Gene–Environment Interaction Involving Recently Identified Colorectal Cancer Susceptibility Loci

Elizabeth D. Kantor1,2,3, Carolyn M. Hutter4, Jessica Minnier2,5, Sonja I. Berndt6, Hermann Brenner7,8, Bette J. Caan9, Peter T. Campbell10, Christopher S. Carlson2,3, Graham Casey11, Andrew T. Chan12,13, Jenny Chang-Claude14, Stephen J. Chanock6, Michelle Cotterchio15, Mengmeng Du2,3,13, David Duggan16, Edward L. Giovannucci1,13,18, Jian Gong2, Tabitha A. Harrison2, Richard B. Hayes19, Brian E. Henderson20, Michael Hoffmeister7, John L. Hopper21, Mark A. Jenkins21, Shuo Jiao2, Laurence N. Kolone22, Loïc Le Marchand22, Mathieu Lemire23, Jing Ma13, Polly A. Newcomb2,3, Cornelia M. Ulrich2,3,30, Greg S. Warnick2, Brent W. Zanke31, Ulrike Peters2,3, Li Hsu2,32, and Emily White2,3

Abstract

Background: Genome-wide association studies have identified several single nucleotide polymorphisms (SNPs) that are associated with risk of colorectal cancer. 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, rs3824999, rs7136702, rs11169552, rs59336, rs3217810, rs4925386, and rs2423279.

Methods: Data on 9,160 cases and 9,280 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, postmenopausal 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 colorectal cancer 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 colorectal cancer taken one at a time. Cancer Epidemiol Biomarkers Prev; 23(9); 1–10. ©2014 AACR.
Introduction

Colorectal cancer 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 colorectal cancer, as finding evidence of gene-environment (G × E) interaction may help guide future prevention strategies. Furthermore, understanding G × E interaction may shed light on the mechanisms by which genetic polymorphisms affect risk of colorectal cancer, as well as the underlying biology of this disease. The SNPs identified to be associated with colorectal cancer thus far only account for a small fraction of the estimated heritability of colorectal cancer (15, 16), and it has been suggested that one factor contributing to this “missing heritability” is G × E interaction (17, 18).

We previously reported on G × E interaction for the first 10 identified susceptibility loci (19). Since the time of that publication, 16 additional SNPs have been associated with colorectal cancer, 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), or 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 G × E interactions were observed in these studies; however, statistical power to detect interaction may have been limited because of insufficient sample sizes.

We have therefore evaluated whether environmental risk factors for colorectal cancer modify the associations between these genetic polymorphisms and colorectal cancer risk using data on 9,160 cases and 9,280 controls in the Colon 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, postmenopausal 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 colorectal cancer 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), CCFR, Colorectal Cancer Studies 2&3 (Colo2&3), and the PMH study within the CCFR (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 Supplementary Methods. All participants gave informed consent and studies were approved by their respective Institutional Review Boards.

Outcome

Colorectal cancer 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 Supplementary Methods.

Genotype data

G × E interaction was evaluated for 16 SNPs located at recently identified colorectal cancer 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). 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), or 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 G × E interactions were observed in these studies; however, statistical power to detect interaction may have been limited because of insufficient sample sizes.

We have therefore evaluated whether environmental risk factors for colorectal cancer modify the associations between these genetic polymorphisms and colorectal cancer risk using data on 9,160 cases and 9,280 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, postmenopausal 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 colorectal cancer risk (25-35).
and imputed SNPs are instead coded in terms of the number of risk alleles, with 0 corresponding to genotyped is provided in Table 2. All SNPs are presented (average $R^2$, Table 2). For each SNP included in our analyses, 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 Supplementary Table S1.

