis

of the Virome and Bacteriome of HIV/HPV Co-Infected Women Reveals Proxies for Immunodeficiency. Viruses. 2019 May 7;11(5):422.

26. McClymont E, Albert AY, Wang C, Dos Santos SJ, Coutlée F, Lee M, et al. Vaginal microbiota associated with oncogenic HPV in a cohort of HPV-vaccinated women living with HIV. Int J STD AIDS. 2022 Aug 1;33(9):847–55.

27. Zhang Y, Xu X, Yu L, Shi X, Min M, Xiong L, et al. Vaginal Microbiota Changes Caused by HPV Infection in Chinese Women. Front Cell Infect Microbiol. 2022;12:814668.

28. 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 Med. 2018 May 2;7(5):2172–9.

29. Usyk M, Schlecht NF, Pickering S, Williams L, Sollecito CC, Gradissimo A, et al. molBV reveals immune landscape of bacterial vaginosis and predicts human papillomavirus infection natural history. Cortes B Jimenez SE, Rodriguez AC, Hildesheim A, Kreimer AR, Lowy DR, Schiffman M, Schiller JT, Sherman M, Wacholder S, Kemp TJ, Sidawy MK, Quint W, van Doorn LJ, Struijk L, Palefsky JM, Darragh TM, Stoler MH GP, editor. Nat Commun. 2022;13(1):233.

30. Usyk M, Zolnik CP, Castle PE, Porras C, Herrero R, Gradissimo A, et al. Cervicovaginal microbiome and natural history of HPV in a longitudinal study. Silvestri G, editor. PLOS Pathog. 2020 Mar 26;16(3):e1008376.

31. 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. Sci Rep. 2017 Dec 31;7(1):10200.

32. So KA, Yang EJ, Kim NR, Hong SR, Lee J-H, Hwang C-S, et al. Changes of vaginal microbiota during cervical carcinogenesis in women with human papillomavirus infection. Consolaro MEL, editor. PLoS One. 2020 Sep 17;15(9):e0238705.

33. 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. Sci Rep. 2018 Oct 19;8(1):15479.

34. Oh HY, Kim B-S, Seo S-S, Kong J-S, Lee J-K, Park S-Y, et al. The association of uterine cervical microbiota with an increased risk for cervical intraepithelial neoplasia in Korea. Clin Microbiol Infect. 2015 Jul;21(7):674.e1-674.e9.

35. 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. Tornesello ML, editor. PLoS One. 2016 Apr 26;11(4):e0153274.

36. Musa J, Maiga M, Green SJ, Magaji FA, Maryam AJ, Okolo M, et al. Vaginal microbiome community state types and high-risk human papillomaviruses in cervical precancer and cancer in North-central Nigeria. BMC Cancer. 2023 Jul 20;23(1):683.

37. Dareng EO, Ma B, Adebamowo SN, Famooto A, Ravel J, Pharoah PP, et al. Vaginal microbiota diversity and paucity of Lactobacillus species are associated with persistent hrHPV infection in HIV negative but not in HIV positive women. Sci Rep. 2020 Nov 5;10(1):19095.

38. DARENG EO, MA B, FAMOOTO AO, AKAROLO-ANTHONY SN, OFFIONG RA, OLANIYAN O, et al. Prevalent high-risk HPV infection and vaginal microbiota in Nigerian women. Epidemiol Infect. 2016 Jan 11;144(1):123–37.

39. Rosário A, Sousa A, Varandas T, Marinho-Dias J, Medeiros R, Martins G, et al. Impact of cervicovaginal microbiome on the risk of cervical abnormalities development. J Med Virol. 2023 May 3;95(5).

40. van de Wijgert JHHM, Gill AC, Chikandiwa A, Verwijs MC, Kelly HA, Omar T, et al. Human papillomavirus infection and cervical dysplasia in HIV-positive women. AIDS. 2020 Jan 1;34(1):115–25.

41. McKee KS, Carter KA, Bassis C, Young VB, Reed B, Harper DM, et al. The vaginal microbiota, high-risk human papillomavirus infection, and cervical cytology: results from a population-based study. Gynecol Pelvic Med. 2020 Sep;3:18–18.

42. Mitra A, MacIntyre DA, Ntritsos G, Smith A, Tsilidis KK, Marchesi JR, et al. The vaginal microbiota associates with the regression of untreated cervical intraepithelial neoplasia 2 lesions. Nat Commun. 2020;11(1):1–13.

43. Piyathilake CJ, Ollberding NJ, Kumar R, Macaluso M, Alvarez RD, Morrow CD. Cervical Microbiota Associated with Higher Grade Cervical Intraepithelial Neoplasia in Women Infected with High-Risk Human Papillomaviruses. Cancer Prev Res. 2016 May 1;9(5):357–66.

44. 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. J Infect Dis. 2016 Nov 1;214(9):1361–9.

45. Brotman RM, Shardell MD, Gajer P, Tracy JK, Zenilman JM, Ravel J, et al. Interplay Between the Temporal Dynamics of the Vaginal Microbiota and Human Papillomavirus Detection. J Infect Dis. 2014 Dec 1;210(11):1723–33.

46. Mortaki D, Gkegkes ID, Psomiadou V, Blontzos N, Prodromidou A, Lefkopoulos F, et al. Vaginal microbiota and human papillomavirus: a systematic review. J Turkish-German Gynecol Assoc. 2020 Sep 1;21(3):193–200.

47. Sims TT, Colbert LE, Klopp AH. The Role of the Cervicovaginal and Gut Microbiome in Cervical Intraepithelial Neoplasia and Cervical Cancer. J Immunother Precis Oncol. 2021;4(2):72–8.

48. Gardella B, Pasquali MF, La Verde M, Cianci S, Torella M, Dominoni M. The Complex Interplay between Vaginal Microbiota, HPV Infection, and Immunological Microenvironment in Cervical Intraepithelial Neoplasia: A Literature Review. Int J Mol Sci. 2022;23(13).

49. Yang X, Da M, Zhang W, Qi Q, Zhang C, Han S. Role of Lactobacillus in cervical cancer. Cancer Manag Res. 2018 May;Volume 10:1219–29.

50. Andrade Pessoa Morales J, Marconi C, El-Zein M, Ravel J, da Silva Pinto GV, Silveira R, et al. Vaginal microbiome components as correlates of cervical human papillomavirus infection. J Infect Dis. 2021;226(6).

51. Yang Q, Wang Y, Wei X, Zhu J, Wang X, Xie X, et al. The Alterations of Vaginal Microbiome in HPV16 Infection as Identified by Shotgun Metagenomic Sequencing. Front Cell Infect Microbiol. 2020 Jun 23;10.

52. Lee YH, Kang G-U, Jeon SY, Tagele SB, Pham HQ, Kim M-S, et al. Vaginal Microbiome-Based Bacterial Signatures for Predicting the Severity of Cervical Intraepithelial Neoplasia. Diagnostics. 2020 Nov 26;10(12):1013.

53. Bloom SM, Mafunda NA, Woolston BM, Hayward MR, Frempong JF, Abai AB, et al. Cysteine dependence of Lactobacillus iners is a potential therapeutic target for vaginal microbiota modulation. Nat Microbiol. 2022 Mar 3;7(3):434–50.

54. Tachedjian G, Aldunate M, Bradshaw CS, Cone RA. The role of lactic acid production by probiotic Lactobacillus species in vaginal health. Res Microbiol. 2017 Nov;168(9–10):782–92.

55. Turovskiy Y, Ludescher RD, Aroutcheva AA, Faro S, Chikindas ML. Lactocin 160, a Bacteriocin Produced by Vaginal Lactobacillus rhamnosus, Targets Cytoplasmic Membranes of the Vaginal Pathogen, Gardnerella vaginalis. Probiotics Antimicrob Proteins. 2009 Jun 29;1(1):67–74.

56. Borgdorff H, van der Veer C, van Houdt R, Alberts CJ, de Vries HJ, Bruisten SM, et al. The association between ethnicity and vaginal microbiota composition in Amsterdam, the Netherlands. Fredricks DN, editor. PLoS One. 2017 Jul 11;12(7):e0181135.

57. Nanda K, McCrory DC, Myers ER, Bastian LA, Hasselblad V, Hickey JD, et al. Accuracy of the Papanicolaou Test in Screening for and Follow-up of Cervical Cytologic Abnormalities. Ann Intern Med. 2000 May 16;132(10):810.

58. Cuzick J, Clavel C, Petry K-U, Meijer CJLM, Hoyer H, Ratnam S, et al. Overview of the European and North American studies on HPV testing in primary cervical cancer screening. Int J Cancer. 2006 Sep 1;119(5):1095–101.

59. Tota JE, Bentley J, Blake J, Coutlée F, Duggan MA, Ferenczy A, et al. Introduction of molecular HPV testing as the primary technology in cervical cancer screening: Acting on evidence to change the current paradigm. Prev Med (Baltim). 2017 May;98:5–14.

60. Mayrand M-H, Duarte-Franco E, Rodrigues I, Walter SD, Hanley J, Ferenczy A, et al. Human Papillomavirus DNA versus Papanicolaou Screening Tests for Cervical Cancer. N Engl J Med. 2007 Oct 18;357(16):1579–88.

61. Mas-Lloret J, Obón-Santacana M, Ibáñez-Sanz G, Guinó E, Pato ML, Rodriguez-Moranta F, et al. Gut microbiome diversity detected by high-coverage 16S and shotgun sequencing of paired stool and colon sample. Sci Data. 2020 Mar 16;7(1):92.

62. Brumfield KD, Huq A, Colwell RR, Olds JL, Leddy MB. Microbial resolution of whole genome shotgun and 16S amplicon metagenomic sequencing using publicly available NEON data. Gyarmati P, editor. PLoS One. 2020 Feb 13;15(2):e0228899.

63. Ranjan R, Rani A, Metwally A, McGee HS, Perkins DL. Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. Biochem Biophys Res Commun. 2016 Jan;469(4):967–77.

64. Clooney AG, Fouhy F, Sleator RD, O' Driscoll A, Stanton C, Cotter PD, et al. Comparing Apples and Oranges?: Next Generation Sequencing and Its Impact on Microbiome Analysis. White BA, editor. PLoS One. 2016 Feb 5;11(2):e0148028.

65. Gonzalez E, Pitre FE, Brereton NJB. ANCHOR: a 16S rRNA gene amplicon pipeline for microbial analysis of multiple environmental samples. Environ Microbiol. 2019;21(7):2440–68.

66. ilumina. Sequence hundreds of organisms with the power of multiplexing [Internet]. Available from: https://www.illumina.com/areas-of-interest/microbiology/microbial-sequencing-methods/microbial-whole-genome-sequencing.html#:~:text=Introduction to Microbial Whole-Genome Sequencing&text=Sequencing entire bacterial%2C viral%2C and,for other comparative ge

67. AAT Bioquest. What are the disadvantages of WGS? [Internet]. Available from: https://www.aatbio.com/resources/faq-frequently-asked-questions/what-are-the-disadvantages-of-whole-genome-sequencing

68. Zymo Research. 16S Sequencing VS Shotgun Metagenomic Sequencing [Internet]. Available from: https://zymoresearch.eu/blogs/blog/16s-sequencing-vs-shotgun-metagenomic-sequencing

69. Ruairi Robertson. 16S rRNA Gene Sequencing vs. Shotgun Metagenomic Sequencing [Internet]. Available from: https://blog.microbiomeinsights.com/16s-rrna-sequencing-vs-shotgun-metagenomic-sequencing

3.2.10. Manuscript 1 Tables

Table 3-1. Observational studies on the association between the CVM and HPV-associated cervical carcinogenesis.

First Author (Year)	Study design (number of visits for microbial characterization)	Country	Outcome(s) assessed[a]	Main findings
Musa (2023)[36]	Cross-sectional	Nigeria	Cytology or biopsy results	-No significant association found between any CST types and HSIL/ICC versus NILM and LSIL cytology in the entire study population or following stratification by HPV positivity
Rosário (2023)[39]	Cross-sectional	Portugal	HPV prevalence; cytology or biopsy results	-*Lactobacillus* species, *Gardnerella vaginalis*, and BV panel found to be significantly protective against the presence of multiple hrHPVs, HPV16 or HPV18, and 9-valent HPVs whereas *Mobiluncus* species significantly increased the risk of 9-valent HPVs Following co-variate adjustment, in three multivariable models: -*Lactobacillus* species, *Gardnerella vaginalis*, *Atopobium vaginae*, and *Mobiluncus* species were significantly protective against cervical abnormalities versus NILM -*Lactobacillus* species and *Gardneralla vaginalis* were significantly protective against cervical lesions versus NILM and ASC-US -*Lactobacillus* species, *Gardnerella vaginalis*, *Atopobium vaginae*, and *Mobiluncus* species were significantly protective against HSIL/ASC-H lesions versus NILM and ASC-US and LSIL
McClymont (2022)[26]	Longitudinal (max 3 visits)[b]	Canada	HPV acquisition	-No significant association found between increasing RA of *Gardnerella swidinski* or *Lactobacillus crispatus* and the odds of incident hrHPV among HIV+ women
Usyk (2022)[29]	Longitudinal (2 visits)	Costa Rica	HPV clearance; cytology or biopsy results	-Transition from a high to low *molBV* score had a significantly lower likelihood of hrHPV clearance relative to those with a consistently low score -Higher *molBV* score significantly increased the odds of hrHPV progression to CIN2+ lesions
Zhang (2022)[27]	Cross-sectional	China	HPV prevalence; cytology or biopsy results	-CST types found significantly protective against hrHPV/lrHPV prevalence -No significant association found between CST types and CIN lesions
Dareng (2020)[37]	Nested-case-control (2 visits)	Nigeria	HPV persistence	-*Lactobacillus* dominance significantly decreased the odds of hrHPV persistence in HIV- women -*Lactobacillus* dominance increased the odds of hrHPV persistence in HIV+ women
McKee (2020)[41]	Cross-sectional	United States	HPV prevalence; cytology or biopsy results	-No significant association between any CST type and the odds of hrHPV positivity or abnormal cytology -No significant association between the presence of *Lactobacillus gasseri* and hrHPV positivity or abnormal cytology following co-variate adjustment
Mitra (2020)[42]	Longitudinal (3 visits)[c]	United States	Cytology or biopsy results	-*Lactobacillus* depleted microbiome significantly associated with increased odds of non-regression of CIN2 lesions after 12 and 24 months of follow up
So (2020)[32]	Cross-sectional	Korea	Cytology or biopsy results	-Presence of *Atopobium vaginae, Dialister invisus, Finegoldia magna, Gardnerella vaginalis, Prevotella buccalis*, and *Prevotella timonensis* significantly increased the odds of CIN2+ lesions or cervical cancer

Usyk (2020)[30]	Longitudinal (2 visits)	Costa Rica	Cytology or biopsy results	-Increasing baseline RA of *Lactobacillus* significantly decreased the odds of progression to CIN2+ lesions -High microbial diversity at visit 2 significantly increased the odds of progression to CIN2+ lesions
Van de Wijgert (2020)[40]	Two nested-case-control studies (Both consisted of 2 visits)[b]	South Africa	HPV acquisition, clearance, prevalence, persistence; cytology or biopsy results	Among HIV+ women: - No significant association between the Simpson index, *Lactobacillus crispatus* or *Lactobaillus jensenii* dominance, *Lactobacillus* RA, and abundance of BV anaerobes in baseline and/or samples at the end of follow up with the odds of hrHPV prevalence -A vaginal microbiome dominated by *Lactobacillus crispatus* or *Lactobacillus jensenii* in samples at the end of follow up significantly decreased the risk of an incident hrHPV infection -Higher microbial diversity, measured by the Simpson index in samples at the end of follow up, significantly increased the risk of hrHPV clearance and the risk of incident CIN2+ lesions and prevalence of CIN2+ lesions -A vaginal microbiome characterized by high diversity BV anaerobes at baseline significantly increased the risk of CIN2+ lesion clearance -Increasing BV-anaerobes RA at baseline significantly increased the risk of CIN2+ lesions at one or both visits -Increasing *Lactobacillus* RA at the end of follow-up significantly decreased the risk of CIN2+ lesions at one or both visits
Siqueira (2019)[25]	Longitudinal (2 visits)	Brazil	HPV prevalence	-Risk of HPV16 prevalence among women with a CST dominated by anaerobic bacteria not significantly different from those with a CST dominated by *Lactobacillus iners* in pregnant women coinfected with HIV and HPV
Arokiyaraj (2018)[33]	Longitudinal (2-5 visits)	Korea	HPV clearance, prevalence, persistence	-Increasing RA of *Eubacterium eligen, Gardnerella vaginalis*, and *Ureoplasma urealyticum* significantly increased the odds of hrHPV clearance -Increasing RA of *Lactobacillus crispatus* significantly increased the odds of hrHPV negativity -Increasing RA of *Lactobacillus johnsonii* significantly increased the odds of hrHPV persistence
Zhang (2018)[28]	Cross-sectional	China	Cytology or biopsy results	-High and middle *Pseudomonas stutzeri* RA significantly decreased the odds of CIN2+ lesions -High *Atopobium vaginae* RA significantly decreased the odds of CIN2+ lesions
Di Paola (2017)[31]	Longitudinal (2 visits)[c]	Italy	HPV persistence	-CST IV-BV significantly increased the odds of hrHPV persistence -CST IV-AV significantly decreased the odds of hrHPV persistence
Audirac-Chalifour (2016)[35]	Cross-sectional	Mexico	Cytology or biopsy results	-No significant association found between high microbial diversity (Shannon diversity index and PD whole tree) and the odds of squamous intraepithelial lesions and cervical cancer
Dareng (2016)[38]	Cross-sectional	Nigeria	HPV prevalence	-No significant association between any CST type and the odds of hrHPV positivity
Piyathilake (2016)[43]	Cross-sectional	United States	Cytology or biopsy results	- A community type dominated by unclassified *Lactobacillus* and *Lactobacillus iners* significantly associated with an increased risk of CIN2+ lesions among women with a hrHPV infection
Reimers (2016)[44]	Longitudinal (2-11 visits)	United States	HPV acquisition, prevalence	-Transition to a CST dominated by *Lactobacillus crispatus* significantly associated with a lower risk of incident HPV following adjustment for HIV study group -High RA of *Lactobacillus crispatus* significantly associated with lower risk of any HPV and hrHPV detection among HIV+/HIV- women
Oh (2015)[34]	Cross-sectional	Korea	Cytology or biopsy results	-Dominance of *Atopobium vaginae, Gardnerella vaginalis* and *Lactobacillus iners* as well as the depletion of *Lactobacillus crispatus* significantly associated with CIN risk -High RA of *Atopobium vaginae* alone and with hrHPV negativity significantly associated with CIN risk
Brotman (2014)[45]	Longitudinal (25-33 visits)	United States	HPV acquisition, clearance	-CST II significantly associated with more rapid clearance of HPV

Abbreviations: ASC-H, Atypical squamous cells - cannot exclude high grade squamous intraepithelial lesion; ASC-US, atypical squamous cells of undetermined significance; BV, bacterial vaginosis; CIN, cervical intraepithelial neoplasia; CST, community state type; CVM, cervicovaginal microbiome; HIV, human immunodeficiency virus; HPV, human papillomavirus; hrHPV, high-risk human papillomavirus; HSIL, high-grade squamous intraepithelial lesion; ICC, invasive cervical cancer; lrHPV, low-risk human papillomavirus; LSIL, low-grade squamous intraepithelial lesion; *molBV* molecular bacterial vaginosis; NILM, negative for intraepithelial lesion or malignancy; RA, relative abundance

[a]Cytology or biopsy results refers to cytological interpretations or biopsy confirmed CIN or cervical cancer as the outcome of interest.
[b]Longitudinal data not collected for entire study sample.
[c] Corresponds to total visits as only baseline samples were used for microbial characterization.

Table 3-2. Systematic reviews on the relationship between the CVM and HPV prevalence, acquisition, persistence, clearance and/or cytology interpretations or biopsy confirmed CIN and cervical cancer.

First Author (Year)	Focus (sample size [n] of studies)	Microbiome characterization[a]	Outcome(s) assessed[b]	Main findings[c]	Limitations acknowledged
Gardella (2022)[48]	Interplay between HPV-associated CIN, the vaginal microbiome and the immune system (n = 6)	Wet-mount microscopy, QiAMp Mini DNA kit, 16S rRNA, PowerSoil DNA Isolation Kit	HPV prevalence, persistence; cytology or biopsy results	-The following were found to be linked to an HPV infection: (1) high diversity, (2) loss of *Lactobacillus*, (3) CST IV, (4) *Shuttleworthia*, (5) *Gemella*, (6) *Olsenella*, (7) common bacterial vaginosis bacteria (*Gardnerella, Prevotella, Atopobium, Megasphaera, Parvimonas, Anaerococcus, Peptostreptococcus, Sneathia*) and (8) aerobic vaginitis and other dysbiosis microbes	-None reported
				-Relationship between dysbiosis, high diversity and/or *Lactobacillus* depletion and HPV persistence	
				-In HSIL and/or cervical cancer, (1) *Lactobacillus* depletion, (2) increase in anaerobes, and (3) bacterial vaginosis are evident -*Shuttleworthia, Gemella, Olsenella, Gardnerella, Prevotella, Atopobium, Megasphaera, Parvimonas, Anaerococcus, Peptostreptococcus, Sneathia* were related to high-grade and low-grade lesions and invasive cervical cancer -Vaginal dysbiosis may occur in low-grade and high-grade disease and *Sneathia sanguineans* may be more frequent in high- than low-grade lesions -*Lactobacillus iners* may be associated with risk of cervical dysplasia -Increased expression of *Lactobacillus jensenii* and *Lactobacillus coleohominis* is evident in LSIL patients -Risk of CST IV (high diversity) may increase with increasing lesion severity (LSIL, HSIL and invasive cancer) and risk of lesion severity was lower with CST I (*Lactobacillus crispatus* dominance) -Transition from CST IV (high diversity) to CST II (*Lactobacillus gasseri*) observed in CIN -Patients with LSIL mainly have a microbial pattern of CST I (*Lactobacillus crispatus*) -Aerobic vaginitis associated with cervical dysplastic lesions -*Sneathia* species may serve as a biomarker for CIN progression and *Delftia* species for squamous intraepithelial lesions -Inconclusive findings in the literature on the association between BV and cervical cancer development	
Sims (2021)[47]	Cervicovaginal and gut microbiome in CIN and Cervical Cancer	Gram staining, 16S rRNA, WMS	HPV prevalence, acquisition, persistence;	-*Lactobacillus crispatus* dominance associated with lower HPV prevalence -*Lactobacillus* species produce lactic acid and a depletion in their abundance could result in an inability to fight off viruses (i.e., HPV)	-None reported
				-Dysbiosis and high diversity may increase the risk of hrHPV	

	(n = 8)[d]		cytology or biopsy results	-Together, high diversity, dysbiosis, and inflammation can facilitate HPV persistence -High diversity linked to HPV persistence	
				-Reduction in *Lactobacillus* species may result in carcinogenic changes -CST of women with CIN may be: (1) *Lactobacillus* depleted, (2) dominated by anaerobic bacteria, and (3) dominated by *Lactobacillus iners* -High bacterial diversity associated with CIN lesion severity -*Sneathia* species linked to CIN lesions -Microbial composition is similar amongst BV and cervical cancer patients -*Fusobacterium* found in CIN and cervical cancer, and more common in cervical cancer	
Mortaki (2020)[46]	Relationship between the vaginal microbiome and HPV (n = 19)[e]	Microarray technology, microscopic evaluation, PCR-DEJA assay, microbiological cultures, vaginal pH, 16S rRNA, 16S rDNA	HPV prevalence, acquisition, persistence, clearance; cytology or biopsy results	-HPV positive women may have: (1) a high diversity microbiome, (2) depletion of *Lactobacillus* species, (3) lower abundance of *Lactobacillus iners and Lactobacillus crispatus,* (4) increase in anaerobic bacteria (i.e. *Prevotella* and *Leptotrichia*) -*Gardnerella vaginalis* and *Lactobacillus gasseri* are common in HPV positive patients -Studies found a relationship between *Lactobacillus crispatus* and a lower prevalence of hrHPV	-Number of existing studies and subsequently patients included in the synthesis is limited -Abstracts, reviews, conference papers, editorials, animal studies and commentaries were not considered
				-CST IV-A (high diversity) may be related to a transition from an HPV negative to an HPV positive state compared to CST I (*Lactobacillus crispatus* dominant)	
				-CST III (*Lactobacillus iners* dominant) and CSTIV-B (high diversity) may result in faster and slower clearance of HPV, respectively	
				-CSTIV-B (high diversity) may increase the risk of HPV persistence	
				-CST III (*Lactobacillus iners* dominant) has been related to severe CIN lesions -CIN lesions linked to: (1) high diversity with *Sneathia* species, and (2) increased abundance of *Lactobacillus and Lactobacillus reuteri* (specifically in CIN2) -*Fusobacterium* species frequently detected in women with cervical cancer	
Yang (2018)[49]	Role of *Lactobacillus* species in cervical cancer (n = 29)	16S rRNA, 16S rDNA, PCR, microscopic evaluation, bacterial isolation and purification[f]	HPV prevalence; cytology or biopsy results	-Lower abundance of *Lactobacillus* species in HPV positive women	- The mechanistic role of *Lactobacillus* in cervical cancer is ill-defined -Paucity of existing RCTs and cohort studies on the topic -Unpublished research and ongoing studies
				-High RA of *Lactobacillus* sp. and *Lactobacillus iners* in precancerous lesions and cervical cancer -Low RA of *Lactobacillus jensenii* and *Lactobacillus crispatus* in precancerous lesions and cervical cancer	

					were not considered

Abbreviations: CIN, cervical intraepithelial neoplasia; CST, community state type; HPV, human papillomavirus; hrHPV, high-risk human papillomavirus; HSIL, high-grade squamous intraepithelial lesions; LSIL, low-grade squamous intraepithelial lesions; PCR, polymerase chain reaction; RCT, randomized controlled trial; RA, relative abundance; sp, singular species; WMS, whole metagenome sequencing; 16S rRNA, 16S ribosomal RNA; 16S rDNA, 16S ribosomal DNA

[a] Refers to the microbial characterization techniques utilized by the studies included in the systematic review.
[b] Cytology or biopsy results refers to cytological interpretations or biopsy confirmed CIN or cervical cancer as the outcome of interest.
[c] Main findings are subdivided according to the outcomes assessed.
[d] Included studies do not include those that assessed the relationship between the gut microbiome and cervical cancer.
[e] Discrepancy whereby authors listed 19 studies in Table 1 but state that only 16 met their inclusion criterion.
[f] This systematic review included finding from both clinical and experimental studies; the tabulated microbial characterization techniques refer to those reported for the clinical studies.

Table 3-3. Meta-analyses on the association between the CVM and HPV prevalence, acquisition, persistence, clearance and/or cytology interpretations or biopsy confirmed CIN and cervical cancer.

First Author (Year) (sample size [n] of studies)	Microbiome characterization[a]	Inclusion criteria *Exclusion criteria*	Exposure	Outcome	Overall effect estimate[d] (95% CI) I^2	Limitations acknowledged
Brusselaers (2019) [b 18] (n = 15), but tabulated risk estimates were based on 2 studies using molecular techniques	Microscopy (Nugent Scoring, wet mount microscopy, Pap-stained smears) and Molecular (16s rRNA and cpn60 gene sequencing)	Original research, longitudinal or nested-case-control studies; Compare women with/without vaginal dysbiosis and assessed risk of HPV incidence, persistence and/or SIL lesions; Minimum of 2 measurement points; Microbial assessment via microscopic or molecular methods *Microbial assessment via vaginal pH measurement*	Non-*Lactobacillus crispatus* dominant	Incident HPV	RR, 1.85 (0.47-7.32) 55.7%	-Confounding -Differences in exposures/outcomes, study populations, and HPV types -Misclassification bias -Potential for reverse causation
			Non-*Lactobacillus crispatus* dominant	Persistent HPV	RR, 1.33 (0.63-2.81) 23.8%	
			Lactobacillus gasseri dominant	Persistent HPV	RR, 0.63 (0.10-3.86) 81.0%	
			Lactobacillus iners dominant	Persistent HPV	RR, 1.06 (0.42-2.63) 0.0%	
			Low lactobacilli, mixed aerobe & anaerobe	Persistent HPV	RR, 1.00 (0.23-4.30) 80.1%	
			Low *lactobacilli* anaerobe	Persistent HPV	RR, 2.00 (1.05-3.81) 0.0%	

Tamarelle (2019)[19] (n = 39), but risk estimate tabulated was based on 20 studies	16S rRNA gene amplicon sequencing, Nugent Score, Amsel's criteria or presence of clue cells	Association between vaginal microbiota and HPV, *C.trachomatis, M. genitalium* and/or *N.gonorroeae*; Microbial characterization via 16S rRNA gene amplicon sequencing, Nugent score, Amsel's criteria or presence of clue cells; Human study population; Cohort, cross-sectional or interventional study; Detect STIs with PCR *HIV positive (study population or >10% of participants); Pregnant; Literature review, letters and/or editorials; Sample size < 30*	Low *lactobacillus* vs. high *lactobacillus v*aginal microbiome	HPV Infection	OR, 1.53 (1.23-1.82) 66.1%	-Methods for STI diagnoses have different sensitivity and specificity -Different methods for microbial characterization -Pooling methods could mask the effect of individual species on STIs -No overall measure of association -Publication bias
Wang (2019)[20] (n = 11), but risk estimate tabulated differed by exposure/outcome considered	PCR amplification of 16S rRNA gene sequencing, Microarray with probes targeting 16S and 18S rRNA genes	Association between cervicovaginal *lactobacilli* and hrHPV/CIN/cervical cancer *Lactobacillus examination with microscopic techniques; Lactobacillus had to be assessed quantitatively in CST*; *Studies in human cell lines or animal models; Abstracts; Non-English articles; Duplicate publications; Studies without numerical value of raw data*	*Lactobacillus* dominant *CST*	hrHPV	OR, 0.64 (0.48-0.87), based on 9 studies 6.0%	-Cross-sectional analysis -Small number of studies and sample size -Variation in 16S hypervariable regions -Analysis restricted to CST types
			Lactobacillus iners dominant CST III	hrHPV	OR, 0.96 (0.69-1.34), based on 8 studies 0.0%	
			Lactobacillus crispatus dominant CST I	hrHPV	OR, 0.49 (0.31-0.79), based on 8 studies 10.0%	
			Lactobacillus dominant CST	CIN	OR, 0.53 (0.34-0.83), based on 6 studies 0.0%	
			Lactobacillus, *iners* dominant CST III	CIN	OR, 0.99 (0.60-1.64), based on 5 studies	

					0.0%	
			Lactobacillus crispatus dominant CST I	CIN	OR, 0.50 (0.29-0.88), based on 5 studies 0.0%	
			Cervicovaginal *Lactobacillus* pre-dominant CST	Cervical Cancer	OR, 0.12 (0.04-0.36), based on 3 studies 0.0%	
			Lactobacillus, *iners* dominant CST III	Cervical Cancer	OR, 0.13 (0.02-1.13), based on 2 studies 0.0%	
			Lactobacillus crispatus dominant CST I	Cervical Cancer	OR, 0.17 (0.03, 1.05), based on 2 studies 0.0%	

Abbreviations: CIN, cervical intraepithelial neoplasia; CST, community state type; HIV, human immunodeficiency virus; HPV, human papillomavirus; hrHPV, high-risk human papillomavirus; OR, odds ratio, PCR, polymerase chain reaction; RR, relative risk; SIL, squamous intraepithelial lesion; STI, sexually transmitted infection; 16S rRNA, 16S ribosomal RNA; 18S rRNA, 18S ribosomal RNA

[a] Refers to the microbial characterization techniques utilized by the studies included in the meta-analysis.
[b] This meta-analysis included findings from studies that characterized the microbiome using molecular and microscopic techniques. We only tabulated the effect estimates for molecular studies.
[c] Exclusion criterion was only for studies included in the quantitative synthesis.

