Published in final edited form as:
Cancer Epidemiol Biomarkers Prev. 2014 September ; 23(9): 1824–1833. doi: 10.1158/1055-9965.EPI-14-0062.

Gene-environment interaction involving recently identified colorectal cancer susceptibility loci

Elizabeth D. Kantor[1,2,3], **Carolyn M. Hutter**[4], **Jessica Minnier**[2,5], **Sonja I. Berndt**[6], **Hermann Brenner**[7,8], **Bette J. Caan**[9], **Peter T. Campbell**[10], **Christopher S. Carlson**[2,3], **Graham Casey**[11], **Andrew T. Chan**[12,13], **Jenny Chang-Claude**[14], **Stephen J. Chanock**[6], **Michelle Cotterchio**[15], **Mengmeng Du**[2,3,13], **David Duggan**[16], **Charles S. Fuchs**[13,17], **Edward L. Giovannucci**[1,13,18], **Jian Gong**[2], **Tabitha A. Harrison**[2], **Richard B. Hayes**[19], **Brian E. Henderson**[20], **Michael Hoffmeister**[7], **John L. Hopper**[21], **Mark A. Jenkins**[21], **Shuo Jiao**[2], **Laurence N. Kolonel**[22], **Loic Le Marchand**[22], **Mathieu Lemire**[23], **Jing Ma**[13], **Polly A. Newcomb**[2,3], **Heather M. Ochs-Balcom**[24], **Bethann M. Pflugeisen**[2], **John D. Potter**[2,3,25], **Anja Rudolph**[26], **Robert E. Schoen**[27], **Daniela Seminara**[4], **Martha L. Slattery**[28], **Deanna L. Stelling**[2], **Fridtjof Thomas**[29], **Mark Thornquist**[2], **Cornelia M. Ulrich**[2,3,30], **Greg S. Warnick**[2], **Brent W. Zanke**[31], **Ulrike Peters**[2,3], **Li Hsu**[2,32], and **Emily White**[2,3]

[1]Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA [2]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA [3]Department of Epidemiology, University of Washington School of Public Health, Seattle, WA, USA [4]Division of Cancer Control and Population Sciences, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA [5]Department of Public Health & Preventive Medicine, Oregon Health & Science University, Portland, OR, USA [6]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA [7]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany [8]German Cancer Consortium (DKTK), Heidelberg, Germany [9]Division of Research, Kaiser Permanente Medical Care Program, Oakland, CA, USA [10]Epidemiology Research Program, American Cancer Society, Atlanta, GA, USA [11]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA [12]Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA [13]Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA [14]Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany [15]Prevention and Cancer Control, Cancer Care Ontario, Toronto, ON, Canada [16]Translational Genomics Research Institute, Phoenix, AZ, USA [17]Department of Medical Oncology, Dana Farber Cancer Institute, Boston, MA, USA [18]Department of Nutrition, Harvard School of Public Health, Boston, MA, USA [19]Division of Epidemiology, Department of Population Health, New York University School of Medicine, New York, NY, USA [20]Keck School of Medicine, University of Southern California, Los Angeles, CA, USA [21]Melbourne School of Population Health, The University of Melbourne, VIC, Australia [22]Epidemiology Program,

Corresponding Author: Elizabeth D. Kantor, Harvard School of Public Health, Department of Epidemiology, 677 Huntington Avenue, Boston, MA, 02115, Phone: 617.432.1194, Fax: 617.566.7805, ekantor@hsph.harvard.edu.

Conflict of Interest: Study authors have no conflict of interest to declare.

University of Hawaii Cancer Center, Honolulu, HI, USA [23]Ontario Institute for Cancer Research, Toronto, ON, Canada [24]Department of Social and Preventive Medicine, University at Buffalo, Buffalo, NY, USA [25]Centre for Public Health Research, Massey University, Wellington, New Zealand [26]Division of Cancer Epidemiology, Unit of Genetic Epidemiology, German Cancer Research Center, Heidelberg, Germany [27]Department of Medicine and Epidemiology, University of Pittsburgh Medical Center, Pittsburgh, PA, USA [28]Department of Internal Medicine, University of Utah Health Sciences Center, Salt Lake City, UT, USA [29]Department of Preventive Medicine, University of Tennessee Health Science Center, University of Tennessee, Memphis, TN, USA [30]Division of Preventive Oncology, National Center for Tumor Diseases and German Cancer Research Center, Heidelberg, Germany [31]Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada [32]Department of Biostatistics, University of Washington School of Public Health, Seattle, WA, USA

Abstract

BACKGROUND—Genome-wide association studies have identified several single nucleotide polymorphisms (SNPs) that are associated with risk of colorectal cancer (CRC). Prior research has evaluated the presence of gene-environment interaction involving the first 10 identified susceptibility loci, but little work has been conducted on interaction involving SNPs at recently identified susceptibility loci, including: rs10911251, rs6691170, rs6687758, rs11903757, rs10936599, rs647161, rs1321311, rs719725, rs1665650, rs3824999, rs7136702, rs11169552, rs59336, rs3217810, rs4925386, and rs2423279.

METHODS—Data on 9160 cases and 9280 controls from the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO) and Colon Cancer Family Registry (CCFR) were used to evaluate the presence of interaction involving the above-listed SNPs and sex, body mass index (BMI), alcohol consumption, smoking, aspirin use, post-menopausal hormone (PMH) use, as well as intake of dietary calcium, dietary fiber, dietary folate, red meat, processed meat, fruit, and vegetables. Interaction was evaluated using a fixed-effects meta-analysis of an efficient Empirical Bayes estimator, and permutation was used to account for multiple comparisons.

RESULTS—None of the permutation-adjusted p-values reached statistical significance.

CONCLUSIONS—The associations between recently identified genetic susceptibility loci and CRC are not strongly modified by sex, BMI, alcohol, smoking, aspirin, PMH use, and various dietary factors.

IMPACT—Results suggest no evidence of strong gene-environment interactions involving the recently identified 16 susceptibility loci for CRC taken one at a time.

Keywords

Colorectal Cancer; Gene-Environment Interaction; Polymorphism; Single Nucleotide; Genetic Predisposition to Disease; Diet

INTRODUCTION

Colorectal cancer (CRC) is the third most common cancer among men and women in the United States [1]. To date, genome-wide association studies (GWAS) have identified a number of single nucleotide polymorphisms (SNPs) that are associated with risk of this cancer [2–14]. There is much interest in identifying whether demographic and lifestyle factors modify the association between genetic variants and CRC, as finding evidence of gene-environment (GxE) interaction may help guide future prevention strategies. Furthermore, understanding GxE interaction may shed light on the mechanisms by which genetic polymorphisms affect risk of CRC, as well as the underlying biology of this disease. The SNPs identified to be associated with CRC thus far only account for a small fraction of the estimated heritability of CRC [15,16], and it has been suggested that one factor contributing to this 'missing heritability' is gene-environment (GxE) interaction [17,18].