### Environmental data and harmonization procedure

Environmental and demographic exposures evaluated for $G \times E$ 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 multistep, iterative data harmonization procedure was used. After the common data elements (CDE) 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 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: $<1$, 1 to $<28$, or $28+ \text{g/d}$. 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.
Table 2. Associations between recently identified SNPs and colorectal cancer in the GECCO and CCFR

<table>
<thead>
<tr>
<th>SNPa,b</th>
<th>Chromosomal location</th>
<th>Gene/locus</th>
<th>Riskc allele</th>
<th>Basec allele</th>
<th>ORd</th>
<th>95% CI</th>
<th>P</th>
<th>Pmeta</th>
<th>No. of study sets genotyped</th>
<th>No. of study sets imputed</th>
<th>MeanRAF</th>
<th>MeanR2</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs10911251</td>
<td>1q25.3</td>
<td>LAMC1</td>
<td>A</td>
<td>C</td>
<td>1.1</td>
<td>1.06-1.16</td>
<td>1.0×10^-5</td>
<td>0.56</td>
<td>0</td>
<td>17</td>
<td>0.57</td>
<td>0.88</td>
</tr>
<tr>
<td>rs6687758</td>
<td>1q41</td>
<td>DUSP10</td>
<td>G</td>
<td>A</td>
<td>1.04</td>
<td>0.99-1.10</td>
<td>0.11</td>
<td>0.09</td>
<td>16</td>
<td>1</td>
<td>0.20</td>
<td>1.00</td>
</tr>
<tr>
<td>rs6691170</td>
<td>1q41</td>
<td>DUSP10</td>
<td>T</td>
<td>G</td>
<td>1.01</td>
<td>0.97-1.06</td>
<td>0.57</td>
<td>0.25</td>
<td>3</td>
<td>14</td>
<td>0.37</td>
<td>0.98</td>
</tr>
<tr>
<td>rs11903757</td>
<td>2q32.3</td>
<td>NABP1/SDPR</td>
<td>C</td>
<td>T</td>
<td>1.14</td>
<td>1.07-1.23</td>
<td>1.8×10^-4</td>
<td>0.32</td>
<td>0</td>
<td>17</td>
<td>0.17</td>
<td>0.69</td>
</tr>
<tr>
<td>rs10936599</td>
<td>3q26.2</td>
<td>MYNN</td>
<td>C</td>
<td>T</td>
<td>1.02</td>
<td>0.97-1.07</td>
<td>0.45</td>
<td>0.74</td>
<td>8</td>
<td>9</td>
<td>0.76</td>
<td>0.98</td>
</tr>
<tr>
<td>rs647161</td>
<td>5q31.1</td>
<td>PTX1/H2AFY</td>
<td>A</td>
<td>C</td>
<td>1.07</td>
<td>1.02-1.12</td>
<td>8.5×10^-3</td>
<td>0.06</td>
<td>0</td>
<td>17</td>
<td>0.67</td>
<td>0.88</td>
</tr>
<tr>
<td>rs1321311</td>
<td>6p21</td>
<td>SRSF3/CDKN1A</td>
<td>A</td>
<td>C</td>
<td>1.07</td>
<td>1.02-1.13</td>
<td>4.2×10^-4</td>
<td>0.27</td>
<td>7</td>
<td>10</td>
<td>0.25</td>
<td>0.96</td>
</tr>
<tr>
<td>rs719725</td>
<td>9p24</td>
<td>TPD52L3/IL-33/ UHRF2/GLDC</td>
<td>A</td>
<td>C</td>
<td>1.08</td>
<td>1.03-1.13</td>
<td>7.1×10^-4</td>
<td>0.28</td>
<td>0</td>
<td>17</td>
<td>0.62</td>
<td>1.00</td>
</tr>
<tr>
<td>rs1665650</td>
<td>10q26.2</td>
<td>HSP112A</td>
<td>T</td>
<td>C</td>
<td>0.95</td>
<td>0.91-1.00</td>
<td>4.9×10^-2</td>
<td>0.20</td>
<td>0</td>
<td>17</td>
<td>0.27</td>
<td>0.97</td>
</tr>
<tr>
<td>rs3824999</td>
<td>11q13.4</td>
<td>POLD3</td>
<td>G</td>
<td>T</td>
<td>1.10</td>
<td>1.06-1.15</td>
<td>7.0×10^-6</td>
<td>0.61</td>
<td>3</td>
<td>14</td>
<td>0.51</td>
<td>1.00</td>
</tr>
<tr>
<td>rs3217810</td>
<td>12p13.32</td>
<td>CCND2</td>
<td>T</td>
<td>C</td>
<td>1.19</td>
<td>1.10-1.29</td>
<td>3.1×10^-5</td>
<td>0.84</td>
<td>3</td>
<td>14</td>
<td>0.16</td>
<td>0.49</td>
</tr>
<tr>
<td>rs7136702</td>
<td>12q13.13</td>
<td>LARP4/DIP2B</td>
<td>T</td>
<td>C</td>
<td>1.10</td>
<td>1.05-1.16</td>
<td>4.9×10^-5</td>
<td>0.45</td>
<td>3</td>
<td>14</td>
<td>0.32</td>
<td>0.87</td>
</tr>
<tr>
<td>rs11169552</td>
<td>12q13.13</td>
<td>DIP2B/ATF1</td>
<td>C</td>
<td>T</td>
<td>1.05</td>
<td>1.00-1.10</td>
<td>4.0×10^-2</td>
<td>0.51</td>
<td>16</td>
<td>1</td>
<td>0.73</td>
<td>1.00</td>
</tr>
<tr>
<td>rs59336</td>
<td>12q24.21</td>
<td>TBX3</td>
<td>T</td>
<td>A</td>
<td>1.15</td>
<td>1.07-1.23</td>
<td>1.4×10^-4</td>
<td>1.5×10^-3</td>
<td>0</td>
<td>17</td>
<td>0.48</td>
<td>0.94</td>
</tr>
<tr>
<td>rs2423279</td>
<td>20p12.3</td>
<td>HAO1/PLCB1</td>
<td>C</td>
<td>T</td>
<td>1.07</td>
<td>1.02-1.12</td>
<td>7.5×10^-3</td>
<td>0.19</td>
<td>10</td>
<td>7</td>
<td>0.25</td>
<td>1.00</td>
</tr>
<tr>
<td>rs4925386</td>
<td>20q13.33</td>
<td>LAMA5</td>
<td>C</td>
<td>T</td>
<td>1.06</td>
<td>1.01-1.11</td>
<td>1.5×10^-2</td>
<td>0.45</td>
<td>1</td>
<td>1</td>
<td>0.69</td>
<td>1.00</td>
</tr>
</tbody>
</table>

