OXFORD

*JNCI J Natl Cancer Inst* (2021) 113(1): djaa058

doi: 10.1093/jnci/djaa058
First published online September 3, 2020
Article

# Association of Body Mass Index With Colorectal Cancer Risk by Genome-Wide Variants

Peter T. Campbell, PhD [1,*] Yi Lin, MS,[2] Stephanie A. Bien, PhD,[2] Jane C. Figueiredo, PhD,[3,4] Tabitha A. Harrison, MPH, [2] Mark A. Guinter [1] Sonja I. Berndt, PharmD, PhD,[5] Hermann Brenner, MD,[6,7,8] Andrew T. Chan, MD,[9,10,11,12,13,14] Jenny Chang-Claude, PhD,[15,16] Steven J. Gallinger, MD,[17] Susan M. Gapstur, PhD,[1] Graham G. Giles, PhD, [18,19,20] Edward Giovannucci, MD, PhD,[13,21] Stephen B. Gruber, MD, PhD,[22] Marc Gunter, PhD,[23] Michael Hoffmeister, PhD, [6] Eric J. Jacobs, PhD, [1] Mark A. Jenkins, PhD, [19] Loic Le Marchand, MD, PhD,[24] Li Li, MD, PhD,[25] John R. McLaughlin, PhD,[26] Neil Murphy, PhD,[23] Roger L. Milne, PhD, [18,19,20] Polly A. Newcomb, PhD,[2] Christina Newton, MSPH, [1] Shuji Ogino, MD, PhD, [12,13,27,28] John D. Potter, MD, PhD, [2] Gad Rennert, MD, PhD, [29,30,31] Hedy S. Rennert, MPH,[29,30,31] Jennifer Robinson, PhD,[32] Lori C. Sakoda, PhD, [2,33] Martha L. Slattery, PhD,[34] Yiqing Song, PhD, [35] Emily White, PhD,[2,36] Michael O. Woods, PhD, [37] Graham Casey, PhD,[38] Li Hsu, PhD,[2] Ulrike Peters, PhD[2,36]

[1]Behavioral and Epidemiology Research Group, American Cancer Society, Atlanta, GA, USA; [2]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA; [3]Department of Medicine, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA; [4]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA; [5]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA; [6]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany; [7]Division of Preventive Oncology, German Cancer Research Center (DKFZ), and National Center for Tumor Diseases (NCT), Heidelberg, Germany; [8]German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany; [9]Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA; [10]Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; [11]Clinical and Translational Epidemiology Unit, Massachusetts General Hospital, and Harvard Medical School, Boston, MA, USA; [12]Broad Institute of Harvard and MIT, Cambridge, MA, USA; [13]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA, USA; [14]Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA, USA; [15]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany; [16]University Medical Centre Hamburg-Eppendorf, University Cancer Centre Hamburg (UCCH), Hamburg, Germany; [17]Lunenfeld Tanenbaum Research Institute, Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada; [18]Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, Victoria, Australia; [19]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia; [20]Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria, Australia; [21]Department of Nutrition, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA, USA; [22]Center for Precision Medicine and Department of Medical Oncology and Therapeutics Research, City of Hope National Medical Center, Duarte, CA, USA; [23]Section of Nutrition and Metabolism, International Agency for Research on Cancer, Lyon, France; [24]University of Hawaii Cancer Center, Honolulu, HI, USA; [25]Department of Family Medicine and Cancer Center, University of Virginia, Charlottesville, VA, USA; [26]Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada; [27]Program in MPE Molecular Pathological Epidemiology, Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; [28]Department of Pathology, Brigham & Women's Hospital, and Harvard Medical School, Boston, MA, USA; [29]Department of Community Medicine and Epidemiology, Lady Davis Carmel Medical Center, Haifa, Israel; [30]Ruth and Bruce Rappaport Faculty of Medicine, Technion-Israel Institute of Technology, Haifa, Israel; [31]Clalit National Cancer Control Center, Haifa, Israel; [32]Department of Epidemiology, College of Public Health, University of Iowa, Iowa City, IA, USA; [33]Division of Research, Kaiser Permanente Northern California, Oakland, CA, USA; [34]Department of Internal Medicine, University of Utah, Salt Lake City, UT, USA; [35]Department of Epidemiology, Richard M. Fairbanks School of Public Health, Indiana University, Indianapolis, IN, USA; [36]Department of Epidemiology, University of Washington School of Public Health, Seattle, WA, USA; [37]Discipline of Genetics, Memorial University of Newfoundland, St. John's, Newfoundland, Canada and [38]Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA

*Correspondence to: Peter Campbell, PhD, Behavioral and Epidemiology Research Group, American Cancer Society, 250 Williams Street, NW, Atlanta, GA 30303-1002, USA (e-mail: peter.campbell@cancer.org).

## Abstract

**Background:** Body mass index (BMI) is a complex phenotype that may interact with genetic variants to influence colorectal cancer risk. **Methods:** We tested multiplicative statistical interactions between BMI (per 5 kg/m$^2$) and approximately 2.7 million single nucleotide polymorphisms with colorectal cancer risk among 14 059 colorectal cancer case (53.2% women) and 14 416 control (53.8% women) participants. All analyses were stratified by sex *a priori*. Statistical methods included 2-step (ie, Cocktail method) and single-step (ie, case-control logistic regression and a joint 2-degree of freedom test) procedures. All statistical tests were two-sided. **Results:** Each 5 kg/m$^2$ increase in BMI was associated with higher risks of colorectal cancer, less

**Received:** December 13, 2019; **Revised:** February 27, 2020; **Accepted:** April 17, 2020

so for women (odds ratio [OR] = 1.14, 95% confidence intervals [CI] = 1.11 to 1.18; $P = 9.75 \times 10^{-17}$) than for men (OR = 1.26, 95% CI = 1.20 to 1.32; $P = 2.13 \times 10^{-24}$). The 2-step Cocktail method identified an interaction for women, but not men, between BMI and a *SMAD7* intronic variant at 18q21.1 (rs4939827; $P_{observed} = .0009$; $P_{threshold} = .005$). A joint 2-degree of freedom test was consistent with this finding for women (joint $P = 2.43 \times 10^{-10}$). Each 5 kg/m$^2$ increase in BMI was more strongly associated with colorectal cancer risk for women with the rs4939827-*CC* genotype (OR = 1.24, 95% CI = 1.16 to 1.32; $P = 2.60 \times 10^{-10}$) than for women with the *CT* (OR = 1.14, 95% CI = 1.09 to 1.19; $P = 1.04 \times 10^{-8}$) or *TT* (OR = 1.07, 95% CI = 1.01 to 1.14; P = .02) genotypes. **Conclusion:** These results provide novel insights on a potential mechanism through which a *SMAD7* variant, previously identified as a susceptibility locus for colorectal cancer, and BMI may influence colorectal cancer risk for women.

Colorectal cancer has a complex etiology involving inherited genetic variants, environmental and behavioral factors, and their interactions. Family studies estimate that inherited variability explains up to 35% of the population variation in colorectal cancer susceptibility (1,2). High-risk genetic syndromes and the common, low-risk variants identified by genome-wide association studies (GWAS) account for an estimated 3% and 12%, respectively, of the disease burden (3–5). Some of the missing heritability may be explained by gene-by-environment (GxE) interactions (6–10).

Body mass index (BMI), a general measure of body fatness, is an established risk factor for colorectal cancer and adenoma, although associations are often higher for men than women and may differ by location in the colorectum or by tumor molecular phenotype (11–16). The precise mechanisms that explain the BMI–colorectal cancer association are unknown; however, they likely involve multiple inflammatory, hormonal, metabolic, and immunologic networks that interact with the local tissue microenvironment. Given this potential for broad biologic interactions, it is important to consider BMI and colorectal cancer risk in the context of germline genetic variants. To date, no consistent GxE interactions have been identified from candidate gene (17–22) or GWA (23) studies of BMI and colorectal cancer risk. Lack of observed interactions may be due to insufficient statistical power in earlier studies.

In this study, we tested for multiplicative statistical interactions between approximately 2.7 million single nucleotide polymorphisms (SNPs) and BMI with risk of colorectal adenocarcinoma using 14 059 colorectal cancer/advanced adenoma case (53.2% women) and 14 416 control (53.8% women) participants.

