UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
FACULDADE DE MEDICINA

Júlia Andrade Pessoa Morales

The vaginal microbiome of reproductive-age women is an indicator of cervical human papillomavirus infection

Dissertação apresentada à Faculdade de Medicina, Universidade Estadual Paulista "Júlio de Mesquita Filho", Câmpus de Botucatu, para otenção do Título de Mestre em Ciências – Área: Patologia.

Orientadora: Profa. Dra. Márcia Guimarães da Silva
Coorientadora: Profa. Dra. Camila Marconi

Botucatu
2020

Júlia Andrade Pessoa Morales

The vaginal microbiome is an indicator of cervical human papillomavirus infection

Dissertação apresentada à Faculdade de Medicina, Universidade Estadual Paulista "Júlio de Mesquita Filho", Câmpus de Botucatu, para otenção do Título de Mestre em Ciências – Área: Patologia.

Orientadora: Profa. Dra. Márcia Guimarães da Silva
Coorientadora: Profa. Dra. Camila Marconi

Botucatu
2020

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ROSEMEIRE APARECIDA VICENTE-CRB 8/5651

Morales, Julia Andrade Pessoa.
The vaginal microbiome of reproductive-age women is an indicator of cervical human papillomavirus infection / Julia Andrade Pessoa Morales. - Botucatu, 2020

Dissertação (mestrado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Faculdade de Medicina de Botucatu
Orientador: Marcia Guimarães da Silva
Coorientador: Camila Marconi
Capes: 40105008

1. Doenças por papilomavírus. 2. Papillomaviridae. 3. Microbiota. 4. Papillomaviridae. 5. Vagina - Infecções.

Palavras-chave: HPV; Microbioma vaginal; Microbiota vaginal; Papillomavírus humano.

Sumário

Dedico este trabalho à minha mãe, Isabel, por ter abdicado tantas vezes de suas próprias vontades em prol da realização dos nossos sonhos, e, por sempre ter me incentivado à estudar. Obrigada pela formação sólida e ética, nada disso seria possível sem você.

Agradecimento

Sou grata a cada uma das pessoas que passou pelo meu caminho até aqui, permitindo que eu me tornasse quem eu sou hoje, do ponto de vista pessoal, profissional e humano.

Diversas pessoas, conversas e vivências compõem a essência de quem eu sou hoje. Meu eterno muito obrigada a todos esses valiosos instantes de tempo, lembranças e sentimentos, compartilhados com tantas pessoas.

A finalização deste ciclo, para mim, representa muito além da obtenção de um título acadêmico; representa também a realização da jornada de um grande encontro comigo mesma.

Agradeço, também, às diversas fontes de fomento que permitiram minha formação e a consolidação desse estudo.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, CNPq e Fapesp (Processos 2018/18469-9, 2012/16800-3, 2012/01278-0).

“Educar é impregnar de sentido o que fazemos a cada instante.”

Paulo Freire

Resumo

Introdução: O entendimento do papel desempenhado pela microbiota vaginal para a saúde reprodutiva tem ganhado destaque na literatura mundial, especialmente no que se refere à associação com a infecção pelo Papilomavírus Humano (HPV). Diversos fatores de risco para a infecção por HPV já foram identificados, com destaque para a vaginose bacteriana (VB). Apesar de tal associação já ter sido demonstrada, diversos aspectos dessa relação permanecem desconhecidos, dada a diversidade microbiana presente no ambiente vaginal. Embora a recente caracterização do microbioma vaginal tenha permitido o conhecimento da real composição microbiana nesse ambiente, a relação entre componentes do microbioma vaginal e a infecção por HPV permanece por ser elucidada. **Objetivo:** O objetivo desse estudo é avaliar a associação entre componentes do microbioma vaginal e a infecção cervical por genótipos de alto risco oncogênico do HPV (hrHPV). **Pacientes e Métodos:** Trata-se de estudo transversal incluindo 546 mulheres oriundas das 5 macrorregiões brasileiras incluídas em estudo prévio intitulado "Caracterização do microbioma vaginal de mulheres brasileiras em idade reprodutiva". O microbioma vaginal foi determinado pelo sequenciamento das regiões V3-V4 do gene bacteriano RNAr 16S em equipamento MiSeq Plataforma 300PE (Illumina, San Diego, CA) e agrupados em *community-state types* (CSTs). Amostras cervicais foram submetidas à pequisa e genotipagem de HPV empregando-se o kit *Linear Array HPV genotyping* (Roche Molecular Systems, Pleasanton, CA). Modelos de regressão logística *stepwise* (forward, P<0,15) foram utilizados para a construção de dois escores lineares para predizer a positividade para hrHPV: um baseado exclusivamente na presença dos taxons bacterianos identificados (*microbiome-based [MB] score*) e o outro baseado exclusivamente em características sociodemográficas, comportamentais e clínicas (*SBC score*). O *MB score* combinou coeficientes de 30 (de 116) espécies retidas no modelo. O *SBC score* reteu 6 (idade, estatus marital, nova parceria sexual, uso de contraceptivo hormonal, índice de massa corporal, e tabagismo) de 25 variáveis candidatas. Foram construídas curvas ROC para os escores para a correlação com hrHPV e comparadas suas AUC e 95% IC para inferir a diferença em suas performances de predição. **Resultados:** A prevalência de hrHPV foi 15,8% (n=86) e 143 (26,2%)

participantes tiverem microbioma com depleção de predomínio lactobacilar. As AUC foram 0,8022 (IC: 0.7517-0.8527) para o *MB score* e 0,7027 (IC: 0.6419-0.7636) para o *SBC score* (P=0,0163 para a diferença entre AUCs). **Conclusão:** O estudo demonstra a forte correlação entre componentes do microbioma vaginal e a infecção por hrHPV, propondo a utilidade dos componentes do microbioma como indicador da infecção por hrHPV, aqui demonstrados através do *MB score*. A utilidade clínica do *score* como ferramenta preditora da infecção por hrHPV deve ser validada posteriormente em estudos longitudinais.

Abstract

Background: The understanding of the role played by vaginal microbiome on reproductive health has gained prominence in literature recently, specially regarding to its association to human papillomavirus (HPV) infection. Several risk factors for cervical HPV infection were already identified, highlighting bacterial vaginosis (BV). Despite the association between HPV and BV is well recognized, most aspects of this relation remain unknown. This is in part due to the great heterogeneity present in vaginal environment. Although the recent vaginal microbiome characterization has allowed the understanding of the real microbial composition in that local, the relation between vaginal microbiome components and HPV infection is to be elucidated. **Aim:** The aim of this study it to evaluate the association between vaginal microbiome components and cervical high-risk HPV (hrHPV) infection. **Patients and Methods:** We cross-sectionally enrolled 546 women from the five macro-regions of Brazil in previous study entitled "Characterization of the vaginal microbiome in Brazilian women of reproductive-age". Vaginal microbiome was then determined by sequencing V3-V4 regions of bacterial rRNA 16S gene through MiSeq Platform 300PE (Illumina, San Diego, CA) and clustered in community-state types (CSTs). Cervical samples were subjected to HPV detection and genotyping using *Linear Array HPV genotyping* kit (Roche Molecular Systems, Pleasanton, CA). We used stepwise (forward, $p<0.15$) logistic regression to construct two linear scores to predict hrHPV positivity: one based exclusively on the presence of individual bacterial taxa (microbiom-base [MB] score) and the other exclusively on participants' sociodemographic, behavioral and clinical (SBC) characteristics. The MB score combined coefficients of 30 (out of 116) species retained in the model. The SBC score retained six (age, marital status, new sex partner, hormonal contraceptive use, body mass index and smoking) out of 25 candidate variables. We constructed receiver operating characteristic curves for the scores as hrHPV correlates and compared the areas under the curve (AUC) and 95% confidence intervals (CI) to infer the difference in predictive performance. **Results**: The prevalence of hrHPV was 15.8% (n=86), and 143 (26.2%) participants had *Lactobacillus*-depleted vaginal microbiome. The AUCs were 0.8022 (CI: 0.7517-0.8527) for the MB score and 0.7027 (CI: 0.6419-0.7636) for the SBC score (P=0.0163 for the difference between AUCs). **Conclusions**:

Our findings demonstrate the strong correlation between vaginal microbiome components and hrHPV positivity, here demonstrated by MB score, warranting further validation of its clinical utility as a hrHPV infection predictor via longitudinal studies.

Revisão de Literatura

O câncer do colo do útero, ou câncer cervical, com 570 mil casos e 311 mil mortes estimadas em 2018 se configura como quarto em incidência e mortalidade por câncer na população feminina mundial. Apesar de seu caráter prevenível baseado em estratégias populacionais de rastreamento e vacinação, o câncer cervical permanece como um grave problema de saúde pública, especialmente em países subdesenvolvidos e em desenvolvimento.(1)

É bem estabelecido o entendimento da história natural da doença, uma vez que é comprovada a relação causal desse tipo de câncer com determinados genótipos de Papillomavírus Humano (HPV). Os papilomavírus humanos são vírus que pertencem à família *Papillomaviridae*. Atualmente, são descritos mais de 140 tipos de HPV delineados pela identidade da sequência de DNA que infectam células epiteliais humanas. Dentre esses, 40 apresentam afinidade para infecção do trato anogenital, que compreendem alguns genótipos que estão envolvidos na etiologia de lesões neoplásicas.(2) Dessa forma, os HPV genitais são classificados de acordo com o risco que conferem ao desenvolvimento de neoplasias. Os principais tipos de HPV genital de baixo risco oncogênico são os HPV6 e HPV11, que estão associados ao desenvolvimento de verrugas genitais. Já os HPV genitais de alto risco, estão associados às lesões intraepiteliais de alto grau e carcinomas invasivos do colo do útero.(3) Dentre os tipos de HPV de alto risco mais prevalentes, destacam-se o HPV16 e o HPV18 que, juntos, são responsáveis por mais de 70% dos casos de carcinomas epidermóides e mais de 80% dos adenocarcinomas do colo do útero.(4, 5)

Os HPV são vírus não-envelopados, constituídos por um capsídeo de simetria icosaédrica, que engloba uma molécula de DNA circular de dupla fita com aproximadamente 8000 bp.(2) O genoma

do HPV é dividido em três regiões. A primeira delas é uma região regulatória (LCR, *long control region*) que contém a origem de replicação do DNA e, portanto, responsável pelo início ciclo replicativo viral. A próxima região é a precoce (*early*) que codifica 6 proteínas (E1, E2, E4, E5, E6 e E7) que possuem ação enzimática ou reguladora no ciclo replicativo viral. Dentre essas, merecem destaque as proteínas E6 e E7, codificadas pelos genes de mesmo nome, que inativam, respectivamente, os supressores tumorais p53 e pRB da célula infectada. (6-8) Dessa forma, tais proteínas estão envolvidas diretamente na imortalização dessas células e, portanto, no desenvolvimento de neoplasias. Finalmente, os genes da região tardia (*late*) que codificam as duas proteínas que compões o capsídeo viral, a L1 e L2, proteínas essas altamente imunogênicas que são utilizadas na produção de vacinas profiláticas. (8)

O potencial oncogênico intrínseco de alguns genótipos de HPV se reflete nos achados de estudos que demonstram que 5% de todos os cânceres diagnosticados anualmente no mundo estão associados à infecção por esse vírus.(9, 10) Embora vários fatores já tenham sido associados ao desenvolvimento do câncer de colo do útero, como o *status* da imunidade do hospedeiro, presença de coinfecções genitais e comportamento sexual, a persistência da infecção por HPV de alto risco é mandatória, visto que o DNA viral está presente em 99,7% dos casos diagnosticados.(11) Por muito tempo, a principal estratégia para prevenção do câncer de colo do útero se baseou em exames citopatológicos para rastreio de lesões pré-neoplásicas, entretanto, dada a relação causal entre a persistência da infecção pelo HPV e o desenvolvimento neoplásico, a vacinação contra o HPV foi implementada em diversas partes do mundo como uma oportunidade sem precedentes para reduzir as taxas de câncer. (12) Atualmente, pelo menos 118 milhões de mulheres já receberam uma dose da vacina de HPV. Embora esse número seja encorajador, representa apenas 3,5% da população mundial. (13)

De forma geral, as infecções por HPVs genitais são altamente prevalentes na população sexualmente ativa, e, possuem caráter predominantemente transitório, visto que a maioria das infecções é eliminada pelo sistema imunológico do hospedeiro num período de 1 a 2 anos. (6) Entretanto, o maior período de persistência da infecção por genótipos de alto risco oncogênico é o principal fator que leva ao desenvolvimento de lesões HPV induzidas no colo do útero. (7) A literatura é consistente ao demonstrar que o tempo de *clearance* para os HPVs de alto risco oncogênico é significativamente maior quando comparado aos de baixo risco. (8)

