

# Cervical Microbiota Associated with Higher Grade Cervical Intraepithelial Neoplasia in Women Infected with High-Risk Human Papillomaviruses

Chandrika J. Piyathilake<sup>1</sup>, Nicholas J. Ollberding<sup>2</sup>, Ranjit Kumar<sup>3</sup>, Maurizio Macaluso<sup>2</sup>, Ronald D. Alvarez<sup>4</sup>, and Casey D. Morrow<sup>5</sup>

## Abstract

It is increasingly recognized that microbes that reside in and on human body sites play major roles in modifying the pathogenesis of several diseases, including cancer. However, specific microbes or microbial communities that can be mechanistically linked to cervical carcinogenesis remain largely unexplored. The purpose of the study was to examine the association between cervical microbiota and high-grade cervical intraepithelial neoplasia (CIN 2+) in women infected with high-risk (HR) human papillomaviruses (HPV) and to assess whether the cervical microbiota are associated with oxidative DNA damage as indicated by the presence of cervical cells positive for 8-hydroxy-2'-deoxyguanosine. The study included 340 women diagnosed with CIN 2+ (cases) and 90 diagnosed with CIN 1 (non-cases). Microbiota composition was determined by Illumina sequencing of the 16S rRNA gene amplified from DNA extracted from cervical mucus samples. Measures

of alpha/beta-diversity were not associated with either CIN severity or oxidative DNA damage. However, a cervical mucosal community type (CT) dominated by *L. iners* and unclassified *Lactobacillus* spp. was associated with CIN 2+ (OR = 3.48; 95% CI, 1.27–9.55). Sequence reads mapping to Lactobacillaceae, *Lactobacillus*, *L. reuteri*, and several sub-genus level *Lactobacillus* operational taxonomic units were also associated with CIN 2+ when examined independently (effect size >2.0;  $P < 0.05$ ). Our 16S rRNA sequencing results need confirmation in independent studies using whole-genome shotgun sequencing and that would allow sharpening the suggested associations at finer taxonomic levels. Our results provide little evidence that DNA oxidative damage mediates the effect of the microbiome on the natural history of HPV infection and CIN severity. *Cancer Prev Res*; 9(5); 357–66. ©2016 AACR.

## Introduction

The human microbiome is gaining attention as a factor in health and disease, including cancer. The gut microbiome is the best characterized and implicated in carcinogenesis in the gut (1) and liver by translocation of bacterial metabolites or systemic inflammation (2). Microbial communities in other anatomical sites may alter cancer risk (3, 4). Studies have demonstrated that microbiota influence cancer susceptibility and progression by modulating inflammation, inducing oxidative stress, and promoting genomic instability of the host cells (5, 6). As the bacterial density is high in mucous membranes, microbiota may also play a role in the development of precancerous or cancerous lesions of the uterine cervix. High-risk (HR) human papillomaviruses

(HPV) cause high-grade cervical intraepithelial neoplasia (CIN 2+), the precursor of invasive cervical cancer (ICC), but only a small proportion of women infected with HR-HPVs develop CIN 2+. It is important to identify the factors that may alter the natural history of HR-HPV infection.

As the microbiome influences acquisition and pathogenicity of viral infections in the gut (7, 8), the same may occur in the genital tract. Bacterial vaginosis (BV), a condition characterized by a decrease in *Lactobacillus* with a concomitant increase in anaerobic bacteria (e.g., *Gardnerella*, *Prevotella*, and *Clostridiales*; ref. 9), is associated with increased risk of acquisition, reactivation, or delayed clearance of HPV infection (9–11). Previous studies, however, have relied only on Gram stain to assess BV. Vaginal microbiome sequencing by quantitative PCR showed greater microbiome diversity and a higher prevalence of *L. gasseri* and *Gardnerella vaginalis* among HPV-positive women (12). Pyrosequencing of 16S rRNA genes revealed a lower proportion of *Lactobacillus* spp. and identified *Fusobacteria*, including *Sneathia* spp., as key bacteria among HPV-positive women (13). A recent report also using 16S rRNA sequencing found that a vaginal microbiome dominated by *L. gasseri* was associated with increased clearance of HPV (14).

The purpose of the current study was to examine the association between the cervicovaginal microbiome and CIN 2+ among women well characterized for HPV infection and histologically confirmed diagnoses of CIN lesions and other risk factors. We also assessed whether the cervical microbiome was associated with the presence of cervical cells positive for 8-hydroxy-2'-deoxyguanosine [8-OHdG], a marker of oxidative DNA damage.

<sup>1</sup>Department of Nutrition Sciences, The University of Alabama at Birmingham (UAB), Birmingham, Alabama. <sup>2</sup>Cincinnati Children's Hospital Medical Center and University of Cincinnati College of Medicine, Cincinnati, Ohio. <sup>3</sup>University of Alabama at Birmingham (UAB) Center for Clinical & Translational Science. <sup>4</sup>Department of Obstetrics and Gynecology, UAB, Birmingham, Alabama. <sup>5</sup>Department of Cell Developmental and Integrative Biology, UAB, Birmingham, Alabama.

**Note:** Supplementary data for this article are available at Cancer Prevention Research Online (<http://cancerprevres.aacrjournals.org/>).

**Corresponding Author:** Chandrika J. Piyathilake, University of Alabama at Birmingham, Wallace Tumor Institute 420 D, 1824 6th Avenue South, Birmingham, AL 35294-3300. Phone: 205-975-5398; Fax: 205-934-7049; E-mail: piyathic@uab.edu

**doi:** 10.1158/1940-6207.CAPR-15-0350

©2016 American Association for Cancer Research.

## Materials and Methods

### Study population

We evaluated data on 430 women enrolled in two completed studies (R01 CA105448 and R01 CA102489). All patients were diagnosed with abnormal cervical cells by cytology at the Health Departments in Birmingham, AL, and were referred to a colposcopy clinic at the University of Alabama at Birmingham for further examination. Women were 19 to 50 years of age; had no previous history of cervical or other lower genital cancer; no previous hysterectomy, or destructive therapy of the cervix; and were not pregnant. A total of 340 were diagnosed with CIN 2+ [cases, including CIN 2 ( $n = 208$ ), CIN 3 ( $n = 132$ )] and 90 with CIN 1 (non-cases). Both cases and non-cases were positive for HR-HPV (any one of 13 types of HR-HPV, HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and 68) based on the Roche Diagnostics Linear Array. Pelvic examinations and biopsies were carried out following established protocols. All research procedures were approved by the Institutional Review Board at the University of Alabama at Birmingham.

### Collection of data and samples

Age at entry, race, level of education, cigarette smoking, parity, and use of hormonal/oral contraceptives were ascertained via interview with study personnel. Height and weight measurements were obtained using standard protocols and used to calculate the body mass index (BMI). Cervical mucus samples were collected using Merocel ophthalmic sponges (Medtronic Xomed, Inc.) following speculum insertion prior to collecting exfoliated cervical cells. The sponge was placed in the cervical os, with care not to touch the vaginal wall and vulva, allowed to absorb secretions for 60 seconds and placed into a sterile cryovial containing 5 mL of phosphate-buffered saline (PBS) and then placed immediately on ice following collection. Exfoliated cervical cells collected from the transformation zone with a cervical brush were immediately rinsed in 10 mL of PBS and were kept cold. Both sample types were transported to the laboratory on ice within 2 hours of collection. In the laboratory, the sponges suspended in 5 mL of PBS were stored at 80°C. Cervical cell suspensions were separated into two aliquots and centrifuged, and supernatant was removed to obtain cell pellets. One cervical cell pellet used for HPV genotyping was stored in PreservCyt Solution at -20°C. The other cervical cell pellet was fixed by resuspending in 500  $\mu$ L of 3% formalin and allowed to stand at room temperature for 15 minutes.