We previously reported on gene-environment interaction for the first 10 identified susceptibility loci [19]. Since the time of that publication, 16 additional SNPs have been associated with CRC, including: rs10911251 (1q25.3), rs6691170 (1q41), rs6687758 (1q41), rs11903757 (2q32.3), rs10936599 (3q26.2), rs647161 (5q31.1), rs1321311 (6p21), rs719725 (9p24), rs1665650 (10q26.12), rs3824999 (11q13.4), rs7136702 (12q13.13), rs11169552 (12q13.13), rs59336 (12q24.21), rs3217810 (12p13.32), rs4925386 (20q13.33), rs2423279 (20p12.3) [3,4,7,8,10,14]. Few studies have evaluated the presence of interaction involving these recently identified susceptibility loci [8,20–24]. Although it has been suggested that sex may interact with rs4925386 [22], no interaction has been observed between sex and rs719725 [8, 21,24], rs6691170 [22], rs10936599 [22], or rs11169552 [22]. Of the newly identified susceptibility loci, only rs719725 [8, 21,23] and SNPs highly correlated with rs719725 [20] have been evaluated for interaction with environmental factors such as body mass index (BMI), alcohol consumption, smoking, medication use, and diet. No statistically significant GxE interactions were observed in these studies; however statistical power to detect interaction may have been limited due to insufficient sample sizes. We have therefore evaluated whether environmental risk factors for CRC modify the associations between these genetic polymorphisms and CRC risk using data on 9160 cases and 9280 controls in the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO) and the Colon Cancer Family Registry (CCFR). The following environmental and demographic factors were included in our study: sex, BMI, alcohol use, smoking, aspirin use, post-menopausal hormone (PMH) use, dietary intake of calcium, fiber, folate, red meat, processed meat, fruit, and vegetables. These 'environmental factors' have been loosely defined so as to include lifestyle factors and personal characteristics associated with CRC risk [25–35].

MATERIALS AND METHODS

Study participants

Study participants were drawn from either case-control studies (Ontario Familial Colorectal Cancer Registry [OFCCR], Darmkrebs: Chancen der Verhuetung durch Screening [DACHS], Diet, Activity and Lifestyle Survey [DALS], Colon Cancer Family Registry [CCFR], Colorectal Cancer Studies 2&3 [Colo2&3], and the Postmenopausal Hormone study within the Colon Cancer Family Registry [PMH-CCFR]) or from case-control studies

nested within prospective cohorts: Health Professionals Follow-up Study [HPFS], Nurses' Health Study [NHS], Physicians' Health Study [PHS], Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial [PLCO], Women's Health Initiative [WHI], Multiethnic Cohort Study [MEC], and the VITamins And Lifestyle [VITAL] study. More detailed information on these studies can be found in Table 1 and in the Supplemental Methods. All participants gave informed consent and studies were approved by their respective Institutional Review Boards.

Outcome

Colorectal cancer (CRC) cases included in this study were defined as invasive colorectal adenocarcinoma (ICD codes 153–154). Cases were confirmed by medical record, pathology report, or death certificate. Controls in these case-control studies and nested case-control studies were selected based on study-specific eligibility and matching criteria, as detailed in the Supplemental Methods.

Genotype Data

Gene-environmental interaction was evaluated for 16 SNPs located at recently identified CRC susceptibility loci, including: rs10911251 (1q25.3), rs6691170 (1q41), rs6687758 (1q41), rs11903757 (2q32.3), rs10936599 (3q26.2), rs647161 (5q31.1), rs1321311 (6p21), rs719725 (9p24), rs1665650 (10q26.12), rs3824999 (11q13.4), rs7136702 (12q13.13), rs11169552 (12q13.13), rs59336 (12q24.21), rs3217810 (12p13.32), rs4925386 (20q13.33), rs2423279 (20p12.3) [3,4,7,8,10,14].

DNA for genotyping was largely obtained from blood samples, though DNA was also obtained from buccal swabs for VITAL participants and for a subset of participants from DACHS, MEC, and PLCO. Genotyping was conducted on several different platforms and several of the studies were genotyped in sets, Therefore, in describing the genotyping platform and in presenting data on genotyping quality in Supplemental Table 1, results are presented by study set. However, we have presented results in tables and figures by overall study population. The Illumina HumanHap BeadChip Array System was used to genotype SNPs for the following studies: Colo2&3, DACHS1, DALS2, MEC, PLCO2, PMH-CCFR, VITAL, WHI2 (300k); DALS1, WHI1 (550k); WHI1 (550kduo); DALS1, WHI1 (610k); DACHS2, HPFS1, HPFS2, NHS1, NHS2, PHS1+2 (730k), as described previously [9]; OFCCR samples were genotyped using Affymetrix platforms [14]. All genotyping underwent quality-control checks, including concordance checks for blinded and unblinded duplicates, as well as examination of sample call rates, SNP call rates, and, in controls, Hardy-Weinberg Equilibrium (HWE). Samples with gender discrepancies were excluded, as were persons who reported a racial/ethnic group other than "white;" European ancestry was confirmed in GWAS samples using principal components analysis.

As not all of the SNPs of interest were genotyped on each platform, we imputed SNPs to the CEPH collection (CEU) population in HapMap II. Imputation was used only if a minor allele frequency (MAF) of >1% could be assumed and satisfactory overall imputation accuracy (R^2 >0.3) was achieved. Imputation quality was high for all SNPs of interest (average R^2>0.85), except rs3217810 (average R^2=0.49) and rs11903757 (average R^2=0.69)

(Table 2). For each SNP included in our analyses, the number of studies in which that SNP was imputed or genotyped is provided in Table 2. All SNPs are presented in terms of number of risk alleles, with 0 corresponding to no risk alleles, and 2 corresponding to 2 risk alleles. Directly genotyped SNPs are coded as 0, 1, or 2 risk alleles, and imputed SNPs are instead coded in terms of the expected number of risk alleles ("dosage" between 0 and 2) [36]. The risk allele designation for each SNP was determined by the discovery studies, as presented in Table 2. The SNP details by study, including the risk allele frequency (RAF), imputation R^2, and HWE among controls are provided in Supplemental Table 1.

Environmental data and harmonization procedure

Environmental and demographic exposures evaluated for GxE interaction include: sex, BMI, alcohol consumption, smoking, aspirin use, PMH use, dietary intake of calcium, fiber, folate, red meat, processed meat, fruit, and vegetables [25–35].

