Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Alexander Nicholas Maillis	4/24/2019
_____	_____
[Student's name typed]	Date

Association of Colorectal Cancer Susceptibility Loci Alone and in Combination with Adherence to the WCRF/AICF Guidelines for Cancer Prevention on Colorectal Cancer Risk

By

Alexander N. Maillis
Master of Public Health

Department of Epidemiology

_____

Dr. Veronika Fedirko, PhD
Committee Chair

Association of Colorectal Cancer Susceptibility Loci Alone and in Combination with Adherence to the WCRF/AICF Guidelines for Cancer Prevention on Colorectal Cancer Risk

By

Alexander N. Maillis
B.S. Biomedical Sciences
University of Central Florida
2016

Thesis Committee Chair: Veronika Fedirko, PhD

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology
2019

Abstract

Association of Colorectal Cancer Susceptibility Loci Alone and in Combination with Adherence to the WCRF/AICF Guidelines for Cancer Prevention on Colorectal Cancer Risk

By Alexander N. Maillis

Background: Colorectal cancer (CRC) is the 3rd most common cancer in men and women combined worldwide. Previous genome-wide association studies (GWAS) identified several common genetic variants associated with CRC risk. Additionally, observational studies demonstrated that healthy lifestyle choices are strongly association with colorectal neoplasms; however, little is known about joint effects of genetics and lifestyle on CRC risk.
Methods: A nested case control study including 1,462 CRC cases and 1,482 controls was conducted within the European Prospective Investigation into Cancer and Nutrition (EPIC) study, a prospective cohort of more than 520,000 participants from 10 Western European countries. Odds ratios and 95% confidence intervals for the risk of CRC by categories of genetic risk score (GRS) alone and in combination with the WCRF/AICR score were estimated from unconditional logistic regression models, with adjustment for age at recruitment, sex, and study center.
Results: Study participants with ≥18 CRC susceptibility loci had a statistically significantly higher risk for CRC (OR = 1.36; 95%: 1.09, 1.69) when compared to those with <14 loci. A high adherence to the WCRF/AICR recommendations was statistically significantly associated with a lower risk for CRC (OR = 0.79; 95%: 0.63, 0.98) compared to study participants with the least concordance. Joint analysis of the GRS and WCRF/AICR recommendation adherence suggested that better concordance with the WCRF/AICR recommendations is associated with lower CRC risk, regardless of genetic risk, so that individuals with low adherence to the WCRF/AICR recommendations and high genetic risk were at a statistically significantly higher risk of CRC compared to those with low genetic risk and high adherence to the WCRF/AICR recommendations (OR = 1.63; 95%: 1.12, 2.38).
Conclusion: Our study suggests that adherence to the WCRF/AICR recommendations is associated with lower CRC risk regardless of the genetic risk.

Association of Colorectal Cancer Susceptibility Loci Alone and in Combination with Adherence to the WCRF/AICF Guidelines for Cancer Prevention on Colorectal Cancer Risk

By

Alexander N. Maillis

B.S., Biomedical Science
University of Central Florida
2016

Thesis Committee Chair: Veronika Fedirko, PhD

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology
2019

Acknowledgements: I would like to acknowledge my thesis advisor, Dr. Veronika Fedirko, for her guidance. I would also like to acknowledge my family and Sarah McDaniel for their patience with me throughout the process. Finally, I would like to acknowledge Emory University, The Rollins School of Public Health, the researchers and study centers responsible for constructing the EPIC cohort, and the individuals within the study that have made the further understanding of cancer etiology possible.

Table of Contents

Background

Colorectal Cancer Incidence and Mortality

As of 2016, colorectal cancer ("CRC") is the third most common cancer diagnosis among men and second most common among women worldwide – making it responsible for approximately 10% of all cancer cases (1, 2). Additionally, CRC had the fourth highest mortality rate of all cancers with 14 deaths per 100,000 people reported in 2015. Incidence rates of CRC have declined over the past decades but, despite this, different trends have emerged (1, 2). Reports from the American Cancer Society show that, from 2006-2015, adults 55 years or older experienced a 3.7% decrease in CRC incidence while those less than 55 years old experienced a 1.8% increase. Mortality rates follow this trend, with those 55 years or older seeing a 2.7% reduction in mortality rates while those less than 55 years old had a 1% increase (2, 3, 4). Current scientific literature suggests that changing exposure to risk factors as well as improvements in, and wider use of, screening techniques such as colonoscopy may explain these trends (1).