### Immunohistochemical evaluation of cervical cells for 8-OHdG

The formalin-fixed cervical cell suspension was centrifuged to get a pellet, and supernatant was removed. The cervical cell pellet was again resuspended in PBS and smeared on microscopic slides and allowed to dry out at 37°C for 20 minutes. After drying, slides were placed in 0.01 mol/L citric acid, pH 6.0, and boiled in a pressure cooker set at full power for 10 minutes. After the slides were cooled, they were rinsed in dH<sub>2</sub>O and placed in 3.5 N HCl for 15 minutes at room temperature to open the DNA. The slides were rinsed with dH<sub>2</sub>O and treated with 3.0% H<sub>2</sub>O<sub>2</sub> for 5 minutes to quench endogenous peroxidase activity. Slides were incubated with preimmune goat serum (1%) for 1 hour at room temperature to suppress nonspecific staining and then subsequently incubated with 8-OHdG antibody (catalog NWA-8OHdG, 1:500 concentration) for 1 hour at room temperature. All cervical cells on each

slide were counted to derive the percentages of cells positive for 8-OHdG.

### Isolation of microbial DNA and creation of 16S V4 amplicon library

Microbial DNA was isolated from cervical mucus samples using the Fecal DNA isolation kit from Zymo Research following the manufacturer's instructions. Once the sample DNA was prepared, PCR was used with unique bar-coded primers to amplify the variable region 4 (V4) region of the 16S rRNA gene to create an amplicon library from individual samples (15, 16).

### DNA sequencing and bioinformatic processing

The PCR product was ~255 bases from the V4 segment of the 16S rDNA gene. Paired-end sequencing (2 × 250) was performed on the Illumina MiSeq (15, 16). FASTQ conversion of the raw data files was performed following de-multiplexing using MiSeq reporter. Quality assessment of the FASTQ files was performed using FASTQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and then quality filtering was done using the FASTX toolkit ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)). High-quality reads (reads where 80% bases have Q score > 20) were selected for analysis, and reads with unknown bases ("N") were discarded. The remainder of the steps were performed with the Quantitative Insights into Microbial Ecology (QIIME) suite, version 1.8 as described below (15, 17, 18). Chimeric sequences were filtered with UCHIME using the Ribosomal Database Program (RDP) database (19, 20). Sequences were grouped into operational taxonomic units (OTU) using UCLUST at a similarity threshold of 97% (19). The RDP classifier trained using the Greengenes (v13.8) 16S rRNA database (21) was used to make taxonomic assignments for all OTUs at confidence threshold of 80% (22). OTUs whose average abundance was less than 0.005% were removed (23). Multiple sequence alignment was performed with PyNAST v1.2 (24) and a phylogenetic tree was constructed using FastTree v2.1 (25, 26). A total of 8,600,000 sequences clustered to 652 OTUs were obtained after quality filtering and subsampling to 20k reads. Alpha and beta diversity metrics were calculated as implemented in QIIME (19, 27, 28). Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt; ref. 29) was used to predict Kyoto Encyclopedia of Genes and Genomes (KEGG) orthologs (i.e., functional gene counts) for each sample and to assign inferred gene counts to KEGG pathways.

### Statistical analysis

Differences in demographic and clinical characteristics of participants according to CIN status were assessed by *t* tests or Wilcoxon rank-sum tests for continuous variables and  $\chi^2$  tests for categorical variables. Differences in the Chao1 index, Shannon index, and Faith's phylogenetic diversity ( $\alpha$ -diversity metrics) according to CIN status were tested by the Wilcoxon rank-sum test. Principal coordinates analysis (PCoA) of the weighted and unweighted Unifrac metrics, as well as the Bray-Curtis, Sorensen, Jaccard, and Morisita-Horn indices ( $\beta$ -diversity), were used to visualize differences in microbial community structure according to CIN and 8-OHdG status. Permutational ANOVA of the distance matrices as implemented by the ADONIS function in the R package vegan (30) was used to test for differences in diversity measures. Non-metric dimensional scaling (NMDS) of the Bray-Curtis and Jaccard distance matrices followed by permutational

ANOVA was used to assess differences in the metagenome according to CIN and 8-OHdG status. Although differences in measures of diversity provide evidence that microbial communities are different, it is possible for measures of diversity to be similar while the composition of microbial communities is different. We used a Dirichlet multinomial mixture (DMM) model to partition samples into metacommunities (cervical mucus community types, CT; ref. 31). The method provides a model-based approach to metagenomic clustering that recognizes the discrete nature of the data and uses an evidence-based clustering approach to estimate the number and composition of metacommunities. First, each specimen is considered the result of a sampling process whereby each taxa in the underlying microbial community has a specific probability of being detected, such that the likelihood of the observed sample is derived from a multinomial distribution with a parameter vector representing the probability that each sequence read is from a given taxa. Next, the model uses a Bayesian framework and the Dirichlet distribution as the natural conjugate prior to estimate the parameters of the multinomial sampling distribution of a common, underlying "metacommunity." The novelty of the model lies in extending the Dirichlet prior to a mixture of Dirichlets and provides the means to view the samples as having been generated by a mixture of metacommunities, facilitating the partitioning of samples into clusters. For this analysis, the DMM model as implemented in Mothur v1.35 (32) was fit to an OTU table aggregated to the genus level to improve clustering. The optimal number of partitions is selected by minimizing the Laplace approximation to the negative log model evidence. The model assigns each sample to one of the CTs (based on its OTU composition) and estimates the weight (i.e., relative frequency,  $\pi$ ) and the precision (i.e., a measure inversely related to the degree of dispersion with respect to OTU composition,  $\theta$ ) of each CT. We reviewed the OTU composition of the CTs identified by the DMM model to assess their biological significance, and selected one CT as the referent group, comparing other partitions to the referent. Odds ratios (OR) and 95% confidence intervals (CI) for CIN2+ and high 8-OHdG status according to cervical mucus CT were calculated using logistic regression. Multivariable models included adjustment for age, BMI, race (non-Hispanic black/non-Hispanic white), education (high school graduate yes/no), parity, hormonal contraceptive use (yes/no), smoking status (current smoker yes/no), and CIN

status where appropriate. The indirect effect of CT on CIN 2+ mediated by 8-OHdG was assessed by decomposing the total effect into direct and indirect components and estimating the indirect effect and 95% CI using the PROCESS macro for SAS (33). Linear Discriminant Analysis Effect Size (LEfSe; ref. 34) and DESeq2 (35) were used to detect differences in OTUs according to CIN and 8-OHdG status. LEfSe was performed on the subsampled OTU table using the default values of  $\alpha = 0.05$  and LDA = 2.0. The negative binomial regression model implemented in DESeq2 was fit to the raw counts adjusting for factors contributing to CIN risk (covariates described above). Log<sub>2</sub>fold changes for OTUs with false discovery rate (FDR) corrected  $P < 0.05$  are reported. LEfSe was also used to assess enrichment in imputed genes and gene pathways according to CIN and 8-OHdG status.

## Results

### Participant characteristics

Demographic and clinical characteristics of participants according to CIN status are provided in Table 1. The mean ( $\pm$ SD) age of participants was 26.1 ( $\pm$ 4.7) years. The majority were non-Hispanic black (53%), had more than a high school education (81%), were not current smokers (53%), and were current users of hormonal contraceptives (84%). Mean BMI and parity were 28.0 ( $\pm$ 8.0) and 1.3 ( $\pm$ 1.0), respectively. CIN 2+ cases were younger, had a lower BMI, and higher rates of HPV 16 infection than CIN 1, non-cases ( $P < 0.05$ ). The proportion of women classified to cervical mucosa CT, and with HPV 18 infection, did not differ. The proportion of 8-OHdG-positive cells was 43% in CIN 2+ cases and 35% in non-cases ( $P = 0.062$ ).