Embora o câncer cervical seja causado por infecções persistentes por genótipos de HPVs de alto risco oncogênico do HPV, a infecção é considerada um fator necessário, mas não suficiente para o desenvolvimento neoplásico. Nesse sentido, outros fatores que alteram o microambiente vaginal têm sido identificados por seu papel desempenhado na progressão da doença. (14) Esse fato é ainda ressaltado considerando o entendimento de que, a maior parte das mulheres irá se infectar com pelo menos um genótipo de HPV ao longo de sua vida, mas a ocorrência de persistência por genótipos de alto risco, a progressão de lesões precursoras e, o desenvolvimento neoplásico propriamente, ocorre somente em uma parcela dos indivíduos, ressaltando o papel desempenhado por outros co-fatores. (15, 16)

Além de fatores comportamentais, outros como dieta pobre em nutrientes, uso de contraceptivo hormonal de progesterona, e a presença de infecções sexualmente transmissíveis foram reportados como fatores associados à persistência viral. Mais recentemente, o papel desempenhado pela microbiota vaginal tem ganhado destaque em seu entendimento como fator associado à oncogênese do câncer cervical. (12, 13, 17-19)

Desta forma, estudos tem demonstrado que mulheres com disbioses vaginais possuem risco aumentado para a aquisição e persistência da infecção por HPV, bem como risco aumentado de progressão pré-neoplásica e neoplásica desse tipo de câncer. (15)

Por definição, a microbiota vaginal saudável é representada pelo predomínio de *Lactobacillus* spp. É reconhecido o papel protetor desempenhado por esta microbiota por meio de diversos mecanismos no ambiente vaginal. Dentre os fatores protetores mediados pela microbiota vaginal, destacam-se a manutenção do ambiente vaginal em pH ácido, especialmente através da produção de ácido lático e peróxido de hidrogênio. (20) No mesmo sentido, a produção de bacteriocinas e biosurfactantes por determinadas espécies de lactobacilos auxiliam na manutenção de uma barreira mucosa íntegra e protetora contra a infecção por vírus e bactérias exógenas. De forma geral, tais mecanismos previnem o crescimento e adesão de outras bactérias oportunistas ao epitélio vaginal. (21-24)

Por outro lado, a presença de disbioses vaginais se configura como fator de risco para diversas complicações ginecológicas e obstétricas. Dentre as alterações de microbiota vaginal, em especial a vaginose bacteriana (VB) é associada a risco aumentado de abortos espontâneos e recorrentes, parto pré-termo, rotura prematura de membranas e maior suceptibilidade à aquisição de diversas infecções sexualmente transmissíveis (IST), dentre elas pelo vírus da Imunodeficiência Humana (HIV). (15, 21, 25)

No que se refere ao entendimento das disbioses vaginais como fator de risco para a aquisição de ISTs, atribui-se a maior susceptibilidade à perda dos mecanismos de proteção mediados por espécies lactobacilares. De forma geral, as disbioses vaginais estão associadas à uma falha na integridade da barreira mucosa, bem como associadas a um cenário inflamatório crônico e níveis

elevados de espécies reativas de oxigênio, demonstrando um contexto geral de estresse oxidativo. (15, 26, 27)

No que tange a associação entre disbioses vaginais, susceptibilidade à aquisição e persistência da infecção por HPV e o desenvolvimento displásico e neoplásico, reconhece-se que os capsídeos virais não são capazes de infectar o epitélio vaginal íntegro. Desta forma, a falha nos mecanismos protetores da barreira mucosa em condições disbióticas pode facilitar tal processo. (15, 26) Além disso, o cenário inflamatório crônico na mucosa pode ser considerado um fator favorável à carcinogênese induzida por HPV. (15, 28)

Um estudo recente que caracterizou o metaboloma vaginal de mulheres infectadas e não infectadas por HPV demonstra que, esse metaboloma vaginal não pode ser agrupado de acordo com o *status* da infecção pelo vírus, mas, paralelamente, pode ser agrupado de forma geral baseado na composição bacteriana desse ambiente. Perfis metabolômicos diferenciais são encontrados frente à infecção por HPV quando considerando um mesmo perfil de microbiota, em especial no que se refere aos metabólitos representativos de estresse oxidativo. (29)

Tendo em vista a compreensão de que as disbioses vaginais compreendem uma série de diversas alterações de microbiota, é importante ressaltar a vaginose bacteriana (VB), a mais prevalente entidade, acomete cerca de 30% da população feminina em idade reprodutiva. Por definição, a VB é caracterizada pela substituição total ou parcial das espécies de lactobacilos por grande diversidade de espécies bacterianas, em sua maioria anaeróbias. Dessa forma, a VB é considerada uma condição polimicrobiana, dada a grande quantidade de espécies a ela associadas e também à heterogeneidade com que se apresenta. (30)

Do ponto de vista clínico, os critérios de Amsel et al. (31) foram amplamente utilizados para o diagnóstico desta condição, baseado em quatro critérios: identificação de conteúdo vaginal fino e

homogêneo; pH vaginal acima de 4,5; identificação da presença de aminas voláteis quando em adição de hidróxido de potássio (*whiff test*); e identificação de células-pista (*clue cells*) à observação microscópica. Entretanto, a subjetividade atribuída a parte dos critérios imputava limitações aos critérios clínicos como metodologia diagnóstica.

No início da década de 90, Nugent et al. (32) propuseram um critério microscópico para diagnóstico da VB, amplamente utilizado e considerado padrão-ouro até os dias de hoje. O critério é baseado na classificação semi-quantitativa dos morfotipos presentes em esfregaços vaginais corados pelo método de Gram e classifica a microbiota vaginal em Flora I (escores 0 a 3), Flora Intermediária (escores 4 a 6) e Vaginose Bacteriana (escores 7 a 10).

Considerando a heterogeneidade intrínseca à VB, especialmente por seu caráter polimicrobiano por definição, o desenvolvimento de novos métodos moleculares para identificação bacteriana, especialmente através do sequenciamento de nova geração do gene bacteriano RNA ribossômico 16S, trouxe grande evolução na determinação da real composição bacteriana presente no ambiente vaginal pela caracterização de seu microbioma.

Nesse sentido, Ravel et al. (33) demonstraram que, apesar da grande diversidade bacteriana presente no ambiente vaginal, esse microbioma pode ser agrupado em 5 tipos de comunidades bacterianas, as denominadas *community-state types* (CSTs), conforme a predominância de determinadas espécies. Das 5 CSTs, 4 são caracterizadas pelo predomínio de espécies lactobacilares: *Lactobacillus crispatus* (CST I)*, L. gasseri* (CST II)*, L. iners* (CST III) e *L. jensenii* (CST V). Na comunidade bacteriana vaginal remanescente, a CST IV, o predomínio lactobacilar é substituído por aumento significativo na diversidade de espécies bacterianas de forma que não há predominância particular de nenhuma espécie. De fato, a maioria dos casos de VB detectada microscopicamente

encontra-se na CST IV, embora um número ainda considerável de casos de VB encontra-se distribuído nas demais CSTs, especialmente na CST III.

A compreensão microscópica e molecular da composição da microbiota vaginal permitiu a ampliação do entendimento da relação entre esse nicho e a infecção por HPV. Ainda, no que se refere à VB, já foi demonstrado que tal alteração da microbiota vaginal está associada à maior prevalência da infecção pelo HPV, já que alguns metabólitos dos microrganismos anaeróbios comprometem a resposta imune inata no ambiente vaginal conferindo tanto maior susceptibilidade à infecção, quanto a diminuição significativa na taxa do *clearance* viral. (34-37)

Brusselaers et al. (15), em uma revisão sistemática e meta-análise recentemente publicada, baseada em estudos longitudinais que incluem a avaliação microscópica e molecular da microbiota, sugerem a ideia de uma associação causal entre disbioses vaginais e o câncer cervical ao longo da aquisição e persistência da infecção, bem como progressão displásica.

Paralelamente, Norenhag et al. (24), em revisão sistemática e meta-análise também recentemente publicada, baseada exclusivamente critérios moleculares para caracterização da microbiota, postulam que CSTs não dominadas por *Lactobacillus* spp. ou CSTs dominadas por *Lactobacillus iners* estão associadas a prevalências mais elevadas de infecção por HPV quando comparadas a comunidades dominadas por *L. crispatus*, associação encontrada tanto para qualquer genótipo de HPV quanto para genótipos de alto risco oncogênico. Nesse mesmo estudo os autores destacam que essa mesma associação é encontrada em casos de displasia e câncer do colo do útero.

A CST não dominada por *Lactobacillus spp* e a dominada por *L. iners* demonstraram de 3 a 5 vezes mais risco para a infecção por qualquer HPV e 2 a 3 vezes mais risco para a identificação de genótipos de alto risco oncogênico e displasia ou câncer cervical, quando comparados a comunidade dominada por *L. crispatus*. (24)

De forma geral, a compreensão do microbioma vaginal permitiu maior refinamento no entendimento da associação entre a composição microbiana e a infecção por HPV, inclusive através da identificação do papel desempenhado por diferentes espécies de *Lactobacillus spp* frente à infecção. Esse entendimento ressalta, também, a possibilidade de compreensão mais profunda das relações estabelecidas por determinadas espécies bacterianas e o HPV, em especial tendo em vista a heterogeneidade intrínseca da microbiota, em especial nos casos de VB e CST IV.

Desta forma, a ideia de uma associação entre determinadas espécies bacterianas e a infecção por HPV, em especial os de alto risco oncogênico, tem sido proposta recentemente. (38)

De forma pontual, *Fusobacterium* spp. tem sido proposto por seu papel como possível facilitador oncogênico e sua associação ao desenvolvimento displásico, (9,20 – 44) em especial de forma paralela aos entendimentos do microbioma intestinal e o papel desempenhado por essa bactéria no câncer colorretal. (24, 39-41)

Paralelamente, as espécies *Gardnerella vaginalis* e *Atopobium vaginae* foram propostos como possíveis marcadores moleculares, considerando o fato de ambos contribuírem para a formação de biofilmes, em especial pelo papel que esta formação pode desempenhar na persistência do HPV. (24, 42-45)

Similarmente, uma associação foi descrita associação entre genótipos de alto risco oncogênico e a abundância de *Prevotella* spp. e *Leptotrichia* spp. (46), enquanto *Snethia* spp. foi descrita por seu possível papel na carcinogênese cervical (47) e reportado como possível marcador para infecções por HPV de alto risco oncogênico. (21)

Adicionalmente, determinadas espécies, dentre elas *Atopobium* spp. (que não *A. vaginae*), *Dialister* spp., *Eggerthella* spp., *Gemella* spp. e *Gardnerella* spp., foram encontradas de forma mais

abundante em mulheres com infecção por HPV classificadas com CST IV, bem como concordantes com a presença de determinadas aminas biogênicas após avaliação metabolômica. (29)

De forma geral, o emergente entendimento da associação entre determinadas espécies e a infecção viral é um campo promissor na identificação de possíveis marcadores biológicos, em especial por seu potencial de aplicação clínica. A identificação de mulheres com determinadas composições microbianas pode direcionar abordagens de seguimento e manejo diferenciais do ponto de vista clínico frente à infecção por HPV.

Embora causalidade não tenha sido determinada entre componentes microbianos e a persistência e progressão de lesões HPV-induzidas, em especial pelos fatores confundidores associados à composição da microbiota, é relevante ressaltar que a composição microbiana é um fator de risco potencialmente modificável. (15)

Nesse sentido, é relevante ressaltar alguns dos diversos fatores que influenciam a composição da microbiota e que impõem algumas limitações ao completo entendimento da relação microbiota – HPV. São fatores inerentes ao hospedeiro: idade, *status* hormonal, fatores sociodemográficos, fatores comportamentais e sexuais, etnia, bem como fatores genéticos associados, presença de coinfecções, entre outros. Paralelamente, a hetetogeneidade associada e classificações diferenciais da microbiota entre estudos são outros fatores confundidores que permeiam a complexidade desse entendimento. (15, 24)

Apesar dessas limitações, estudos reiteram o papel desempenhado pela microbiota na persistência da infecção e no desenvolvimento do câncer cervical. Nesse aspecto, é importante ressaltar o papel do manejo das disbioses em favor do estabelecimento de um ambiente lactobacilar como um fator de intervenção no controle das alterações HPV-induzidas no colo uterino. É relevante

o potencial de tais intervenções como uma possível ferramenta auxiliar na redução de procedimentos ablativos e excisionais na prática ginecológica. (15)

As disbioses vaginais são ainda pouco estudadas e devidamente diagnosticas, tendo em vista seu papel como fator de risco na epidemiologia do HPV. A avaliação e manejo das disbioses, associadas a estratégias vacinais, possivelmente contribuiriam significativamente na redução do câncer cervical. (15)

Referências

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394-424.
2. zur Hausen H. Papillomavirus infections--a major cause of human cancers. *Biochim Biophys Acta* 1996;1288:F55-78.
3. Muñoz N, Bosch FX, de Sanjosé S, Herrero R, Castellsagué X, Shah KV, et al. Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N Engl J Med* 2003;348:518-27.
4. Bosch FX, de Sanjosé S. Chapter 1: Human papillomavirus and cervical cancer--burden and assessment of causality. *J Natl Cancer Inst Monogr* 2003:3-13.
5. Smith JS, Lindsay L, Hoots B, Keys J, Franceschi S, Winer R, et al. Human papillomavirus type distribution in invasive cervical cancer and high-grade cervical lesions: a meta-analysis update. *Int J Cancer* 2007;121:621-32.
6. Münger K, Howley PM. Human papillomavirus immortalization and transformation functions. *Virus Res* 2002;89:213-28.
7. Scheffner M, Werness BA, Huibregtse JM, Levine AJ, Howley PM. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell* 1990;63:1129-36.
8. Münger K, Werness BA, Dyson N, Phelps WC, Harlow E, Howley PM. Complex formation of human papillomavirus E7 proteins with the retinoblastoma tumor suppressor gene product. *EMBO J* 1989;8:4099-105.
9. Parkin DM. The global health burden of infection-associated cancers in the year 2002. *Int J Cancer* 2006;118:3030-44.
10. Dalton-Griffin L, Kellam P. Infectious causes of cancer and their detection. *J Biol* 2009;8:67.
11. Bosch FX, Lorincz A, Muñoz N, Meijer CJ, Shah KV. The causal relation between human papillomavirus and cervical cancer. *J Clin Pathol* 2002;55:244-65.

