The association of dietary factors and bacterial alpha diversity in the colorectal tumor-associated microbiome

Courtney M. Hill

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2022

Committee:

Amanda I. Phipps

Polly A. Newcomb

Meredith A. J. Hullar

Program Authorized to Offer Degree:

Epidemiology

©Copyright 2022
Courtney M. Hill

University of Washington

Abstract

The association of dietary factors and bacterial alpha diversity in the colorectal tumor-associated microbiome

Courtney M. Hill

Chair of the Supervisory Committee:
Amanda I. Phipps
Department of Epidemiology

Background: Aspects of the gut microbiome, such as microbial diversity, are related to the development and prognosis of colorectal cancer (CRC). Diet has been linked to the gut microbiome but no studies have considered the association of diet and microbial diversity in the tumor-associated microbiome. The goal of this study was to determine if dietary factors are associated with bacterial alpha diversity in CRC tumor tissue.

Methods: Patient-matched CRC tumor and normal tissue samples were obtained from a subset of participants in the Puget Sound Colorectal Cancer Cohort study (n=446). Self-reported dietary factors included vegetable, fruit, red meat, and alcohol intake. Tumor tissue alpha diversity was measured using the Shannon index and was dichotomized two ways: “low” (tumor tissue alpha diversity less than the sample mean) and “depleted” (tumor tissue alpha diversity less than patient-matched adjacent normal tissue alpha diversity). Logistic regression was used to estimate

odds ratios (OR) and 95% confidence intervals (CI) for the association between dietary factors and tumor tissue alpha diversity adjusted for age, sex, and smoking status.

Results: About three-fifths of the sample had depleted tumor tissue alpha diversity. Participants who reported that they had <1 serving of vegetables/day had 2.0 times the odds of having a low tumor alpha diversity (95% CI 1.00, 4.16; p=.05). Higher red meat intake tended to be associated with higher odds of low and depleted tumor tissue alpha diversity, although results were not statistically significant. Fruit intake and alcohol intake were not associated with tumor tissue alpha diversity.

Conclusions: Our study failed to detect an association between self-reported diet and bacterial alpha diversity in the CRC tumor tissue, which may have resulted from the crude measure of diet. Regardless, diet is an important factor in CRC risk and further work is needed to verify the findings of this study using more robust measures of diet.

INTRODUCTION

Colorectal cancer (CRC) is the second most common cause of cancer death in the United States (US) and at least 148,000 individuals were newly diagnosed with CRC in the US in 2020.[1] Several dietary factors have been associated with CRC risk.[2,3] In particular, consumption of vegetables, fruits, and whole grains is associated with lower risk of CRC, while consumption of red meat and processed meat is associated with higher risk.[4]

Dietary factors may plausibly impact CRC risk via an impact on the composition and balance of the gut microbial community. Diet shapes the gut microbiome by selecting for bacteria that can metabolize nutrients made available through the diet.[5,6] Some bacterial metabolites, including vitamins and short-chain fatty acids (SCFA), which are by products of fiber found in vegetables and fruits, can suppress inflammation and cancer.[6–10] While others, such as secondary bile acids, which are associated with red meat consumption, can promote carcinogenesis.[11] Since concentration of these protective or damaging microbial metabolites is influenced by diet, diet may drive the development of CRC through changes to the microbial community (e.g., via enrichment of bacteria that metabolize carcinogenic compounds).

Alpha diversity is one measure of the microbiome community that captures bacterial richness and evenness.[12] Low alpha diversity may be an indicator of microbiome dysbiosis, which can be characterized by changes to the microbial community relative to the microbial community in healthy individuals.[13] Examples include the presence of few unique bacterial taxa and/or depletion or enrichment of select bacterial taxa. At least one prior study has indicated that alpha

diversity is associated CRC stage at diagnosis.[14] However, despite the plausible link between diet, the gut microbiome, and CRC, no work has considered how diet influences alpha diversity in the CRC tumor-associated microbiome.

The goal of this study was to determine if four dietary factors (vegetable, fruit, red meat, and alcohol intake) were associated with bacterial alpha diversity in CRC tumor tissue. We also examined whether dietary factors were associated with depleted alpha diversity in tumor tissue relative to patient-matched adjacent normal tissue. This study contributes to the literature by leveraging 16S rRNA gene sequencing of bacterial genomic DNA from tumor tissue and by using patient-matched adjacent normal tissue as a control.

METHODS

Study population

Data for this study came from the population-based Puget Sound Colorectal Cancer Cohort (PSCCC), which is described in detail elsewhere.[15] The PSCCC includes two successive cohorts of individuals diagnosed with incident CRC from 1998-2007. Eligibility criteria included being aged 18-74 years and residing in western Washington State at the time of CRC diagnosis. Eligible participants were identified via the Surveillance, Epidemiology, and End Results (SEER) cancer registry. The present study is based on a subset of individuals enrolled in the PSCCC for whom both tumor and adjacent normal tissue specimens were available for bacterial DNA extraction and analysis. All study participants included in the present analysis provided

informed consent for the collection and analysis of their study materials. This study was approved by the Institutional Review Board of the Fred Hutchinson Cancer Center.