Table 3-4. Diagnostic performance of the CVM to detect HPV and high-grade lesions in studies reporting a receiver operating characteristic curve analysis.

First Author (Year)	**Study Design** **Population** **(sample size [n] of participants; *breakdown by outcome*)** **Age**	**Microbiome characterization** **Hypervariable region** **Sequencing**	**HPV genotyping method** **Types Detected**	**Analysis approach to select microbial components** ***Outcome***	**Number of microbial components included** ***Enumeration of microbial components***	**AUC (95% CI)**
Morales (2021)[50]	Cross-sectional Reproductive aged women in Brazil (n = 546; *86 hrHPV+, 460 hrHPV-)* Ages, 18-51 years	16S rRNA V3-V4 Illumina MiSeq	Linear Array HPV genotyping hrHPVs; 16, 18, 26, 31, 33, 35, 45, 51, 52, 53, 56, 58, 59, 66, 68, 69, 70, 73, 82	Multivariable logistic regression with a stepwise forward selection algorithm (P < 0.15 for variable retention) *hrHPV Infection*	30 out of 116 bacterial taxa *Shuttleworthia satelles, Sutterella stercoricanis, Peptoniphilus, Eubacterium saphenum, Lactobacillus salivarius, Sutterella morbirenis, Pediococcus acidilactici, Aerococcus viridans,* BVAB3, *Prevotella* genogroup 3, *Streptococcus intermedius, Corynebacterium accolens, Dialister* sp type 2, *Megasphaera* sp type 2, *Dialister propionicifaciens*, *Eubacterium siraeum, Bacteroides uniformis, Prevotella* genogroup 2, *Leptotrichia amnionii, Acinetobacter calcoaceticus, Arcanobacterium hippocoleae, Roseburia intestinalis, Porphyromonas endodontalis, Enterococcus faecalis, Varibaculum cambriense, Raoultella planticola, Staphylococcus lugdunensis, Streptococcus anginosus, Mycoplasma genitalium, Streptococcus mutans*	0.802 (0.752 – 0.853)
Lee (2020)[52]	Cross-sectional Healthy women and those with CIN in Korea (n = 66; *42 CIN2+, 24 CIN1-*)[a] Age, mean 45.1	16S rRNA V3 Ion Torrent PGM for 1250 flows with Ion PGM Hi Q Sequencing Kit	Anyplex II HPV 28 assay kit 28 HPVs; 18 hrHPVs and 8 lrHPVs hrHPVs: 16, 18, 26, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68, 69, 73, 82 lrHPVs: 6, 11, 40, 42, 44, 53, 54, and 70	Random forest model using grid search on a five-run 10-fold cross-validation to select top 40 optimal bacterial species which were gradually added in the final model *CIN2+ vs. CIN1-*	Abundance of 33 bacterial species *Lactobacillus iners, Gardnerella vaginalis, Atopobium vaginae, Lactobacillus gasseri, Ureaplasma parvum, Streptococcus agalactiae, Lactobacillus amylovorus, Streptocccus anginosus, Lactobacillus casei, Lactobacillus helveticus, Lactobacillus acetolerans, Lactobacillus acidophilus, Prevotella disiens, Aerococcus christensenii, Lactobacillus vaginalis, Bifidobacterium dentium, Atopobium minutum, Lactobacillus johnsonii, Finegoldia magna, Lactobacillus pontis,* Unclassified *Dialister, Corynebacterium tuberculostearicum, Sneathia sanguinegens, Lactobacillus jensenii, Streptococcus salivarius, Enteroccocus faecalis, Prevotella bivia, Lactobacillus delbrueckii subsp Bulgaricus,* Unclassified *Megasphaera*, *Prevotella timonensis*,	0.952 (0.820-1.000)

					Unclassified *Prevotella*, *Streptococcus canis*, *Clostridium* BVAB2	
Yang (2020)[51]	Cross-sectional Women attending a colposcopy clinic following screening for cervical cancer (n = 52; *27 HPV16+, 25 HPV-*) Ages, 25-42 years	Shotgun metagenomic sequencing N/A Hiseq X-ten platform	Hybribio Rapid GenoArray test kit 15 hrHPVs, 6 lrHPVs hrHPVs: 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, 66, and 53 lrHPVs: 6, 11, 42, 43, 44, and CP8304	Random forest ensemble learning based on the mean decrease accuracy *HPV16+ vs. HPV-*	17 genera NR[b]	0.819 (0.684-0.954)
					7 species NR[b]	0.918 (0.839-0.997)

Abbreviations: AUC, area under the curve; HPV, human papillomavirus; hrHPV indicates high-risk HPV; CIN, cervical intraepithelial neoplasia; lrHPV indicates low-risk HPV; NR indicates not reported; 16S rRNA, 16S ribosomal RNA

[a] CIN2+ group consisted of CIN2 lesions to cervical cancer whereas CIN1- consisted of healthy controls to CIN1 lesions.
[b] Putative genera and species were listed in the article but those used in ROC analysis were not specified.

Table 3-5. Comparison between 16S rRNA gene sequencing and WMS for CVM characterization.

Features	16S rRNA Gene Sequencing	WMS
Technique and process[61–64]	Collection of cervical and/or vaginal sample	
	Extraction of genomic DNA from the cervical and/or vaginal sample	
	PCR amplification of hypervariable region(s) of the 16S rRNA bacterial gene	DNA shearing of all genomic DNA
	-	Preparation of WMS library depending on the target organism(s) and DNA fragmenting
	Amplicon sequencing using a sequencing platform (i.e., Pyrosequencing, Illumina MiSeq, Illumina nextSeq, PacBlo Seqel II/IIe, Nanopore)	Sequencing all genomic DNA using a sequencing platform (i.e., Illumina novaSeq, HiSeq**)**
	Taxonomic assignment using curated 16S rRNA databases	Taxonomic assignment using non-redundant databases or marker databases (metaphlan)
	Assessment of RA, and diversity of taxon between samples	Bioinformatics analysis to assess the microbial makeup of a sample
Ability to identify specific microbes[24]	Bacteriome (i.e. bacteria present in a sample at the genus and/or species level)	Complete microbiome (i.e. archaea, bacteria, eukaryotes, viruses, fungi, plasmids and gene content)
CVM Identification[53]	Bacteriome	
Advantages *Disadvantages*[65–69]	Cost-effective, time efficient, well-established databases *Generally limited taxonomic resolution of standard bioinformatics pipelines*	Comprehensive, high taxonomic resolution, identifies variants and mutations *Large quantity of noisy data, expensive, timely, databases are relatively novel*

Abbreviations: PCR, polymerase chain reaction; RA, relative abundance; WMS, whole metagenome sequencing; 16S rRNA, 16S ribosomal RNA

3.2.11. Manuscript 1 Figures

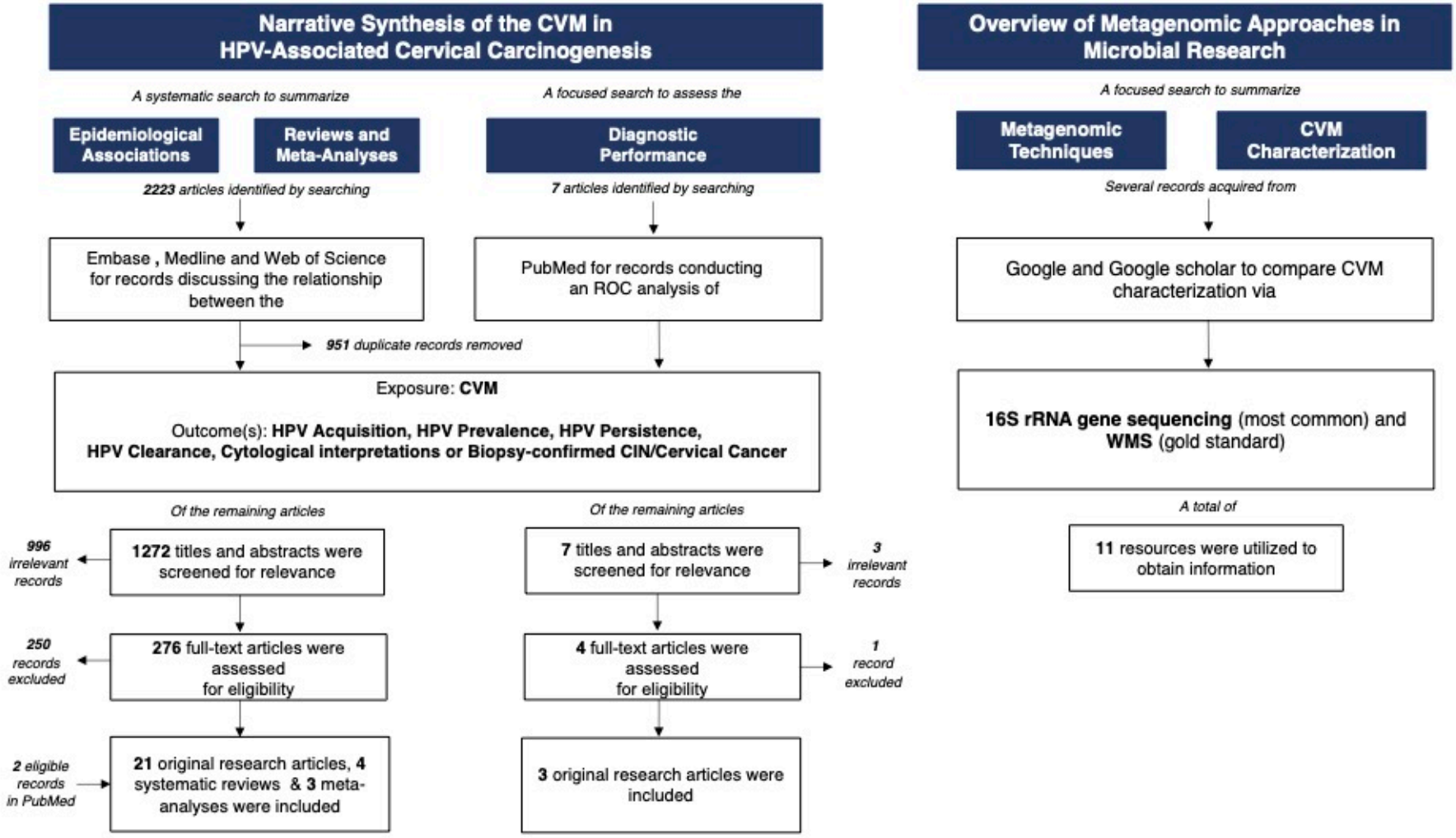


Figure 3-1. Mapping of the search strategy, methodology, and results.

Figure 3-1 Legend

To assess the etiological role of the CVM in HPV-associated cervical carcinogenesis, 2223 records were identified in Embase, Medline and Web of Science, of which 951 duplicates were excluded. 1272 articles were screened by title and abstract for relevance and the full texts of 276 articles were assessed for eligibility. Two additional articles were identified in PubMed following the systematic search. A total of 21 observational studies, 4 systematic reviews, and 3 meta-analyses were included. To assess the diagnostic performance of the CVM in HPV-associated cervical carcinogenesis, 7 relevant articles were identified in PubMed, of which 3 were excluded. Of the 4 full texts assessed for eligibility, one was excluded as it only assessed the diagnostic performance of bacterial genes (not taxa). The corresponding search strategies for the aforementioned searches are detailed in Supplemental Table 1. To identify common metagenomic techniques in microbial research, 11 resources were identified through a focused search of Google and Google Scholar to identify webpages and scholarly articles, respectively.

Abbreviations: CVM, cervicovaginal microbiome; HPV, human papillomavirus; ROC, receiver operating characteristic; WMS, whole metagenome sequencing; 16S rRNA, 16S ribosomal RNA.

CHAPTER 4: EMPIRICAL RESEARCH MANUSCRIPT

4.1. Preface

Chapter 3 provided a comprehensive review of the literature assessing the epidemiological and clinical role of CVM communities (identified using metagenomic methods) in HPV-associated cervical carcinogenesis. In summary, despite the utilization of comparable microbial characterization methods, epidemiological associations were inconsistent across the literature. Nevertheless, CVM species consistently served as good correlates of HPV and CIN lesions. Overall, in order to identify similar CVM communities and promote scientific replicability in future studies, we highlight a need for a harmonious technique for CVM characterization and high-resolution assessment of microbial communities.

The high-resolution ANCHOR pipeline has been shown to characterize the majority of 16S reads at the species-level.[21] Thus, we saw the opportunity to capitalize upon this novel bioinformatics pipeline and fully ascertain the microbial species constituting the CVM of 186 women enrolled in the MARKER study.[77] Participants were referred for colposcopy following abnormal cervical cytology results and samples were tested for the presence of HPV DNA. In Chapter 4 we analyze this data and present the findings in an empirical research manuscript titled, *Species-level characterization of the cervicovaginal microbiota and its role in human papillomavirus-associated cervical carcinogenesis.* The overall study methodology including the selection of the study population, bioinformatic characterization methods and epidemiological analyses are outlined in **S-Figure 4-0**. This figure is not to serve as redundant material, but to further explain the parent studies and to complement the methodology detailed in Section 4.2. To the best of our knowledge this is the first study to comprehensively characterize the CVM at the species-level with 16S rRNA gene sequencing and explore correlations between CVM species and hrHPV as well as CIN; these results could have important implications for cervical cancer prevention. Findings from this manuscript were presented at the *McGill Centre for Microbiome Research Symposium* (December 2022), *EUROGIN – Multidisciplinary HPV Congress* (February 2023), and *CORTO - Celebration of Research and Training in Oncology Conference* (June 2023). This manuscript was submitted to the *Journal of Medical Virology* following the initial submission of this thesis.

4.2. MANUSCRIPT 2: SPECIES-LEVEL CHARACTERIZATION OF THE CERVICOVAGINAL MICROBIOTA AND ITS ROLE IN HPV-ASSOCIATED CERVICAL CARCINOGENESIS

Authors:

Margaret Logel,[1] Mariam El-Zein,[1] Eduardo L. Franco,[1] Emmanuel Gonzalez[2,3]

Affiliations:

[1]Division of Cancer Epidemiology, McGill University, Montréal, Québec, Canada

[2]Canadian Centre for Computational Genomics (C3G), Department of Human Genetics, McGill University, Montréal, Québec, Canada

[3]Microbiome Research Platform, McGill Interdisciplinary Initiative in Infection and Immunity (MI4), Genome Centre, McGill University, Montreal, Montreal, Canada

Correspondence:

Eduardo L. Franco

Division of Cancer Epidemiology, McGill University

5100 Maisonneuve Blvd West, Suite 720

Montreal, QC, H4A 3T2, Canada E-mail address: eduardo.franco@mcgill.ca

4.2.1. Abstract

Background: Human papillomavirus (HPV) infection is a necessary but not sufficient cause of cervical cancer and its precancerous lesion, cervical intraepithelial neoplasia (CIN). The cervicovaginal microbiome may contribute to carcinogenesis, but studies have been limited by low-resolution computational analysis methods of microbial data. Using a high-resolution bioinformatics pipeline, we evaluated the relationship of the cervicovaginal microbiome with HPV and CIN.

Methods: The cervicovaginal microbiome of 186 women (ages 21-67), referred to colposcopy following abnormal cytology, was characterized by sequencing two 16S rRNA gene regions (V3-V4 and V5-V6) and annotated using the high-resolution ANCHOR pipeline. Samples were genotyped for HPV using the Roche-cobas 4800 assay. We fitted logistic regression models with a stepwise forward selection algorithm to select species (presence/absence) as correlates of CIN1+ and constructed a linear microbiome score based on the regression coefficients. Cytology-based (ASC-US; LSIL; HSIL+; missing) and HPV-based (HPV16/18; 12 other high-risk HPVs) scores were calculated from two separate logistic regression models to detect CIN1+. Receiver operating characteristic curve analyses were performed, and the area under the curve (AUC) and 95% confidence intervals (CI) were compared between scores.

Results: Overall, 66.7% of participants were HPV-positive. Of the 168 women with valid cytology results, 35 (20.8%) and 58 (34.5%) had normal or HSIL+, respectively. A total of 77 species were identified: 8 unique to V3-V4, 48 unique to V5-V6, and 21 shared between regions. Of these, 12 species (5 shared and 7 unique to V5-V6) were retained in the stepwise regression. The AUCs for the microbiome-, cytology- and HPV-based scores were 0.7656 (95%CI 0.6885-0.8426), 0.8524 (95%CI 0.7899-0.9149), and 0.7529 (95%CI 0.6855-0.8204), respectively.

Conclusion: Although cytology performed the best, the microbiome- and HPV-based scores had similar performance for CIN1+ detection, suggesting that bacterial species may also be implicated in cervical carcinogenesis.

4.2.2. Introduction

Cervical cancer is the fourth most common female cancer with approximately 604 000 new cases and 342 000 deaths in 2020.[1] Infection with high-risk human papillomavirus (HPV) types (particularly HPV16 and HPV18)[2] is a necessary but not sufficient cause of invasive cervical cancer and its pre-cancerous lesion, cervical intraepithelial neoplasia (CIN),[3] suggesting the implication of other factors, including the local microbiota.[4] Although most HPV infections are cleared or become latent,[5] those that persist can progress to CIN. Three CIN grades exist on a biological continuum of increasing disease severity; CIN1 (low-grade, mild dysplasia), CIN2 (high-grade, moderate dysplasia), and CIN3 (high-grade, severe dysplasia). Without treatment, CIN3 can progress to invasive cervical cancer; less than 20% of these lesions have been shown to regress.[6]

Increasing evidence suggests that the composition of the cervicovaginal microbiome contributes to the natural history of HPV infections and cervical cancer development.[4,7] The vaginal microbiome of a healthy female is dominated by species of the *Lactobacillus* genus, corresponding to low microbial diversity.[8] Bacterial vaginosis, a pathogenic state of the cervicovaginal microbiome, is characterized by a shift from *Lactobacillus* dominance to an overgrowth of anaerobic bacteria including *Gardnerella, Prevotella, Atopobium, Mobiluncus, Bifidobacterium, Sneathia, Leptotrichia,* and *Clostridiales*.[9] A loss of *Lactobacillus* dominance and an increase in anaerobic bacteria has been observed in certain disease-associated states.[9]

Methodological advances in microbiome sequencing techniques and analysis have enabled reliable microbial characterization and quantification. Much of the published literature on the composition of the cervicovaginal microbiota and its role in HPV-associated cervical carcinogenesis has used 16S ribosomal RNA (16S rRNA) gene sequencing. In 2011, using this technique and amplifying one 16S region (V1-V2), Ravel et al. categorized the vaginal microbiome of reproductive-age women into five community state types (CST) after annotating all reads at the genus- or species-level.[10] Four are dominated by different *Lactobacilli* species: CST I (characterized by *L.crispatus*), CST II (characterized by *L.gasseri*), CST III (characterized by *L.iners*), and CST V (characterized by *L. jensenii*).[10] Conversely, CST IV is *Lactobacilli* depleted with an increase in anaerobic bacteria. CST IV was associated with HPV persistence,[11] and non-regression of CIN2 lesions.[12]

Bacterial vaginosis is characterized by a community structure that is aligned with CST IV.[9,10] In 2022, Usyk et al. developed a molecular score for bacterial vaginosis (*molBV*) using 16S rRNA gene sequencing data (amplifying the V4 region) to assess microbial variability on a continuous scale based on the log ratios between the abundance of *Lactobacilli* and pre-specified genus-level bacterial vaginosis taxa.[13] An increasing *molBV* score (corresponding to higher microbial diversity) was associated with a higher risk of progression to CIN2+ lesions among women with an incident high-risk HPV infection.[13]

However, standard 16S bioinformatics pipelines such as QIIME1, MOTHUR, DADA2, and UPARSE are often limited to genus- rather than species-level taxonomic identification.[14] Moreover, the categorical CSTs are only able to pinpoint the dominant communities and the continuous *molBV* can only provide an overall assessment of microbial diversity, but neither provides detailed identification of all bacterial species. One consideration with 16S rRNA gene sequencing is the choice of the hypervariable region to amplify. Most studies assessing the association between individual microbial species and CIN or cervical cancer amplified V1-V3 or V3-V4 regions.[15–17] The presence of *Atopbium vaginae*, *Dialister invisus, Finegoldia magna, Gardnerella vaginalis*, *Prevotella buccalis*, and *Prevotella timonensis* was significantly associated with higher odds of CIN2+ or cervical cancer.[17] Similarly, a higher relative abundance of *Atopbium vaginae* was associated with a significantly higher risk of CIN.[16] A medium and/or high relative abundance of *Pseudomonas stutzeri*, *Bacteroides fragilis*, *Lactobacillus delbrueckii*, and *Streptococcus agalactiae* were associated with higher odds of CIN2+ lesions, however, most risk estimates were not significant.[15] Another consideration with 16S rRNA gene sequencing is the potential identification of different bacterial taxa depending on the amplified region; more bacterial species were found using V3-V4 compared to V1-V2 and missing or underestimating species may overestimate the abundance of other bacterial taxa.[18]

In this context, we characterized at the species-level the cervicovaginal microbiome in women – who were referred for colposcopy due to abnormal cytology – by amplifying two 16S regions to increase the resolving power for identifying bacterial taxa and using a novel high-resolution bioinformatics pipeline.[14] We also described microbial diversity in relation to the severity of CIN lesions and assessed the diagnostic performance of the cervicovaginal microbiome for the detection of CIN lesions and high-risk HPV infections. Finally, we calculated a *molBV* score to assess the correlation with CIN and high-risk HPV.

4.2.3. Methods

Study Design, Population, and Procedures

We used data from 186 physician-collected cervical samples (54 biopsy-confirmed normal, 50 CIN1, 40 CIN2, and 42 CIN3) from the Methylation Analysis Revealing Key Epigenetic Regulation (MARKER) study whose aim was to identify methylation markers that permitted discriminating among CIN grades. Details of the MARKER study design and population have been described elsewhere.[19] Briefly, women referred for abnormal cytology or for initial treatment of a cervical lesion in a colposcopy clinic at a McGill University- affiliated hospital in Montreal, Quebec, Canada were enrolled. The 186 women (aged 21-67 years) were selected at random among 480 women who had valid HPV and histology results. To enhance statistical efficiency, normal and CIN1 samples were under-sampled and CIN2 and CIN3 were over-sampled. Ethics approval was obtained from Institutional Review Boards at McGill University and the Jewish General Hospital. All women provided written informed consent for the original study and to utilize their samples for additional testing. Cervical samples were collected by the attending gynecologist using a ThinPrep Pap Test (Hologic Inc., Marlborough, MA). Cytological results were interpreted using the Bethesda classification as negative for intraepithelial lesion or malignancy (NILM), atypical squamous cells of undetermined significance (ASC-US), low-grade squamous intraepithelial lesion (LSIL), high-grade squamous intraepithelial lesion (HSIL), atypical squamous cells-cannot exclude HSIL (ASC-H), atypical glandular cells (AGC), or cervical cancer.[20] Specimens were genotyped for HPV DNA using the Roche Cobas 4800 HPV assay, which detects 14 high-risk HPV types via three separate channels: one for HPV16, one for HPV18, and a third, pooled result for 12 other HPV types (31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66 and 68). Cervical lesions were biopsied by senior gynecological pathologists.

Microbiome Characterization and Annotation

Genomic material was extracted from cervical samples according to the National Institutes of Health' Human Microbiome protocols.[21] Two distinct 16S rRNA hypervariable regions, V3-V4 and V5-V6, were amplified using PCR primers. The V3-V4 region was amplified using the 341F and 805R primer set and the V5-V6 region using the P609D and P699R primer set. PCR amplification was performed by Genome Québec and products were sequenced

using Illumina MiSeq v2 platform to produce 2 x 300 bp paired end reads. Sequences were processed and annotated using the high-resolution ANCHOR pipeline.[14] Briefly, after raw sequences were aligned, amplicon filtering occurred at a length of 465 for V3-V4 and 331 for V5-V6. Exact Sequence Variants (ESVs) were selected based on a count threshold of 31 for all 186 samples and denoised using ANCHOR algorithms. Sequences were compared to the National Center for Biotechnology Institute (NCBI) curated bacterial and archaeal reference sequence database (2022) and non-redundant nucleotide databases.[22] Annotation was characterized using BLASTn with a criteria of >99% identity and coverage for a family, genera or species-level inference. NCBI curated bacterial and archaeal refseq was given priority when at 100% identity and coverage. Multiple, equally good (highest identity/coverage), annotation was retained and reported. Amplicons with low counts (< 31) were grouped to high-count sequences in a second BLASTn, using a lower threshold of 98% identity and coverage. All annotations should be considered as putative and interpreted with caution as databases contain errors and are subject to change. ESVs were grouped into species based on taxonomic identification and those annotated at the genus-level or higher, or corresponding to unknown, were not included in the epidemiological analyses.

Bioinformatic Analyses and *molBV* Calculation

Raw counts were transformed across samples for comparison (rlog function, R Phyloseq package).[23] Sparsity and low-abundance cut-offs were applied where an ESV count in a single sample was < 90% of the counts in all samples, and ESV counts were > 2 across samples within a comparison. Alpha diversity, measured using the Shannon index with the Vegan R package,[24] was compared between normal and CIN samples using a Mann-Whitney U (non-parametric) test. Supervised ordination analysis (CCA) was performed using rlog transformed data with the Vegan package.[24] Principal component analysis (PCoA) unsupervised ordination was calculated based on Bray-Curtis distances and transformed counts. Dispersion ellipses were drawn using veganCovEllipse function from the Vegan package in R (R Development Core Team, 2014).[24] Multivariate analysis was conducted to identify significant variances between normal and CIN samples (PERMANOVA, $p < 0.05$).

Using the 16S rRNA sequencing data (logarithmic transformed), we calculated a continuous *mol*BV measurement using the code developed by Usyk et al. (2022).[13] The *molBV*

score ranges from 0 to 10, with higher values corresponding to increased microbial dysbiosis. Bioinformatic analyses and calculation of the *molBV* score were performed in R version 4.2.2.

Statistical Analyses

We described the characteristics (age, cytology, and HPV positivity) of study participants, overall and by histological endpoints. We differentiated between the number of species identified solely using V3-V4 and those using V5-V6 as well as the species that were shared between the two hypervariable regions (i.e., identified in both regions), overall and by age, cytology, and histology. The presence of bacterial species was compared across histological endpoints using the Fisher's exact test for heterogeneity. We also assessed microbial variability on a continuous scale in terms of the species raw abundance and calculated descriptive statistics (range, mean, geometric mean, median, and interquartile range, IQR). For species that had a raw abundance of zero, the median and IQR were calculated based on samples with a raw abundance greater than zero. We generated a correlation matrix based on species raw abundance in samples with normal histology; for species that were shared between V3-V4 and V5-V6, the raw abundance from the region with the highest maximum raw abundance was considered. The strength of the rank correlation between bacterial species was assessed using the Spearman correlation coefficient; ρ ranges from -1 to 1 indicating a perfect positive and negative relationship, respectively.