Data on environmental exposures were self-reported at either in-person interview or in structured self-administered questionnaires. As data collection instruments differed across studies, a multi-step, iterative data harmonization procedure was used. After the common data elements (CDEs) were identified, the questionnaires and data dictionaries of each study were examined to identify specific elements that could be mapped to these CDEs. These data elements were then written to a common data platform and then transformed via an SQL programming script, allowing these variables to be combined into a single dataset with common definitions, standardized coding, and standardized permissible values. This mapping procedure and resulting values were reviewed for quality assurance, with range and logic checks performed to assess data distributions within and between studies. After examining the data, outlying samples were truncated to the minimum or maximum value of the established range for each variable.

The harmonized alcohol variable was categorized as follows: <1 g/day, 1-<28 grams/day, or 28+ grams/day. BMI was modeled as a scaled variable (BMI [kg/m^2]/10), with underweight persons (BMI<18.5) excluded in analyses of BMI to avoid concern that underweight persons may have had occult disease at the time of exposure assessment.

Smoking was defined in two ways, a binary never/ever variable and a 5-level pack-year variable (never smoking, 4 study-specific quartiles of pack-years smoked). Aspirin use was defined as a binary variable, with yes indicating regular use of aspirin at the time of reference (with study-specific definitions varying across studies); similarly, PMH use was defined as a binary variable, with yes indicating any current use of PMH at the time of reference, and analyses of PMH use were limited to women.

All dietary variables (dietary calcium intake, dietary fiber intake, dietary folate intake, red meat consumption, processed meat consumption, vegetable consumption, fruit consumption) were categorized into quartiles. Calcium, fiber, and folate were limited to dietary intake. These quartiles were sex- and study-specific, with the coding of the quartiles corresponding to the median value of the quartile within each sex and study. After combining data across studies, we then scaled these variables to a unit reflective of the distribution of each dietary variable; the scaled units are as follows: calcium (500 mg/day), fiber (10 g/day), folate (500

mcg/day), processed meat (servings/day), red meat (servings/day), vegetable (5 servings/day), and fruit (5 servings/day). As some of the studies included in our meta-analysis collected information in categories that did not allow for conversion to these quartiles, we have also examined consumption of processed meat, red meat, vegetable, and fruit as less-rich (but more inclusive) binary variables, with the threshold between low and high consumption defined by sex-and study-specific medians. HPFS and NHS were excluded from analyses of fiber and the 4-level processed meat variable, as comparable data for these variables were not available at the time of study initiation. DACHS was excluded from analyses pertaining to the 4-level fruit and vegetable variables due to substantial differences in how these variables were assessed and defined. For all environmental exposures, the referent group corresponds to the lowest level of exposure.

Statistical analysis

Analyses of main effects of SNPs and environmental factors and GxE interaction were adjusted for age, sex, and study center. Analyses involving genetic data were further adjusted for population substructure (first 3 principal components using EIGENSTRAT [37]); analyses corresponding to the following dietary variables were further adjusted for energy intake if available: calcium, fiber, folate, fruit consumption, and vegetable consumption. Analyses of the Physicians' Health Study were further adjusted for smoking, as participants were matched on smoking status.

To assess the best model fit for each SNP, we compared an unrestricted model to log-additive, dominant, and recessive models using a likelihood ratio test [19]. All SNPs were best modeled using a log-additive model, except for rs59336; this SNP was modeled dominantly, given that the unrestricted model outperformed both the additive and recessive models.

The model form of environmental variables was also assessed. The best model form for the alcohol variable and 4-level dietary variables was assessed using a likelihood ratio test to compare a model with unrestricted categorical variables to a reduced model with a single linear variable. The likelihood ratio test indicated that modeling alcohol categorically significantly outperformed the linear alcohol variable; therefore, alcohol was modeled using unrestricted categorical variables. However, all of the 4-level dietary variables (fruit consumption, vegetable consumption, red meat consumption, processed meat consumption, fiber intake, folate intake, and calcium intake) were modeled as single linear variables, given that the unrestricted categorical variable did not outperform the linear variable. To assess the best model form for BMI ([kg/m^2]/10) and pack-years smoked (5-level variable), we used a likelihood ratio test to compare a model with and without a quadratic term; the addition of the quadratic term did not improve the model fit for either of these variables, and therefore both BMI ([kg/m^2]/10) and smoking (5-level variable) were modeled linearly.

To test for interaction, an efficient Empirical Bayes (EB) shrinkage method was used, which is a weighted sum of the case-only test and the traditional case-control method [38]. In the event that the assumption of gene-environment independence appears to hold, more weight is given to the more powerful case-only method; if this assumption is violated, more weight is given to the case-control estimate, which does not assume gene-environment

independence. This approach affords the greater power of the case-only analysis, while protecting against bias in the event of gene-environment dependence. All results for meta-analyses were obtained using a fixed-effects model, and for each meta-analysis performed, we examined the corresponding p-value for heterogeneity across studies (Supplemental Table 2).

Given that 288 tests were performed (16 SNPs*18 environmental factors) and some of the environmental variables were correlated with one another, permutation was used to account for multiple testing and correlations among variables. Each analysis was performed 2000 times using a permuted case-control status in each run, after which the Westfall and Young stepdown procedure was applied to derive an adjusted p-value for each interaction [39]. These adjusted p-values were then used to assess the presence of interaction at the alpha=0.05 level. All other p-values are termed nominal p-values.

Data harmonization was performed in SAS and T-SQL, while all other analyses were performed in R.

RESULTS

Our study population included a total of 18,440 persons, including 9160 cases and 9280 controls. Of the 18,440 persons included, 8110 (44.0%) were male and 10,330 (56.0%) were female.

The marginal associations of the SNPs with CRC risk are presented in Table 2. In this consortium of studies, of the 16 SNPs studied, 12 showed evidence of association with CRC risk as initially discovered, with p-values <0.05. Though not statistically significant, three of the remaining SNPs (rs6687758, rs6691170, and rs10936599) showed evidence of association in the expected direction [4]. However, for one SNP, rs1665650, the significant risk allele in our study (C) did not match the risk allele as it was discovered (T)[7]. One SNP, rs59336, showed evidence of heterogeneity across studies in its marginal association with CRC (p-het: 1.5×10^{-3}).

The marginal associations between the environmental factors and CRC are presented in Table 3. Increasing folate intake, NSAID use, PMH use, low alcohol intake, and increasing consumption of calcium, vegetable, fruit, and fiber were associated with reduced risk of CRC, whereas high alcohol consumption, increasing red and processed meat consumption, smoking, and high BMI were associated with increased CRC risk. The main effect of sex is not presented due to matching on this variable. As can be seen in Supplemental Table 3, the main effects of the environmental variables tend to be stronger in case-control studies than in cohort studies.