Data collection

This study draws data from two main sources. Dietary intake data were collected at enrollment into the PSCCC via a diet history questionnaire.[16,17] The questionnaire was completed by telephone interview and participants were asked about average dietary intake from a time approximately two years prior to CRC diagnosis. Vegetable, fruit, and red meat intake were assessed by asking participants, "Would you please tell me how often per day, per week, or per month you ate the following foods?" Alcohol intake was assessed in each decade of life by first asking participants, "Since turning [age]…/Think back to the period when you were in your [age decade]. During that time… Did you ever drink any alcoholic beverages at least once a week for 6 months or longer?" and if yes, "How many alcoholic beverages a week did you typically drink when you consumed at least 1 drink a week for 6 months or longer?"

Tumor and adjacent normal tissue biopsy samples were available from treating hospitals on formalin-fixed paraffin-embedded (FFPE) slides. A DNA extraction protocol for measuring bacterial DNA from FFPE slides was used.[18] The protocol incorporates both mechanical and enzymatic lysis, which is necessary to detect prokaryotic DNA through degradation of bacterial cell walls. For tumor tissue FPPE slides comprised of <80% lesional tissue, macrodissection was used to ensure that tissue was taken from the tumor with a high degree of precision. A microbiome-specific DNA extraction kit was used to maximize bacterial DNA coverage (Qiagen

QIAamp DNA FFPE Tissue Kit and QIAamp DNA Mini Kit). Finally, 16S rRNA gene amplicon sequencing of the V4 region was performed using paired-end reads on the Illumina MiSeq platform. DNA sequences were processed using USEARCH-UNOISE3 (usearch v11.0.667_i86linux64) and QIIME 2. An amplicon sequence variant (ASV) minimum relative abundance filter was applied using the threshold 0.001%. Bacterial taxonomy was assigned using the classify-sklearn module in QIIME2 with the SILVA 138 taxonomy. Samples with less than 10000 counts were removed.

Variables

Dietary factors. There were four dietary factors: vegetable intake, fruit intake, red meat intake, and alcohol intake. Each factor was grouped into three categories. For vegetables and fruits, intake was categorized as <1 serving per day, 1 serving per day, and >1 than one serving per day. Red meat intake was categorized as <2 servings per week, 2-4 servings per week, and >4 servings per week. We used the most recent decade of life to categorize alcohol intake and categories were no drinks per week, 1-6 drink(s) per week, and ≥7 drinks per week.

Alpha diversity. Alpha diversity was assessed using the Shannon diversity index (H).[12] Per study goals, tumor tissue alpha diversity was evaluated two ways for each study participant: 1) low CRC tumor alpha diversity (categorized using the mean for the entire study sample as a cut off) and 2) depleted tumor tissue alpha diversity. Depleted tumor tissue alpha diversity was based on the difference in alpha diversity between tumor and adjacent normal tissue patient-matched samples, where values less than zero indicated the adjacent normal tissue had a greater

alpha diversity. All observations with a difference less than zero were considered to have depleted tumor tissue alpha diversity.

Confounder selection. Confounders were identified as variables hypothesized to contribute to exposure (i.e., dietary factors) and outcome (i.e., alpha diversity in tumor), and that do not lie on the causal pathway. Based on these criteria, the following confounders were selected: age at diagnosis (in ten year categories), sex (male vs. female), and smoking status at time of diagnosis (current smoker, former smoker, never smoker).

Tumor characteristics. Additional tumor clinicopathologic and molecular characteristics variables were included to further describe the study sample. Tumor site and stage at diagnosis were obtained from SEER records. Tumor site was categorized as "right sided," "left sided," or "rectal" based on ICD-O-3 codes: tumors located in the cecum, ascending colon, hepatic flexure and transverse colon were grouped together as right-sided tumors (ICD-O-3 codes were C18.0, C18.2, C18.3, and C18.4 respectively); tumors located in the splenic fixture, descending colon, and sigmoid colon were grouped together as left-sided tumors (ICD-O-3 codes were C18.5, C18.6, and C18.7 respectively); tumors located in the rectosigmoid junction and rectum were grouped together as rectal tumors (ICD-O-3 codes were C19.9 and C20.9 respectively).[19] Stage at diagnosis was classified as "localized," "regional," or "distant" according SEER summary stage definitions. Tumor molecular characteristics were collected by the PSCCC and included microsatellite instability (MSI) status (dichotomized as "microsatellite stable or MSI-low" and "MSI-high"), *BRAF* v600E somatic mutation status, and *KRAS* somatic mutation status, as described elsewhere.[19–22]

Statistical analysis

First, we generated descriptive statistics (mean, standard deviation [SD] or n, %) to describe the study sample. Then, we generated confounder-adjusted odds ratios (OR) and 95% confidence intervals (CIs) to examine the relationship between each dietary factor and tumor tissue alpha diversity using multivariable logistic regression. Dietary factor reference categories were set to levels hypothesized to be associated with higher alpha diversity (i.e., high vegetable and fruit intake and low red meat and alcohol intake). Observations with missing data for the analytical variables were removed for a complete-case analysis (N=446 to 319). Because there were four dietary factors of interest in this study, we used an alpha of 0.0125 (i.e., .50/4) to determine statistical significance. All analyses were conducted in R version 4.03.

