INNOVATIONS JOURNALS

Research Article

Assessment of Distinct Gut Microbiome Signatures in a Diverse Cohort of Patients Undergoing Definitive Treatment for Rectal Cancer

David M. Hein,[1] Laura A. Coughlin,[2] Nicole Poulides,[2] Andrew Y. Koh,[2,3,4] Nina N. Sanford[1,3]

[1]Department of Radiation Oncology, The University of Texas Southwestern Medical Center, Dallas, TX, USA
[2]Department of Pediatrics, Division of Hematology/Oncology, The University of Texas Southwestern Medical Center, Dallas, TX, USA
[3]Harold C. Simmons Comprehensive Cancer Center, The University of Texas Southwestern Medical Center, Dallas, TX, USA
[4]Department of Microbiology, The University of Texas Southwestern Medical Center, Dallas, TX, USA

Nina N. Sanford and Andrew Y. Koh contributed equally to this work.

Address correspondence to Andrew Y. Koh (Andrew.Koh@UTSouthwestern.edu).

Source of Support: Andrew Y. Koh received financial support from the National Institutes of Health (grant K24 AI123163) and the University of Texas Southwestern Medical Center and Children's Health Cellular and Immuno Therapeutics Program. Andrew Y. Koh and Nina N. Sanford received financial support from Harold C. Simmons Comprehensive Cancer Center Early Onset Colorectal Cancer Pilot Grant.

Conflict of Interest: Andrew Y. Koh is a consultant for Prolacta Biosciences, received research funding from Novartis and Prolacta, and is a cofounder of Aumenta. The remaining authors have no disclosures.

Submitted: Aug 25, 2023; First Revision Received: Nov 29, 2023; Accepted: Dec 3, 2023

Hein DM, Coughlin LA, Poulides N, Koh AY, Sanford NN. Assessment of distinct gut microbiome signatures in a diverse cohort of patients undergoing definitive treatment for rectal cancer. *J Immunother Precis Oncol*. Published online. DOI: 10.36401/JIPO-23-30.

ABSTRACT

Introduction: Disparities in incidence and outcome of rectal cancer are multifactorial in etiology but may be due, in part, to differences in gut microbiome composition. We used serial robust statistical approaches to assess baseline gut microbiome composition in a diverse cohort of patients with rectal cancer receiving definitive treatment. **Methods:** Microbiome composition was compared by age at diagnosis (< 50 vs ≥ 50 years), race and ethnicity (White Hispanic vs non-Hispanic), and response to therapy. Alpha diversity was assessed using the Shannon, Chao1, and Simpson diversity measures. Beta diversity was explored using both Bray-Curtis dissimilarity and Aitchison distance with principal coordinate analysis. To minimize false-positive findings, we used two distinct methods for differential abundance testing: LinDA and MaAsLin2 (all statistics two-sided, Benjamini-Hochberg corrected false discovery rate < 0.05). **Results:** Among 64 patients (47% White Hispanic) with median age 51 years, beta diversity metrics showed significant clustering by race and ethnicity ($p < 0.001$ by both metrics) and by onset (Aitchison $p = 0.022$, Bray-Curtis $p = 0.035$). White Hispanic patients had enrichment of bacterial family *Prevotellaceae* (LinDA fold change 5.32, MaAsLin2 fold change 5.11, combined adjusted $p = 0.0007$). No significant differences in microbiome composition were associated with neoadjuvant therapy response. **Conclusion:** We identified distinct gut microbiome signatures associated with race and ethnicity and age of onset in a diverse cohort of patients undergoing definitive treatment for rectal cancer.

Keywords: microbiome, rectal cancer, early onset, healthcare disparities

INTRODUCTION

The increasing incidence of early-onset colorectal cancer (EOCRC), defined as a diagnosis of CRC in patients aged less than 50 years, has become a growing concern over the last four decades.[1] This trend is particularly associated with rectal tumors, with notable racial and ethnic disparities in presentation and outcome.[2] For instance, Black individuals have the highest EOCRC incidence and mortality rates,[3] whereas Hispanic patients, despite overall lower overall incidence, tend to be diagnosed at younger ages compared to non-Hispanic White individuals.[4,5]

The mechanisms behind this increasing incidence and observed disparities are not understood. Genomic studies have not yet identified meaningful molecular differences by age, and heritable conditions only account for a fraction of cases in young patients.[6] This scenario suggests a potential role for other factors, such as the gut microbiome, in the pathogenesis of EOCRC.[7] Previous studies have linked specific microbes with CRC, with some data supporting causality in animal models.[8,9] Regarding age, prior research suggests that younger patients tend to gain "harmful" taxa in contrast to older patients who lose "beneficial" taxa.[10,11] Furthermore, healthy humans have distinct microbiome profiles by race and ethnicity, with one study showing an overabundance of CRC-associated bacteria in Black individuals.[12,13]

However there remains a scarcity of research on microbiome variations by race and ethnicity among patients with cancer, in part due to minority underrepresentation in clinical trials and biospecimen repositories.[14,15] Understanding microbiome differences, such as differential taxa abundance or degree of diversity, could provide insight into the mechanisms underlying the increasing incidence of rectal cancer and the disparities in outcomes. As such, we assessed differences in gut microbiome composition and association with race and ethnicity, age of onset, and treatment outcome among a diverse rectal cancer patient cohort.

