Genetic Predictors for Fecal Propionate and Butyrate-Producing Microbiome Pathway Are Not Associated with Colorectal Cancer Risk: A Mendelian Randomization Analysis

Yujia Lu[1], Yu Chen Zhao[1], Jenny Chang-Claude[2,3], Stephen B. Gruber[4], Andrea Gsur[5], Kenneth Offit[6,7,8], Ludmila Vodickova[9,10,11], Michael O. Woods[12], Long H. Nguyen[13,14], Kaitlin H. Wade[15], Robert Carreras-Torres[16,17], Victor Moreno[16,18,19,20], Daniel D. Buchanan[21,22,23], Michelle Cotterchio[24,25], Andrew T. Chan[13,26,27], Amanda I. Phipps[28,29], Ulrike Peters[28,29], and Mingyang Song[1,13,30]

ABSTRACT

Background: Mechanistic data indicate the benefit of short-chain fatty acids (SCFA) produced by gut microbial fermentation of fiber on colorectal cancer, but direct epidemiologic evidence is limited. A recent study identified SNPs for two SCFA traits (fecal propionate and butyrate-producing microbiome pathway PWY-5022) in Europeans and showed metabolic benefits.

Methods: We conducted a two-sample Mendelian randomization analysis of the genetic instruments for the two SCFA traits (three SNPs for fecal propionate and nine for PWY-5022) in relation to colorectal cancer risk in three large European genetic consortia of 58,131 colorectal cancer cases and 67,347 controls. We estimated the risk of overall colorectal cancer and conducted subgroup analyses by sex, age, and anatomic subsites of colorectal cancer.

Results: We did not observe strong evidence for an association of the genetic predictors for fecal propionate levels and the abundance of PWY-5022 with the risk of overall colorectal cancer, colorectal cancer by sex, or early-onset colorectal cancer (diagnosed at <50 years), with no evidence of heterogeneity or pleiotropy. When assessed by tumor subsites, we found weak evidence for an association between PWY-5022 and risk of rectal cancer (OR per 1-SD, 0.95; 95% confidence intervals, 0.91–0.99; $P = 0.03$) but it did not surpass multiple testing of subgroup analysis.

Conclusions: Genetic instruments for fecal propionate levels and the abundance of PWY-5022 were not associated with colorectal cancer risk.

Impact: Fecal propionate and PWY-5022 may not have a substantial influence on colorectal cancer risk. Future research is warranted to comprehensively investigate the effects of SCFA-producing bacteria and SCFAs on colorectal cancer risk.

Introduction

Colorectal cancer is the third most common cancer and a leading cause of cancer-related deaths in the world (1). Short-chain fatty acids (SCFA), including butyrate, acetate, and propionate, are products of gut bacterial fermentation of dietary fiber and may have anti–colorectal cancer effects through immune and metabolic regulation (1). SCFA-producing gut

[1]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts. [2]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany. [3]University Cancer Center Hamburg (UCCH), Medical Center Hamburg-Eppendorf, Hamburg, Germany. [4]Department of Medical Oncology & Therapeutics Research, City of Hope National Medical Center, Duarte, California. [5]Center for Cancer Research, Medical University of Vienna, Vienna, Austria. [6]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts. [7]Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, New York. [8]Department of Medicine, Weill Cornell Medical College, New York, New York. [9]Department of Molecular Biology of Cancer, Institute of Experimental Medicine of the Czech Academy of Sciences, Prague, Czech Republic. [10]Institute of Biology and Medical Genetics, First Faculty of Medicine, Charles University, Prague, Czech Republic. [11]Biomedical Centre, Faculty of Medicine in Pilsen, Charles University, Pilsen, Czech Republic. [12]Division of Biomedical Sciences, Memorial University of Newfoundland, St. John's, Newfoundland and Labrador, Canada. [13]Division of Gastroenterology, Clinical and Translational Epidemiology Unit, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts. [14]Clinical and Translational Epidemiology Unit, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts. [15]Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom. [16]Colorectal Cancer Group, ONCOBELL Program, Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, Barcelona, Spain. [17]Digestive Diseases and Microbiota Group, Girona Biomedical Research Institute (IDIBGI), Salt, Girona, Spain. [18]Oncology Data Analytics Program, Catalan Institute of Oncology, L'Hospitalet de Llobregat, Barcelona, Spain. [19]Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), Madrid, Spain. [20]Department of Clinical Sciences, Faculty of Medicine, University of Barcelona, Barcelona, Spain. [21]Colorectal Oncogenomic Group, Department of Clinical Pathology, University of Melbourne, Parkville, Victoria, Australia. [22]Victorian Comprehensive Cancer Centre, University of Melbourne Centre for Cancer Research, Parkville, Victoria, Australia. [23]Genetic Medicine and Family Clinic, The Royal Melbourne Hospital, Parkville, Victoria, Australia. [24]Prevention and Cancer Control, Clinical Institutes and Quality Programs, Ontario Health (Cancer Care Ontario), Toronto, Ontario, Canada. [25]Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada. [26]Department of Medicine, Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts. [27]Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, Massachusetts. [28]Department of Epidemiology, University of Washington School of Public Health, Seattle, Washington. [29]Public Health Sciences Division, Fred Hutchinson Cancer Center, Seattle, Washington. [30]Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, Massachusetts.