We fitted logistic regression models with a stepwise forward selection algorithm ($P < 0.15$ for variable retention) to select species (based on presence or absence considering shared species to be present if raw abundance > 0 in one or both regions; species raw abundance considering the abundance of shared species from the region with the highest maximum abundance) as potential correlates of CIN grades (CIN1+ versus normal; CIN2+ versus normal and CIN1) and high-risk HPV infection (positive versus negative). We used the regression coefficients of the retained species to construct corresponding linear microbiome-based scores. We also constructed cytology-based (ASC-US; LSIL; HSIL, ASC-H, AGC, and cervical cancer; missing), HPV-based (HPV16 and/or 18; 12 other high-risk HPVs), and microbiome-HPV-based scores using the regression coefficients from three separate logistic regression models. The microbiome-HPV-based score included the retained bacterial species (from the stepwise

selection) and the two HPV variables (HPV16 and/or 18; 12 other high-risk HPVs). We plotted receiver operating characteristic (ROC) curves to compare the performance of the scores to detect CIN1+, CIN2+, or high-risk HPV. The area under the curve (AUC) and 95% confidence intervals (CI) were calculated based on the asymptotic normal. The AUCs were compared across test scores using a χ^2 test for the equality of ROC areas; p-values < 0.05 were considered statistically significant. We performed an additional analysis by restricting to women who tested positive for high-risk HPV and similarly calculated the microbiome-based (species presence or absence) and cytology-based scores for CIN1+ and CIN2+ detection.

We performed ROC analyses for the performance of *molBV* scores using a subset of samples (n = 183; 3 samples were excluded due to sparsity filtering). Separately, we constructed $molBV_{V3\text{-}V4}$ and $molBV_{V5\text{-}V6}$ scores based on the regression coefficients from univariate logistic regression models for the detection of CIN1+, CIN2+, and high-risk HPV.

All epidemiological analyses were performed using Stata BE 17.0 (StataCorp LLC., TX).

4.2.4. Results

As shown in **Table 4-1**, most study participants were aged between 30 and 50 years (60.8%) and almost two thirds had LSIL or worse lesions (60.2%) and were positive for any high-risk HPV types (66.7%). As expected, a larger proportion of participants with CIN2 (90.0%) and CIN3 (88.1%) lesions were positive for any high-risk HPV compared to those with CIN1 (62.0%) or normal (37.0%) histology. Positivity for HPV16 and/or 18 increased among patients with increasing lesion severity: normal (3.7%), CIN1 (14.0%), CIN2 (37.5%), and CIN3 (47.6%).

Figure 4-1 presents descriptive bioinformatics results. *Firmicutes* was the most relative abundant phylum in V3-V4 and V5-V6 regions (**Figure 4-1 A and B**). Of the total raw reads sequenced based on V3-V4 (raw reads = 5,014,566) and V5-V6 (raw reads = 5,001,428) regions, 71 (56 species, 6 genus, 2 family, and 7 unknown) and 222 (115 species, 11 genus, 4 family, and 92 unknown) ESVs were inferred, representing 3,947,186 and 4,284,291 total counts respectively. The majority (92.6%) of raw counts were annotated at the species level for both regions (**Figure 4-1C**). Alpha diversity comparing CIN and normal samples revealed a higher, but not significant, median Shannon index across CIN samples (P > 0.05) for both regions (**Figure 4-1D**). With respect to Beta diversity, no significant difference between CIN and normal

samples was found via PCoA for V3-V4 or V5-V6 ($P > 0.05$) (**Figure 4-1E**) or CCA for V3-V4 ($P > 0.05$) (**Figure 4-1F**). CCA showed a significant separation of normal and CIN samples based only on V5-V6 ($P < 0.05$) (**Figure 4-1F**).

Table 4-2 shows that a total of 77 unique species were identified; 8 were exclusive to V3-V4, 48 to V5-V6, and 21 were shared between regions. More cervicovaginal microbial species were identified by amplification of the V5-V6 region (n= 69) compared to V3-V4 (n= 29). The number of species identified did not differ by age, cytology, and histology. **S-Tables 4-1 and 4-2** summarize the identified species by histological endpoint and their raw abundance. A statistically significant difference by histology was found for two bacterial species: *Finegoldia magna* ($P = 0.040$, based on V3-V4) and *Mycoplasma hominis* ($P = 0.032$, based on V5-V6). In both regions, raw abundance varied greatly across the identified species and was highest for *Lactobacillus iners* (max = 32976, mean = 6227.1 in V3-V4; max = 37804, mean = 6702.5 in V5-V6). A total of 70 bacterial species were present in normal samples; generally, there was a weak, negative correlation between *Lactobacilli* and non-*Lactobacillus* species, as expected (**S-Figure 4-1**).

The stepwise logistic regression coefficients for bacterial species (presence versus absence) constituting the microbiome-based scores to detect CIN1+, CIN2+, high-risk HPV, and CIN2+ among women who tested HPV positive are respectively shown in **S-Table 4-3** (12/77 species retained, 6 species absent in normal samples were not included as predictors due to model convergence errors), **S-Table 4-4** (5/77 species retained), **S-Table 4-5** (11/77 species retained), and **S-Table 4-6** (6/77 species retained). **S-Tables 4-3 to 4-6** also present the regression coefficients for the cytology-, HPV-, and microbiome-HPV-based scores. **Table 4-3** and **Figures 4-2** and **4-3** show the corresponding ROC curve analysis results. For CIN1+ detection, the microbiome-based score performed similarly well to the HPV- (AUC = 0.7656 vs. 0.7529, $P = 0.8103$) and cytology- (AUC = 0.7656 vs. 0.8524, $P = 0.1078$) based scores whereas the microbiome-HPV-based score performed the best; it was significantly more accurate than the HPV-based score (AUC = 0.8749 vs. 0.7529, $P < 0.001$). Considering CIN2+, the microbiome-based score was significantly less accurate than the HPV- (AUC = 0.6409 vs. 0.7591, $P = 0.0111$) and cytology- (AUC = 0.6409 vs. 0.8431, $P = 0.0441$) based scores. The microbiome-HPV-based score had similar performance to the HPV- (AUC = 0.7984 vs. 0.7591, $P = 0.0630$) and cytology- (AUC = 0.7984 vs. 0.8431, $P = 0.3066$) based scores for CIN2+ detection. In

terms of high-risk HPV positivity, the microbiome- and cytology-based scores performed the same (AUC 0.7703 vs. 0.7733, P = 0.9498). Upon restricting the analysis to women who tested positive for high-risk HPV, the microbiome-based score was less accurate than the cytology-based score for CIN2+ detection (AUC = 0.7044 vs. 0.8255, P = 0.0350). For the detection of CIN1+, the models failed to converge.

Likewise, the stepwise logistic regression coefficients for bacterial species (raw abundance) constituting the microbiome-based scores to detect CIN1+, CIN2+, and high-risk HPV are respectively shown in **S-Table 4-7** (5/77 species retained), **S-Table 4-8** (6/77 species retained), and **S-Table 4-9** (4/77 species retained), along with the coefficients from the corresponding cytology-, HPV-, and microbiome-HPV-based scores. **S-Table 4-10** and **S-Figure 4-2** show the corresponding ROC curve analysis results. For CIN1+ detection, the microbiome-based score was significantly less accurate than the HPV- (AUC = 0.5252 vs. 0.7529, P < 0.001) and cytology- (AUC = 0.5252 vs. 0.8524, P < 0.001) based scores. By contrast, the microbiome-HPV-based score had similar performance to the HPV-based score (AUC = 0.7623 vs. 0.7529, P = 0.5574) but was significantly less accurate than the cytology-based score (AUC = 0.7623 vs. 0.8524, P = 0.0389). Considering CIN2+, the microbiome-based score was significantly less accurate than the HPV- (AUC = 0.6377 vs. 0.7591, P = 0.0197) and cytology- (AUC = 0.6377 vs. 0.8431, P < 0.001) based scores. The microbiome-HPV-based score performed significantly better than the HPV-based score (AUC = 0.8130 vs. 0.7591, P = 0.0120) and similarly to the cytology-based score (AUC = 0.8130 vs. 0.8431, P = 0.4708). In terms of high-risk HPV positivity, the microbiome-based score was significantly less accurate than the cytology-based score (AUC = 0.6052 vs. 0.7733, P = 0.0044).

The logistic regression coefficients for the continuous $molBV_{V3\text{-}V4}$- and $molBV_{V5\text{-}V6}$-based scores are shown in **S-Table 4-11**. Both scores were unable to detect CIN1+, CIN2+, or high-risk HPV (AUCs ≤ 0.5808) **(S-Table 4-12 and S-Figure 4-3)**.

4.2.5. Discussion

To the best of our knowledge, our study is the first to provide detailed species-level taxonomic identification of 16S reads. Utilizing the high-resolution ANCHOR pipeline,[14] and amplifying two unique 16S regions (V3-V4 and V5-V6) allowed us to comprehensively ascertain and analyze the species that constitute the cervicovaginal microbiome and examine their

associations with CIN and high-risk HPV. Among women referred for colposcopy following abnormal cytology, we found that microbial species, cytology and HPV results performed similarly for the detection of any CIN, whereas joint positivity for HPV and microbial species performed similarly to cytology and better than HPV for the detection of high-grade CIN. Moreover, we identified some microbial species that correlated well with CIN and high-risk HPV.

We found a significant gain in the ability to detect CIN1+ and CIN2+ when considering joint positivity for microbial species and high-risk HPV compared to high-risk HPV alone. Based on the sufficient-component cause model, there is rarely a singular, sufficient cause of an outcome but rather several component causes which must be present and act together to cause disease.[25] Although infection with high-risk HPV types is well-established as a necessary cause of invasive cervical cancer,[3] only a small proportion of HPV infections progress to pre-cancerous lesions and an even smaller proportion cause cancer,[5] suggesting that these infections are a single, albeit critical, component cause of cervical cancer. It is plausible to hypothesize that in addition to high-risk HPV infections, microbial species may be an additional factor contributing to the development of pre-cancerous lesions and subsequently cervical cancer.

In our study, *Lactobacillus delbrueckii* and *Prevotella timonensis*, species negatively associated with persistent HPV infections,[26] and positively associated with CIN2/3 lesions or cervical cancer,[17] respectively, were only identified in the V5-V6 region, highlighting the importance of the choice of 16S region utilized. The prevalence of *Finegoldia magna* (16 normal, 9 CIN1, 17 CIN2, 17 CIN3; P = 0.040; based on V3-V4) and *Mycoplasma hominis* (13 normal, 6 CIN1, 9 CIN2, 2 CIN3; P = 0.032; based on V5-V6) differed significantly by histologic endpoint. As we identified a large number of bacterial species, these associations may have arisen due to chance. However, *Mycoplasma hominis* was retained as a variable in the microbiome-based score to detect CIN1+, as was *Finegoldia magna* for the detection of CIN1+ and high-risk HPV. Moreover, the presence of *Finegoldia magna* has been significantly associated with a 6-fold increase in the odds of CIN2+.[17]

Our finding that a higher abundance of *Lactobacillus* species in women with normal histology correlated with a lower abundance of several other species corroborates previous research. Most women of reproductive age have a vaginal microbiome dominated by species of the *Lactobacillus* genus.[10] Indicative of health, *Lactobacillus* species produce lactic acid[10] and

bacteriocins such as lactocin,[27] which could aid in protecting against harmful bacteria and viruses. Bacterial vaginosis, characterized by a community structure that is consistent with that of CST IV (*Lactobacillus*-depleted),[9,10] has been associated with the acquisition of sexually transmitted infections including human immunodeficiency virus[28] and HPV.[29] We did not find that *molBV* was able to detect high-risk HPV or CIN contrary to findings by Usyk et al. in their prospective evaluation of microbial variability using *molBV* (n = 307 women, 2 study visits).[13] After adjusting for age, smoking status and HPV16, the authors found that an increasing *molBV* score was significantly associated with a 24% increase in the odds of progression to CIN2+ among women with an incident high-risk HPV infection.[13] However, this study was limited by low taxonomic resolution.

Our study had a few limitations that need to be acknowledged. First, our sample size was relatively small. However, we had sufficient precision and power as we had equal proportions of histological endpoints due to over-sampling of CIN2/3 and under-sampling of normal and CIN1. In the general population, the number of normal and CIN1 lesions are much greater than CIN2 and the latter larger than CIN3; this would represent sampling approximately 30 000 average risk women to gain the same precision.[30] Furthermore, the controls were biopsy-confirmed normal and represent the true counterfactual contrast. Second, our results pertaining to the species raw abundance and sensitivity (restricting to women who tested positive for high-risk HPV) analyses need to be interpreted with caution due to errors with model convergence. These errors could be attributed to the consideration of several bacterial species with a raw abundance of zero for the former analysis and a smaller sample size for the HPV-restricted analyses. Moreover, the raw abundance analysis yielded computational errors where the exposure variables perfectly predicted the outcome; a situation which can lead to estimation errors. Finally, the cross-sectional nature of our study limits our ability to assess temporality of events and understanding of the role of the cervicovaginal microbiome in the persistence or regression of HPV and subsequent disease. As a required next step, a prospective cohort study characterizing the cervicovaginal microbiome at the species-level would enable a better assessment of causation and studying detailed changes in microbial composition throughout different stages of HPV-associated cervical carcinogenesis.

Cervical cytology, characterized by its high specificity (96-98%), but lower sensitivity (51-53%), has been the gold standard for cervical cancer screening for many decades.[31–33] More

recently, HPV DNA testing has been replacing cytological assessment as the primary screening modality due to its superior sensitivity despite a lower specificity.[31,34] However, countries implementing HPV-based screening are faced with the challenge of how to triage women who test positive to identify those at highest risk of disease. Microbial testing may prove to be an additional tool for triaging women who test positive for HPV especially in low-resource settings where the implementation of HPV vaccination and cervical cancer screening presents several challenges due to financial and cultural barriers as well as a lack of critical infrastructure. Indeed, less than 30% of low- and middle- income countries have established HPV vaccination programs and only 44% of women have been screened for cervical cancer.[1] Alternatively, following the etiological establishment of microbial components in cervical carcinogenesis, microbiome modulation could represent a low-cost therapeutic strategy for correcting disturbances to the vaginal microbiota.[35,36]

4.2.6. Declarations of Interests and Sources of Funding

ELF reports grants and/or consulting fees from Merck and the Canadian Institutes of Health Research (CIHR), outside of the submitted work. ELF and MZ hold a patent related to the discovery of "DNA methylation markers for early detection of cervical cancer", registered at the Office of Innovation and Partnerships, McGill University, Montreal, Quebec, Canada (October 2018). ELF is an Advisory Board Member of Cancer Research Organizations (Camargo Cancer Center, Charbonneau Cancer Institute, Universidade do Minho). ML received a travel award from the McGill Centre for Viral Diseases. ELF receives an honorarium as an editor for various journals (Oxford University Press, Elsevier and Elifesciences Ltd). MZ receives an honorarium from Elsevier as an Associate Editor for Preventive Medicine. ML receives an honorarium from Elsevier as an Assistant Editor for Preventive Medicine and Preventive Medicine Reports. EG has nothing to disclose. The present work was supported financially by the CIHR (Foundation Grant FDN – 143347 to ELF) and Fonds de Recherche du Québec – Santé (Master's Fellowship to ML).

4.2.7. Acknowledgements

We would like to thank Dr. Robert Burk and Mykhaylo Usyk for their insights and for making publicly available for use their continuous *molBV* score.

4.2.8. Author Contributions

ELF, MZ, and EG formulated the research question. ELF and MZ designed the parent study. EG performed the microbial characterization and bioinformatics analyses. ML assisted with the bioinformatics analyses, performed the statistical analyses, and drafted the manuscript under the supervision of MZ, ELF, and EG. All authors revised the manuscript and approved the final version that was submitted for publication.

4.2.9. Manuscript 2 References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021 May 4;71(3):209–49.

2. Muñoz N, Bosch FX, Castellsagué X, Díaz M, de Sanjose S, Hammouda D, et al. Against which human papillomavirus types shall we vaccinate and screen? the international perspective. Int J Cancer. 2004 Aug 20;111(2):278–85.

3. Walboomers JMM, Jacobs M V., Manos MM, Bosch FX, Kummer JA, Shah K V., et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. J Pathol. 1999;189(1):12–9.

4. Norenhag J, Du J, Olovsson M, Verstraelen H, Engstrand L, Brusselaers N. The vaginal microbiota, human papillomavirus and cervical dysplasia: a systematic review and network meta-analysis. BJOG An Int J Obstet Gynaecol. 2020 Jan 17;127(2):171–80.

5. Schiffman M, Castle PE, Jeronimo J, Rodriguez AC, Wacholder S. Human papillomavirus and cervical cancer. Lancet. 2007 Sep;370(9590):890–907.

6. Loopik DL, Bentley HA, Eijgenraam MN, IntHout J, Bekkers RLM, Bentley JR. The Natural History of Cervical Intraepithelial Neoplasia Grades 1, 2, and 3: A Systematic Review and Meta-analysis. J Low Genit Tract Dis. 2021 Jul;25(3):221–31.

7. Brusselaers N, Shrestha S, van de Wijgert J, Verstraelen H. Vaginal dysbiosis and the risk of human papillomavirus and cervical cancer: systematic review and meta-analysis. Am J Obstet Gynecol. 2019 Jul;221(1):9-18.e8.

8. Mancabelli L, Tarracchini C, Milani C, Lugli GA, Fontana F, Turroni F, et al. Vaginotypes of the human vaginal microbiome. Environ Microbiol. 2021 Mar 1;23(3):1780–92.

9. Chen X, Lu Y, Chen T, Li R. The Female Vaginal Microbiome in Health and Bacterial Vaginosis. Front Cell Infect Microbiol. 2021 Apr 7;11:631972.

10. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SSK, McCulle SL, et al. Vaginal microbiome of reproductive-age women. Proc Natl Acad Sci U S A. 2011 Mar 15;108(SUPPL. 1):4680–7.

11. 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. Sci Rep. 2017 Dec 31;7(1):10200.

12. Mitra A, MacIntyre DA, Ntritsos G, Smith A, Tsilidis KK, Marchesi JR, et al. The vaginal microbiota associates with the regression of untreated cervical intraepithelial neoplasia 2 lesions. Nat Commun. 2020;11(1):1–13.

13. Usyk M, Schlecht NF, Pickering S, Williams L, Sollecito CC, Gradissimo A, et al. molBV reveals immune landscape of bacterial vaginosis and predicts human papillomavirus infection natural history. Cortes B Jimenez SE, Rodriguez AC, Hildesheim A, Kreimer AR, Lowy DR, Schiffman M, Schiller JT, Sherman M, Wacholder S, Kemp TJ, Sidawy

MK, Quint W, van Doorn LJ, Struijk L, Palefsky JM, Darragh TM, Stoler MH GP, editor. Nat Commun. 2022;13(1):233.

14. Gonzalez E, Pitre FE, Brereton NJB. ANCHOR: a 16S rRNA gene amplicon pipeline for microbial analysis of multiple environmental samples. Environ Microbiol. 2019;21(7):2440–68.

15. 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 Med. 2018 May 2;7(5):2172–9.

16. Oh HY, Kim B-S, Seo S-S, Kong J-S, Lee J-K, Park S-Y, et al. The association of uterine cervical microbiota with an increased risk for cervical intraepithelial neoplasia in Korea. Clin Microbiol Infect. 2015 Jul;21(7):674.e1-674.e9.

17. So KA, Yang EJ, Kim NR, Hong SR, Lee J-H, Hwang C-S, et al. Changes of vaginal microbiota during cervical carcinogenesis in women with human papillomavirus infection. Consolaro MEL, editor. PLoS One. 2020 Sep 17;15(9):e0238705.

18. Graspeuntner S, Loeper N, Künzel S, Baines JF, Rupp J. Selection of validated hypervariable regions is crucial in 16S-based microbiota studies of the female genital tract. Sci Rep. 2018 Jun 26;8(1):9678.

19. El-Zein M, Cheishvili D, Gotlieb W, Gilbert L, Hemmings R, Behr MA, et al. Genome-wide DNA methylation profiling identifies two novel genes in cervical neoplasia. Int J Cancer. 2020;147(5):1264–74.

20. SMITH JHF. Bethesda 2001. Cytopathology. 2002 Feb;13(1):4–10.

21. National Institutes of Health. NIH Human Microbiome Project.

22. National Library of Medicine (Organization). 16S RefSeq records processing and curation [Internet]. Available from: https://www.ncbi.nlm.nih.gov/refseq/targetedloci/16S_process/

23. McMurdie PJ, Holmes S. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS One. 2013;8(4).

24. Oksanen r J, Simpson GL, Blanchet FG, Solymos P, Stevens MHH, Szoecs E, et al. Title Community Ecology Package. Vegan Community Ecol Packag. 2022.

25. Rothman KJ, Greenland S. Causation and causal inference in epidemiology. Am J Public Health. 2005;95(SUPPL. 1):144–50.

26. Ritu W, Enqi W, Zheng S, Wang J, Ling Y, Wang Y. Evaluation of the Associations Between Cervical Microbiota and HPV Infection, Clearance, and Persistence in Cytologically Normal Women. Cancer Prev Res. 2019 Jan 1;12(1):43–56.

27. Turovskiy Y, Ludescher RD, Aroutcheva AA, Faro S, Chikindas ML. Lactocin 160, a Bacteriocin Produced by Vaginal Lactobacillus rhamnosus, Targets Cytoplasmic Membranes of the Vaginal Pathogen, Gardnerella vaginalis. Probiotics Antimicrob Proteins. 2009;1(1):67–74.

28. Atashili J, Poole C, Ndumbe PM, Adimora AA, Smith JS. Bacterial vaginosis and HIV acquisition: a meta-analysis of published studies. AIDS. 2008 Jul 31;22(12):1493–501.

29. King CC, Jamieson DJ, Wiener J, Cu-Uvin S, Klein RS, Rompalo AM, et al. Bacterial Vaginosis and the Natural History of Human Papillomavirus. Infect Dis Obstet Gynecol. 2011;2011:1–8.

30. El-Zein M, Bouten S, Louvanto K, Gilbert L, Gotlieb W, Hemmings R, et al. Validation of a new HPV self-sampling device for cervical cancer screening: The Cervical and Self-Sample In Screening (CASSIS) study. Gynecol Oncol. 2018 Jun;149(3):491–7.

31. Tota JE, Bentley J, Blake J, Coutlée F, Duggan MA, Ferenczy A, et al. Introduction of molecular HPV testing as the primary technology in cervical cancer screening: Acting on evidence to change the current paradigm. Prev Med (Baltim). 2017 May;98:5–14.

32. Nanda K, McCrory DC, Myers ER, Bastian LA, Hasselblad V, Hickey JD, et al. Accuracy of the papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: A systematic review. Ann Intern Med. 2000 May 16 [cited 2021 Apr 5];132(10):810–9.

33. Cuzick J, Clavel C, Petry K-U, Meijer CJLM, Hoyer H, Ratnam S, et al. Overview of the European and North American studies on HPV testing in primary cervical cancer screening. Int J Cancer. 2006 Sep 1;119(5):1095–101.

34. World Health Organization (WHO). New recommendations for screening and treatment to prevent cervical cancer [Internet]. [cited 2023 Dec 14]. Available from: https://www.who.int/news/item/06-07-2021-new-recommendations-for-screening-and-treatment-to-prevent-cervical-cancer#:~:text=WHO suggests using either of,every 5 to 10 years.

35. Ou Y-C, Fu H-C, Tseng C-W, Wu C-H, Tsai C-C, Lin H. The influence of probiotics on genital high-risk human papilloma virus clearance and quality of cervical smear: a randomized placebo-controlled trial. BMC Womens Health. 2019 Dec 24;19(1):103.

36. Verhoeven V, Renard N, Makar A, Royen P Van, Bogers JP, Lardon F, et al. Probiotics enhance the clearance of human papillomavirus-related cervical lesions: A prospective controlled pilot study. Eur J Cancer Prev. 2013;22(1):46–51.

4.2.10. Manuscript 2 Tables

Table 4-1. Characteristics [n (%)] of study participants, overall and by histological endpoints.

Variables	Categories	Overall (N = 186)	Histology			
			Normal (n = 54)	CIN1 (n = 50)	CIN2 (n = 40)	CIN3 (n = 42)
Age	< 30	47 (25.3)	11 (20.4)	14 (28.0)	14 (35.0)	8 (19.0)
	30 - 50	113 (60.8)	34 (63.0)	29 (58.0)	25 (62.5)	25 (59.5)
	>50	26 (14.0)	9 (16.7)	7 (14.0)	1 (2.5)	9 (21.4)
Cytology	NILM	35 (18.8)	31 (57.4)	2 (4.0)	0 (0.0)	2 (4.8)
	ASC-US	21 (11.3)	3 (5.6)	8 (16.0)	8 (20.0)	2 (4.8)
	LSIL	54 (29.0)	11 (20.4)	30 (60.0)	11 (27.5)	2 (4.8)
	HSIL or worse[a]	58 (31.2)	3 (5.6)	5 (10.0)	19 (47.5)	31 (73.8)
	Missing	18 (9.7)	6 (11.1)	5 (10.0)	2 (5.0)	5 (11.9)
HPV positivity (no mutually exclusive categories)	HPV16	41 (22.0)	2 (3.7)	6 (12.0)	13 (32.5)	20 (47.6)
	HPV18	4 (2.2)	0 (0.0)	2 (4.0)	2 (5.0)	0 (0.0)
	HPV16 and/or HPV18	44 (23.7)	2 (3.7)	7 (14.0)	15 (37.5)	20 (47.6)
	Other high-risk HPV[b]	101 (54.3)	19 (35.2)	30 (60.0)	28 (70.0)	24 (57.1)
	Any high-risk HPV[c]	124 (66.7)	20 (37.0)	31 (62.0)	36 (90.0)	37 (88.1)

Abbreviations: ASC-US, atypical squamous cells of undetermined significance; CIN, cervical intraepithelial neoplasia; HSIL, high-grade squamous intraepithelial lesion; HPV, human papillomavirus; LSIL, low-grade squamous intraepithelial lesion, NILM, negative for intraepithelial lesion or malignancy

[a] Includes 16 ASC-H, atypical squamous cells-cannot exclude HSIL, (3 among CIN1, 5 among CIN2, and 8 among CIN3), 2 AGC, atypical glandular cells, among normal, and 1 cancer case among CIN3.
[b] Includes a pooled positivity result for any of HPVs 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and/or 68.
[c] Includes positivity for any of HPVs 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and/or 68.

Table 4-2. Number of species identified based on the hypervariable region (total and unshared or shared between V3-V4 and V5-V6), overall and by characteristics of study participants.

Variables	Categories	n Species, V3-V4		n Species, V5-V6		n Species, V3-V4 & V5-V6
		Total	Unshared	Total	Unshared	Shared
	Overall	29	8	69	48	21
Age	< 30	29	10	59	40	19
	30 – 50	28	8	66	46	20
	> 50	27	8	65	46	19
Cytology	NILM	26	7	61	42	19
	ASC-US	27	9	59	41	18
	LSIL	29	8	64	43	21
	HSIL or worse[a]	29	8	67	46	21
	Missing	25	9	48	32	16
Histology	Normal	28	8	63	43	20
	CIN1	29	8	66	45	21
	CIN2	29	10	60	41	19
	CIN3	29	8	65	44	21

Abbreviations: ASC-US, atypical squamous cells of undetermined significance; CIN, cervical intraepithelial neoplasia; HSIL, high-grade squamous intraepithelial lesion; LSIL, low-grade squamous intraepithelial lesion, NILM, negative for intraepithelial lesion or malignancy

[a] Includes 16 ASC-H, atypical squamous cells-cannot rule out HSIL, 2 AGC, atypical glandular cells, and 1 cancer case.

Table 4-3. ROC curve analysis for the performance of cytology-, HPV-, and microbiome- (species presence/absence) based scores to detect CIN lesions and high-risk HPV infections.

Study population	Contrast groups	Test-based score	AUC	95% CI[a]	P-value, HPV-based score as comparator[b]	P-value, cytology-based score as comparator[b]
All participants, n = 186	CIN1+ vs. Normal	Cytology[c]	0.8524	0.7899 - 0.9149	0.0151	-
		HPV[d]	0.7529	0.6855 - 0.8204	-	0.0151
		Microbiome[e]	0.7656	0.6885 - 0.8426	0.8103	0.1078
		Microbiome-HPV[f]	0.8749	0.8207 - 0.9291	< 0.001	0.5910
	CIN2+ vs. Normal & CIN1	Cytology[c]	0.8431	0.7879 - 0.8983	0.0441	-
		HPV[d]	0.7591	0.6947 - 0.8235	-	0.0441
		Microbiome[g]	0.6409	0.5723 - 0.7095	0.0111	< 0.001
		Microbiome-HPV[h]	0.7984	0.7352 - 0.8616	0.0630	0.3066
	HPV+ vs. HPV-	Cytology[c]	0.7733	0.7011 – 0.8456	-	-
		Microbiome[i]	0.7703	0.7022 – 0.8384	-	0.9498
HPV+ participants, n = 124	CIN2+ vs. Normal & CIN1	Cytology[c]	0.8255	0.7542 - 0.8969	-	-
		Microbiome[j]	0.7044	0.6124 - 0.7965	-	0.0350

Abbreviations: AUC, Area under the curve; CI, confidence interval; CIN, cervical intraepithelial neoplasia; HPV, human papillomavirus; ROC, receiver operating characteristic

[a] 95% confidence interval based on the asymptotic normal.
[b] P-value represents the equality of AUC based on a chi2 test, $P < 0.05$ was considered significant.
[c] Calculated based on categories for ASC-US, LSIL, HSIL or worse, and missing cytology results. Refer to methods section and Supplemental Table 3 for the linear scores.
[d] Calculated based on positivity for HPV16 and/or 18 and 12 other high-risk HPVs.
[e] Calculated based on the presence of 12/71 bacterial species, retained using a stepwise (forward selection) logistic regression (dependent variable: presence/absence of microbial species). Refer to methods and Supplemental Table 3 for retained species and the linear scores.
[f] Calculated based on the presence of the 12 bacterial species that were retained (refer to footnote e) and two additional predictors (positivity for HPV16 and/or 18 and 12 other high-risk HPVs). Refer to methods and Supplemental Table 3 for the linear scores.
[g] Calculated based on the presence of 5/77 bacterial species, retained using a stepwise (forward selection) logistic regression (dependent variable: presence/absence of microbial species). Refer to methods and Supplemental Table 4 for retained variables and the linear scores.
[h] Calculated based on the presence of the 5 bacterial species that were retained (refer to footnote g) and positivity for HPV16 and/or 18 and 12 other high-risk HPVs. Refer to methods and Supplemental Table 4 for the linear scores.
[i] Calculated based on 11/77 bacterial species, retained using a stepwise (forward selection) logistic regression (dependent variable: presence/absence of microbial species). Refer to methods and Supplemental Table 5 for retained species and the linear scores.
[j] Calculated based on 6/77 bacterial species, retained using a stepwise (forward selection) logistic regression (dependent variable: presence/absence of microbial species). Refer to methods and Supplemental Table 6 for retained variables and the linear scores. Interpret with caution due to model convergence concerns when selecting species.

