

Evaluation of Buccal Cell Samples for Studies of Oral Microbiota

Guoqin Yu¹, Steve Phillips², Mitchell H. Gail³, James J. Goedert⁴, Michael Humphrys⁵, Jacques Ravel⁵, Yanfang Ren², and Neil E. Caporaso¹

Abstract

Background: The human microbiota is postulated to affect cancer risk, but collecting microbiota specimens with prospective follow-up for diseases will take time. Buccal cell samples have been obtained from mouthwash for the study of human genomic DNA in many cohort studies. Here, we evaluate the feasibility of using buccal cell samples to examine associations of human microbiota and disease risk.

Methods: We obtained buccal cells from mouthwash in 41 healthy participants using a protocol that is widely employed to obtain buccal cells for the study of human DNA. We compared oral microbiota from buccal cells with that from eight other oral sample types collected by following the protocols of the Human Microbiome Project. Microbiota profiles were determined by sequencing 16S rRNA gene V3–V4 region.

Results: Compared with each of the eight other oral samples, the buccal cell samples had significantly more observed species

($P < 0.002$) and higher alpha diversity (Shannon index, $P < 0.02$). The microbial communities were more similar (smaller beta diversity) among buccal cells samples than in the other samples ($P < 0.001$ for 12 of 16 weighted and unweighted UniFrac distance comparisons). Buccal cell microbial profiles closely resembled saliva but were distinct from dental plaque and tongue dorsum.

Conclusions: Stored buccal cell samples in prospective cohort studies are a promising resource to study associations of oral microbiota with disease.

Impact: The feasibility of using existing buccal cell collections in large prospective cohorts allows investigations of the role of oral microbiota in chronic disease etiology in large population studies possible today. *Cancer Epidemiol Biomarkers Prev*; 26(2); 249–53. ©2016 AACR.

Introduction

Recent studies have revealed associations between the features of human microbial communities (human microbiota) and various diseases, including autoimmune disorders, diabetes, obesity, colon cancer, and even psychiatric conditions (1–7). Large prospective, population-based studies involving thousands of participants are needed to confirm those results and systematically study the role of the human microbiota in health. Profiles of the microbiota can be generated with next-generation DNA sequencing methods, but very few large long-term population-based cohorts have collected biological specimens for the purpose of microbiota research.

¹Genetic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, DHHS, Bethesda, Maryland. ²Eastman Institute of Oral Health, University of Rochester, Rochester, New York. ³Bioinformatics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, DHHS, Bethesda, Maryland. ⁴Infections and Immunoepidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, DHHS, Bethesda, Maryland. ⁵Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, Maryland.

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (<http://cebp.aacrjournals.org/>).

Corresponding Authors: Guoqin Yu, Genetic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, NIH/NCI, Rockville, MD 20852. Phone: 240-276-7249; Fax: 240-276-7832; E-mail: guoqin.yu@nih.gov; and Neil E. Caporaso, Genetic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Rockville, MD, 20852. Phone: 240-276-7228; Fax: 240-276-7832; E-mail: caporasn@mail.nih.gov

doi: 10.1158/1055-9965.EPI-16-0538

©2016 American Association for Cancer Research.

Buccal cells, which have been obtained from mouthwash in many cohorts for the study of human genomic DNA, might be employed to study the oral microbiota. The purpose of the current study was to evaluate whether buccal cell specimens, collected by the protocol used in a prospective cancer cohort (<http://prevention.cancer.gov/major-programs/prostate-lung-colorectal>), could be used for oral microbiota research by comparing buccal cell specimens with eight other oral sample types from the same subjects.

Materials and Methods

Participants

Forty-three subjects were recruited at Eastman Institute of Oral Health, University of Rochester (Rochester, NY). All the subjects signed informed consent and filled out questionnaires. Results are based on 41 subjects after quality control–determined exclusions. Individuals with antibiotic usage or professional dental cleaning within the last 3 months or diagnosed with severe periodontal disease or cancer were excluded. The study was approved by Institutional Review Boards of the NCI (Rockville, MD) and University of Rochester (Rochester, NY).