### Cervical mucosa CTs

The DMM model identified four CTs with weights  $\pi = (0.15, 0.41, 0.16, 0.28)$  and precision  $\theta = (7.7, 89.8, 283.0, 55.6)$ . Partition 1 had the lowest precision and was composed of diverse taxa, including unclassified *Lactobacillus*, *L. iners*, Bifidobacteriaceae, Clostridiales, and *Allobaculum* (Fig. 1). Partition 2 was the most prevalent, while *Lactobacillus* dominated, uniquely characterized by the high diversity and proportion of rare taxa. Partition 3 was the least varied and dominated by unclassified *Lactobacillus* (43%) and *L. iners* (40%) organisms. The percentages of sequence reads mapping to *L. iners* in partitions 1, 2, and 4 were 15%, 8%, and 6%, respectively. Partition 4 was the only CT lacking

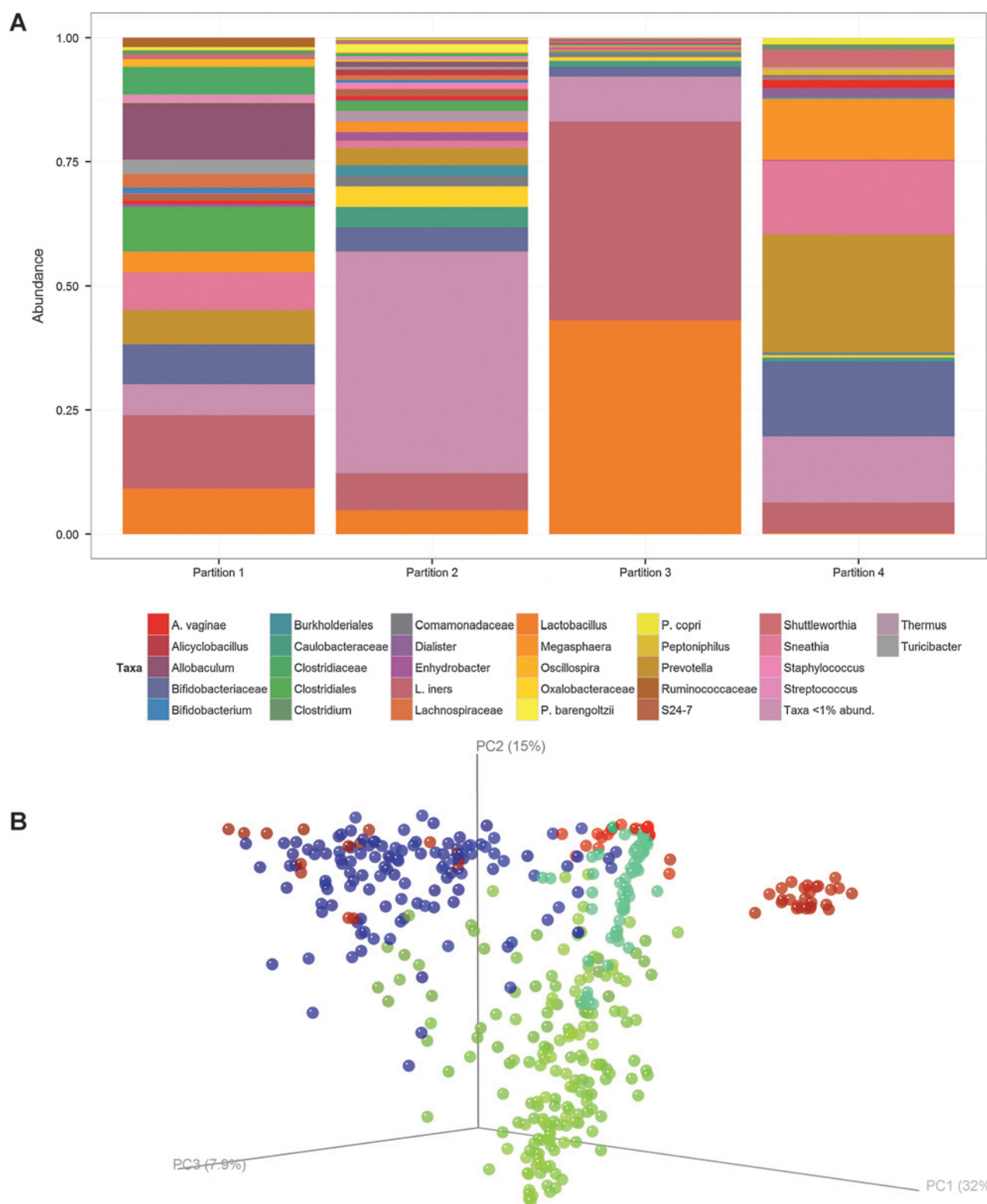
**Table 1.** Participant characteristics according to CIN status

Characteristics	CIN 1 (n = 90)	CIN 2+ (n = 340)	P <sup>a</sup>
Age (y), mean (SD)	29.5 (4.9)	25.2 (4.3)	<0.001
Non-Hispanic black, n (%)	51 (56.7)	178 (52.4)	0.466
High school graduate, n (%)	76 (84.4)	273 (80.3)	0.371
Body mass index (kg/m <sup>2</sup> ), mean (SD)	29.8 (8.1)	27.6 (7.9)	0.011
Current smoker, n (%)	42 (46.7)	161 (47.4)	0.908
Parity, mean (SD)	1.6 (1.2)	1.3 (1.0)	0.050
Current use of hormonal contraceptives, n (%)	70 (77.8)	285 (83.8)	0.258
Cervical mucosa community type, n (%)			
Partition 1	17 (18.9)	46 (13.5)	0.208
Partition 2	34 (37.8)	141 (41.5)	
Partition 3	10 (11.1)	62 (18.2)	
Partition 4	29 (32.2)	91 (26.8)	
Percentage of cervical cells positive for 8-OHdG, mean (SD) <sup>b</sup>	34.8 (19.5)	43.1 (25.3)	0.062
HPV16 positive, n (%)	25 (27.8)	161 (47.4)	0.001
HPV18 positive, n (%)	8 (8.9)	34 (10.0)	0.752

Abbreviations: CIN, cervical intraepithelial neoplasia; HPV, human papillomavirus; SD, standard deviation.

<sup>a</sup>P value for the *t* test or Wilcoxon rank-sum test for continuous variables and  $\chi^2$  test for categorical variables.

<sup>b</sup>Data available for a subset of participants ( $n = 160$ ).



**Figure 1.** A, relative bacterial abundance according to cervical mucosa community type (CT) partition. B, ordination of the PCoA performed on the weighted UniFrac metric according to CT (red = partition 1, green = partition 2, turquoise = partition 3, blue = partition 4; ADONIS  $P < 0.001$ ).