Correspondence Author: Mingyang Song, Department of Epidemiology, Harvard T.H. Chan School of Public Health, 655 Huntington Avenue, Boston, MA 02115, USA. Phone: 617-432-1301; E-mail: msong@hsph.harvard.edu

Cancer Epidemiol Biomarkers Prev 2023;XX:XX–XX

doi: 10.1158/1055-9965.EPI-22-0861

microbiota, like *Bifidobacterium, Clostridium,* and *Roseburia* species, have been found to be depleted in colorectal cancer cases compared with healthy controls (1–3). Despite these data, however, direct epidemiologic evidence on the association between SCFA and colorectal cancer risk remains limited and inconsistent, and prospective studies on causal relationship is lacking. Recently, an integrated study using bidirectional Mendelian randomization (MR) analysis demonstrated potential causal effects of two SCFA-related microbiome features, including abundance of butyrate-producing microbiome pathway PWY-5022 (4-aminobutanoate degradation V pathway) and fecal propionate levels, on improved insulin sensitivity and lowered risk of type 2 diabetes (4). Using the genetic instruments for PWY-5022 and fecal propionate identified in that study, we conducted a MR study in three large genetic consortia totaling 58,131 colorectal cancer cases and 67,347 controls.

Materials and Methods

Genetic predictors of SCFA levels and data on colorectal cancer

Genetic predictors comprising SNPs for PWY-5022 and fecal propionate (variance explained = 16% and 6.3%, respectively) were identified at $P < 1 \times 10^{-5}$ from a prospective study (**Table 1**; ref. 4). Summary data for the associations of these SNPs with the risk of colorectal cancer overall, early-onset colorectal cancer, colorectal cancer by sex, and subsites (proximal colon, distal colon, and rectum) were obtained from genome-wide association studies (GWAS) in three consortia of European ancestry only, including the ColoRectal Transdisciplinary Study (CORECT), the Colon Cancer Family Registry (CCFR), and studies within the Genetics and Epidemiology of Colorectal Cancer (GECCO) consortium (**Table 1**). Details of the consortia and GWAS data have been described previously (5, 6).

Statistical analysis

We conducted random-effects inverse variance weighted (IVW) tests in a two-sample MR analysis, details of which have been described previously (4). To investigate the robustness of effect estimates to pleiotropy, we conducted sensitivity analyses using additional MR methods, including the weighted median, MR-Egger and MR-PRESSO methods (4). We evaluated the presence of horizontal pleiotropy by the MR-PRESSO Global and MR-Egger intercept tests (4). In the secondary analysis, heterogeneity by sex, age, and anatomic subsites of colorectal cancer was evaluated using the contrast test (7). Cochran's *Q* statistics were used to assess the heterogeneity across individual SNPs (5). All statistical analyses were conducted using the *MendelianRandomization* R package (v0.6.0; ref. 5). For all analyses, *P* values were interpreted as continuous indicators of evidence strength and conclusions were drawn on the basis of effect sizes and their precision. For the secondary analysis, we accounted for multiple testing (7 subgroups * 2 exposures = 14) by using the Bonferroni-corrected α level ($\alpha = 0.05/14 = 0.004$). Power calculation was conducted using an online tool (RRID:SCR_022156) at https://shiny.cnsgenomics.com/mRnd/.

Data availability

The summary statistics used in this study are available within the paper and its Supplementary Data files. The availability of the original GWAS data was described previously (6).

Results

We did not find strong evidence for any association between the genetic predictors for fecal propionate levels and the abundance of PWY-5022 and risk of overall colorectal cancer (**Table 2**). The null association was further confirmed by the null OR using the polygenic

Table 1. Summary of the SNPs associated with the two study traits (abundance of the PWY-5022 pathway and fecal propionate levels) and colorectal cancer risk.

				Association with the trait		Association with colorectal cancer risk	
SCFA trait	dbSNP (or PRS)	Effect allele	Other allele	β (95% CI)[a]	*P* value	OR (95% CI)[b]	*P* value
PWY-5022	rs9423658	C	A	0.33 (0.19–0.48)	5.70E-06	0.99 (0.97–1.02)	0.66
PWY-5022	rs881390	C	T	0.40 (0.23–0.57)	5.70E-06	1.00 (0.97–1.03)	0.99
PWY-5022	rs2089222	A	G	0.56 (0.32–0.79)	3.70E-06	0.99 (0.94–1.03)	0.52
PWY-5022	rs9904981	G	A	0.25 (0.14–0.36)	6.00E-06	0.99 (0.97–1.02)	0.54
PWY-5022	rs10483112	T	C	0.59 (0.34–0.84)	5.20E-06	1.00 (0.94–1.07)	0.90
PWY-5022	rs12994030	T	C	0.24 (0.14–0.34)	3.50E-06	0.99 (0.97–1.01)	0.27
PWY-5022	rs2056208	T	C	0.24 (0.14–0.35)	9.10E-06	1.01 (0.99–1.02)	0.59
PWY-5022	rs10019739	C	T	0.24 (0.13–0.34)	8.20E-06	1.00 (0.98–1.02)	0.74
PWY-5022	rs7743827	G	A	0.27 (0.15–0.38)	4.90E-06	0.99 (0.97–1.02)	0.59
PWY-5022	PRS[c]					1.00 (0.99–1.00)	0.51
Propionate	rs7142308	G	A	0.24 (0.14–0.34)	2.10E-06	0.99 (0.98–1.01)	0.58
Propionate	rs12050534	C	A	0.31 (0.18–0.44)	6.40E-06	1.00 (0.97–1.02)	0.80
Propionate	rs1400566	G	T	−0.22 (−0.31 to −0.12)	9.60E-06	1.00 (0.98–1.02)	0.94
Propionate	PRS[c]					1.00 (0.99–1.01)	0.70

Abbreviations: CI, confidence intervals; CRC, colorectal cancer; OR, odds ratio; PRS, polygenic risk score; SCFA, short-chain fatty acid; SNPs, single nucleotide polymorphisms.

