Genotype–Environment Interactions in Microsatellite Stable/Microsatellite Instability-Low Colorectal Cancer: Results from a Genome-Wide Association Study

Jane C. Figueiredo¹, Juan Pablo Lewinger¹, Chi Song¹, Peter T. Campbell³, David V. Conti¹, Christopher K. Edlund², Dave J. Duggan⁴, Jagadish Rangrej⁵, Mathieu Lemire⁵, Thomas Hudson⁵, Hutchinson Cancer Research Center, Seattle, Washington; ¹Departments of Medicine, University of Southern California, Los Angeles, California; ²USC Epigenome Center, University of Southern California, Los Angeles, California; ³Department of Epidemiology, American Cancer Society, Atlanta, Georgia; ⁴Translational Genomics Research Institute, Phoenix, Arizona; ⁵Ontario Institute for Cancer Research, The MaRS Center; ⁶Cancer Care Ontario; ⁷Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto; ⁸University of Ottawa Faculty of Medicine, Ottawa, Ontario, Canada; ⁹School of Population Health, University of Melbourne, Victoria, Australia; ¹⁰Fred Hutchinson Cancer Research Center, Seattle, Washington; ¹¹Departments of Medicine and Community and Family Medicine, Dartmouth Medical School, Lebanon, New Hampshire; and ¹²Cancer Research Center of the MaRS Center, Toronto, Canada

Abstract

Background: Genome-wide association studies (GWAS) have led to the identification of a number of common susceptibility loci for colorectal cancer (CRC); however, none of these GWAS have considered gene-environment (G × E) interactions. Therefore, it is unclear whether current hits are modified by environmental exposures or whether there are additional hits whose effects are dependent on environmental exposures.

Methods: We conducted a systematic search for G × E interactions using genome-wide data from the Colon Cancer Family Registry that included 1,191 cases of microsatellite stable (MSS) or microsatellite instability–low (MSI-L) CRC and 999 controls genotyped using either the Illuma HumanM1 or HumanM1-Duo BeadChip. We tested for interactions between genotypes and 14 environmental factors using 3 methods: a traditional case–control test, a case-only test, and the recently proposed 2-step method by Murcray and colleagues. All potentially significant findings were replicated in the ARCTIC Study.

Results: No G × E interactions were identified that reached genome-wide significance by any of the 3 methods. When analyzing previously reported susceptibility loci, 7 significant G × E interactions were found at a 5% significance level. We investigated these 7 interactions in an independent sample and none of the interactions were replicated.

Conclusions: Identifying G × E interactions will present challenges in a GWAS setting. Our power calculations illustrate the need for larger sample sizes; however, as CRC is a heterogeneous disease, a tradeoff between increasing sample size and heterogeneity needs to be considered.

Impact: The results from this first genome-wide analysis of G × E in CRC identify several challenges, which may be addressed by large consortium efforts. Cancer Epidemiol Biomarkers Prev; 20(5); 758–66. ©2011 AACR.

Introduction

Recently, several genome-wide association studies (GWAS) have led to the identification and replication of a number of susceptibility loci for CRC (1–6). Incorporating environmental exposures into GWAS data may aid in the identification of additional susceptibility alleles that would be otherwise masked by heterogeneity in subgroups, and would also clarify whether certain environmental exposures may modulate risk in susceptible individuals. However, there are limited data on the interaction between other susceptibility alleles and environmental risk factors for CRC. To date, no studies have examined the interaction between a wide range of environmental factors and genome-wide genotype data with respect to cancer risk.

Detecting gene–environment (G × E) interactions using a standard case–control test is challenging in a genome-wide context because of the stringent significance level required to adjust for multiple testing and because only weak G × E interactions are expected. The case-only test is known to be more powerful than the case–control test but in the presence of population level G × E association it can yield a severely inflated type I error (7). Recently, new methods to test for G × E interactions in GWAS have
been proposed. Murcray and colleagues introduced an efficient 2-step approach that is carried out independently of any initial scans for main effects (8). The method expands on the traditional test for G × E interaction in a case–control study by incorporating a preliminary screening step constructed to efficiently use all available information. This method has been shown to be more powerful for a wide range of environmental exposures, minor allele frequencies, and genetic effects compared to the traditional 1-step test (8). In this study, we take advantage of these methodologies to systematically search for G × E interactions within a GWAS of MSS (microsatellite stable)/MSI-L (microsatellite instability–low) CRC from the Colon Cancer Family Registry considering lifestyle and environmental exposures known to be involved in the etiology of CRC.