4.2.11. Manuscript 2 Figures

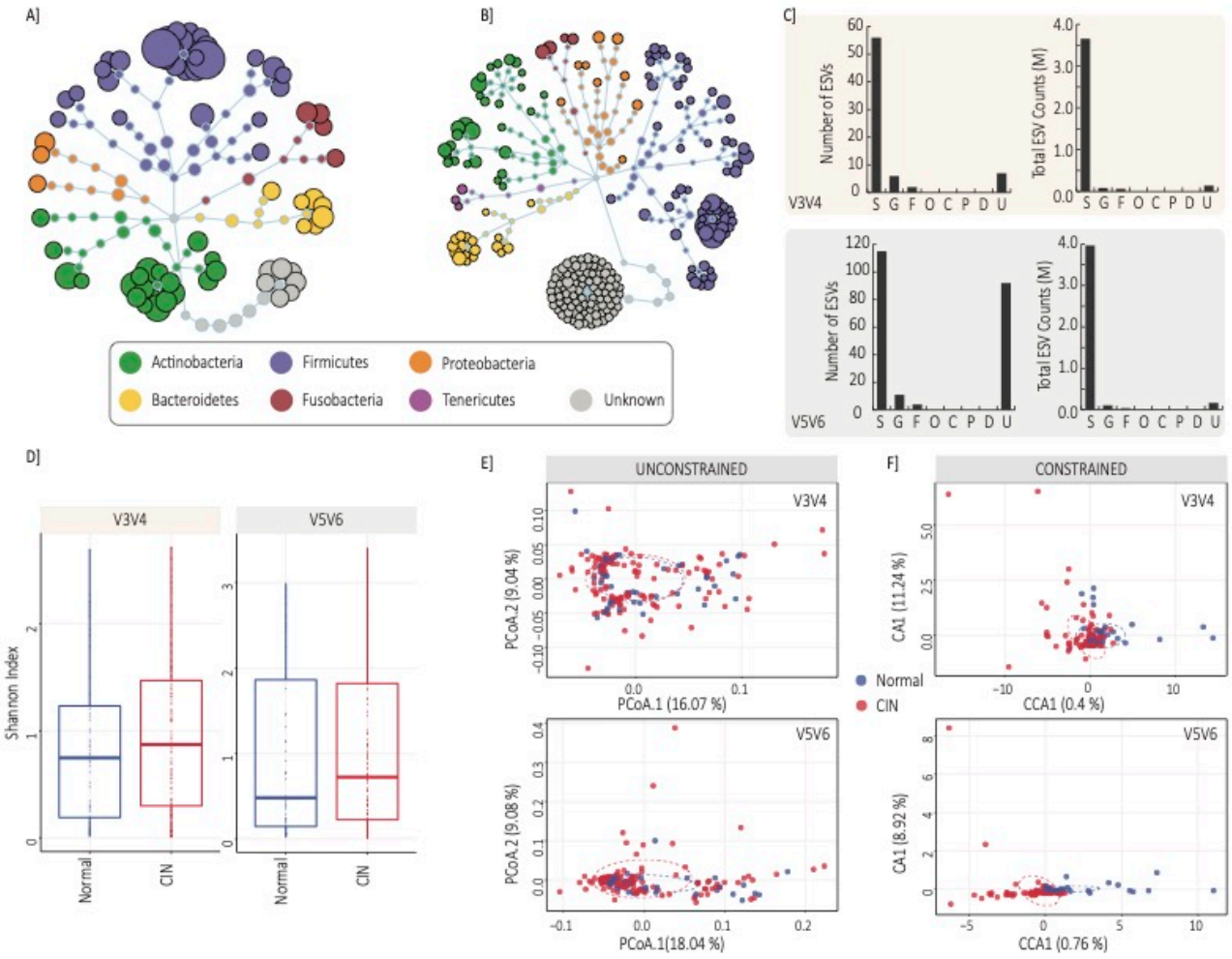


Figure 4-1. Descriptive bioinformatics results based on the V3-V4 and V5-V6 hypervariable regions.

Figure 4-1 Legend

A) Flower diagram representing V3-V4 microbial diversity (ESVs coloured by phyla)
B) Flower diagram representing V5-V6 microbial diversity (ESVs coloured by phyla)
C) ANCHOR annotation (ESVs and total counts are shown on the y-axis and their ESV annotation level on the x-axis)
D) Distribution of alpha diversity (Shannon index) comparing CIN grades to normal samples
E) Beta diversity between CIN and normal histologically-confirmed samples using unsupervised ordination (Permanova test, $p > 0.05$ for both regions)
F) Beta diversity between CIN and normal histologically-confirmed samples using supervised ordination (Permanova test, $p > 0.05$ for V3-V4, $p < 0.05$ for V5-V6)

Abbreviations: CIN, cervical intraepithelial neoplasia; ESV, exact sequence variants

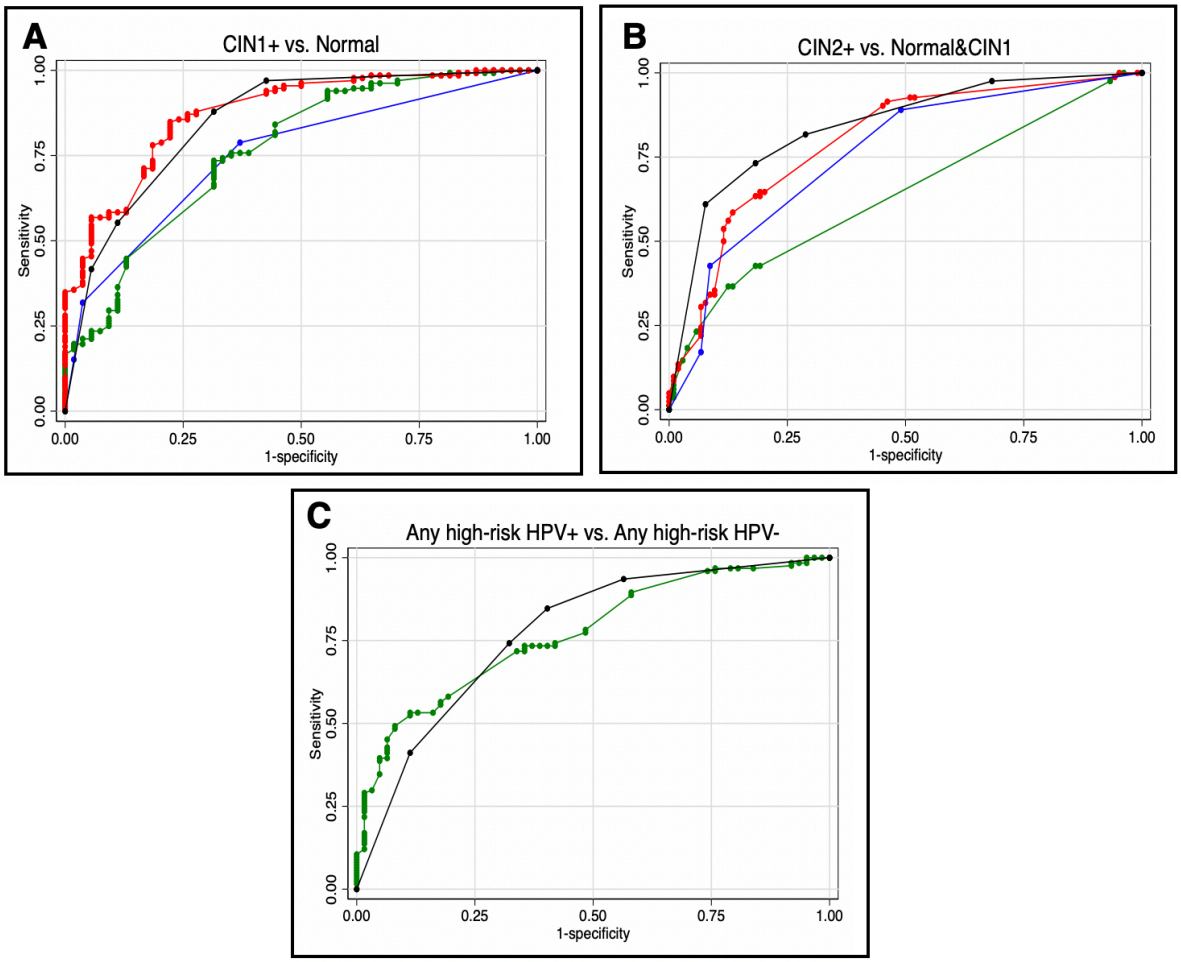


Figure 4-2. Performance of cytology-, HPV-, and microbiome- (species presence/absence) based scores to detect CIN and high-risk HPV.

Figure 4-2 Legend

The ROC curves plot the performance of cytology- (black), HPV- (blue), microbiome- (green), and microbiome-HPV- (red) based scores for the detection of CIN1+ (Panel A) and CIN2+ (Panel B). Panel C plots the performance of cytology- (black) and microbiome- (green) based scores for the detection of any high-risk HPV positivity. Refer to Table 3 footnotes for the calculation of the different scores.

Abbreviations: CIN, cervical intraepithelial neoplasia; HPV, human papillomavirus; ROC, receiver operating characteristic

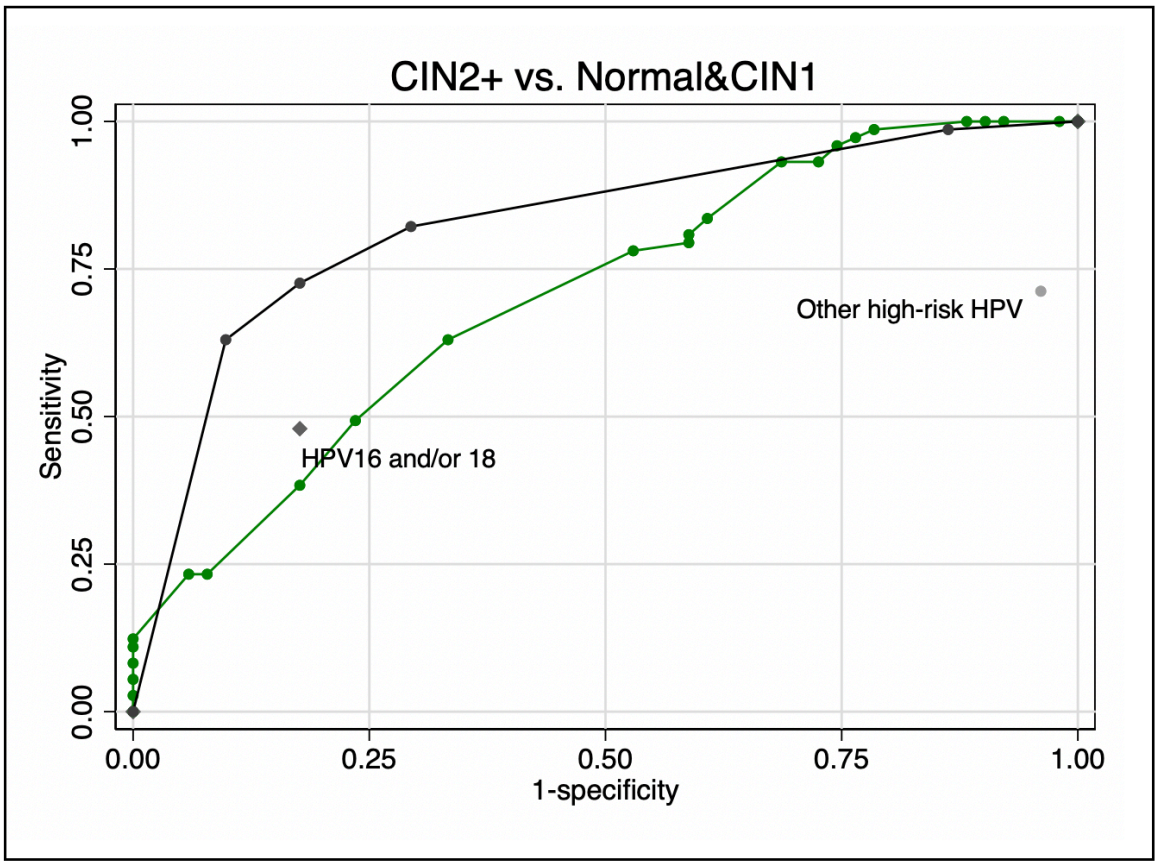


Figure 4-3. Performance of cytology- and microbiome- (species presence/absence) based scores to detect CIN2+ among women who tested HPV-positive.

Figure 4-3 Legend

The ROC curve plots the performance of cytology- (black) and microbiome- (green) based scores. Refer to Table 3 footnotes for the calculation of the scores. The sensitivity and specificity of HPV16 and/or 18 and those of other high-risk HPVs are denoted by the black diamond and grey circle, respectively.

Abbreviations: CIN, cervical intraepithelial neoplasia; HPV, human papillomavirus; ROC, receiver operating characteristic

CHAPTER 5: DISCUSSION

5.1. Key Findings

The manuscripts that constitute this thesis provide evidence that the bacterial composition of the CVM is implicated in HPV-associated cervical carcinogenesis; these findings could have important implications for cervical cancer prevention.

In manuscript 1 we conducted a narrative review on the relationship between CVM communities in HPV-associated cervical carcinogenesis by focusing on studies utilizing metagenomics for CVM characterization. Our findings are aligned with the assumption that *Lactobacillus* species colonize the vaginal microbiome of reproductive-aged women and produce substances which may aid in defending against disease;[6,70] after comparing epidemiological associations this state generally protected against HPV prevalence, acquisition, persistence and cervical lesions/cancer. By contrast a trend of high-diversity, or a loss of this dominant state increased the risk of adverse outcomes in cervical carcinogenesis. Nevertheless, the directionality and statistical significance of associations were inconsistent throughout the literature which could be attributed to differences in sample sizes, incomparable study populations, or misclassification bias. The majority of studies utilized 16S rRNA gene sequencing, and standard 16S algorithms infer sequences from parametric error models, which modify sequence nucleotides resulting in low taxonomic resolution.[21] Furthermore, numerous 16S databases which may contain errors and are subject to change as novel taxonomy is identified are available for taxonomic assignment. Errors in the identification of microbial communities could lead to incorrect classification of subjects according to their microbial exposure. Assuming that all women have the same chance of misclassification, this bias would be non-differential and effect estimates biased towards the null. Thus, across included studies the true associations between the CVM and outcomes of interest may have been masked. Although this bias is likely intrinsic to metagenomics as this technique relies on inferences regarding the microbial make-up of a sample, establishment of the most accurate microbial characterization technique with high taxonomic resolution as a harmonious method for CVM characterization could aid in minimizing biases.

In manuscript 1, only 8 of the identified studies assessed associations between individual bacterial species and the outcomes of interest.[78–85] Six of these investigations used 16S rRNA gene sequencing to identify bacterial species and if specified amplified the V1-V2, V1-V3, V3-V4, or V4 region of the 16S gene.[78,80–83,85] The remaining studies utilized PCR amplification or the *Allplex*™ Bacterian Vaginosis Assay.[79,84] Due to the low taxonomic resolution of standard 16S algorithms,[21] these studies likely failed to annotate all reads at the species-level. Moreover, the choice of primer set can lead to the identification of different bacterial species; for CVM characterization amplification of the V3-V4 region results in the identification of more bacterial taxa compared to V1-V2.[86] Thus, previous literature may not have comprehensively identified all of the bacterial species constituting the CVM and species implicated in HPV-associated cervical carcinogenesis may not have been detected nor identified as important predictors in these processes. In manuscript 2, we expanded upon the findings from these studies by amplifying at two 16S regions (V3-V4 and V5-V6) and using the high-resolution ANCHOR pipeline to analyze sequencing data.[21] ANCHOR is a 16S algorithm which does not rely on sequence modification and identifies microbial taxa at high resolution,[21] and in our study the majority of raw counts were annotated at the species-level. Moreover, we found that the V5-V6 region was more diverse than V3-V4 and after integrating the results from both regions, we comprehensively identified numerous bacterial species of the CVM (n = 77 overall, n = 8 unique to V3-V4, n = 48 unique to V5-V6). Different combinations of these 77 bacterial species were correlated with hrHPV infections and CIN. Overall, our unique methodology allowed us to identify several CVM species and demonstrate that CVM composition may differ in hrHPV and CIN samples compared to healthy controls.

A common objective of manuscript 1 and 2 was to assess the diagnostic value of the CVM in cervical carcinogenesis. In manuscript 1, 3 observational studies identified microbial taxa as excellent correlates of HPV16, hrHPV or CIN lesions (AUC range 0.802 – 0.952).[87-89] Similarly in manuscript 2, we performed our own analyses and found that bacterial species had an acceptable detection ability of hrHPV (AUC = 0.770) and CIN1+ (AUC = 0.766), whereas CIN2+ was lower (AUC = 0.641). The AUCs and 95% CIs from the identified studies (manuscript 1) and our original analyses (manuscript 2) are compared in **Table 5-1**. For the

detection of hrHPV, our results were in accordance with Morales et al.[88] However, our findings contradicted Lee et al.,[89] who demonstrated that bacterial species were excellent correlates of CIN2+. Discrepancies could be a result of dissimilarities in microbial characterization methods leading to the identification of different species (i.e., shogun metagenomics vs. 16S rRNA gene sequencing), and different microbial exposures (i.e., bacterial species abundance vs. presence/absence). Although we explored correlations based on abundance, modeling errors warrant caution in interpretation of these estimates and as such Table 5 only includes our findings based on the presence/absence of bacterial species. Although AUC estimates varied across studies, collectively these analyses suggest that different microbial components can discriminate between hrHPV infections and CIN lesions. Thus, there is evidence suggesting a potential role of the CVM in clinical practice. However, there is a need for critical evaluation and for further research due to a lack of longitudinal assessment and independent validation of the AUCs presented in both manuscripts.

Table 5-1. Comparison of the diagnostic performance of CVM components to detect prevalent HPV infections and/or CIN lesions across scientific studies.

Outcome	First Author (Year)	Exposure	Specific Outcome	AUC (95% CI)
HPV Prevalence	Logel (2023)[a]	11 bacterial species	hrHPV+	0.770 (0.702 – 0.838)
	Morales (2022)[88]	30 bacterial taxa	hrHPV+	0.802 (0.752 – 0.853)
	Yang (2020)[87]	17 genera	HPV16+	0.819 (0.684 – 0.954)
		7 species	HPV16+	0.918 (0.839 – 0.997)
CIN Lesions	Logel (2023)[a]	12 bacterial species	CIN1+ vs. Normal	0.766 (0.689 – 0.843)
		5 bacterial species	CIN2+ vs. CIN1-	0.641 (0.572 – 0.710)
	Lee (2020)[89]	Abundance of 33 bacterial species	CIN2+ vs. CIN1-	0.952 (0.820 – 1.000)

[a] Study not published, refers to the analyses presented in Chapter 4 of this thesis.

Abbreviations: AUC, area under the curve; CIN, cervical intraepithelial neoplasia; HPV, human papillomavirus; hrHPV, high-risk human papillomavirus.

In manuscript 2 we further investigated the clinical value of the CVM in HPV-associated cervical carcinogenesis by comparing the ability of CVM species to detect CIN lesions with cervical cytology and HPV DNA testing (i.e., common modalities for cervical cancer screening).

Notably, the presence of 12 bacterial species performed similarly to cervical cytology and hrHPV genotyping results for the detection of CIN1+. Similarly, combined positivity for 5 bacterial species and hrHPV exhibited a comparable ability to detect CIN2+ to that of cervical cytology. The congruency between microbial species and common screening modalities for the detection of CIN could have important implications in the clinical setting as there may be a future need for additional clinical tests. HPV vaccination coverage increased steadily from 2010 to 2019,[90] and the prevalence of HPV vaccine targeted genotypes as well as CIN2+ lesions have decreased substantially among HPV-vaccinated women followed for up to eight-years.[91] Assuming a continual increase in HPV vaccination coverage and decrease in HPV prevalence, the PPV and NPV of cytological assessment and HPV DNA testing will continue to decrease and increase, respectively. Despite these fluctuations in validity measurements, cervical cytology and HPV DNA testing will continue to be utilized for cervical cancer screening. However, microbial testing could be a complementary tool for the post-screening management of women. Follow-up testing for microbial species could aid in identifying women at highest risk for cervical lesions.

5.2. Strengths and Limitations

Our narrative review (manuscript 1) summarized the associations between the CVM and a variety of outcomes (HPV acquisition, prevalence, persistence, clearance, and biopsy/cytology confirmed lesions or cervical cancer). Although this was a narrative review, we systematically searched 3 databases and our search strategies were developed with the assistance of a librarian. An inherent strength was the exclusion of articles using microscopic techniques to characterize the CVM; relative to other reviews our restriction to metagenomic methods likely allowed for increased comparability of findings across studies. There was no clear trend suggesting that findings varied between cross-sectional and longitudinal studies. Nevertheless, the possibility of reverse causation from cross-sectional studies should be carefully considered when interpreting our findings. Our review may be limited due to a language restriction, and the exclusion of articles that did not provide a relative or absolute measure of association. Additionally, performing logistic regression analyses using the available data was beyond the scope of this review. A meta-analysis was not performed as we considered several exposures and outcomes which would have likely led to high heterogeneity. Moreover, pooling of CVM exposures may result in a loss of taxonomic resolution (i.e., due to a limited number of studies assessing

associations at the species-level, categorization would have likely occurred at the genus-level or by grouping into CSTs). Thus, meta-analyzing the data would have limited the comprehensiveness of our review. We recognize that there are several useful outcomes aside from the AUC which can be utilized to detect clinical performance. By limiting our search to studies conducting a ROC curve analysis, we may have missed important literature assessing the diagnostic ability of the CVM. However, our goal was to align manuscript 1 with the remainder of this thesis and effectively introduce the analysis conducted in manuscript 2 where AUCs are the primary outcome.

There are two major strengths in manuscript 2. Importantly, we characterized the CVM at high taxonomic resolution and identified several bacterial species putative to the CVM. Additionally, under-sampling of normal and CIN1 lesions and over-sampling of CIN2 and CIN3 lesions allowed for acceptable precision and power, despite a relatively small sample size (n = 186). The most important limitation of this study was the utilization of cross-sectional data. Moreover, independent validation was not performed as we did not have access to an external dataset with microbial data. Our estimates may be inflated as the stepwise selection of microbial species and ROC analyses were calculated with the same dataset. Additionally, the presented AUCs may have been biased towards cytology due to the utilization of a cytology-referral population. Despite these considerations, our findings are comparable to previous research suggesting that CVM taxa can detect hrHPV and CIN.[87–89]

5.3. Conclusion & Future Directions

This thesis provides evidence that CVM species are involved in HPV-associated cervical carcinogenesis. Notably, using a novel high-resolution 16S algorithm,[21] we show a cross-sectional correlation between CVM species and hrHPV as well as CIN. As a required next step, a prospective cohort study could allow for assessment of temporal changes in CVM species as a hrHPV infection is acquired, persists or develops into CIN and subsequently cervical cancer. This would allow for the establishment of a possible causal role of the CVM in cervical carcinogenesis. Clinically, HPV vaccination and screening are important strategies for cervical cancer prevention. However, due to inequities in their uptake (particularly in LMICs),[3] and the potential need for additional screening in the post-HPV vaccination era, establishment of CVM

species as a cause in cervical cancer could lead to the development of additional low-cost prevention techniques.

REFERENCES

1. Walboomers JMM, Jacobs M V., Manos MM, Bosch FX, Kummer JA, Shah K V., et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. J Pathol. 1999 Sep;189(1):12–9.

2. Trottier H, Franco EL. The epidemiology of genital human papillomavirus infection. Vaccine. 2006 Mar;24:S4–15.

3. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021 May 4;71(3):209–49.

4. Castellsagué X. Natural history and epidemiology of HPV infection and cervical cancer. Gynecol Oncol. 2008 Sep;110(3):S4–7.

5. Richardson H, Kelsall G, Tellier P, Voyer H, Abrahamowicz M, Ferenczy A, et al. The natural history of type-specific human papillomavirus infections in female university students. Cancer Epidemiol Biomarkers Prev. 2003 Jun;12(6):485–90.

6. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SSK, Mcculle SL, et al. Vaginal microbiome of reproductive-age women. Proc Natl Acad Sci U S A. 2011 Mar 15;108 Suppl 1(Suppl 1):4680-7.

7. Chen X, Lu Y, Chen T, Li R. The Female Vaginal Microbiome in Health and Bacterial Vaginosis. Front Cell Infect Microbiol. 2021 Apr 7;11:631972.

8. Gallo MF, Macaluso M, Warner L, Fleenor ME, Hook EW, Brill I, et al. Bacterial Vaginosis, Gonorrhea, and Chlamydial Infection Among Women Attending a Sexually Transmitted Disease Clinic: A Longitudinal Analysis of Possible Causal Links. Ann Epidemiol. 2012 Mar;22(3):213–20.

9. Atashili J, Poole C, Ndumbe PM, Adimora AA, Smith JS. Bacterial vaginosis and HIV acquisition: a meta-analysis of published studies. AIDS. 2008 Jul 31;22(12):1493–501.

10. King CC, Jamieson DJ, Wiener J, Cu-Uvin S, Klein RS, Rompalo AM, et al. Bacterial Vaginosis and the Natural History of Human Papillomavirus. Infect Dis Obstet Gynecol. 2011;2011:1–8.

11. Watts DH, Fazarri M, Minkoff H, Hillier SL, Sha B, Glesby M, et al. Effects of Bacterial Vaginosis and Other Genital Infections on the Natural History of Human Papillomavirus Infection in HIV-1–Infected and High-Risk HIV-1–Uninfected Women. J Infect Dis. 2005 Apr;191(7):1129–39.

12. Moscicki A-B, Ma Y, Jonte J, Miller-Benningfield S, Hanson E, Jay J, et al. The Role of Sexual Behavior and Human Papillomavirus Persistence in Predicting Repeated Infections with New Human Papillomavirus Types. Cancer Epidemiol Biomarkers Prev. 2010 Aug 1;19(8):2055–65.

13. Cherpes TL, Meyn LA, Krohn MA, Lurie JG, Hillier SL. Association between Acquisition of Herpes Simplex Virus Type 2 in Women and Bacterial Vaginosis. Clin Infect Dis. 2003 Aug 1;37(3):319–25.

14. Mortaki D, Gkegkes ID, Psomiadou V, Blontzos N, Prodromidou A, Lefkopoulos F, et al. Vaginal microbiota and human papillomavirus: a systematic review. J Turkish-German Gynecol Assoc. 2020 Sep 1;21(3):193–200.

15. Yang X, Da M, Zhang W, Qi Q, Zhang C, Han S. Role of Lactobacillus in cervical cancer. Cancer Manag Res. 2018 May;Volume 10:1219–29.

16. Sims TT, Colbert LE, Klopp AH. The Role of the Cervicovaginal and Gut Microbiome in Cervical Intraepithelial Neoplasia and Cervical Cancer. J Immunother Precis Oncol. 2021;4(2):72–8.

17. Gardella B, Pasquali MF, La Verde M, Cianci S, Torella M, Dominoni M. The Complex Interplay between Vaginal Microbiota, HPV Infection, and Immunological Microenvironment in Cervical Intraepithelial Neoplasia: A Literature Review. Int J Mol Sci. 2022 Jun 28;23(13):7174.

18. Brusselaers N, Shrestha S, van de Wijgert J, Verstraelen H. Vaginal dysbiosis and the risk of human papillomavirus and cervical cancer: systematic review and meta-analysis. Am J Obstet Gynecol. 2019 Jul;221(1):9-18.e8.

19. Tamarelle J, Thiébaut ACM, de Barbeyrac B, Bébéar C, Ravel J, Delarocque-Astagneau E. The vaginal microbiota and its association with human papillomavirus, Chlamydia trachomatis, Neisseria gonorrhoeae and Mycoplasma genitalium infections: a systematic review and meta-analysis. Clin Microbiol Infect. 2019 Jan;25(1):35–47.