[a]The beta coefficients with 95% CI from regression of the SCFA traits on the genetic variant by a linear mixed model, derived from the genome-wide association analysis in 952 normo-glycemic LifeLines individuals (PWY-5022) and from the genome-wide association analysis in 898 normo-glycemic LL-DEEP individuals for which fecal propionates levels were available.

[b]The OR with 95% CI from regression of colorectal cancer risk on the genetic variant (or the PRS) by a logistic regression additive model after adjusting for age, sex, and study/genotyping project-specific covariates, which were described previously by Huyghe et al. in their Supplementary Table S12,[6] derived from the GWAS data integrating CORECT, the CCFR, and studies within the GECCO consortium.

[c]The PRS for each individual was calculated by summing the product of the beta coefficient associating each SNP with the trait and the number of effect alleles each person carries in each selected SNP. For instance, the score for propionate = 0.24*rs7142308_G + 0.31* rs12050534_C + (-0.22)* rs1400566_G.

Table 2. MR estimates for the associations of PWY-5022 and propionate with risk of colorectal cancer.

SCFA trait	IVW random effects OR (95% CI)	*P* value[a]	Weighted median OR (95% CI)	MR-Egger OR (95% CI)	MR-Egger intercept p[b]	MR-PRESSO OR (95% CI)	MR-PRESSO global p[c]	$P_{heterogeneity}$
PWY-5022								
Colorectal cancer								
All	0.99 (0.96–1.01)	0.31	0.98 (0.95–1.02)	0.99 (0.90–1.10)	0.87	0.99 (0.97–1.00)	0.98	
Male	1.01 (0.97–1.04)	0.74	1.01 (0.96–1.06)	0.98 (0.85–1.12)	0.67	1.01 (0.98–1.04)	0.76	0.16[d]
Female	0.97 (0.93–1.01)	0.09	0.97 (0.92–1.02)	1.00 (0.87–1.16)	0.59	0.97 (0.95–0.99)	0.96	
Age <50	0.97 (0.91–1.03)	0.29	0.96 (0.89–1.04)	1.13 (0.91–1.42)	0.15	0.97 (0.92–1.02)	0.58	
Colon cancer	0.99 (0.96–1.02)	0.55	0.99 (0.95–1.03)	0.96 (0.84–1.08)	0.56	0.99 (0.97–1.01)	0.96	
Distal colon cancer	1.00 (0.95–1.04)	0.86	1.02 (0.96–1.08)	0.95 (0.80–1.11)	0.52	1.00 (0.96–1.04)	0.58	0.35[e]
Proximal colon cancer	0.97 (0.94–1.02)	0.23	0.97 (0.92–1.03)	0.93 (0.79–1.09)	0.56	0.97 (0.94–1.01)	0.69	
Rectal cancer	0.95 (0.91–0.99)	0.03	0.97 (0.91–1.03)	0.89 (0.75–1.07)	0.48	0.95 (0.91–1.00)	0.35	
Propionate								
Colorectal cancer								
All	0.99 (0.95–1.03)	0.67	0.99 (0.94–1.04)	0.97 (0.71–1.33)	0.91	NA*	NA*	
Male	1.01 (0.95–1.08)	0.66	1.02 (0.94–1.09)	0.98 (0.63–1.51)	0.87	NA*	NA*	0.31[d]
Female	0.97 (0.91–1.03)	0.33	0.98 (0.91–1.05)	0.99 (0.63–1.55)	0.93	NA*	NA*	
Age <50	0.96 (0.86–1.07)	0.49	0.98 (0.85–1.13)	0.53 (0.26–1.09)	0.10	NA*	NA*	
Colon cancer	0.99 (0.93–1.04)	0.59	0.99 (0.93–1.05)	0.99 (0.68–1.45)	0.97	NA*	NA*	
Distal colon cancer	0.96 (0.89–1.03)	0.26	0.93 (0.85–1.02)	0.75 (0.43–1.31)	0.39	NA*	NA*	0.76[e]
Proximal colon cancer	0.99 (0.93–1.06)	0.85	1.01 (0.93–1.11)	1.27 (0.77–2.08)	0.33	NA*	NA*	
Rectal cancer	0.99 (0.92–1.06)	0.75	0.99 (0.91–1.08)	1.09 (0.66–1.79)	0.69	NA*	NA*	

Abbreviations: CI, confidence interval; IVW, inverse-variance-weighted; MR, Mendelian randomization; OR, odds ratio, which represent changes in the odds of cancer risk per one standard deviation increase in relative abundance of PWY-5022 or fecal propionates levels; SCFA, short-chain fatty acid.