12. Luckett R, Feldman S. Impact of 2-, 4- and 9-valent HPV vaccines on morbidity and mortality from cervical cancer. *Hum Vaccin Immunother* 2016;12:1332-42.
13. Bruni L, Diaz M, Barrionuevo-Rosas L, Herrero R, Bray F, Bosch FX, et al. Global estimates of human papillomavirus vaccination coverage by region and income level: a pooled analysis. *Lancet Glob Health* 2016;4:e453-63.
14. Audirac-Chalifour A, Torres-Poveda K, Bahena-Román M, Téllez-Sosa J, Martínez-Barnetche J, Cortina-Ceballos B, et al. Cervical Microbiome and Cytokine Profile at Various Stages of Cervical Cancer: A Pilot Study. *PLoS One* 2016;11:e0153274.
15. 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;221:9-18.e8.
16. Sudenga SL, Shrestha S. Key considerations and current perspectives of epidemiological studies on human papillomavirus persistence, the intermediate phenotype to cervical cancer. *Int J Infect Dis* 2013;17:e216-20.
17. Ginindza TG, Dlamini X, Almonte M, Herrero R, Jolly PE, Tsoka-Gwegweni JM, et al. Prevalence of and Associated Risk Factors for High Risk Human Papillomavirus among Sexually Active Women, Swaziland. *PLoS One* 2017;12:e0170189.
18. Bogani G, Taverna F, Lombardo C, Borghi C, Martinelli F, Signorelli M, et al. Retrospective study of the influence of HPV persistence on outcomes among women with high-risk HPV infections and negative cytology. *Int J Gynaecol Obstet* 2017;138:62-8.
19. Molano M, Van den Brule A, Plummer M, Weiderpass E, Posso H, Arslan A, et al. Determinants of clearance of human papillomavirus infections in Colombian women with normal cytology: a population-based, 5-year follow-up study. *Am J Epidemiol* 2003;158:486-94.
20. Graver MA, Wade JJ. The role of acidification in the inhibition of Neisseria gonorrhoeae by vaginal lactobacilli during anaerobic growth. *Ann Clin Microbiol Antimicrob* 2011;10:8.
21. Lee JE, Lee S, Lee H, Song YM, Lee K, Han MJ, et al. Association of the vaginal microbiota with human papillomavirus infection in a Korean twin cohort. *PLoS One* 2013;8:e63514.
22. Boris S, Barbés C. Role played by lactobacilli in controlling the population of vaginal pathogens. *Microbes Infect* 2000;2:543-6.
23. Aroutcheva A, Gariti D, Simon M, Shott S, Faro J, Simoes JA, et al. Defense factors of vaginal lactobacilli. *Am J Obstet Gynecol* 2001;185:375-9.
24. 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* 2019.
25. Unemo M, Bradshaw CS, Hocking JS, de Vries HJC, Francis SC, Mabey D, et al. Sexually transmitted infections: challenges ahead. *Lancet Infect Dis* 2017;17:e235-e79.
26. Schwabe RF, Jobin C. The microbiome and cancer. *Nat Rev Cancer* 2013;13:800-12.
27. Garrett WS. Cancer and the microbiota. *Science* 2015;348:80-6.
28. Fernandes JV, DE Medeiros Fernandes TA, DE Azevedo JC, Cobucci RN, DE Carvalho MG, Andrade VS, et al. Link between chronic inflammation and human papillomavirus-induced carcinogenesis (Review). *Oncol Lett* 2015;9:1015-26.
29. Borgogna JC, Shardell MD, Santori EK, Nelson TM, Rath JM, Glover ED, et al. The vaginal metabolome and microbiota of cervical HPV-positive and HPV-negative women: a cross-sectional analysis. *BJOG* 2020;127:182-92.

30. Mitchell C, Marrazzo J. Bacterial vaginosis and the cervicovaginal immune response. *Am J Reprod Immunol* 2014;71:555-63.
31. Amsel R, Totten PA, Spiegel CA, Chen KC, Eschenbach D, Holmes KK. Nonspecific vaginitis. Diagnostic criteria and microbial and epidemiologic associations. *Am J Med* 1983;74:14-22.
32. 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;29:297-301.
33. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, McCulle SL, et al. Vaginal microbiome of reproductive-age women. *Proc Natl Acad Sci USA* 2011;108 Suppl 1:4680-7.
34. Watts DH, Fazzari M, Fazarri M, Minkoff H, Hillier SL, Sha B, 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;191:1129-39.
35. Gillet E, Meys JF, Verstraelen H, Bosire C, De Sutter P, Temmerman M, et al. Bacterial vaginosis is associated with uterine cervical human papillomavirus infection: a meta-analysis. *BMC Infect Dis* 2011;11:10.
36. 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;2011:319460.
37. Guo YL, You K, Qiao J, Zhao YM, Geng L. Bacterial vaginosis is conducive to the persistence of HPV infection. *Int J STD AIDS* 2012;23:581-4.
38. Huang X, Li C, Li F, Zhao J, Wan X, Wang K. Cervicovaginal microbiota composition correlates with the acquisition of high-risk human papillomavirus types. *Int J Cancer* 2018;143:621-34.
39. Marchesi JR, Dutilh BE, Hall N, Peters WH, Roelofs R, Boleij A, et al. Towards the human colorectal cancer microbiome. *PLoS One* 2011;6:e20447.
40. Kostic AD, Chun E, Robertson L, Glickman JN, Gallini CA, Michaud M, et al. Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe* 2013;14:207-15.
41. Viaud S, Saccheri F, Mignot G, Yamazaki T, Daillère R, Hannani D, et al. The intestinal microbiota modulates the anticancer immune effects of cyclophosphamide. *Science*. 2013;342:971-6.
42. 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;7:10200.
43. Machado A, Cerca N. Influence of Biofilm Formation by Gardnerella vaginalis and Other Anaerobes on Bacterial Vaginosis. *J Infect Dis* 2015;212:1856-61.
44. Swidsinski A, Mendling W, Loening-Baucke V, Ladhoff A, Swidsinski S, Hale LP, et al. Adherent biofilms in bacterial vaginosis. *Obstet Gynecol* 2005;106:1013-23.
45. Swidsinski A, Loening-Baucke V, Mendling W, Dörffel Y, Schilling J, Halwani Z, et al. Infection through structured polymicrobial Gardnerella biofilms (StPM-GB). *Histol Histopathol* 2014;29:567-87.
46. Dareng EO, Ma B, Famooto AO, Adebamowo SN, Offiong RA, Olaniyan O, et al. Prevalent high-risk HPV infection and vaginal microbiota in Nigerian women. *Epidemiol Infect* 2016;144:123-37.

47. Łaniewski P, Barnes D, Goulder A, Cui H, Roe DJ, Chase DM, et al. Linking cervicovaginal immune signatures, HPV and microbiota composition in cervical carcinogenesis in non-Hispanic and Hispanic women. *Sci Rep* 2018;8:7593.

Artigo Científico

Manuscrito apresentado de acordo com as normas para submissão ao periódico *Clinical Infectious Diseases*.

Vaginal microbiome components are an indicator of cervical human papillomavirus infection

Running title: Vaginal microbiome and HPV infection

Keywords: vaginal microbiome, HPV infection, high-risk HPV, 16S RNA sequencing

Summary: This cross-sectional study demonstrates the strong correlation between vaginal microbiome components and cervical high-risk HPV positity, suggesting its utility as hrHPV indicator.

Authors: Julia Andrade Pessoa Morales,[1] Camila Marconi,[1,2] Mariam El-Zein,[3] Jacques Ravel,[4] Gabriel Victor da Silva Pinto,[1] Rosana Silveira,[1] Moisés Diogo de Lima,[5] Newton Sérgio de Carvalho,[2] Rosane Ribeiro Figueiredo Alves,[6] Cristina Maria Garcia de Lima Parada,[7] Sandra Helena Morais Leite,[8] Luisa L. Villa,[9] Eduardo L. Franco,[3] Márcia Guimarães da Silva[1]

Afilliation:

[1] Department of Pathology, Botucatu Medical School, Sao Paulo State University (UNESP), Botucatu, São Paulo, Brazil.
[2] Department of Basic Pathology, Federal University of Paraná (UFPR), Curitiba, Paraná, Brazil
[3] Division of Cancer Epidemiology, McGill University, Montréal, Québec, Canada.
[4] Institute of Genomic Science, University of Maryland School of Medicine, Baltimore, Maryland, USA.
[5] Department of Gynecology and Obstetrics, Federal University of Paraíba (UFPB), João Pessoa, Paraíba, Brazil.

[6] Department of Gynecology and Obstetrics, Federal University of Goiás (UFG), Goiania, Goiás, Brazil.
[7] Department of Nursing, Botucatu Medical School, Sao Paulo State University (UNESP), Botucatu, São Paulo, Brazil.
[8] Department of Gynecology and Obstetrics, State University of Pará (UEPA), Belém, Pará, Brazil.
[9] Center for Translational Investigation in Oncology, Instituto do Câncer do Estado de São Paulo, Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil.

1. Abstract

Background: The association between vaginal microbiome and human papillomavirus (HPV) remains unclear, partly due to the heterogeneity of the microbiota across population. We evaluated the association between microbiome components and cervical high-risk HPV (hrHPV) infection. **Methods**: We enrolled 546 women aged 18-51 years in a cross-sectional study conducted in five Brazilian regions. Cervicovaginal samples were genotyped for HPV using Roche's Linear Array test. For vaginal microbiome analysis, the V3-V4 region of 16S rRNA gene was sequenced (Illumina). We used stepwise (forward, $p<0.15$) logistic regression to construct two linear scores to predict hrHPV positivity: one based exclusively on the presence of individual bacterial taxa (microbiome-based [MB] score) and the other exclusively on participants' sociodemographic, behavioral and clinical (SBC) characteristics. The MB score combined coefficients of 30 (out of 116) species retained in the model. The SBC score retained six out of 25 candidate variables. We constructed receiver operating characteristic curves for the scores as hrHPV correlates and compared the areas under the curve (AUC) and 95% confidence intervals (CI) to infer the difference in predictive performance. **Results**: The prevalence of hrHPV was 15.8% (n=86), and 143 (26.2%) participants had *Lactobacillus*-depleted vaginal microbiome. The AUCs were 0.8022

(CI: 0.7517-0.8527) for the MB score and 0.7027 (CI: 0.6419-0.7636) for the SBC score (P=0.0163 for the difference between AUCs). **Conclusions**: Our findings suggest that the composition of the vaginal microbiome is strongly correlated with hrHPV positivity, warranting further validation of its clinical utility via longitudinal studies.

2. Introduction

Microbiome signatures of diseases have provided cutting-edge information for uncovering microbiome-host relations. Recent studies of the integrative phase of the Human Microbiome Project (HMP) have established microbial signatures of the gut and vagina for colorectal cancer and preterm labor, respectively.(1, 2) Although no causal relation between the microbiome components and disease was addressed, these studies have identified the microbial taxa that are shared by individuals with the aforementioned diseases.