METHODS

Patient Cohort

The University of Texas Southwestern institutional review board approved this prospective study, and consent was obtained from patients participating. Adults (age > 18 years) with newly diagnosed stage II–IV rectal adenocarcinoma undergoing definitive intent treatment were included in this study. Specifically, this included patients with limited stage IV disease who were treated with curative local therapy including radiation. Patients who had undergone up-front abdominal perineal resection or diverting colostomy for obstruction were excluded due to the effects of these surgeries on their microbiomes, which could skew analyses. Data collected included cancer TNM stage (AJCC 8th edition), age at diagnosis, sex, self-reported race and ethnicity, body mass index (BMI), treatment regimen, medical oncology facility, detailed antibiotic use, and follow-up magnetic resonance imaging (MRI) and pathology reports. Antibiotic use was stratified into three groups: (1) patients with no antibiotics or a single cefazolin administration more than 90 days prior to stool collection; (2) those with a single cefazolin administration within the past 90 days; and (3) patients who received a broad-spectrum antibiotic active against anaerobes in the previous 180 days. Parkland Health and Hospital System is the sole safety-net health system in Dallas, TX and provides oncology care for a diverse population of uninsured persons in North Texas.

Response Grading

Response was defined primarily pathologically, with radiographic grading used for patients without primary surgery. The pathologic specimen was graded via the modified Ryan scheme for tumor regression, which includes the following four categories: complete response, near complete response, partial response, and poor or no response.[16] For patients who did not undergo primary surgery, response was graded based upon restaging MRI performed after chemotherapy and radiation. MRI tumor regression grade includes the following categories: complete response, near complete response, moderate response, and slight response. The reason for the patient not undergoing primary surgery was also recorded. For the purposes of our study, both radiographic and pathologic response were dichotomized into complete or near complete versus partial, poor, moderate, or slight response.

Stool Collection and Processing

Patients provided stool samples prior to radiation initiation, which were subsequently stored at −80°C. Genomic DNA was extracted from these samples using the MagAttract Power Microbiome DNA/RNA KF kit (Qiagen) and Kingfisher Flex (Thermo Fisher Scientific). From each sample, 16S rRNA genes (variable region 4, V4) were amplified using uniquely bar-coded primers. Polymerase chain reaction (PCR) reactions consisted of Accuprime Pfx Supermix, primers, and template. Following amplification, PCR products were verified, cleaned, and normalized. Pooled samples were sequenced using Illumina MiSeq (PE-250).[17] Postsequencing, raw sequences were quality filtered and primer mismatches or ambiguous bases were removed. Alignment and read count from FASTQ files were conducted with DADA2 in R.[18] Taxa were assigned to the genus level with the Silva nr 99 v138.1 training data, and species were assigned with exact matches to ASVs with the Silva species assignment v138.1 dataset.[19]

Statistical Analysis

Differential abundance (DA) testing was conducted at all taxonomic levels except domain using LinDA and

Table 1. Patient characteristics with breakdown by therapy response

	Complete or Near Complete (*N* = 33, 52%)	**Partial or Poor (*N* = 31, 48%)**	**Overall (*N* = 64)**	***p*-value**
Sex, *n* (%)				
Female	12 (36.4)	14 (45.2)	26 (40.6)	0.611
Male	21 (63.6)	17 (54.8)	38 (59.4)	
Race and Ethnicity, *n* (%)				
Non-Hispanic Asian	4 (12.1)	4 (12.9)	8 (12.5)	0.943
Non-Hispanic Black	3 (9.1)	4 (12.9)	7 (10.9)	
White Hispanic	15 (45.5)	15 (48.4)	30 (46.9)	
Non-Hispanic White	11 (33.3)	8 (25.8)	19 (29.7)	
Onset, *n* (%)				
AOCRC	17 (51.5)	17 (54.8)	34 (53.1)	0.808
EOCRC	16 (48.5)	14 (45.2)	30 (46.9)	
Age (y) at Diagnosis				
Mean (SD)	51.5 (11.3)	52.6 (11.9)	52.1 (11.5)	0.682
Median (range)	51.0 (27.0, 79.0)	51.0 (27.0, 78.0)	51.0 (27.0, 79.0)	
BMI				
Mean (SD)	28.4 (5.55)	28.1 (5.21)	28.2 (5.34)	0.968
Median (range)	28.0 (20.5, 42.2)	28.0 (19.7, 42.3)	28.0 (19.7, 42.3)	
Facility, *n* (%)				
Parkland	18 (54.5)	23 (74.2)	41 (64.1)	0.123
UTSW	15 (45.5)	8 (25.8)	23 (35.9)	
Initial T stage, *n* (%)				
3a–b	21 (63.6)	10 (32.3)	31 (48.4)	0.0424
3c–d	3 (9.1)	6 (19.4)	9 (14.1)	
4	9 (27.3)	15 (48.4)	24 (37.5)	
Initial N stage, *n* (%)				
N0	6 (18.2)	4 (12.9)	10 (15.6)	0.482
N1	11 (33.3)	7 (22.6)	18 (28.1)	
N2	16 (48.5)	20 (64.5)	36 (56.3)	
Initial treatment, *n* (%)				
TNT Long	6 (18.2)	7 (22.6)	13 (20.3)	0.744
TNT Short	15 (45.5)	15 (48.4)	30 (46.9)	
Other	10 (30.3)	6 (19.4)	16 (25.0)	
Preop RT	2 (6.1)	3 (9.7)	5 (7.8)	
Antibiotics, *n* (%)				
None	25 (75.8)	21 (67.7)	46 (71.9)	0.731
Recent (< 90 d) IV cefazolin	3 (9.1)	5 (16.1)	8 (12.5)	
Broad spectrum (< 1 y)	5 (15.2)	5 (16.1)	10 (15.6)	
Detailed Response				
Complete or near complete, *n* (%)				
MRI (watch and wait)	N/A	N/A	14 (31.3)	N/A
Pathological	N/A	N/A	19 (29.7)	
Partial or poor, *n* (%)				
Died	N/A	N/A	2 (3.2)	N/A
MRI lost to follow-up	N/A	N/A	3 (4.7)	
MRI progressed	N/A	N/A	5 (7.8)	
Pathological	N/A	N/A	21 (32.8)	