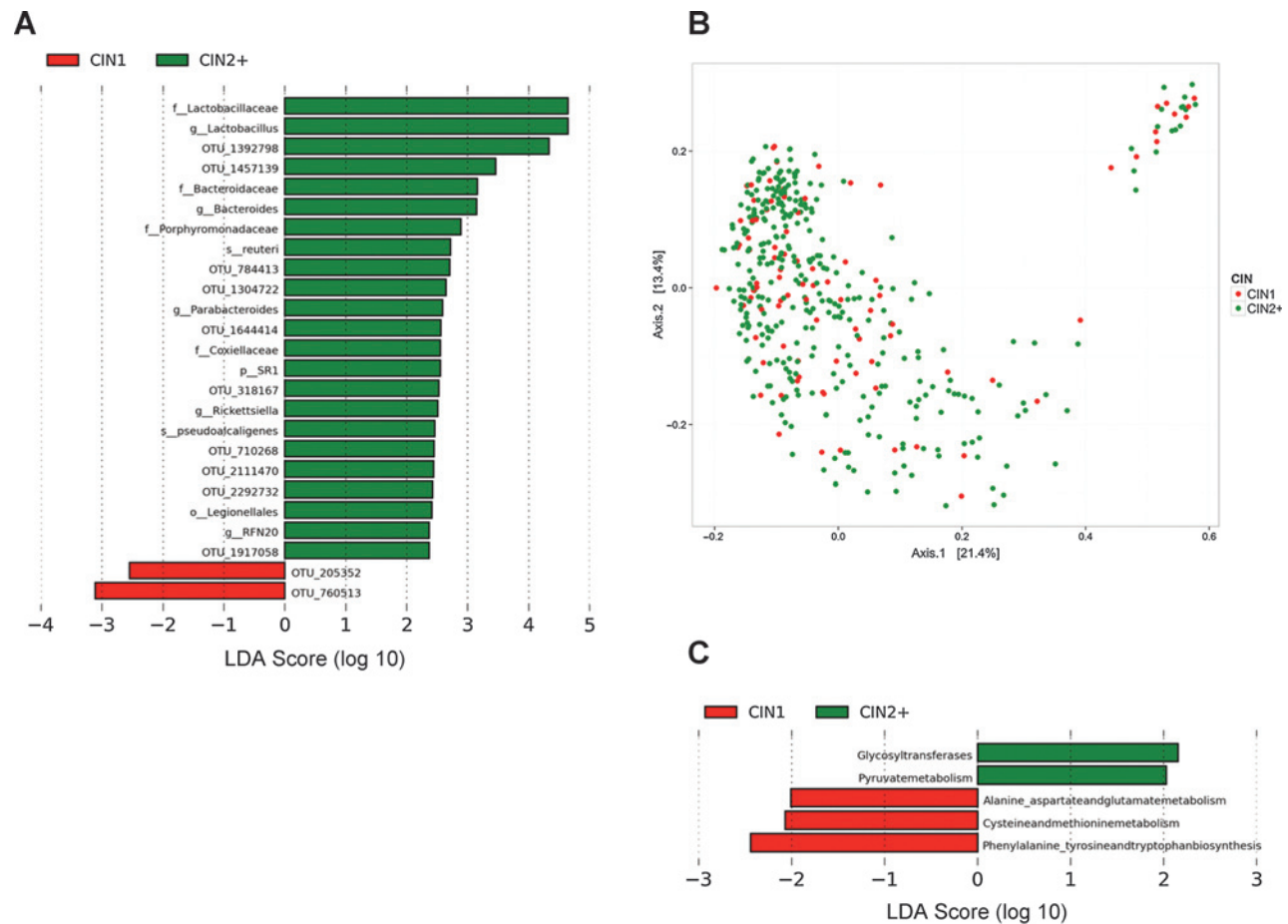
*Lactobacillus* dominance and primarily composed of Bifidobacteriaceae, *Prevotella*, *Sneathia*, and *Megasphaera*. Based on *a priori* considerations about the biological potential of these CTs, we considered partition 1 to be "low risk" and selected it as the referent group, and compared the other three partitions to the referent. Overall, *Lactobacillus* genera were dominant (represented >5% of all sequence reads) in 55% of samples. The proportion of African American women (67%) was higher in partition 4 compared with Caucasian American women (33%). Age, educational status, BMI, smoking status, parity, use of hormonal contraceptives, and the percentage of women positive for HPV 16 or 18 did not differ by partition (data not shown).

#### Cervical mucus microbiota and CIN status

Alpha-diversity metrics did not differ in cases and non-cases ( $P > 0.39$ ). Ordinations based on beta-diversity distance matrices and permutational ANOVA (e.g., ADONIS) did not support differences in community structure (Fig. 2). Similar findings were obtained when comparing CIN 3+ with CIN 1 (data not shown). Whereas overall CT partitioning did not significantly differ

between CIN 2+ and CIN 1 (Table 1,  $P = 0.208$ ), pairwise comparisons between "high-risk" CTs and the referent were significant.

In unadjusted models, samples clustered to partition 3 had a 2.5-fold higher odds [OR = 2.48 (1.01; 6.07)] of being diagnosed with CIN 2+ when compared with samples clustered to partition 1 (Table 2). The OR increased to 3.5 (1.27; 9.55) in models adjusted for factors contributing to CIN risk. Compared with partition 1, ORs for samples clustered to partition 2 were 1.63 (0.83; 3.21) and 1.84 (0.86; 3.93) for unadjusted and multivariable models, respectively. ORs for CIN 3+ were similar to those obtained for CIN 2+, but failed to achieve statistical significance. The odds of CIN 2+ and CIN 3+ for partition 4 were similar to those for partition 1. In the subset of samples ( $n = 156$ ) for which 8-OHdG was assayed the odds ratio for the indirect effect of partition 3 operating through 8-OHdG on CIN 2+ was 1.12 (0.88; 2.44; ~12.5% of total effect mediated; data not shown). Thus, a relatively small proportion of the association between CIN 2+ diagnosis and partition 3 appears to be



**Figure 2.**

A, linear discriminant effect size (LEfSe) analysis comparing differentially abundant taxa according to CIN status (top 25 shown). OTU classifications: *Lactobacillus*; unclassified (OTUs 1304722, 1392798, 1457139), *L. reuteri* (OTU\_784413), *Bacteroides*; unclassified (OTUs 2292732, 318167), *RFN20* (OTU\_1917058), SR1 (OTU\_1644414), *P. pseudoaicaligenes* (OTU\_2111470), *Fusobacterium*; unclassified (OTU\_710268), Ruminococcaceae; unclassified (OTU\_205352), Clostridiales; unclassified (OTU\_760513). B, ordination of the PCoA performed on the unweighted UniFrac metric. C, LEfSe comparing differentially abundant imputed KEGG pathways according to CIN status.

**Table 2.** Odds ratios and 95% CIs for CIN 2+, CIN 3, and 8-OH status according to cervical mucosa community type

Community type	Partition 1	Partition 2	Partition 3	Partition 4
$\pi$ ( $\theta$ ) <sup>a</sup>	0.15 (7.74)	0.41 (89.78)	0.16 (282.98)	0.28 (55.59)
CIN 2+ vs. CIN 1				
Total number (% CIN 2+)	62 (73%)	170 (81%)	68 (87%)	119 (76%)
Unadjusted model	1.00	1.63 (0.83–3.21)	2.48 (1.01–6.07)	1.17 (0.58–2.36)
Multivariable model <sup>b</sup>	1.00	1.84 (0.86–3.93)	3.48 (1.27–9.55)	1.13 (0.51–2.52)
CIN 3+ vs. CIN 1				
Total number (% CIN 3)	34 (50%)	78 (59%)	32 (72%)	65 (55%)
Unadjusted model	1.00	1.44 (0.64–3.23)	2.55 (0.92–7.10)	1.24 (0.54–2.85)
Multivariable model <sup>b</sup>	1.00	1.68 (0.66–4.28)	3.24 (0.98–10.68)	1.20 (0.45–3.20)
8-OHdG (>38.8% vs. $\leq$ 38.8%, median)				
Total number (% above median 8-OHdG)	30 (40%)	42 (57%)	27 (63%)	57 (44%)
Unadjusted model	1.00	2.00 (0.77–5.18)	2.55 (0.88–7.43)	1.17 (0.48–2.88)
Multivariable model <sup>c</sup>	1.00	2.23 (0.82–6.07)	2.61 (0.85–8.03)	1.42 (0.53–3.79)

NOTE: Partitioning of samples was performed using the Dirichlet multinomial mixture model (31) as implemented in Mothur v1.35 (32).