## Methods

### Study Participants

The overall GWA study design has been described previously (22,24,25). In brief, this analysis is based on GWA studies from the multicentered Colon Cancer Family Registry, the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO), and the ColoRectal Transdisciplinary Study (Supplementary Methods, available online). Study-specific data for age and BMI are shown in Table 1.

All case participants with invasive colorectal adenocarcinoma were confirmed by abstraction of medical records, pathology reports, cancer registry linkage, or death certificates. Control participants were selected based on study-specific eligibility and matching criteria (eg, sex and age). Advanced colorectal adenoma cases were confirmed by review of medical records or pathology reports (women, n = 494; men, n = 304). Controls for adenoma cases had a negative colonoscopy or a negative sigmoidoscopy (for the latter, controls were matched only to cases who were diagnosed with distal adenoma). All studies were approved by their respective institutional review boards.

### Genotyping, Quality Assurance, Quality Control, and Imputation

Detailed information on genotyping, imputation, quality assurance, and quality control are presented elsewhere (25). In brief, genotyped SNPs were excluded based on call rate (<98%), lack of Hardy–Weinberg Equilibrium in controls ($P < 1 \times 10^{-4}$), and low minor allele count. Because our analysis is focused on common variants, we imputed the autosomal SNPs of all studies to the CEU population in HapMap II. SNPs were restricted to those with a per-study minor allele count greater than 5 and good imputation accuracy ($R^2 > 0.3$). After imputation and quality control analyses, a total of more than 2.7 million SNPs were used. All analyses were restricted to samples that clustered in principal component analysis with the Utah residents of northern and western European ancestry from the CEU population.

### Harmonization of Epidemiologic Data

Information on demographics and potential risk factors were collected by interviews and/or structured questionnaires. We carried out a multistep data-harmonization procedure at the GECCO coordinating center (Fred Hutchinson Cancer Research Center) as described previously (6–10).

The reference time for cohort studies was time of blood draw or buccal collection. The reference time for case-control studies was generally the period 1–2 years prior to diagnosis (cases) or enrollment (controls) to avoid bias from illness-associated weight loss. BMI was calculated from self-reports or direct measures of body weight (kg) divided by height (m$^2$). World Health Organization definitions for normal, overweight, and obese BMI were used for categorical analyses (26). Men and women with BMI less than 18.5 kg/m$^2$ (n = 247) were excluded from this analysis because of observed nonlinear associations at the lower end of the BMI continuum in these data and in other studies (27,28). Participants with missing BMI were excluded (n = 1626).

### Statistical Methods

All statistical analyses were conducted centrally at the GECCO coordinating center on individual-level data using the R programming language. Unless otherwise indicated, we adjusted for age at the reference time, study center, and the first 3 principal components from EIGENSTRAT. Each directly genotyped SNP was coded as 0, 1, or 2 copies of the variant allele. For imputed SNPs, we used the expected number of copies of the variant allele, which has been shown to give unbiased test statistics (29). Genotyped and imputed SNPs were treated as continuous

**Table 1.** Descriptive characteristics of study participants included in the interaction analysis of body mass index and genome-wide variants with colorectal cancer[a]

| | Women | | | | | | Men | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Case participants | | | Control participants | | | Case participants | | | Control participants | | |
| Study | No. | Mean BMI (SD), kg/m$^2$ | Mean age at diagnosis (SD), years | No. | Mean BMI (SD) | Age at enrollment (SD) | No. | Mean BMI (SD) | Mean age at diagnosis (SD) | No. | Mean BMI (SD) | Age at enrollment (SD) |
| CCFR Set 1 | 430 | 27.47 (6.73) | 52.97 (11.19) | 505 | 26.83 (6.89) | 59.6 (11.27) | 504 | 29.34 (7.72) | 54.97 (11.02) | 464 | 27.19 (4.94) | 60.56 (10.27) |
| OFCCR | 335 | 25.8 (4.44) | 61.14 (7.31) | 220 | 25.67 (4.74) | 62.19 (8.11) | 199 | 27.24 (3.57) | 60.73 (7.9) | 291 | 26.97 (4.07) | 63.07 (7.25) |
| DALS Set 1 | 297 | 27.33 (5.71) | 65.78 (10.01) | 303 | 26.16 (4.81) | 64.44 (9.84) | 397 | 28.2 (4.77) | 64.53 (9.41) | 399 | 26.78 (3.79) | 63.19 (9.92) |
| DALS Set 2 | 190 | 27.3 (5.74) | 64.65 (10.56) | 219 | 26.39 (5.39) | 64.34 (9.81) | 218 | 28.17 (4.58) | 65.25 (10.14) | 238 | 26.67 (3.72) | 64.35 (10.18) |
| PLCO Set 1 | 214 | 28.14 (5.97) | 68.94 (5.93) | 196 | 26.54 (4.99) | 64.21 (5.39) | 281 | 27.78 (4.14) | 69.3 (5.78) | 318 | 28 (4.42) | 64.92 (5.09) |
| PLCO Set 2 | 200 | 27.08 (4.69) | 70.83 (6.51) | 169 | 26.95 (4.74) | 63.74 (5.01) | 271 | 27.87 (4.16) | 70.45 (6.84) | 237 | 27.66 (3.83) | 63.57 (5.25) |
| WHI Set 1 | 451 | 28.03 (5.62) | 70.89 (7.08) | 504 | 26.7 (4.99) | 67.89 (6.73) | — | — | — | — | — | — |
| WHI Set 2 | 997 | 28.47 (5.6) | 71.96 (7.26) | 1000 | 28.22 (5.6) | 65.61 (6.24) | — | — | — | — | — | — |
| DACHS Set 1 | 674 | 26.57 (4.6) | 69.03 (10.82) | 670 | 25.71 (4.11) | 68.97 (9.95) | 993 | 27.42 (3.61) | 67.75 (9.71) | 1014 | 26.79 (3.25) | 68.61 (10.3) |
| DACHS Set 2 | 255 | 26.77 (4.33) | 69.55 (11.92) | 172 | 25.45 (4.37) | 70.76 (10.35) | 404 | 27.41 (3.64) | 68.13 (10.52) | 315 | 26.63 (3.28) | 68.96 (9.39) |
| HPFS Set 1 | — | — | — | — | — | — | 221 | 26.08 (3.1) | 70.4 (9.01) | 218 | 25.26 (3.29) | 65.93 (8.94) |
| HPFS Set 2 | — | — | — | — | — | — | 168 | 26.58 (3.24) | 66.26 (9.07) | 156 | 25.71 (3.08) | 64.04 (8.85) |
| HPFS adenoma | — | — | — | — | — | — | 304 | 25.99 (3.13) | 67.24 (8.64) | 336 | 25.68 (3.01) | 60.26 (8.21) |
| MEC | 148 | 26.7 (5.77) | 70.32 (8) | 160 | 25.83 (4.68) | 63.34 (7.62) | 176 | 27.11 (4.27) | 69.59 (8.6) | 182 | 26.35 (3.84) | 62.9 (8.32) |
| NHS Set 1 | 374 | 25.42 (4.27) | 65.22 (9.44) | 750 | 25.49 (4.26) | 59.87 (6.55) | — | — | — | — | — | — |
| NHS Set 2 | 146 | 25.96 (4.65) | 66.71 (8.53) | 171 | 26.11 (3.75) | 59.33 (6.4) | — | — | — | — | — | — |
| NHS adenoma | 494 | 25.88 (4.42) | 66.54 (7.2) | 556 | 25.1 (4.33) | 57 (6.77) | — | — | — | — | — | — |
| PHS Set 1 | — | — | — | — | — | — | 323 | 25.44 (2.8) | 70.58 (9.58) | 332 | 24.71 (2.65) | 59.17 (8.7) |
| VITAL | 122 | 28.08 (6.25) | 70.5 (6.47) | 123 | 26.79 (4.91) | 67.8 (5.99) | 146 | 28.44 (5.01) | 69.31 (6.77) | 147 | 27.17 (4.09) | 66.46 (6.43) |
| SCCFR | 276 | 27.89 (6.03) | 64.39 (6.93) | 115 | 25.94 (4.55) | 61.58 (7.46) | — | — | — | — | — | — |
| CCFR-CORECT | 403 | 27.13 (6.25) | 50.86 (11.34) | 370 | 26.51 (6.23) | 50.61 (11.75) | 437 | 27.86 (4.79) | 50.63 (10.19) | 315 | 27.49 (4.34) | 50.59 (10.85) |
| CPS-II | 268 | 26.75 (5.16) | 75.22 (5.85) | 250 | 25.67 (4.23) | 68.16 (5.61) | 271 | 27.05 (3.89) | 75.49 (5.26) | 271 | 26.32 (3.36) | 69.06 (5.35) |
| MCCS | 261 | 27.08 (4.52) | 70.2 (8.79) | 224 | 26.62 (4.34) | 59.96 (7.24) | 277 | 27.84 (3.39) | 69.69 (8.55) | 242 | 27.3 (3.44) | 60.23 (8.01) |
| MECC | 390 | 27.73 (5.02) | 70.55 (11.11) | 364 | 26.92 (4.65) | 71.92 (11.09) | 420 | 26.87 (4) | 71.66 (10.42) | 405 | 26.57 (3.44) | 74.71 (10.6) |
| Kentucky | 481 | 28.85 (6.71) | 62.57 (10.54) | 532 | 27.69 (5.97) | 66.75 (6.61) | 458 | 29.53 (5.45) | 62.86 (10.15) | 497 | 28.51 (5.07) | 60.51 (9) |
| NFCCR | 70 | 28.58 (5.41) | 60.16 (8.57) | 189 | 27.25 (5.06) | 58.17 (8.56) | 115 | 29.24 (4.62) | 62.2 (8.63) | 277 | 27.45 (3.91) | 60.05 (9.34) |
| Total | 7476 | — | — | 7762 | — | — | 6583 | — | — | 6654 | — | — |