Sample collection and processing

We collected 8 oral samples, including 2 dental plaque samples (supra- and subgingival plaque), raw saliva, and swabs from 5 soft tissue sites (keratinized gingiva, hard palate, buccal mucosa, palatine tonsil, and tongue dorsum) by following the protocols of the Human Microbiome Project (HMP; ref. 8; http://hmpdacc.org/doc/HMP_MOP_Version12_0_072910.pdf; see Supplementary Fig. S1 for detailed sampling locations). In addition, buccal cells from mouthwash were also collected following the protocol

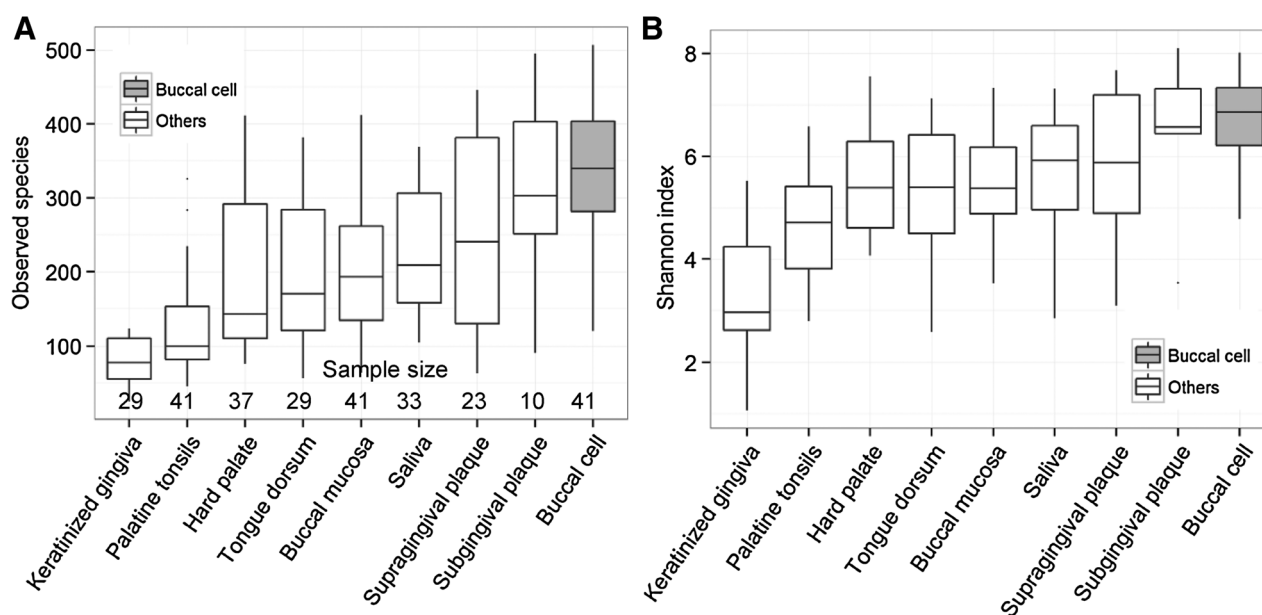


Figure 1. Comparison of alpha diversity measures number of observed species per sample (A) and Shannon index (B) between buccal cell and 8 other oral sample types. The sample size for each sample type was shown at the bottom of each bar. The sample types were ordered by average number of observed species. Boxes, interquartile range (IQR); median values, bands within the boxes; lines outside the boxes, 1.5 times IQR; dots, outliers.

used in the Prostate, Lung, Colorectal and Ovarian (PLCO) cancer cohort (Supplementary Methods).

16S rRNA gene sequence analysis

DNA was extracted and purified as described previously (9). Briefly, samples were thawed on ice and then vortexed for 5 minutes to resuspend the cells. A 0.5-mL aliquot was transferred to a sterile 2.0-mL tube for cell lysis chemically and physically (bead-beating). The resulting crude lysate was processed using QIAamp DNA Mini Kit (Qiagen) according to the manufacturer's guidance. The samples were eluted with 2 × 200 µL of AE buffer (10 mmol/L Tris-Cl, 0.5 mmol/L EDTA; pH 9.0) into separate tubes. The DNA concentrations in the samples were measured using the Quant-iT PicoGreen dsDNA Assay Kit from Molecular Probes (Invitrogen).