[a]2-sided *P* value for IVW point estimate.

[b]MR-Egger intercept test (2-sided *P* value).

[c]MR-PRESSO global test (2-sided *P* value).

[d]χ^2 test for heterogeneity (2-sided *P* value) by sex (male vs. female).

[e]χ^2 test for heterogeneity (2-sided *P* value) by subsites (distal colon vs. proximal colon vs. rectal cancer).

*Lack of MR_PRESSO test result for propionate is due to insufficient instrumental variables, which should be at least 4 SNPs.

risk scores (PRS) of the SNPs (**Table 1**). Such null association was unlikely to be due to limited statistical power—based on the strength of the genetic instrument and the large sample size, we expect to have 80% power to detect an OR per SD of 0.96 for the abundance of PWY-5022 and 0.94 for fecal propionate levels in relation to colorectal cancer risk (Supplementary Table S1).

When stratified by tumor subsites, PWY-5022 was weakly associated with a lower risk of rectal cancer [OR per 1-SD, 0.95; 95% confidence interval (CI), 0.91–0.99; $P = 0.03$], but this did not surpass multiple testing of subgroup analysis. There was no strong evidence for associations for early-onset colorectal cancer or colorectal cancer by sex or for heterogeneity by sex or tumor subsites. In addition, there was no evidence of directional pleiotropy for any of the associations by MR-Egger ($P > 0.05$) or MR_PRESSO global test ($P > 0.05$; **Table 2**).

Discussion

Leveraging the large sample size from three genetic consortia, our study indicates that the two SCFA traits were likely not associated with risk of colorectal cancer. Although previous studies support the hypothesis that SCFAs have anti–colorectal cancer risk effects, epidemiologic evidence remains limited and inconsistent. A cross-sectional study showed that fecal levels of acetate, propionate and butyrate were considerably lower in individuals with advanced colorectal adenoma ($n = 344$) compared with healthy controls ($n = 344$; all $P = 0.001$; ref. 3). However, another cross-sectional study reported no difference in fecal concentrations of acetate, propionate, or butyrate across groups of colonic adenomas ($n = 120$), colonic adenomas ($n = 198$), and healthy colons ($n = 172$; all $P > 0.15$; ref. 8). The most plausible explanation for our finding is that PWY-5022 and propionate might not be the predominant pathways underlying the benefit of SCFA for colorectal cancer, although the weak inverse association between PWY-5022 and rectal cancer deserves further investigation. The limitations of this study include that the genetic instruments used for analysis were selected on the basis of the association with SCFA traits and may not be relevance to colorectal cancer. The genetic predictors explain a modest amount of variation in SCFA traits, which are also influenced by other factors such as diet. Our findings do not exclude the possibility that other SCFAs and microbial pathways are associated with colorectal cancer. In conclusion, our findings indicate that genetic predictors for fecal propionate levels and the abundance of butyrate-producing PWY-5022 microbial pathway were not associated with colorectal cancer risk.

Authors' Disclosures

V. Moreno reports grants from Agency for Management of University and Research Grants (AGAUR), Instituto de Salud Carlos III; and grants from Spanish Association Against Cancer (AECC) Scientific Foundation during the conduct of the study. A.T. Chan reports personal fees from Bayer Pharma AG, Boehringer Ingelheim; grants and personal fees from Pfizer Inc.; grants from Zoe Ltd.; and grants from Freenome outside the submitted work. No disclosures were reported by the other authors.

Authors' Contributions

Y. Lu: Formal analysis, investigation, methodology, writing–original draft, writing–review and editing. **Y.C. Zhao:** Formal analysis, investigation, writing–original draft, writing–review and editing. **J. Chang-Claude:** Data

curation, funding acquisition, writing–review and editing. **S.B. Gruber:** Data curation, funding acquisition, writing–review and editing. **A. Gsur:** Data curation, funding acquisition, writing–review and editing. **K. Offit:** Data curation, funding acquisition, writing–review and editing. **L. Vodickova:** Data curation, funding acquisition, writing–review and editing. **M.O. Woods:** Data curation, funding acquisition, writing–review and editing. **L.H. Nguyen:** Writing–review and editing. **K.H. Wade:** Writing–review and editing. **R. Carreras-Torres:** Data curation, funding acquisition, writing–review and editing. **V. Moreno:** Data curation, funding acquisition, writing–review and editing. **D.D. Buchanan:** Data curation, funding acquisition, writing–review and editing. **M. Cotterchio:** Data curation, funding acquisition, writing–review and editing. **A.T. Chan:** Data curation, funding acquisition, writing–review and editing. **A.I. Phipps:** Resources, data curation, funding acquisition, project administration, writing–review and editing. **U. Peters:** Resources, data curation, funding acquisition, project administration, writing–review and editing. **M. Song:** Conceptualization, supervision, funding acquisition, methodology, project administration, writing–review and editing.

Acknowledgments

ASTERISK: We are very grateful to Dr. Bruno Buecher without whom this project would not have existed. We also thank all those who agreed to participate in this study, including the patients and the healthy control persons, as well as all the physicians, technicians, and students.