AOCRC: Average-onset colorectal cancer; BMI: body mass index; EOCRC: early-onset colorectal cancer; IV: intravenous; MRI: magnetic resonance imaging; N/A: not available; Preop RT: preoperative radiotherapy followed by adjuvant chemotherapy; TNT long: total neoadjuvant therapy with neoadjuvant chemotherapy and long-course chemoradiation; TNT short: total neoadjuvant therapy with neoadjuvant chemotherapy and short-course radiotherapy; UTSW: University of Texas Southwestern.

MaAsLin2, with 20% abundance filtering performed at each level.[20–22] Four contrasts were investigated: early vs average onset, White Hispanic vs non-Hispanic race and ethnicity, complete or near complete response vs partial or poor response, and broad-spectrum vs cefazolin vs no antibiotic use. Fold change *p*-values from both methods were combined using the Cauchy combination test[23] and adjusted for multiple testing with the Benjamini-Hochberg method ($p < 0.05$ considered significant). In all DA tests, mixed-effects models were used with sequencing batch as a random effect and antibiotic use as a fixed effect.[24]

Baseline patient characteristics and potential associations with response were assessed using Fisher exact tests or Wilcoxon rank-sum tests. The Wilcoxon rank-sum test additionally evaluated alpha diversity indices (Shannon, Chao1, and Simpson, significance considered at $p < 0.05$). Beta diversity was assessed with principal coordinate

Figure 1. White Hispanic patients who develop colorectal cancer exhibit an enrichment of bacterial family *Prevotellaceae*. All significantly different taxonomic abundances between White Hispanic ($n = 30$) and non-Hispanic ($n = 34$) patients with fold changes and +2 standard errors. Analyzed by LinDA and MaAsLin2 (20% abundance filtering, Benjamini-Hochberg method; p-values combined with Cauchy combination test; adjusted p-value < 0.05 deemed significant).

analysis (PCoA) of both the Bray-Curtis dissimilarity and the robust Aitchison[25] distance (10% abundance filtering, counts adjusted for batch effects using the *ConQuR*[26] package with default settings for logistic LASSO correction including batch size).

Significant clustering on PCoA was assessed through permutational multivariate ANOVA (PERMANOVA), and differences in dispersion were evaluated using the betadispr function, both from the *vegan* package. All permutation tests were performed with 100,000 permutations and all calculations done in R (version 4.2.1). The sequencing data generated in this study are publicly available in the NIH SRA (SUB12910880), and all code for computations available in the public GitHub repository (DavidHein96/microbiome_crc_workflow).

RESULTS

Between October 2020 and August 2022, 64 patients met inclusion criteria for response analyses, including 30 (47%) White Hispanic patients (Table 1), and approximately half (47%) of patients were diagnosed before the age of 50. Median age was 49 years for White Hispanic and 52.5 for non-Hispanic patients ($p = 0.13$, Wilcoxon). Overall, 52% had a complete or near complete response to therapy when categorized binarily including pathologic and radiographic response (Table 1). White Hispanic patients had a significantly higher median BMI than non-Hispanic patients (29.55 vs 25.4, $p = 0.008$).