The V3–V4 regions of the 16S rRNA gene were amplified and sequenced on an Illumina MiSeq instrument using the 300 paired-end protocol at the Institute for Genome Sciences, Genomic Resource Center, University of Maryland School of Medicine (Baltimore, MD; ref. 10). The sequence reads were processed by removing the low-quality and short-sequence reads and samples with <1,000 reads (11) and then clustered into species-level operational taxonomy units (OTU; species-level OTUs or observed species) at 97% identity in the Quantitative Insights Into Microbial Ecology (QIIME 1.8.0) pipeline (12). The sequence data have been submitted to NCBI (BioProject accession number PRJNA316469).

Alpha diversity measures of number of observed species (species-level OTUs), and Shannon index (adjusted for species relative abundance; ref. 13) were estimated by averaging more than 20 rarefied tables (1,000 sequence reads/sample). Beta diversity, which measures the pairwise difference among samples, was estimated as the UniFrac distance, unweighted and also weighted by the relative abundance of each species (14). To rule out the batch effects, 19 random samples were duplicated within and between batches. No

differences within and between batches were found in alpha and beta diversity measures (Supplementary Fig. S2).

Differences in alpha diversity between buccal cells and each of the other sample types were compared by Wilcoxon signed-rank tests. To examine differences in beta diversity, we compared the UniFrac distance within buccal cell samples (within-buccal cell) with the UniFrac distance within each of the other sample types (within-other), and the distance between buccal cells and each of the other sample types (between-groups) by permutation tests (1,000 Monte Carlo repetitions).

To construct the dendrogram for similarity and difference among the oral sample types, Bray–Curtis distance matrices were calculated from the species-level profiles by sample type and then clustered by the complete linkage clustering method. Bootstrap statistics with 1,000 bootstrap samples were calculated to estimate the proportion of samples in which each cluster was identified. A heatmap is prepared to visualize the genus-level profiles and their difference by sample types. To better show difference by samples types in each genus, z-score $[(x - \mu)/SD]$ was calculated for each genus, where x is a particular sample's relative abundance, μ and SD are mean and standard deviation of relative abundance of all samples.

Results

The 41 participants ranged in age from 25 to 66 years; they included 20 African Americans and 21 Caucasians, 22 men, 19 women, 22 current smokers, and 19 never smokers. As reported elsewhere (submitted), essentially no differences in oral microbiome metrics were found by age, race, sex, or smoking status. All 41 participants had successful sequencing (with >1,000 reads per sample) for buccal cell, buccal mucosa, and palatine tonsil samples, whereas samples from the other six oral sites were successfully sequenced in as few as 10 participants for subgingival plaque and up to 37 participants for hard palate (Fig. 1).