Methods

Study design and sample

Participants included in this analysis were recruited from 3 population-based registries based at the Fred Hutchinson Cancer Research Center (FH CRC, Seattle, WA), Cancer Care Ontario (Ontario Familial Colorectal Cancer Family Registry (OFCCR), Toronto, Canada), and the University of Melbourne (Victoria, Australia), which recruited families from both Australia and New Zealand as part of the Colon Cancer Family Registry (Colon CFR; ref. 9).

Cases from these registries met the following eligibility criteria: invasive CRC; self-identified as non-Hispanic White; no identified germline mutations in mismatch repair (MMR) genes; MSS or MSI-L CRC, and/or MMR protein immunohistochemistry positive determined using standard methods (10). All cases meeting these criteria and under age 50 or who had an affected first-degree relative with CRC were included, together with a 20% random sample of those over age 50 with no affected first-degree relative.

Population-based controls were randomly sampled from these same catchment areas as the 3 registries, frequency matched on age, as described recently (9). All controls were self-identified as non-Hispanic White and reported no personal or family history of CRC.

Written, informed consent was obtained from all participants. The study was approved by the Institutional Review Board at each of the institutions.

Data collection and variable definitions

All participants completed mailed questionnaires (Cancer Care Ontario) or a telephone-based or face-to-face interview (FH CRC, University of Melbourne) at study enrollment information. Questions focused on exposures 2-years before the date of diagnosis for cases and 2-years before the date of recruitment for controls. Data were collected on personal and family histories of colorectal (cancer and polyps) and other cancers and colon polyps, and lifestyle risk factors, including, med-

Ever-use (yes, no) of selected supplements (multivitamins, folic acid, and calcium) and medications (non-steroidal anti-inflammatory drugs, NSAIDs) were defined as use at least 2 times per week for more than a month during a participant’s lifetime. NSAIDs included ibuprofen and aspirin. Because folic acid is contained in nearly all multivitamins, the derived variable for folic acid included use of folic acid supplements and multivitamins. Alcohol use was defined as the consumption of any alcoholic beverage (beer, wine, hard cider, sake, liquor, spirits, mixed drinks, or cocktails) at least once a week for 6 months or longer during the most recent decade of life at enrollment. Being an ever smoker was defined as ever smoking at least 1 cigarette per day for 3 months or longer. Pack-years of smoking was calculated based on the number of cigarettes smoked per day and the number of years smoked. A person was considered to be physically active if they reported more than 20 metabolic equivalent (MET) hours per week of physical activity during the most recent decade of life at enrollment. The number of servings per week of fruits, vegetables, and red meat were also calculated. Body mass index (BMI) was calculated as the person’s weight (kg) 2 years prior to study recruitment divided by adult height (m) squared.

Genotyping

All participants provided a blood sample at the time of recruitment. DNA samples were genotyped with the Illumina Human1M (n individuals = 1,973; m = 1,072,820 SNPs) or Human1M-Duo (n individuals = 374; m = 1,199,187 SNPs) BeadChip platforms. Samples with GenCall scores less than 0.15 at any locus were considered “no calls.” Each 96-well plate included 1 inter-plate positive quality control sample (NA06990—Coriell Cell Repositories). In addition, 27 blinded and 22 unblinded quality control replicates from the study sample were genotyped. SNP data obtained from both the Coriell and study sample replicates showed a very high concordance rate of called genotypes: 99.95% and more than 99.94%, respectively (for samples with call rates >90%). The Human1M and Human1M-Duo contain 415 and 436 SNPs, respectively, that were genotyped as part of a candidate gene study on the Illumina GoldenGate platform on a subset of the individuals genotyped in this study (N = 444). A high concordance rate (>98%) was observed for more than 99% of the samples with a call rate more than 90%.