aAll SNPs modeled additively, with the exception of rs59336, which was modeled dominantly.
bSNPs identified to be associated with colorectal cancer risk in the following studies: rs10911251 [Peters et al. (10)]; rs6687758 [Houlston et al. (4)]; rs6691170 [Houlston et al. (4)]; rs11903757 [Peters et al. (10)]; rs10936599 [Houlston et al. (4)]; rs647161 [Jia et al. (7)]; rs1321311 [Dunlop et al. (3)]; rs719725 [Zanke et al. (14) and Kocarnik et al. (8)]; rs1665650 [Jia et al. (2013)]; rs3824999 [Dunlop et al. (3)]; rs3217810 [Peters et al. (10)]; rs7136702 [Houlston et al. (4)]; rs11169552 [Houlston et al. (4)]; rs59336 [Peters et al. (10)]; rs2423279 [Jia et al. (7)]; rs4925386 [Houlston et al. (4)].
cRisk/base allele designation based on the literature.
dAdjusted for age, sex, study center, and population substructure (principal components 1–3).
Smoking was defined in 2 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; 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/d), fiber (10 g/d), folate (500 mcg/d), 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, 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, 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 dietary variables were 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/m²)/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/m²)/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 G × E independence seems 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 G × E independence. This approach affords the greater power of the case-only analysis, while protecting against bias in the event of G × E 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 (Supplementary Table S2).

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 2,000 times using a permuted case–control status in each run, after which the Westfall and Young step-down 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 α = 0.05 level. All other P values are termed nominal P values.

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

**Results**

Our study population included a total of 18,440 persons, including 9,160 cases and 9,280 controls. Of the 18,440 persons included, 8,110 (44.0%) were male and 10,330 (56.0%) were female.

The marginal associations of the SNPs with colorectal cancer risk are presented in Table 2. In this consortium of studies, of the 16 SNPs studied, 12 showed evidence of association with colorectal cancer risk as initially

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/m²)/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/m²)/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 G × E independence seems 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 G × E independence. This approach affords the greater power of the case-only analysis, while protecting against bias in the event of G × E 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 (Supplementary Table S2).

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 2,000 times using a permuted case–control status in each run, after which the Westfall and Young step-down 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 α = 0.05 level. All other P values are termed nominal P values.