20. Wang H, Ma Y, Li R, Chen X, Wan L, Zhao W. Associations of Cervicovaginal Lactobacilli With High-Risk Human Papillomavirus Infection, Cervical Intraepithelial Neoplasia, and Cancer: A Systematic Review and Meta-Analysis. J Infect Dis. 2019 Sep 13;220(8):1243–54.

21. Gonzalez E, Pitre FE, Brereton NJB. ANCHOR: a 16S rRNA gene amplicon pipeline for microbial analysis of multiple environmental samples. Environ Microbiol. 2019;21(7):2440–68.

22. Burchell AN, Winer RL, de Sanjosé S, Franco EL. Chapter 6: Epidemiology and

transmission dynamics of genital HPV infection. Vaccine. 2006 Aug;24:S52–61.

23. Trottier H, Burchell AN. Epidemiology of Mucosal Human Papillomavirus Infection and Associated Diseases. Public Health Genomics. 2009;12(5–6):291–307.

24. de Sanjosé S, Diaz M, Castellsagué X, Clifford G, Bruni L, Muñoz N, et al. Worldwide prevalence and genotype distribution of cervical human papillomavirus DNA in women with normal cytology: a meta-analysis. Lancet Infect Dis [Internet]. 2007 Jul;7(7):453–9.

25. WHO. INTERNATIONAL AGENCY FOR RESEARCH ON CANCER IARC Monographs on the Evaluation of Carcinogenic Risks to Humans VOLUME 90 Human Papillomaviruses [Internet]. Vol. 90, Iarc Monographs On The Evaluation Of Carcinogenic Risks To Humans. 2007. 670 p.

26. Muñoz N, Bosch FX, Castellsagué X, Díaz M, de Sanjose S, Hammouda D, et al. Against which human papillomavirus types shall we vaccinate and screen? the international perspective. Int J Cancer. 2004 Aug 20;111(2):278–85.

27. Schiffman M, Castle PE, Jeronimo J, Rodriguez AC, Wacholder S. Human papillomavirus and cervical cancer. Lancet. 2007 Sep;370(9590):890–907.

28. de Martel C, Plummer M, Vignat J, Franceschi S. Worldwide burden of cancer attributable to HPV by site, country and HPV type. Int J Cancer. 2017;141(4):664–70.

29. Burd EM. Human papillomavirus and cervical cancer. Vol. 16, Clinical Microbiology Reviews. American Society for Microbiology (ASM); 2003 Jan; 16(1):1-7.

30. Stanley M. Pathology and epidemiology of HPV infection in females. Gynecol Oncol. 2010 May;117(2):S5–10.

31. Stanley MA. Genital human papillomavirus infections: current and prospective therapies. J Gen Virol. 2012 Apr 1;93(4):681–91.

32. Narisawa-Saito M, Kiyono T. Basic mechanisms of high-risk human papillomavirus-induced carcinogenesis: Roles of E6 and E7 proteins. Cancer Sci. 2007 Oct 23;98(10):1505–11.

33. Carcinoma of the cervix and tobacco smoking: Collaborative reanalysis of individual data on 13,541 women with carcinoma of the cervix and 23,017 women without carcinoma of the cervix from 23 epidemiological studies. Int J Cancer. 2006 Mar 15;118(6):1481–95.

34. Cervical carcinoma and reproductive factors: Collaborative reanalysis of individual data on 16,563 women with cervical carcinoma and 33,542 women without cervical carcinoma from 25 epidemiological studies. Int J Cancer. 2006 Sep 5;119(5):1108–24.

35. Smith JS, Green J, de Gonzalez AB, Appleby P, Peto J, Plummer M, et al. Cervical cancer and use of hormonal contraceptives: a systematic review. Lancet. 2003 Apr;361(9364):1159–67.

36. Gravitt P, Winer R. Natural History of HPV Infection across the Lifespan: Role of Viral Latency. Viruses. 2017 Sep 21;9(10):267.

37. Stanley M. Chapter 17: Genital Human Papillomavirus Infections--Current and Prospective Therapies. JNCI Monogr. 2003 Jun 1;2003(31):117–24.

38. Stanley M. Immune responses to human papillomavirus. Vaccine. 2006 Mar;24:S16–22.

39. Schlecht NF, Kulaga S, Robitaille J, Ferreira S, Santos M, Miyamura RA, et al. Persistent human papillomavirus infection as a predictor of cervical intraepithelial neoplasia. JAMA. 2001 Dec 26;286(24):3106–14.

40. Rositch AF, Koshiol J, Hudgens MG, Razzaghi H, Backes DM, Pimenta JM, et al. Patterns of persistent genital human papillomavirus infection among women worldwide: A literature review and meta-analysis. Int J Cancer [Internet]. 2013 Sep 15;133(6):1271–85.

41. Loopik DL, Bentley HA, Eijgenraam MN, IntHout J, Bekkers RLM, Bentley JR. The Natural History of Cervical Intraepithelial Neoplasia Grades 1, 2, and 3: A Systematic Review and Meta-analysis. J Low Genit Tract Dis. 2021 Jul;25(3):221–31.

42. Malagón T, Volesky KD, Bouten S, Laprise C, El-Zein M, Franco EL. Cumulative risk of cervical intraepithelial neoplasia for women with normal cytology but positive for human papillomavirus: Systematic review and meta-analysis. Int J Cancer. 2020 Nov 15;147(10):2695–707.

43. World Health Organization. Weekly epidemiological record: Human papillomavirus vaccines: WHO position paper (2022 Update). Wkly Epidemiol Rec. 2022;97(50):645672.

44. Stanley M. Immunobiology of HPV and HPV vaccines. Gynecol Oncol. 2008 May;109(2):S15–21.

45. Logel M, Laurie C, El-Zein M, Guichon J, Franco EL. A Review of Ethical and Legal Aspects of Gender-Neutral Human Papillomavirus Vaccination. Cancer Epidemiol Biomarkers Prev. 2022 May 4;31(5):919–31.

46. Paavonen J, Jenkins D, Bosch FX, Naud P, Salmerón J, Wheeler CM, et al. Efficacy of a prophylactic adjuvanted bivalent L1 virus-like-particle vaccine against infection with

human papillomavirus types 16 and 18 in young women: an interim analysis of a phase III double-blind, randomised controlled trial. Lancet. 2007 Jun;369(9580):2161–70.

47. Quadrivalent Vaccine against Human Papillomavirus to Prevent High-Grade Cervical Lesions. N Engl J Med. 2007 May 10;356(19):1915–27.

48. Paavonen J, Naud P, Salmerón J, Wheeler C, Chow S-N, Apter D, et al. Efficacy of human papillomavirus (HPV)-16/18 AS04-adjuvanted vaccine against cervical infection and precancer caused by oncogenic HPV types (PATRICIA): final analysis of a double-blind, randomised study in young women. Lancet. 2009 Jul;374(9686):301–14.

49. Muñoz N, Manalastas R, Pitisuttithum P, Tresukosol D, Monsonego J, Ault K, et al. Safety, immunogenicity, and efficacy of quadrivalent human papillomavirus (types 6, 11, 16, 18) recombinant vaccine in women aged 24–45 years: a randomised, double-blind trial. Lancet. 2009 Jun;373(9679):1949–57.

50. Vesikari T, Brodszki N, van Damme P, Diez-Domingo J, Icardi G, Petersen LK, et al. A Randomized, Double-Blind, Phase III Study of the Immunogenicity and Safety of a 9-Valent Human Papillomavirus L1 Virus-Like Particle Vaccine (V503) Versus Gardasil® in 9–15-Year-Old Girls. Pediatr Infect Dis J. 2015 Sep;34(9):992–8.

51. Kavanagh K, Pollock KG, Cuschieri K, Palmer T, Cameron RL, Watt C, et al. Changes in the prevalence of human papillomavirus following a national bivalent human papillomavirus vaccination programme in Scotland: a 7-year cross-sectional study. Lancet Infect Dis. 2017 Dec;17(12):1293–302.

52. Palmer T, Wallace L, Pollock KG, Cuschieri K, Robertson C, Kavanagh K, et al. Prevalence of cervical disease at age 20 after immunisation with bivalent HPV vaccine at age 12-13 in Scotland: retrospective population study. BMJ. 2019 Apr 3;l1161.

53. World Health Organization (WHO). GLOBAL MARKET STUDY - HPV. 2022;(March):1–7.

54. Winer RL, Hughes JP, Feng Q, O'Reilly S, Kiviat NB, Holmes KK, et al. Condom Use and the Risk of Genital Human Papillomavirus Infection in Young Women. N Engl J Med. 2006 Jun 22;354(25):2645–54.

55. Nanda K, McCrory DC, Myers ER, Bastian LA, Hasselblad V, Hickey JD, et al. Accuracy of the papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: A systematic review. Ann Intern Med. 2000 May 16 [cited 2021 Apr 5];132(10):810–9.

56. Cuzick J, Clavel C, Petry K-U, Meijer CJLM, Hoyer H, Ratnam S, et al. Overview of the European and North American studies on HPV testing in primary cervical cancer

screening. Int J Cancer. 2006 Sep 1;119(5):1095–101.

57. Tota JE, Bentley J, Blake J, Coutlée F, Duggan MA, Ferenczy A, et al. Introduction of molecular HPV testing as the primary technology in cervical cancer screening: Acting on evidence to change the current paradigm. Prev Med (Baltim). 2017 May;98:5–14.

58. Mayrand M-H, Duarte-Franco E, Rodrigues I, Walter SD, Hanley J, Ferenczy A, et al. Human Papillomavirus DNA versus Papanicolaou Screening Tests for Cervical Cancer. N Engl J Med. 2007 Oct 18;357(16):1579–88.

59. Tota JE, Isidean SD, Franco EL. Defining benchmarks for tolerable risk thresholds in cancer screening: Impact of <scp>HPV</scp> vaccination on the future of cervical cancer screening. Int J Cancer. 2020 Dec 15;147(12):3305–12.

60. Franco EL, Mahmud SM, Tota J, Ferenczy A, Coutlée F. The expected impact of HPV vaccination on the accuracy of cervical cancer screening: the need for a paradigm change. Arch Med Res. 2009 Aug;40(6):478–85.

61. Ursell LK, Metcalf JL, Parfrey LW, Knight R. Defining the human microbiome. Nutr Rev. 2012 Aug;70:S38–44.

62. Ogunrinola GA, Oyewale JO, Oshamika OO, Olasehinde GI. The Human Microbiome and Its Impacts on Health. Int J Microbiol. 2020 Jun 12;2020:1–7.

63. Stern J, Miller G, Li X, Saxena D. Virome and bacteriome: two sides of the same coin. Curr Opin Virol. 2019 Aug;37:37–43.

64. Gacesa R, Kurilshikov A, Vich Vila A, Sinha T, Klaassen MAY, Bolte LA, et al. Environmental factors shaping the gut microbiome in a Dutch population. Nature. 2022 Apr 28;604(7907):732–9.

65. Raza MH, Gul K, Arshad A, Riaz N, Waheed U, Rauf A, et al. Microbiota in cancer development and treatment. J Cancer Res Clin Oncol. 2019 Jan 12;145(1):49–63.

66. Galloway-Peña J, Hanson B. Tools for Analysis of the Microbiome. Dig Dis Sci. 2020 Mar 31;65(3):674–85.

67. Zymo Research. 16S Sequencing VS Shotgun Metagenomic Sequencing [Internet]. Available from: https://zymoresearch.eu/blogs/blog/16s-sequencing-vs-shotgun-metagenomic-sequencing

68. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016 Jul 23;13(7):581–3.

69. Smith WL, Hedges SR, Mordechai E, Adelson ME, Trama JP, Gygax SE, et al. Cervical and Vaginal Flora Specimens Are Highly Concordant with Respect to Bacterial Vaginosis-Associated Organisms and Commensal Lactobacillus Species in Women of Reproductive Age. Forbes BA, editor. J Clin Microbiol. 2014 Aug;52(8):3078–81.

70. Boskey ER, Cone RA, Whaley KJ, Moench TR. Origins of vaginal acidity: high d/l lactate ratio is consistent with bacteria being the primary source. Hum Reprod. 2001 Sep;16(9):1809–13.

71. Allsworth JE, Peipert JF. Prevalence of Bacterial Vaginosis. Obstet Gynecol. 2007 Jan;109(1):114–20.

72. Brotman RM. Vaginal microbiome and sexually transmitted infections: an epidemiologic perspective. J Clin Invest. 2011 Dec 1;121(12):4610–7.

73. Nugent RP, Krohn MA, Hillier SL. Reliability of diagnosing bacterial vaginosis is improved by a standardized method of gram stain interpretation. J Clin Microbiol. 1991 Feb;29(2):297–301.

74. Amsel R, Totten PA, Spiegel CA, Chen KCS, Eschenbach D, Holmes KK. Nonspecific vaginitis. Am J Med. 1983 Jan;74(1):14–22.

75. Platz-Christensen JJ, Larsson P-G, Sundström E, Wlqvlst N. Detection of bacterial vaginosis in wet mount, Papanicolaou stained vaginal smears and in Gram stained smears. Acta Obstet Gynecol Scand. 1995 Jan;74(1):67–70.

76. Usyk M, Schlecht NF, Pickering S, Williams L, Sollecito CC, Gradissimo A, et al. molBV reveals immune landscape of bacterial vaginosis and predicts human papillomavirus infection natural history. Cortes B Jimenez SE, Rodriguez AC, Hildesheim A, Kreimer AR, Lowy DR, Schiffman M, Schiller JT, Sherman M, Wacholder S, Kemp TJ, Sidawy MK, Quint W, van Doorn LJ, Struijk L, Palefsky JM, Darragh TM, Stoler MH GP, editor. Nat Commun. 2022;13(1):233.

77. El-Zein M, Cheishvili D, Gotlieb W, Gilbert L, Hemmings R, Behr MA, et al. Genome-wide DNA methylation profiling identifies two novel genes in cervical neoplasia. Int J Cancer. 2020;147(5):1264–74.

78. So KA, Yang EJ, Kim NR, Hong SR, Lee J-H, Hwang C-S, et al. Changes of vaginal microbiota during cervical carcinogenesis in women with human papillomavirus infection. Consolaro MEL, editor. PLoS One. 2020 Sep 17;15(9):e0238705.

79. McClymont E, Albert AY, Wang C, Dos Santos SJ, Coutlée F, Lee M, et al. Vaginal microbiota associated with oncogenic HPV in a cohort of HPV-vaccinated women living

with HIV. Int J STD AIDS. 2022 Aug 1;33(9):847–55.

80. McKee KS, Carter KA, Bassis C, Young VB, Reed B, Harper DM, et al. The vaginal microbiota, high-risk human papillomavirus infection, and cervical cytology: results from a population-based study. Gynecol Pelvic Med. 2020 Sep;3:18–18.

81. 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. Sci Rep. 2018;8(1):1–9.

82. 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 Med. 2018 May 2;7(5):2172–9.

83. Oh HY, Kim B-S, Seo S-S, Kong J-S, Lee J-K, Park S-Y, et al. The association of uterine cervical microbiota with an increased risk for cervical intraepithelial neoplasia in Korea. Clin Microbiol Infect. 2015 Jul;21(7):674.e1-674.e9.

84. Rosário A, Sousa A, Varandas T, Marinho-Dias J, Medeiros R, Martins G, et al. Impact of cervicovaginal microbiome on the risk of cervical abnormalities development. J Med Virol. 2023 May 3;95(5).

85. 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. J Infect Dis. 2016 Nov 1;214(9):1361–9.

86. Graspeuntner S, Loeper N, Künzel S, Baines JF, Rupp J. Selection of validated hypervariable regions is crucial in 16S-based microbiota studies of the female genital tract. Sci Rep. 2018 Jun 26;8(1):9678.

87. Yang Q, Wang Y, Wei X, Zhu J, Wang X, Xie X, et al. The Alterations of Vaginal Microbiome in HPV16 Infection as Identified by Shotgun Metagenomic Sequencing. Front Cell Infect Microbiol. 2020 Jun 23;10.

88. Andrade Pessoa Morales J, Marconi C, El-Zein M, Ravel J, da Silva Pinto GV, Silveira R, et al. Vaginal microbiome components as correlates of cervical human papillomavirus infection. J Infect Dis. J Infect Dis. 2021;226(6).

89. Lee YH, Kang G-U, Jeon SY, Tagele SB, Pham HQ, Kim M-S, et al. Vaginal Microbiome-Based Bacterial Signatures for Predicting the Severity of Cervical Intraepithelial Neoplasia. Diagnostics. 2020 Nov 26;10(12):1013.

90. Bruni L, Saura-Lázaro A, Montoliu A, Brotons M, Alemany L, Diallo MS, et al. HPV vaccination introduction worldwide and WHO and UNICEF estimates of national HPV

immunization coverage 2010–2019. Prev Med (Baltim). 2021;144.

91. Drolet M, Bénard É, Pérez N, Brisson M, Ali H, Boily M-C, et al. Population-level impact and herd effects following the introduction of human papillomavirus vaccination programmes: updated systematic review and meta-analysis. Lancet. 2019 Aug;394(10197):497–509.

92. El-Zein M, Bouten S, Louvanto K, Gilbert L, Gotlieb W, Hemmings R, et al. Validation of a new HPV self-sampling device for cervical cancer screening: The Cervical and Self-Sample In Screening (CASSIS) study. Gynecol Oncol. 2018 Jun;149(3):491

APPENDICES

A.1. Manuscript 1 Supplemental Material

S-Table 3-1. Search strategies to examine the etiological and clinical role of the CVM in HPV-associated cervical carcinogenesis.

Database *(Purpose)*	**Search #**	**Exposure or Outcome**	**Search Terms**
Medline *(Etiologic role)*	**#1**	HPV	Papillomavirus Infections/ or exp Alphapapillomavirus/
	#2	HPV	(human papillomavirus or HPV* or papillomavirus infection*).mp.
	#3	Cervical Cancer	Uterine Cervical Neoplasms/
	#4	Cervical Cancer	((cervi* or uterine) adj (cancer or neoplasm*)).mp.
	#5	Human Papillomavirus **OR** Cervical Cancer	#1 OR #2 OR #3 OR #4
	#6	Cervicovaginal Microbiome	((cervi* or vagina*) adj microbio*).mp. or microbiota/
	#7	Cervicovaginal Microbiome	Lactobacillus/ or lactobacillus.mp. or Vaginosis, Bacterial/ or bacterial vagin*.mp.
	#8	Microbiome	#6 or #7
	#9	HPV, Cervical Cancer and Microbiome	#5 and #8
Embase *(Etiologic role)*	**#1**	HPV	"Papillomavirus Infection"/ or exp Alphapapillomavirus/
	#2	HPV	("human papillomavirus" or HPV* or "papillomavirus infection*").mp.
	#3	Cervical Cancer	exp uterine cervix cancer/
	#4	Cervical Cancer	((cervi* or uterine) adj (cancer or neoplasm*)).mp.
	#5	HPV **OR** Cervical Cancer	#1 OR #2 OR #3 OR #4
	#6	Cervicovaginal Microbiome	((cervi* or vagina*) adj microbio*).mp. or exp microbiome/
	#7	Cervicovaginal Microbiome	exp Lactobacillus/ or lactobacillus.mp. or "bacterial vagin*".mp.
	#8	Microbiome	#6 or #7
	#9	HPV, Cervical Cancer and Microbiome	#5 and #8
Web of Science *(Etiologic role)*	**-**	-	((ALL=("human papillomavirus" OR HPV* OR "papillomavirus infection*")) OR (ALL=((cervi* OR uterine) "NEAR/0" (cancer OR neoplasm*)))) AND ((ALL=((cervi* OR vagina*) "NEAR/0" microbio*)) OR (ALL=lactobacillus OR ALL="bacterial vagin*"))
PubMed *(Clinical role)*	-	-	((AUC[Title/Abstract] OR ROC[Title/Abstract]) AND (microbiome[Title/Abstract] OR cervicovaginal microbiome[Title/Abstract] OR vaginal microbiome[Title/Abstract])) AND (HPV[Title/Abstract] OR human papillomavirus[Title/Abstract] OR cervical cancer[Title/Abstract] OR cervical intraepithelial neoplasia[Title/Abstract])

S-Table 3-2. Observational studies on the association between the CVM and HPV prevalence, acquisition, persistence, clearance and/or cytology interpretations or biopsy confirmed CIN and cervical cancer.

First Author (Year)	Study Population (sample size [n] of participants) Age	Microbial Characterization Method Region Sequencing	HPV Genotyping Method HPV Types	Microbiome Categorization	Exposure	Outcome	Crude Effect (95%CI)	Adjusted Effect (95% CI)	Variable Adjustment
Musa (2023)	Women being screened for cervical cancer, colposcopy or evaluation for ICC (n = 151) Ages, median 52	16S rRNA V3-V4 Illumina MiSeq	Anyplex™ II HPV28 detection kit 28 HPV types (hrHPVs 16, 18, 26, 31, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 68, 69, 73, 82; lrHPVs 6, 11, 40, 42, 43, 44, 54, 61, 70)	VALENICA algorithm to cluster into 5 CST types (CST I, *Lactobacillus crispatus*; CST II, *L.gasseri*; CST III, *L.iners*; CST IV, diverse; CST-V, *L.jensenii*)	CST IV vs. CST I	HSIL/ICC vs. NILM/LSIL	OR, 1.79 (0.63-5.04)	aOR, 1.31 (0.39-4.41)	Age, HIV, HPV, education, births
					CST III vs. CST I	HSIL/ICC vs. NILM/LSIL	OR, 1.2 (0.37-3.87)	aOR, 1.13 (0.27-4.67)	
					CST IV vs. CST I	HSIL/ICC vs. NILM/LSIL	NR	aOR, 0.89 (0.17-4.39), based on 53 HPV- women	HIV, HIV, education, births
					CST IV vs. CST I	HSIL/ICC vs. NILM/LSIL	NR	aOR, 3.64 (0.63-20.9), based on 62 HPV+ women	
					CST III vs. CST I	HSIL/ICC vs. NILM/LSIL	NR	aOR, 0.94 (0.15-6.00), based on 53 HPV- women	
					CST III vs. CST I	HSIL/ICC vs. NILM/LSIL	NR	aOR, 6.69 (0.67-66.6), based on 62 HPV+ women	
Rosário (2023)[a,b]	hrHPV positive women with cytological assessment from an organized cervical cancer screening program (n = 807)	*Allplex*™ Bacterian Vaginosis Assay NA NA	Anyplex II HPV HR Detection kit 14 HPV types (hrHPVs 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68)	BV panel (Detection of *Megasphaera* Type 1, *Lactobacillus* spp., *Bacteroids fragilis*, *Gardnerella vaginalis*, Bacteria associated to BV,	*Lactobacillus* spp.	Multiple hrHPV	RR, 0.69 (0.49-0.96)	NR	NA
					Gardnerella vaginalis	Multiple hrHPV	RR, 0.71 (0.52-0.98)	NR	
					Lactobacillus spp.	HPV16/18	RR, 0.51 (0.34-0.76)	NR	NA
					Gardnerella vaginalis	HPV16/18	RR, 0.50 (0.34-0.76)	NR	
					BV panel	HPV16/18	RR, 0.60	NR	

				Atopobium vaginae and *Mobiluncus* spp.)			(0.43-0.84)		
	Ages, 25-60				*Lactobacillus* spp.	HPV 9-val	RR, 0.50 (0.36-0.69)	NR	NA
					Gardnerella vaginalis	HPV 9-val	RR, 0.68 (0.50-0.92)	NR	
					Mobiluncus spp.	HPV 9-val	RR, 1.85 (1.07-3.20)	NR	
					BV panel	HPV 9-val	RR, 0.59 (0.43-0.81)	NR	
					Megaphaera Type 1	Cervical abnormalities vs. NILM	OR, 0.19 (0.12-0.30)	aOR, 1.64 (0.48-5.62)	Age, hrHPV
					Lactobacillus spp.	Cervical abnormalities vs. NILM	OR, 0.29 (0.21-0.40)	aOR, 0.33 (0.18-0.60)	
					Gardnerella vaginalis	Cervical abnormalities vs. NILM	OR, 0.11 (0.08-0.16)	aOR, 0.41 (0.21-0.82)	
					Bacteria associated to BV	Cervical abnormalities vs. NILM	OR, 0.23 (0.15-0.35)	aOR, 1.25 (0.34-4.58)	
					Atopobium vaginae	Cervical abnormalities vs. NILM	OR, 0.43 (0.32-0.58)	aOR, 0.53 (0.30-0.95)	
					Mobiluncus spp.	Cervical abnormalities vs. NILM	OR, 0.23 (0.14-0.38)	aOR, 0.29 (0.10-0.83)	
					Megaphaera Type 1	Cervical lesions vs. NILM + ASC-US	OR, 0.30 (0.19-0.49)	aOR, 0.63 (0.26-1.52)	Age, hrHPV
					Lactobacillus spp.	Cervical lesions vs. NILM + ASC-US	OR, 0.25 (0.17-0.35)	aOR, 0.27 (0.18-0.40)	
					Gardnerella vaginalis	Cervical lesions vs. NILM + ASC-US	OR, 0.16 (0.11-0.23)	aOR, 0.29 (0.19-0.46)	
					Bacteria associated to	Cervical lesions vs.	OR, 0.34 (0.21-0.54)	aOR, 1.22 (0.50-3.00)	

					BV	NILM + ASC-US			
					Atopobium vaginae	Cervical lesions vs. NILM + ASC-US	OR, 0.35 (0.25-0.48)	aOR, 0.77 (0.51-1.17)	
					Mobiluncus spp.	Cervical lesions vs. NILM + ASC-US	OR, 0.30 (0.17-0.52)	aOR, 0.55 (0.26-1.14)	
					Megaphaera Type 1	ASC-H/HSIL vs. NILM + ASC-US + LSIL	OR, 0.56 (0.31-1.00)	aOR, 1.66 (0.48-5.67)	Age, hrHPV. *Ureoplasma parvum, Chlamydia trachomatis*
					Lactobacillus spp.	ASC-H/HSIL vs. NILM + ASC-US + LSIL	OR, 0.24 (0.14-0.40)	aOR, 0.30 (0.17-0.55)	
					Gardnerella vaginalis	ASC-H/HSIL vs. NILM + ASC-US + LSIL	OR, 0.24 (0.14-0.39)	aOR, 0.42 (0.21-0.84)	
					Bacteria associated to BV	ASC-H/HSIL vs. NILM + ASC-US + LSIL	OR, 0.52 (0.29-0.96)	aOR, 1.24 (0.34-4.55)	
					Atopobium vaginae	ASC-H/HSIL vs. NILM + ASC-US + LSIL	OR, 0.35 (0.22-0.56)	aOR, 0.55 (0.31-0.99)	
					Mobiluncus spp.	ASC-H/HSIL vs. NILM + ASC-US + LSIL	OR, 0.31 (0.13-0.73)	aOR, 0.31 (0.11-0.91)	

McClymont (2022)	Women and girls with HIV enrolled in an HPV vaccine study to receive three vaccine dosages (Enrollment, NR; Follow-up, maximum 8 years) (n = 172) Ages, ≥ 9	PCR amplification N/A MiSeq	Roche Linear Array 36 HPV types, (Oncogenic HPVs 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, and 82)	Hierarchical clustering into six CST types (CST IVA, diversity, *Lactobacillus* and *Megasphaera* depletion*;* CST IVC, high RA of *Gardnerella vaginalis* and *Gardnerella swidskinskii*; CST IVD.1, high RA of *Megasphaera, Clostridiales* spp., *Prevotella* spp,, *Dialister pneuosintes* and *Porphyromonas uenonis*; CST IVD.2, depleted of *Megasphaera* and high RA of *Clostridiales* sp., *Prevotella* spp., and *Porphyromonas uenonis*; CST III/V, high RA of *Lactobacillus iners* and/or *Lactobacillus jensenii;* CST I dominated by *Lactobacillus crispatus*)	*Gardnerella Swidinski* RA	Incident oncogenic HPV	OR, 1.10 (0.98-1.22)	NR	NA
					Lactobacillus crispatus RA	Incident oncogenic HPV	OR, 0.91 (0.84-1.01)	NR	
Usyk (2022)	Women in the placebo arm of an HPV vaccine trial with an incident hrHPV infection (Enrollment, 2004-2005; Follow-up, 2 visits)	16S rRNA V4 Illumina MiSeq	AmpliTaq DNA polymerase (MY-Taq)[c] 12 HPV types (HPVs 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58 and/or	*molBV* (A 16S amplicon score for BV with equivalent accuracy to the Nugent score for BV diagnosis)	*molBV* became high vs. sustained low[d,e]	hrHPV clearance	NR	aHR, 0.84 (0.51-1.38)	Age, smoking status, HPV16
					molBV became low vs. sustained low[d,e]	hrHPV clearance	NR	aHR, 0.55 (0.30-0.97)	