[a]BMI = body mass index; CCFR = Colon Cancer Family Registry; OFCCR = Ontario Familial Colorectal Cancer Registry; DALS = Diet, Activity, and Lifestyle Study; PLCO = Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial; WHI = Women's Health Initiative; DACHS = Darmkrebs: Chancen der Verhütung durch Screening; HPFS = Health Professionals Follow-up Study; MEC = Multiethnic Cohort Study; NHS = Nurses' Health Study; PHS = Physicians' Health Study; VITAL = VITamins And Lifestyle; SCCFR = Seattle Colon Cancer Family Registry; CCFR-CORECT = Colon Cancer Family Registry participants from the ColoRectal Transdisciplinary Study (CORECT); CPS-II = American Cancer Society Cancer Prevention Study-II Nutrition cohort; MCCS = Melbourne Collaborative Cohort Study; MECC = Molecular Epidemiology of Colorectal Cancer study; Kentucky = Kentucky Case-Control study; NFCCR = Newfoundland Familial Colorectal Cancer Registry.

variables (ie, log-additive effects). Each study was analyzed separately, and study-specific results were combined using fixed-effects meta-analysis to obtain summary odds ratios (ORs) and 95% confidence intervals (CIs). We calculated the heterogeneity *P* values by Cochran Q statistics (30). Quantile–quantile plots assessed whether the distributions of the *P* values were consistent with the null distribution (except for the extreme tail).

To test for multiplicative statistical interactions between each SNP and BMI, we used conventional case-control logistic regression analysis and the Cocktail method (31) as primary analytic methods. We also used a 2-degree of freedom (*df*) joint test (32). We did not use any case-only statistical methods because BMI is a heritable trait (33,34) and G-E independence cannot be assumed.

For the logistic regression case-control analysis, we modeled the SNPxBMI interaction with the product of the SNP and BMI (per 5 $kg/m^2$), while including both the main effects variables for the SNP and BMI (and other covariates) in the same model. For these analyses, a 2-sided *P* less than $5 \times 10^{-8}$ was considered statistically significant.

The Cocktail method (31) consists of 2 steps. In the initial screening step, the 2.7 million SNPs are individually ranked according to their lowest *P* values from either a marginal association test of each SNP with colorectal cancer risk (35) or by a correlation between each SNP and BMI in cases and controls combined (36). Next, we used a weighted hypothesis-testing framework that ranks SNPs based on lowest *P* values from the screening step: SNPs with lower *P* values from the screening steps have less stringent alpha thresholds for the interaction test (eg, the top 5 SNPs with the lowest *P* values from the screening steps have an interaction alpha threshold of less than .005, whereas the next group of 10 SNPs has an alpha threshold of less than .00125, and so on). A marked advantage of a 2-step procedure that uses weighted hypothesis testing over a single-step procedure (eg, case-control logistic regression with Bonferroni correction) is that the former maintains an overall genome-wide error rate (consistent with the Bonferroni approach) while reallocating type 1 error to SNPs that are more likely to show a multiplicative interaction based on screening statistics. In contrast, a Bonferroni correction simply assumes all SNPs, regardless of any evidence for GxE from screening statistics, have an equal probability of GxE interaction (37). The last step of the Cocktail method is the testing step for statistical interaction, which, in this case, is a case-control logistic regression model.

For interactions highlighted here, we adjusted for additional covariates in the logistic regression case-control model (ie, smoking history, alcohol consumption, physical activity, and red meat consumption) because many of the environmental variables are correlated with BMI, and they are also associated with colorectal cancer risk. We also examined the main effects of BMI when stratified by the genotype of interest.

We also used the 2-*df* joint test (32), which simultaneously tests for a main effect of each SNP on colorectal cancer risk and a GxE interaction; this includes a 2-*df* $\chi^2$ test, which is the sum of the square of the z-statistic for the marginal association of each SNP with colorectal cancer risk and the square of the z-statistic from the case-control analysis of GxE interaction. For the 2-*df* joint test, a 2-sided *P* less than $5 \times 10^{-8}$ was considered statistically significant; manual review of each result below the alpha threshold is required to ensure that it is not simply the result of a low marginal association. All statistical tests were two-sided.

We performed bioinformatic follow-up for loci that were deemed statistically significant. Noncoding function was investigated using normal colorectal epigenomes (Roadmap, n = 3; International Human Epigenome Consortium [IHEC], n = 6), adenocarcinomas (IHEC, n = 6), colorectal cancer cell lines from ENCODE (n = 2), and regional annotations of enhancers gained or lost in tumor vs normal tissue (ie, variant enhancer loci, 3 normal crypts vs 10 colorectal cancer cell lines) (38). Variant effects on gene expression was investigated using normal colorectal expression data (GTEx transverse, n = 169, GTEx sigmoid, n = 124, and The Cancer Genome Atlas (TCGA) paired solid tissue normal, n = 51) and primary colon and rectal tumor samples from TCGA (n = 380). Annotation was performed for all variants tagged by a given locus ($r^2 \geq 0.5$ 1000 Genomes Project EUR) using Haploreg and the University of California at Santa Cruz (UCSC) genome browser. Lastly, to explore potential issues surrounding colorectal cancer somatic tumor heterogeneity, for loci that showed statistical evidence for GxBMI interaction, we investigated specific molecular phenotypes of colorectal cancer (eg, methylation markers, somatic mutations) in a subsample of case participants with available data.

## Results

Descriptive characteristics of case and control participants in this study are shown in Table 1. Each 5 $kg/m^2$ increase in BMI was associated with higher risks of colorectal cancer for women (OR = 1.14, 95% CI = 1.11 to 1.18; $P = 9.75 \times 10^{-17}$; Figure 1) and, more so, for men (OR = 1.26, 95% CI = 1.20 to 1.32; $P = 2.13 \times 10^{-24}$; Figure 2).

Statistical interaction results are summarized in Table 2 for 2 loci of interest. From traditional case-control logistic regression models with a Bonferroni correction for multiple testing, we did not identify any statistically significant interactions between BMI and any variant for women or men (data not shown). For women, the Cocktail method identified a statistically significant interaction between BMI and a *SMAD7* intronic variant at 18q21.1 (rs4939827: $P_{observed} = .0009$; $P_{threshold} = .005$) and a second suggestive finding, albeit above the alpha threshold, with a *PIK3CG* variant at chromosome 7 (rs849389: $P_{observed} = .016$; $P_{threshold} = .00125$). The former was 1 of 5 loci in the first Cocktail grouping ($P_{threshold} = .005$), whereas the latter was 1 of 10 loci in the second Cocktail grouping ($P_{threshold} = .00125$). The Cocktail method did not identify any statistically significant interactions for men. For women, using the 2-*df* joint test, we again identified a statistically significant interaction between BMI and rs4939827 (joint $P_{observed} = 2.43 \times 10^{-10}$): the low joint *P* value was the result of both a strong marginal association for rs4939827 and colorectal cancer risk ($P = 7.7 \times 10^{-9}$) and a low *P* value for the case-control interaction term (*P* = .0009). The 2-*df* joint test for men did not detect any statistically significant interactions.