The results for the 288 gene-environment interactions tested are presented in Supplemental Table 2. In analyses adjusted for age, sex, study center, and population substructure (principal components), six interactions had a nominal p-value <0.01: rs6691170*PMH use (no/yes), rs3217810*dietary fiber intake (per 10 g/day), rs3217810*dietary folate intake (per 500 mcg/day), rs7137602*vegetable consumption (per 5 servings/day), rs10936599*sex, and rs719725*fruit consumption (high vs low) (Table 4). The strongest interaction was

between rs6691170 and PMH, with an interaction odds ratio (OR) of 1.22 (95% CI: 1.08–1.39), and a nominal p-value of 1.74×10^{-3} (p-value heterogeneity=0.18; results presented in Table 4). After accounting for multiple comparisons, the adjusted p-value for the PMH-rs6691170 interaction did not reach statistical significance (adjusted p-value=0.30) (Table 4). No other interactions were statistically significant after accounting for multiple comparisons.

DISCUSSION

In our meta-analysis of 9160 CRC cases and 9280 controls, after adjustment for multiple comparisons, we found no statistical evidence to support that the associations between recently identified susceptibility loci and CRC are modified by environmental factors, including sex, BMI, smoking, alcohol, aspirin use, PMH use, and various dietary factors.

We confirmed expected associations between CRC and environmental factors studied, as well as between CRC and 12 of the recently identified SNPs. Four variants did not replicate in this study population, including SNPs located at 1q41 (rs6687758, rs6691170), 3q26.2 (rs10936599), and 10q26.2 (rs1665650); nonetheless, the direction of association for three of these SNPs (rs6687758, rs6691170, and rs10936599) was the same in our study as prior studies [4,22,40]. However, the risk allele for rs1665650 in our study did not match the one reported [7]. This may be due to differences in the underlying linkage patterns given the ethnic differences in populations studied (the discovery study by Jia et al. was conducted among Asian populations, whereas our study included only persons of European descent). However, it remains unclear why rs6687758, rs6691170, and rs10936599 did not replicate in GECCO. It may be that the distribution of environmental factors in our population differs from that of the populations in which these genetic variants were discovered, though, as noted, none of the environmental factors studied here interacted with these genetic variants.

None of the interactions studied was statistically significant after adjustment for multiple comparisons. This may be because there is truly no interaction between these genetic and environmental factors or it may be that power is still limited to detect modest or weak interactions despite our large sample size. In our analyses of 9160 cases and 9280 controls, we are adequately powered to detect interactions with an interaction OR in the range of 1.21 to 1.29 for MAF in the observed range (0.16–0.49), assuming a main effect of 1.08 for log-additive SNPs, a main effect of 1.22 for binary environmental risk factors, and an alpha of 1.74×10^{-4} (Bonferonni p-value of 1.74×10^{-4}= [0.05/288]). However, as analyses of PMH use were limited to women (4284 cases, 4695 controls), we were underpowered to detect an OR in this range and therefore a larger sample size may be needed to more thoroughly evaluate interactions between PMH use and recently identified susceptibility loci. This evidence builds upon a prior analysis conducted within GECCO in which we examined the presence of gene-environment interaction for the first 10 identified susceptibility loci [19]. In that paper, we observed only one statistically significant gene-environment interaction, between rs16892766 (8q23.3) and vegetable consumption. Taken together, there is very little evidence for gene-environment interaction involving known susceptibility loci within GECCO, though a larger sample size may be needed to evaluate interaction. Our data suggests that GxE with known susceptibility loci may not account for the missing

heritability. It is possible, though, that perhaps more complicated multifactorial interactions account for this missing heritability. It is also possible that gene-environment interaction may be present for SNPs not identified to be associated with CRC risk through genome-wide screens of marginal SNP associations; such gene-environment interaction might only become apparent when using a genome-wide GxE approach. Currently, our consortium is investigating the presence of genome-wide GxE interaction with a variety of environmental factors. It may also be informative to evaluate GxE by anatomic subsite or by molecular characteristics, such as microsatellite instability; however, an even larger sample size would be needed for such analyses.

One of the major strengths of this study is the large sample size. This is especially important, as this is the largest study examining GxE involving these SNPs and prior studies have cited the need for a larger sample size when evaluating gene-environment interaction [20,23]. Furthermore, we used an Empirical Bayes approach so as to derive additional power from the use of case-only analyses [38]. Another advantage is that we used a standardized harmonization procedure to combine environmental data across studies.

Nonetheless, a limitation of this study is measurement error. As measurement error can bias estimates of interaction in GxE analyses [41,42], we evaluated the best model form for environmental and genetic factors to minimize the measurement error present in our variables. Regardless, harmonizing data across studies necessarily yields simpler variables, potentially leading to some loss of information in our environmental data and attenuation of effect estimates. For example, our PMH use variable is limited to a binary variable and does not incorporate information on other potentially important characteristics of use.

Furthermore, our consortium includes both retrospective and prospective studies, and these types of studies have different sources of error. The main exposure effects varied somewhat by study design (Supplemental Table 3), likely due to differential measurement error and/or selection bias in case-controls studies or the variable time period between baseline questionnaire and cancer diagnosis in the prospective studies (the average time between baseline and cancer diagnosis ranged from approximately 3–11 years across prospective studies). However, gene-exposure interactions are not subject to selection bias under the assumption that genotype does not influence participation (conditional on exposure and disease status) [43]. Despite these concerns, the associations between all environmental variables and CRC were in the expected directions. Indeed, it is notable that the environmental variables show relationships almost entirely consistent with the large body of earlier epidemiologic work [25–35]. Even so, this loss of richness of environmental data is a limitation common to consortia-based studies; however, it is this harmonization of environmental data which allows for the sample size needed to evaluate GxE.

Finally, we examined GWAS-identified SNPs and therefore our analyses do not include all genetic polymorphisms associated with CRC risk. These GWAS-identified SNPs are unlikely to be the underlying functional (i.e., disease-causing) variant; instead, they tag correlated variants that may have functional importance in CRC development. If these causal SNPs are not well tagged, a study that directly genotypes these causal SNPs would yield stronger associations [44,45] and improve power to detect GxE interactions.

In conclusion, our study suggests that the associations between recently identified CRC susceptibility loci and CRC are not strongly modified by known environmental factors. Our findings, along with those of our prior GxE paper [19] suggest that there may be limited gene-environment interaction involving the first 26 identified susceptibility loci and common CRC risk factors. However, large studies incorporating richer harmonized environmental data and causal SNPs may be needed to uncover the presence of weak to moderate gene-environment interaction. Further work is needed to evaluate the presence of genome-wide GxE interaction involving rare variants and multifactorial interaction.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

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

GECCO: The authors would like to thank all those at the GECCO Coordinating Center for helping bring together the data and people that made this project possible.