Abbreviations: CIN, cervical intraepithelial neoplasia; 8-OHdG, 8 hydroxy deoxyguanosine.

<sup>a</sup> $\pi$  reflects proportion of samples clustered to the community type.  $\theta$  reflects variance with small values reflecting greater variation.

<sup>b</sup>Adjusted for age, BMI, race, education, parity, hormonal contraceptive use and smoking status.

<sup>c</sup>Adjusted for age, BMI, race, education, parity, hormonal contraceptive use, smoking status, and CIN status.

mediated by oxidative DNA damage. The indirect effect of 8-OHdG was not testable for the other CT partitions.

LEfSe analyses assessing taxa discriminating on CIN status are shown in Fig. 2 (top 25) and Supplementary Table S1. Samples from CIN 2+ were enriched in Lactobacillaceae, *Lactobacillus*, *L. reuteri* and several sub-genus level *Lactobacillus* OTUs (effect size >2.0;  $P < 0.05$ ). Bacterial families, including Bacteroidaceae, Porphyromonadaceae, and Coxiellaceae; genera, including *Bacteroides*, *Parabacteroides*, *Rickettsiella*, and *RFN20*; and numerous OTUs ( $n = 20$ ) were also enriched in CIN 2+ samples. Two OTUs classified to Ruminococcaceae and Clostridiales were enriched in samples obtained from CIN1 non-cases. Similar findings for OTUs were obtained from negative binomial regression after accounting for differences in sequencing depth and adjustment for factors contributing to CIN risk (Supplementary Table S2).

PICRUSt was used to infer the functional metagenomic content of each sample and to classify the inferred genes to KEGG pathways. NMDS of the Bray-Curtis and Jaccard distance matrices and permutational ANOVA did not support differences in the inferred metagenome by CIN status ( $P = 0.20$ ; Supplementary Fig. S1). LEfSe analysis of the inferred metagenome identified five KO identifiers enriched in women diagnosed with CIN 1, including K02003 (ATP-binding protein), K09687 (antibiotic transport system ATP-binding protein), K02004 (ABC transport system permease protein), K06147 (ATP-binding cassette), and K07133 (unclassified; data not shown). Analysis of inferred functional pathways identified glycosyltransferase enzymes and pyruvate metabolism enriched in CIN 2+ samples and enrichment in alanine, aspartate, and glutamate metabolism; cysteine and methionine metabolism; and phenylalanine, tyrosine, and tryptophan biosynthesis in CIN1 (effect size >2.0;  $P < 0.05$ ; Fig. 2).

### Cervical mucus microbiota and 8-OHdG

The percentage of 8-OHdG-positive cervical cells was not associated with microbial diversity when examined as a continuous variable or dichotomized at the median (median = 38.8%;  $P < 0.23$  all models; data not shown). Compared with partition 1, odds ratio for samples clustered to partition 2 and 3 were 2.23 (0.82; 6.70) and 2.61 (0.85; 8.03) in multivariable models, respectively (Table 2). Enrichment in Betaproteobacteria, Bur-

kholderiales, and Rhizobiales; several bacterial families including Psuedomonadaceae, Oxalobacteraceae, Rhizobiaceae; and several genera including *Pseudomonas*, *Argobacterium*, *limnohabitans*, and *Phenylobacterium* was found in samples from women with above median values for 8-OHdG (effect size >2.0;  $P < 0.05$ ; Supplementary Fig. S2). Samples from women with low 8-OHdG levels were enriched in Clostridiaceae. Analysis of imputed metagenomes identified two genes enriched in non-cases including K08884 (serine/threonine protein kinase) and K01897 (long-chain acyl-CoA synthetase). Analysis of the imputed KEGG pathways identified enrichment in two component system, signal transduction, lipid metabolism, transcription factors, secretion system, metabolism of other amino acids, and aminobenzoate degradation in samples from women with a greater percentage of 8-OHdG-positive cells (Supplementary Fig. S2). Enrichment in alanine, aspartate, and glutamate metabolism; terpenoid backbone synthesis; thiamin metabolism; one carbon pool by folate; cysteine and methionine metabolism; translation proteins; amino acid-related enzymes; genetic information processing; and energy metabolism was found for samples from women below the median percentage cells positive for 8-OHdG.

### Discussion

We studied a well-characterized group of women to evaluate the association between cervical microbiota and CIN 2+ and to assess whether cervical microbiota are associated with oxidative DNA damage. Measures of alpha- or beta-diversity were not associated with either CIN severity or oxidative DNA damage. A cervical mucosal community type dominated by *L. iners* and unclassified *Lactobacillus* spp. was associated with CIN 2+. Sequence reads mapping to Lactobacillaceae, *Lactobacillus*, *L. reuteri*, and several sub-genus level *Lactobacillus* OTUs were also associated with CIN 2+. We found little evidence that DNA oxidative damage mediates the effect of the microbiome on the natural history of HPV infection and CIN severity.

Healthy vaginal flora is described as *Lactobacillus* dominated with low bacterial diversity and accompanying low pH, while "unhealthy" flora is often characterized by high diversity bacterial populations with increased anaerobic bacteria and lactobacilli present at lower levels (36). High diversity communities are

associated with pro-inflammatory cytokine concentrations (37). Recent data from diverse human populations have suggested that this characterization is overly simplistic. Anahtar and colleagues recently documented that in a cohort of young, healthy South African women, only 37% had a lactobacillus-dominated vaginal microbiota (37). Although recent research has clarified host-commensal dynamics in the gastrointestinal tract and to a certain extent in the vaginal microbiome in relation to pregnancy complications and pelvic inflammatory diseases (38–40), much less is understood about the cervical microbiome in relation to cervical cancer risk. This gap must be filled if we want to explore the potential for microbiota-based interventions to prevent cervical cancer.

Microbial communities in the cervical mucosa have not been well studied, and comparing our findings with previous research is problematic. Because of the close proximity of the organs, however, given that the cervicovaginal microbiome plays a role in the acquisition and natural history of HPV infections, it is plausible that the microbial communities of the cervical mucosa modify the risk of developing HR-HPV-associated CIN 2+. No comprehensive studies have evaluated this hypothesis. A study compared Korean women diagnosed with CIN 1 ( $n = 55$ ) and CIN 2+ ( $n = 15$ ) with those with normal cytology or ASCUS ( $n = 50$ ) and reported that a cervical cell microbiome with a predominance of *A. vaginae*, *G. vaginalis*, and *L. iners* and a concomitant paucity of *L. crispatus*, in combination with oncogenic HPVs, may be a risk factor for cervical neoplasia (41). The study is of limited value because CIN 1 is not a "true" pre-neoplastic lesion, and some of the women diagnosed with ASCUS are likely to have underlying CIN 2+ lesions. Further, the results reported were not adjusted by other risk factors.

We observed that a cervical mucosa CT dominated by *L. iners* and unclassified *Lactobacillus* was associated with CIN 2+. Smith and colleagues also observed that cervical microbiota clustered in six distinct CTs, one of which was characterized by a predominance of *L. iners* (42). *L. iners*, discovered by Falsen and colleagues (43), has escaped attention because it grows only on blood agar. *L. iners*-dominated samples have lower concentrations of D-lactic acid, which is protective against infection by pathogens (44). *L. iners* has also been reported to induce IL8 secretion *in vivo* and thereby potentially moderate pro-inflammatory activity in the cervix, which may influence CIN progression (36). We did not observe an association between *L. iners* and CIN 2+ when examining single OTUs, but did so when aggregating taxa to Lactobacillaceae and *Lactobacillus*, as well as for *L. reuteri* and several other sub-genus level *Lactobacillus* OTUs. Thus, if the association between *L. iners* and CIN 2+ is real, our findings suggest that the broader community structure may influence the effect of *L. iners* on CIN 2+ risk. Studies incorporating whole-genome shotgun sequencing, finer taxonomic resolution, and metatranscriptomics may provide greater insight into the nature of this association.