Sensitivity Analyses

We performed a number of sensitivity analyses. First, we dichotomized tumor tissue alpha diversity using more extreme cut-offs. Specifically, we conducted an analysis of "very low" tumor tissue alpha diversity to "high" tumor tissue alpha diversity, where we compared the distribution of dietary factors for cases with tumors in the highest vs. lowest quartile of tumor tissue alpha diversity. Similarly, we conducted analyses of "highly depleted" alpha diversity in tumor, where we identified the median level of depletion among cases with tumors defined as having depleted alpha diversity and restricted our analysis of a comparison of those cases above the median level of depletion (vs. those with no depletion).

Second, we modeled tumor tissue alpha diversity as a continuous variable. For alpha diversity depletion, the difference was also modeled as a continuous variable and included normal tissue alpha diversity as a covariate.

Third, we explored the impact of using two other alpha diversity indices: the Chao1 index, a measure of microbial abundance that corrects for variance, and the Simpson Index which captures concentration of different bacterial taxa.[23]

RESULTS

Sample characteristics

Data were available for 446 participants of the PSCCC. About 45% of the sample was under 50 years old at diagnosis, slightly more than half (58%) were female, 12% were current smokers, 43% were former smokers and 45% were never smokers (**Table 1**). Right-sided tumor site was most common (47%) and the majority of participants were diagnosed with regional stage (i.e., stage 2/3) tumors (59%). One-fifth of the sample had MSI-H tumors, 13% had *BRAF*-mutated tumors and 32% had *KRAS*-mutated tumors.

The mean Shannon index for tumor tissue alpha diversity was 4.7 (SD 1.1) and the mean difference in alpha diversity between tumor and adjacent normal tissue was -0.2 (SD 0.9) **(Table**

2). About 60% of the sample had depleted tumor tissue alpha diversity. Twenty-nine percent of the sample reported that they had less than one serving of vegetables per day, 43% reported that they had less than one serving of fruit per day, 41% reported that they had more than 4 servings of red meat per week, and 15% of the sample reported that they had seven or more alcoholic drinks per week.

Association between dietary factors and tumor tissue alpha diversity

After adjusting for confounders and other dietary factors, participants who reported that they had one serving of vegetables per day and less than one serving of vegetables per day had 1.97 and 2.02 times the odds of having a low tumor alpha diversity, respectively (95% CI 1.08, 3.65; p=.03 and 95% CI 1.00 4.16; p=.05) **(Table 3)**. Fruit and vegetable intake associations were less pronounced, and not statistically significant, when examining depleted tumor tissue alpha diversity as an outcome and also not statistically significant. Although failing to reach significance at the alpha 0.0125 level, higher reported red meat intake tended to be associated with higher odds of low and depleted tumor tissue alpha diversity (OR 1.33; 95% CI 0.69, 2.58; p=.39 and OR 1.2; 95% CI 0.62, 2.38; p=.57). Self-reported fruit intake and alcohol intake were not associated with tumor tissue alpha diversity.

Sensitivity analyses

Sensitivity analysis results were consistent with the main analyses when comparing more extreme cut-offs for low tumor tissue alpha diversity and tumor tissue depletion

(Supplementary Table 1) and when evaluating alpha diversity as a continuous measure **(Supplementary Table 2)**. Results were also similar when using the Chao1 and Simpson indices as measures of alpha diversity **(Supplementary Tables 3-5)**

DISCUSSION

The goal of this study was to examine the association between dietary factors and tumor tissue alpha diversity among CRC patients. We examined four self-reported dietary factors: vegetable intake, fruit intake, red meat intake, and alcohol intake with two tumor tissue alpha diversity outcomes: low tumor tissue alpha diversity and depleted tumor tissue alpha diversity relative to patient-matched adjacent normal tissue. While results were not statistically significant, low vegetable intake and higher red meat intake tended to be associated with higher odds of low tumor tissue alpha diversity. Diet is an important CRC risk factor and further work is needed to verify the findings of this study using more robust measures of diet.

This is the first known study that examined the association of self-reported diet and bacterial alpha diversity in CRC tissue and patient-matched adjacent normal tissue. We found that the four self-reported dietary factors of interest in this study were not statistically significantly associated with low tumor alpha diversity. Studies of the normal gut microbiome have found that alpha diversity is influenced by diet, with at least two studies reporting that plasma carotenoid concentration (an objective dietary biomarker for vegetable and fruit intake) is positively associated with alpha diversity.[24,25] Other work, which used a dietary score reflecting overall healthfulness of the diet, found that high intake of plants and fibers is associated with richer gut

microbiome communities.[26] While our findings suggest that self-reported diet was not related to tumor tissue alpha diversity in the sample of CRC patients, null findings may have resulted from the substantial measurement error associated self-reported diet[27] and the sparse dietary intake data used in this study. Additional research, potentially using temporally-relevant dietary biomarkers (that do not rely on self-reported diet), may provide more conclusive findings. Alternatively, our null results may have precipitated from poor timing in study measures. The gut microbial community experiences considerable changes during the development and progression of CRC and any differences in bacterial alpha diversity due to diet may have already been replaced by a microbiome community indicative of CRC development at the time that the tissue samples were collected for this study.