In the context of vaginal microbiota, dominance of *Lactobacillus* spp. over other bacteria has long been considered as an indicator of healthy state, mainly due to the acidic environment originating from glycogen fermentation that protects against pathogens, including sexually transmitted infections (STI).(3, 4) Conversely, women with vaginal dysbiosis, generally referred to as bacterial vaginosis, are at an increased risk of acquiring several STIs, including the human papillomavirus (HPV).(5-8) HPV-related diseases have an enormous impact on public health, especially in regards to women reproductive health. Cervical cancer is caused by persistent infection with oncogenic high-risk HPV (hrHPV) genotypes, and is the fourth most frequent type of cancer in women, excluding non-melanoma skin tumours.(9)

Bacterial vaginosis (BV) has been associated with incident(6-8) and persistent HPV infection(6) as well as with progression to premalignant and malignant cervical lesions(10, 11). Up to now, a few studies have used a microbiome-based approach to evaluate the association between vaginal bacteria and HPV.(12-14) Despite being limited to small sample sizes and targeting specific populations, the latter studies have corroborated previous reports of an association between BV and persistent HPV infection (6-8). However, it remains unknown if certain bacterial taxa are actually associated with cervical hrHPV genotypes. We aimed to evaluate the association between vaginal microbiome components, especially bacterial taxa identified by high throughput 16S rRNA sequencing, and cervical hrHPV infection.

Material and Methods

Study design and population

The study population consisted of a subset of participants enrolled in a cross-sectional study (2013-2015) that aimed to characterize and compare the vaginal microbiome in reproductive-age women from five geopolitical regions of Brazil. Participants were enrolled in the North (Belém/PA, n=133), Northeast (João Pessoa/PB, n=108), Central-West (Goiânia/GO, n=119), Southeast (Botucatu/SP, n=140) and South (Curitiba/PR, n=109) regions. Recruitment in all regions, except for the South, took place in primary healthcare units (community-based population) by approaching women who were attending their routine Pap-testing. In the South, participants were recruited in outpatient clinics (general gynecology, pre- and post-operative care) at a university referral hospital. The study was approved by the review board of all participating centers (approval number 306.547), and all participants signed an informed consent form.

We recruited women meeting the following eligibility criteria upon enrollment: aged ≥18 years, not pregnant; having regular monthly menstrual periods and had their last period since at least 5 days; no vaginal sex or medical procedures (e.g., an ultrasound examination) for a minimum of 72 hours; and not reporting urinary loss, infection, use of intrauterine device (IUD) and vaginal ring as a contraceptive method, topic and/or systemic antimicrobial drugs in the last 45 days, or a previous positive test for HIV or syphilis. None of the participants had been previously vaccinated against HPV.

Data collection and sampling procedures

We collected data on sociodemographic, behavioral and clinical characteristics using a structured questionnaire in face-to-face interviews. For the physical exam, we used a non-lubricated sterile speculum and measured the vaginal pH by placing a pH strip (range 4.0 to 7.0, Merck, Darmstadt, Germany) against the vaginal wall for 1 minute. We collected two samples from the mid-vaginal wall, the first using Eswabs (Copan, Brescia, Italy) was stored at -80º C for microbiome analysis, and the second using sterile cotton swab was smeared onto microscopic glass slides for microscopic classification. After smears confection, we added three droplets of a 10% [v/v] potassium hydroxide (KOH) to the swab to check for presence of volatile amines indicative of microbial anaerobic metabolism. Finally, we collected endocervical brush samples that were stored at -20º C for *Chlamydia trachomatis*, *Neisseria gonorrhoeae*, *Trichomonas vaginalis* and HPV molecular testing.

Vaginal microbiota assessment

We performed Gram-staining of vaginal smears for microscopic observation. Microbiota was classified according to the Nugent score system that comprises three categories: normal (scores 0 to 3), intermediate (scores 4 to 6) or BV (scores 7 to 10).(15)

For microbiome assessment, we extracted total bacterial DNA using MoBio PowerSoil kit (MoBio, Carlsbad, CA, USA), according to manufacturers' instruction. We further performed PCR targeting V3-V4 regions of the 16S rRNA gene using Phusion High-Fidelity PCR Master (Thermo Scientific, Waltham, MA, USA) and purified amplicons using SPRIselect Reagent Kit (Beckman Coulter Life Sciences, Indianapolis, IN, USA). We performed DNA sequencing of amplicon libraries using 300 PE protocol in Illumina MiSeq instrument (Illumina, San Diego, CA, USA), using methods specifically developed for vaginal samples by Fadrosh et al.(16) After checking and trimming for quality, reads were paired and assembled using FLASH with overlap by ~90 bp.(17) We processed and de-multiplexed raw reads in QIIME (version 1.8.0)(18) and performed *de novo* and reference-based chimera detection using UCHIME (v5.1)(19) with greengenes database (Aug, 2013 version).(20) For precise species level assignments, we performed species and genera assignments of 16S rRNA sequences using an in-house fifth-order Markov Chain model and a pre-compiled database of bacterial taxa already identified in vaginal microbiota.(4) Based on the dominant bacterial taxa and their relative abundance, we stratified samples into five community state types (CSTs I to V) as described by Ravel et al.(4) using Jensen-Shannon divergence metrics. (4, 21, 22) Four CSTs were *Lactobacillus*-dominated: CST I by *L. crispatus*, CST II by *L. gasseri*, CST III by *L. iners*, and CST V by *L. jensenii*. CST IV was characterized by the lack of significant taxa dominance and a great microbial diversity. We defined bacterial richness as the number of taxa sequenced in each sample, and alpha-diversity as the Shannon-Weiner index using the *vegan* package in R.(23)

HPV genotyping and detection of other infections

For molecular testing, we extracted DNA from cervical brush samples using *AmpliLute Liquid Media Extraction* kit (Roche Molecular Systems, Inc., Pleasanton, CA, USA). We performed HPV genotyping using the *Linear Array HPV genotyping tests* kit (Roche Molecular Systems, Pleasanton, CA), according to manufacturer's instructions. This method is based on the amplification of target DNA by using HPV primers to the L1 gene followed by hybridization of the amplified products to oligonucleotide probes. This technique allows the identification of 37 anogenital HPV genotypes, including low-risk HPV types (lrHPVs) and hrHPVs. Additionally, we tested cervical samples for *C. trachomatis* and *N. gonorrhoeae* infection by PCR, as previously described.(24, 25) To assess *T. vaginalis* status, we performed PCRs using primers set TVA5 (5'-GATCATGTTCTATCTTTTCA-3') and TVA6 (5'-GATCACCACCTTAGTTTACA-3') (26) at 10 μM (0.25 μL volume each), added to 12.5 μL GoTaq Green Master Mix (Promega, Madison, WI) and 2.0 μL DNA template. We used a 40 cycles reaction of 94ºC for 1 minute, followed by 52ºC for 1 minute, and by 72ºC for another minute. We submitted PCR reactions to electrophoresis and considered as positive samples those with a 102 bp-sized amplicon observed through ultraviolet transilluminator.

Statistical analysis

We compared sociodemographic, behavioral and clinical characteristics of the study population as well HPV infection status across enrollment regions; categorical variables were compared using Pearson's chi-square test and continuous variables were compared using the Kruskal-Wallis test. We also examined the distribution of HPV infections across CSTs.

We used logistic regression analyses to estimate age- and region- adjusted and multivariable adjusted odds ratios (OR) and their 95% confidence intervals (CI) for the association between participants' characteristics and HPV infection. We fitted two stepwise multivariable logistic regression models to assess the association between each of (1) microbiome-derived taxa and (2) sociodemographic, behavioral, and clinical characteristics with the occurrence of high-risk HPV infection. We defined these variables for the second logistic regression as these are known to be associated to HPV infection in some extent, and as a way of comparison to evaluate the association between microbiome component and hrHPV infection. Potential a priori covariates were retained if they were significant at the 0.15 level. Using the regression coefficients, we constructed two corresponding linear scores for predicting hrHPV positivity: (1) a microbiome-based (MB) score and (2) a score based on sociodemographic, behavioral, and clinical (SBC) characteristics. We constructed a receiver operating curve (ROC) to compare the performance of both scores in predicting hrHPV infection by calculating for MB and SBC scores their area under the curve (AUC) and 95% CI. Considering the potential source of bias by South region differential recruitment setting, statistical analyses were conducted repeatedly excluding this center in order to check the consistency of the results. All analyses were performed using Stata/SE (version 15.1, StataCorp, College Station, TX); P-values <0.05 were considered significant.

3. Results

Study population: overall and by enrollment centers

Of the 609 enrolled participants, 63 women (10.3%) were excluded (37 had insufficient cervical samples and 26 tested negative for beta-globin internal control of DNA extraction), resulting in 546 participants included in the current analysis. Table 1 shows their sociodemographic,

behavioral and clinical characteristics, overall and by recruitment region. The overall median age was 34 years (range 18-51). More than half identified themselves as Black/Pardo (55.7%), had at least completed secondary education (>11 years, 57.7%), and were living with a partner at enrollment (64.4%). A small proportion were smokers (11.2%), had two or more sex partners during the previous year (10.8%), and a new sex partner 2 months before enrollment (7.7%). A total of 257 women (47.1%) had been treated at least once for abnormal discharge with antimicrobial drugs. Current abnormal discharge was reported by 44.7% of participants and more than half (52%) had abnormal vaginal pH (> 4.5). Cervical infections caused by *C. trachomatis, N. gonorrhoeae* or *T. vaginalis* were detected in 30 (5.5%) participants. The overall prevalence of microscopic (Nugent) BV was 27.5% (n=150), which significantly differed by region, with participants from the North showing the highest prevalence (37.5%). In regard to 16S rRNA sequencing findings, CST I (n=166, 30.4%), CST III (n=204, 37.4%) and CST IV (n=143, 26.2%) were the most prevalent, while CST V was only found in 1.1% (n=6) of the population. The prevalence of CSTs also differed according to region. *Lactobacillus iners*-dominated CST III was the most prevalent in all regions, except for the North where CST IV was the most frequent type. The South region showed the lowest prevalence of CST IV (7.3%) and was the only region where CST II (*L. gasseri*-dominated) had a fairly expressive prevalence (12.6%, n=12).

Cervical HPV infection: oncogenic risk, species and genotypes

As displayed in Table 2, the overall positivity for any HPV infection was 21.4% (n=117), while prevalence of lrHPV and hrHPV were 10.8% (n=59) and 15.8% (n=86), respectively. The prevalence of any HPV and lrHPV was significantly higher in the South and Central regions compared to other regions ($p<0.05$). HPV species 9 was the most prevalent (9.9%) amongst the 10 HPV species identified in the study. The prevalence of HPV species did not differ by region.

Among the 32 HPV genotypes detected, the most prevalent was HPV16 (5.1%), followed by HPV6 (3.1%) and HPV52 (2.4%). We observed that only the prevalence of HPVs 6, 11 and 16 differed across regions ($p<0.0001$).

When comparing HPV prevalence rates stratified by microbiome-defined CSTs, no significant difference was observed (Table 3). However, at the species level, we observed that the prevalence of HPV species 1 and 10 differed across CSTs ($p<0.05$).

Association between population characteristics associated with HPV infection

The associations between characteristics of the study population and HPV infection are presented in Table 4. Women aged between 26 and 39 years (fully adjusted OR=0.45, 95% CI: 0.26-0.77) and those who were living with a partner upon enrollment (fully adjusted OR= 0.54, 95% CI: 0.35-0.84) had a reduced risk of any HPV infection. Women who reported having a new sex partner two months prior to enrollment were more likely to have hrHPV (OR= 2.23, 95% CI: 1.06-4.66), independently of other factors. The use of injectable hormone was protective against HPV (fully adjusted OR=0.21, 95% CI: 0.06-0.72) and hrHPV (fully adjusted OR=0.08, 95% CI: 0.01-0.63), while no association was found with oral contraceptives. Condom use was positively associated with having any HPV infection in age- and region- analysis (OR=1.58, 95% CI: 1.01-2.50); no associations were observed with hrHPV. Bacterial vaginosis using Nugent microscopic classification was associated with any HPV (OR= 1.89, 95% CI:1.15-3.03) and hrHPV (OR= 1.90, 95% CI: 1.13-3.31) in age- and region- adjusted models but were not retained in multivariable analyses. Women with CST IV were at an increased risk of having any HPV infection (age- and region- adjusted OR= 1.96, 95% CI: 1.11-3.48). No association was found between CSTs and hrHPV positivity.

Vaginal microbiome components as indicators of hrHPV

Supplementary Table 1 presents the stepwise multivariate logistic regression coefficients of the MB score used to predict cervical hrHPV. Out of 116 bacterial taxa identified in the study, 30 were retained in the model for hrHPV prediction. Using a similar statistical approach, Supplementary Table 2 presents the coefficients of six (age, marital status, new sex partner, hormonal contraceptive use, body mass index and smoking) out of 25 candidate SBC characteristics. Figure 1 plots the ROC curves for the MB and SBC for predicting hrHPV positivity. The MB score performed better (AUC=0.8022, 95% CI: 0.7517-0.8527) than the SBC score (AUC=0.7027, 95% CI: 0.6419 – 0.7636) for predicting hrHPV infection ($p<0.001$), with no overlap in the 95% CIs. Results remain the same when excluding South region as a potential source of bias.