The limited precision of the ORs estimated for the cervical mucosa CTs may reflect the inherent challenge of attempting to reduce a high-dimensional community structure to a few clusters. Figure 1 shows clear clustering of samples into CTs, but the large within-CT spread suggests residual variation in community structure. Potential contamination of specimens during sampling, processing, or sequencing may have also reduced our ability to detect associations with CIN 2+. Primer selection and library preparation choices also represent a potential source of bias (45).

On the other hand, clustering may capture the interactions of specific microbes operating in complex community structures that may not be well represented using general diversity measures based on underlying distance matrices, and strengthens the power of our analyses. There is a growing body of literature in the field of microbial ecology pertaining to the use of Dirichlet multinomial models (DMM) to identify naturally occurring clusters of samples based on microbial composition. These methods have been used in numerous studies to cluster samples into "community types" or "enterotypes" (46–49) and readily implemented in commonly used software programs, including Mothur and R (16, 50). DMM clustering provides information that is distinctly different from ordinations performed on dissimilarity matrices, as it is performed directly on read counts (i.e., do not rely on a continuous approximation to the discrete nature of the data), takes into account sequence variation, does not rely on an underlying distance matrix, and is not limited to defining a fixed number of clusters. The DMM described by Holmes and colleagues (31) uses a Bayesian framework to assign samples to the community type for which it has the highest posterior probability of membership with the optimal number of clusters based on minimizing model error. The relative abundance of taxa in each cluster can then be used to facilitate interpretation of major organisms comprising the group. The DMM approach used here has been shown to reveal important CTs when applied to data from the Human Microbiome Project and superior performance to alternative clustering methods (51).

Whereas previous studies have suggested that BV may alter the natural history of HPV infection, in this study, a "BV-like" CT (partition 4) was not associated with CIN 2+. As we did not use conventional BV diagnostic tests, it is possible that women in partition 4 did not have clinical BV. The cervical microbiome of CIN 2+ also tended to be enriched in potentially beneficial microbes such as *L. reuteri* (52, 53), but also microbes found in the intestinal tract such as *Bacteroides*, which have been associated with colorectal cancer (54) and cause significant pathology when they escape the intestinal tract (55). In contrast, samples of non-cases were enriched with Ruminococcaceae, one of the most abundant gut Firmicute families associated with good gut health (56, 57) and Clostridia, common gut microbiota involved in the maintenance of gut function (58). In healthy women, microbiota in the gut and cervicovaginal area are confined to the mucosa by a multilevel barrier system that includes a mucus layer, secretion of soluble immune mediators, and intact epithelium with tight junctions. Failure of this barrier may lead to translocation of microbiota (59), possibly inducing inflammation, oxidative DNA damage, and ultimately the development of CIN 2+. Mucin-degrading enzymes secreted by these microbes may destroy the mucosal barrier and increase susceptibility to HPV (60). Many bacterial species colonize both the gastrointestinal and the reproductive tract, and the rectum serves as a reservoir of organisms that colonize the cervicovaginal area (61–63).

Whereas we did not find a statistically significant association between the percentage of cervical cells positive for 8-OHdG and microbial diversity or community structure, OR estimates were  $>2$ . We note that the highest percentage of women with above median 8-OHdG was found in partition 3, which was associated with CIN 2+. Thus, a biologically plausible association may exist despite our inconclusive findings. Oxidative DNA damage is an inevitable consequence of cellular metabolism and is increased following toxic insult. 8-OHdG is an established cancer

biomarker of oxidative stress (64): elevated levels of 8-OHdG were reported in invasive ductal carcinoma (65), primary breast cancer (66), and colorectal cancer (67) compared with normal cells of the corresponding organ, and in precancerous lesions of the cervix compared with normal cervical epithelial cells (68, 69). Assessing the impact of the cervical microbiota on 8-OHdG levels is important to understand the role of microbial community composition and oxidative stress in the progression of HPV-related cervical lesions. The microbiome influences reactive oxygen species (ROS) production in the gut epithelium (70) and chronically high ROS levels can outpace DNA repair mechanisms, leading to DNA damage (71). *In vitro* studies have shown that ROS production is associated with 8-OHdG formation and increased DNA damage (72). ROS induce double-strand breaks in the host DNA and in the HPV episome, facilitating integration of the virus and carcinogenesis (63). There is accumulating evidence that microbial metabolites may induce cancer via DNA damage (73).

Glycosyltransferase enzymes and pyruvate metabolism were enriched in CIN 2+ samples. Although aberrant pyruvate metabolism (74) and glycosyltransferases (75) play a role in carcinogenesis, the impact of bacterial pyruvate metabolism on CIN is unclear. Some of the genes identified in non-cases and in women with lower 8-OHdG, such as methionine metabolism, play an important role in the pathway linking folate/vitamin B12 metabolism with DNA methylation and DNA repair: we have previously reported that a higher degree of DNA methylation is inversely associated with CIN 2+ (76). Because the gut microbiome supplies its host with up to a third of the recommended intake of B vitamins (77, 78), and because the rectum is a reservoir for organisms that colonize the vagina, our previous observations could be explained by differences in gut microbial profiles by CIN status. The folate biosynthesis pathway is present in nearly all Bacteroidetes genomes, as well as in most Fusobacteria and Proteobacteria while vitamin B12 synthesis of cobalamin is present in all Fusobacteria, but is rare in Actinobacteria and Proteobacteria (79). Like colonocytes, cervical epithelial cells contain transporters for these micronutrients (80, 81). Thus, cervical microbiota may influence the natural history of HPV infection and CIN by altering the status of these micronutrients in the cervical environment.

Some of the genes identified in women with lower 8-OHdG such as one-carbon pool by folate and methionine metabolism play an important role in the folate metabolic pathway, which contributes to DNA synthesis and repair. It is biologically plausible that pathways such as signal transduction are enriched in women with a greater percentage cells positive for 8-OHdG

because DNA damage triggers a signal transduction pathway that appears to be a kinase cascade (82).

In summary, our study did not show an association between microbiome diversity and CIN severity or oxidative DNA damage, but provided suggestive evidence that a cervical microbiome characterized by a predominance of *Lactobacillus* and *L. iners* is associated with CIN 2+ in women infected with HR-HPVs. Future studies are needed to confirm our 16S rRNA sequencing results and to resolve associations for specific organisms and at finer taxonomic levels using whole-genome shotgun sequencing together with measures of cervical microbiome-induced molecular pathways that may alter the risk of CIN 2+. Our data suggest that only a small proportion of the effect of the microbiome on CIN may be mediated through oxidative DNA damage.

### Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

### Authors' Contributions

**Conception and design:** C.J. Piyathilake, M. Macaluso

**Development of methodology:** C.J. Piyathilake, C.D. Morrow

**Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.):** C.J. Piyathilake

**Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis):** C.J. Piyathilake, N.J. Ollberding, R. Kumar, M. Macaluso, C.D. Morrow

**Writing, review, and/or revision of the manuscript:** C.J. Piyathilake, N.J. Ollberding, R. Kumar, M. Macaluso, R.D. Alvarez, C.D. Morrow

**Study supervision:** C.J. Piyathilake

### Acknowledgments

The authors acknowledge the excellent technical assistance provided by Peter Eipers, Michelle Moses Chambers, and Suguna Badiga.