About three-fifths of the CRC participants in this study had depleted tumor tissue alpha diversity relative to patient-matched adjacent normal tissue. The literature has consistently reported that gut microbiome dysbiosis is present in CRC patients when compared to healthy controls and to patient-matched normal tissue,[28,29] so it is not surprising that a majority of the sample in this study had depleted tumor tissue alpha diversity relative to patient-matched adjacent normal tissue. For the remaining 40% of the sample that did not have depleted tumor tissue alpha diversity, it is possible that there were global changes to the gut microbiome that universally impacted diversity in both tumor tissue and adjacent normal tissue. Recent work has identified patient-microbe interactions that predict alpha diversity in CRC patients. Zhao et al.[29] report that due to the heterogeneity of CRC, there are distinct microbial subtypes and only one of the microbial subtypes has reduced tumor tissue microbial alpha diversity compared to adjacent

normal tissues. Further work is needed to identify the patient, tumor, and epidemiological factors that contribute to tumor tissue alpha diversity.

There is strong evidence that dietary factors, including red meat and alcohol intake, are associated with higher risk of CRC.[30,31] On the other hand, while meta-analyses have reported that vegetable and fruit intake are associated with lower CRC risk,[32] the World Cancer Research Fund/American Institute for Cancer Research concludes that the current evidence is not convincing. [30,31] Recent work has hypothesized that the equivocal effects of vegetable and fruit intake could be a result of heterogeneity in CRC molecular subtypes.[33] While the findings from our study fail to support the hypothesis that diet influences the CRC-associated tumor microbiome alpha diversity, it is biologically plausible that diet could increase CRC risk through the enrichment of bacterial taxa that metabolize the carcinogenic or chemopreventive compounds obtained from the diet. For example, bacterial metabolites like SCFA that suppress inflammation and cancer[6–10] and are present in higher concentrations in populations that have a plant and fiber-rich diet.[26] Examining bacterial functional pathways in the development of CRC could clarify this relationship.

This study presents multiple areas for further research. First, we used a crude measure of average self-reported dietary intake from a time two years prior to CRC diagnosis which is subject to multiple biases. Future research may benefit from the incorporation of more robust dietary measures, including objective dietary biomarkers and longitudinal designs that collect diet at multiple time points. Second, while we examined the associations between vegetable, fruit, red meat, and alcohol intake with CRC tumor tissue alpha diversity, other dietary factors may be

relevant. For example, fiber, dietary calcium, and yogurt intake are all protective factors for CRC,[30,31] and should be considered in future work. Lifestyle factors may also play a role in CRC tumor tissue alpha diversity. At least one study has reported that microbiome dysbiosis was not a significant factor in obesity-associated CRC,[34] but other lifestyle factors, like smoking and physical activity also warrant inspection.

Strengths and Limitations

Study findings should be interpreted in light of limitations. We expect some level of non-differential misclassification in the dietary intake measures because obtaining an accurate historical measure of diet is a methodological challenge that is highly susceptible to participant recall. While the cross-sectional data preclude determination of temporality between dietary factors and tumor tissue alpha diversity, participants were asked about dietary information prior to the time at which the tissue samples under study were collected. Tumor and patient-matched adjacent normal samples may have been exposed to contamination given they were stored for up to 20 years which will potentially bias the genomic sequencing results. However, we took steps to address this limitation by using a direct comparison of tumor and adjacent normal tissue samples, which were collected and assayed at the same time and stored under the same conditions.

There are also several strengths to this study. A key strength is the use of both tumor and adjacent patient-matched normal tissue to define alpha diversity depletion. This approach allows for an evaluation of both tumor-specific microbiome characteristics in addition to characteristics

of the microbiome in adjacent normal tissue which can inform whether microbial changes are isolated to the tumor tissue or act globally. We also performed a number of sensitivity analyses to explore different parameterizations of alpha diversity, including an examination of extreme cut-offs and as a continuous variable. This can be used to inform how similar variables are parameterized in future work.

CONCLUSION

In this study, we leveraged 16S rRNA gene sequencing of bacterial genomic DNA from CRC tumor tissue and patient-matched adjacent normal tissue to examine the association between self-reported dietary factors and bacterial alpha diversity. We failed to detect a statistically significant association between self-reported diet and bacterial alpha diversity in CRC tumor tissue. The null results may have resulted from the crude measure of diet. Nevertheless, diet is an important factor in CRC risk and further work is needed to verify the findings of this study using more robust measures of diet in addition to identifying other relevant lifestyle factors.

REFERENCES

1. Siegel RL, Miller KD, Goding Sauer A, et al. Colorectal cancer statistics, 2020. *CA Cancer J Clin*. 2020;70(3):145-164. doi:10.3322/CAAC.21601
2. Aleksandrova K, Pischon T, Jenab M, et al. Combined impact of healthy lifestyle factors on colorectal cancer: A large European cohort study. *BMC Medicine*. 2014;12(1):1-15. doi:10.1186/S12916-014-0168-4/FIGURES/2
3. Carr PR, Weigl K, Jansen L, et al. Healthy Lifestyle Factors Associated With Lower Risk of Colorectal Cancer Irrespective of Genetic Risk. *Gastroenterology*. 2018;155(6):1805-1815.e5. doi:10.1053/J.GASTRO.2018.08.044
4. Schwingshackl L, Schwedhelm C, Hoffmann G, et al. Food groups and risk of colorectal cancer. *Int J Cancer*. 2018;142(9):1748-1758. doi:10.1002/IJC.31198
5. Shanahan F, van Sinderen D, O'Toole PW, Stanton C. Feeding the microbiota: transducer of nutrient signals for the host. *Gut*. 2017;66(9):1709-1717. doi:10.1136/GUTJNL-2017-313872
6. Louis P, Hold GL, Flint HJ. The gut microbiota, bacterial metabolites and colorectal cancer. *Nat Rev Microbiol*. 2014;12(10):661-672. doi:10.1038/NRMICRO3344
7. Neish AS. Microbes in gastrointestinal health and disease. *Gastroenterology*. 2009;136(1):65-80. doi:10.1053/J.GASTRO.2008.10.080
8. Maslowski KM, MacKay CR. Diet, gut microbiota and immune responses. *Nat Immunol*. 2011;12(1):5-9. doi:10.1038/NI0111-5
9. Brestoff JR, Artis D. Commensal bacteria at the interface of host metabolism and the immune system. *Nat Immunol*. 2013;14(7):676-684. doi:10.1038/NI.2640