4. Discussion

We demonstrate the strong correlation between vaginal microbiome components and hrHPV infection, whereby proposing the utility of vaginal microbiome components, here demonstrated by MB score, as an indicator of cervical hrHPV infection.

Until now, studies have demonstrated the protective role played by some *Lactobacillus* species against HPV infection in the vaginal environment. Within the protective factors mediated by the microbiota, the maintenance of an acidic environment, especially through the production of lactic acid and hydrogen peroxide, as well as bacteriocins and biosurfactant by some *Lactobacillus* species help the provision of a protective mucosal barrier and epithelial integrity. (21, 27)

In contrast, vaginal dysbiosis are known to lack the protective factors mediated by Lactobacillus, as well as to be associated with disruption of the integrity of the mucosal barrier, a chronic inflammatory context and oxidative stress. (28) Highly diverse vaginal microbiome communities

that lack *Lactobacillus* dominance, especially CST IV, have been linked to HPV infection persistence and increased CIN severity. (21, 28)

Despite the described association between vaginal dysbiosis, HPV infection and CIN progression, an idea of an association between certain bacterial species and HPV infection, especially hrHPV, has been brought recently.(29)

Among the 30 discriminative taxa retained in the model for its association with hrHPV positivity, some have been already acknowledged as vaginal colonizers associated with disrupted microbiota, such as *Prevotella* spp., *Dialister* spp., *Megasphaera* spp., *Leptotrichia* spp., Bacterial Vaginosis-Associated Bacteria-3 (BVAB3), *Mycoplasma genitalium* and others. (21, 30, 31)

In addition, it is remarkable that some of the bacterial taxa that compose the MB score have been directly linked to HPV infection in recent studies. *Prevotella* spp. *and Leptotrichia* spp. (30) have been associated with hrHPV infection. *Prevotella* spp., a microbe with a close symbiotic metabolism relation to *Gardnerella vaginalis*, is well known for its potential to disrupt the vaginal immunity. (33, 34) Along with *Gardnerella* spp., increased baseline abundance of *Prevotella* spp. was associated with an persistence of hrHPV infection. (12)

Recently, Chen et al (35) found *Bifidobacterium* spp., *Bacillus* spp., *Megasphaera* spp., *Snethia* spp., *Prevotella* spp., *Gardnerella* spp., *Fastidiospila* spp. and *Dialister* spp. to be biomarkers for hrHPV infection in non-pregnant women in Chinese cohorts.

At the same time, *Gardnerella vaginalis* and *Atopobium vaginae* have been proposed as potential molecular markers, especially by its contribution in biofilm formation and its association to viral persistence.(12, 21, 36-38)

As a whole, an idea that hrHPV infection might be linked to specific agents, regardless of its abundance has been brought recently. Huang et al demonstrates that it is not a diverse or common

vaginal microbiota itself, but rather the presence of certain microbiome components that influence the acquisition of hrHPV genotypes. (29)

Still in that context, a recent study that characterizes the vaginal metabolome of HPV positive and negative women demonstrates that, the vaginal metabolome can not be clustered based on HPV status, but it can be, in general, characterized based on the bacterial composition. Considering that, differential metabolome profiles can be identified between HPV positive and negative women within the same CST, especially regarding to oxidative stress metabolites.(31)

All together, these findings demonstrate the emerging understanding of the interplay between HPV infection and certain vaginal microbiome components. Thus, MB score itself, proposed by this study, highlights the association between those entities.

We also used *SBC score* as a way to compare the performance of our newly developed *MB score* and to evaluate the association of microbiome components to hrHPV infection. Concerning that comparison, retained variables from the multivariate logistic regression model that composed *SBC* score must be pointed out. Those variables are *a priori* cofounders for HPV infection that reflect the only information that may be available for clinicians during cervical cancer screening practice. Among the six retained variables that composed SBC score, five (age, living with partner, new sex partner, use of hormonal contraceptive, and smoking habits) are variables that have already been linked to HPV infection in some extent (39) and might be taken into account as relevant information during clinical assessment. The remaining variable, BMI, is the only one that has not been associated with HPV infection. (40) The retained SBC variables themselves corroborate the previous clinical understanding associated with HPV infection.

Altogether, *MB score* highlights the strong association between vaginal microbiome components and HPV infection compared to well-understood clinical variables presented by *SBC score*.

Regarding to the characterization of our population, the overall prevalence of BV was 27.5% in our population, in line with the 30% average reported in the literature.(41) The prevalence of BV is known to vary across different populations; in our study population it varied considerably from 12.6% (South) to 37.5% (North). Among vaginal microbiome characterization, nearly 75% of participants had a *Lactobacillus*-dominated CSTs, while the remainder had CST IV, similar to the general prevalence of BV. CST IV varied across region with lower and higher rates in the South (7.3%) and North (35.2%), respectively.

The prevalence of hrHPV (15.8%) was very similar to the previously reported prevalence (14.9%) in a Brazilian study that enrolled more than 16,000 women attending primary healthcare clinics for cervical cancer screening.(42)

Women with ages between 26 and 39 years and those who reported to cohabitate with a partner (including those married) were less likely to be HPV-positive. Older age and being in a marital relationship were negatively associated with cervical HPV. (43, 44) We found a protective effect of injectable hormones on HPV cervical infection, contrary to a previous Brazilian-cohort study. (45) Condom use was positively associated with any HPV positivity, but the literature is controversial on this association.(45, 46) One possible explanation for our findings could be that participants in more steady relationships would opt for injections, while others would be more prone to use condoms. It is also relevant the limitations in evaluating the consistency of condom use in this type of questionnaire.

Several sexual behaviors such as age at coitarche, number of lifetime sex partners, and having recent casual sex, have been shown to increase the risk of cervical HPV infection.(43) In our population, participants who had a new sex partner in the 2 months prior to study enrolment were

more likely to be infected with any HPV and/or hrHPV. We did not observe an association with the number of sex partners in the past year or the frequency of sexual intercourse.

We found an association between BV and HPV in bivariate analyses in agreement with previous reports.(6-8) Similarly, CST IV was positively associated with any HPV infection. Data from previous studies indicate an association between *Lactobacillus*-depleted microbiome and HPV persistence.(30, 47, 48) Also, presence of certain *Lactobacillus* spp., such as *L. crispatus*, may be more beneficial than others in clearing HPV infections.(14)

The emerging understanding of the association between certain bacterial species and HPV infection is a promising field in the identification of biological markers, especially for its potential in clinical application. The identification of women with certain microbial compositions can lead to differential approaches and management from a clinical perspective of HPV infection. Additionally, the understanding of the role played by some vaginal microbiome components can translate into some auxiliary tools in gynecological practice in the follow-up of hrHPV positive patients.

Although causality has not been determined between vaginal microbiome components, HPV infection, persistence and cervical cancer progression, it must be pointed that the microbiota is a potentially modifiable risk factor. (28)

From that perspective, it is relevant to highlight some of the several factors that influence the composition of the microbiota and that impose some limitations to the complete understanding of the microbiota – HPV interplay.

Some of the factors inherent to the host are: age; hormonal status; sociodemographic factors; behavioral and sexual factors; ethnicity, as well as associated genetic factors; presence of coinfections, among others. At the same time, the inherent heterogeneity associated with the

microbiota, as well as differential classifications between studies are other cofounding factors that permeate the complexity of this understanding. (21, 28)

From that perspective, the methodologies used in this study, in line with the reference studies in this field; the population size as well as its extent; the requirement of reproductive-age women; as well as the characterization of our population are all strengths that highlight the impact of our findings.

Despite some of the intrinsic limitations in this field of study, studies corroborate the role played by vaginal microbiota in the progression of infection and cervical cancer development. In this regard, it is important to highlight the role of proper vaginal dysbiosis diagnosis and management in favor of establishing a healthy microbiota as an associated intervention tool in the burden of HPV-associated changes in the cervix in gynecological practice.

Vaginal dysbiosis are still a matter of studies and proper diagnosis, considering its role as a risk factor in the epidemiology of HPV. The appropriate management of dysbiosis, associated with vaccination strategies would possibly contribute significantly to the reduction of cervical cancer and HPV-related diseases. Additionally, the complete understanding of the role played by some species in the vaginal environment can translate into auxiliary tools for cancer development risk assessment for HPV positive women in gynecological practice.

Altogether the identification of the strong correlation between vaginal microbiome and hrHPV infection, shown by *MB score*, contributes to futures insights in this field of study, especially through the demonstrating the relevance of vaginal microbiome components compared to sociodemographic, clinical and behavioral characteristics. These results might also contribute to the better understanding of the interplay between vaginal microbiome and HPV infection. Further

validation of *MB score* clinical utility as a predictor tool must be confirmed via longitudinal studies.

5. Conclusion

Our findings demonstrate the strong correlation between vaginal microbiome components and hrHPV infection, whereby proposing the utility of vaginal microbiome components, here demonstrated by *MB score*, as an indicator of hrHPV infection. It warrants further validation of its clinical utility as a predictor via longitudinal studies.

Funding

This work was supported by Sao Paulo Research Foundation (Fapesp), Grants: 2012/01278-0, 2012/16800-3, 2012/10403-2, 2018/18469-9.

6. References

1. Fettweis JM, Serrano MG, Brooks JP, Edwards DJ, Girerd PH, Parikh HI, et al. The vaginal microbiome and preterm birth. Nat Med. 2019;25(6):1012-21.
2. Thomas AM, Manghi P, Asnicar F, Pasolli E, Armanini F, Zolfo M, et al. Author Correction: Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. Nat Med. 2019.
3. Mirmonsef P, Hotton AL, Gilbert D, Burgad D, Landay A, Weber KM, et al. Free glycogen in vaginal fluids is associated with Lactobacillus colonization and low vaginal pH. PLoS One. 2014;9(7):e102467.
4. Ravel J, Gajer P, Abdo Z, Schneider GM, Koenig SS, McCulle SL, et al. Vaginal microbiome of reproductive-age women. Proc Natl Acad Sci U S A. 2011;108 Suppl 1:4680-7.
5. 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;22(3):213-20.
6. 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;2011:319460.
7. Watts DH, Fazzari M, Fazarri M, Minkoff H, Hillier SL, Sha B, 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;191(7):1129-39.

8. Moscicki AB, 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;19(8):2055-65.
9. Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Piñeros M, et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. Int J Cancer. 2019;144(8):1941-53.
10. Engberts MK, Verbruggen BS, Boon ME, van Haaften M, Heintz AP. Candida and dysbacteriosis: a cytologic, population-based study of 100,605 asymptomatic women concerning cervical carcinogenesis. Cancer. 2007;111(5):269-74.
11. Denslow SA, Westreich DJ, Firnhaber C, Michelow P, Williams S, Smith JS. Bacterial vaginosis as a risk factor for high-grade cervical lesions and cancer in HIV-seropositive women. Int J Gynaecol Obstet. 2011;114(3):273-7.
12. 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;7(1):10200.
13. 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;214(9):1361-9.
14. 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;210(11):1723-33.
15. 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;29(2):297-301.
16. Fadrosh DW, Ma B, Gajer P, Sengamalay N, Ott S, Brotman RM, et al. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. Microbiome. 2014;2(1):6.
17. Magoč T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics. 2011;27(21):2957-63.
18. Kuczynski J, Stombaugh J, Walters WA, González A, Caporaso JG, Knight R. Using QIIME to analyze 16S rRNA gene sequences from microbial communities. Curr Protoc Bioinformatics. 2011;Chapter 10:Unit 10.7.
19. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics. 2011;27(16):2194-200.
20. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, et al. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J. 2012;6(3):610-8.
21. 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. 2019.
22. Gajer P, Brotman RM, Bai G, Sakamoto J, Schütte UM, Zhong X, et al. Temporal dynamics of the human vaginal microbiota. Sci Transl Med. 2012;4(132):132ra52.
23. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, et al. Vegan: Community Ecology Package. 2013.