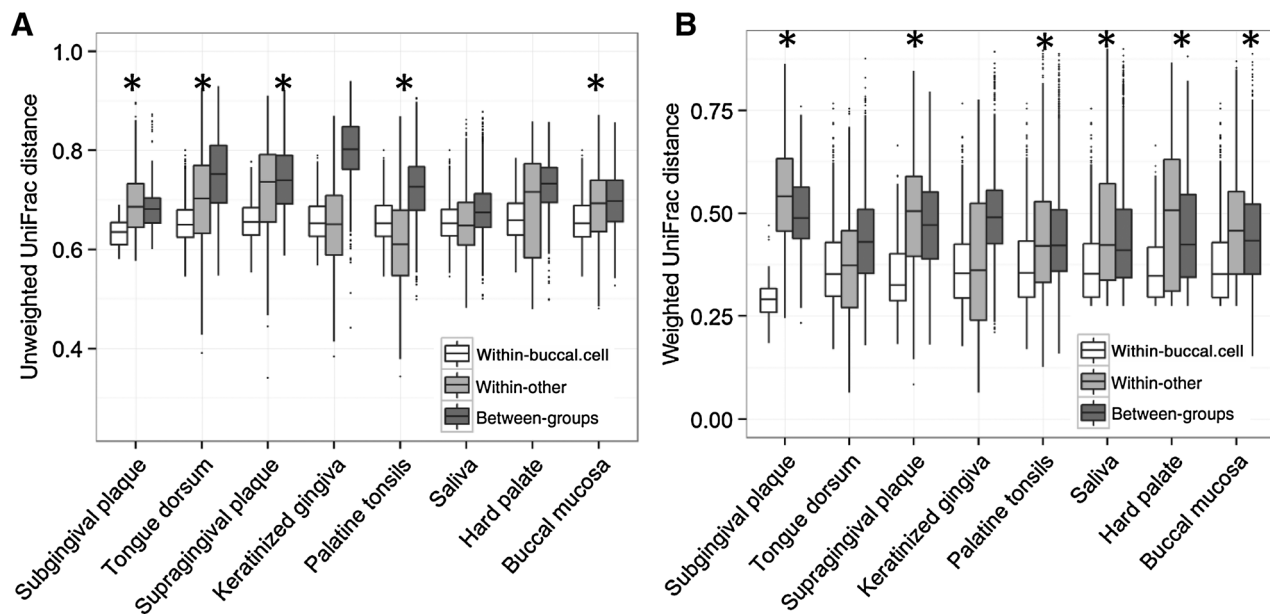


Figure 2.

Comparison of beta diversity measures unweighted (A) and weighted UniFrac distance (B) between buccal cell and 8 other oral sample types. Boxes, interquartile ranges (IQR); median values, bands within the boxes; lines outside the boxes, 1.5 times IQR; dots, outliers. Comparison of distances within buccal cell samples (within-buccal.cell, white boxes), within the other sample types (within-other, light gray boxes), and between buccal cells and the other sample types (between-group, dark gray boxes) were made by 1,000 Monte Carlo permutations (Supplementary Table S2). * $P < 0.05$ for the comparison of within-other distances (light gray boxes) and within-buccal cell distances (white boxes) according to permutation tests with Bonferroni correction. The difference between within-buccal.cell distance (white boxes) and between-group (dark gray boxes) was statistically significant across all shown sample types ($P < 0.05$). Comparisons of any two sample types were made using only subjects with both sample types, which causes variation in the buccal cell bars (white boxes) across comparisons with other sample types.

Across the nine oral sample types, buccal cells had the highest alpha diversity by both observed species (unadjusted for relative abundance) and Shannon index (adjusted for relative abundance), followed by supra- and subgingival plaque, saliva and, at the low end, the five oral soft tissue swabs (Fig. 1). Of the five soft tissue swab types, buccal mucosa had the highest alpha diversity, followed by tongue dorsum, hard palate, palatine tonsils, and keratinized gingiva (Fig. 1). Buccal cells differed from each of the eight other sample types ($P < 0.05$) according to Wilcoxon signed-rank tests with Bonferroni correction for multiple comparisons (Supplementary Table S1).

The microbiota composition within buccal cell samples tended to be more similar across participants, as suggested by relatively smaller within-sample weighted and unweighted UniFrac distances (Fig. 2, white boxes). Specifically, for all sites except keratinized gingiva, palatine tonsils, saliva, and hard palate, within-buccal-cell samples unweighted distance was smaller ($P < 0.05$) than each of the within-other sample type unweighted distances (Fig. 2A, white vs. light gray boxes, Supplementary Table S2). Likewise for weighted distances (Fig. 2B), within-buccal cell samples distance (white boxes) was smaller than each of the within-other sample type (light gray boxes) weighted distances except tongue dorsum and keratinized gingiva.

Compared with UniFrac distances within buccal cell samples, UniFrac distances between buccal cells and each of the other oral sample types were larger ($P < 0.05$; Supplementary Table S2) as indicated in Fig. 2 (white vs. dark gray boxes), implying site-specific differences in microbiota composition.

According to the cladogram based on species-level profiles (Fig. 3, top), the oral sample types formed at least two clusters

with bootstrap support $>70\%$. One cluster included two dental plaques and tongue dorsum. Another included all the other sample types except keratinized gingiva swab. The cladogram suggested that buccal cell samples most closely resembled saliva, then hard palate and buccal mucosa, distinct from dental plaques and tongue dorsum. As shown in the heatmap (Fig. 3) for visualization of the average genus-level profile by sample type, the cluster of buccal cell, saliva, hard palate, and buccal mucosa sample had the genus-level profile different from other sample types. In particular, it featured with relative less abundance of *Streptococcus*.