Table 3 shows associations of BMI with colorectal cancer for women according to rs4939827 genotype. BMI per 5 $kg/m^2$ was more strongly associated with colorectal cancer risk for women with the *CC* genotype (OR = 1.24, 95% CI = 1.16 to 1.32; $P = 2.60 \times 10^{-10}$) than for women with the *CT* (OR = 1.14, 95% CI = 1.09 to 1.19; $P = 1.04 \times 10^{-8}$) or *TT* (OR = 1.07, 95% CI = 1.01 to 1.14; *P* = .02) genotypes. The BMIx*SMAD7* rs4939827 interaction result in men (Table 4) was not consistent with the multiplicative interaction observed in women; indeed, the suggested trend in men is for marginally weaker associations between BMI and colorectal cancer risk for the *CC* genotype (OR = 1.18, 95% CI = 1.07 to

| Study | N | Cases | Beta | OR | P–value |
|---|---|---|---|---|---|
| CCFR Set 1 | 935 | 430 | 0.09 | 1.09 | 0.09 |
| OFCCR | 555 | 335 | 0.03 | 1.03 | 0.75 |
| DALS Set 1 | 600 | 297 | 0.22 | 1.24 | 0.007 |
| DALS Set 2 | 409 | 190 | 0.15 | 1.16 | 0.11 |
| PLCO Set 1 | 410 | 214 | 0.25 | 1.29 | 0.009 |
| PLCO Set 2 | 369 | 200 | 0.04 | 1.04 | 0.74 |
| WHI Set 1 | 955 | 451 | 0.23 | 1.26 | $2.69 \times 10^{-4}$ |
| WHI Set 2 | 1997 | 997 | 0.04 | 1.04 | 0.29 |
| DACHS Set 1 | 1344 | 674 | 0.23 | 1.26 | $3.02 \times 10^{-4}$ |
| DACHS Set 2 | 427 | 255 | 0.42 | 1.52 | 0.001 |
| MEC | 308 | 148 | 0.15 | 1.16 | 0.18 |
| NHS adenoma | 1050 | 494 | 0.20 | 1.23 | 0.004 |
| NHS Set 1 | 1124 | 374 | −0.02 | 0.98 | 0.81 |
| NHS Set 2 | 317 | 146 | −0.05 | 0.95 | 0.72 |
| VITAL | 245 | 122 | 0.20 | 1.23 | 0.09 |
| SCCFR | 391 | 276 | 0.34 | 1.40 | 0.002 |
| CCFR–CORECT | 773 | 403 | 0.11 | 1.11 | 0.09 |
| CPS–II | 518 | 268 | 0.27 | 1.31 | 0.006 |
| MCCS | 485 | 261 | 0.12 | 1.13 | 0.25 |
| MECC | 754 | 390 | 0.16 | 1.18 | 0.03 |
| Kentucky | 1013 | 481 | 0.10 | 1.10 | 0.07 |
| NFCCR | 259 | 70 | 0.26 | 1.30 | 0.05 |
| **Meta–analysis** | **15238** | **7476** | **0.13** | **1.14** | **$9.75 \times 10^{-17}$** |
| **P–heterogeneity = 0.03** | | | | | |

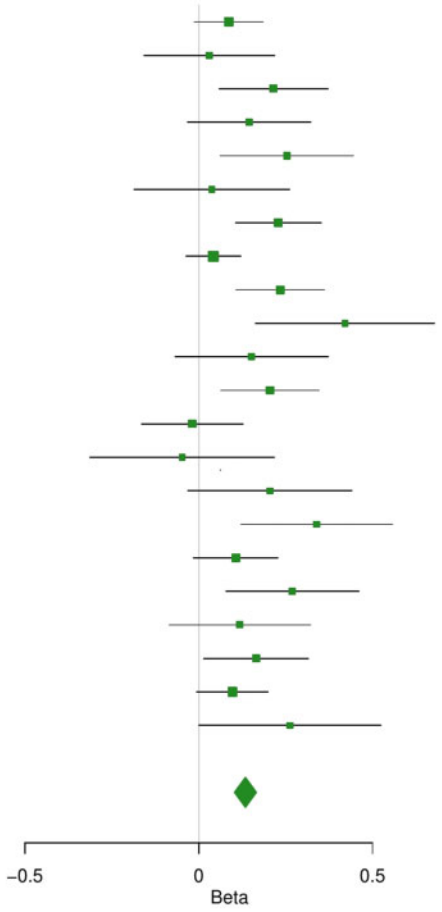


**Figure 1.** Forest plot for BMI (per 5 kg/m$^2$) and colorectal cancer risk in women. Study-specific betas, odds ratios (OR) and 95% confidence intervals (CI) were estimated with logistic regression case-control models. The summary OR was calculated using fixed-effects meta-analysis. CCFR = Colon Cancer Family Registry; CCFR-CORECT = Colon Cancer Family Registry participants from the ColoRectal Transdisciplinary Study (CORECT); CPS-II = American Cancer Society Cancer Prevention Study-II Nutrition cohort; DACHS = Darmkrebs: Chancen der Verhütung durch Screening; DALS = Diet, Activity, and Lifestyle Study; Kentucky = Kentucky Case-Control study; MEC = Multiethnic Cohort Study; MECC = Molecular Epidemiology of Colorectal Cancer study; MCCS = Melbourne Collaborative Cohort Study; NFCCR = Newfoundland Familial Colorectal Cancer Registry; NHS = Nurses' Health Study; OFCCR = Ontario Familial Colorectal Cancer Registry; PLCO = Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial; SCCFR = Seattle Colon Cancer Family Registry; VITAL = VITamins And Lifestyle; WHI = Women's Health Initiative.

ARTICLE

1.29) than for the *CT* (OR = 1.27, 95% CI = 1.20 to 1.35) or *TT* (OR = 1.32, 95% CI = 1.22 to 1.43) genotypes, although the *P* value for this interaction was above any threshold for statistical significance ($P = .10$).

We performed bioinformatic analysis of rs4939827. Annotation was performed for all variants tagged by rs4939827 ($r^2 \geq 0.5$ 1000 Genomes Project European) using Haploreg and the UCSC genome browser. Data from the TCGA Xena Browser (https://xenabrowser.net) showed reduced expression for SMAD7 in 380 primary tumor COAD-READ samples compared with 52 paired normal samples ($P = 6.8 \times 10^{-20}$; Supplementary Figure 1, available online). Supplementary Figure 2 (available online) shows UCSC genome browser results for the *SMAD7* gene: the rs4939827 locus (shown in first track) is in linkage disequilibrium (LD) with rs34007497, a locus associated with allele-specific expression in colon transverse tissue from GTEx. The additional 3 GTEx tracks below the UCSC gene annotation of *SMAD7* show that SMAD7 is expressed in normal colon tissues. Furthermore, the locus appears to overlap an enhancer that is more active in normal colon tissues than cancer cell lines (ie, variant enhancer locus). We did not observe any clear patterns of association according to strata of selected tumor molecular phenotypes to add insight on the observed interaction for rs4939827 and BMI in the 680 (or fewer) women case participants for whom these data were available (data not shown).

## Discussion

Consistent with overwhelming evidence from many studies (11,27,28), we found that higher BMI was associated with increased risk of colorectal cancer, more so for men than for women. We extend these established findings by reporting a novel GxE interaction between BMI and an intronic locus of *SMAD7* (rs4939827) for women. It is important to note that higher BMI was associated with risk of colorectal cancer for women in all 3 *SMAD7* genotype groups at this locus; that is, these results suggest that the magnitude of this association varies by genotype, not the direction.