	(n = 307) Ages, NR		59)		*molBV* sustained high vs. sustained low[d,e]	hrHPV clearance	NR	aHR, 0.85 (0.56-1.28)	
					Continuous *molBV* score (visit 2)	Progression to CIN2+ following visit 2	NR	aOR, 1.24 (1.02-1.55)	Age, smoking status, HPV16
Zhang (2022)	Chinese women, unvaccinated against HPV (n = 356) Ages, 20–70	16S rRNA NR NovaSeq	Hybrid Capture 2 assay NR	Heatmap analysis into clusters (Cluster I, *Lactobacillus crispatus* dominant; Cluster II, *L. gasseri* dominant; Cluster III, *L. iners* dominant; Cluster IV-A, modest proportions of either *L. crispatus, L. iners* or *Lactobacillus* spp.; Cluster IV-B, non-*Lactobacillus* dominant)	CST Type	HPV	NR	aOR, 0.74 (0.59-0.93)	Age, pregnancy, gestation, number of sexual partners, smoking, vaginal douche, contraception, IUD implantation
					CST Type	CIN	NR	aOR, 1.09 (0.85-1.41)	Age, pregnancy, gestation, number of sexual partners, smoking, vaginal douche, contraception, IUD implantation, HPV infection
Dareng (2020)	HIV-negative and -positive Nigerian women (Enrollment 2012-2013; Follow-up, baseline and 6-months for sample collection) (n = 211), a total of 353 samples for microbial	16S rRNA V3-V4 Illumina MiSeq	$SPF_{10}LiPA_{25}$ System 25 HPV Types (hrHPVs 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59)	Hierarchical clustering into CST types (CST I, *Lactobacillus crispatus*; CST II, *Lactobacillus gasseri*; CST III, *Lactobacillus iners*; CST I-B, moderately high proportions of	≥70% vs. < 70% *Lactobacillus* dominant microbiota in HIV-	hrHPV persistence	NR	aOR, 0.35 (0.14-0.89)	Age
					≥ 70% vs. < 70% *Lactobacillus* dominant microbiota in	hrHPV persistence	NR	aOR, 1.25 (0.73-2.14)	

	assessment Ages, mean 37.8			*Lactobacillus* species and low proportions of pathobionts; CST IV-B; high diversity, low proportions of *Lactobacillus* and pathobionts and BV associated taxa)	HIV+				
					≥70% vs. < 70% *Lactobacillus crispatus* dominant microbiota in HIV-	hrHPV persistence	NR	aOR, 0.22 (0.03-1.43)	
					≥70% vs. < 70% *Lactobacillus crispatus* dominant microbiota in HIV+	hrHPV persistence	NR	aOR, 1.20 (0.82-1.76)	
					CST I-B vs. CST IV-B in HIV-	hrHPV persistence	NR	aOR, 0.95 (0.50-1.82)	
					CST III vs. CST IV-B in HIV-	hrHPV persistence	NR	aOR, 0.67 (0.28-1.61)	
					CST III vs. CST IV-B in HIV+	hrHPV persistence	NR	aOR, 1.03 (0.39-2.70)	
					CST II vs. CST IV-B in HIV-	hrHPV persistence	NR	aOR, 0.28 (0.04-2.02)	
					CST II vs. CST IV-B in HIV+	hrHPV persistence	NR	aOR, 1.0 (0.40-2.55)	
					CST I vs. CST IV-B in HIV-	hrHPV persistence	NR	aOR, 0.29 (0.06-1.46)	
					CST I vs. CST IV-B in HIV+	hrHPV persistence	NR	aOR, 1.30 (0.40-2.54)	
McKee (2020)	Cytology samples from women attending a clinic in Ohio or West Virginia	16S rRNA V4 IlluminaMiSeq	NR	Partitioning into 3 CST Types (CST 1, *Lactobacillus crispatus;* CST 2,	CST 2 vs. CST 1	hrHPV+ vs. NILM/HPV-	OR, 0.61 (0.28-1.34)	aOR, 0.67 (0.29-1.57), based on 284 samples	Age, race, current smoking, >= 2 male partners in the past year

	(n = 308) Age, mean 26			*Lactobacillus iners;* CST 3, diverse with a higher RA of *Gardnerella vaginalis*)	CST 3 vs. CST 1	hrHPV+ vs. NILM/HPV-	OR, 1.44 (0.65-3.20)	aOR, 1.53 (0.62-3.76), based on 284 samples	
					Lactobacillus gasseri presence vs. absence	hrHPV+ vs. NILM/HPV-	OR, 0.37 (0.20-0.70)	aOR, 0.50 (0.25-1.02), based on 257 samples	Age, race, smoking, >= 2 male partners in the past year, current condoms use
					CST 2 vs. CST 1	Abnormal cytology vs. NILM/HPV-	OR, 0.65 (0.29-1.46)	aOR, 0.67 (0.28-1.59), based on 284 samples	Age, race, current smoking, >= 2 male partners in the past year
					CST 3 vs. CST 1	Abnormal cytology vs. NILM/HPV-	OR, 1.65 (0.74-3.70)	aOR, 1.63 (0.66-4.03), based on 284 samples	
					Lactobacillus gasseri presence vs. absence	Abnormal cytology vs. NILM/HPV-	OR, 0.61 (0.32-1.16)	aOR, 0.88 (0.42-1.83), based on 257 samples	Age, race, smoking, >= 2 male partners in the past year, current condoms use
Mitra (2020)	Women with histologically confirmed CIN2 lesions at clinics in Northern California (Enrollment, 2002-2007; Follow-up, 3 visits baseline, 12 months and 24 months) (n = 87), a total of 573 samples	16S rRNA V1-V2 Illumina MiSeq	Roche Linear Array 37 HPV types (HPVs 6, 11, 16, 18, 26, 31, 33, 35, 39, 40, 42, 45, 51, 52, 53, 54, 55, 56, 58, 59, 61, 62, 64, 66, 67, 68, 69, 70, 71, 72, 73, 81, 82, 83, 84 and 89)	Genus level: Hierarchical clustering into *Lactobacillus* dominant and *Lactobacillus* depleted microbiome Species level: Hierarchical clustering into 5 vaginal CSTs (CST I, *Lactobacillus*	*Lactobacillus* depleted vs. dominant (Follow up 0-12 moths)	CIN2 non-regression vs. regression	OR, 3.21 (1.32-8.22)	aOR, 3.56 (1.31-9.60)	Age, ethnicity, smoking, douching, contraception, HPV16/18 positivity
					Lactobacillus depleted vs. dominant (Follow up 0-24 moths)	CIN2 non-regression vs. regression	OR, 2.50 (0.95-6.59)	aOR, 2.85 (1.03-7.92)	
					Lactobacillus depleted vs. *Lactobacillus* dominant	CIN2 non-regression vs. regression	OR, 2.50 (0.61-10.26)	aOR, 3.06 (0.54-17.14)	

	Ages, 16-26			*crispatus*; CST II, *Lactobacillus* gasseri; CST III, *Lactobacillus iners*; CST IV, *Lactobacillus* species depletion; CST V, *Lactobacillus jensenii*)	(Follow up 12-24 moths)				
					CST III vs. CST I (Follow up 0-12 moths)	CIN2 non-regression vs. regression	OR, 1.30 (0.43-3.90)	aOR, 1.14 (0.32-4.05)	
					CST IV vs. CST I (Follow up 0-12 moths)	CIN2 non-regression vs. regression	OR, 3.79 (1.17-12.30)	aOR, 3.85 (1.10-13.42)	
					CST III vs. CST I (Follow up 0-24 moths)	CIN2 non-regression vs. regression	OR, 2.00 (0.47-8.41)	aOR, 1.86 (0.39-8.81)	
					CST IV vs. CST I (Follow up 0-24 moths)	CIN2 non-regression vs. regression	OR, 4.00 (0.96-16.61)	aOR, 4.25 (0.98-18.50)	
					CST III vs. CST I (Follow up 12-24 moths)	CIN2 non-regression vs. regression	OR, 1.82 (0.15-20.71)	aOR, 1.30 (0.06-27.07)	
					CST IV vs. CST I (Follow up 12-24 moths)	CIN2 non-regression vs. regression	OR, 4.67 (0.40-53.95)	aOR, 4.94 (0.26-94.86)	
So (2020)	Women from the department of obstetrics and gynecology at a hospital in Korea (n = 50) Ages, 20-50	16S rRNA V3-V4 MiSeq platform	NR	Presence or absence of bacterial species	*Atopobium vaginae* + vs. -	CIN2/3 or cervical cancer vs. Normal	OR, 4.33 (1.15-16.32)	NR	N/A
					Dialister invisus + vs. -	CIN2/3 or cervical cancer vs. Normal	OR, 4.89 (1.20-19.94)	NR	
					Finegoldia magna + vs. -	CIN2/3 or cervical cancer vs. Normal	OR, 6.00 (1.08-33.27)	NR	
					Gardnerella vaginalis + vs. -	CIN2/3 or Cervical Cancer vs. Normal	OR, 7.43 (1.78-31.04)	NR	
					Prevotella buccalis	CIN2/3 or cervical cancer	OR, 11.00 (2.00-	NR	

					+ vs. -	vs. Normal	60.57)		
					Prevotella timonensis + vs. -	CIN2/3 or cervical cancer vs. Normal	OR, 6.00 (1.46-24.69)	NR	
Usyk (2020)	Placebo arm of an HPV vaccine trial with an incident hrHPV infection (Enrollment, 2004-2005; Follow-up, 2 visits) (n =273), a total of 539 samples with only 266 women with data at 2 visits Ages, 18-25	16S rRNA V4 Illumina MiSeq	AmpliTaq DNA polymerase (MY-Taq)[c] 12 HPV types (HPVs 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59)	Hierarchical clustering into four CST types (First CST, *Lactobacillus iners* dominant; Second CST, *Lactobacillus crispatus* dominant; Third CST, *Gardnerella vaginalis* dominant; Fourth CST, high diversity)	*Lactobacillus* abundance (visit 1)	Progression to CIN2+ vs. clearance	NR	aOR, 0.41 (0.22-0.79)	Age, CST, smoking, HPV16, *Gardnerella* abundance (visit 1), fungal observed operational taxonomic units (visit 1), cell motility (visit 1)
					Microbial diversity measured by the Shannon index (visit 2)	Progression to CIN2+ vs. clearance	NR	aOR, 1.17 (1.02-1.29)	
Van de Wijgert (2020)[f]	HIV positive women living in Johannesburg with CIN1- at both visits (for hrHPV outcomes) or CIN2+ at one visit (for cases in the CIN outcomes) (Enrollment, 2011-2012; Follow-up, every 6 months for a median of 16 months; microbial analyses using baseline and samples at the end of follow-up) (n = 304), for tabulated estimates with hrHPV	16S rRNA V3-V4 Illumina HiSeq	LiPA HPV Genotyping Extra 13 HPV types (hrHPVs 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and 68)	Hierarchal clustering to identify 7 VMB types (1, *Lactobacillus iners* dominant; 2, *Lactobacillus crispatus* or *Lactobacillus jensenii* dominant; 3, *Lactobacilli* and BV-anaerobes; 4, High diversity BV-anaerobes; 5, BV-anaerobe dominant; 6, Pathobionts-characterized; 7, *Bifidobacterium* dominant)	VMB Type 2 vs. 1 (end of follow-up sample)	Incident hrHPV vs. no hrHPV at both visits	RRR, 0.125 (P=0.019)[g]	NR	N/A
					Simpson index (end of follow-up sample)	Cleared hrHPV vs. no hrHPV at both visits	RRR, 3.856 (P= 0.034)[g]	NR	N/A
					Simpson index (end of follow-up sample)	Incident CIN2+ vs. </ CIN1 and hrHPV at both visits	RRR, 7.352 (P=0.028)[g]	NR	N/A
					VMB Type 4 vs. 1 (baseline sample)	Cleared CIN2+ </ CIN1 and hrHPV at both visits	RRR, 8.662 (P=0.021)[g]	NR	
					BV-anaerobes RA (baseline sample)	CIN2+ at one or both visits vs. no hrHPV at both visits	RRR, 2.561 (P=0.049)[g]	NR	

	outcomes (n = 236), for tabulated estimates with CIN outcomes Ages, median 34				Simpson index (end of follow-up sample)	CIN2+ at one or both visits vs. no hrHPV at both visits	RRR, 5.981 (P=0.003)[g]	NR	
					Lactobacilli RA (end of follow-up sample)	CIN2+ at one or both visits vs. no hrHPV at both visits	RRR, 0.352 (P=0.025)[g]	NR	
Siqueira (2019)	HIV/HPV co-infected pregnant women (Enrollment, 2009; Follow-up, until 2011; bacteriome analyses 6- and 12-months after delivery) (n = 12), a total of 24 samples Age, average 28	16S rRNA V3-V6 NR	PCR and Sanger sequencing NR	Hierarchical clustering into 2 CST types (CST III; *Lactobacillus iners*, CST IV; anaerobic bacteria)	CST IV vs. CST III	HPV16	RR, 0.75 (0.44-1.26)	NR	N/A
Arokiyaraj (2018)	Normal and ASC-US patients enrolled in an HPV cohort study (Enrollment, 2006-2013; Follow-up, 6 month-intervals with a maximum 5 visits) (n = 41), a total of 107 samples Ages, 18-65	16S rRNA NR Pyrosequencing	Digene HC2 high-risk DNA test 13 HPV types (hrHPVs 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68)	Bacterial species RA	*Lactobacillus johnsonii* RA	hrHPV persistence vs. clearance	NR	aOR, 16.4 (1.77-152.2)	Age, menopausal status, oral contraceptive use and smoking habit
					Lactobacillus crispatus RA	hrHPV negative vs. positive	NR	aOR, 8.25 (2.13-32.0)	Age, menopausal status, oral contraceptive use and smoking habit
					Eubacterium eligen RA	hrHPV clearance vs. negative & persistence	NR	aOR, 11.5 (1.31-101.4)	Age, menopausal status, oral contraceptive

					Gardnerella vaginalis RA	hrHPV clearance vs. negative	NR	aOR, 17.0 (2.18-131.8)	use and smoking habit
					Ureaplasma urealyticum	hrHPV clearance vs. negative	NR	aOR, 7.42 (1.30-42.5)	
Zhang (2018)	Biopsy-confirmed cervical samples from women in Beijing, China Ages, NR (n = 166)	16S rRNA V3-V4 Illumina HiSeq	PCR Amplification of L1 open reading frame with probes for 17 HPV types 17 HPV types (HPVs 6, 11, 16, 18, 31, 33, 35, 39, 45, 52, 56, 58, 59, 66, 68, 69, 82)	Visualization via four clusters at the species level (Cluster I, *Lactobacillus iners*; Cluster II, *Lactobacillus crispatus*; Cluster III; various dominant microbes and higher RA of *Atopobium vaginae, Escherichia coli, Streptococcus agalactiae* and other microbes; Cluster IV, *Lactobacillus iners* and *Lactobacillus crispatus*)	*Pseudomonas stutzeri* middle vs. low[h]	CIN2+ vs. CIN1-	OR, 0.39 (0.17-0.90)	NR	N/A
					Pseudomonas stutzeri high vs. low[h]	CIN2+ vs. CIN1-	OR, 0.36 (0.15-0.85)	NR	
					Bacteroides fragilis high vs. low[h]	CIN2+ vs. CIN1-	OR, 1.11 (0.57-2.17)	NR	
					Lactobacillus delbrueckii middle vs. low[h]	CIN2+ vs. CIN1-	OR, 1.72 (0.80-3.69)	NR	
					Lactobacillus delbrueckii high vs. low[h]	CIN2+ vs. CIN1-	OR, 1.35 (0.37-4.99)	NR	
					Atopbium vaginae middle vs. low[h]	CIN2+ vs. CIN1-	OR, 0.50 (0.20-1.21)	NR	
					Atopbium vaginae high vs. low[h]	CIN2+ vs. CIN1-	OR, 0.28 (0.08-0.95)	NR	
					Streptococcus agalactiae middle vs. low[h]	CIN2+ vs. CIN1-	OR, 2.24 (0.68-7.43)	NR	
					Streptococcus agalactiae high vs. low[h]	CIN2+ vs. CIN1-	OR, 2.99 (0.84-10.61)	NR	
Di Paola (2017)	Reproductive and post-menopausal women	16S rRNA V3-V5	Hybrid Capture 2 assay	CST types based on RA	CST IV-BV vs. NR	hrHPV persistence vs. clearance	OR, 9.38 (1.85-47.52)	NR	N/A

	unvaccinated against HPV in a study to assess the efficacy of HPV DNA testing for cervical cancer screening (Enrollment, NR; Follow up, 1 year) (n = 72) Ages, 26-64	Pyrosequencing	13 HPV types (hrHPVs 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68)	(CST I, *Lactobacillus crispatus*; CST II, *Lactobacillus gasseri*; CST III, *Lactobacillus iners*; CST IV-BV, *Lactobacillus* depletion and dominated by aerobic/anaerobic bacteria; CST IV-AV; *Lactobacillus* depletion aerobic/anaerobic bacteria)	CST IV-AV vs. NR	hrHPV persistence vs. clearance	OR, 0.11 (0.01-0.93)	NR	
					CST III vs. NR	hrHPV persistence vs. clearance	OR, 2.08 (0.59-7.30)	NR	
					CST II vs. NR	hrHPV persistence vs. clearance	OR, 0.21 (0.02-2.04)	NR	
					CST I vs. NR	hrHPV persistence vs. clearance	OR, 0.43 (0.12-1.53)	NR	
Audirac-Chalifour (2016)	HPV positive women with non-cervical lesions, SIL, and/or cervical cancer (n = 32) Ages, 22-61	16S rRNA V3-V4 Genome Sequencer Titanium Roche-454	NR	CST based on predominant taxa (CST I, *Lactobacillus crispatus*; CST II, *Lactobacillus iners*; CST III, *Pseudomonas oleovorans*; CST IV *Sneathia* spp.; CST V, *Gardnerella vaginalis*; CST VI, *Streptococcus agalactiae*; CST VII, *Fusobacterium necrophorum*; CST VII, *Fusobacterium* spp.)	Shannon diversity index	SIL & cervical cancer vs. non-cervical lesions	NR	aOR, 3.35 (0.64-17.65)	Age, contraceptive method, HPV-genotype
					PD whole tree	SIL & cervical cancer vs. non-cervical lesions	NR	aOR, 3.30 (0.76-14.49)	
Dareng (2016)[i]	Women who underwent gynecological examination in a	16S rRNA V4	Roche Linear Array 37 HPV types (hrHPVs 16, 18, 31,	Hierarchical clustering into CST types	CST I vs. CST IV-B	hrHPV+ vs. hrHPV-	OR, 0.60 (0.10-2.00)	aOR, 0.40 (0.10-1.70)	Age, age of sexual debut, number of sex partners in the

	cervical cancer screening program (n = 278) Ages, 18+	Illumina MiSeq	33, 35, 39, 45, 51, 52, 56, 58, 59, 68; lrHPVs 6, 11, 26, 40, 42, 53, 54, 55, 61, 62, 64, 66, 67, 69, 70, 71, 72, 73, 81, 82, 83, 84, IS39, CP6108)	(CST I, *Lactobacillus crispatus;* CST III, *Lactobacillus iners;* CST IV-B, low *Lactobacillus*; CST VI, *Proteobacteria*)	CST III vs. CST IV-B	hrHPV+ vs. hrHPV-	OR, 0.80 (0.40-1.40)	aOR, 0.70 (0.30-1.30)	last year, visual inspection with acetic acid results, HIV, male condom use
					CST VI vs. CST IV-B	hrHPV+ vs. hrHPV-	OR, 0.20 (0.0-1.70)	NR	
Piyathilake (2016)	Patients with abnormal cytology, hrHPV positivity and diagnosis of CIN (n = 430) Ages, 19-50	16S rDNA V4 Illumina MiSeq	Roche Diagnostics Linear Arrary 13 HPV types (hrHPVs 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68)	Dirichlet multinomial mixture (DMM) model to partition into cervical mucosa CT types (Partition 1, diverse taxa including unclassified *Lactobacillus*, *Lactobacillus iners*, *Bifidobacteriaceae, Clostridiales* and *Allobaculum;* Partition 2, *Lactobacillus* dominant, high diversity and rare taxa; Patition 3, Unclassified *Lactobacillus* and *Lactobacillus iners* dominant; Parition 4, Lack of *Lactobacillus* dominance, composed of *Bifidobacteriaceae, Prevotella, Sneathia* and *Megasphaera*)	Partition 2 vs. Partition 1	CIN2+ vs. CIN1	OR, 1.63 (0.83-3.21)	aOR, 1.84 (0.86-3.93)	Age, body mass index, race, education, parity, hormonal contraceptive use and smoking status
					Partition 3 vs. Partition 1	CIN2+ vs. CIN1	OR, 2.48 (1.01-6.07)	aOR, 3.48 (1.27-9.55)	
					Partition 4 vs. Partition 1	CIN2+ vs. CIN1	OR, 1.17 (0.58-2.36)	aOR, 1.13 (0.51-2.52)	
					Partition 2 vs. Partition 1	CIN3+ vs. CIN1	OR, 1.44 (0.64-3.23)	aOR, 1.68 (0.66-4.28)	
					Partition 3 vs. Partition 1	CIN3+ vs. CIN1	OR, 2.55 (0.92-7.10)	aOR, 3.24 (0.98-10.68)	
					Partition 4 vs. Partition 1	CIN3+ vs. CIN1	OR, 1.24 (0.54-2.85)	aOR, 1.20 (0.45-3.20)	
Reimers (2016)[j]	HIV-positive and -negative African American, premenopausal	16S rRNA V1-V2	PCR Assay, probed for presence of HPV DNA	Hierarchical clustering into CST types	Transition to *Lactobacillus crispatus* dominant	Incident HPV detection	TRR, 0.17 (0.04-0.79)[k]	aTRR, 0.20 (0.03-1.14)	Additional covariates NR

	women (Enrollment, 1994-1995 and 2001-2002; Follow up, 8-10 years, semi-annual cervicovaginal lavage sample) (n = 64), a total of 398 samples from 22 HIV- women, 22 HIV+ women with stable $CD4^+$ T-cell count, 20 HIV + women with progressive immunosuppression Age, average 32.1	Pyrosequencing	42 HPV types (HPVs 6, 11, 13, 16, 18, 26, 31, 32, 33, 34, 35, 39, 40, 42, 45, 51, 52, 53, 54, 55, 56, 57, 58, 59, 61, 62, 64, 66, 67, 68, 69, 70, 71, 72, 73, 74, 81, 82, 83, 84, 85, 89)	(CST I, *Lactobacillus crispatus*; CST II, *Lactobacillus gasseri*; CST III, *Lactobacillus iners*; CST IV-A medium RA of *Lactobacillus iners*; CST V *Lactobacillus jensenii*; CST IV-B variety of genera and low RA of *Lactobacillus*)	CST				
					CST I vs. CST IV-B	Detection of any HPV	NR	aOR, 0.58 (0.29-1.16)	Continuous pH, age, HIV study group, highly active antiretroviral therapy use, number of recent sex partners, smoking status, condom use, lifetime sex partners, *Trichomonas vaginalis* or *Candida* or other STI infection in the past 6 months
					CST II vs. CST IV-B	Detection of any HPV	NR	aOR, 0.84 (0.44-1.61)	
					CST III vs. CST IV-B	Detection of any HPV	NR	aOR, 1.10 (0.72-1.69)	
					CST IV-A vs. CST IV-B	Detection of any HPV	NR	aOR, 1.34 (0.82-2.19)	
					CST V vs. CST IV-B	Detection of any HPV	NR	aOR, 1.12 (0.61-2.08)	
					CST I vs. CST IV-B	Detection of oncogenic HPV	NR	aOR, 0.34 (0.09-1.33)	
					CST II vs. CST IV-B	Detection of Oncogenic HPV	NR	aOR, 0.61 (0.06-6.00)	
					CST III vs. CST IV-B	Detection of oncogenic HPV	NR	aOR, 1.42 (0.66-3.04)	
					CST IV-A vs. CST IV-B	Detection of oncogenic HPV	NR	aOR, 1.84 (0.78-4.34)	
					CST V vs. CST IV-B	Detection of oncogenic HPV	NR	aOR, 1.09 (0.37-3.18)	
					Lactobacillus crispatus RA medium vs. low	Detection of any HPV	NR	aOR, 0.24 (0.07-0.81)	
					Lactobacillus crispatus RA high vs. low	Detection of any HPV	NR	aOR, 0.47 (0.24-0.90)	
					Lactobacillus crispatus RA medium vs.	Detection of oncogenic HPV	NR	aOR, 0.52 (0.24-1.11)	

					low				
					Lactobacillus crispatus RA high vs. low	Detection of oncogenic HPV	NR	aOR, 0.14 (0.01-1.65)	
Oh (2015)[l]	Korean women from gynecological oncology clinics (n = 120) Ages, 18 - 65	16S rRNA V1-V3 Pyrosequencing	Chemiluminescent HPV DNA Test 13 hrHPV types	Risky Microbial Pattern (Defined by predominance of *Atopobium vaginae, Lactobacillus iners Gardnerella vaginalis* and depletion of *Lactobacillus crispatus*)	Risky microbial pattern (T3)	CIN vs. Normal	NR	aOR, 5.80 (1.73-19.4)	Age, marital status, menopausal status, oral contraceptive use and smoking
					Atopium vaginae RA (T3)	CIN vs. Normal	NR	aOR, 6.63 (1.61-27.2)	
					Gardnerella vaginalis RA (T3)	CIN vs. Normal	NR	aOR, 3.21 (0.95-10.80)	
					Lactobacillus iners RA (T3)	CIN vs. Normal	NR	aOR, 2.49 (0.88-7.08)	
					Lactobacillus crispatus RA (T3)	CIN vs. Normal	NR	aOR, 1.88 (0.65-5.38)	
					Risky microbial pattern (high tertile) and hrHPV- vs. risky microbial pattern (low-medium tertile) and hrHPV-	CIN vs. Normal	NR	aOR, 10.8 (1.71-68.8)	Age, marital status, menopausal status, oral contraceptive use and smoking
					Atopobium vaginae (high tertile) and hrHPV- vs. *Atopobium vaginae* (low-	CIN vs. Normal	NR	aOR, 7.76 (1.26-48.0)	

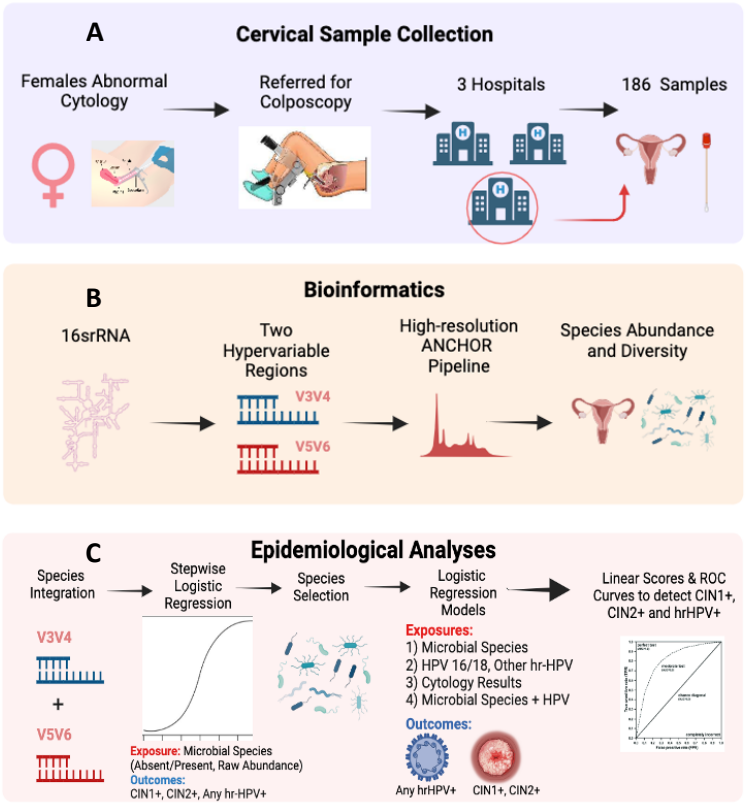
					middle tertile) and hrHPV-				
					Gardnerella vaginalis (high tertile) and hrHPV- vs. *Gardnerella vaginalis (*low-middle tertile) and hrHPV-	CIN vs. Normal	NR	aOR, 3.14 (0.55-18.1)	
					Lactobacillus iners (high tertile) and hrHPV- vs. *Lactobacillus iners (*low-middle tertile) and hrHPV-	CIN vs. Normal	NR	aOR, 1.55 (0.24-10.1)	
Brotman (2014)	Women from a vaginal douching cessation study in Baltimore, Maryland (Enrollment, 2005 – 2007; Follow-up, 16 weeks, sample collection twice weekly) (n = 32), a total of 937 samples with microbiome data and 930 samples with HPV results	16SrRNA V1-V2 Pyrosequencing	Roche Linear Array 37 HPV types (hrHPVs 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68)	CST (CST I, *Lactobacillus crispatus*; CST II, *Lactobacillus gasseri*; CST III, *Lactobacillus iners;* CST V, *Lactobacillus jensenii,* CST IV-A; anaerobic bacteria; CST IV-B, Increase in *Atopobium, Gardneralla, Prevotella* and others)	CST II vs. CST I	Transition from HPV- to HPV+	TRR, 1.13 (0.31-4.15)	aTRR,0.34 (0.06-1.85)	Age, ethnicity/race, hormonal contraceptives, study phase, lubricant use, vaginal sex (with/without condoms), menses, normalized menstrual cycle time
					CST III vs. CST I	Transition from HPV- to HPV+	TRR, 1.94 (0.85-4.46)	aTRR, 1.79 (0.71-4.51)	
					CST IV-A vs. CST I	Transition from HPV- to HPV+	TRR, 4.00 (1.46-11.01)	aTRR, 1.86 (0.52-6.74)	
					CST IV-B vs. CST I	Transition from HPV- to HPV+	TRR, 1.06 (0.37-3.04)	aTRR, 0.76 (0.26-2.24)	
					CST II vs.	Transition from HPV+ to HPV-	TRR, 5.19 (1.45-	aTRR, 4.43 (1.11-17.7)	Age, ethnicity/race,

	Ages, 22 - 53				CST I		18.50)		hormonal contraceptive, study phase, lubricant use, vaginal sex (with/without condoms), menses, normalized menstrual cycle time
					CST III vs. CST I	Transition from HPV+ to HPV-	TRR, 0.55 (0.25-1.19)	aTRR, 0.79 (0.28-2.2)	
					CST IV-A vs. CST I	Transition from HPV+ to HPV-	TRR, 0.98 (0.33-2.86)	aTRR, 2.38 (0.69-8.17)	
					CST IV-B vs. CST I	Transition from HPV+ to HPV-	TRR, 0.30 (0.12-0.77)	aTRR, 0.33 (0.12-1.19)	

Abbreviations: aHR, adjusted hazard ratio; aOR, adjusted odds ratio; aTRR, adjusted transition rate ratio; ASC-H, Atypical squamous cells - cannot exclude high grade squamous intraepithelial lesion; ASC-US, atypical squamous cells of undetermined significance; BV, bacterial vaginosis; CI, confidence interval; CIN, cervical intraepithelial neoplasia; CST, community state type; CT; community type; HPV, human papillomavirus; HIV, human immunodeficiency virus; hrHPV, high-risk human papillomavirus; HSIL, high-grade squamous intraepithelial lesions; ICC, invasive cervical cancer; IUD, intrauterine device; LSIL; low-grade squamous intraepithelial lesion; lrHPV, low-risk human papillomavirs; PCR, polymerase chain reaction; *molBV*, molecular bacterial vaginosis NILM, negative for intraepithelial lesion or malignancy; NA, not applicable; NR, not reported; OR, odds ratio; RA, relative abundance; RR, relative risk; RRR, relative risk ratio; sp., singular species; spp., multiple species; SIL, squamous intraepithelial lesion; TRR, transition rate ratio; 16S rRNA, 16S ribosomal RNA

[a] Only statistically significant effect estimates were tabulated for associations between the CVM and HPV prevalence.
[b] Only statistically significant unadjusted effect estimates were tabulated for associations between the CVM and cytological interpretations.
[c] HPV detection method was extracted from the original cohort study.
[d] High and low refer to above and below the median, respectively.
[e] Sustained refers to agreement of the measurement as either above or below the median at visit 1 and visit 2.
[f] Only statistically significant effect estimates were tabulated.
[g] 95% CI not reported.
[h] RA was categorized into three groups (low, middle and high) according the 25th and 75th percentile.
[i] Unadjusted effect estimates correspond to an exact logistic regression model. Findings from the univariate regression were not tabulated.
[j] Authors constructed multiple models with pH as continuous, binary and ordinal, only effect estimated based on continuous pH were tabulated and analyzed.
[k] Unadjusted estimate was adjusted for HIV study group.
[l] Authors constructed multiple models by dividing RA of microbial components into tertials, effect estimates were tabulated and analyzed based on the third tertial (T3).
[m] All unadjusted estimates was adjusted for all adjusted variables except for menstrual cycle time.