### Grant Support

This work was supported by grants R21 CA188066 (C. Piyathilake), R01 CA102489 (C. Piyathilake), and R01 CA105448 (C. Piyathilake) of the National Cancer Institute. The following are acknowledged for their support of the Microbiome Resource at the University of Alabama at Birmingham: School of Medicine, Comprehensive Cancer Center (P30AR050948), Center for AIDS Research (5P30AI027767) and Center for Clinical Translational Science (UL1TR000165). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received September 22, 2015; revised February 10, 2016; accepted February 28, 2016; published OnlineFirst March 2, 2016.

### References

- Feng Q, Liang S, Jia H, Stadlmayr A, Tang L, Lan Z, et al. Gut microbiome development along the colorectal adenoma-carcinoma sequence. *Nat Commun* 2015;6:6528.
- Dapito DH, Mencin A, Gwak GY, Pradere JP, Jang MK, Mederacke I, et al. Promotion of hepatocellular carcinoma by the intestinal microbiota and TLR4. *Cancer Cell* 2012;21:504–16.
- Bultman SJ. Emerging roles of the microbiome in cancer. *Carcinogenesis* 2014;35:249–55.
- Yu Y, Champer J, Beynet D, Kim J, Friedman AJ. The role of the cutaneous microbiome in skin cancer: lessons learned from the gut. *J Drugs Dermatol* 2015;14:461–5.
- Mangerich A, Knutson CG, Parry NM, Muthupalani S, Ye W, Prestwich E, et al. Infection-induced colitis in mice causes dynamic and tissue-specific changes in stress response and DNA damage leading to colon cancer. *Proc Natl Acad Sci USA* 2012;109:E1820–9.
- Bultman SJ. Emerging roles of the microbiome in cancer. *Carcinogenesis* 2014;35:249–55.
- Kuss SK, Best GT, Etheredge CA, Pruijssers AJ, Frierson JM, Hooper LV, et al. Intestinal microbiota promote enteric virus replication and systemic pathogenesis. *Science* 2011;334:249–52.
- Kane M, Case LK, Kopaskie K, Kozlova A, MacDearmid C, Chervonsky AV, et al. Successful transmission of a retrovirus depends on the commensal microbiota. *Science* 2011;334:245–9.
- Fredricks DN, Fiedler TL, Marrazzo JM. Molecular identification of bacteria associated with bacterial vaginosis. *N Engl J Med* 2005;353:1899–911.



10. Watts DH, Fazzari M, Minkoff H, Hillier SL, Sha B, Glesby M, et al. Effects of bacterial vaginosis and other genital infections on the natural history of human papillomavirus infection in HIV-1-infected and high-risk HIV-1-uninfected women. *J Infect Dis* 2005;191:1129–39.
11. 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.
12. Gao W, Weng J, Gao Y, Chen X. Comparison of the vaginal microbiota diversity of women with and without human papillomavirus infection: a cross-sectional study. *BMC Infect Dis* 2013;13:271.
13. 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.
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;210:1723–33.
15. Kumar R, Eipers P, Little RB, Crowley M, Crossman DK, Lefkowitz EJ, et al. Getting started with microbiome analysis: sample acquisition to bioinformatics. *Curr Protoc Hum Genet* 2014;82:18.8.1–18.8.29.
16. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* 2013;79:5112–20.
17. Lozupone CA, Hamady M, Kelley ST, Knight R. Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Appl Environ Microbiol* 2007;73:1576–85.
18. Navas-Molina JA, Peralta-Sanchez JM, Gonzalez A, McMurdie PJ, Vazquez-Baeza Y, Xu Z, et al. Advancing our understanding of the human microbiome using QIIME. *Methods Enzymol* 2013;531:371–444.
19. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* (Oxford, England) 2010;26:2460–61.
20. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, et al. Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res* 2014;42 (Database issue):D633–42.
21. 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;6:610–18.
22. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 2007;73:5261–7.
23. Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R, et al. Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nature Methods* 2013;10:57–9.
24. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* (Oxford, England) 2010;26:266–7.
25. Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS one* 2010;5:e9490.
26. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evolution* 2009;26:1641–50.
27. Lozupone C, Hamady M, Knight R. UniFrac—an online tool for comparing microbial community diversity in a phylogenetic context. *BMC Bioinformatics* 2006;7:371.
28. Vazquez-Baeza Y, Pihrrung M, Gonzalez A, Knight R. EMPERor: a tool for visualizing high-throughput microbial community data. *Gigascience* 2013;2:16.
29. Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnol* 2013;31:814–21.
30. Oksanen Jari, Guillaume Blanchet F, Kindt Roeland, Legendre Pierre, Minchin Peter R, O'Hara R B, et al. (2015). *vegan: Community Ecology Package*. R package version 2.3-0. <http://CRAN.R-project.org/package=vegan>
31. Holmes I, Harris K, Quince C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. *PLoS One* 2012;7:e30126.
32. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 2009;75:7537–41.
33. Hayes A. Introduction to Mediation, Moderation and conditional process analysis. Chapter 4. New York, NY: The Guilford Press; 2013.
34. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. *Genome Biol* 2011;12:R60.
35. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15:550.
36. Brotman RM. Vaginal microbiome and sexually transmitted infections: an epidemiologic perspective. *J Clin Invest* 2011;121:4610–17.
37. Anahtar MN, Byrne EH, Doherty KE, Bowman BA, Yamamoto HS, Soumillon M, et al. Cervicovaginal bacteria are a major modulator of host inflammatory responses in the female genital tract. *Immunity* 2015;42:965–76.
38. Hillier SL, Nugent RP, Eschenbach DA, Krohn MA, Gibbs RS, Martin DH, et al. Association between bacterial vaginosis and preterm delivery of a low-birth-weight infant. The Vaginal Infections and Prematurity Study Group. *N Engl J Med* 1995;333:1737–42.
39. Leitich H, Bodner-Adler B, Brunbauer M, Kaider A, Egarter C, Husslein P. Bacterial vaginosis as a risk factor for preterm delivery: a meta-analysis. *Am J Obstet Gynecol* 2003;189:139–47.
40. Wilks M, Wiggins R, Whiley A, Hennessy E, Warwick S, Porter H, et al. Identification and H<sub>2</sub>O<sub>2</sub> production of vaginal lactobacilli from pregnant women at high risk of preterm birth and relation with outcome. *J Clin Microbiol* 2004;42:713–7.
41. Oh HY, Kim BS, Seo SS, Kong JS, Lee JK, Park SY, et al. The association of uterine cervical microbiota with an increased risk for cervical intraepithelial neoplasia in Korea. *Clin Microbiol Infect* 2015;21:674.e1–9.
42. Smith BC, McAndrew T, Chen Z, Harari A, Barris DM, Viswanathan S, et al. The cervical microbiome over 7 years and a comparison of methodologies for its characterization. *PLoS ONE* 2012;7:e40425.
43. Falsen E, Pascual C, Sjöden B, Ohlén M, Collins MD. Phenotypic and phylogenetic characterization of a novel *Lactobacillus* species from human sources: description of *Lactobacillus iners* sp. nov. *Int J Syst Bacteriol* 1999;49:217–21.
44. Witkin SS, Mendes-Soares H, Linhares IM, Jayaram A, Ledger WJ, Forney LJ. Influence of vaginal bacteria and D- and L-lactic acid isomers on vaginal extracellular matrix metalloproteinase inducer: implications for protection against upper genital tract infections. *MBio*. 2013;4:e00460–13.
45. Schirmer M, Ijaz UZ, D'Amore R, Hall N, Sloan WT, Quince C. Schirmer. Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Research* 2015;43:e37.
46. Nakatsu G, Li X, Zhou H, Sheng J, Wong SH, Wu WK, et al. Gut mucosal microbiome across stages of colorectal carcinogenesis. *Nat Commun* 2015;6:8727.
47. Uchihashi M, Bergin I, Bassis C, Hashway S, Chai D, Bell J. Influence of age, reproductive cycling status, and menstruation on the vaginal microbiome in baboons (*Papio anubis*). *Am J Primatol* 2015;77:563–578.
48. Yooshep S, Kirkness EF, Tran TM, Harkins DM, Jones MB, Torralba MG, et al. Stool microbiota composition is associated with the prospective risk of *Plasmodium falciparum* infection. *Bmc Genomics* 2015;16:631.
49. Vandeputte D, Falony G, Vieira-Silva S, Tito RY, Joossens M, Raes J. Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates. *Gut Microotia* 2016;65:57–62.
50. Martin Morgan. DirichletMultinomial: Dirichlet-multinomial mixture model machine learning for microbiome data. R package version 1.12.1.
51. Ding T, Schloss PD. Dynamics and associations of microbial community types across the human body. *Nature* 2014;509:357–360.
52. Lakritz JR, Poutahidis T, Levkovich T, Varian BJ, Ibrahim YM, Chatzigiagkos A, et al. Beneficial bacteria stimulate host immune cells to counteract dietary and genetic predisposition to mammary cancer in mice. *Int J Cancer* 2014;135:529–40.
53. Szajewska H, Gyrzczuk E, Horvath A. *Lactobacillus reuteri* DSM 17938 for the management of infantile colic in breastfed infants: a randomized, double-blind, placebo-controlled trial. *J Pediatr*. 2013;162:257–62.
54. Wu N, Yang X, Zhang R, Li J, Xiao X, Hu Y, et al. Dysbiosis signature of fecal microbiota in colorectal cancer patients. *Microb Ecol* 2013;66:462–70.
55. Wexler HM. Bacteroides: the good, the bad, and the nitty-gritty. *Clin Microbiol Rev* 2007;20:593–621.
56. Jalanka-Tuovinen J, Salonen A, Nikkilä J, Immonen O, Kekkonen R, Lahti L, et al. Intestinal microbiota in healthy adults: temporal analysis reveals