The underlying pathophysiology of the BMI–colorectal cancer association is not fully understood, but it likely includes

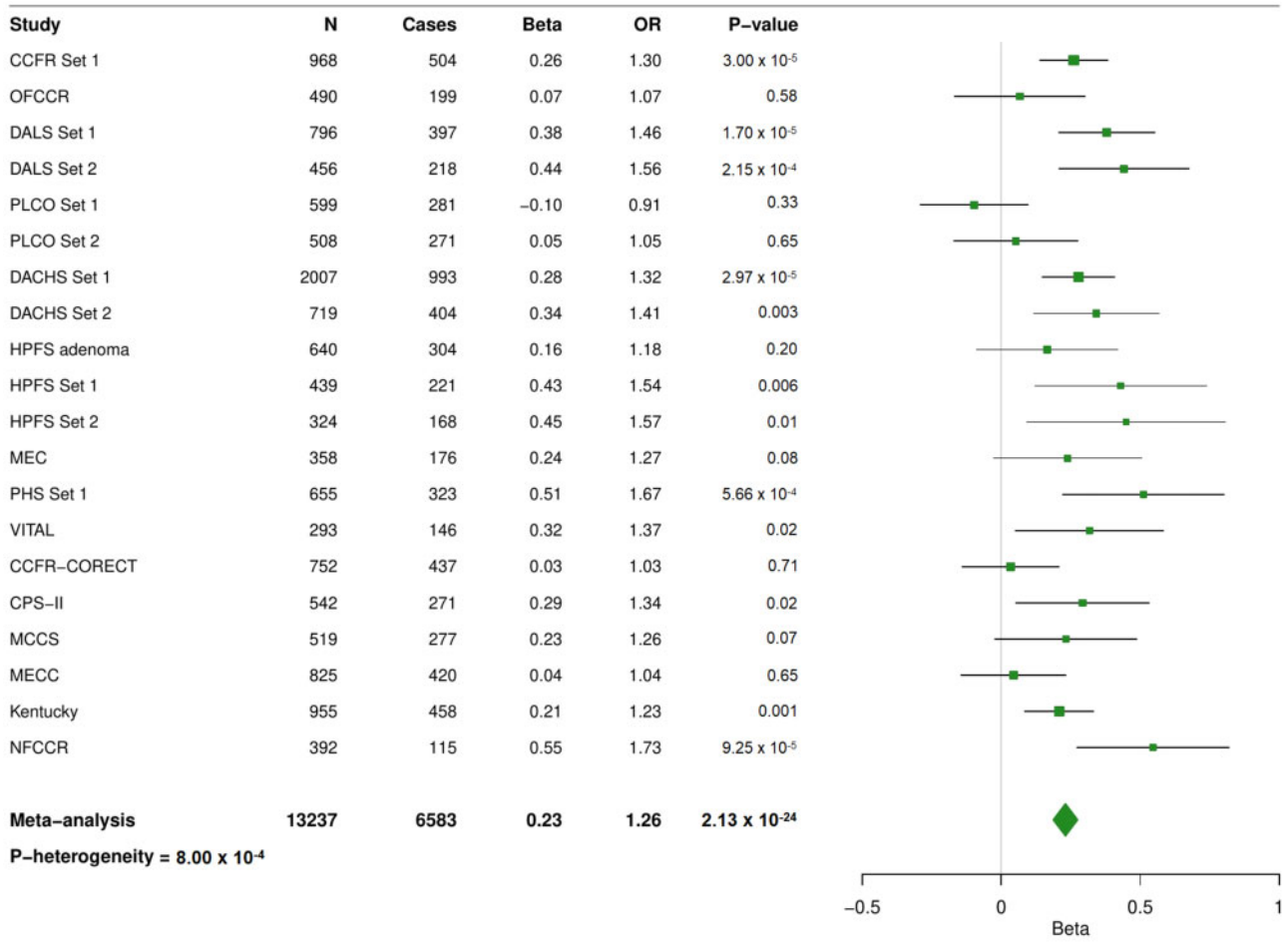


**Figure 2.** Forest plot for BMI (per 5 kg/m$^2$) and colorectal cancer risk in men. Study-specific betas, odds ratios (OR) and 95% confidence intervals (CI) were estimated with logistic regression case-control models. The summary OR was calculated using fixed-effects meta-analysis. CCFR = Colon Cancer Family Registry; CCFR-CORECT = Colon Cancer Family Registry participants from the ColoRectal Transdisciplinary Study (CORECT); CPS-II = American Cancer Society Cancer Prevention Study-II Nutrition cohort; DACHS = Darmkrebs: Chancen der Verhütung durch Screening; DALS = Diet, Activity, and Lifestyle Study; Kentucky = Kentucky Case-Control study; MEC = Multiethnic Cohort Study; MECC = Molecular Epidemiology of Colorectal Cancer study; MCCS = Melbourne Collaborative Cohort Study; NFCCR = Newfoundland Familial Colorectal Cancer Registry; NHS = Nurses' Health Study; OFCCR = Ontario Familial Colorectal Cancer Registry; PLCO = Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial; SCCFR = Seattle Colon Cancer Family Registry; VITAL = VITamins And Lifestyle; WHI = Women's Health Initiative.

**Table 2.** Main results for genome-wide interaction analyses and body mass index (BMI) with colorectal cancer risk among women[a]

| SNP | Chr | BP Position | Gene | Count allele | Count allele frequency | Statistical method used to detect the GxBMI interaction[b] | $P_{threshold}$ for GxBMI interaction | $P_{observed}$ for GxBMI interaction | *P* for heterogeneity | No. of studies included |
|---|---|---|---|---|---|---|---|---|---|---|
| Rs4939827 | 18 | 46453463 | *SMAD7* | C | 0.54 | Cocktail test | .005 | .0009 | .32 | 22 |
| Rs4939827 | 18 | 46453463 | *SMAD7* | C | 0.54 | 2 *df* Joint test | $5 \times 10^{-8}$ | $2.40 \times 10^{-10}$ | .32 | 22 |
| Rs849389 | 7 | 106508978 | *PIK3CG* | A | 0.97 | Cocktail test | .00125 | .016 | .90 | 22 |

[a]BMI was modeled per 5 kg/m$^2$, excluding values less than 18.5. Directly genotyped SNPs were coded as 0, 1, or 2 copies of the count allele. Imputed SNPs were coded as expected gene dosage. Multiplicative interaction terms were modeled as the product of BMI and each SNP of interest. BP position = base pair position based on NCBI Build37; Chr = chromosome; SNP = single nucleotide polymorphism; NCBI= National Center for Biotechnology Information.
[b]All statistical tests were 2-sided.

roles for dysfunctional white adipose tissue on creating a protumor microenvironment via increased levels of inflammatory cytokines (eg, IL-6, TNF-$\alpha$); unfavorable profiles of glucose homeostasis markers (eg, glucose, insulin, insulin-related growth factors); and hypoxemia-angiogenesis dysregulation (eg, TGF-$\beta$, HIF1-$\alpha$). Among these suggested mechanisms, SMAD7 is relevant in the context of a GxBMI interaction because it negatively regulates both TGF-$\beta$ transcription (39) and glucose/lipid metabolism involving the ASK1/TGF-$\beta$/p53 pathways (40). In turn, both the TGF-$\beta$ (41) and p53 (42) pathways are principally involved in colorectal carcinogenesis.