When comparing race and ethnicity, we found that White Hispanic patients had significant enrichment of

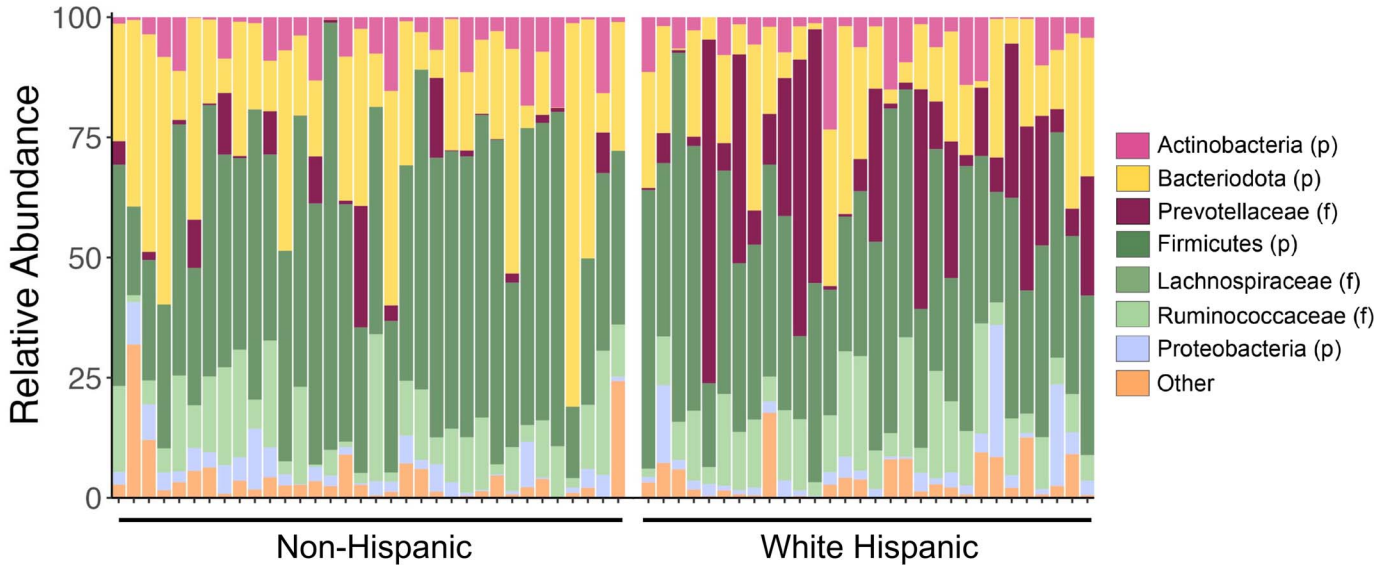

Figure 2. White Hispanic ($n = 30$) vs non-Hispanic ($n = 32$) patients with rectal adenocarcinoma are enriched with *Prevotellaceae*. Histogram of relative abundance of gut bacteria taxonomic groups for the race and ethnicity cohort as determined by 16S rRNA sequencing (V4 region).

bacterial family *Prevotellaceae* (LinDA fold change = 5.32, MaAsLin2 fold change 5.11, combined adjusted $p < 0.001$, Figs. 1 and 2, Supplemental Table S1) compared to non-Hispanic patients. Additionally, both beta diversity metrics showed significant clustering by race and ethnicity (Figs. 3A and B, $p < 0.001$ in both metrics, PERMANOVA). In contrast, there were no differences in alpha diversity (Fig. 4A).

We did not observe any significantly differentially abundant taxa by CRC age of onset category. We did, however, observe significantly lower Shannon diversity in EOCRC patients ($p = 0.029$; Fig. 4B) and both beta diversity metrics showed modest clustering (Aitchison $p = 0.022$, Bray-Curtis $p = 0.035$; Figs. 3C and D).

Variables associated with response were initial T stage ($p = 0.042$, Table 1) and medical oncology facility ($p = 0.123$, Table 1). These variables were included as covariates in the DA analysis for response. However, both before and after controlling for T stage and facility, there were no significant differences in taxa abundance or beta and alpha diversity between complete or near complete and partial or poor responders (Figs. 3E, 3F, and 4C).

Finally, patients with broad spectrum antibiotic use were enriched in family Enterococcaceae (LinDA fold change = 4.15, MaAsLin2 fold change 2.79, combined adjusted $p = 0.005$, reference group no recent antibiotic use, Supplemental Table S2).

DISCUSSION

In this study of baseline microbiome features in rectal cancers receiving definitive therapy, we found clustering of microbiome composition by race and ethnicity with significant enrichment of *Prevotellaceae* among White Hispanic patients. Abundance of *Prevotella*, the major genus within *Prevotellaceae*, has been associated with positive traits, such as plant-based diets and improvements in glucose metabolism.[27,28] However, other studies have documented negative associations with *Prevotellaceae*, including increased risk of inflammatory disease, higher rates of chemotherapy-induced toxicity, and decreased response to chemotherapy in CRC mouse models.[27–31] The various reported effects of *Prevotellaceae* are likely due to their large strain diversity.[32] Related to our findings, a study of Hispanic individuals showed that a higher *Prevotella* to *Bacteroides* ratio was associated with obesity.[33] Interestingly, an analysis of healthy subjects in the American Gut Project did not find *Prevotellaceae* to be differentially abundant across race and ethnicity.[12] In our cohort, White Hispanic patients in our cohort had significantly higher BMI, which is a potential risk factor for colorectal cancers.[34]

We found a modest degree of clustering by beta diversity (comparison of gut microbiome populations between groups) and no significant differences in specific taxa abundance between early- and average-onset patients. A prior meta-analysis suggested that younger patients have increased abundance of "carcinogenic" taxa.[10] Multiple studies have compared microbiome composition in early- versus average-onset CRC to determine whether microbiome profiles can serve as a diagnostic biomarker and have prognostic or predictive value.[10,11,35,36] The results are discordant, with some studies suggesting association of EOCRC with specific bacterial colonization and others showing no differences, including our cohort.