CRC is the result of abnormal cell growth within the lumen of the gut. For most, this growth starts as an accumulation of cells on the inner lining of the colon called a polyp. This growth is noncancerous initially and can grow on the lining of the colon for up to 10-20 years (2). The most common type of polyp within the colon is an adenomatous polyp, or adenoma. One-third to one-half of all Americans will have an adenomatous polyp present at some point but only an estimated 10% of these growths progress to cancer (2). Cancer occurring from the formation of a cancerous adenoma is referred to as an adenocarcinoma; these account for 96% of all CRC cases. The remaining 4% of cases are either mucinous carcinomas or adenosquamous carcinomas (2). Once cancerous, the

tumor has the ability to grow into the lining of the colon where the cancerous cells can spread to the nearby lymph nodes, blood vessels or even surrounding organs of the body. Once the cancer cells spread past the lining of the colon into surrounding structures it is considered metastatic disease (1, 2).

Colorectal Cancer Risk Factors

There are multiple established and suspected risk factors associated with CRC development. Current scientific literature suggests that CRC does have some hereditary components that we discuss below; however, the majority of cases are sporadic and develop over a period of years – suggesting that our environment plays a substantial role in CRC (5, 6). The World Cancer Research Fund and American Institute for Cancer Research (WRCF/AICR) report that physical activity has been convincingly associated with decreased risk of CRC - as does partaking in a diet high in wholegrains, dietary fiber, dairy products and calcium (3). Dietary and behavioral choices such as high intake of processed/red meats, long-term smoking, increased alcohol intake and increased body fatness have convincingly been associated with increased risk of CRC (3, 5). Other factors that have shown a correlation for decreased risk, although limited in suggestion, are increased intake of foods containing Vitamin C, Fish, Vitamin D and multi-vitamin use while low vegetable intake, low fruit intake and high intake of iron-containing foods have shown a limited-suggestive negative correlation (3, 5). Other studies have looked at additional factors such as Vitamin A intake, low-fat diets, total energy intake, meal frequency and dietary pattern but no definitive results have been observed (3). At the population level, CRC risk tends to increase with increasing age with the median age of diagnosis being 68 in men and 72 in women in the colon, and 63 years for both men and

women within the rectum. Variation by sex is also noted, with men reported to have a 30% higher incidence rate and 40% higher mortality rate relative to women (1, 2, 3). Hypothesized mechanisms behind this observation, such as differences between sex hormones, have been inconclusive so far - leaving the disparity attributed to differences in risk factor exposure (2, 5). Finally, differences in CRC incidence and mortality can be observed within different racial/ethnic groups. Non-Hispanic blacks have the highest incidence and mortality rates within the United States while Asians/Pacific Islanders have the lowest (1, 2). Statistically, Non-Hispanic Blacks have a 20% increased incidence compared to Non-Hispanic Whites and 50% higher rates when compared to Asian/Pacific Islanders. The disparity between these ethnic groups is complex, but is theorized to be secondary to differences in socioeconomic status (1, 2). Additional observations show that particular ethnic groups (Alaskan Natives and American Indians) have an even higher incidence and mortality rate than Non-Hispanic Blacks; with reports from 2017 suggesting it to be almost 80% higher than Non-Hispanic Blacks (1). Reasons such as high-fat diet, low Vitamin D exposure, smoking/alcohol use prevalence and abnormal exposure to the bacterium Helicobacter Pylori have all been suggested as explanations for the increased burden of CRC in Alaska (1, 2).

Colorectal Cancer Screening

In addition to avoiding particularly high-risk lifestyle activities, regular screening in accordance with those of the American Cancer Association (and your physician) is recommended. Current recommendations by the American Cancer Society suggest that an individual should start receiving CRC screening at the age of 45 and continue until approximately 85 (1, 2). There are several screening techniques currently used - all of

which have different mechanisms of detecting cancerous polyps and require different periods between rescreening. A colonoscopy can examine your entire colon and is considered the most precise, but is also expensive and can only be performed by a medical doctor; these tests are performed every 10 years unless determined otherwise by your gastroenterologist (1, 2). Computed Tomographic Colonography is performed every 5 years and is less invasive than a colonoscopy but has only been effective in accurately diagnosing large polyps as cancerous, and necessitates a colonoscopy if a polyp is identified (1, 2). Double-contrast enema and flexible sigmoidoscopy are additional visual-based examinations, both of which are performed every 5 years (1, 2). Double-contrast enema's have become less common screening procedures since they also necessitate a colonoscopy if abnormal findings are present, while flexible sigmoidoscopies are less common due to their restriction of only visualizing the upper 1/3rd of the colon (1, 2). Recent medical advances have brought about the increased use of stool tests as a cheaper and less complex alternative for colorectal screening. Tests such as the Fecal immune-chemical test (FIT), High-sensitivity guaiac-based fecal occult blood test (gFOBT) and FIT-DNA test can all be performed at home on an annual basis – and all require a colonoscopy with abnormal findings (1 ,2). The FIT and gFOBT test can identify hemoglobin, or hidden blood, within the stool. The tests are not highly sensitive when it comes to small polyp identification, causing some providers to lean towards a combination stool test/flexible sigmoidoscopy (2). The FIT-DNA test combines hemoglobin identification with the ability to identify genetic mutations within cells shed in the stool. All stool tests currently in-market have shown to have high false-positive rates with the FIT-DNA test having the highest among the three (2). As