**Table 3.** Statistical interaction results for rs4939827 and body mass index (BMI) with colorectal cancer risk in women based on models with one common reference group and/or stratified by genotype

| | rs4939827 genotype | | | | | | | | | Per C allele within strata of BMI categories | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *TT* | | | *CT* | | | CC | | | | |
| BMI | No. Case/Control | OR (95% CI)[a] | $P^b$ | No. Case/Control | OR (95% CI)[a] | $P^b$ | No. Case/Control | OR (95% CI)[a] | $P^b$ | OR (95% CI)[a] | $P^b$ |
| BMI and genotype categories with a common reference group | | | | | | | | | | | |
| Normal | 926/931 | 1.00 (referent) | — | 1460/1762 | 0.81 (0.72 to 0.92) | $7.0 \times 10^{-4}$ | 572/826 | 0.68 (0.59 to 0.78) | $1.70 \times 10^{-7}$ | 0.82 (0.76 to 0.88) | $8.6 \times 10^{-8}$ |
| Overweight | 814/703 | 1.16 (1.01 to 1.33) | .04 | 1301/1352 | 0.96 (0.85 to 1.08) | .48 | 539/613 | 0.87 (0.75 to 1.01) | .07 | 0.86 (0.79 to 0.93) | $2.5 \times 10^{-4}$ |
| Obese | 556/439 | 1.24 (1.05 to 1.45) | .009 | 876/788 | 1.07 (0.93 to 1.23) | .35 | 432/348 | 1.21 (1.02 to 1.45) | .03 | 0.99 (0.89 to 1.09) | .77 |
| BMI categories within strata of genotype | | | | | | | | | | | |
| Normal | — | 1.00 (referent) | — | — | 1.00 (referent) | — | — | 1.00 (referent) | — | — | — |
| Overweight | — | 1.16 (1.01 to 1.33) | .04 | — | 1.18 (1.06 to 1.31) | .003 | — | 1.28 (1.09 to 1.51) | .003 | — | — |
| Obese | — | 1.24 (1.05 to 1.45) | .009 | — | 1.31 (1.16 to 1.48) | $1.5 \times 10^{-05}$ | — | 1.79 (1.49 to 2.15) | $5.0 \times 10^{-10}$ | — | — |
| BMI per 5 kg/m$^2$ within strata of genotype | | | | | | | | | | | |
| BMI continuous | — | 1.07 (1.01 to 1.14) | .02 | — | 1.14 (1.09 to 1.19) | $1.04 \times 10^{-8}$ | — | 1.24 (1.16 to 1.32) | $2.6 \times 10^{-10}$ | — | — |

[a]ORs and 95% CIs are adjusted for age, first 3 principal components of genetic ancestral markers, smoking history, alcohol consumption, physical activity, and red meat consumption. CI = confidence intervals; OR = odds ratio.
[b]All *P* values were derived from case-control logistic regression models and are 2-sided.

**Table 4.** Statistical interaction results for rs4939827 and body mass index (BMI) with colorectal cancer risk in men based on models with one common reference group and/or stratified by genotype

| | rs4939827 genotype | | | | | | | | | per C allele within strata of BMI categories | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *TT* | | | *CT* | | | CC | | | | |
| BMI | No. Case/Control | OR (95% CI)[a] | $P^b$ | No. Case/Control | OR (95% CI)[a] | $P^b$ | No. Case/Control | OR (95% CI)[a] | $P^b$ | OR (95% CI)[a] | $P^b$ |
| BMI and genotype categories with a common reference group | | | | | | | | | | | |
| Normal | 546/614 | 1.00 (referent) | — | 906/1115 | 0.92 (0.79 to 1.07) | .29 | 390/540 | 0.83 (0.69 to 0.99) | .04 | 0.91 (0.83 to 1.01) | .05 |
| Overweight | 1011/899 | 1.29 (1.11 to 1.50) | $7.9 \times 10^{-4}$ | 1602/1612 | 1.14 (0.99 to 1.31) | .06 | 661/728 | 1.03 (0.88 to 1.22) | .68 | 0.89 (0.83 to 0.96) | .002 |
| Obese | 438/312 | 1.65 (1.36 to 2.00) | $3.3 \times 10^{-7}$ | 739/580 | 1.49 (1.2 to 1.75) | $2.3 \times 10^{-6}$ | 290/254 | 1.27 (1.03 to 1.57) | .03 | 0.88 (0.78 to 0.99) | .03 |
| BMI categories within strata of genotype | | | | | | | | | | | |
| Normal | — | 1.00 (referent) | — | — | 1.00 (referent) | — | — | 1.00 (referent) | — | — | — |
| Overweight | — | 1.29 (1.11 to 1.50) | $7.9 \times 10^{-4}$ | — | 1.24 (1.10 to 1.39) | $2.9 \times 10^{-4}$ | — | 1.25 (1.05 to 1.48) | .01 | — | — |
| Obese | — | 1.65 (1.36 to 2.00) | $3.3 \times 10^{-7}$ | — | 1.61 (1.39 to 1.87) | $1.3 \times 10^{-10}$ | — | 1.53 (1.23 to 1.91) | $1.6 \times 10^{-4}$ | — | — |
| BMI per 5 kg/m$^2$ within strata of genotype | | | | | | | | | | | |
| BMI continuous | — | 1.32 (1.22 to 1.43) | $3.4 \times 10^{-11}$ | — | 1.27 (1.20 to 1.35) | $1.2 \times 10^{-14}$ | — | 1.18 (1.07 to 1.29) | $5.6 \times 10^{-4}$ | — | — |

[a]ORs and 95% CIs are adjusted for age, first 3 principal components of genetic ancestral markers, smoking history, alcohol consumption, physical activity, and red meat consumption. CI = confidence intervals; OR = odds ratio.
[b]All *P* values are derived from case-control logistic regression models and are 2-sided.

Although our interaction results for rs849389, in *PIK3CG*, were not statistically significant, future studies should examine these findings closely because of the role of PIK3CG on maintaining tissue homeostasis in the colonic epithelium and its inhibitory role on the PI3-kinase/Akt pathway (43), which, among other functions, is an essential downstream mediator of metabolic signaling from insulin, glucose, and related growth factors.

Rs4939827 was first identified as a risk locus for colorectal cancer in 2007 (44); this finding has been replicated by many subsequent studies that generally show an approximate 15% increased risk for each *T* allele (4,5). The complexity of the *SMAD7*–colorectal cancer relationship is exhibited by studies on survival outcomes after colorectal cancer diagnosis where the *C* allele, compared with the *T* allele, is associated with worse prognosis (45,46), in direct contrast to data from incidence studies where it is the *T* allele that is associated with higher risk. To explore this paradox, Garcia-Albeniz et al. (47) reported that part of the poorer prognosis associated with the *C* allele at this locus might be explained by a higher proportion of patients with later-stage tumors and more frequent methylation (ie, inactivation) at *RUNX3*. The genomic region wherein rs4939827 maps includes transcription factor binding sites for *RUNX3*, *SRY*, and *PAX4* (48), adding plausibility to this association. Further, RUNX3 was recently shown to act as a direct antioxidant barrier against TGF-$\beta$–induced genomic instability in the colon (49). TGF-$\beta$ super-family members, in turn, are increased with excess body fatness (50). Collectively, these studies provide support for an interaction mechanism whereby women with obese BMI and the *CC* genotype at rs4939827 may be at especially higher risk of colorectal cancer because of the combined effects of increased TGF-$\beta$ signaling and predilection toward methylation at *RUNX3* in colonic epithelial cells. Experimental studies are required to directly test this hypothesis.

The rs4939827 tagging SNP (tagSNP) is in LD with 4 reportedly functional SNPs (rs6507874, rs6507875, rs8085824, and rs58920878) that show allele-specific enhancer activity in colon cancer cell lines (51). Specifically, a haplotype containing the *C* allele had higher enhancer activity in 2 colorectal cancer cell lines, but this did not translate to higher SMAD7 expression levels (51). Our bioinformatics results showed that rs4939827 was also in LD with rs34007497, which has allele-specific enhancer activity in normal colon tissue. It is curious that this variant enhancer activity was more pronounced in normal colon tissues than in colon cancer cell lines, suggesting the effect may vary according to the colon microenvironment. It is important to note that it is especially difficult to interpret tumor expression data for SNPs on chromosomes, including chromosome 18, that are often lost through aneuploidy in cancer tissues.

Although this study included more than 14 000 case participants, relatively few participants also had tumor molecular phenotype data, and the study was underpowered to look at specific tumor phenotypic profiles according to sex-specific strata of rs4939827 and BMI—a limitation that future studies should address. Additionally, this study also relied largely on self-reported height and weight, which are prone to some misreporting, although the expected degree of underreporting for weight and overreporting for height (52–54) is unlikely to materially affect our results, particularly because that misreporting is unlikely to differ by genotype. This study was restricted, by design, to participants with northern and western European genetic ancestry; future studies should examine GxE interactions in other racial and ethnic groups, especially in populations that experience high rates of colorectal cancer. We chose *a priori* to not split our study sample into discovery and replication sets; future GWASxBMI studies will need to confirm or refute the interaction detected here.