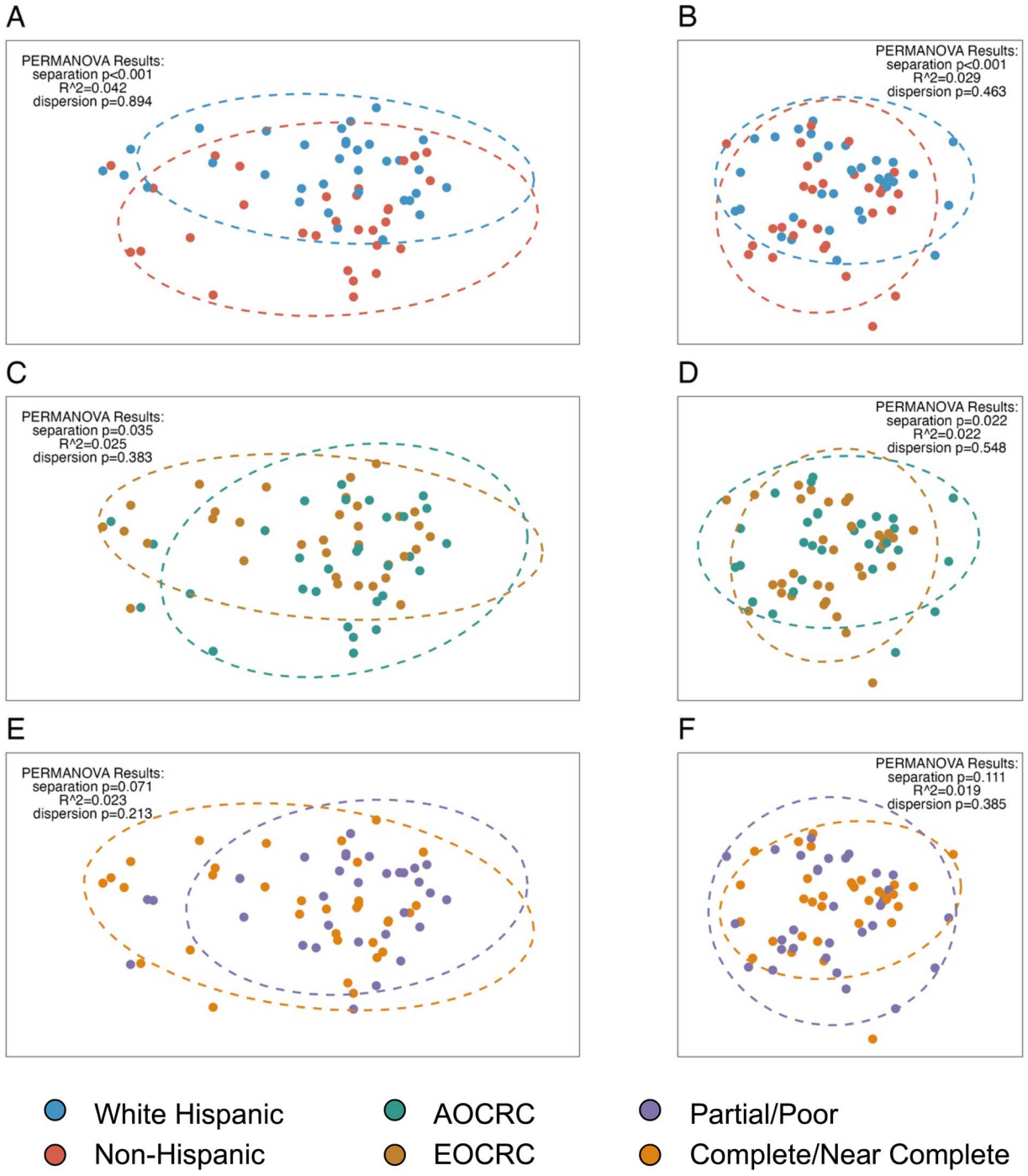

Figure 3. Significant differences in gut microbiome composition (beta diversity) are identified when comparing race and ethnicity and age of onset in patients with colorectal cancer (CRC). Beta diversity analysis of 16S rRNA sequencing data using principal coordinate analysis (PCoA) of both Aitchison distance (right-hand column) and Bray-Curtis dissimilarity (left-hand column). (**A**, **B**) White Hispanic ($n = 30$) vs non-Hispanic ($n = 34$). (**C**, **D**) Average-onset CRC (AOCRC, age > 50 y, $n = 34$) vs early-onset CRC (EOCRC, $n = 30$). (**E**, **F**) Partial or poor responders ($n = 31$) vs complete or near-complete responders ($n = 33$). The proportion of variance accounted by each principal component is indicated. Permutational multivariate ANOVA (PERMANOVA) clustering and differences in dispersion results are indicated, along with a 95% CI (*t* distribution).