Individuals were excluded with (Fig. 1)

1. missing phenotype data (n = 2);
2. self-identified as Caucasian (n = 29);
3. poor concordance (>98%) with genotypes on selected candidate genes genotyped on GoldenGate platform (n = 3);
(4) chromosome X or Y anomalies (n = 3 gender misclassifications, n = 1 low male chromosome Y intensity); (5) a call rate less than 95% (n = 75); (6) any stripe call rate less than 80% (n = 9); (7) high identity-by-descent with another subject (n = 2); or (8) non-European admixture as estimated by STRUCTURE (11; n = 33).

SNPs were excluded from analysis if

(1) they were not on both the Human1M and Human1M-Duo (m = 190,301); (2) they were annotated as “Intensity Only” on either the Human1M or Human1M-Duo (m = 8,263); (3) had a call rate less than 90% on the Human1M or Human1M-Duo (m = 7,836 and 1,393, respectively); (4) had a call rate less than 90% in any study center/case status group (m = 12,695); (5) they departed from Hardy–Weinberg equilibrium (P \(\leq 1 \times 10^{-4}\)) in controls (m = 2,601); or (6) had a minor allele frequency (MAF) less than 0.05 (m = 238,198).

A total of 2,190 individuals and 770,098 SNPs were used in the final analysis.

All SNPs with borderline significant associations underwent additional quality control checks including:

(1) visual inspection of genotype clusters;
(2) verification of genotype concordance between HapMap samples genotyped by Illumina on the Human1M and Human1M-Duo and with HapMap phase II release 24;
(3) verification of no major MAF difference in controls genotyped on the Human1M and Human1M-Duo.

**Replication sample**

We replicated significant interactions using data from the ARCTIC Study. Details of the ARCTIC Study are provided elsewhere (5). The selected SNPs were extracted from a larger set of SNPs genotyped in 2,433 unique samples on a custom 10,640-bead iSelect array from Illumina. All eligible cases of CRC were included irrespective of MSI status. After excluding samples from the Colon CFR that were included in this study and samples not self-identified as white, we were left with a total of 872 cases and 810 controls. The selected SNPs were extracted from a larger set of SNPs genotyped in 2,433 unique samples on a custom 10,640-bead iSelect array from Illumina designed for 7,703 SNPs. The call rate for this panel was 99.96% after excluding 3 failed DNAs (call rate < 69%) and 332 failed SNPs (4.3%). There were no discordant genotypes in 23 pairs of duplicates. Data collection on environmental risk factors and variable definitions were carried out in the same manner as described above. All subjects provided written informed consent. This study was approved by the ethics review boards of the Toronto Academic Health Sciences Council.

**Statistical analysis**

We considered 3 approaches for genome-wide G \(\times\) E testing. In the first 2 approaches, we exhaustively tested every SNP in the GWAS panel for G \(\times\) E interaction with each of the environmental exposures using either a case-control test or a case-only test. The case–control test is based on the logistic regression model:

\[
\log \left( \frac{Pr(Y = 1|G, E, Z)}{Pr(Y = 0|G, E, Z)} \right) = \alpha + \beta_x E + \beta_y G + \beta_{xy} G \times E + \eta Z
\]  

(A)

where Y indicates disease status, with Y = 1 for cases and Y = 0 for controls, E is an environmental exposure, G is the genotype at a particular SNP, and Z represents any additional covariates to be adjusted for such as sex, age (continuous) and center.
For a binary exposure, the case-only test is based on logistic regression model:

$$\log \left( \frac{Pr(E = 1|G, Z)}{Pr(E = 0|G, Z)} \right) = \alpha + \gamma G + \eta Z$$

(B)

For a quantitative exposure, the case-only test is based on logistic regression model:

$$E[E|G] = \alpha + \gamma G + \eta Z$$

(C)

We used an additive coding for the genotypes for both the case–control and the case-only tests, i.e., $G$ indicates the number of copies of the reference allele ($G = 0, 1, 2$). The additive model is known to have good power for a very wide range of true modes of action (recessive, dominant, multiplicative; ref. 12). For the case–control test, the hypothesis of no SNP–environment (SNP × E) interaction corresponds to the $H_0: \beta_{GE} = 0$ in model (A). For the case-only test the hypothesis of no SNP × E interaction corresponds to $H_0: \beta = 0$ in models (B) or (C) depending on whether the exposure is binary or quantitative. We tested every SNP that passed quality control and was polymorphic in our sample using the case–control and case-only tests. We refer to these scans for SNP × E interactions as exhaustive case–control or exhaustive case-only respectively, in contrast with the 2-step scan described below. To control for multiple testing we used a simple Bonferroni correction for the number of SNPs that were actually tested ($0.05/770,098 \approx 6.5 \times 10^{-5}$). Because each exposure was considered an independent a priori hypothesis we corrected across SNPs for each exposure, but not across exposures. For continuous exposures, we also report the stratified estimates of risk by dichotomizing at the median unless otherwise reported.