24. Ferreira CST, Marconi C, Parada CMdLG, Duarte MTC, Gonçalves APO, Rudge MVC, et al. Bacterial vaginosis in pregnant adolescents: proinflammatory cytokine and bacterial sialidase profile. Cross-sectional study. Sao Paulo Medical Journal. 2015;133:465-70.
25. Marconi C, Santos-Greatti MM, Parada CM, Pontes A, Pontes AG, Giraldo PC, et al. Cervicovaginal levels of proinflammatory cytokines are increased during chlamydial infection in bacterial vaginosis but not in lactobacilli-dominated flora. J Low Genit Tract Dis. 2014;18(3):261-5.
26. Riley DE, Roberts MC, Takayama T, Krieger JN. Development of a polymerase chain reaction-based diagnosis of Trichomonas vaginalis. J Clin Microbiol. 1992;30(2):465-72.
27. Graver MA, Wade JJ. The role of acidification in the inhibition of Neisseria gonorrhoeae by vaginal lactobacilli during anaerobic growth. Ann Clin Microbiol Antimicrob. 2011;10:8.
28. 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;221(1):9-18.e8.
29. Huang X, Li C, Li F, Zhao J, Wan X, Wang K. Cervicovaginal microbiota composition correlates with the acquisition of high-risk human papillomavirus types. Int J Cancer. 2018;143(3):621-34.
30. Dareng EO, Ma B, Famooto AO, Adebamowo SN, Offiong RA, Olaniyan O, et al. Prevalent high-risk HPV infection and vaginal microbiota in Nigerian women. Epidemiol Infect. 2016;144(1):123-37.
31. Borgogna JC, Shardell MD, Santori EK, Nelson TM, Rath JM, Glover ED, et al. The vaginal metabolome and microbiota of cervical HPV-positive and HPV-negative women: a cross-sectional analysis. BJOG. 2020;127(2):182-92.
32. Lee JE, Lee S, Lee H, Song YM, Lee K, Han MJ, et al. Association of the vaginal microbiota with human papillomavirus infection in a Korean twin cohort. PLoS One. 2013;8(5):e63514.
33. Pybus V, Onderdonk AB. Evidence for a commensal, symbiotic relationship between Gardnerella vaginalis and Prevotella bivia involving ammonia: potential significance for bacterial vaginosis. J Infect Dis. 1997;175(2):406-13.
34. Lewis WG, Robinson LS, Perry J, Bick JL, Peipert JF, Allsworth JE, et al. Hydrolysis of secreted sialoglycoprotein immunoglobulin A (IgA) in ex vivo and biochemical models of bacterial vaginosis. J Biol Chem. 2012;287(3):2079-89.
35. Chen Y, Hong Z, Wang W, Gu L, Gao H, Qiu L, et al. Association between the vaginal microbiome and high-risk human papillomavirus infection in pregnant Chinese women. BMC Infect Dis. 2019;19(1):677.
36. Machado A, Cerca N. Influence of Biofilm Formation by Gardnerella vaginalis and Other Anaerobes on Bacterial Vaginosis. J Infect Dis. 2015;212(12):1856-61.
37. Swidsinski A, Mendling W, Loening-Baucke V, Ladhoff A, Swidsinski S, Hale LP, et al. Adherent biofilms in bacterial vaginosis. Obstet Gynecol. 2005;106(5 Pt 1):1013-23.
38. Swidsinski A, Loening-Baucke V, Mendling W, Dörffel Y, Schilling J, Halwani Z, et al. Infection through structured polymicrobial Gardnerella biofilms (StPM-GB). Histol Histopathol. 2014;29(5):567-87.
39. Chelimo C, Wouldes TA, Cameron LD, Elwood JM. Risk factors for and prevention of human papillomaviruses (HPV), genital warts and cervical cancer. J Infect. 2013;66(3):207-17.
40. Wee CC, Huang A, Huskey KW, McCarthy EP. Obesity and the likelihood of sexual behavioral risk factors for HPV and cervical cancer. Obesity (Silver Spring). 2008;16(11):2552-5.

41. Marconi C, Duarte MT, Silva DC, Silva MG. Prevalence of and risk factors for bacterial vaginosis among women of reproductive age attending cervical screening in southeastern Brazil. Int J Gynaecol Obstet. 2015;131(2):137-41.
42. Levi JE, Martins TR, Longatto-Filho A, Cohen DD, Cury L, Fuza LM, et al. High-Risk HPV Testing in Primary Screening for Cervical Cancer in the Public Health System, São Paulo, Brazil. Cancer Prev Res (Phila). 2019;12(8):539-46.
43. Burk RD, Ho GY, Beardsley L, Lempa M, Peters M, Bierman R. Sexual behavior and partner characteristics are the predominant risk factors for genital human papillomavirus infection in young women. J Infect Dis. 1996;174(4):679-89.
44. Smith JS, Melendy A, Rana RK, Pimenta JM. Age-specific prevalence of infection with human papillomavirus in females: a global review. J Adolesc Health. 2008;43(4 Suppl):S5-25, S.e1-41.
45. Shaw E, Ramanakumar AV, El-Zein M, Silva FR, Galan L, Baggio ML, et al. Reproductive and genital health and risk of cervical human papillomavirus infection: results from the Ludwig-McGill cohort study. BMC Infect Dis. 2016;16:116.
46. Manhart LE, Koutsky LA. Do condoms prevent genital HPV infection, external genital warts, or cervical neoplasia? A meta-analysis. Sex Transm Dis. 2002;29(11):725-35.
47. Mitra A, MacIntyre DA, Lee YS, Smith A, Marchesi JR, Lehne B, et al. Cervical intraepithelial neoplasia disease progression is associated with increased vaginal microbiome diversity. Sci Rep. 2015;5:16865.
48. Audirac-Chalifour A, Torres-Poveda K, Bahena-Román M, Téllez-Sosa J, Martínez-Barnetche J, Cortina-Ceballos B, et al. Cervical Microbiome and Cytokine Profile at Various Stages of Cervical Cancer: A Pilot Study. PLoS One. 2016;11(4):e0153274.

Table 1. Sociodemographic, behavioral and clinical characteristics of the population, overall and by region.

	Overall (n= 546)	South (n= 95)	Southeast (n= 109)	Central (n= 111)	Northeast (n= 103)	North (n= 128)	P-value[a]
Age, years							0.66
Median (range)	34 (18-51)	35 (18-50)	33 (18-51)	35 (18-50)	34 (18-49)	34 (19-49)	
Age, years							0.07
<25	109 (20.0)	22 (23.2)	31 (28.4)	21 (18.9)	15 (14.6)	20 (15.6)	
26-39	278 (50.9)	42 (44.2)	43 (39.5)	59 (53.2)	61 (59.2)	73 (57.0)	
≥40	159 (29.1)	31 (32.6)	55 (32.1)	31 (27.9)	27 (26.2)	35 (27.4)	
Ethnicity							<0.0001
Black/pardo	304 (55.7)	24 (25.3)	42 (38.5)	70 (63.1)	67 (65.1)	101 (78.9)	
White	219 (40.1)	71 (74.7)	65 (59.6)	30 (27.0)	33 (32.0)	20 (15.6)	
Other	23 (4.2)	0 (0.0)	32(1.8)	11 (9.9)	3 (2.9)	7 (5.5)	
Living with partner							0.01
No	194 (35.6)	39 (41.1)	40 (36.7)	51 (46.0)	27 (26.2)	37 (28.7)	
Yes	352 (64.4)	56 (58.9)	69 (63.3)	60 (54.0)	76 (73.8)	91 (71.1)	
School, years completed							0.0001
Median (range)	11 (0-20)	11 (2-20)	9 (0-16)	11 (0-17)	9 (0-15)	11 (2-20)	
Schooling, years completed							<0.0001
≤11	231 (42.3)	29 (29.5)	68 (62.4)	32 (28.8)	54 (52.4)	79 (38.7)	
>11	315 (57.7)	71 (70.5)	41 (37.6)	79 (71.2)	49 (47.6)	79 (61.3)	
Has personal income?							0.005

	Overall (n= 546)	South (n= 95)	Southeast (n= 109)	Central (n= 111)	Northeast (n= 103)	North (n= 128)	P-value[a]
No	193 (35.6)	24 (25.3)	43 (39.4)	28 (25.2)	44 (42.7)	55 (42.2)	
Yes	353 (64.4)	71 (74.7)	66 (60.6)	83 (74.8)	59 (57.3)	74 (57.8)	
Body mass index							0.05
Median (range)	25.0 (14.7-53.0)	24.3 (16.2-40.3)	26.1 (17.6-53.0)	24.7 (17.8-41.5)	25.6 (16.9-41.1)	24.6 (14.7-39.2)	
Body mass index[b]							0.30
Underweight/Normal	275 (50.4)	53 (55.8)	47 (43.2)	58 (52.3)	46 (44.7)	71 (55.5)	
Overweight	177 (32.4)	29 (30.5)	37 (33.9)	34 (30.6)	35 (34.0)	42 (32.8)	
Obese	94 (17.2)	13 (13.7)	25 (22.9)	19 (17.1)	22 (21.3)	15 (11.7)	
Smoking							0.006
No	485 (88.8)	83 (87.4)	87 (79.8)	105 (94.6)	92 (89.3)	118 (92.2)	
Yes	61 (11.2)	12 (12.6)	22 (20.2)	6 (5.4)	11 (10.7)	10 (7.8)	
Alcohol							0.007
No	388 (71.1)	60 (63.2)	90 (82.6)	81 (73.0)	77 (74.8)	80 (62.5)	
Yes	158 (28.9)	35 (36.8)	24 (17.4)	30 (27.0)	26 (25.2)	48 (38.5)	
Intimate soap							<0.0001
No	273 (50.0)	46 (48.2)	80 (73.4)	46 (41.4)	59 (57.3)	42 (32.8)	
Yes	273 (50.0)	49 (51.6)	29 (26.6)	65 (58.6)	44 (42.7)	86 (67.2)	
Douching							0.13
No	466 (85.3)	78 (82.1)	196 (88.1)	97 (87.4)	93 (90.3)	102 (79.7)	
Yes	80 (14.7)	17 (18.9)	13 (11.9)	14 (12.6)	10 (9.7)	26 (20.3)	
Sitz-bathing							<0.0001
No	500 (91.6)	93 (98.9)	109 (100.0)	111 (100.0)	92 (89.3)	96 (74.2)	
Yes	46 (8.4)	2 (2.1)	0 (0.0)	0 (0.0)	11 (10.7)	33 (25.8)	
Number of sex partners[c]							0.29
None	29 (5.3)	8 (8.4)	3 (2.4)	7 (6.3)	6 (5.8)	5 (3.9)	
1	458 (83.9)	85 (86.3)	95 (87.2)	92 (82.9)	86 (83.5)	104 (80.5)	
2+	59 (10.8)	5 (5.2)	11 (10.4)	12 (10.8)	11 (10.7)	20 (15.6)	
New sex partner[d]							<0.0001
No	504 (92.3)	83 (84.4)	108 (97.3)	108 (97.3)	102 (99.0)	119 (93.0)	
Yes	42 (7.7)	13 (12.6)	17 (2.7)	3 (2.7)	1 (1.0)	9 (7.0)	
Sexual intercourse/week							0.002
0	96 (17.6)	26 (27.4)	15 (13.8)	22 (19.8)	13 (12.6)	20 (15.6)	
1-2	242 (44.3)	49 (51.6)	42 (28.5)	44 (39.6)	55 (53.4)	52 (40.7)	
3+	208 (38.1)	20 (21.0)	52 (47.7)	45 (40.5)	35 (34.0)	56 (43.8)	
Hormonal contraceptive							<0.0001
None	337 (61.7)	40 (42.1)	61 (56.0)	69 (62.1)	62 (60.2)	105 (82.0)	
Oral	166 (30.4)	49 (51.6)	38 (34.5)	36 (32.4)	28 (27.2)	15 (11.7)	
Injectable	43 (7.9)	6 (6.3)	10 (9.2)	6 (5.4)	13 (12.6)	8 (6.3)	
Condom use							0.002
No	371 (68.0)	66 (69.5)	81 (74.3)	63 (56.8)	82 (79.6)	79 (61.7)	
Yes	175 (32.0)	29 (30.5)	28 (25.7)	48 (44.2)	21 (20.4)	50 (38.3)	
Ever pregnant							<0.0001
No	120 (22.0)	36 (37.9)	28 (25.7)	21 (18.9)	13 (12.6)	22 (17.2)	
Yes	426 (78.0)	59 (62.1)	81 (74.3)	90 (81.1)	90 (87.4)	107 (82.8)	
Menstrual cycle phase							0.003
Follicular	237 (43.4)	30 (31.6)	50 (45.9)	48 (43.2)	52 (50.5)	57 (4.5)	
Luteal	255 (46.7)	54 (56.8)	51 (46.8)	42 (37.8)	43 (41.7)	65 (50.8)	
Suppressed[e]	54 (9.9)	11 (11.6)	8 (7.3)	21 (18.9)	8 (7.8)	6 (4.7)	
Dyspareunia							0.61
No	348 (63.7)	59 (62.1)	68 (62.4)	78 (70.3)	65 (63.1)	78 (60.9)	