10. Ben Q, Sun Y, Chai R, Qian A, Xu B, Yuan Y. Dietary fiber intake reduces risk for colorectal adenoma: a meta-analysis. *Gastroenterology*. 2014;146(3). doi:10.1053/J.GASTRO.2013.11.003

11. Ajouz H, Mukherji D, Shamseddine A. Secondary bile acids: an underrecognized cause of colon cancer. *World J Surg Oncol*. 2014;12(1). doi:10.1186/1477-7819-12-164

12. Li K, Bihan M, Yooseph S, Methé BA. Analyses of the microbial diversity across the human microbiome. *PLoS One*. 2012;7(6). doi:10.1371/JOURNAL.PONE.0032118

13. Petersen C, Round JL. Defining dysbiosis and its influence on host immunity and disease. *Cellular Microbiology*. 2014;16(7):1024. doi:10.1111/CMI.12308

14. Ai D, Pan H, Li X, Gao Y, Liu G, Xia LC. Identifying gut microbiota associated with colorectal cancer using a zero-inflated lognormal model. *Frontiers in Microbiology*. 2019;10(APR):826. doi:10.3389/FMICB.2019.00826/BIBTEX

15. Mima K, Cao Y, Chan AT, et al. Fusobacterium nucleatum in Colorectal Carcinoma Tissue According to Tumor Location. *Clin Transl Gastroenterol*. 2016;7(11). doi:10.1038/CTG.2016.53

16. Newcomb PA, Baron J, Cotterchio M, et al. Colon Cancer Family Registry: an international resource for studies of the genetic epidemiology of colon cancer. *Cancer Epidemiol Biomarkers Prev*. 2007;16(11):2331-2343. doi:10.1158/1055-9965.EPI-07-0648

17. Stram DO, Hankin JH, Wilkens LR, et al. Calibration of the Dietary Questionnaire for a Multiethnic Cohort in Hawaii and Los Angeles. *Am J Epidemiol*. 2000;151(4):358. doi:10.1093/OXFORDJOURNALS.AJE.A010214

18. Flores Bueso Y, Walker SP, Tangney M. Characterization of FFPE-induced bacterial DNA damage and development of a repair method. *Biol Methods Protoc*. 2020;5(1):1-14. doi:10.1093/BIOMETHODS/BPAA015

19. Phipps AI, Limburg PJ, Baron JA, et al. Association between molecular subtypes of colorectal cancer and patient survival. *Gastroenterology*. 2015;148(1):77-87.e2. doi:10.1053/J.GASTRO.2014.09.038

20. Phipps AI, Buchanan DD, Makar KW, et al. BRAF mutation status and survival after colorectal cancer diagnosis according to patient and tumor characteristics. *Cancer Epidemiol Biomarkers Prev*. 2012;21(10):1792. doi:10.1158/1055-9965.EPI-12-0674

21. Phipps AI, Buchanan DD, Makar KW, et al. KRAS-mutation status in relation to colorectal cancer survival: the joint impact of correlated tumour markers. *Br J Cancer*. 2013;108(8):1757-1764. doi:10.1038/BJC.2013.118

22. Newcomb PA, Zheng Y, Chia VM, et al. Estrogen plus progestin use, microsatellite instability, and the risk of colorectal cancer in women. *Cancer Res*. 2007;67(15):7534-7539. doi:10.1158/0008-5472.CAN-06-4275

23. Willis AD. Rarefaction, alpha diversity, and statistics. *Frontiers in Microbiology*. 2019;10(OCT):2407. doi:10.3389/FMICB.2019.02407/BIBTEX

24. Schmidt KM, Haddad EN, Sugino KY, et al. Dietary and plasma carotenoids are positively associated with alpha diversity in the fecal microbiota of pregnant women. *J Food Sci*. 2021;86(2):602-613. doi:10.1111/1750-3841.15586

25. Frankenfeld CL, Hullar MAJ, Maskarinec G, et al. The Gut Microbiome Is Associated with Circulating Dietary Biomarkers of Fruit and Vegetable Intake in a Multiethnic Cohort. *J Acad Nutr Diet*. 2022;122(1):78-98. doi:10.1016/J.JAND.2021.05.023

26. Koponen KK, Salosensaari A, Ruuskanen MO, et al. Associations of healthy food choices with gut microbiota profiles. *Am J Clin Nutr*. 2021;114(2):605-616. doi:10.1093/AJCN/NQAB077

27. Ravelli MN, Schoeller DA. Traditional Self-Reported Dietary Instruments Are Prone to Inaccuracies and New Approaches Are Needed. *Frontiers in Nutrition*. 2020;7:90. doi:10.3389/FNUT.2020.00090