Antibiotics can have profound effects on gut microbiome populations. We separated patients who received multiple doses of broad-spectrum antibiotics due to active infection versus those who received only a single intravenous administration of cefazolin as a procedure prophylaxis. We found that exposure to broad-spectrum antibiotics, specifically those that are effective in killing anaerobic taxa, was associated with expansion of *Enterococcaceae*. Interestingly, *Enterococcus* spp. expansion was also observed in adult and pediatric stem cell transplant

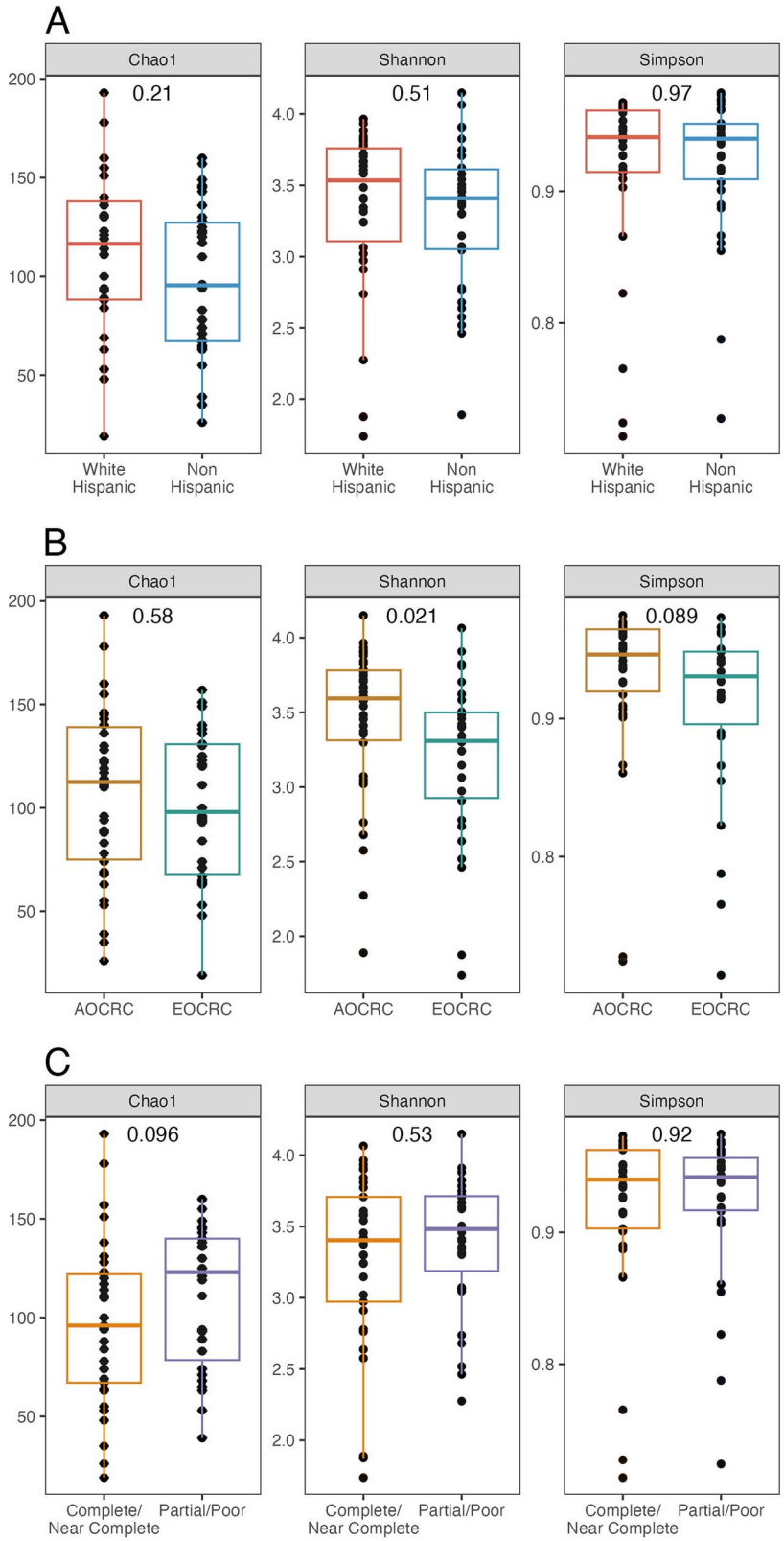

Figure 4. Alpha diversity analysis of 16S rRNA sequencing data. (**A**) White Hispanic ($n = 30$) vs non-Hispanic patients ($n = 34$). (**B**) Average-onset colorectal cancer (AOCRC, age > 50 y, $n = 34$) vs early-onset CRC (EOCRC, $n = 30$). (**C**) Partial or poor responders ($n = 31$) vs complete or near-complete responders ($n = 33$). Differences assessed with Wilcoxon rank sum test, *p*-values shown. Box plots represent 25th, 50th, and 75th percentile plus whisker extension to 1.5× interquartile range. Chao1, Shannon, and Simpson metrics were calculated.