A.2. Manuscript 2 Supplemental Methodology



S-Figure 4-0. Overall methodology for the empirical research manuscript. *Figure designed using Biorender: Scientific Image and Illustration Software.*

S-Figure 4-0 Legend.

The study population for this analysis consisted of 186 women enrolled in the MARKER study at a McGill University affiliated hospital in Montreal, Quebec, Canada.[77] These 186 samples were selected from one of three hospitals participating in the Cervical and Self-Sampling in Screening Study designed to compare the effectiveness of the HerSwab™ self-sampling device with physician-collected samples for the detection of CIN and cervical cancer **(Panel A)**.[92] Using 16S rRNA gene sequencing, with the high-resolution ANCHOR pipeline and amplifying at two hypervariable regions (V3-V4 and V5-V6) we identified bacterial species constituting the CVM of these 186 samples **(Panel B).** We integrated the species identified in the V3-V4 and V5-V6 primer sets and performed stepwise logistic regression to select bacterial species and construct linear scores based on the regression coefficients to assess correlations with CIN1+, CIN2+ and hrHPV infections. Additionally, the regression coefficients from three separate logistic regression models were used to build

linear scores for: (1) cytology, (2) hrHPV, and (3) combined positivity for selected bacterial species and hrHPV to assess correlations with the aforementioned outcomes. For all linear scores, we plotted ROC curves and compared the performance of the scores for the detection of CIN and hrHPV **(Panel C).**

Abbreviations: CIN, cervical intraepithelial neoplasia; CVM, cervicovaginal microbiome; HPV, human papillomavirus; hrHPV, high-risk human papillomavirus; MARKER, Methylation Analysis Revealing Key Epigenic Regulation; ROC, receiver operating characteristic; 16S rRNA, 16S ribosomal RNA

A.3. Manuscript 2 Supplemental Material

S-Table 4-1. Distribution of bacterial species by histology and descriptive statistics of their raw abundance based on V3-V4 primer set.

	Histological endpoints, n (%)					Raw abundance				
Bacterial Species	**Normal (n = 54)**	**CIN1 (n = 50)**	**CIN2 (n = 40)**	**CIN3 (n = 42)**	**P-value[a]**	**Range[b]**	**Mean[c]**	**Geometric Mean**	**Median[d]**	**Q1 - Q3[c]**
Alloscardovia omnicolens	4 (7.4)	3 (6.0)	1 (2.5)	4 (9.5)	0.627	1-19169	109.8	20.7	12.5	1-352
Anaerococcus vaginalis	6 (11.1)	4 (8.0)	4 (10)	5 (11.9)	0.950	1-1795	20.5	16.6	6.0	2 -107
Bifidobacterium breve	6 (11.1)	6 (12.0)	1 (2.5)	5 (11.9)	0.342	1-11270	253.2	34.4	11.0	1-6239
Ezakiella coagulans	3 (5.6)	2 (4.0)	3 (7.5)	3 (7.1)	0.883	1-2546	15.9	6.6	2.0	1-16
Fannyhessea vaginae	24 (44.4)	28 (56.0)	18 (45.0)	18 (42.9)	0.560	1-8397	614.5	82.6	158.5	1-1797
Finegoldia magna	16 (29.6)	9 (18.0)	17 (42.5)	17 (40.5)	0.040	1-1214	22.9	5.6	3.0	2-10
Gardnerella leopoldii	24 (44.4)	25 (50.0)	19 (47.5)	15 (35.7)	0.559	1-8477	334.5	51.8	54.0	4-670
Gardnerella swidsinskii	26 (48.2)	20 (40.0)	21 (52.5)	15 (35.7)	0.390	1-17190	670.4	55.1	43.5	2-1061
Gardnerella vaginalis	37 (68.5)	35 (70.0)	32 (80.0)	27 (64.3)	0.447	1-24607	1771.5	103.7	190.0	4-1754
Lactobacillus acidophilus	24 (44.4)	25 (50.0)	18 (45.0)	18 (42.9)	0.909	1-8014	94.4	35.6	30.0	13-141
Lactobacillus crispatus	35 (64.8)	31 (62.0)	22 (55.0)	23 (54.8)	0.683	1-23141	813.5	75.0	134.0	6-213
Lactobacillus fornicalis	24 (44.4)	33 (66.0)	23 (57.5)	18 (42.9)	0.070	1-6946	247.4	80.3	138.0	23-383
Lactobacillus gasseri	16 (29.6)	23 (46.0)	11 (27.5)	12 (28.6)	0.191	1-25580	420	18.5	8.5	1-71
Lactobacillus iners	49 (90.7)	47 (94.0)	39 (97.5)	39 (92.9)	0.638	1-32976	6227.1	369.1	915.0	12-13613
Lactobacillus jensenii	24 (44.4)	32 (64.0)	26 (65.0)	21 (50.0)	0.110	1-15584	281.1	45.2	62.0	5-243
Lactobacillus kalixensis	2 (3.7)	3 (6.0)	2 (5.0)	4 (9.5)	0.699	1-14688	80.5	6.1	1.0	1-42
Mageeibacillus indolicus	4 (7.4)	4 (8.0)	5 (12.5)	2 (4.8)	0.659	1-1104	15.5	42.5	69.0	8-193
Mobiluncus mulieris	11 (20.4)	5 (10.0)	7 (17.5)	6 (14.3)	0.513	1-1678	41	22.1	12.0	2-247
Porphyromonas asaccharolytica	7 (13.0)	6 (12.0)	8 (20.0)	4 (9.5)	0.568	1-5350	57.4	26.4	50.0	1-183
Prevotella amnii	12 (22.2)	8 (16.0)	10 (25.0)	7 (16.7)	0.675	1-8377	174.3	65.0	154.0	1-913
Prevotella bivia	11 (20.4)	8 (16.0)	9 (22.5)	8 (19.1)	0.884	1-3026	67.7	23.3	22.5	2.5-113
Prevotella colorans	6 (11.1)	3 (6.0)	2 (5.0)	5 (11.9)	0.579	1-3696	23.7	4.7	2.0	1.5-4.5
Prevotella melaninogenica	3 (5.6)	1 (2.0)	2 (5.0)	1 (2.4)	0.742	1-2597	18.1	32.0	16.0	1-609
Prevotella timonensis	20 (37.0)	15 (30.0)	15 (37.5)	10 (23.8)	0.465	1-7127	217	43.4	24.0	4-427
Sneathia amnii	10 (18.5)	8 (16.0)	8 (20.0)	8 (19.1)	0.963	1-5577	51.8	16.6	15.5	2-103
Sneathia sanguinegens	19 (35.2)	11 (22.0)	12 (30.0)	10 (23.8)	0.447	1-10043	140.7	31.2	48.0	1.5-497.5
Streptococcus agalactiae	5 (9.3)	3 (6.0)	3 (7.5)	9 (21.4)	0.117	1-18950	145.1	34.0	19.5	3-310
Streptococcus anginosus	7 (13.0)	5 (10.0)	9 (22.5)	8 (19.1)	0.348	1-4252	25.8	5.6	3.0	1-13
Streptococcus urinalis	0 (0.0)	1 (2.0)	3 (7.5)	1 (2.4)	0.146	1-22124	119.0	7.4	1.0	1-1

Abbreviations: CIN, cervical intraepithelial neoplasia; Q1, first quartile; Q3, third quartile

[a]Fisher's exact test used to assess heterogeneity, $P < 0.05$ was considered significant.
[b]The minimum value for each bacterial species was 0 (absent), minimum listed is that other than 0.
[c]Mean may be unreliable due to skewness in species' abundance.
[d]Calculated based on samples with a raw abundance greater than zero.

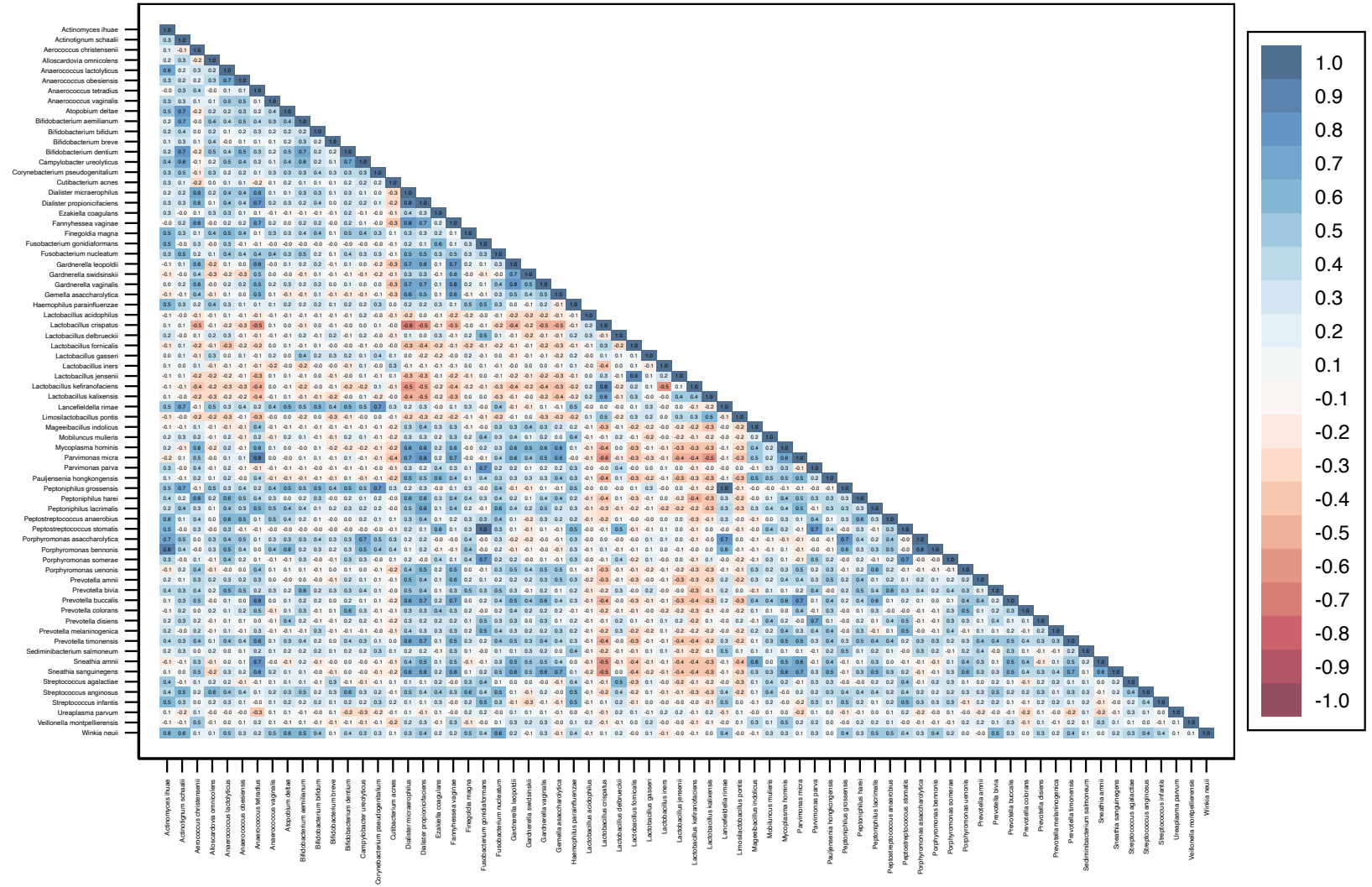
S-Table 4-2. Distribution of bacterial species by histology and descriptive statistics of their raw abundance based on V5-V6 primer set.

	Histological endpoints, n (%)					Raw abundance				
Bacterial Species	**Normal (n = 54)**	**CIN1 (n = 50)**	**CIN2 (n = 40)**	**CIN3 (n = 42)**	**P-value**	**Range**	**Mean[b]**	**Geometric Mean**	**Median[c]**	**Q1-Q3[c]**
Actinomyces ihuae	4 (7.4)	1 (2.0)	2 (5.0)	2 (4.8)	0.647	0-267	1.9	6.2	5.0	1-18
Actinotignum schaalii	2 (3.7)	2 (4.0)	0 (0.0)	1 (2.4)	0.765	0-2268	12.9	31.9	12.0	11-111
Aerococcus christensenii	15 (27.8)	23 (46.0)	16 (40.0)	12 (28.6)	0.173	0-1679	57.6	35.0	33.5	10-203
Alloscardovia omnicolens	3 (5.6)	2 (4.0)	0 (0.0)	5 (11.9)	0.110	0-14585	82.6	25.0	17.0	1-162
Anaerococcus lactolyticus	10 (18.5)	6 (12.0)	9 (22.5)	6 (14.3)	0.570	0-167	2.8	5.7	6.0	1-22
Anaerococcus obesiensis	7 (13.0)	5 (10.0)	7 (17.5)	6 (14.3)	0.760	0-976	8.0	6.0	2.0	1-34
Anaerococcus rubeinfantis	0 (0.0)	0 (0.0)	0 (0.0)	2 (4.8)	0.095	0-1377	7.4	37.1	689.0	1-1377
Anaerococcus tetradius	14 (25.9)	9 (18.0)	11 (27.5)	11 (26.2)	0.670	0-3056	25.5	13.8	14.0	5-44
Arcanobacterium urinimassiliense	0 (0.0)	2 (4.0)	1 (2.5)	0 (0.0)	0.334	0-500	2.7	10.0	2.0	1-500
Atopobium deltae	4 (7.4)	3 (6.0)	6 (15.0)	0 (0.0)	0.054	0-344	2.1	3.3	2.0	1-6
Bifidobacterium aemilianum	5 (9.3)	4 (8.0)	1 (2.5)	2 (4.8)	0.563	0-599	6.1	4.1	1.0	1-6.5
Bifidobacterium bifidum	3 (5.6)	2 (4.0)	3 (7.5)	6 (14.3)	0.296	0-532	6.8	15.8	18.0	6-50
Bifidobacterium breve	3 (5.6)	2 (4.0)	0 (0.0)	4 (9.5)	0.255	0-4968	28.5	18.8	7.0	2-79
Bifidobacterium dentium	4 (7.4)	2 (4.0)	0 (0.0)	3 (7.1)	0.294	0-2081	11.8	15.0	11.0	6-26
Campylobacter ureolyticus	5 (9.3)	3 (6.0)	4 (10.0)	4 (9.5)	0.897	0-585	5.5	6.8	3.0	1-56.5
Corynebacterium pseudogenitalium	2 (3.7)	5 (10.0)	3 (7.5)	6 (14.3)	0.295	0-512	3.3	3.8	2.0	1-7
Cutibacterium acnes	15 (27.8)	7 (14.0)	12 (30.0)	14 (33.3)	0.129	0-1728	18.3	4.0	2.0	1-6
Dialister micraerophilus	24 (44.4)	21 (42.0)	19 (47.5)	18 (42.9)	0.964	0-600	31.8	23.0	49.0	4-100
Dialister pneumosintes	0 (0.0)	0 (0.0)	0 (0.0)	1 (2.4)	0.441	0-789	4.2	789.0	789.0	789-789
Dialister propionicifaciens	22 (40.7)	13 (26.0)	19 (47.5)	16 (38.1)	0.187	0-2893	62.5	33.9	60.0	5-195
Ezakiella coagulans	2 (3.7)	5 (10.0)	2 (5.0)	1 (2.4)	0.418	0-1062	5.8	3.7	2.0	1-4
Fannyhessea vaginae	27 (50.0)	27 (54.0)	22 (55.0)	19 (45.2)	0.798	0-8508	484.1	73.2	111.0	3-1611
Finegoldia magna	24 (44.4)	18 (36.0)	17 (42.5)	18 (42.9)	0.841	0-510	12.5	4.6	3.0	1-9
Fusobacterium gonidiaformans	1 (1.9)	3 (6.0)	3 (7.5)	1 (2.4)	0.500	0-894	6.4	19.4	16.5	3-127
Fusobacterium nucleatum	7 (13.0)	6 (12.0)	3 (7.5)	8 (19.1)	0.497	0-7285	40.6	6.3	5.5	1-20.5
Gardnerella vaginalis	41 (75.9)	35 (70.0)	28 (70.0)	29 (69.1)	0.864	0-18168	992.4	83.9	148.0	4-1380
Gemella asaccharolytica	8 (14.8)	9 (18.0)	13 (32.5)	4 (9.5)	0.059	0-544	10.9	18.7	28.5.0	5-52
Haemophilus parahaemolyticus	0 (0.0)	3 (6.0)	3 (7.5)	2 (4.8)	0.204	0-371	3.4	12.4	6.0	2-126
Haemophilus parainfluenzae	4 (7.4)	4 (8.0)	5 (12.5)	3 (7.1)	0.839	0-297	3.3	6.5	3.5	1.5-28
Lactobacillus acidophilus	3 (5.6)	4 (8.0)	2 (5.0)	3 (7.1)	0.934	0-14281	76.8	2.4	1.0	1-1
Lactobacillus crispatus	54 (100.0)	50 (100.0)	40 (100.0)	42 (100.0)	N/A	1-33211	8129	319.1	104.5	10-19734
Lactobacillus delbrueckii	3 (5.6)	2 (4.0)	1 (2.5)	1 (2.4)	0.914	0-8199	45.1	12.0	3.0	1-187
Lactobacillus iners	54 (100.0)	50 (100.0)	40 (100.0)	42 (100.0)	N/A	2-37804	6702.5	551.9	890.5	26-12858

Lactobacillus jensenii	34 (63.0)	39 (78.0)	28 (70.0)	28 (66.7)	0.394	0-22181	709.5	67.4	173.0	2-817
Lactobacillus kalixensis	37 (68.5)	34 (68.0)	24 (60.0)	24 (57.1)	0.589	0-16356	90.8	2.6	2.0	2-3
Lactobacillus kefiranofaciens	20 (37.0)	26 (52.0)	17 (42.5)	14 (33.3)	0.282	0-420	5.3	5.7	7.0	4-9
Lancefieldella rimae	1 (1.9)	0 (0.0)	0 (0.0)	1 (2.4)	0.843	0-553	3.0	33.3	277.5	2-553
Limosilactobacillus pontis	22 (40.7)	28 (56.0)	18 (45.0)	18 (42.9)	0.438	0-516	38.4	39.6	66.5	15-118
Mageeibacillus indolicus	6 (11.1)	4 (8.0)	4 (10.0)	5 (11.9)	0.950	0-785	11.3	32	56.0	16-103
Mobiluncus mulieris	5 (9.3)	3 (6.0)	1 (2.5)	4 (9.5)	0.552	0-2488	36.2	31.3	9.0	3-761
Mycoplasma hominis	13 (24.1)	6 (12.0)	9 (22.5)	2 (4.76)	0.032	0-1667	19.1	18.3	23.0	4-82
Parvimonas micra	14 (26.0)	8 (16.0)	8 (20.0)	12 (28.6)	0.462	0-1011	45.4	48.1	116.0	4-275
Parvimonas parva	2 (3.7)	3 (6.0)	4 (10.0)	2 (4.8)	0.643	0-148	2.5	10.1	5.0	1-100
Pauljensenia hongkongensis	7 (13.0)	3 (6.0)	4 (10.0)	4 (9.5)	0.717	0-47	2.0	13.7	20.5	9-29
Peptoniphilus grossensis	1 (1.9)	2 (4.0)	1 (2.5)	2 (4.8)	0.854	0-268	1.7	9.8	8.0	2-30
Peptoniphilus harei	13 (24.1)	15 (30.0)	13 (32.5)	12 (28.6)	0.824	0-327	8.0	6.5	4.0	2-12
Peptoniphilus lacrimalis	9 (16.7)	11 (22.0)	8 (20.0)	8 (19.1)	0.928	0-1128	18.6	16.2	11.5	3.5-84
Peptostreptococcus anaerobius	11 (20.4)	11 (22.0)	8 (20.0)	4 (9.5)	0.390	0-118	4.9	10.9	11.0	4-41
Peptostreptococcus stomatis	1 (1.9)	3 (6.0)	3 (7.5)	2 (4.8)	0.563	0-338	2.5	9.3	9.0	2-29
Porphyromonas asaccharolytica	2 (3.7)	7 (14.0)	3 (7.5)	3 (7.1)	0.305	0-5768	42.2	8.8	3.0	2-16
Porphyromonas bennonis	3 (5.6)	2 (4.0)	0 (0.0)	2 (4.8)	0.553	0-465	4.8	10.4	7.0	1-400
Porphyromonas somerae	2 (3.7)	3 (6.0)	0 (0.0)	0 (0.0)	0.229	0-613	3.3	3.6	1.0	1-1
Porphyromonas uenonis	7 (13.0)	6 (12.0)	7 (17.5)	6 (14.3)	0.884	0-1436	25.9	27.8	21.0	4-287
Prevotella amnii	19 (35.2)	15 (30.0)	11 (27.5)	10 (23.8)	0.675	0-8897	247.1	25.5	6.0	1-1043
Prevotella bivia	12 (22.2)	22 (44.0)	14 (35.0)	14 (33.3)	0.130	0-4995	90.0	12.3	6.0	1-83
Prevotella buccalis	14 (25.9)	9 (18.0)	10 (25.0)	8 (19.1)	0.723	0-2877	64.8	44.4	69.0	7-213
Prevotella disiens	4 (7.4)	11 (22.0)	8 (20.0)	5 (11.9)	0.134	0-1226	14.5	5.8	2.5	1-23.5
Prevotella ihumii	0 (0.0)	2 (4.0)	1 (2.5)	0 (0.0)	0.334	0-459	2.9	31.9	71.0	1-459
Prevotella timonensis	22 (40.7)	20 (40.0)	20 (50.0)	14 (33.3)	0.503	0-3943	198.7	26.5	13.0	2-389.5
Sediminibacterium salmoneum	5 (9.3)	2 (4.0)	2 (5.0)	8 (19.1)	0.084	0-133	2.3	5.5	2.0	2-13
Sneathia sanguinegens	21 (38.9)	17 (34.0)	13 (32.5)	8 (19.1)	0.199	0-8087	224	45.4	68.0	1-1209
Streptococcus agalactiae	4 (7.4)	6 (12.0)	7 (17.5)	11 (26.2)	0.073	0-17963	141.2	14.6	4.0	1.5-165.5
Streptococcus anginosus	5 (9.3)	9 (18.0)	5 (12.5)	4 (9.5)	0.537	0-283	3.1	6.4	4.0	2-21
Streptococcus gwangjuense	0 (0.0)	2 (4.0)	1 (2.5)	1 (2.4)	0.531	0-1018	5.7	14.7	23.5	1-532
Streptococcus infantis	4 (7.4)	4 (8.0)	4 (10.0)	5 (11.9)	0.882	0-361	2.5	4.9	3.0	2-16
Streptococcus urinalis	1 (1.9)	1 (2.0)	5 (12.5)	3 (7.1)	0.093	0-10466	56.7	5.2	1.0	1-6
Ureaplasma parvum	22 (40.7)	21 (42.0)	11 (27.5)	16 (38.1)	0.499	0-74	6.3	9.7	13.0	5-22
Veillonella montpellierensis	6 (11.1)	6 (12.0)	3 (7.5)	3 (7.1)	0.836	0-485	7.6	15.0	15.0	2-78
Winkia neuii	6 (11.1)	5 (10.0)	5 (12.5)	8 (19.1)	0.613	0-484	4.7	5.5	2.5	1-18

Abbreviations: CIN, cervical intraepithelial neoplasia; Q1, first quartile; Q3, third quartile

[a]Fisher's exact test used to assess heterogeneity, $P < 0.05$ was considered significant.
[b]Mean may be unreliable due to skewness in species' abundance.



S-Figure 4-1. Correlation between bacterial species in normal samples.

S-Figure 4-1 Legend

The correlation matrix shows the Spearman correlation coefficients between the 70 bacterial species present (listed alphabetically, considering the abundance from the region with the highest maximum raw abundance for species shared between the V3-V4 and V5-V6 regions) in biopsy-confirmed normal samples (n = 54). A value of +1 (blue) suggests a perfect positive correlation between two bacterial species, a value of -1 (red) a perfect negative association, and 0 (white) suggests no correlation.

S-Table 4-3. Stepwise logistic regression coefficients of cytology-, HPV-, and microbiome- (species presence/absence) based scores used to construct linear scores, comparing CIN1+ to normal histology.

Test-based score	Variables included	Prevalence in overall study population, n (%)	Coefficient	Standard Error	P-value
Cytology-based (constant = -2.05)	ASC-US	21 (11.29)	3.84	0.82	< 0.001
	LSIL	54 (29.03)	3.41	0.63	< 0.001
	HSIL or worse[a]	58 (31.18)	4.96	0.80	< 0.001
	Missing	18 (9.68)	2.74	0.73	< 0.001
HPV-based (constant= -0.16)	HPV16 and/or HPV18	44 (23.66)	2.75	0.76	< 0.001
	Other high-risk HPV[b]	101 (54.30)	1.35	0.36	< 0.001
Microbiome-based[c] (constant = 0.43)	*Lactobacillus jenseniii*	142 (76.34)	0.88	0.43	0.041
	Prevotella disiens	28 (15.05)	1.33	0.70	0.059
	Mycoplasma hominis	30 (16.13)	-2.04	0.70	0.003
	Aerococcus christensenii	66 (35.48)	1.44	0.57	0.011
	Streptococcus anginosus	43 (23.12)	1.09	0.53	0.038
	Finegoldia magna	86 (46.24)	-0.90	0.40	0.026
	Sneathia sanguingens	71 (38.17)	-0.75	0.48	0.121
	Gemella asaccharolytica	34 (18.28)	1.72	0.79	0.029
	Mobiluncus mulieris	33 (17.74)	-1.43	0.69	0.037
	Peptoniphilus lacrimalis	36 (19.35)	1.24	0.65	0.055
	Peptostreptococcus anaerobius	34 (18.28)	-1.28	0.70	0.068
	Peptostreptococcus stomatis	9 (4.84)	2.04	1.32	0.122
Microbiome-HPV-based[d] (constant = -0.85)	*Lactobacillus jensenii*	142 (76.34)	0.95	0.52	0.065
	Prevotella disiens	28 (15.05)	1.12	0.79	0.156
	Mycoplasma hominis	30 (16.13)	-2.60	0.82	0.002
	Aerococcus christensenii	66 (35.48)	1.45	0.62	0.020
	Streptococcus anginosus	43 (23.12)	1.35	0.64	0.036
	Finegoldia magna	86 (46.24)	-0.66	0.46	0.152
	Sneathia sanguingens	71 (38.17)	-0.81	0.58	0.163
	Gemella asaccharolytica	34 (18.28)	2.38	0.93	0.011
	Mobiluncus mulieris	33 (17.74)	-2.84	0.91	0.002
	Peptoniphilus lacrimalis	36 (19.35)	2.01	0.78	0.009
	Peptostreptococcus anaerobius	34 (18.28)	-1.29	0.82	0.115
	Peptostreptococcus stomatis	9 (4.84)	2.64	1.55	0.089
	HPV16 and/or HPV18	44 (23.66)	3.68	0.93	< 0.001
	Other high-risk HPV[b]	101 (54.30)	1.43	0.46	0.002

Abbreviations: ASC-US, atypical squamous cells of undetermined significance; HSIL, high-grade squamous intraepithelial lesion; HPV, human papillomavirus; HPV, human papillomavirus; LSIL, low-grade squamous intraepithelial lesion

[a] Includes 16 ASC-H, atypical squamous cells-cannot rule out HSIL, 2 AGC, atypical glandular cells, and 1 cancer case.
[b] Includes a pooled positivity result for any of HPVs 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and/or 68.
[c] Calculated based on the presence of 12/71 bacterial species, retained using a stepwise (forward selection) logistic regression (dependent variable: presence/absence of microbial species).
[d] Calculated based on the presence of the 12 bacterial species that were retained (refer to footnote c) and two additional predictors (positivity for HPV16 and/or 18 and 12 other high-risk HPVs).