The third method for genome-wide G × E testing we considered was the approach of Murcray and colleagues (8). This 2-step method consists of a screening first step followed by a formal test of interaction. Specifically, in the first step a test of association between the exposure $E$ and SNP $G$ is carried out on the combined sample of cases and controls based on the logistic regression model:

$$\log \left( \frac{Pr(E = 1|G, Z)}{Pr(E = 0|G, Z)} \right) = \alpha + \gamma G + \eta Z$$

(B)

for binary exposures, or the linear regression model:

$$E[E|G] = \alpha + \gamma G + \eta Z$$

(C)

for continuous exposures.

The hypothesis $H_0$: $\gamma = 0$ is tested for each SNP at significance level $\alpha_1 = 0.001$ using a $\chi^2$ 1df Wald test. The $m$ SNPs achieving $\alpha_1$ significance (i.e., with $P$-value $< \alpha_1$) pass the screening step and are tested for G × E interaction using model (1). To preserve a genome-wide type I error of $\alpha = 0.05$ of the overall 2-step procedure, Murcray and colleagues (8) showed that it suffices to correct in step 2 by the $m$ SNPs that pass the screening, i.e., by testing at significance level $\alpha/m$. For details of the rationale behind the screening step and the validity of the 2-step approach see Murcray and colleagues (8). All the genome-wide G × E analyses were carried out with the software PLINK (13).

In addition to the genome-wide G × E analyses using the exhaustive case–control, exhaustive case-only, and 2-step methods described above, we carried out focused testing of previously reported and replicated genetic variants associated with CRC from 5 published GWAS (1–5) and a meta-analysis (6) for G × E interaction with each of the exposures of interest: 8q23.3 (rs16892766, EIF3H; ref. 2); 8q24 (rs6983267, rs7014346; refs. 1, 4, 5, 14–16), 10p14 (rs10795668 ; ref. 2), 11q23 (rs3802824; ref. 1), 14q22.2 (rs4444235; BMP4; ref. 6); 15q13 (rs4779584; ref. 17); 16q22.1 (rs9929218, CDH1(6); and 18q21 (rs4939827, SMAD7; refs. 1, 3), 19q13.1 (rs10411210, RHPN2; ref. 6), and 20p12.3 (rs961253; ref. 6). 8q24 rs1050477 and 9p24-rs719725 (5, 14) were not available on the Illumina Human1M or Human1M-Duo. We considered 9p24-rs7025295 and 9p24-rs7857628 as surrogates for the missing 9p24-rs719725 ($r^2 = 0.965, r^2 = 0.966$ using HapMap2_r2 CEU, respectively). We tested each variant using the case–control based on model (A) and case-only test based on model (B) at 5% significance level. In addition to testing individual SNPs, we tested a score that combines information from all the 13 SNPs into a single variable for interaction with the exposures of interest. For each subject, the score was constructed by counting the number of CRC risk-increasing variants across the 13 SNPs (i.e., the score ranges from 0 to 26). We tested the interaction of the score and the exposures using the standard logistic regression model (A), with $G$ representing now the quantitative score.

Only interactions that were identified to be significant from this GWA study were tested in the ARCTIC Study using the 1-step test on model (A) at 5% significance level. All models were adjusted for age, center, and sex unless otherwise specified.

Results

This study included 1,191 population-based cases of MSS/MSI-L CRC and 999 unrelated population-based controls. Table 1 shows the distribution of selected characteristics for the study population. After adjustment for age, sex, and study center, we found BMI, smoking, and red meat intake were positively associated with risk of CRC. Ever use of folic acid and multivitamins were associated with an increased risk of disease in unadjusted models only. Alcohol use, NSAID use and calcium use were associated with statistically significant decreased risk of CRC. Among women, ever use of postmenopausal hormones or oral contraceptives were associated with statistically significant decreased disease risks. Servings of fruits and vegetables, physical activity and height were not associated with risk of CRC.