Discussion

This study found that buccal cell specimens, which were collected in mouthwash following the PLCO cohort protocol, had significantly more observed species, higher alpha diversity, and microbial communities varied less across healthy participants (beta diversity) compared with saliva and with swabs/scrapings from seven other oral sites. Thus, microbiota from buccal cell samples is diverse and relatively stable across healthy subjects. Differences were found in microbial profiles between buccal cell and other oral communities. Some differences were large, as between buccal cells and dental plaques or tongue dorsum, indicating that the microbial profile of buccal cell cannot represent the profiles of all other oral samples. Buccal cell microbiota resembles microbiota from saliva more closely than from other sites. Taken together, our study indicates that the mouthwash-based buccal specimens that have been collected and are being collected in prospective cohort studies have a diverse microbial profile that lends itself to studies of associations with cancer and other diseases.

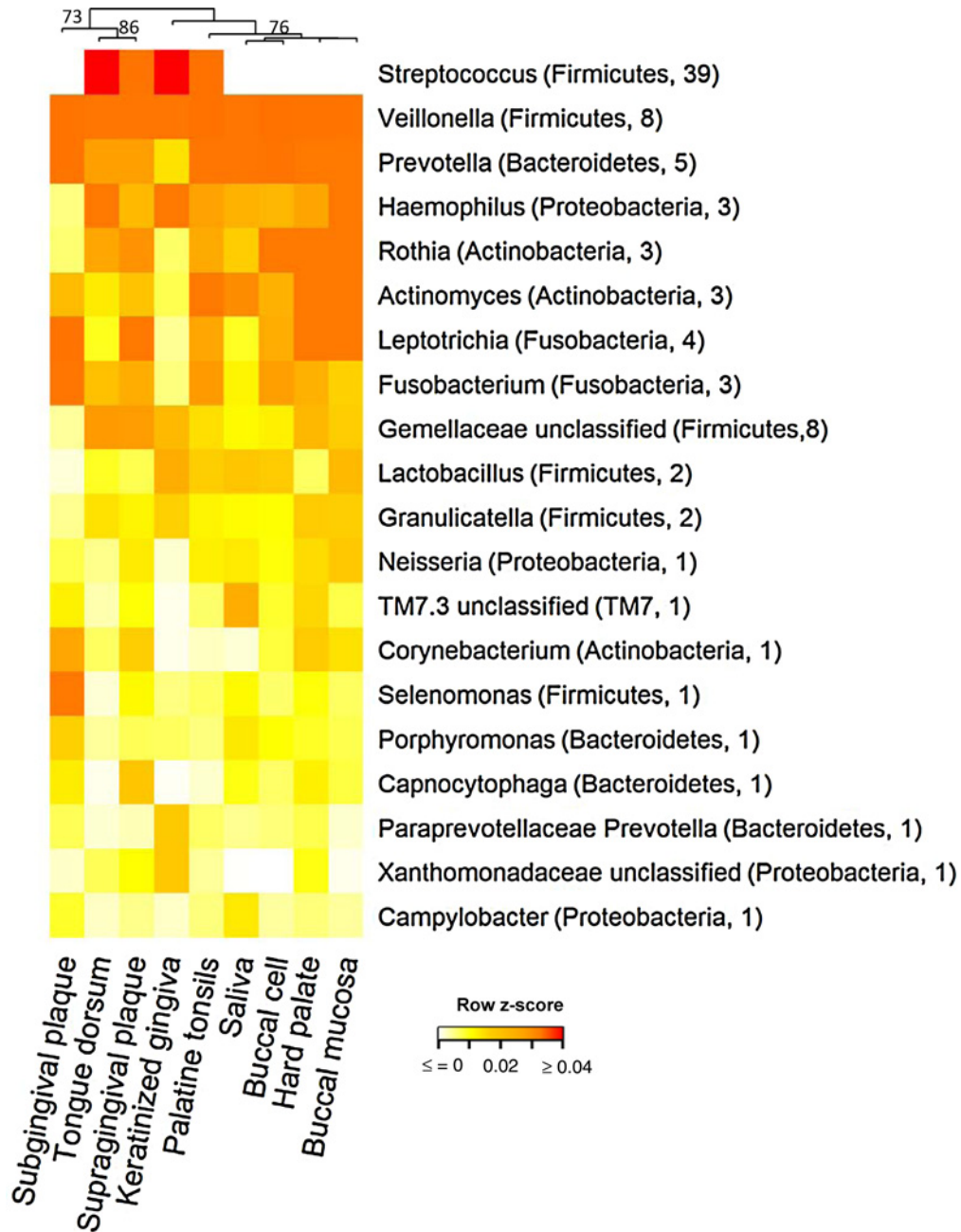


Figure 3.

The dendrogram and heatmap for comparison and visualization of taxonomic profiles among 9 oral sample types. Top, dendrogram shows similarity by sample types based on Bray–Curtis distance of species-level profiles; bootstrap statistics (>70%) indicated the proportion of bootstrap samples in which a given subcluster is identified. Right legend, genus names with their associated phylum names and average relative abundance (%) in the parenthesis. Only the genera with minimum relative abundance 1% were shown. The colors show differences among samples by each genus (row).

Mouthwash collection of buccal cells for human DNA has been employed for more than a decade (15, 16). Thus, there are repositories of buccal cells that can be used to study associations of microbiota with disease. This may prove fruitful, because cancer associations have been established with bacteria in other oral sites. For example, periodontal disease is associated with several aerodigestive malignancies and two periodontal pathogens, *Fusobacterium nucleatum* and *Porphyromonas gingivalis*, that

have been associated with colorectal cancer in cross-sectional studies (6, 17–19). In addition, risk of incident pancreas cancer was increased 2-fold among participants in the European Prospective Investigation into Cancer and Nutrition study who had elevated serum antibodies against *Porphyromonas gingivalis* (20).

The distal gut harbors some 90% of the human microbiota, and methods for collecting and stabilizing feces for microbiome research have been described (21, 22) and are being used for a