	Overall (n= 546)	South (n= 95)	Southeast (n= 109)	Central (n= 111)	Northeast (n= 103)	North (n= 128)	P-value[a]
Yes	198 (36.3)	36 (37.9)	41 (37.6)	33 (29.7)	38 (36.9)	50 (39.1)	
Previously treated abnormal discharge							<0.0001
No	289 (52.9)	43 (45.3)	59 (54.1)	92 (82.9)	48 (46.6)	47 (36.7)	
Yes	257 (47.1)	52 (54.7)	50 (45.9)	19 (17.1)	55 (53.4)	81 (63.3)	
Abnormal discharge							0.002
No	302 (55.3)	66 (69.5)	57 (52.3)	56 (48.7)	63 (61.2)	62 (48.4)	
Yes	244 (44.7)	29 (30.5)	52 (47.7)	57 (51.4)	40 (38.8)	66 (51.6)	
Vaginal pH							0.007
Median (range)	4.7 (4.0-6.1)	4.4 (4.0-6.1)	4.7 (4.0-5.8)	4.4 (4.0-5.5)	4.4 (4.0-5.5)	4.4 (4.0-5.5)	
Vaginal pH							<0.0001
Normal (≤4.5)	262 (48.0)	53 (55.8)	30 (27.5)	60 (54.0)	52 (50.5)	67 (52.3)	
Abnormal (>4.5)	284 (52.0)	42 (44.2)	79 (72.5)	51 (46.0)	51 (49.5)	61 (47.7)	
Vulvitis							0.003
No	528 (96.7)	86 (90.5)	105 (96.3)	110 (99.1)	100 (97.1)	127 (99.2)	
Yes	18 (3.3)	9 (9.5)	4 (3.7)	1 (0.9)	3 (2.9)	1 (0.8)	
Signs of cervicitis[f]							0.001
No	495 (90.7)	89 (93.7)	103 (94.5)	110 (99.1)	81 (78.6)	112 (87.5)	
Yes	51 (9.3)	6 (6.3)	6 (5.5)	1 (0.9)	22 (21.4)	16 (12.5)	
Number of inflammatory cells[g]							<0.0001
<5/field (x1000)	512 (93.8)	93 (97.9)	99 (90.8)	109 (98.2)	99 (96.1)	112 (87.5)	
≥5/field (x1000)	34 (6.2)	2 (2.1)	10 (9.2)	10 (1.8)	4 (3.9)	16 (12.5)	
Positive for CT, NG or TV							0.17
No	516 (94.5)	94 (99.9)	102 (93.6)	106 (95.5)	97 (94.2)	117 (91.4)	
Yes	30 (5.5)	1 (1.1)	7 (6.4)	5 (4.5)	6 (5.8)	11 (8.6)	
Microscopic status of microbiota[h]							0.002
Normal	341 (62.4)	74 (77.9)	61 (56.0)	75 (67.6)	66 (64.1)	65 (50.8)	
Intermediate	55 (10.1)	9 (9.5)	15 (13.8)	6 (5.4)	10 (9.7)	15 (11.7)	
Bacterial vaginosis	150 (27.5)	12 (12.6)	33 (30.2)	30 (27.0)	27 (26.2)	48 (37.5)	
Vaginal microbiome							<0.0001
CST I	166 (30.4)	37 (39.0)	29 (26.6)	31 (27.9)	25 (24.3)	44 (34.4)	
CST II	27 (4.9)	12 (12.6)	4 (3.7)	3 (2.7)	6 (5.8)	2 (1.5)	
CST III	204 (37.4)	36 (37.9)	44 (40.4)	42 (37.8)	45 (43.7)	37 (28.9)	
CST IV	143 (26.2)	7 (7.3)	31 (28.4)	35 (31.5)	25 (24.3)	45 (35.2)	
CST V	6 (1.1)	3 (3.2)	1 (0.9)	0 (0.0)	1 (1.9)	0 (0.0)	

[a]Categorical and continuous variables were compared, respectively, by chi-squared and Kruskal-Wallis tests among regions, with P<0.05 considered as significant; [b]BMI classification according to World Health Organization; [c]12 months prior enrollment; [d]2 months prior to enrollment; [e]Continuous hormonal contraceptive use; [f]Bleeding or mucopurulent secretion [g]Cell count on Gram-stained vaginal smears overserved under immersion oil and (100x) objective lens, [h]Classification of Gram-stained smears according to Nugent score (0-3: normal, 4-6: intermediate, 7-10: bacterial vaginosis)
CT: *Chlamydia trachomatis*; NG: *Neisseria gonorrhoeae*; TV: *Trichomonas vaginalis*, CST: community-state type.

Table 2. Positivity for any HPV infection, low- and high- oncogenic risk HPV, and species, overall and by region.

	Overall (n= 546)	**South (n= 95)**	**Southeast (n= 109)**	**Central (n= 111)**	**Northeast (n= 103)**	**North (n= 128)**	**P-value**
Any HPV infection	117 (21.4)	30 (31.6)	17 (15.6)	31 (27.9)	20 (19.4)	19 (14.8)	0.006
HPV oncogenic risk							
Low-risk HPV	59 (10.8)	17 (17.9)	8 (7.3)	16 (14.4)	10 (9.7)	8 (6.3)	0.03
High-risk HPV	86 (15.8)	21 (22.1)	14 (12.8)	23 (20.7)	13 (12.6)	15 (11.7)	0.09
HPV species							
Species 1	2 (0.4)	1 (1.0)	0 (0.0)	0 (0.0)	1 (1.0)	0 (0.0)	0.47
Species 3	27 (5.0)	4 (4.2)	7 (6.4)	6 (5.4)	4 (3.9)	6 (4.7)	0.92
Species 5	14 (2.6)	3 (3.2)	3 (2.8)	6 (5.4)	2 (1.9)	0 (0.0)	0.12
Species 6	17 (3.1)	4 (4.2)	2 (1.8)	6 (5.4)	4 (3.9)	1 (0.8)	0.25
Species 7	22 (4.0)	4 (4.2)	4 (3.7)	6 (5.4)	3 (2.9)	5 (3.9)	0.92
Species 9	54 (9.9)	17 (17.9)	10 (9.2)	9 (8.1)	8 (7.8)	10 (7.8)	0.08
Species 10	26 (4.8)	13 (13.7)	2 (1.8)	9 (8.1)	2 (1.9)	0 (0.0)	<0.001

	Overall (n= 546)	South (n= 95)	Southeast (n= 109)	Central (n= 111)	Northeast (n= 103)	North (n= 128)	P-value
Species 11	3 (0.6)	0 (0.0)	1 (0.9)	1 (0.9)	1 (1.0)	0 (0.0)	0.72
Species 13	5 (0.9)	0 (0.0)	1 (0.9)	1 (0.9)	2 (1.9)	1 (0.8)	0.72
Species 14	4 (0.7)	0 (0.0)	0 (0.0)	1 (0.9)	1 (1.0)	2 (1.6)	0.59
HPV genotypes							
HPV6	17 (3.1)	12 (12.6)	2 (1.8)	2 (1.8)	1 (1.0)	0 (0.0)	<0.0001
HPV11	8 (1.5)	1 (1.1)	0 (0.0)	7 (6.3)	0 (0.0)	0 (0.0)	<0.0001
HPV16	28 (5.1)	14 (14.7)	5 (4.6)	3 (2.7)	3 (2.9)	3 (2.3)	<0.0001
HPV18	9 (1.7)	4 (4.2)	1 (0.9)	2 (1.8)	0 (0.0)	2 (1.6)	0.20
HPV26	1 (0.2)	0 (0.0)	0 (0.0)	0 (0.0)	1 (1.0)	0 (0.0)	0.37
HPV31	3 (0.6)	1 (1.0)	1 (0.9)	0 (0.0)	0 (0.0)	1 (0.8)	0.73
HPV33	2 (0.4)	0 (0.0)	0 (0.0)	1 (0.9)	0 (0.0)	1 (0.8)	0.63
HPV35	5 (0.9)	1 (1.0)	1 (0.9)	1 (0.9)	2 (1.9)	0 (0.0)	0.66
HPV42	2 (0.4)	1 (1.0)	0 (0.0)	0 (0.0)	1 (1.0)	0 (0.0)	0.47
HPV44	2 (0.4)	1 (1.0)	0 (0.0)	0 (0.0)	1 (1.0)	0 (0.0)	0.47
HPV45	5 (0.9)	1 (1.0)	2 (1.8)	1 (0.9)	1 (1.0)	0 (0.0)	0.70
HPV51	11 (2.0)	2 (2.1)	2 (1.8)	6 (5.4)	1 (1.0)	0 (0.0)	0.05
HPV52	13 (2.4)	0 (0.0)	5 (5.6)	2 (1.8)	2 (1.9)	4 (3.1)	0.27
HPV53	8 (1.5)	2 (2.1)	0 (0.0)	3 (2.7)	3 (2.9)	0 (0.0)	0.17
HPV54	5 (0.9)	0 (0.0)	1 (0.9)	1 (0.9)	2 (1.9)	1 (0.8)	0.72
HPV56	2 (0.4)	0 (0.0)	0 (0.0)	0 (0.0)	1 (1.0)	1 (0.8)	0.59
HPV58	10 (1.8)	2 (2.1)	1 (0.9)	3 (2.7)	2 (1.9)	2 (1.6)	0.90
HPV59	2 (0.4)	0 (0.0)	0 (0.0)	1 (0.9)	0 (0.0)	1 (0.8)	0.63
HPV61	4 (0.7)	0 (0.0)	1 (0.9)	2 (1.8)	0 (0.0)	1 (0.8)	0.52
HPV62	4 (0.7)	2 (2.1)	0 (0.0)	2 (1.8)	0 (0.0)	0 (0.0)	0.15
HPV66	8 (1.5)	3 (3.2)	2 (1.8)	3 (2.7)	0 (0.0)	0 (0.0)	0.16
HPV68	5 (0.9)	0 (0.0)	1 (0.9)	1 (0.9)	1 (1.0)	2 (1.6)	0.83
HPV69	1 (0.2)	1 (1.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.31
HPV70	2 (0.4)	0 (0.0)	0 (0.0)	1 (0.9)	1 (1.0)	0 (0.0)	0.54
HPV71	4 (0.7)	0 (0.0)	0 (0.0)	1 (0.9)	1 (1.0)	2 (1.6)	0.59
HPV72	3 (0.6)	0 (0.0)	1 (0.9)	0 (0.0)	0 (0.0)	2 (1.6)	0.34
HPV73	3 (0.6)	0 (0.0)	1 (0.9)	1 (0.9)	1 (1.0)	0 (0.0)	0.72
HPV81	5 (0.9)	1 (1.0)	3 (2.8)	1 (0.9)	0 (0.0)	0 (0.0)	0.18
HPV82	3 (0.9)	0 (0.0)	1 (0.9)	2 (1.8)	0 (0.0)	0 (0.0)	0.26
HPV83	4 (0.7)	1 (1.0)	0 (0.0)	0 (0.0)	3 (2.9)	0 (0.0)	0.05
HPV84	6 (1.1)	0 (0.0)	2 (1.8)	1 (0.9)	0 (0.0)	3 (2.3)	0.33

	Overall (n= 546)	South (n= 95)	Southeast (n= 109)	Central (n= 111)	Northeast (n= 103)	North (n= 128)	P-value
HPV89	5 (0.9)	0 (0.0)	3 (2.8)	0 (0.0)	1 (1.0)	1 (0.8)	0.20

Table 3. Frequency of HPV genotypes, subgenus and species, overall and by vaginal microbiome community-state types (CST).

	CST I (n=166)	CST II (n=27)	CST III (n=204)	CST IV (n=143)	CST V (n=6)	P-value[a]
Any HPV	27 (16.3)	6 (22.2)	48 (23.5)	35 (24.5)	1 (16.7)	0.40
Oncogenic risk						
Low risk HPV	15 (9.0)	6 (22.2)	24 (11.8)	14 (9.8)	0 (0.0)	0.26
High risk HPV	20 (12.0)	4 (14.8)	36 (17.7)	25 (17.5)	1 (16.7)	0.63
HPV species						
Species 1	1 (0.6)	1 (3.7)	0 (0.0)	0 (0.0)	0 (0.0)	0.04
Species 3	6 (3.6)	1 (3.7)	11 (5.4)	9 (6.3)	0 (0.0)	0.80
Species 5	2 (1.2)	1 (3.7)	7 (3.4)	4 (2.8)	0 (0.0)	0.70
Species 6	8 (4.8)	1 (3.7)	5 (2.4)	3 (2.1)	0 (0.0)	0.62
Species 7	3 (1.8)	1 (1.8)	9 (4.4)	9 (6.3)	0 (0.0)	0.36

	CST I (n=166)	**CST II** (n=27)	**CST III** (n=204)	**CST IV** (n=143)	**CST V** (n=6)	**P-value[a]**
Species 9	9 (5.4)	3 (11.1)	24 (11.8)	17 (11.9)	1 (16.7)	0.23
Species 10	8 (4.8)	4 (14.8)	12 (5.9)	2 (1.4)	0 (0.0)	0.03
Species 11	1 (0.6)	0 (0.0)	0 (0.0)	2 (1.4)	0 (0.0)	0.52
Species 13	3 (1.8)	0 (0.0)	1 (0.5)	1 (0.7)	0 (0.0)	0.69
Species 14	0 (0.0)	0 (0.0)	2 (1.0)	2 (1.4)	0 (0.0)	0.64

[a]Variables compared across region using chi-square test, $P<0.05$ considered as significant.