HPFS, NHS and PHS: We would like to acknowledge Patrice Soule and Hardeep Ranu of the Dana Farber Harvard Cancer Center High-Throughput Polymorphism Core who assisted in the genotyping for NHS, HPFS, and PHS under the supervision of Dr. Immaculata Devivo and Dr. David Hunter, Qin (Carolyn) Guo and Lixue Zhu who assisted in programming for NHS and HPFS, and Haiyan Zhang who assisted in programming for the PHS. We would like to thank the participants and staff of the Nurses' Health Study and the Health Professionals Follow-Up Study, for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY.

PLCO: The authors thank Drs. Christine Berg and Philip Prorok, Division of Cancer Prevention, National Cancer Institute, the Screening Center investigators and staff or the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial, Mr. Tom Riley and staff, Information Management Services, Inc., Ms. Barbara O'Brien and staff, Westat, Inc., and Drs. Bill Kopp, Wen Shao, and staff, SAIC-Frederick. Most importantly, we acknowledge the study participants for their contributions to making this study possible.

A subset of control samples were genotyped as part of the Cancer Genetic Markers of Susceptibility (CGEMS) Prostate Cancer GWAS [46], Colon CGEMS pancreatic cancer scan (PanScan) [47, 48], and the Lung Cancer and Smoking study. The prostate and PanScan study datasets were accessed with appropriate approval through the dbGaP online resource (http://cgems.cancer.gov/data/) accession numbers phs000207v.1p1 and phs000206.v3.p2, respectively, and the lung datasets were accessed from the dbGaP website (http://www.ncbi.nlm.nih.gov/gap) through accession number phs000093 v2.p2. For the lung study, the GENEVA Coordinating Center provided assistance with genotype cleaning and general study coordination, and the Johns Hopkins University Center for Inherited Disease Research conducted genotyping.

PMH: The authors would like to thank the study participants and staff of the Hormones and Colon Cancer study.

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: https://cleo.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Short%20List.pdf

Financial Support:

C.S. Carlson, M. Du, J. Gong, T.A. Harrison, L. Hsu, C.M. Hutter, S. Jiao, J. Minnier, B.M. Pflugeisen, U. Peters, D.L. Stelling, M. Thornquist, G.S. Warnick, and C.M. Ulrich are affiliated with GECCO, which is supported by the following grants from the National Cancer Institute, National Institutes of Health, U.S. Department of Health and Human Services: U01 CA137088 and R01 CA059045.

L. Le Marchand is affiliated with COLO2&3, which is supported by the National Institutes of Health (R01 CA60987).

G. Casey, J.L. Hopper, M.A. Jenkins, and P.A. Newcomb are affiliated with CCFR, which is supported by the National Institutes of Health (RFA # CA-95-011) and through cooperative agreements with members of the Colon Cancer Family Registry and P.I.s. This genome wide scan was supported by the National Cancer Institute, National Institutes of Health by U01 CA122839. The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the CFRs, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government or the CFR. The following Colon CFR centers contributed data to this manuscript and were supported by National Institutes of Health: Australasian Colorectal Cancer Family Registry (U01 CA097735), Ontario Registry for Studies of Familial Colorectal Cancer (U01 CA074783), and Seattle Colorectal Cancer Family Registry (U01 CA074794).

H. Brenner, J. Chang-Claude, M. Hoffmeister, and A. Rudolph are affiliated with DACHS, which was supported by grants from the German Research Council (Deutsche Forschungsgemeinschaft, BR 1704/6-1, BR 1704/6-3, BR 1704/6-4 and CH 117/1-1), and the German Federal Ministry of Education and Research (01KH0404 and 01ER0814).

B.J. Caan, J.D. Potter, and M.L. Slattery are affiliated with DALS, which was supported by the National Institutes of Health (R01 CA48998 to M.L. Slattery).

A.T. Chan, C.S. Fuchs, E.L. Giovannucci, and J. Ma are affiliated with HPFS, NHS, and PHS. HPFS was supported by the National Institutes of Health (P01 CA 055075, UM1 CA167552, R01 137178, and P50 CA 127003), NHS by the National Institutes of Health (R01 CA137178, P01 CA 087969, and P50 CA 127003) and PHS by the National Institutes of Health (CA42182).

B.E. Henderson, L.N. Kolonel, and L. Le Marchand are affiliated with MEC, which is supported by the following grants from the National Institutes of Health: R37 CA54281, P01 CA033619, and R01 CA63464.

M. Cotterchio, M. Lemire, and B.W. Zanke are affiliated with OFCCR, which is supported by the National Institutes of Health, through funding allocated to the Ontario Registry for Studies of Familial Colorectal Cancer (U01 CA074783); see CCFR section above. Additional funding toward genetic analyses of OFCCR includes the Ontario Research Fund, the Canadian Institutes of Health Research, and the Ontario Institute for Cancer Research, through generous support from the Ontario Ministry of Research and Innovation.

S. I. Berndt, S.J. Chanock, R.B. Hayes, and R.E. Schoen are affiliated with PLCO, which was supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics and supported by contracts from the Division of Cancer Prevention, National Cancer Institute, NIH, DHHS. Funding for the Lung Cancer and Smoking study was provided by National Institutes of Health (NIH), Genes, Environment and Health Initiative (GEI) Z01 CP 010200, NIH U01 HG004446, and NIH GEI U01 HG 004438.

P.A. Newcomb is affiliated with PMH, which is supported by the National Institutes of Health (R01 CA076366 to P.A. Newcomb).

E. White is affiliated with VITAL, which is supported in part by the National Institutes of Health (K05 CA154337) from the National Cancer Institute and Office of Dietary Supplements.

H.M. Ochs-Balcom and F. Thomas are affiliated with WHI. The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C.

P. Campbell is at the American Cancer Society (ACS) and funded through ACS.

M. Du is supported by the National Cancer Institute, National Institutes of Health (R25CA94880).

D. Duggan is affiliated with TGEN and funded through a subaward with GECCO (R01 CA059045).

E.D. Kantor is supported by the National Cancer Institute, National Institutes of Health (R25CA94880 and T32CA009001).

D. Seminara is a Senior Scientist and Consortia Coordinator at the Epidemiology and Genetics Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, National Institutes of Health.