wide range of clinical research (23). Promising studies with fecal specimens are only beginning, although it will require some time to achieve adequate prospective follow-up for disease outcome. In the meantime, hypotheses involving cancer and other conditions in relation to the oral microbiome could be developed from retrospective cohort studies with previously collected buccal cells. One could also conduct cross-sectional studies to determine whether variations in buccal cell microbiota are associated with variations in gut microbiota.

A logical extension of the current study is to characterize buccal cell samples that were previously collected during the course of a prospective cohort study. This will permit measurements of microbiota in relation to time of storage, collections and storage characteristics, and diverse covariates. Larger sample sizes and additional repeat studies can help better characterize random variation. We explicitly followed the protocol used by the PLCO cohort, so the available results will generalize to this large representative study. A limitation of our work is that usable sequence data were not obtained from all participants for 6 of the oral sample types. Nonetheless, investigation of the primary objective, the performance of buccal cell specimens for microbiome research, was effectively addressed, because sequencing was successful from buccal cells for all participants.

In summary, microbial DNA was successfully sequenced from buccal cell samples derived from mouthwash. Moreover, buccal cells from healthy patients in a dental clinic had microbial profiles with higher alpha diversity and less variation across participants (beta diversity) compared with other types of oral specimens. Thus, stored buccal cell samples from previous cohorts are a promising resource for studies of oral microbiota to find microbial

associations with disease, although additional studies are needed to establish stability in storage. Sampling from specific oral sites may be required when dysbiosis is postulated to act locally, as may be the case for periodontitis or hairy leukoplakia (24).

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: G. Yu, J.J. Goedert, Y. Ren, N.E. Caporaso

Development of methodology: G. Yu, S. Phillips, M.H. Gail, J. Ravel, N.E. Caporaso

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): G. Yu, S. Phillips, M. Humphrys, J. Ravel, Y. Ren, N.E. Caporaso

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): G. Yu, M.H. Gail, J. Ravel, N.E. Caporaso

Writing, review, and/or revision of the manuscript: G. Yu, M.H. Gail, J.J. Goedert, Y. Ren, N.E. Caporaso

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): G. Yu, S. Phillips, J. Ravel, Y. Ren, N.E. Caporaso

Study supervision: G. Yu, Y. Ren, N.E. Caporaso

Grant Support

This work was supported by intramural research funding at Division of Cancer Epidemiology and Genetics, NCI at the NIH.