We chose *a priori* to stratify all analyses by sex because of consistently observed differences in the magnitude of the BMI-colorectal cancer association for men compared with women, suggesting differing etiologies. The attenuation of this association for women has been attributed to an offsetting effect of adipose-derived estrogens (55,56); circulating estrogens in postmenopausal women, but not in men, are associated with lowered colorectal cancer risk (57,58). We are not aware of an obvious explanation for the sex-specific BMIx *SMAD7* interaction, but it seems plausible that steroid hormones may be involved. This hypothesis is further supported by findings that 17$\beta$-estradiol treatment has direct effects on TGF-$\beta$ signaling and SMAD7 protein expression in diabetic rat models (59).

This study has several strengths. First, with more than 28 000 case and control participants, this study was large enough to detect a statistically significant BMIx*SMAD7* interaction. The interaction locus detected at rs4939827 is reasonably well-characterized for its potential influence on colorectal carcinogenesis (39,48,51). The rs4939827 locus was directly genotyped in 15 of the 16 different GWAS platforms used in studies that included women in this analysis, and it was imputed with high accuracy in the one remaining study. Other strengths of this study include the selection of an environmental variable that is straightforward to harmonize and less prone to between-study heterogeneity.

In conclusion, we report a novel association whereby a common variant in *SMAD7* and BMI may jointly influence colorectal cancer risk for women. SMAD7 has a complex role in colorectal carcinogenesis with both tumor suppressive and oncogenic properties; thus, an interaction with BMI, an exposure that influences many of the same biologic pathways as SMAD7, seems plausible. This interaction may involve RUNX3 expression, the TGF-$\beta$ and p53 pathways, or other tumor-specific markers. From a public health perspective, these findings serve as an example on how a well-established risk factor for colorectal cancer may interact with genetic variants.

## Funding

For a complete list of funding, please see the Supplementary Materials (available online).

## Notes

**Role of the funder:** The funders had no role in the design of the study; the collection, analysis, and interpretation of the data; the writing of the manuscript; and the decision to submit the manuscript for publication.

**Disclosures:** All authors have no conflicts of interest to disclose.

**Acknowledgments:** Please see the Supplementary Materials (available online).

**Author contributions:** **YL**: Formal analysis; Methodology; Writing—review & editing. **SAB:** Formal analysis; Investigation; Methodology; Writing—original draft; Writing—review & editing. **JCF:** Investigation; Methodology; Project administration; Resources; Writing—review & editing. **TAH:** Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Supervision; Writing—review & editing. **MJG:** Formal analysis; Investigation; Methodology; Writing—

review & editing. **SIB:** Data curation; Funding acquisition; Writing—review & editing. **HB:** Data curation; Project administration; Supervision; Writing—review & editing. **ATC:** Data curation; Funding acquisition; Investigation; Resources; Writing—review & editing. **JC-C:** Data curation; Investigation; Project administration; Writing—review & editing. **SJG:** Data curation; Funding acquisition; Supervision; Writing—review & editing. **SMG:** Data curation; Funding acquisition; Supervision; Writing—review & editing. **GGG:** Data curation; Funding acquisition; Project administration; Writing—review & editing. **EG:** Conceptualization; Data curation; Funding acquisition; Project administration; Writing—review & editing. **SBG:** Data curation; Funding acquisition; Writing—review & editing. **MG:** Data curation; Funding acquisition; Project administration; Writing—review & editing. **MH:** Data curation; Funding acquisition; Project administration; Writing—review & editing. **EJJ:** Data curation; Methodology; Writing—review & editing. **MAJ:** Data curation; Funding acquisition; Project administration; Writing—review & editing. **LLM:** Data curation; Funding acquisition; Writing—review & editing. **LL:** Data curation; Funding acquisition; Writing—review & editing. **JRM:** Data curation; Funding acquisition; Project administration; Writing—review & editing. **NM:** Data curation; Investigation; Writing—review & editing. **RLM:** Data curation; Project administration; Writing—review & editing. **PAN:** Data curation; Funding acquisition; Project administration; Writing—review & editing. **CN:** Data curation; Formal analysis; Methodology; Writing—review & editing. **SO:** Conceptualization; Data curation; Funding acquisition; Supervision; Writing—review & editing; **JDP:** Conceptualization; Data curation; Funding acquisition; Project administration; Resources; Writing—review & editing; **GR:** Data curation; Funding acquisition; Project administration; Writing—review & editing. **HSR:** Data curation; Funding acquisition; Writing—review & editing. **JR:** Data curation; Funding acquisition; Writing—review & editing. **LCS:** Data curation; Funding acquisition; Project administration; Writing—review & editing. **MLS:** Data curation; Funding acquisition; Writing—review & editing. **YS:** Data curation; Project administration; Writing—review & editing. **EW:** Data curation; Funding acquisition; **MOW:** Data curation; Funding acquisition; Writing—review & editing. **GC:** Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Writing—original draft; Writing—review & editing. **LH:** Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Writing—review & editing. **UP:** Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Writing—original draft; Writing—review & editing.

## References


1. Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med*. 2000;343(2):78–85.
2. Czene K, Lichtenstein P, Hemminki K. Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database. *Int J Cancer*. 2002;99(2):260–266.
3. Foulkes WD. Inherited susceptibility to common cancers. *N Engl J Med*. 2008; 359(20):2143–2153.
4. Schmit SL, Edlund CK, Schumacher FR, et al. Novel common genetic susceptibility loci for colorectal cancer. *J Natl Cancer Inst*. 2019;111(2):146–157.
5. Huyghe JR, Bien SA, Harrison TA, et al. Discovery of common and rare genetic risk variants for colorectal cancer. *Nat Genet*. 2019;51(1):76–87.
6. Figueiredo JC, Hsu L, Hutter CM, et al. Genome-wide diet-gene interaction analyses for risk of colorectal cancer. *PLoS Genet*. 2014;10(4):e1004228.
7. Nan H, Hutter CM, Lin Y, et al. Association of aspirin and NSAID use with risk of colorectal cancer according to genetic variants. *JAMA*. 2015;313(11): 1133–1142.
8. Lemire M, Qu C, Loo LW, et al. A genome-wide association study for colorectal cancer identifies a risk locus in 14q23.1. *Hum Genet*. 2015;134(11-12): 1249–1262.
9. Garcia-Albeniz X, Rudolph A, Hutter C, et al. CYP24A1 variant modifies the association between use of oestrogen plus progestogen therapy and colorectal cancer risk. *Br J Cancer*. 2016;114(2):221–229.
10. Gong J, Hutter CM, Newcomb PA, et al. Genome-wide interaction analyses between genetic variants and alcohol consumption and smoking for risk of colorectal cancer. *PLoS Genet*. 2016;12(10):e1006296.
11. Lauby-Secretan B, Scoccianti C, Loomis D, et al. Body fatness and cancer—viewpoint of the IARC Working Group. *N Engl J Med*. 2016;375(8):794–798.
12. Ning Y, Wang L, Giovannucci EL. A quantitative analysis of body mass index and colorectal cancer: findings from 56 observational studies. *Obes Rev*. 2010; 11(1):19–30.
13. Campbell PT, Jacobs ET, Ulrich CM, et al. Case-control study of overweight, obesity, and colorectal cancer risk, overall and by tumor microsatellite instability status. *J Natl Cancer Inst*. 2010;102(6):391–400.
14. Kuchiba A, Morikawa T, Yamauchi M, et al. Body mass index and risk of colorectal cancer according to fatty acid synthase expression in the nurses' health study. *J Natl Cancer Inst*. 2012;104(5):415–420.
15. Hughes LA, Williamson EJ, van Engeland M, et al. Body size and risk for colorectal cancers showing BRAF mutations or microsatellite instability: a pooled analysis. *Int J Epidemiol*. 2012;41(4):1060–1072.
16. Carr PR, Alwers E, Bienert S, et al. Lifestyle factors and risk of sporadic colorectal cancer by microsatellite instability status: a systematic review and meta-analyses. *Ann Oncol*. 2018;29(4):825–834.
17. Slattery ML, Murtaugh MA, Sweeney C, et al. PPARgamma, energy balance, and associations with colon and rectal cancer. *Nutr Cancer*. 2005;51(2): 155–161.
18. Slattery ML, Murtaugh M, Caan B, et al. Energy balance, insulin-related genes and risk of colon and rectal cancer. *Int J Cancer*. 2005;115(1):148–154.
19. Gunter MJ, Hayes RB, Chatterjee N, et al. Insulin resistance-related genes and advanced left-sided colorectal adenoma. *Cancer Epidemiol Biomarkers Prev*. 2007;16(4):703–708.
20. Liu L, Zhong R, Wei S, et al. Interactions between genetic variants in the adiponectin, adiponectin receptor 1 and environmental factors on the risk of colorectal cancer. *PLoS One*. 2011;6(11):e27301.
21. Campbell PT, Curtin K, Ulrich C, et al. Mismatch repair polymorphisms and risk of colon cancer, tumor microsatellite instability, and interactions with lifestyle factors. *Gut*. 2009;58(5):661–667.
22. Hutter CM, Chang-Claude J, Slattery ML, et al. Characterization of gene-environment interactions for colorectal cancer susceptibility loci. *Cancer Res*. 2012;72(8):2036–2044.
23. Figueiredo JC, Lewinger JP, Song C, et al. Genotype-environment interactions in microsatellite stable/microsatellite instability-low colorectal cancer: results from a genome-wide association study. *Cancer Epidemiol Biomarkers Prev*. 2011;20(5):758–766.
24. Peters U, Hutter CM, Hsu L, et al. Meta-analysis of new genome-wide association studies of colorectal cancer risk. *Hum Genet*. 2012;131(2):217–234.
25. Peters U, Jiao S, Schumacher FR, et al. Identification of genetic susceptibility loci for colorectal tumors in a genome-wide meta-analysis. *Gastroenterology*. 2013;144(4):799–807.e24.
26. World Health Organization. Report of a WHO consultation on obesity. *Obesity: Preventing and Managing the Global Epidemic*. Geneva: WHO; 1998.
27. Zheng W, McLerran DF, Rolland B, et al. Association between body-mass index and risk of death in more than 1 million Asians. *N Engl J Med*. 2011;364(8): 719–729.
28. Di Angelantonio E, Bhupathiraju SN, Wormser D, et al. Body-mass index and all-cause mortality: individual-participant-data meta-analysis of 239 prospective studies in four continents. *Lancet*. 2016;388(10046):776–786.
29. Jiao S, Hsu L, Hutter CM, et al. The use of imputed values in the meta-analysis of genome-wide association studies. *Genet Epidemiol*. 2011;35(7):597–605.
30. Cochran WG. The combination of estimates from different experiments. *Biometrics*. 1954;10(1):101–129.
31. Hsu L, Jiao S, Dai JY, Hutter C, Peters U, Kooperberg C. Powerful cocktail methods for detecting genome-wide gene-environment interaction. *Genet Epidemiol*. 2012;36(3):183–194.
32. Dai JY, Logsdon BA, Huang Y, et al. Simultaneously testing for marginal genetic association and gene-environment interaction. *Am J Epidemiol*. 2012; 176(2):164–173.
33. Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015;518(7538):197–206.
34. Justice AE, Winkler TW, Feitosa MF, et al. Genome-wide meta-analysis of 241,258 adults accounting for smoking behaviour identifies novel loci for obesity traits. *Nat Commun*. 2017;8(1):14977.
35. Kooperberg C, Leblanc M. Increasing the power of identifying gene x gene interactions in genome-wide association studies. *Genet Epidemiol*. 2008;32(3): 255–263.
36. Murcray CE, Lewinger JP, Gauderman WJ. Gene-environment interaction in genome-wide association studies. *Am J Epidemiol*. 2008;169(2):219–226.