REFERENCES

1. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. CA Cancer J Clin. 2010; 60:277–300. [PubMed: 20610543]

2. Broderick P, Carvajal-Carmona L, Pittman AM, Webb E, Howarth K, Rowan A, et al. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. Nat Genet. 2007; 39:1315–1317. [PubMed: 17934461]
3. Dunlop MG, Dobbins SE, Farrington SM, Jones AM, Palles C, Whiffin N, et al. Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. Nat Genet. 2012; 44:770–776. [PubMed: 22634755]
4. Houlston RS, Cheadle J, Dobbins SE, Tenesa A, Jones AM, Howarth K, et al. Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. Nat Genet. 2010; 42:973–977. [PubMed: 20972440]
5. Hutter CM, Slattery ML, Duggan DJ, Muehling J, Curtin K, Hsu L, et al. Characterization of the association between 8q24 and colon cancer: gene-environment exploration and meta-analysis. BMC Cancer. 2010; 10:670. [PubMed: 21129217]
6. Jaeger E, Webb E, Howarth K, Carvajal-Carmona L, Rowan A, Broderick P, et al. Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. Nat Genet. 2008; 40:26–28. [PubMed: 18084292]
7. Jia WH, Zhang B, Matsuo K, Shin A, Xiang YB, Jee SH, et al. Genome-wide association analyses in East Asians identify new susceptibility loci for colorectal cancer. Nat Genet. 2013; 45:191–196. [PubMed: 23263487]
8. Kocarnik JD, Hutter CM, Slattery ML, Berndt SI, Hsu L, Duggan DJ, et al. Characterization of 9p24 risk locus and colorectal adenoma and cancer: gene-environment interaction and meta-analysis. Cancer Epidemiol Biomarkers Prev. 2010; 19:3131–3139. [PubMed: 20978172]
9. Peters U, Hutter CM, Hsu L, Schumacher FR, Conti DV, Carlson CS, et al. Meta-analysis of new genome-wide association studies of colorectal cancer risk. Hum Genet. 2012; 131:217–234. [PubMed: 21761138]
10. Peters U, Jiao S, Schumacher FR, Hutter CM, Aragaki AK, Baron JA, et al. Identification of Genetic Susceptibility Loci for Colorectal Tumors in a Genome-Wide Meta-analysis. Gastroenterology. 2013; 144:799–807. e724. [PubMed: 23266556]
11. Tenesa A, Farrington SM, Prendergast JG, Porteous ME, Walker M, Haq N, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. Nat Genet. 2008; 40:631–637. [PubMed: 18372901]
12. Tomlinson I, Webb E, Carvajal-Carmona L, Broderick P, Kemp Z, Spain S, et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. Nat Genet. 2007; 39:984–988. [PubMed: 17618284]
13. Tomlinson IP, Webb E, Carvajal-Carmona L, Broderick P, Howarth K, Pittman AM, et al. A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. Nat Genet. 2008; 40:623–630. [PubMed: 18372905]
14. Zanke BW, Greenwood CM, Rangrej J, Kustra R, Tenesa A, Farrington SM, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. Nat Genet. 2007; 39:989–994. [PubMed: 17618283]
15. Houlston RS, Webb E, Broderick P, Pittman AM, Di Bernardo MC, Lubbe S, et al. Meta-analysis of genome- wide association data identifies four new susceptibility loci for colorectal cancer. Nat Genet. 2008; 40:1426–1435. [PubMed: 19011631]
16. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, 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; 343:78–85. [PubMed: 10891514]
17. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009; 461:747–753. [PubMed: 19812666]
18. Thomas D. Gene--environment-wide association studies: emerging approaches. Nat Rev Genet. 2010; 11:259–272. [PubMed: 20212493]
19. Hutter CM, Chang-Claude J, Slattery ML, Pflugeisen BM, Lin Y, Duggan D, et al. Characterization of gene environment interactions for colorectal cancer susceptibility loci. Cancer Res. 2012; 72:2036–2044. [PubMed: 22367214]
20. Figueiredo JC, Lewinger JP, Song C, Campbell PT, Conti DV, Edlund CK, 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; 20:758–766. [PubMed: 21357381]

21. He J, Wilkens LR, Stram DO, Kolonel LN, Henderson BE, Wu AH, et al. Generalizability and epidemiologic characterization of eleven colorectal cancer GWAS hits in multiple populations. Cancer Epidemiol Biomarkers Prev. 2011; 20:70–81. [PubMed: 21071539]
22. Lubbe SJ, Di Bernardo MC, Broderick P, Chandler I, Houlston RS. Comprehensive evaluation of the impact of 14 genetic variants on colorectal cancer phenotype and risk. Am J Epidemiol. 2012; 175:1–10. [PubMed: 22156018]
23. Siegert S, Hampe J, Schafmayer C, von Schönfels W, Egberts JH, Försti A, et al. Genome-wide investigation of gene-environment interactions in colorectal cancer. Hum Genet. 2013; 132:219–231. [PubMed: 23114982]
24. von Holst S, Picelli S, Edler D, Lenander C, Dalén J, Hjern F, et al. Association studies on 11 published colorectal cancer risk loci. Br J Cancer. 2010; 103:575–580. [PubMed: 20648012]
25. Aune D, Chan DS, Lau R, Vieira R, Greenwood DC, Kampman E, et al. Dietary fibre, whole grains, and risk of colorectal cancer: systematic review and dose-response meta-analysis of prospective studies. BMJ. 2011; 343:d6617. [PubMed: 22074852]
26. Chan DS, Lau R, Aune D, Vieira R, Greenwood DC, Kampman E, et al. Red and processed meat and colorectal cancer incidence: meta-analysis of prospective studies. PLoS One. 2011; 6:e20456. [PubMed: 21674008]
27. Fedirko V, Tramacere I, Bagnardi V, Rota M, Scotti L, Islami F, et al. Alcohol drinking and colorectal cancer risk: an overall and dose-response meta-analysis of published studies. Ann Oncol. 2011; 22:1958–1972. [PubMed: 21307158]
28. Huncharek M, Muscat J, Kupelnick B. Colorectal cancer risk and dietary intake of calcium, vitamin D, and dairy products: a meta-analysis of 26,335 cases from 60 observational studies. Nutr Cancer. 2009; 61:47–69. [PubMed: 19116875]
29. Johnson CM, Wei C, Ensor JE, Smolenski DJ, Amos CI, Levin B, et al. Meta-analyses of colorectal cancer risk factors. Cancer Causes Control. 2013; 24:1207–1222. [PubMed: 23563998]
30. Kim DH, Smith-Warner SA, Spiegelman D, Yaun SS, Colditz GA, Freudenheim JL, et al. Pooled analyses of 13 prospective cohort studies on folate intake and colon cancer. Cancer Causes Control. 2010; 21:1919–1930. [PubMed: 20820900]
31. Ma Y, Yang Y, Wang F, Zhang P, Shi C, Zou Y, et al. Obesity and risk of colorectal cancer: a systematic review of prospective studies. PLoS One. 2013; 8:e53916. [PubMed: 23349764]
32. Nelson HD, Humphrey LL, Nygren P, Teutsch SM, Allan JD. Postmenopausal hormone replacement therapy: scientific review. JAMA. 2002; 288:872–881. [PubMed: 12186605]
33. Nguyen SP, Bent S, Chen YH, Terdiman JP. Gender as a risk factor for advanced neoplasia and colorectal cancer: a systematic review and meta-analysis. Clin Gastroenterol Hepatol. 2009; 7:676–681. e1-3. [PubMed: 19514116]
34. Rothwell PM, Wilson M, Elwin CE, Norrving B, Algra A, Warlow CP, et al. Long-term effect of aspirin on colorectal cancer incidence and mortality: 20-year follow-up of five randomised trials. Lancet. 2010; 376:1741–1750. [PubMed: 20970847]
35. Tsoi KK, Pau CY, Wu WK, Chan FK, Griffiths S, Sung JJ, et al. Cigarette smoking and the risk of colorectal cancer: a meta-analysis of prospective cohort studies. Clin Gastroenterol Hepatol. 2009; 7:682–688. e1-5. [PubMed: 19245853]
36. Jiao S, Hsu L, Berndt S, Bezieau S, Brenner H, Buchanan D, et al. Genome-wide search for gene-gene interactions in colorectal cancer. PLoS One. 2012; 7:e52535. [PubMed: 23300701]
37. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects For stratification in genome-wide association studies. Nat Genet. 2006; 38:904–909. [PubMed: 16862161]
38. Mukherjee B, Chatterjee N. Exploiting gene-environment independence for analysis of case-control studies: an empirical Bayes-type shrinkage estimator to trade-off between bias and efficiency. Biometrics. 2008; 64:685–694. [PubMed: 18162111]
39. Westfall, PH.; Young, SS. Probability and Mathematical Statistics. New York: Wiley; 1993. Resampling-based multiple testing: examples and methods for p-value adjustment.