Using the exhaustive case–control and case-only tests, we observed no statistically significant interactions with
Table 1. Characteristics of the study sample

|                        | Cases  
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( n = 1,191 )</td>
</tr>
<tr>
<td>Mean age ± SD, y</td>
<td>52.6 ± 11.3</td>
</tr>
<tr>
<td>Sex, ( n ) (%)</td>
<td>Male 617 (51.6)</td>
</tr>
<tr>
<td></td>
<td>Female 574 (48.2)</td>
</tr>
<tr>
<td>Center, ( n ) (%)</td>
<td>Ontario, Canada 417 (35.0)</td>
</tr>
<tr>
<td></td>
<td>Melbourne, Australia 330 (27.7)</td>
</tr>
<tr>
<td></td>
<td>Seattle 444 (37.3)</td>
</tr>
<tr>
<td>Mean height ± SD, m</td>
<td>1.71 ± 0.10</td>
</tr>
<tr>
<td>Unknown/missing</td>
<td>1</td>
</tr>
<tr>
<td>BMI (kg/m²), ( n )  ( % )</td>
<td>18.5–25.0 (normal) 379 (32.8)</td>
</tr>
<tr>
<td></td>
<td>25.1–30 (overweight) 490 (42.4)</td>
</tr>
<tr>
<td></td>
<td>30+ (obese) 287 (24.8)</td>
</tr>
<tr>
<td>Mean ± SD, kg/m²</td>
<td>27.40 ± 5.43</td>
</tr>
<tr>
<td>Alcohol use, ( n )  ( % )</td>
<td>&lt;1 per wk 428 (36.1)</td>
</tr>
<tr>
<td></td>
<td>≥1 per wk 757 (63.9)</td>
</tr>
<tr>
<td>Smoking, ( n ) (%)</td>
<td>Never 508 (43.2)</td>
</tr>
<tr>
<td></td>
<td>Ever 668 (56.8)</td>
</tr>
<tr>
<td>Mean pack-years ± SD</td>
<td>13.14 ± 20.36</td>
</tr>
<tr>
<td>Unknown/missing</td>
<td>15</td>
</tr>
<tr>
<td>Folic acid, ( b ) ( n ) (%)</td>
<td>Never 558 (47.0)</td>
</tr>
<tr>
<td></td>
<td>Ever 630 (53.0)</td>
</tr>
<tr>
<td>Multivitamins, ( n ) (%)</td>
<td>Never 765 (64.3)</td>
</tr>
<tr>
<td></td>
<td>Ever 425 (35.7)</td>
</tr>
<tr>
<td>Calcium, ( n ) (%)</td>
<td>Never 608 (51.4)</td>
</tr>
<tr>
<td></td>
<td>Ever 576 (48.6)</td>
</tr>
<tr>
<td>Mean number of servings per wk of red meat ± SD</td>
<td>4.75 ± 4.34</td>
</tr>
<tr>
<td>Unknown/missing</td>
<td>41</td>
</tr>
</tbody>
</table>

\( ^a \) Adjusted for age, sex, and center.

\( ^b \) Includes multivitamin users.

\( ^c \) Females only.
any SNP at a genome-wide significance level of $6.5 \times 10^{-8}$ with any environmental exposure. The lowest interaction $P$ values were between: oral contraceptive use and rs17329226 ($P = 7.0E^{-07}$); and ever smoker and rs2486540 ($P = 3.1E^{-07}$), rs2486538 ($P = 3.7E^{-07}$), and rs538835 ($P = 5.3E^{-07}$).

Using the 2-step method, between 662 and 1,004 SNPs (depending on the exposure variable) passed the significance level in the screening step and were carried on to the second step. Therefore, the appropriate number of corrections for multiple testing varied by exposure, dependent on the total number of SNPs in the second step, from $5.0 \times 10^{-5}$ to $7.6 \times 10^{-3}$. We identified no significant G × E interactions with $P$ value less than $10^{-4}$.