CCFR: The CCFR graciously thanks the generous contributions of their study participants, dedication of study staff, and the financial support from the NCI, without which this important registry would not exist. The authors would like to thank the study participants and staff of the Seattle CCFR and the Hormones and Colon Cancer study (CORE Studies).

CLUE II: We thank the participants of Clue II and appreciate the continued efforts of the staff at the Johns Hopkins George W. Comstock Center for Public Health Research and Prevention in the conduct of the Clue II Cohort Study.

COLON and NQplus: the authors would like to thank the COLON and NQplus investigators at Wageningen University & Research and the involved clinicians in the participating hospitals.

CORSA: We kindly thank all those who contributed to the screening project Burgenland against colorectal cancer. Furthermore, we are grateful to Doris Mejri and Monika Hunjadi for laboratory assistance.

CPS-II: The authors thank the CPS-II participants and Study Management Group for their invaluable contributions to this research. The authors would also like to acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention National Program of Cancer Registries, and cancer registries supported by the NCI Surveillance Epidemiology and End Results (SEER) program.

Czech Republic CCS: We are thankful to all clinicians in major hospitals in the Czech Republic, without whom the study would not be practicable. We are also sincerely grateful to all patients participating in this study.

DACHS: We thank all participants and cooperating clinicians, and Ute Handte-Daub, Utz Benscheid, Muhabbet Celik, and Ursula Eilber for excellent technical assistance.

EDRN: We acknowledge all the following contributors to the development of the resource: University of Pittsburgh School of Medicine, Department of Gastroenterology, Hepatology and Nutrition: Lynda Dzubinski; University of Pittsburgh School of Medicine, Department of Pathology: Michelle Bisceglia; and University of Pittsburgh School of Medicine, Department of Biomedical Informatics.

EPIC: Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer/World Health Organization.

The EPIC-Norfolk study: we are grateful to all the participants who have been part of the project and to the many members of the study teams at the University of Cambridge who have enabled this research.

EPICOLON: We are sincerely grateful to all patients participating in this study who were recruited as part of the EPICOLON project. We acknowledge the Spanish National DNA Bank, Biobank of Hospital Clínic–IDIBAPS, and Biobanco Vasco for the availability of the samples. The work was carried out (in part) at the Esther Koplowitz Centre, Barcelona.

Harvard cohorts (HPFS, NHS, PHS): The study protocol was approved by the institutional review boards of the Brigham and Women's Hospital and Harvard T.H. Chan School of Public Health, and those of participating registries as required. We acknowledge Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital as home of the NHS.

The authors would like to acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention's National Program of Cancer Registries (NPCR) and/or the NCI's SEER Program. Central registries may also be supported by state agencies, universities, and cancer centers. Participating central cancer registries include the following: Alabama, Alaska, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, Florida, Georgia, Hawaii, Idaho, Indiana, Iowa, Kentucky, Louisiana, Massachusetts, Maine, Maryland, Michigan, Mississippi, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Puerto Rico, Rhode Island, Seattle SEER Registry, South Carolina, Tennessee, Texas, Utah, Virginia, West Virginia, Wyoming. The authors assume full responsibility for analyses and interpretation of these data.

Kentucky: We would like to acknowledge the staff at the Kentucky Cancer Registry.

LCCS: We acknowledge the contributions of Jennifer Barrett, Robin Waxman, Gillian Smith and Emma Northwood in conducting this study.

NCCCS I & II: We would like to thank the study participants, and the NC Colorectal Cancer Study staff.

NSHDS investigators thank the Biobank Research Unit at Umeå University, the Västerbotten Intervention Programme, the Northern Sweden MONICA study and Region Västerbotten for providing data and samples and acknowledge the contribution from Biobank Sweden, supported by the Swedish Research Council (VR 2017-00650).

PLCO: The authors thank the PLCO Cancer Screening Trial screening center investigators and the staff from Information Management Services Inc and Westat Inc. Most importantly, we thank the study participants for their contributions that made this study possible.

SEARCH: We thank the SEARCH team.

SELECT: We thank the research and clinical staff at the sites that participated on SELECT study, without whom the trial would not have been successful. We are also grateful to the 35,533 dedicated men who participated in SELECT.

UK Biobank: We would like to thank the participants and researchers UK Biobank for their participation and acquisition of data.

WHI: The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: http://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Short%20List.pdf.

ASTERISK: a Hospital Clinical Research Program (PHRC-BRD09/C) from the University Hospital Center of Nantes (CHU de Nantes) and supported by the Regional Council of Pays de la Loire, the Groupement des Entreprises Françaises dans la Lutte contre le Cancer (GEFLUC), the Association Anne de Bretagne Génétique and the Ligue Régionale Contre le Cancer (LRCC).

The ATBC Study is supported by the Intramural Research Program of the NCI, NIH, Department of Health and Human Services.

Maryland Cancer Registry (MCR) Cancer data was provided by the Maryland Cancer Registry, Center for Cancer Prevention and Control, Maryland Department of Health, with funding from the State of Maryland and the Maryland Cigarette Restitution Fund. The collection and availability of cancer registry data is also supported by the Cooperative Agreement NU58DP006333, funded by the Centers for Disease Control and Prevention. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention or the Department of Health and Human Services.