37. Dai JY, Hsu L, Kooperberg C. Two-stage testing for genome-wide gene-environment interactions. *Handbook of Statistical Methods for Case-Control Studies*. London: Chapman and Hall; 2018:459–474.
38. Akhtar-Zaidi B, Cowper-Sal-Lari R, Corradin O, et al. Epigenomic enhancer profiling defines a signature of colon cancer. *Science*. 2012;336(6082):736–739.
39. Troncone E, Monteleone G. Smad7 and colorectal carcinogenesis: a double-edged sword. *Cancers (Basel)*. 2019;11(5):612.
40. Seong HA, Manoharan R, Ha H. SMAD proteins differentially regulate obesity-induced glucose and lipid abnormalities and inflammation via class-specific control of AMPK-related kinase MPK38/MELK activity. *Cell Death Dis*. 2018;9(5):471.
41. Jung B, Staudacher JJ, Beauchamp D. Transforming growth factor beta superfamily signaling in development of colorectal cancer. *Gastroenterology*. 2017; 152(1):36–52.
42. Iacopetta B. TP53 mutation in colorectal cancer. *Hum Mutat*. 2003;21(3): 271–276.
43. Semba S, Itoh N, Ito M, et al. Down-regulation of PIK3CG, a catalytic subunit of phosphatidylinositol 3-OH kinase, by CpG hypermethylation in human colorectal carcinoma. *Clin Cancer Res*. 2002;8(12):3824–3831.
44. Broderick P, Carvajal-Carmona L, Pittman AM, et al. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. *Nat Genet*. 2007;39(11):1315–1317.
45. Phipps AI, Newcomb PA, Garcia-Albeniz X, et al. Association between colorectal cancer susceptibility loci and survival time after diagnosis with colorectal cancer. *Gastroenterology*. 2012;143(1):51–54 e4.
46. Passarelli MN, Coghill AE, Hutter CM, et al. Common colorectal cancer risk variants in SMAD7 are associated with survival among prediagnostic nonsteroidal anti-inflammatory drug users: a population-based study of postmenopausal women. *Genes Chromosom Cancer*. 2011;50(11): 875–886.
47. Garcia-Albeniz X, Nan H, Valeri L, et al. Phenotypic and tumor molecular characterization of colorectal cancer in relation to a susceptibility SMAD7 variant associated with survival. *Carcinogenesis*. 2013;34(2):292–298.
48. Pittman AM, Naranjo S, Webb E, et al. The colorectal cancer risk at 18q21 is caused by a novel variant altering SMAD7 expression. *Genome Res*. 2009;19(6): 987–993.
49. Krishnan V, Chong YL, Tan TZ, et al. TGFbeta promotes genomic instability after loss of RUNX3. *Cancer Res*. 2018;78(1):88–102.
50. Lee MJ. Transforming growth factor beta superfamily regulation of adipose tissue biology in obesity. *Biochim Biophys Acta Mol Basis Dis*. 2018;1864(4): 1160–1171.
51. Fortini BK, Tring S, Plummer SJ, et al. Multiple functional risk variants in a SMAD7 enhancer implicate a colorectal cancer risk haplotype. *PLoS One*. 2014; 9(11):e111914.
52. Shields M, Gorber SC, Tremblay MS. Effects of measurement on obesity and morbidity. *Health Rep*. 2008;19(2):77–84.
53. McAdams MA, Van Dam RM, Hu FB. Comparison of self-reported and measured BMI as correlates of disease markers in US adults. *Obesity (Silver Spring)*. 2007;15(1):188–196.
54. Spencer EA, Appleby PN, Davey GK, et al. Validity of self-reported height and weight in 4808 EPIC-Oxford participants. *Public Health Nutr*. 2002;5(4):561–565.
55. Campbell PT, Cotterchio M, Dicks E, et al. Excess body weight and colorectal cancer risk in Canada: Associations in subgroups of clinically defined familial risk of cancer. *Cancer Epidemiol Biomarkers Prev*. 2007;16(9):1735–1744.
56. Simpson ER, Bulun SE, Nichols JE, et al. Estrogen biosynthesis in adipose tissue: regulation by paracrine and autocrine mechanisms. *J Endocrinol*. 1996; 150 (Suppl):S51–S57.
57. Murphy N, Strickler HD, Stanczyk FZ, et al. A prospective evaluation of endogenous sex hormone levels and colorectal cancer risk in postmenopausal women. *J Natl Cancer Inst*. 2015;107(10):djv210.
58. Lin JH, Zhang SM, Rexrode KM, et al. Association between sex hormones and colorectal cancer risk in men and women. *Clin Gastroenterol Hepatol*. 2013; 11(4):419–424 e1.
59. Dixon A, Maric C. 17beta-estradiol attenuates diabetic kidney disease by regulating extracellular matrix and transforming growth factor-beta protein expression and signaling. *Am J Physiol Renal Physiol*. 2007;293(5):F1678–90.