Table 2 lists the known hits for CRC indentified through published GWA studies (1–6). We tested whether any of these SNPs showed a significant interaction with the selected environmental exposures. We identified the following interactions (using a case–control test): rs3802842 and postmenopausal hormones ($P = 0.01$); rs10795668 and oral contraceptive-use ($P = 0.04$), rs961253 and oral contraceptive-use ($P = 0.01$); rs9929218 and height ($P = 0.02$); rs9929218 and alcohol-use ($P = 0.04$); rs4939827 and servings of vegetable intake ($P = 0.01$); and rs9929218 and calcium-use ($P = 0.045$). In our replication sample, none of the interactions with the individual SNPs were significant at the 5% level. When we tested the interaction of the environmental covariates and the score that combines the previously reported GWAS hits, we only found 2 marginal significant interactions with red meat consumption ($P = 0.01$) and calcium ($P = 0.05$). We did not attempt replication of these interactions.

Discussion

In this GWAS of early-onset MSS/MSI-L CRC, we identified no selected personal or lifestyle characteristic that significantly modified the effect of genetic variants on the risk of CRC at a strict genome-wide level of less than $6.5 \times 10^{-8}$ using an exhaustive case–control or case–only test or the appropriate significance levels for 2-step methods (20, 21). In our analysis, we identified a few significant interactions at the less than 5% level, but none of these were significant in an independent case–control and case–only analysis (20, 21), but there has been no formal comparison of these methods.

Achieving sufficient statistical power is challenging in a genome-wide context, even with these recently described methodologies. Our power calculations highlight this point, especially where the expected gene, exposure and interaction effects are modest. Figure 2 shows the sample size required to attain 80% power with the 2-step approach for various combinations of minor allele frequencies, exposure prevalences, and interaction odds ratios. In this context, it was assumed that there were no SNP main effects, corresponding to the scenario where a G × E scan could detect a SNP that a standard GWAS based on SNP main effects would not. We found that a using data from a typical GWAS of 1,000 cases and 1,000 controls would detect interaction odds ratios of 2 or higher, with highly prevalent exposures and allele frequencies. There are likely to be many G × E interactions, but our study is underpowered to detect them. International consortia gathering GWAS data in CRC may aid in this effort if environmental covariates are available and there is potential for harmonization of variable definitions. However, even this increased sample size will not suffice to detect interaction odds ratios below 1.4, especially for less frequent exposures and lower allele frequencies.

We also investigated whether any of the recently reported and robustly replicated susceptibility loci identified through GWA studies of CRC were modulated by selected environmental factors. We considered only replicated susceptibility variants from published GWA studies of independent CRC cases and unaffected controls (1–6). We identified a few significant interactions at the less than 5% level, but none of these were significant in an independent case–control study of CRC that had collected epidemiologic data using the same questionnaires from individuals in one of the same geographic regions. One potential reason for our failure to replicate could be that we were unable to restrict our replication sample to only cases with early-onset MSS or MSI-L cancers. Common environmental exposures, such as alcohol intake, cigarette smoking, and obesity, have been reported to differ by MSI strata (22, 23). Furthermore, for 4 known susceptibility alleles we found no association with CRC in the Colon CFR and in the absence of a main effect the prospects of identifying a G × E interaction may be lower.

There are some limitations to this study. The main concern is the limited statistical power to investigate G × E interactions.
Table 2. Significant interactions at 5% level between known CRC-associated variants and environmental factors