40. Spain SL, Carvajal-Carmona LG, Howarth KM, Jones AM, Su Z, Cazier JB, et al. Refinement of the associations between risk of colorectal cancer and polymorphisms on chromosomes 1q41 and 12q13.13. Hum Mol Genet. 2012; 21:934–946. [PubMed: 22076443]
41. Garcia-Closas M, Rothman N, Lubin J. Misclassification in case-control studies of gene-environment interactions: assessment of bias and sample size. Cancer Epidemiol Biomarkers Prev. 1999; 8:1043–1050. [PubMed: 10613335]
42. Prentice RL. Empirical evaluation of gene and environment interactions: methods and potential. J Natl Cancer Inst. 2011; 103:1209–1210. [PubMed: 21791675]
43. Morimoto LM, White E, Newcomb PA. Selection bias in the assessment of gene-environment interaction in case-control studies. Am J Epidemiol. 2003; 158:259–263. [PubMed: 12882948]
44. Hein R, Beckmann L, Chang-Claude J. Sample size requirements for indirect association studies of gene environment interactions (G x E). Genet Epidemiol. 2008; 32:235–245. [PubMed: 18163529]
45. Nickels S, Truong T, Hein R, Stevens K, Buck K, Behrens S, et al. Evidence of gene-environment interactions between common breast cancer susceptibility loci and established environmental risk factors. PLoS Genet. 2013; 9:e1003284. [PubMed: 23544014]
46. Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, Wacholder S, et al. Genome-wide association study of Prostate cancer identifies a second risk locus at 8q24. Nat. Genet. 2007; 39:645–649. [PubMed: 17401363]
47. Amundadottir L, Kraft P, Stolzenberg-Solomon RZ, Fuchs CS, Petersen GM, Arslan AA, et al. Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. Nat Genet. 2009; 41:986–990. [PubMed: 19648918]
48. Petersen GM, Amundadottir L, Fuchs CS, Kraft P, Stolzenberg-Solomon RZ, Jacobs KB, et al. A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. Nat Genet. 2010; 42:224–228. [PubMed: 20101243]

Cancer Epidemiol Biomarkers Prev. Author manuscript; available in PMC 2015 September 01.

Table 1

General characteristics of included studies

Study	Study Design	Case N	Control N	Male N	Female N	Age Range	Colon Cancer N	Rectal Cancer N	Total N
CCFR	Case-control	1163	977	1072	1068	20–81	445	286	2140
Colo 2&3	Case-control	87	125	117	95	38–86	59	27	212
DACHS	Case-control	2376	2206	2752	1830	33–99	1422	949	4582
DALS	Case-control	1116	1174	1261	1029	28–79	1112	0	2290
HPFS	Nested case-control	173	230	403	0	48–81	113	41	403
MEC	Nested case-control	328	346	361	313	45–76	241	81	674
NHS	Nested case-control	375	955	0	1330	44–69	285	86	1330
PHS	Nested case-control	375	389	764	0	40–85	286	84	764
PLCO	Nested case-control	486	415	518	383	55–75	320	161	901
PMH-CCFR	Case-control	280	122	0	402	48–73	206	64	402
OFCCR	Case-control	650	522	562	610	29–77	433	197	1172
VITAL	Nested case-control	285	288	300	273	50–76	215	66	573
WHI	Nested case-control	1466	1531	0	2997	50–79	1149	261	2997
Total		*9160*	*9280*	*8110*	*10330*	*20–99*	*6286*	*2303*	*18440*

Table 2

Associations between recently identified single nucleotide polymorphisms and colorectal cancer in the Genetics and Epidemiology of Colorectal Cancer Consortium and Colon Cancer Family Registry

SNP[a, b]	Chromosomal Location	Gene/ Locus	Risk[c] Allele	Base[c] Allele	OR[d]	95% CI	p-value	pv.het	# Study Sets genotyped	# Study sets mputed	Mean RAF	Mean R^2
rs10911251	1q25.3	*LAMC1*	A	C	1.11	1.06–1.16	1.0×10^{-5}	0.56	0	17	0.57	0.88
rs6687758	1q41	*DUSP10*	G	A	1.04	0.99–1.10	0.11	0.09	16	1	0.20	1.00
rs6691170	1q41	*DUSP10*	T	G	1.01	0.97–1.06	0.57	0.25	3	14	0.37	0.98
rs11903757	2q32.3	*NABP1/ SDPR*	C	T	1.14	1.07–1.23	1.8×10^{-4}	0.32	0	17	0.17	0.69
rs10936599	3q26.2	*MYNN*	C	T	1.02	0.97–1.07	0.45	0.74	8	9	0.76	0.98
rs647161	5q31.1	*PITX1/ H2AFY*	A	C	1.07	1.02–1.12	8.5×10^{-3}	0.06	0	17	0.67	0.88
rs1321311	6p21	*SRSF3/ CDKN1A*	A	C	1.07	1.02–1.13	4.2×10^{-3}	0.27	7	10	0.25	0.96
rs719725	9p24	*TPD52L3/ IL-33/ UHRF2/ GLDC*	A	C	1.08	1.03–1.13	7.1×10^{-4}	0.28	0	17	0.62	1.00
rs1665650	10q26.2	*HSPA12A*	T	C	0.95	0.91–1.00	4.9×10^{-2}	0.20	0	17	0.27	0.97
rs3824999	11q13.4	*POLD3*	G	T	1.10	1.06–1.15	7.0×10^{-6}	0.61	3	14	0.51	1.00
rs3217810	12p13.32	*CCND2*	T	C	1.19	1.10–1.29	3.1×10^{-5}	0.84	3	14	0.16	0.49
rs7136702	12q13.13	*LARP4/ DIP2B*	T	C	1.10	1.05–1.16	4.9×10^{-5}	0.45	3	14	0.32	0.87
rs11169552	12q13.13	*DIP2B/ ATF1*	C	T	1.05	1.00–1.10	4.0×10^{-2}	0.51	16	1	0.73	1.00
rs59336	12q24.21	*TBX3*	T	A	1.15	1.07–1.23	1.4×10^{-4}	1.5×10^{-3}	0	17	0.48	0.94
rs2423279	20p12.3	*HAO1/ PLCB1*	C	T	1.07	1.02–1.12	7.5×10^{-3}	0.19	10	7	0.25	1.00
rs4925386	20q13.33	*LAMA5*	C	T	1.06	1.01–1.11	1.5×10^{-2}	0.45	16	1	0.69	1.00