patients who received broad spectrum antibiotics and developed the posttransplant autoimmune complication graft-versus-host disease, a result that was phenocopied in a preclinical GVHD model.[37,38] A recent study demonstrated that an *Enterococcus faecalis*–derived metabolite was able to promote colorectal cancer progression in vitro.[39]

Limitations of this study include modest sample size, use of the shorter-term end point of treatment response,[40] and observational study design. The negative findings on treatment response are complicated by the wide range of treatments received and varying baseline characteristics of our patients. A lack of overlap on and incomplete capture of confounding variables precludes our ability to perform further analysis on treatment response. Future studies with larger sample sizes, longitudinal sampling and microbiome profiling, and use of in vitro and in vivo laboratory approaches to investigate causality are warranted.

Most microbiome studies use a single DA method, which can result in a high false-discovery rate.[41–44] A strength of our study is the reporting of taxa identified as enriched or depleted by both of two DA methods, as well as the use of mixed-effects models to control for antibiotic use and batch effects. The two DA methods chosen demonstrate further robustness in our results as they each use different normalization methods, cumulative sum scaling in MaAsLin2 and centered log ratio with correction for library size bias in LinDA. Furthermore, we used both a traditional (Bray-Curtis dissimilarity) and a compositional (Aitchison distance) beta diversity metric and controlled for batch effects using ConQuR.

CONCLUSION

We identified microbiome composition differences by race and ethnicity in a diverse cohort of patients undergoing definitive treatment for rectal cancer. Future studies are warranted to examine potential mechanisms by which gut microbiome composition may affect CRC carcinogenesis and treatment effect. This could lead the way to therapeutic interventions that improve outcomes, particularly in traditionally understudied and underserved populations.

Supplemental Material

Supplemental materials are available online with the article.

References

1. Sinicrope FA. Increasing Incidence of early-onset colorectal cancer. *N Engl J Med*. 2022;386:1547–1558.
2. Murphy CC, Wallace K, Sandler RS, et al. Racial disparities in incidence of young-onset colorectal cancer and patient survival. *Gastroenterology*. 2019;156:958–965.
3. Siegel RL, Miller KD, Goding Sauer A, et al. Colorectal cancer statistics, 2020. *CA Cancer J Clin*. 2020;70:145–164.
4. Rahman R, Schmaltz C, Jackson CS, et al. Increased risk for colorectal cancer under age 50 in racial and ethnic minorities living in the United States. *Cancer Med*. 2015;4:1863–1870.
5. Garcia S, Pruitt SL, Singal AG, et al. Colorectal cancer incidence among Hispanics and non-Hispanic Whites in the United States. *Cancer Causes Control*. 2018;29:1039–1046.
6. Mork ME, You YN, Ying J, et al. High prevalence of hereditary cancer syndromes in adolescents and young adults with colorectal cancer. *J Clin Oncol*. 2015;33:3544–3549.
7. REACCT Collaborative, Zaborowski AM, Abdile A, et al. Characteristics of early-onset vs late-onset colorectal cancer: a review. *JAMA Surg*. 2021;156:865–874.
8. Kostic AD, Chun E, Robertson L, et al. *Fusobacterium nucleatum* potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe*. 2013;14:207–215.
9. Wirbel J, Pyl PT, Kartal E, et al. Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nat Med*. 2019;25:679–689.
10. Ghosh TS, Das M, Jeffery IB, et al. Adjusting for age improves identification of gut microbiome alterations in multiple diseases. *Elife*. 2020;9:e50240.
11. Abdullah M, Sukartini N, Nursyirwan SA, et al. Gut microbiota profiles in early- and late-onset colorectal cancer: a potential diagnostic biomarker in the future. *Digestion*. 2021;102:823–832.
12. Brooks AW, Priya S, Blekhman R, et al. Gut microbiota diversity across ethnicities in the United States. *PLoS Biol*. 2018;16:e2006842.
13. Farhana L, Antaki F, Murshed F, et al. Gut microbiome profiling and colorectal cancer in African Americans and Caucasian Americans. *World J Gastrointest Pathophysiol*. 2018;9:47–58.
14. Dang JH, Rodriguez EM, Luque JS, et al. Engaging diverse populations about biospecimen donation for cancer research. *J Community Genet*. 2014;5:313–327.
15. Spratt DE, Chan T, Waldron L, et al. Racial/ethnic disparities in genomic sequencing. *JAMA Oncol*. 2016;2:1070–1074.
16. Pai RK, Pai RK. Pathologic assessment of gastrointestinal tract and pancreatic carcinoma after neoadjuvant therapy. *Mod Pathol*. 2018;31:4–23.
17. Kozich JJ, Westcott SL, Baxter NT, et al. 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;79:5112–5120.
18. Callahan BJ, McMurdie PJ, Rosen MJ, et al. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods*. 2016;13:581–583.
19. Quast C, Pruesse E, Yilmaz P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*. 2013;41(database issue):D590–D596.
20. Mallick H, Rahnavard A, McIver LJ, et al. Multivariable association discovery in population-scale meta-omics studies. *PLoS Comput Biol*. 2021;17:e1009442.
21. Zhou H, He K, Chen J, et al. LinDA: linear models for differential abundance analysis of microbiome compositional data. *Genome Biol*. 2022;23:95.