- individual and common core and relation to intestinal symptoms. *PLoS One* 2011;6:e23035.
57. Tap J, Mondot S, Levenez F, Pelletier E, Caron C, Furet JP, et al. Towards the human intestinal microbiota phylogenetic core. *Environ Microbiol* 2009;11:2574–84.
  58. Lopetuso LR, Scaldaferri F, Petito V, Gasbarrini A. Commensal Clostridia: leading players in the maintenance of gut homeostasis. *Gut Pathog* 2013;5:23.
  59. Schwabe RF, Jobin C. The microbiome and cancer. *Nat Rev Cancer* 2013;13:800–12.
  60. 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.
  61. Antonio MA, Rabe LK, Hillier SL. Colonization of the rectum by *Lactobacillus* species and decreased risk of bacterial vaginosis. *J Infect Dis* 2005;192:394–8.
  62. El Aila NA, Tency I, Claeys G, Verstraelen H, Saerens B, Santiago GL, et al. Identification and genotyping of bacteria from paired vaginal and rectal samples from pregnant women indicates similarity between vaginal and rectal microflora. *BMC Infect Dis* 2009;9:167.
  63. Petricevic L, Domig KJ, Nierscher FJ, Krondorfer I, Janitschek C, Kneifel W, et al. Characterisation of the oral, vaginal and rectal *Lactobacillus* flora in healthy pregnant and postmenopausal women. *Eur J Obstet Gynecol Reprod Biol* 2012;160:93–9.
  64. Toyokuni S, Okamoto K, Yodoi J, Hiai H. Persistent oxidative stress in cancer. *FEBS Lett*. 1995;358:1–3.
  65. Musarrat J, Arezina-Wilson J, Wani AA. Prognostic and aetiological relevance of 8-hydroxyguanosine in human breast carcinogenesis. *Eur J Cancer* 1996;32A:1209–14.
  66. Matsui A, Ikeda T, Enomoto K, Hosoda K, Nakashima H, Omae K, et al. Increased formation of oxidative DNA damage, 8-hydroxy-2'-deoxyguanosine, in human breast cancer tissue and its relationship to GSTP1 and COMT genotypes. *Cancer Lett* 2000;151:87–95.
  67. Oliva MR, Ripoll F, Muñoz P, Iradi A, Trullenque R, Valls V, et al. Genetic alterations and oxidative metabolism in sporadic colorectal tumors from a Spanish community. *Mol Carcinog*. 1997;18:232–43.
  68. Romano G, Sgambato A, Mancini R, Capelli G, Giovagnoli MR, Flamini G, et al. 8-Hydroxy-2-deoxyguanosine in cervical cells: correlation with grade of dysplasia and human papillomavirus infection. *Carcinogenesis* 2000;21:1143–7.
  69. Visalli G, Riso R, Facciola A, Mondello P, Caruso C, Picerno I, et al. Higher levels of oxidative DNA damage in cervical cells are correlated with the grade of dysplasia and HPV infection. *J Med Virol* 2016;88:336–44.
  70. Neish AS. Redox signaling mediated by the gut microbiota. *Free Radical Res* 2013;47:950–7.
  71. Garrett WS. Cancer and the microbiota. *Science*. 2015;348:80–6.
  72. Zhang X, Wu RS, Fu W, Xu L, Lam PK. Production of reactive oxygen species and 8-hydroxy-2'-deoxyguanosine in KB cells co-exposed to benzo[a]pyrene and UV-A radiation. *Chemosphere* 2004;55:1303–8.
  73. Louis P, Hold GL, Flint HJ. The gut microbiota, bacterial metabolites and colorectal cancer. *Nat Rev Microbiol* 2014;12:661–72.
  74. Gray LR, Tompkins SC, Taylor EB. Regulation of pyruvate metabolism and human disease. *Cell Mol Life Sci* 2014;71:2577–604.
  75. Andergassen U, Liesche F, Kölbl AC, Ilmer M, Hutter S, Friese K, Jeschke U. Glycosyltransferases as markers for early tumorigenesis. *Biomed Res Int* 2015;2015:792672.
  76. Piyathilake CJ, Macaluso M, Alvarez RD, Chen M, Badiga S, Siddiqui NR, et al. A higher degree of LINE-1 methylation in peripheral blood mononuclear cells, a one-carbon nutrient related epigenetic alteration, is associated with a lower risk of developing cervical intraepithelial neoplasia. *Nutrition* 2011;27:513–9.
  77. Hill MJ. Intestinal flora and endogenous vitamin synthesis. *Eur J Cancer Prev* 1997;6:S43–5.
  78. Rossi M, Amaretti A, Raimondi S. Folate production by probiotic bacteria. *Nutrients* 2011;3:118–34.
  79. Magnúsdóttir S, Ravcheev D, de Crécy-Lagard V, Thiele I. Systematic genome assessment of B-vitamin biosynthesis suggests co-operation among gut microbes. *Front Genet* 2015;6:148.
  80. Crott JW, Liu Z, Keyes MK, Choi SW, Jang H, Moyer MP, et al. Moderate folate depletion modulates the expression of selected genes involved in cell cycle, intracellular signaling and folate uptake in human colonic epithelial cell lines. *J Nutr Biochem* 2008;19:328–35.
  81. Wu M, Gunning W, Ratnam M. Expression of folate receptor type alpha in relation to cell type, malignancy, and differentiation in ovary, uterus, and cervix. *Cancer Epidemiol Biomarkers Prev* 1999;8:775–82.
  82. Harper JW, Elledge SJ. The DNA damage response: ten years after. *Mol Cell* 2007;28:739–45.