28. Gao Z, Guo B, Gao R, Zhu Q, Qin H. Microbiota disbiosis is associated with colorectal cancer. *Frontiers in Microbiology*. 2015;6(FEB). doi:10.3389/FMICB.2015.00020/ABSTRACT

29. Zhao L, Cho WC, Nicolls MR. Colorectal Cancer-Associated Microbiome Patterns and Signatures. *Frontiers in Genetics*. 2021;12:2687. doi:10.3389/FGENE.2021.787176/BIBTEX

30. Veettil SK, Wong TY, Loo YS, et al. Role of Diet in Colorectal Cancer Incidence: Umbrella Review of Meta-analyses of Prospective Observational Studies. *JAMA Network Open*. 2021;4(2):e2037341-e2037341. doi:10.1001/JAMANETWORKOPEN.2020.37341

31. The World Cancer Research Fund (WCRF. World Cancer Research Fund/American Institute for Cancer Research. Diet, Nutrition, Physical Activity and Cancer: a Global Perspective. Continuous Update Project Expert Report 2018. Continuous Update Project Expert Report 2018.

32. Schwingshackl L, Schwedhelm C, Hoffmann G, et al. Food groups and risk of colorectal cancer. *Int J Cancer*. 2018;142(9):1748-1758. doi:10.1002/IJC.31198

33. Hidaka A, Harrison TA, Cao Y, et al. Intake of dietary fruit, vegetables, and fiber and risk of colorectal cancer according to molecular subtypes: A pooled analysis of 9 studies. *Cancer Res*. 2020;80(20):4578. doi:10.1158/0008-5472.CAN-20-0168

34. Greathouse KL, White JR, Padgett RN, et al. Gut microbiome meta-analysis reveals dysbiosis is independent of body mass index in predicting risk of obesity-associated CRC. *BMJ Open Gastroenterology*. 2019;6(1). doi:10.1136/BMJGAST-2018-000247

Table 1. Characteristics of Puget Sound Colorectal Cancer Cohort participants diagnosed with incident colorectal cancer, 1998-2007 (N=446)

Characteristic	N = 446 n (%)
Age at diagnosis (years)	
<40	54 (12%)
40-49	143 (32%)
50-59	68 (15%)
60-69	116 (26%)
70-74	65 (15%)
Sex	
Female	259 (58%)
Male	187 (42%)
Smoking status at diagnosis	
Current	54 (12%)
Former	191 (43%)
Never	201 (45%)
Tumor site	
Left-Sided	118 (27%)
Rectal	117 (27%)
Right-Sided	205 (47%)
Missing	6
Stage at diagnosis	
Distant	62 (14%)
Local	119 (27%)
Regional	261 (59%)
Missing	4
MSI status	
MSI-H	82 (19%)
MSS/MSI-L	341 (81%)
Missing	23
BRAF-mutation status	
Mutated	54 (13%)
Wildtype	366 (87%)
Missing	26
KRAS-mutation status	
Mutated	127 (32%)
Wildtype	271 (68%)
Missing	48

MSI, microsatellite instability; MSI-H, High Microsatellite Instability; MSS, Microsatellite Stable; MSI-L, Low Microsatellite Instability.

Table 2. Colorectal cancer tumor tissue alpha diversity and dietary factors among Puget Sound Colorectal Cancer Cohort participants (N=446)

Characteristic	N = 446 mean (SD) or n (%)
Alpha diversity in tumor tissue[1]	4.72 (1.06)
Low tumor tissue alpha diversity[2]	219 (49%)
Difference in alpha diversity[3]	-0.21 (0.93)
Depleted tumor tissue alpha diversity[4]	274 (61%)
Vegetable intake (servings per day)	
>1	153 (35%)
1	163 (37%)
<1	126 (29%)
Missing	4
Fruit intake (servings per day)	
>1	125 (29%)
1	121 (28%)
<1	187 (43%)
Missing	13
Red meat intake (servings per week)	
<2	85 (20%)
2-4	163 (39%)
>4	174 (41%)
Missing	24
Alcohol intake (drinks per week)	
None	232 (67%)
1-6	63 (18%)
≥7	53 (15%)
Missing	98

SD, standard deviation.

[1]Alpha diversity was assessed using the Shannon diversity index (H).

[2]Low CRC tumor alpha diversity was defined as an alpha diversity less than the sample mean.

[3]Difference in alpha diversity was calculated as the difference in tumor and adjacent normal tissue patient-matched samples alpha diversity, where values less than zero indicated the adjacent normal tissue had a greater alpha diversity.

[4]Depleted tumor tissue alpha diversity was based on the difference in alpha diversity between patient-matched tumor and adjacent normal tissue and defined as observations where the adjacent normal tissue had a greater alpha diversity.