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

**Results**

Our study population included a total of 18,440 persons, including 9,160 cases and 9,280 controls. Of the 18,440 persons included, 8,110 (44.0%) were male and 10,330 (56.0%) were female.

The marginal associations of the SNPs with colorectal cancer risk are presented in Table 2. In this consortium of studies, of the 16 SNPs studied, 12 showed evidence of association with colorectal cancer risk as initially
discovered, with P values <0.05. Although not statistically significant, 3 of the remaining SNPs (rs66877758, rs6691170, and rs10936599) showed evidence of association in the expected direction (4). However, for 1 SNP, rs1665650, the significant risk allele in our study (C) did not match the risk allele as it was discovered (T; ref. 7). One SNP, rs59336, showed evidence of heterogeneity across studies in its marginal association with colorectal cancer ($P_{	ext{het}} = 1.5 \times 10^{-5}$).

The marginal associations between the environmental factors and colorectal cancer 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 colorectal cancer, whereas high alcohol consumption, increasing red and processed meat consumption, smoking, and high BMI were associated with increased colorectal cancer risk. The main effect of sex is not presented because of matching on this variable. As can be seen in Supplementary Table S3, the main effects of the environmental variables tend to be stronger in case–control studies than in cohort studies.

The results for the 288 G \( \times \) E interactions tested are presented in Supplementary Table S2. In analyses adjusted for age, sex, study center, and population substructure (principal components), 6 interactions had a nominal $P$-value <0.01: rs6691170$^\text{a}$PMH use (no/yes), rs3217810$^\text{a}$dietary fiber intake (per 10 g/d), rs3217810$^\text{a}$dietary folate intake (per 500 mcg/d), rs7137602$^\text{a}$vegetable consumption (per 5 servings/day), rs10936599$^\text{a}$sex, and rs719725$^\text{a}$fruit consumption (high vs. low; Table 4). The strongest interaction was between rs6691170 and PMH, with an interaction odds ratio (OR) of $1.22 (95\% \text{ CI}, 1.08–1.39)$, and a nominal $P$ value of $1.74 \times 10^{-3}$ ($P_{\text{het}} = 0.18$; 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 9,160 colorectal cancer cases and 9,280 controls, after adjustment for multiple comparisons, we found no statistical evidence to support that the associations between recently identified susceptibility loci and colorectal cancer are modified by environmental factors, including sex, BMI, smoking, alcohol, aspirin use, PMH use, and various dietary factors.

We confirmed expected associations between colorectal cancer and environmental factors studied, as well as between colorectal cancer and 12 of the recently identified SNPs. Four variants did not replicate in this study population, including SNPs located at 1q41 (rs66877758, rs6691170), 3q26.2 (rs10936599), and 10q26.2 (rs1665650); nonetheless, the direction of association for 3 of these SNPs (rs66877758, 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

<table>
<thead>
<tr>
<th>Table 3. Association between environmental factors and colorectal cancer in GECCO</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Environmental variables</strong></td>
</tr>
<tr>
<td>---------------------------------------------</td>
</tr>
<tr>
<td>BMI (per 10 kg/m$^2$)</td>
</tr>
<tr>
<td>Alcohol 1–28 g vs. none</td>
</tr>
<tr>
<td>Alcohol 29+ g vs. none</td>
</tr>
<tr>
<td>Smoking (ever vs. never)</td>
</tr>
<tr>
<td>Smoking (per increase in pack-year grouping)$^b$</td>
</tr>
<tr>
<td>Aspirin use (yes vs. no during reference year)</td>
</tr>
<tr>
<td>PMH use (yes vs. no at referent time)</td>
</tr>
<tr>
<td>Dietary calcium (per 500 mg/d)$^c$</td>
</tr>
<tr>
<td>Dietary fiber (per 10 g/d)$^c$</td>
</tr>
<tr>
<td>Dietary folate (per 500 mcg/d)$^c$</td>
</tr>
<tr>
<td>Red meat (per serving/day)</td>
</tr>
<tr>
<td>Red meat (upper vs. lower half)</td>
</tr>
<tr>
<td>Processed meat (servings/day)</td>
</tr>
<tr>
<td>Processed meat (upper vs. lower half)</td>
</tr>
<tr>
<td>Fruit (per 5 servings/day)$^c$</td>
</tr>
<tr>
<td>Fruit (upper vs. lower half)$^c$</td>
</tr>
<tr>
<td>Vegetable (per 5 servings/day)$^c$</td>
</tr>
<tr>
<td>Vegetable (upper vs. lower half)$^c$</td>
</tr>
</tbody>
</table>

$^a$Analyses adjusted for age, sex, and study center.