ColoCare: This work was supported by the NIH grant numbers R01 CA189184 (Li/Ulrich), U01 CA206110 (Ulrich/Li/Siegel/Figueiredo/Colditz, 2P30CA015704- 40 (Gilliland), R01 CA207371 (Ulrich/Li)), the Matthias Lackas-Foundation, the German Consortium for Translational Cancer Research, and the EU TRANSCAN initiative.

The CCFR (www.coloncfr.org) is supported in part by funding from the NCI, NIH (award U01 CA167551). Support for case ascertainment was provided in part from the SEER Program and the following U.S. state cancer registries: AZ, CO, MN, NC, NH; and by the Victoria Cancer Registry (Australia) and Ontario Cancer Registry (Canada). The CCFR Set-1 (Illumina 1M/1M-Duo) and Set-2 (Illumina Omni1-Quad) scans were supported by NIH awards U01 CA122839 and R01 CA143247 (to GC). The CCFR Set-3 (Affymetrix Axiom CORECT Set array) was supported by NIH award U19 CA148107 and R01 CA81488 (to S.B. Gruber). The CCFR Set-4 (Illumina OncoArray 600K SNP array) was supported by NIH award U19 CA148107 (to S.B. Gruber) and by the Center for Inherited Disease Research (CIDR), which is funded by the NIH to the Johns Hopkins University, contract

number HHSN2682012000008I. Additional funding for the OFCCR/ARCTIC was through award GL201-043 from the Ontario Research Fund, award 112746 from the Canadian Institutes of Health Research, through a Cancer Risk Evaluation (CaRE) Program grant from the Canadian Cancer Society (to S.B. Gruber), and through generous support from the Ontario Ministry of Research and Innovation. The SFCCR Illumina HumanCytoSNP array was supported in part through NCI/NIH awards U01/U24 CA074794 and R01 CA076366. The content of this manuscript does not necessarily reflect the views or policies of the NCI, NIH or any of the collaborating centers in the CCFR, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government, any cancer registry, or the CCFR.

COLON: The COLON study is sponsored by Wereld Kanker Onderzoek Fonds, including funds from grant 2014/1179 as part of the World Cancer Research Fund International Regular Grant Programme, by Alpe d'Huzes and the Dutch Cancer Society (UM 2012-5653, UW 2013-5927, UW2015-7946), and by TRANSCAN (JTC2012-MetaboCCC, JTC2013-FOCUS). The Nqplus study is sponsored by a ZonMW investment grant (98-10030); by PREVIEW, the project PREVention of diabetes through lifestyle intervention and population studies in Europe and around the World (PREVIEW) project which received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant no. 312057; by funds from TI Food and Nutrition (cardiovascular health theme), a public–private partnership on precompetitive research in food and nutrition; and by FOODBALL, the Food Biomarker Alliance, a project from JPI Healthy Diet for a Healthy Life.

Colorectal Cancer Transdisciplinary (CORECT) Study: The CORECT Study was supported by the NCI/NIH, U.S. Department of Health and Human Services (grant numbers U19 CA148107, R01 CA081488, P30 CA014089, R01 CA197350; P01 CA196569; R01 CA201407; R01 CA242218), National Institutes of Environmental Health Sciences, NIH (grant number T32 ES013678), and a generous gift from Daniel and Maryann Fong.

CORSA: The CORSA study was funded by Austrian Research Funding Agency (FFG) BRIDGE (grant 829675, to Andrea Gsur), the "Herzfelder'sche Familienstiftung" (grant to Andrea Gsur) and was supported by COST Action BM1206.

CPS-II: The American Cancer Society funds the creation, maintenance, and updating of the Cancer Prevention Study-II (CPS-II) cohort. This study was conducted with Institutional Review Board approval.

CRCGEN: Colorectal Cancer Genetics & Genomics, Spanish study was supported by Instituto de Salud Carlos III, co-funded by FEDER funds –a way to build Europe– (grants PI14-613 and PI09-1286), Agency for Management of University and Research Grants (AGAUR) of the Catalan Government (grant 2017SGR723), Junta de Castilla y León (grant LE22A10-2), the Spanish Association Against Cancer (AECC) Scientific Foundation grant GCTRA18022MORE and the Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), action Genrisk. Sample collection of this work was supported by the Xarxa de Bancs de Tumors de Catalunya sponsored by Pla Director d'Oncologia de Catalunya (XBTC), Plataforma Biobancos PT13/0010/0013 and ICOBIOBANC, sponsored by the Catalan Institute of Oncology. We thank CERCA Programme, Generalitat de Catalunya for institutional support.

Czech Republic CCS: This work was supported by the Czech Science Foundation 21-27902S, by National operation Program: National Institute for cancer research LX22NPO05102, by Operational Program Integrated Infrastructure for the project: Integrative strategy in development of personalized medicine of selected malignant tumors and its impact on quality of life, IMTS: 313011V446, co-financed by the European Regional Development Fund and by the Charles University Research Fund (Cooperatio No. 43 -Surgical Disciplines).

DACHS: This work was supported by the German Research Council (BR 1704/6-1, BR 1704/6-3, BR 1704/6-4, CH 117/1-1, HO 5117/2-1, HE 5998/2-1, KL 2354/3-1, RO 2270/8-1 and BR 1704/17-1), the Interdisciplinary Research Program of the National Center for Tumor Diseases (NCT), Germany, and the German Federal Ministry of Education and Research (01KH0404, 01ER0814, 01ER0815, 01ER1505A and 01ER1505B).