Table 3. Associations between dietary factors and colorectal cancer tumor tissue alpha diversity for Puget Sound Colorectal Cancer Cohort participants

Dietary factor	Low tumor tissue alpha diversity[1]		Depleted tumor tissue alpha diversity[2]	
	OR (95% CI)[3]	p	OR (95% CI)	p
Vegetable intake (servings per day)				
>1	Ref		Ref	Ref
1	1.97 (1.08, 3.65)	.03	0.85 (0.46, 1.57)	.61
<1	2.02 (1.00, 4.16)	.05	1.18 (0.57, 2.47)	.65
Fruit intake (servings per day)				
>1	Ref			
1	0.56 (0.29, 1.09)	.09	1.17 (0.59, 2.30)	.65
<1	0.64 (0.31, 1.29)	.21	1.02 (0.50, 2.07)	.96
Red meat intake (servings per week)				
<2	Ref			
2-4	1.11 (0.58, 2.14)	.75	0.90 (0.46, 1.74)	.76
>4	1.33 (0.69, 2.58)	.39	1.22 (0.62, 2.38)	.57
Alcohol intake (drinks per week)				
None	Ref			
1-6	0.91 (0.47, 1.77)	.78	1.33 (0.68, 2.70)	.41
≥7	1.06 (0.52, 2.18)	.87	1.13 (0.55, 2.39)	.74

CRC, colorectal cancer; OR, odds ratio; CI, confidence interval

[1]Alpha diversity was assessed using the Shannon diversity index (H). Low CRC tumor alpha diversity was defined as an alpha diversity less than the sample mean.

[2]Depleted tumor tissue alpha diversity was based on the difference in alpha diversity between patient-matched tumor and adjacent normal tissue and defined as observations where the adjacent normal tissue had a greater alpha diversity.

[3]Multivariable logistic regression was used to generate odds ratios. All models were adjusted for confounders (patient age at diagnosis, sex, and smoking status at time of diagnosis) and all dietary factors. All observations with missing data were excluded for a complete-case analysis.

Supplementary Table 1. Associations between dietary factors and lowest quartile of colorectal cancer tumor tissue alpha diversity and highly depleted tumor tissue alpha diversity for Puget Sound Colorectal Cancer Cohort participants

Dietary factor	Lowest quartile of tumor tissue alpha diversity[1]		Highly depleted tumor tissue alpha diversity[2]	
	OR (95% CI)[3]	p	OR (95% CI)	p
Vegetable intake (servings per day)				
>1	Ref		Ref	
1	1.12 (0.42, 3.02)	.82	1.62 (0.75, 3.61)	.23
<1	1.34 (0.46, 3.99)	.59	1.19 (0.46, 3.14)	.72
Fruit intake (servings per day)				
>1	Ref		Ref	
1	0.78 (0.27, 2.18)	.63	0.99 (0.42, 2.35)	.99
<1	0.72 (0.24, 2.11)	.56	0.82 (0.32, 2.10)	.67
Red meat intake (servings per week)				
<2	Ref		Ref	
2-4	0.57 (0.20, 1.53)	.27	1.18 (0.51, 2.85)	.70
>4	0.59 (0.21, 1.59)	.30	0.91 (0.39, 2.21)	.82
Alcohol intake (drinks per week)				
None	Ref		Ref	
1-6	0.72 (0.26, 1.94)	.52	1.21 (0.50, 2.82)	.67
≥7	0.62 (0.24, 1.59)	.32	1.02 (0.38, 2.57)	.96

CRC, colorectal cancer; OR, odds ratio; CI, confidence interval

[1]Alpha diversity was assessed using the Shannon diversity index (H). Analysis was conducted among participants with tumors in the highest vs. lowest quartile of tumor tissue alpha diversity. The outcome was binary and coded 1 for observations in the lowest quartile.

[2]In this analysis, we identified the median level of depletion among cases with tumors defined as having depleted alpha diversity and restricted our analysis to a comparison of those cases above the median level of depletion (termed “highly depleted”) vs. those with no depletion.

[3]Multivariable logistic regression was used to generate odds ratios. All models were adjusted for confounders (patient age at diagnosis, sex, and smoking status at time of diagnosis) and all dietary factors. All observations with missing data were excluded for a complete-case analysis.

Supplementary Table 2. Associations between dietary factors and continuous colorectal cancer tumor tissue alpha diversity for Puget Sound Colorectal Cancer Cohort participants

Dietary factor	Tumor tissue alpha diversity[1]		Difference in tumor and normal tissue alpha diversity[2]	
	Beta (95% CI)[3]	p	Beta (95% CI)	p
Vegetable intake (servings per day)				
>1	Ref		Ref	
1	-0.18 (-0.49, 0.14)	.27	0.04 (-0.20, 0.29)	.74
<1	-0.18 (-0.55, 0.20)	.35	-0.03 (-0.32, 0.26)	.85
Fruit intake (servings per day)				
>1	Ref		Ref	
1	0.09 (-0.26, 0.44)	.61	0.03 (-0.24, 0.30)	.84
<1	0.07 (-0.30, 0.44)	.71	-0.09 (-0.38, 0.19)	.52
Red meat intake (servings per week)				
<2	Ref		Ref	
2-4	0.12 (-0.22, 0.47)	.48	0.11 (-0.15, 0.38)	.41
>4	0.09 (-0.25, 0.44)	.60	0.13 (-0.14, 0.39)	.36
Alcohol intake (drinks per week)				
None	Ref		Ref	
1-6	0.12 (-0.23, 0.47)	.50	0.05 (-0.22, 0.32)	.71
≥7	0.21 (-0.16, 0.59)	.26	0.04 (-0.25, 0.33)	.77

CRC, colorectal cancer; CI, confidence interval

[1]Alpha diversity was assessed using the Shannon diversity index (H) and was treated as a continuous variable.