$^b$Pack-year variable categorized into 5 groups: never smokers and study-specific quartiles of pack-years smoked.

$^c$Analyses further adjusted for energy intake where available.
is investigating the presence of genome-wide \( G \times E \) interactions with nominal interaction \( P \) value <0.01

<table>
<thead>
<tr>
<th>SNP/chromosomal location environmental variable</th>
<th>Gene/locus</th>
<th>OR (^{a,b})</th>
<th>CI</th>
<th>Nominal ( P ) value</th>
<th>Adjusted ( P ) value</th>
<th>( P_{\text{net}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs6691170/1q41/PMH use at reference (yes/no)</td>
<td>DUSP10</td>
<td>1.22</td>
<td>1.08–1.39</td>
<td>( 1.74 \times 10^{-3} )</td>
<td>0.30</td>
<td>0.18</td>
</tr>
<tr>
<td>rs3217810/12p13.32/ Dietary fiber (per 10 g/d)</td>
<td>CCND2</td>
<td>0.77</td>
<td>0.65–0.91</td>
<td>( 2.98 \times 10^{-3} )</td>
<td>0.45</td>
<td>0.20</td>
</tr>
<tr>
<td>rs3217810/12p13.32/ Dietary folate (per 500 mcg/d)</td>
<td>CCND2</td>
<td>0.60</td>
<td>0.42–0.85</td>
<td>( 4.11 \times 10^{-3} )</td>
<td>0.56</td>
<td>0.12</td>
</tr>
<tr>
<td>rs7136702/12q13.13/ Vegetable consumption (per 5 servings/day)</td>
<td>LARP4/DIP2B</td>
<td>1.28</td>
<td>1.07–1.54</td>
<td>( 7.39 \times 10^{-3} )</td>
<td>0.77</td>
<td>0.68</td>
</tr>
<tr>
<td>rs10936599/3q26.2/ Sex (female/male)</td>
<td>MYNN</td>
<td>0.86</td>
<td>0.77–0.96</td>
<td>( 7.73 \times 10^{-3} )</td>
<td>0.78</td>
<td>0.31</td>
</tr>
<tr>
<td>rs719725/9p24/ Fruit consumption (high vs. low)</td>
<td>TPDS2L3/IL-33/UHRF2/GLDC</td>
<td>1.10</td>
<td>1.02–1.19</td>
<td>( 9.59 \times 10^{-3} )</td>
<td>0.86</td>
<td>0.90</td>
</tr>
</tbody>
</table>

\(^a\)Interaction OR for SNP (log-additive for number of risk alleles) \( \times \) 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).

underlying linkage patterns given the ethnic differences in populations studied (the discovery study by Jia and colleagues 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, although, 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 9,160 cases and 9,280 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 \( 1.74 \times 10^{-4} \) \( [\text{Bonferroni } P \text{ value of } 1.74 \times 10^{-4} = (0.05/288)] \). However, as analyses of PMH use were limited to women (4,284 cases, 4,695 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 \( G \times E \) interaction for the first 10 identified susceptibility loci (19). In that paper, we observed only one statistically significant \( G \times E \) interaction, between rs16892766 (8q23.3) and vegetable consumption. Taken together, there is very little evidence for \( G \times E \) interaction involving known susceptibility loci within GECCO, although a larger sample size may be needed to evaluate interaction. Our data suggest that \( G \times E \) with known susceptibility loci may not account for the missing heritability. It is possible, although, that perhaps more complicated multifactorial interactions account for this missing heritability. It is also possible that \( G \times E \) interaction may be present for SNPs not identified to be associated with colorectal cancer risk through genome-wide screens of marginal SNP associations; such \( G \times E \) interaction might only become apparent when using a genome-wide \( G \times E \) approach. Currently, our consortium is investigating the presence of genome-wide \( G \times E \) interaction with a variety of environmental factors. It may also be informative to evaluate \( G \times E \) 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 \( G \times E \) involving these SNPs and prior studies have cited the need for a larger sample size when evaluating \( G \times E \) interaction (20, 23). Furthermore, we used an EB 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 $G \times E$ 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 (Supplementary Table S3), 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 to 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 colorectal cancer were in the expected directions. Indeed, it is notable that these [environmental variables and colorectal cancer] were in

interaction involving rare variants and multifactorial interaction.