ABBREVIATIONS: OR (odds ratio); RAF (risk allele frequency)

[a] All SNPs modeled additively, with the exception of rs59336, which was modeled dominantly

[b] SNPs identified to be associated with colorectal cancer risk in the following studies: rs10911251 (Peters et al. Gasteroenterology, 2013 [10]); rs6687758 (Houlston et al. Nat Genet, 2010 [4]); rs6691170 (Houlston et al. Nat Genet, 2010 [4]); rs11903757 (Peters et al. Gasteroenterology, 2013 [10]); rs10936599 (Houlston et al. Nat Genet, 2010 [4]); rs647161 (Jia et al. Nat Genet, 2013 [7]); rs1321311

(Dunlop et al. Nat Genet, 2012 [3]); rs719725 (Zanke et a., Nat Genet, 2007 [14]; Kocarnik et al. Cancer Epidemiol Biomarkers Prev, 2010 [8]); rs1665650 (Jia et al. Nat Genet, 2013 [2013]); rs3824999 (Dunlop et al. Nat Genet, 2012 [3]); rs3217810 (Peters et al. Gasteroenterology, 2013 [10]); rs7136702 (Houlston et al. Nat Genet, 2010 [4]); rs11169552 (Houlston et al. Nat Genet, 2010 [4]); rs59336 (Peters et al. Gasteroenterology, 2013 [10]); rs2423279 (Jia et al. Nat Genet, 2013 [7]); rs4925386 (Houlston et al. Nat Genet, 2010 [4])

[c] Risk/base allele designation based on the literature

[d] Adjusted for age, sex, study center, and population substructure (principal components 1–3)

Table 3

Association between environmental factors and CRC in GECCO

Environmental Variables	OR[a]	95% CI	p-value	pv.het
BMI (per 10 kg/m^2)	1.43	1.34–1.53	1.0×10^{-25}	3.4×10^{-5}
Alcohol 1-28g vs. none	0.90	0.83–0.97	6.9×10^{-3}	0.21
Alcohol 28+g vs. none	1.21	1.07–1.37	2.3×10^{-3}	0.28
Smoking (ever vs. never)	1.21	1.14–1.29	7.8×10^{-10}	0.67
Smoking (per increase in pack-year grouping)[b]	1.09	1.06–1.11	1.0×10^{-14}	0.19
Aspirin Use (yes vs. no during reference year)	0.71	0.67–0.76	8.0×10^{-25}	7.2×10^{-4}
PMH Use (yes vs. no at referent time)	0.69	0.63–0.76	5.1×10^{-14}	0.16
Dietary calcium (per 500 mg/day)[c]	0.80	0.75–0.85	4.3×10^{-13}	0.28
Dietary fiber (per 10 g/day)[c]	0.83	0.76–0.90	2.0×10^{-5}	0.42
Dietary folate (per 500 mcg/day)[c]	0.70	0.59–0.83	5.3×10^{-5}	0.45
Red meat (per serving/day)	1.33	1.23–1.44	7.9×10^{-13}	2.9×10^{-6}
Red meat (upper vs. lower half)	1.25	1.18–1.34	2.0×10^{-12}	3.7×10^{-3}
Processed meat (servings/day)	1.48	1.30–1.70	1.0×10^{-8}	8.1×10^{-3}
Processed meat (upper vs. lower half)	1.21	1.13–1.30	7.1×10^{-8}	5.9×10^{-3}
Fruit (per 5 servings/day)[c]	0.82	0.69–0.97	0.02	0.79
Fruit (upper vs. lower half)[c]	0.83	0.78–0.89	2.5×10^{-8}	8.9×10^{-3}
Vegetable (per 5 servings/day)[c]	0.82	0.70–0.95	9.0×10^{-3}	0.03
Vegetable (upper vs. lower half)[c]	0.86	0.80–0.92	2.8×10^{-5}	0.15

ABBREVIATION: PMH (post-menopausal hormone); OR (odds ratio)

[a] Analyses adjusted for age, sex, and study center

[b] Pack-year variable categorized into five groups: never smokers and study-specific quartiles of pack-years smoked

[c] Analyses further adjusted for energy intake where available

Table 4

Gene-environment interactions with nominal interaction p-value <0.01

SNP/Chromosomal Location Environmental Variable	Gene/ Locus	OR[a,b]	CI	Nominal p-value	Adjusted p-value	pv.het
rs6691170/1q41 PMH use at reference (yes/no)	*DUSP10*	1.22	1.08–1.39	1.74×10^{-3}	0.30	0.18
rs3217810/12p13.32[c] Dietary fiber (per 10 g/day)	*CCND2*	0.77	0.65–0.91	2.98×10^{-3}	0.45	0.20
rs3217810/12p13.32[c] Dietary folate (per 500 mcg/day)	*CCND2*	0.60	0.42–0.85	4.11×10^{-3}	0.56	0.12
rs7136702/12q13.13[c] Vegetable consumption (per 5 servings/day)	*LARP4/ DIP2B*	1.28	1.07–1.54	7.39×10^{-3}	0.77	0.68
rs10936599/3q26.2 Sex (female/male)	*MYNN*	0.86	0.77–0.96	7.73×10^{-3}	0.78	0.31
rs719725/9p24[c] Fruit consumption (high vs. low)	*TPD52L3/ IL-33/ UHRF2/ GLDC*	1.10	1.02–1.19	9.59×10^{-3}	0.86	0.90

ABBREVIATIONS: PMH (post-menopausal hormone use); OR (odds ratio)

[a] Interaction OR for SNP (log-additive for number of risk alleles) * exposure (as categorized above)

[b] Adjusted for age, sex, study center, and population substructure (principal components 1–3)

[c] Analyses further adjusted for energy intake (where available)

Cancer Epidemiol Biomarkers Prev. Author manuscript; available in PMC 2015 September 01.