22. Cao Q, Sun X, Rajesh K, et al. Effects of rare microbiome taxa filtering on statistical analysis. *Front Microbiol*. 2020;11:607325.
23. Song H, Ling W, Zhao N, et al. Accommodating multiple potential normalizations in microbiome associations studies. *BMC Bioinformatics*. 2023;24:22.
24. Wang Y, LeCao KA. Managing batch effects in microbiome data. *Brief Bioinform*. 2020;21:1954–1970.
25. Martino C, Morton JT, Marotz CA, et al. A novel sparse compositional technique reveals microbial perturbations. *mSystems*. 2019;4:e00016–e00019.
26. Ling W, Lu J, Zhao N, et al. Batch effects removal for microbiome data via conditional quantile regression. *Nat Commun*. 2022;13:5418.
27. Kovatcheva-Datchary P, Nilsson A, Akrami R, et al. Dietary fiber-induced improvement in glucose metabolism is associated with increased abundance of *Prevotella*. *Cell Metab*. 2015;22:971–982.
28. De Filippis F, Pellegrini N, Vannini L, et al. High-level adherence to a Mediterranean diet beneficially impacts the gut microbiota and associated metabolome. *Gut*. 2016;65:1812–1821.
29. Dillon SM, Lee EJ, Kotter CV, et al. Gut dendritic cell activation links an altered colonic microbiome to mucosal and systemic T-cell activation in untreated HIV-1 infection. *Mucosal Immunol*. 2016;9:24–37.
30. Scher JU, Sczesnak A, Longman RS, et al. Expansion of intestinal *Prevotella copri* correlates with enhanced susceptibility to arthritis. *Elife*. 2013;2:e01202.
31. Yu C, Zhou B, Xia X, et al. *Prevotella copri* is associated with carboplatin-induced gut toxicity. *Cell Death Dis*. 2019;10:714.
32. Ley RE. Gut microbiota in 2015: Prevotella in the gut: choose carefully. *Nat Rev Gastroenterol Hepatol*. 2016;13:69–70.
33. Kaplan RC, Wang Z, Usyk M, et al. Gut microbiome composition in the Hispanic Community Health Study/Study of Latinos is shaped by geographic relocation, environmental factors, and obesity. *Genome Biol*. 2019;20:219.
34. Liu PH, Wu K, Ng K, et al. Association of obesity with risk of early-onset colorectal cancer among women. *JAMA Oncol*. 2019;5:37–44.
35. Mima K, Nishihara R, Qian ZR, et al. *Fusobacterium nucleatum* in colorectal carcinoma tissue and patient prognosis. *Gut*. 2016;65:1973–1980.
36. Yang Y, Du L, Shi D, et al. Dysbiosis of human gut microbiome in young-onset colorectal cancer. *Nat Commun*. 2021;12:6757.
37. Shono Y, Docampo MD, Peled JU, et al. Increased GVHD-related mortality with broad-spectrum antibiotic use after allogeneic hematopoietic stem cell transplantation in human patients and mice. *Sci Transl Med*. 2016;8:339ra71.
38. Simms-Waldrip TR, Sunkersett G, Coughlin LA, et al. Antibiotic-induced depletion of anti-inflammatory Clostridia is associated with the development of graft-versus-host disease in pediatric stem cell transplantation patients. *Biol Blood Marrow Transplant*. 2017;23:820–829.
39. Zhang L, Liu J, Deng M, et al. *Enterococcus faecalis* promotes the progression of colorectal cancer via its metabolite: biliverdin. *J Transl Med*. 2023;21:72.
40. Sell NM, Qwaider YZ, Goldstone RN, et al. Ten-year survival after pathologic complete response in rectal adenocarcinoma. *J Surg Oncol*. 2021;123:293–298.
41. Cappellato M, Baruzzo G, Di Camillo B. Investigating differential abundance methods in microbiome data: A benchmark study. *PLoS Comput Biol*. 2022;18:e1010467.
42. Lin H, Peddada SD. Analysis of microbial compositions: a review of normalization and differential abundance analysis. *NPJ Biofilms Microbiomes*. 2020;6:60.
43. Nearing JT, Douglas GM, Hayes MG, et al. Microbiome differential abundance methods produce different results across 38 datasets. *Nat Commun*. 2022;13:342.
44. Yang L, Chen J. A comprehensive evaluation of microbial differential abundance analysis methods: current status and potential solutions. *Microbiome*. 2022;10:130.