Table 4. Odds ratios and 95% confidence intervals for the association between characteristics of the study population and HPV infection

	Any HPV		High-risk HPV	
	Age- and region-adjusted	Multivariable[a]	Age- and region-adjusted	Multivariable[b]
Age, years				
≤ 25 years	Ref	Ref	Ref	Ref
26-39 years	0.77 (0.33-1.85)	0.45 (0.26-0.77)	0.81 (0.31-2.09)	0.44 (0.25-0.79)
≥40 years	1.57 (0.34-7.34)	0.55 (0.29-1.03)	1.28 (0.29-7.17)	0.41 (0.20-0.82)
Ethnicity				
Black/pardo	Ref	--	Ref	--
White	1.04 (0.64-1.70)		1.11 (0.64-1.91)	
Other	1.7 (0.65-4.50)		2.07 (0.75-5.74)	
Living with partner				
No	Ref	Ref	Ref	Ref
Yes	0.58 (0.37-0.9)	0.54 (0.35-0.84)	0.59 (0.36-0.97)	0.56 (0.34-0.93)
School, years completed				
≤11	Ref	--	Ref	--
>11	0.95 (0.61-1.48)		0.73 (0.44-1.20)	
Has personal income?				
No	Ref	--	Ref	--
Yes	1.14 (0.72-1.81)		1.26 (0.75-2.11)	
Smoking habit				
No	Ref	--	Ref	--
Yes	1.44 (0.75-2.78)		1.62 (0.80-3.31)	
Alcohol intake				
No	Ref	--	Ref	--
Yes	1.11 (0.70-1.76)		1.01 (0.61-1.71)	
Regular use of intimate soap				
No	Ref	--	Ref	--
Yes	1.04 (0.68-1.61)		0.92 (0.57-1.50)	
Douching				
No	Ref	--	Ref	--
Yes	1.13 (0.63-2.04)		1.23 (0.64-2.35)	
Sitz-bathing				
No	Ref	--	Ref	--
Yes	1.48 (0.63-3.43)		0.98 (0.35-2.75)	
Number of sexual partners[c]				
0	1.0 (0.38-2.58)	--	0.67 (0.19-2.34)	--
1	Ref		Ref	
2+	1.48 (0.78-2.81)		1.56 (0.79-3.13)	
New sexual partner[d]				
No	Ref	--	Ref	Ref
Yes	2.46 (1.21-5.01)		2.85 (1.36-5.96)	2.23 (1.06-4.66)
Sexual intercourses/week				
0	1.31 (0.75-2.28)	--	1.51 (0.82-2.80)	--
1-2	Ref		Ref	
3+	0.77 (0.47-1.26)		0.90 (0.52-1.56)	
Hormonal contraceptive use				

	Any HPV		High-risk HPV	
	Age- and region-adjusted	Multivariable[a]	Age- and region-adjusted	Multivariable[b]
None	Ref	Ref	Ref	Ref
Oral	0.85 (0.52-1.39)	1.11 (0.70-1.78)	0.81 (0.47-1.40)	0.94 (0.56-1.60)
Injectable	0.20 (0.06-0.68)	0.21 (0.06-0.72)	0.09 (0.01-.0.69)	0.08 (0.01-0.63)
Condom use				
No	Ref	Ref	Ref	--
Yes	1.58 (1.01-2.5)	1.47 (0.94-2.29)	1.55 (0.94-2.55)	
Body mass index[e]				
<25.0	Ref	--	Ref	--
25.0-29.9	0.77 (0.48-1.26)		0.69 (0.39-1.20)	
≤30	0.64 (0.33-1.23)		0.64 (0.30-1.36)	
Menstrual cycle phase				
Follicular	Ref	--	Ref	--
Luteal	0.93 (0.60-1.47)		0.88 (0.53-1.45)	
Supressed[f]	0.91 (0.44-1.88)		0.81 (0.35-1.83)	
Ever pregnant				
No	Ref	--	Ref	--
Yes	1.03 (0.59-1.79)		1.09 (0.59-2.03)	
Dyspareunia				
No	Ref	--	Ref	--
Yes	0.66 (0.42-1.05)		0.84 (0.51-1.38)	
Previously treated abnormal discharge				
No	Ref	--	Ref	--
Yes	1.34 (0.85-2.10)		1.34 (0.81-2.23)	
Current abnormal discharge				
No	Ref	--	Ref	--
Yes	0.96 (0.62-1.48)		1.01 (0.63-1.65)	
Discharge with bad odor				
No	Ref	--	Ref	--
Yes	1.00 (0.60-1.66)		1.13 (0.64-1.97)	
Abnormal pH (>4.5)				
No	Ref	--	Ref	--
Yes	1.16 (0.76-1.79)		1.16 (0.71-1.87)	
Number of inflammatory cells[g]				
<5/field (x1000)	1.27 (0.52-3.11)	--	Ref	--
≥5/field (x1000)			1.14 (0.41-3.15)	
Microscopic classification vaginal microbiota[h]				
Normal	Ref	--	Ref	--
Intermediate	1.92 (0.96-3.86)		2.1 (0.98-4.5)	
Bacterial vaginosis	1.89 (1.15-3.03)		1.9 (1.13-3.31)	
Positive for CT, NG or TV				
No	Ref	--	Ref	--
Yes	1.20 (0.49-2.95)		1.38 (0.53-3.56)	
Cervicitis[i]				
No	Ref	--	Ref	--
Yes	1.20 (0.57-2.52)		0.64 (0.24-1.73)	

	Any HPV		High-risk HPV	
	Age- and region-adjusted	Multivariable[a]	Age- and region-adjusted	Multivariable[b]
Vulvitis				
No	Ref	--	Ref	--
Yes	0.57 (0.15-2.13)		0.89 (0.23-3.35)	
Vaginal microbiome				
CST I, II or V	Ref	--	Ref	--
CST III	1.61 (0.97-2.68)		1.56 (0.89-2.79)	
CST IV	1.96 (1.11-3.48)		1.73 (0.91-3.28)	

[a]1Likehood ratio (LR) chi-square (6) =34.14; [b]LR chi-square (6) =34.80; [c]12 months prior enrollment; [d]2 months prior to enrollment. [e]BMI classification according to World Health Organization; [f]Continous use of hormonal contraceptive; [g]Cell count on Gram-stained vaginal smears overserved under immersion oil and (100x) objective lens; [h]Classification of Gram-stained smears using Nugent score (0-3: normal; 4-6: intermediate; 7-10: bacterial vaginosis); [i]Bleeding and/or mucopurulent secretion observed during sampling.
CT: *Chlamydia trachomatis*; NG: *Neisseria gonorrhoeae*; TV: *Trichomonas vaginalis*; CST: Community state type; --indicates covariates that were not retained in the final multivariable model.

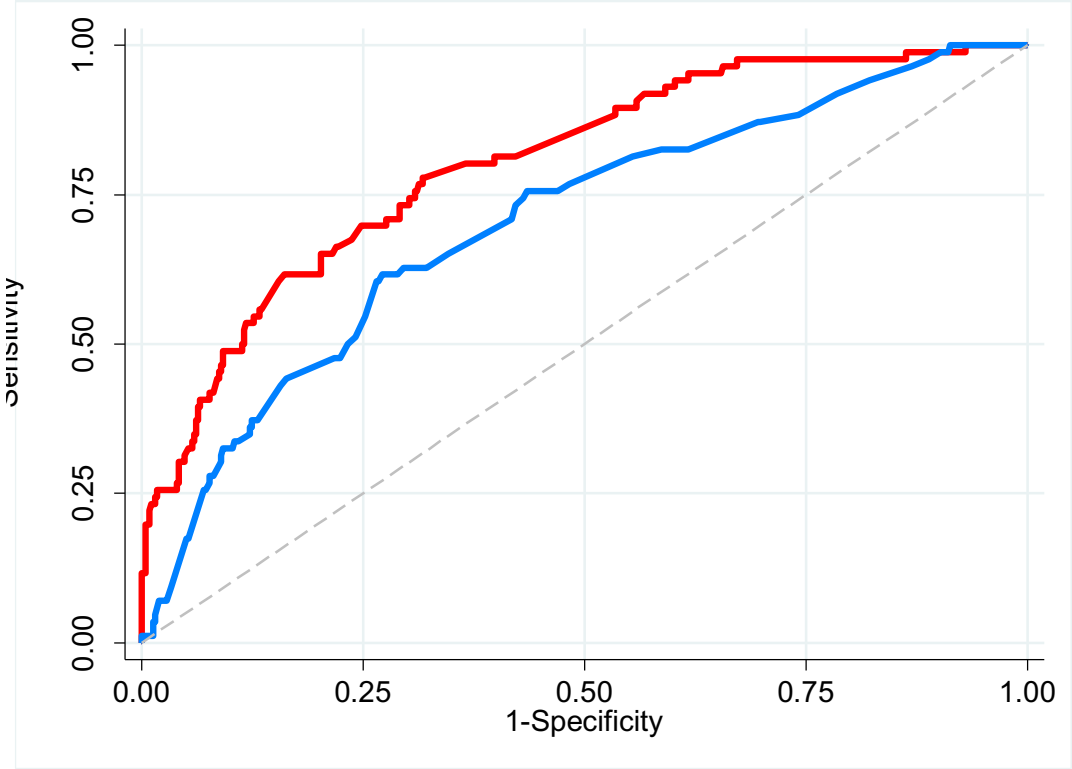

Figure 1. Red curve represents the microbiome-based (MB) score, area under the curve (AUC): 0.8022; 95% confidence interval (CI): 0.7517 - 0.8527). Blue curve represents the sociodemographic, behavioral, and clinical (SBC) variables that were retained in the multivariable model for predicting high-risk HPV infection (AUC 0.7027; 95% CI: 0.6419 - 0.7636). Grey dashed line is the reference. (ROC comparison: P=0.0163)

Supplementary Table 1- Stepwise multivariate logistic regression coefficients of the microbiome-based (MB) score, high-risk HPV as the dependent variable and the 116 bacterial taxa identified as the independent variables.

	Coefficients	Standard error
Shuttleworthia satelles	1.4582	0.7479
Sutterella stercoricanis	0.7375	0.3625
Peptoniphilus	-1.7478	0.6058
Eubacterium saphenum	1.9054	0.7444
Lactobacillus salivarius	2.5021	0.8566
Sutterella morbirenis	0.3369	0.3815
Pediococcus acidilactici	-0.9141	0.3508
Aerococcus viridans	0.9017	0.4694
BVAB3	-1.6070	0.6378
Prevotella genogroup 3	1.6915	0.4858
Streptococcus intermedius	1.8406	1.0690
Corynebacterium accolens	-0.6384	0.3572
Dialister sp type 2	1.5527	0.8458
Megasphaera sp type 2	0.9455	0.4325
Dialister propionicifaciens	-1.2439	0.5761
Eubacterium siraeum	2.7479	0.8471
Bacteroides uniformis	-0.9613	0.5405
Prevotella genogroup 2	0.7865	0.3792
Leptotrichia amnionii	-0.6973	0.3686
Acinetobacter calcoaceticus	1.3697	0.7718
Arcanobacterium hippocoleae	-1.2174	0.6594
Roseburia intestinalis	5.8203	1.9726
Porphyromonas endodontalis	-3.1477	1.6844
Enterococcus faecalis	-1.4769	0.7422
Varibaculum cambriense	1.2476	0.5991
Raoultella planticola	0.7092	0.4070
Staphylococcus lugdunensis	1.1539	0.5878
Streptococcus anginosus	-0.9912	0.5214
Mycoplasma genitalium	0.5829	0.3466
Streptococcus mutans	2.5658	1.5664
Constant	-2.1986	0.2865

A total of 116 species were considered in the model.

Supplementary Table 2- Stepwise multivariate logistic regression coefficients of sociodemographic, behavioral, and clinical (SBC) characteristics, high-risk HPV as the dependent variable

	Coefficients	Standard error
Age, years		
26-39	-0.7888	0.3035
≥40	-0.8679	0.3695
Living with partner at enrollment	-0.5022	0.2584
New sexual partner	0.8827	0.3809
Use of hormonal contraceptive		
Oral	-0.0935	0.2730
Injectable	-2.6580	1.0431
Body mass index		
Overweight	-0.4829	0.2870
Obese	-0.5487	0.3857
Smoking	0.5311	0.3632
Constant	-0.5648	0.3042

A total of 25 variables were considered in the model. Age, ethnicity, living with partner, years at school, personal income, smoking, alcohol, use of intimate soap, vaginal douching, sitz bath, number of sexual partner (12 months), sexual partner (2 months), number of sexual intercourse per week, use of hormonal contraceptives, condom use, body mass index, phase of menstrual cycle at enrollment, was ever pregnant, dyspareunia, abnormal discharge, discharge with bad odor, abnormal pH, signs of cervicitis, vulvitis, history of bacterial vaginosis.