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

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Authors’ Contributions
Development of methodology: J. Gong, S. Jiao, P.A. Newcomb, C.M. Ulrich


Study supervision: P.A. Newcomb, M. Thortonquist, U. Peters, E. White

Acknowledgments
DACHS: The authors thank all the participants and cooperating clinicians, and Ute Handte-Daub, Renate Hettler-Jensen, Utz Benschmidt, Muhabett Celik, and Ursula Eibler for excellent technical assistance.

GECCO: The authors 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: The authors thank Patrice Snow and Hardeep Ranu of the Dana Farber Harvard Cancer Center High-Throughput Polymorphism Core who assisted in the genotyping for NIH, HPFS, and PHS under the supervision of Dr. Immaculata Devivo and Dr. David Hunter, Qin (Carolyn) Guo, and Luxue Zhu who assisted in programming for NHS and HPFS, and Haiyan Zhang who assisted in programming for the PHS. The authors 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 in collecting the baseline data required for their reports: 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, and 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, the authors thank 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; refs. 47 and 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://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/summary?study=phs000206.v3.p2, respectively), and the lung datasets were accessed from the dbGaP website (http://www.ncbi.nlm.nih.gov/gap) through accession number phs000993.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 thank the study participants and staff of the Hormones and Colon Cancer study.

WH: 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%20to%20Write%20a%20Paper/WHI%20Investigator%20Short%20List.pdf.
Grant Support

C.S. Carlson, M. Du, J. Gong, T.A. Harrison, L. Hu, C.M. Hutter, S. Jiao, J. Minnier, B.M. Pflegeisen, U. Peters, D.L. Stelling, M. Thornquist, G.S. Warnick, and C.M. Ulrich are affiliated with GECCO, which is supported by the NIH, through funding allocated to CA63464. M. Cotterchio, M. Lemire, and B.W. Zanke are affiliated with the Ontario Registry for Studies of Familial Colorectal Cancer (U01 OFCCR, which is supported by the NIH, through funding allocated to CA48998 to M.L. Slattery). A. T. Chan, C. S. Fuchs, E. L. Giovannucci, and J. Ma are affiliated with HPFS, NIH, and PHS. HPFS was supported by the NIH (R01 CA51095, U01 CA167532, R01 CA173718, and P50 CA 127003). NHS by the NIH (R01 CA31718, P01 CA 087969, and P50 CA 127003) and PHS by the NIH (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 Cancer Institute, NIH: R01 CA 097735, Ontario Registry for Studies of Familial Colorectal Cancer Family Registry (U01 CA097735), Ontario Registry for Studies of the Colon CFR centers contributed data to this manuscript and were supported by NIH: Australasian Colorectal Cancer Family Registry (U01 CA097735), Ontario Registry for Studies of Familial Colorectal Cancer (U01 CA074793), and Seattle Colorectal Cancer Family Registry (U01 CA074794). H. Brenner, J. Chang-Claude, M. Hofrichter, 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. Jenkins, and P.A. Newcomb are affiliated with CCFR, which is supported by the NIH (R25CA94880). G. Casey, J.L. Hopper, M.A. 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 2007;131:341–57.

References


Gene–Environment Interaction Involving Recently Identified Colorectal Cancer Susceptibility Loci

Elizabeth D. Kantor, Carolyn M. Hutter, Jessica Minnier, et al.

Cancer Epidemiol Biomarkers Prev  Published OnlineFirst July 3, 2014.

Updated version  Access the most recent version of this article at:
doi:10.1158/1055-9965.EPI-14-0062

Supplementary Material  Access the most recent supplemental material at:
http://cebp.aacrjournals.org/content/suppl/2014/07/14/1055-9965.EPI-14-0062.DC1

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.