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 July 7, 2016; revised September 30, 2016; accepted October 3, 2016; published OnlineFirst October 21, 2016.

References

- Round JL, Mazmanian SK. The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immunol* 2009; 9:313–23.
- Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* 2012;490:55–60.
- Ley RE, Turnbaugh PJ, Klein S, Gordon JJ. Microbial ecology: human gut microbes associated with obesity. *Nature* 2006;444:1022–3.
- Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, et al. A core gut microbiome in obese and lean twins. *Nature* 2009; 457:480–4.
- Foster J, Neufeld KA. Gut-brain axis: How the microbiome influences anxiety and depression. *Int J Neuropsychopharmacol* 2014;17:27–.
- Ahn J, Sinha R, Pei Z, Dominianni C, Wu J, Shi J, et al. Human gut microbiome and risk for colorectal cancer. *J Natl Cancer Inst* 2013;105:1907–11.
- Scher JU, Abramson SB. The microbiome and rheumatoid arthritis. *Nat Rev Rheumatol* 2011;7:569–78.
- Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* 2012;486:207–14.
- 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.
- Fadrosh DW, Ma B, Gajer P, Sengamaly 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;2:6.
- Yu GQ, Fadrosh D, Goedert JJ, Ravel J, Goldstein AM. Nested PCR biases in interpreting microbial community structure in 16S rRNA gene sequence datasets. *PLoS One* 2015;10:e0132253.
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010;7:335–6.
- Shannon CE. The mathematical theory of communication. 1963. *MD Comput* 1997;14:306–17.
- Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. UniFrac: an effective distance metric for microbial community comparison. *ISME J* 2011;5:169–72.
- Garcia-Closas M, Egan KM, Abruzzo J, Newcomb PA, Titus-Ernstoff L, Franklin T, et al. Collection of genomic DNA from adults in epidemiological studies by buccal cytobrush and mouthwash. *Cancer Epidemiol Biomarkers Prev* 2001;10:687–96.
- Feigelson HS, Rodriguez C, Welch R, Hutchinson A, Shao W, Jacobs K, et al. Successful genome-wide scan in paired blood and buccal samples. *Cancer Epidemiol Biomarkers Prev* 2007;16:1023–5.
- Kostic AD, Gevers D, Pedamallu CS, Michaud M, Duke F, Earl AM, et al. Genomic analysis identifies association of *Fusobacterium* with colorectal carcinoma. *Genome Res* 2012;22:292–8.
- Castellarin M, Warren RL, Freeman JD, Dreolini L, Krzywinski M, Strauss J, et al. *Fusobacterium nucleatum* infection is prevalent in human colorectal carcinoma. *Genome Res* 2012;22:299–306.
- Vogtmann E, Goedert JJ. Epidemiologic studies of the human microbiome and cancer. *Br J Cancer* 2016;114:237–42.
- Michaud DS, Izard J, Wilhelm-Benartzi CS, You DH, Grote VA, Tjonneland A, et al. Plasma antibodies to oral bacteria and risk of pancreatic cancer in a large European prospective cohort study. *Gut* 2013;62:1764–70.
- Sinha R, Chen J, Amir A, Vogtmann E, Shi JX, Inman KS, et al. Collecting fecal samples for microbiome analyses in epidemiology studies. *Cancer Epidemiol Biomarkers Prev* 2016;25:407–16.
- Flores R, Shi J, Gail MH, Gajer P, Ravel J, Goedert JJ. Assessment of the human faecal microbiota: II. Reproducibility and associations of 16S rRNA pyrosequences. *Eur J Clin Invest* 2012;42:855–63.
- Shreiner AB, Kao JY, Young VB. The gut microbiome in health and in disease. *Curr Opin Gastroenterol* 2015;31:69–75.
- Goncalves LS, Goncalves BM, Fontes TV. Periodontal disease in HIV-infected adults in the HAART era: clinical, immunological, and microbiological aspects. *Arch Oral Biol* 2013;58:1385–96.