S-Table 4-4. Stepwise logistic regression coefficients of cytology-, HPV-, and microbiome- (species presence/absence) based scores used to construct linear scores, comparing CIN2+ to normal and CIN1 histology.

Test-based score	Variables included	Prevalence in overall study population, n (%)	Coefficient	Standard Error	P-value
Cytology-based (constant = -2.80)	ASC-US	21 (11.29)	2.71	0.85	0.001
	LSIL	54 (29.03)	1.65	0.79	0.037
	HSIL or worse[b]	58 (31.18)	4.64	0.82	< 0.001
	Missing	18 (9.68)	2.35	0.87	0.007
HPV-based (constant = -1.29)	HPV16 and/or HPV18	44 (23.66)	2.26	0.44	< 0.001
	Other high-risk HPV[a]	101 (54.30)	0.97	0.35	0.005
Microbiome-based[c] (constant = -0.51)	*Bifidobacterium aemilianum*	12 (6.45)	-1.99	0.90	0.027
	Bifidobacterium bifidum	14 (7.53)	1.35	0.72	0.061
	Sediminibacterium salmoneum	17 (9.14)	0.48	0.55	0.383
	Streptococcus agalactiae	33 (17.74)	0.87	0.44	0.046
	Streptococcus urinalis	13 (6.99)	1.42	0.69	0.040
Microbiome-HPV-based[d] (constant = -1.55)	*Bifidobacterium aemilianum*	12 (6.45)	-1.72	0.91	0.060
	Bifidobacterium bifidum	14 (7.53)	1.64	0.76	0.030
	Sediminibacterium salmoneum	17 (9.14)	0.86	0.60	0.151
	Streptococcus agalactiae	33 (17.74)	0.51	0.47	0.278
	Streptococcus urinalis	13 (6.99)	0.86	0.76	0.257
	HPV16 and/or HPV18	44 (23.66)	2.21	0.45	< 0.001
	Other high-risk HPV[b]	101 (54.30)	1.00	0.37	0.007

Abbreviations: ASC-US, atypical squamous cells of undetermined significance; HSIL, high-grade squamous intraepithelial lesion; HPV, human papillomavirus; HPV, human papillomavirus; LSIL, low-grade squamous intraepithelial lesion

[a] Includes 16 ASC-H, atypical squamous cells-cannot rule out HSIL, 2 AGC, atypical glandular cells, and 1 cancer case.
[b] Includes a pooled positivity result for any of HPVs 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and/or 68.
[c] Calculated based on the presence of 5/77 bacterial species, retained using a stepwise (forward selection) logistic regression (dependent variable: presence/absence of microbial species).
[d] Calculated based on the presence of the 5 bacterial species that were retained (refer to footnote c) and two additional predictors (positivity for HPV16 and/or 18 and 12 other high-risk HPVs).

S-Table 4-5. Stepwise logistic regression coefficients of cytology-, HPV-, and microbiome- (species presence/absence) based scores used to construct linear scores, comparing any high-risk HPV positive to negative.

Test-based score	Variables included	Prevalence in overall study population, n (%)	Coefficient	Standard Error	P-value
Cytology-based (constant = -1.22)	ASC-US	21 (11.29)	1.31	0.59	0.027
	LSIL	54 (29.03)	2.37	0.51	< 0.001
	HSIL or worse[a]	58 (31.18)	3.20	0.57	< 0.001
	Missing	18 (9.68)	2.17	0.66	0.001
Microbiome-based[b] (constant = 0.71)	*Streptococcus agalactiae*	33 (17.74)	1.83	0.67	0.006
	Finegoldia magna	86 (46.24)	-0.94	0.38	0.012
	Anaerococcus tetradius	45 (24.19)	1.52	0.61	0.012
	Porphyromonas bennonis	7 (3.76)	-2.17	1.18	0.065
	Prevotella colorans	16 (8.60)	-1.55	0.71	0.030
	Prevotella timonensis	82 (44.09)	1.00	0.46	0.028
	Prevotella buccalis	41 (22.04)	-1.23	0.67	0.067
	Streptococcus urinalis	13 (6.99)	2.24	1.19	0.059
	Lactobacillus gasseri	62 (33.33)	-0.55	0.38	0.145
	Peptoniphilus lacrimalis	36 (19.35)	1.69	0.84	0.045
	Porphyromonas asaccharolytica	32 (17.20)	-1.43	0.75	0.058

Abbreviations: ASC-US, atypical squamous cells of undetermined significance; HSIL, high-grade squamous intraepithelial lesion; HPV, human papillomavirus; HPV, human papillomavirus; LSIL, low-grade squamous intraepithelial lesion

[a] Includes 16 ASC-H, atypical squamous cells-cannot rule out HSIL, 2 AGC, atypical glandular cells, and 1 cancer case.
[b] Calculated based on the presence of 11/77 bacterial species, retained using a stepwise (forward selection) logistic regression (dependent variable: presence/absence of bacterial species).

S-Table 4-6. Stepwise logistic regression coefficients of cytology-, HPV-, and microbiome- (species presence/absence) based scores used to construct linear scores, comparing CIN2+ to normal and CIN1 histology among women who tested positive for high-risk HPV.

Test-based score	Variables included	Prevalence in High-risk HPV+ women, n (%)	Coefficient	Standard Error	P-value
Cytology-based (constant = -1.95)	ASC-US	11 (8.87)	2.51	1.24	0.043
	LSIL	41 (33.06)	1.06	1.12	0.344
	HSIL or worse[a]	51 (41.13)	4.17	1.17	< 0.001
	Missing	13 (10.48)	2.10	1.21	0.081
Microbiome-based [b] (constant = 0.92)	*Finegoldia magna*	52 (41.94)	0.63	0.42	0.136
	Veillonella montpellierensis	15 (12.10)	-1.65	0.70	0.019
	Sediminibacterium salmoneum	10 (8.06)	2.36	1.12	0.035
	Prevotella amnii	45 (36.29)	-1.08	0.46	0.019
	Lactobacillus kefiranofaciens	49 (39.52)	-0.70	0.44	0.109
	Bifidobacterium breve	12 (9.68)	-0.88	0.69	0.202

Abbreviations: ASC-US, atypical squamous cells of undetermined significance; HSIL, high-grade squamous intraepithelial lesion; HPV, human papillomavirus; HPV, human papillomavirus; LSIL, low-grade squamous intraepithelial lesion

[a] Includes 16 ASC-H, atypical squamous cells-cannot rule out HSIL, 2 AGC, atypical glandular cells, and 1 cancer case.
[b] Calculated based on the presence of 6/77 bacterial species, retained using a stepwise (forward selection) logistic regression (dependent variable: presence/absence of microbial species). Interpret with caution due to model convergence concerns when selecting species.

S-Table 4-7. Stepwise logistic regression coefficients of cytology-, HPV-, and microbiome- (species raw abundance) based scores used to construct linear scores, comparing CIN1+ to normal histology.

Test-based score	Variables included	Prevalence in overall study population, n (%)	Coefficient	Standard Error	P-value
Cytology-based (constant = -2.05)	ASC-US	21 (11.29)	3.84	0.82	< 0.001
	LSIL	54 (29.03)	3.41	0.63	< 0.001
	HSIL or worse[a]	58 (31.18)	4.96	0.80	< 0.001
	Missing	18 (9.68)	2.74	0.73	< 0.001
HPV-based (constant = -0.16)	HPV16 and/or 18	44 (23.66)	2.75	0.76	< 0.001
	Other high-risk HPV[b]	101 (54.30)	1.35	0.36	< 0.001
Microbiome-based[c] (constant = 0.89)	*Mycoplasma hominis*	30 (16.13)	0.0012488	0.0025998	0.631
	Prevotella colorans	16 (8.60)	0.0046594	0.0110028	0.672
	Gardnerella vaginalis	153 (82.26)	0.000042	0.0000465	0.366
	Fusobacterium nucleatum	24 (12.90)	0.0005636	0.0016751	0.737
	Lactobacillus iners	186 (100.00)	-0.0000142	0.000017	0.404
Microbiome-HPV-based[d] (constant = -0.11)	*Mycoplasma hominis*	30 (16.13)	0.000576	0.0019523	0.768
	Prevotella colorans	16 (8.60)	0.0064891	0.0123864	0.600
	Gardnerella vaginalis	153 (82.26)	0.0000203	0.0000517	0.695
	Fusobacterium nucleatum	24 (12.90)	0.0003873	0.0020845	0.853
	Lactobacillus iners	186 (100.00)	-0.0000122	0.0000187	0.514
	HPV16 and/or 18	44 (23.66)	2.73	0.7593015	< 0.001
	Other high-risk HPV[b]	101 (54.30)	1.29	0.3731896	0.001

Abbreviations: ASC-US, atypical squamous cells of undetermined significance; HSIL, high-grade squamous intraepithelial lesion; HPV, human papillomavirus; HPV, human papillomavirus; LSIL, low-grade squamous intraepithelial lesion

[a] Includes 16 ASC-H, atypical squamous cells-cannot rule out HSIL, 2 AGC, atypical glandular cells, and 1 cancer case.
[b] Includes a pooled positivity result for any of HPVs 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and/or 68.
[c] Calculated based on the raw abundance of 5/77 bacterial species, retained using a stepwise (forward selection) logistic regression (dependent variable: raw abundance of bacterial species). Interpret with caution due to model convergence concerns when selecting species and one or more independent variables perfectly predicting the outcome.
[d] Calculated based on the raw abundance of the 5 bacterial species that were retained (refer to footnote c) and two additional predictors (positivity for HPV16 and/or 18 and 12 other high-risk HPVs). Interpret with caution due to one or more independent variable perfectly predicting the outcome.

S-Table 4-8. Stepwise logistic regression coefficients of cytology-, HPV-, and microbiome- (species raw abundance) based scores used to construct linear scores, comparing CIN2+ to normal and CIN1 histology.

Test-based score	Variables included	Prevalence in overall study population, n (%)	Coefficient	Standard Error	P-value
Cytology-based[a] (constant = -2.80)	ASC-US	21 (11.29)	2.71	0.85	0.001
	LSIL	54 (29.03)	1.65	0.79	0.037
	HSIL or worse[a]	58 (31.18)	4.64	0.82	< 0.001
	Missing	18 (9.68)	2.35	0.87	0.007
HPV-based (constant = -1.29)	HPV16 and/or 18	44 (23.66)	2.26	0.44	< 0.001
	Other high-risk HPV[b]	101 (54.30)	0.97	0.35	0.005
Microbiome-based[c,d] (constant = -0.10)	*Anaerococcus tetradius*	45 (24.19)	0.02	0.0111918	0.046
	Fannyhessea vaginae	111 (59.68)	-0.001	0.0003787	0.005
	Limosilactobacillus pontis	86 (46.24)	-0.004	0.002537	0.102
	Prevotella amnii	66 (35.48)	0.0005	0.0003023	0.098
	Gemella asaccharolytica	34 (18.28)	0.007	0.0041996	0.092
	Winkia neuii	24 (12.90)	0.01	0.0066745	0.091
Microbiome-HPV based[d] (constant = -1.21)	*Anaerococcus tetradius*	45 (24.19)	0.02	0.0129518	0.061
	Fannyhessea vaginae	111 (59.68)	-0.001	0.0004192	0.007
	Limosilactobacillus pontis	86 (46.24)	-0.003	0.0029326	0.264
	Prevotella amnii	66 (35.48)	0.0006	0.0003411	0.064
	Gemella asaccharolytica	34 (18.28)	0.008	0.0048738	0.090
	Winkia neuii	24 (12.90)	0.01	0.010699	0.213
	HPV16 and/or 18	44 (23.66)	2.28	0.4478686	< 0.001
	Other high-risk HPV[b]	101 (54.30)	0.94	0.3709355	0.011

Abbreviations: ASC-US, atypical squamous cells of undetermined significance; HSIL, high-grade squamous intraepithelial lesion; HPV, human papillomavirus; HPV, human papillomavirus; LSIL, low-grade squamous intraepithelial lesion

[a] Includes 16 ASC-H, atypical squamous cells-cannot rule out HSIL, 2 AGC, atypical glandular cells, and 1 cancer case.
[b] Includes a pooled positivity result for any of HPVs 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and/or 68.
[c] Calculated based on the raw abundance of 6/77 bacterial species, retained using a stepwise (forward selection) logistic regression (dependent variable: raw abundance of bacterial species). Interpret with caution due to model convergence concerns when selecting species and one or more independent variables perfectly predicting the outcome.
[d] Calculated based on the raw abundance of the 6 bacterial species that were retained (refer to footnote c) and two additional predictors (positivity for HPV16 and/or 18 and 12 other high-risk HPVs). Interpret with caution due to one or more independent variable perfectly predicting the outcome.

S-Table 4-9. Stepwise logistic regression coefficients of cytology-, HPV-, and microbiome- (species raw abundance) based scores used to construct linear scores, comparing any high-risk HPV positive to negative.

Test-based score	Variables included	Prevalence in overall study population, n (%)	Coefficient	Standard Error	P-value
Cytology-based (constant = -1.22)	ASC-US	21 (11.29)	1.31	0.59	0.027
	LSIL	54 (29.03)	2.37	0.51	< 0.001
	HSIL or worse[a]	58 (31.18)	3.20	0.57	< 0.001
	Missing	18 (9.68)	2.17	0.66	0.001
Microbiome-based[b], (constant = 0.91)	*Gardnerella vaginalis*	153 (82.26)	0.0001455	0.000069	0.035
	Prevotella colorans	16 (8.60)	0.0002916	0.0007694	0.705
	Prevotella melaninogenica	7 (3.76)	0.0019986	0.0043088	0.643
	Limosilactobacillus pontis	86 (46.24)	-0.0028757	0.0021234	0.176

Abbreviations: ASC-US, atypical squamous cells of undetermined significance; HSIL, high-grade squamous intraepithelial lesion; HPV, human papillomavirus; HPV, human papillomavirus; LSIL, low-grade squamous intraepithelial lesion

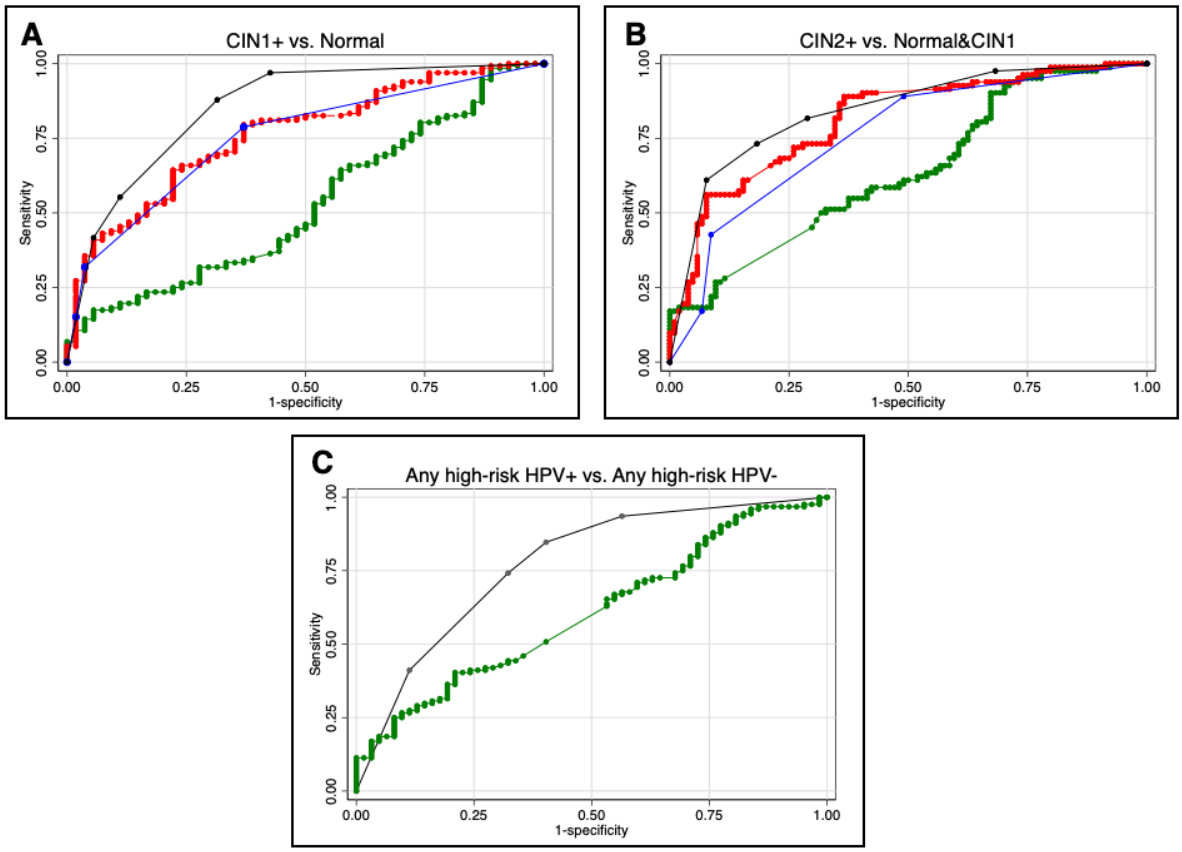
[a] Includes 16 ASC-H, atypical squamous cells-cannot rule out HSIL, 2 AGC, atypical glandular cells, and 1 cancer case.
[b] Calculated based on the raw abundance of 4/77 bacterial species, retained using a stepwise (forward selection) logistic regression (dependent variable: raw abundance of microbial species). Interpret with caution due to one or more independent variables perfectly predicting the outcome during species selection, however dropped observations were added back into a logistic regression model which had no perfect predictors of the outcome.

S-Table 4-10. ROC curve analysis for the performance of cytology-, HPV-, and microbiome- (species raw abundance) based scores to detect CIN lesions and high-risk HPV infections.

Contrast groups	Test-based score	AUC	95% CI[a]	P-value, HPV score as comparator[b]	P-value, cytology score as comparator[b]
CIN1+ vs. Normal	Cytology[c]	0.8524	0.7899 - 0.9149	0.0151	-
	HPV[d]	0.7529	0.6855 - 0.8204	-	0.0151
	Microbiome[e]	0.5252	0.4332 - 0.6172	< 0.001	< 0.001
	Microbiome-HPV[f]	0.7623	0.6894 - 0.8352	0.5574	0.0389
CIN2+ vs. Normal & CIN1	Cytology[c]	0.8431	0.7879 - 0.8983	0.0441	-
	HPV[d]	0.7591	0.6947 - 0.8235	-	0.0441
	Microbiome[g]	0.6377	0.5584 - 0.7171	0.0197	< 0.001
	Microbiome-HPV[h]	0.8130	0.7510 - 0.8749	0.0120	0.4708
HPV+ vs. HPV-	Cytology[c]	0.7733	0.7011 - 0.8456	-	-
	Microbiome[i]	0.6052	0.5210 - 0.6893	-	0.0044

Abbreviations: AUC, Area under the curve; CIN, cervical intraepithelial neoplasia; hrHPV; high-risk HPV; ROC, receiver operating characteristic

[a] 95% confidence interval based on the asymptotic normal.
[b] P-value represents the equality of AUC based on a chi2 test, $P < 0.05$ was considered significant.
[c] Calculated based on categories for ASC-US, LSIL, HSIL or worse, and missing cytology results. Refer to methods and Supplemental Table 7 for the linear scores.
[d] Calculated based on positivity for HPV16 and/or 18 and 12 other high-risk HPVs.
[e] Calculated based on the raw abundance of 5/77 bacterial species, retained using a stepwise (forward selection) logistic regression (dependent variable: raw abundance of microbial species). Interpret with caution due to model convergence concerns when selecting species and one or more predictor perfectly predicting the outcome. Refer to methods and Supplemental Table 7 for the retained species and linear scores.
[f] Calculated based on the raw abundance of 5 bacterial species that were retained (refer to footnote e) and two additional predictors (positivity for HPV16 and/or 18 and 12 other high-risk HPVs).
[g] Calculated based on the raw abundance of 6/77 bacterial species, retained using a stepwise (forward selection) logistic regression (dependent variable: raw abundance of microbial species). Interpret with caution due to one or more predictor perfectly predicting the outcome.
[h] Calculated based on the raw abundance of the 6 bacterial species that were retained (refer to footnote g) and positivity for HPV16 and/or 18 and 12 other high-risk HPVs.
[i] Calculated based on the raw abundance of 4/77, retained using a stepwise (forward selection) logistic regression (dependent variable: raw abundance of microbial species). Interpret with caution due to one or more independent variables perfectly predicting the outcome during species selection, however dropped observations were added back into a logistic regression model which had no perfect predictors of the outcome.



S-Figure 4-2. Performance of cytology-, HPV-, and microbiome- (species raw abundance) based scores to detect CIN and high-risk HPV.

Supplemental Figure 2 Legend

The ROC curves plot the performance of cytology- (black), HPV- (blue), microbiome- (green), and microbiome-HPV- (red) based scores for the detection of CIN1+ (Panel A) and CIN2+ (Panel B). Panel C plots the performance of cytology- (black) and microbiome- (green) based scores for the detection of any high-risk HPV. Refer to Supplemental Table 10 footnotes for the calculation of the different scores.

Abbreviations: CIN, cervical intraepithelial neoplasia; HPV, human papillomavirus; ROC, receiver operating characteristic

S-Table 4-11. Univariate logistic regression coefficients of *molBV*$_{V3-V4}$- and *molBV*$_{V5-V6}$-based scores used to construct linear scores for the contrast groups.

Contrast groups	Score[a]	Coefficient	Standard Error	P-Value
CIN1+ (n = 130) vs. Normal (n = 53)	*molBV*$_{V3-V4}$ (constant = 1.02)	-0.02	0.12	0.844
	molBV$_{V5-V6}$ (constant = 1.06)	-0.04	0.11	0.742
CIN2+ (n = 80) vs. Normal & CIN1 (n=103)	*molBV*$_{V3-V4}$ (constant = -1.09)	0.16	0.11	0.141
	molBV$_{V5-V6}$ (constant = -0.84)	0.13	0.10	0.182
HPV+ (n = 122) vs. HPV- (n = 61)	*molBV*$_{V3-V4}$ (constant = -0.38)	0.21	0.12	0.077
	molBV$_{V5-V6}$ (constant = -0.43)	0.26	0.12	0.025

Abbreviation: CIN, cervical intraepithelial neoplasia; HPV, human papillomavirus; *molBV*, molecular bacterial vaginosis

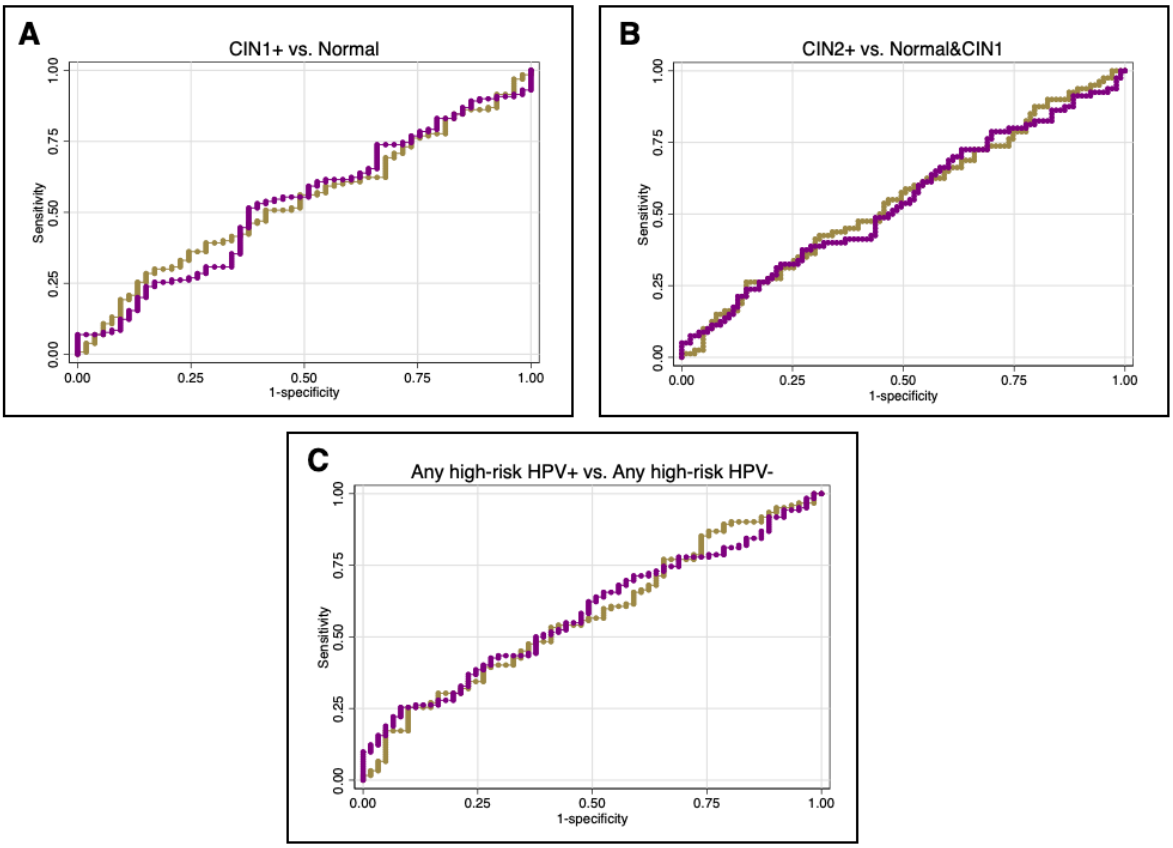
[a] *molBV*$_{V3-V4}$- and *molBV*$_{V5-V6}$- based scores were constructed using the regression coefficients from univariate logistic regression models.

S-Table 4-12. ROC curve analysis for the performance of *molBV*$_{V3-V4}$ - and *molBV*$_{V5-V6}$ - based scores to detect CIN and high-risk HPV infections.

Contrast groups	Test-based scores	AUC	95% confidence interval[a]	P-value, *molBV*$_{V3-V4}$ *as comparator*[b]
CIN1+ vs. Normal	*molBV*$_{V3-V4}$[c]	0.5348	0.4461 - 0.6236	-
	molBV$_{V5-V6}$[d]	0.5318	0.4407 - 0.6229	0.9158
CIN2+ vs. Normal & CIN1	*molBV*$_{V3-V4}$[c]	0.5522	0.4677 - 0.6367	-
	molBV$_{V5-V6}$[d]	0.5434	0.4583 - 0.6286	0.7505
HPV+ vs. HPV-	*molBV*$_{V3-V4}$[c]	0.5752	0.4884 - 0.6621	-
	molBV$_{V5-V6}$[d]	0.5808	0.4962 - 0.6654	0.8539

Abbreviations: AUC, Area under the curve; CIN, cervical intraepithelial neoplasia; HPV, human papillomavirus; molBV, molecular bacterial vaginosis; ROC, receiver operating characteristic curve

[a] Calculated based on the asymptotic normal test.
[b] P-value represents the equality of ROC areas based on a chi2, $P < 0.05$ was considered significant.
[c] Calculated based on a logistic regression with the *molBV* results from the V3V4 region. Refer to methods and Supplemental Table 11 for the linear scores.
[d] Calculated based on a logistic regression with the *molBV* results from the V5V6 region. Refer to methods and Supplemental Table 11 for the linear scores.



S-Figure 4-3. Performance of *molBV*$_{V3\text{-}V4}$ - and *molBV*$_{V5\text{-}V6}$ -based scores to detect CIN and high-risk HPV.

S-Figure 4-3 Legend

The ROC curves plot the performance of *molBV*$_{V3\text{-}V4}$ - (brown), and *molBV*$_{V5\text{-}V6}$ - (purple) based scores for the detection of CIN1+ (Panel A) and CIN2+ (Panel B). Panel C plots the performance of *molBV*$_{V3\text{-}V4\text{-}}$ (brown) and *molBV*$_{V5\text{-}V6\text{-}}$ (purple) based scores for the detection of any high-risk HPV. Refer to Supplemental Table 10 footnote for the calculation of the different scores.

Abbreviations: CIN, cervical intraepithelial neoplasia; HPV, human papillomavirus; *molBV*, molecular bacterial vaginosis; ROC, receiver operating characteristic