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr</th>
<th>Gene</th>
<th>MAFa</th>
<th>Main SNP effect OR (95% CI)</th>
<th>Environment factor</th>
<th>Category</th>
<th>Stratified OR (95% CI)</th>
<th>P-intb</th>
<th>Replication sample: ARCTIC Study</th>
<th>Stratified OR (95% CI)</th>
<th>P-intb</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs16892766</td>
<td>8q23.3</td>
<td>EIF3H (flanking 3’ end)</td>
<td>0.08</td>
<td>1.28 (1.01–1.62)</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs6983267</td>
<td>8q24.21</td>
<td>POU5F1B (flanking 5’ end)</td>
<td>0.48</td>
<td>0.81 (0.70–0.93)</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7014346</td>
<td>8q24.21</td>
<td>POU5F1B (flanking 5’ end)</td>
<td>0.38</td>
<td>1.17 (1.02–1.34)</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7025295</td>
<td>9p24.1</td>
<td>UHRF2 (flanking 5’ end)</td>
<td>0.40</td>
<td>0.83 (0.73–0.96)</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7857628</td>
<td>9p24.1</td>
<td>UHRF2 (flanking 5’ end)</td>
<td>0.41</td>
<td>0.82 (0.71–0.94)</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs10795688</td>
<td>10p14</td>
<td>Gene desert</td>
<td>0.33</td>
<td>0.84 (0.73–0.98)</td>
<td>OC use</td>
<td>No</td>
<td>0.97 (0.70–1.34)</td>
<td>0.04</td>
<td>0.97 (0.71–1.32)</td>
<td>0.30</td>
<td></td>
</tr>
<tr>
<td>rs3802842</td>
<td>11q23.1</td>
<td>C11orf93 (intron)</td>
<td>0.28</td>
<td>1.22 (1.06–1.41)</td>
<td>PMH use</td>
<td>No</td>
<td>1.48 (1.11–1.98)</td>
<td>0.01</td>
<td>1.08 (0.80–1.47)</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td>rs4444235</td>
<td>14q22.2</td>
<td>BMP4 (flanking 3’ end)</td>
<td>0.48</td>
<td>1.11 (0.97–1.26)</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4779584</td>
<td>15q13.3</td>
<td>SCG5 (flanking 3’ end)</td>
<td>0.19</td>
<td>1.22 (1.03–1.44)</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs9929218</td>
<td>16q22.1</td>
<td>CHD1 (intron)</td>
<td>0.29</td>
<td>0.93 (0.80–1.08)</td>
<td>Height &lt;1.7 m</td>
<td>1.13 (0.91–1.41)</td>
<td>0.02</td>
<td>0.79 (0.62–0.99)</td>
<td>0.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: PMH, postmenopausal hormones; OC, oral contraceptives; –, no G × E interaction detected at significance level less than 5% for the selected environmental factors.

aMAF determined in unrelated controls.

bP-value for interaction between G and E.


Furthermore, because of the study design, we were unable to investigate the potential effects of ethnicity, family history of CRC, or other phenotypes of CRC (i.e., MSI high). Lastly, there is no consensus about the correct statistical method to model G × E interactions and more research is required.

In summary, we identified no genome-wide significant G × E interactions in this GWAS of early-onset MSS/MSI-L CRC. Much of the evidence from descriptive epidemiology, migrant studies, and changes in CRC rates in countries undergoing rapid economic development (most obviously Japan in the second half of the twentieth century; Japan now has the highest rates of CRC in the world) points to environmental risk factors as the major determinants of the international variation in CRC. It is crucial therefore that we gain a better understanding of susceptibility to these environmental factors. This, in turn, underscores the need to detect G × E interactions, which will require large collaborations of GWA studies with adequate data collection on exposures.

Disclosure of Potential Conflicts of Interest

The content of this article 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.

Grant Support

This work was supported by the National Cancer Institute, NIH, under RFA no. CA-55-011 and through cooperative agreements with the Australasian Colorectal Cancer Family Registry (U01 CA097733), the Ontario Registry for Studies of Familial Colorectal Cancer (U01 CA074783), and the Seattle Colorectal Cancer Family Registry (U01 CA074794), as well as NIH/NCI U01CA128390 (G. Casey) and the Canadian Cancer Society Research Institute, the Ontario Institute for Cancer Research, and the Ontario Ministry of Research and Innovation.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked ‘advertisement’ in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received June 24, 2010; revised October 25, 2010; accepted January 26, 2011; published OnlineFirst February 25, 2011.
of the genetic epidemiology of colon cancer. Cancer Epidemiol
Lower cancer incidence in Amsterdam-I criteria families without mismatch repair
Genotype–Environment Interactions in Microsatellite Stable/Microsatellite Instability-Low Colorectal Cancer: Results from a Genome-Wide Association Study

Jane C. Figueiredo, Juan Pablo Lewinger, Chi Song, et al.

Cancer Epidemiol Biomarkers Prev 2011;20:758-766. Published OnlineFirst February 25, 2011.

Updated version
Access the most recent version of this article at:
doi:10.1158/1055-9965.EPI-10-0675

Cited articles
This article cites 22 articles, 2 of which you can access for free at:
http://cebp.aacrjournals.org/content/20/5/758.full#ref-list-1

Citing articles
This article has been cited by 7 HighWire-hosted articles. Access the articles at:
http://cebp.aacrjournals.org/content/20/5/758.full#related-urls

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.