DALS: NIH (R01 CA48998 to M.L. Slattery).

EDRN: This work is funded and supported by the NCI, EDRN Grant (U01 CA 84968-06).

EPIC: The coordination of EPIC is financially supported by International Agency for Research on Cancer (IARC) and also by the Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London which has additional infrastructure support provided by the NIHR Imperial Biomedical Research Centre (BRC). The national cohorts are supported by: Danish Cancer Society (Denmark); Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Education Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research Center (DKFZ), German Institute of Human Nutrition Potsdam- Rehbruecke (DIfE), Federal Ministry of Education and Research (BMBF) (Germany); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy, Compagnia di SanPaolo and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands); Health Research Fund (FIS) - Instituto de Salud Carlos III (ISCIII), Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, and the Catalan Institute of Oncology - ICO (Spain); Swedish Cancer Society, Swedish Research Council and County Councils of Skåne and Västerbotten (Sweden); Cancer Research UK (14136 to EPIC-Norfolk; C8221/A29017 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk; MR/M012190/1 to EPIC-Oxford). (United Kingdom).M

EPICOLON: This work was supported by grants from Fondo de Investigación Sanitaria/FEDER (PI08/0024, PI08/1276, PS09/02368, P111/00219, PI11/00681, PI14/00173, PI14/00230, PI17/00509, 17/00878, PI20/00113, PI20/00226, Acción Transversal de Cáncer), Xunta de Galicia (PGIDIT07PXIB9101209PR), Ministerio de Economia y Competitividad (SAF07-64873, SAF 2010-19273, SAF2014-54453R), Fundación Científica de la Asociación Española contra el Cáncer (GCB13131592CAST), Beca Grupo de Trabajo "Oncología" AEG (Asociación Española de Gastroenterología), Fundación Privada Olga Torres, FP7 CHIBCHA Consortium, Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR, Generalitat de Catalunya, 2014SGR135, 2014SGR255, 2017SGR21, 2017SGR653), Catalan Tumour Bank Network (Pla Director d'Oncologia, Generalitat de Catalunya), PERIS (SLT002/16/00398, Generalitat de Catalunya), CERCA Programme (Generalitat de Catalunya) and COST Action BM1206 and CA17118. CIBERehd is funded by the Instituto de Salud Carlos III.

ESTHER/VERDI. This work was supported by grants from the Baden-Württemberg Ministry of Science, Research and Arts and the German Cancer Aid.

Harvard cohorts: HPFS is supported by the NIH (P01 CA055075, UM1 CA167552, U01 CA167552, R01 CA137178, R01 CA151993, and R35 CA197735), NHS by the NIH (P01 CA087969, UM1 CA186107, R01 CA137178, R01 CA151993, and R35 CA197735), and PHS by the NIH (R01 CA042182).

Hawaii Adenoma Study: NCI grants R01 CA72520.

HCES-CRC: the Hwasun Cancer Epidemiology Study–Colon and Rectum Cancer (HCES-CRC; grants from Chonnam National University Hwasun Hospital, HCRI15011-1).

Kentucky: This work was supported by the following grant support: Clinical Investigator Award from Damon Runyon Cancer Research Foundation (CI-8); NCI R01CA136726.

LCCS: The Leeds Colorectal Cancer Study was funded by the Food Standards Agency and Cancer Research UK Programme Award (C588/A19167).

MCCS cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further supported by Australian NHMRC grants 509348, 209057, 251553 and 504711 and by infrastructure provided by Cancer Council Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry (VCR) and the Australian Institute of Health and Welfare (AIHW), including the National Death Index and the Australian Cancer Database.

MEC: NIH (R37 CA54281, P01 CA033619, and R01 CA063464).

MECC: This work was supported by the NIH, U.S. Department of Health and Human Services (R01 CA081488, R01 CA197350, U19 CA148107, R01 CA242218, and a generous gift from Daniel and Maryann Fong.

MSKCC: The work at Sloan Kettering in New York was supported by the Robert and Kate Niehaus Center for Inherited Cancer Genomics and the Romeo Milio Foundation. Moffitt: This work was supported by funding from the NIH (grant numbers R01 CA189184, P30 CA076292), Florida Department of Health Bankhead-Coley Grant 09BN-13, and the University of South Florida Oehler Foundation. Moffitt contributions were supported in part by the Total Cancer Care Initiative, Collaborative Data Services Core, and Tissue Core at the H. Lee Moffitt Cancer Center & Research Institute, a NCI-designated Comprehensive Cancer Center (grant number P30 CA076292).

NCCCS I & II: We acknowledge funding support for this project from the NIH, R01 CA66635 and P30 DK034987.

NFCCR: This work was supported by an Interdisciplinary Health Research Team award from the Canadian Institutes of Health Research (CRT 43821); the NIH, U.S. Department of Health and Human Serivces (U01 CA74783); and NCI of Canada grants (18223 and 18226). The authors wish to acknowledge the contribution of Alexandre Belisle and the genotyping team of the McGill University and Génome Québec Innovation Centre, Montréal, Canada, for genotyping the Sequenom panel in

the NFCCR samples. Funding was provided to Michael O. Woods by the Canadian Cancer Society Research Institute.