[2] Defined as tumor tissue alpha diversity- normal tissue alpha diversity (i.e., a negative value would indicate that the tumor tissue alpha diversity is depleted relative to the normal tissue alpha diversity).

[3]Multivariable linear regression was used to generate beta. All models were adjusted for confounders (patient age at diagnosis, sex, and smoking status at time of diagnosis) and all dietary factors. The depleted tumor tissue alpha diversity was also adjusted for normal tissue alpha diversity. All observations with missing data were excluded for a complete-case analysis.

Supplementary Table 3. Colorectal cancer tumor tissue alpha diversity measured with the Chao1 and Simpson index among Puget Sound Colorectal Cancer Cohort participants (N=446)

Characteristic	N = 446 mean (SD) or n (%)
Alpha diversity in tumor tissue (Chao1)	402 (190)
Depleted tumor tissue alpha diversity[1] (Chao1)	251 (56%)
Alpha diversity in tumor tissue (Simpson)	0.89 (0.12)
Depleted tumor tissue alpha diversity (Simpson)	248 (56%)

SD, standard deviation.

[1]Depleted tumor tissue alpha diversity was based on the difference in alpha diversity between patient-matched tumor and adjacent normal tissue and defined as observations where the patient-matched adjacent normal tissue had a greater alpha diversity.

Supplementary Table 4. Associations between dietary factors and colorectal cancer tumor tissue alpha diversity as measured by the Chao1 index for Puget Sound Colorectal Cancer Cohort participants

Dietary factor	Low tumor tissue alpha diversity[1]		Depleted tumor tissue alpha diversity[2]	
	OR (95% CI)[3]	p	OR (95% CI)	p
Vegetable intake (servings per day)				
>1	Ref		Ref	
1	0.61 (0.34, 1.10)	.10	1.23 (0.68, 2.23)	.49
<1	0.67 (0.33, 1.35)	.27	1.54 (0.76, 3.14)	.23
Fruit intake (servings per day)				
>1	Ref		Ref	
1	0.90 (0.47, 1.72)	.74	1.07 (0.56, 2.07)	.83
<1	0.80 (0.40, 1.60)	.52	1.03 (0.51, 2.06)	.94
Red meat intake (servings per week)				
<2	Ref		Ref	
2-4	0.85 (0.44, 1.62)	.61	0.86 (0.44, 1.64)	.64
>4	1.13 (0.59, 2.16)	.72	0.7 (0.36, 1.35)	.29
Alcohol intake (drinks per week)				
None	Ref		Ref	
1-6	0.75 (0.38, 1.45)	.39	1.72 (0.88, 3.45)	.12
≥7	1.02 (0.51, 2.06)	.96	1.08 (0.54, 2.19)	.83

CRC, colorectal cancer; OR, odds ratio; CI, confidence interval

[1]Alpha diversity was assessed using the Chao1 index Low CRC tumor alpha diversity was defined as an alpha diversity less than the sample mean.

[2]Depleted tumor tissue alpha diversity was based on the difference in alpha diversity between patient-matched tumor and adjacent normal tissue and defined as observations where the adjacent normal tissue had a greater alpha diversity.

[3]Multivariable logistic regression was used to generate odds ratios. All models were adjusted for confounders (patient age at diagnosis, sex, and smoking status at time of diagnosis) and all dietary factors. All observations with missing data were excluded for a complete-case analysis.

Supplementary Table 5. Associations between dietary factors and colorectal cancer tumor tissue alpha diversity as measured by the Simpson index for Puget Sound Colorectal Cancer Cohort participants

Dietary factor	Low tumor tissue alpha diversity[1]		Depleted tumor tissue alpha diversity[2]	
	OR (95% CI)[3]	p	OR (95% CI)	p
Vegetable intake (servings per day)				
>1	Ref		Ref	
1	1.38 (0.71, 2.69)	.34	0.64 (0.35, 1.16)	.14
<1	1.66 (0.77, 3.66)	.20	0.76 (0.37, 1.54)	.45
Fruit intake (servings per day)				
>1	Ref		Ref	
1	0.71 (0.34, 1.46)	.35	1.35 (0.70, 2.62)	.37
<1	0.71 (0.33, 1.52)	.38	1.35 (0.68, 2.73)	.39
Red meat intake (servings per week)				
<2	Ref		Ref	
2-4	0.72 (0.36, 1.45)	.36	0.91 (0.48, 1.73)	.77
>4	0.63 (0.31, 1.27)	.19	1.24 (0.64, 2.38)	.52
Alcohol intake (drinks per week)				
None	Ref		Ref	
1-6	1.26 (0.60, 2.59)	.53	1.42 (0.73, 2.79)	.30
≥7	1.18 (0.55, 2.48)	.66	0.93 (0.46, 1.88)	.83

CRC, colorectal cancer; OR, odds ratio; CI, confidence interval

[1]Alpha diversity was assessed using the Simpson index. Low CRC tumor alpha diversity was defined as an alpha diversity less than the sample mean.

[2]Depleted tumor tissue alpha diversity was based on the difference in alpha diversity between patient-matched tumor and adjacent normal tissue and defined as observations where the adjacent normal tissue had a greater alpha diversity.

[3]Multivariable logistic regression was used to generate odds ratios. All models were adjusted for confounders (patient age at diagnosis, sex, and smoking status at time of diagnosis) and all dietary factors. All observations with missing data were excluded for a complete-case analysis.