NSHDS: The research was supported by Biobank Sweden through funding from the Swedish Research Council (VR 2017-00650, VR 2017-01737), the Swedish Cancer Society (CAN 2017/581), Region Västerbotten (VLL-841671, VLL-833291), Knut and Alice Wallenberg Foundation (VLL-765961), and the Lion's Cancer Research Foundation (several grants) and Insamlingsstiftelsen, both at Umeå University.

OSUMC: OCCPI funding was provided by Pelotonia and HNPCC funding was provided by the NCI (CA16058 and CA67941).

PLCO: Intramural Research Program of the Division of Cancer Epidemiology and Genetics and supported by contracts from the Division of Cancer Prevention, NCI, NIH, DHHS. Funding was provided by NIH, Genes, Environment and Health Initiative (GEI) Z01 CP 010200, NIH U01 HG004446, and NIH GEI U01 HG 004438.

SEARCH: The University of Cambridge has received salary support in respect of PDPP from the NHS in the East of England through the Clinical Academic Reserve. Cancer Research UK (C490/A16561); the UK National Institute for Health Research Biomedical Research Centres at the University of Cambridge.

SELECT: Research reported in this publication was supported in part by the NCI of the NIH under Award Numbers U10 CA37429 (CD Blanke), and UM1 CA182883 (CM Tangen/IM Thompson). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

SMS and REACH: This work was supported by the NCI (grant P01 CA074184 to J.D.P. and P.A.N., grants R01 CA097325, R03 CA153323, and K05 CA152715 to P.A.N., and the National Center for Advancing Translational Sciences at the NIH (grant KL2 TR000421 to A.N.B.-H.)

The Swedish Low-risk Colorectal Cancer Study: The study was supported by grants from the Swedish research council; K2015-55X-22674-01-4, K2008-55X-20157-03-3, K2006-72X-20157-01-2 and the Stockholm County Council (ALF project).

Swedish Mammography Cohort and Cohort of Swedish Men: This work is supported by the Swedish Research Council /Infrastructure grant, the Swedish Cancer Foundation, and the Karolinska Institutet́s Distinguished Professor Award to Alicja Wolk.

UK Biobank: This research has been conducted using the UK Biobank Resource under Application Number 8614.

VITAL: NIH (K05 CA154337).

WHI: The WHI program is funded by the National Heart, Lung, and Blood Institute, NIH, U.S. Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C.

Genotyping/Sequencing services were provided by the Center for Inherited Disease Research (CIDR) contract number HHSN268201700006I and HHSN268201200008I.

The content of this manuscript does not necessarily reflect the views or policies of the NCI, NIH or any of the collaborating centers in the CCFR, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government, any cancer registry, or the CCFR.

This study is supported by: NIH (U01CA261961 to U. Peters).

GECCO Consortium: NCI, NIH, U.S. Department of Health and Human Services (U01 CA137088, R01 CA059045, U01 CA164930, R21 CA191312, R01201407 to U. Peters).

NIH/NCI Cancer Center Support Grant P30 CA015704 (to U. Peters). Scientific Computing Infrastructure at Fred Hutch funded by ORIP grant S10OD028685 (to U. Peters).

CLUE II funding was from the NCI (U01 CA86308, Early Detection Research Network; P30 CA006973 to U. Peters), National Institute on Aging (U01 AG18033 to U. Peters), and the American Institute for Cancer Research. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US government.

The publication costs of this article were defrayed in part by the payment of publication fees. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

Note

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

Received August 8, 2022; revised October 14, 2022; accepted December 6, 2022; published first December 13, 2022.

References

1. Song M, Chan AT, Sun J. Influence of the gut microbiome, diet, and environment on risk of colorectal cancer. Gastroenterology 2020;158:322–40.
2. Thomas AM, Manghi P, Asnicar F, Pasolli E, Armanini F, Zolfo M, et al. Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. Nat Med 2019;25:667–78.
3. Chen HM, Yu YN, Wang JL, Lin YW, Kong X, Yang CQ, et al. Decreased dietary fiber intake and structural alteration of gut microbiota in patients with advanced colorectal adenoma. Am J Clin Nutr 2013;97:1044–52.
4. Sanna S, van Zuydam NR, Mahajan A, Kurilshikov A, Vila AV, Võsa U, et al. Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. Nat Genet 2019;51:600–5.
5. Murphy N, Song M, Papadimitriou N, Carreras-Torres R, Langenberg C, Martin RM, et al. Associations between glycemic traits and colorectal cancer: a Mendelian randomization analysis. J Natl Cancer Inst 2022;114:740–52.
6. Huyghe JR, Bien SA, Harrison TA, Kang HM, Chen S, Schmit SL, et al. Discovery of common and rare genetic risk variants for colorectal cancer. Nat Genet 2019;51:76–87.
7. Wang M, Spiegelman D, Kuchiba A, Lochhead P, Kim S, Chan AT, et al. Statistical methods for studying disease subtype heterogeneity. Stat Med 2016;35:782–800.
8. Sze MA, Topçuoğlu BD, Lesniak NA, Ruffin MT, Schloss PD, Fraser CM. Fecal Short-chain fatty acids are not predictive of colonic tumor status and cannot be predicted based on bacterial community structure. mBio 2019;10:e01454–01419.