mentioned earlier, screening frequency varies by the individual but is typically a result of the patient's medical history. Some factors that may increase the frequency of screening include a family history of CRC, obesity, or presence of gastrointestinal diseases such as Crohn's or Familial Adenomatous Polyposis (2). Observational studies of colonoscopy efficacy suggest that colonoscopies have reduced CRC incidence by approximately 50% and CRC mortality by approximately 40% (2, 5). Similar observational studies on the flexible sigmoidoscopy estimate that they have resulted in a 30% incidence reduction and 20% mortality reduction (2, 5). The implementation and increasingly improving access to these screening tests have played a crucial role in reducing CRC incidence and mortality. From 2000-2015, the percentage of individuals 50+ years old regularly receiving colorectal screening increased from 34% to 63% (2). Initiatives currently in place include the "80% by 2018 initiative" organized by the National Colorectal Council Roundtable. This initiative aimed to raise CRC screening to 80% nationwide and, if met, predicted to prevent approximately 277,000 CRC cases and 203,000 CRC deaths by 2030 (2).

Genetic Risk Alleles of Colorectal Cancer

Single nucleotide polymorphisms ("SNPs") are variations of a base pair, or nucleotide, within our DNA sequence. These alterations represent the most common genetic variations amongst humans. On a macro level there is some degree of these polymorphisms within all of the population, however; some individuals possess more or less of these mispairings. To a certain degree, SNP variation can be used as a genetic marker to follow patterns of inheritance through generations (6). Effects of these SNPs vary and are not guaranteed to induce disease individually; however, current literature

suggests that a combination of SNP's may produce a phenotypically appreciable outcome. There is substantial literature to suggest that CRC does have some genetic components – which is further strengthened by prior Genome-Wide Association Studies ("GWAS") that have identified multiple SNPs associated with a moderate increase in CRC risk (7). It is purported that these SNPs could explain at least some fraction of CRC cases; approximately 17% according to one meta-analysis (7). To date, there is a large amount of variance between SNP selection in these studies, but several SNPs have consistently suggested a moderate association with CRC risk. One such SNP is rs3802842. This variant has been continuously associated with elevated CRC risk across multiple demographic populations: including China, Denmark, European American, and Hawaiian individuals (8). Additional meta-analyses by Lubbe et. al., in conjunction with Peters et. al., identified several additional SNP's on various chromosomal regions (7, 9). These include rs4779584, rs16892766, rs961253, rs9929218, rs6983267, rs4939827 and rs7315438 – all of which have been included in our genetic risk score explained below. The majority of the SNPs chosen are from Peters et al.; 17 of the 24 SNPs within our score are from their analysis. Peters et. al. conducted a GWAS meta-analysis coupled replication study using 10 journals that had previously published SNPs statistically significantly associated with CRC (7). This study contained almost entirely Caucasian individuals from the United States and Europe. Dunlop et. al. is another large European study of approximately 44,389 subjects from predominantly European descent (10). Individually, the association of these SNP's with CRC is marginal ($0.8<OR<1.3$), but we hypothesize that the accumulation of multiple variants may modify CRC risk to a greater extent. Similar hypotheses have been suggested, such as those by Zhang et al.. This study

incorporated 20 variants (three of which are repeatedly associated with CRC occurrence) into a genetic risk score. The score was broken into quintiles and all measurements compared to the lowest quintile of genetic risk. This study, although not entirely compromised of the same variants, demonstrated genetic risk by SNP variant is marginal, but accumulation of risk variants more greatly modifies your risk of CRC (11).

WCRF/AICR Recommendations

Previous literature has suggested that lifestyle choices such as physical activity, smoking, and dietary patterns are associated with CRC occurrence. In 2007, The World Cancer Research Fund and American Institute of Cancer Research (WCRF/AICR, respectively) issued eight general recommendations, and two special recommendations, on diet, physical activity, and weight management for cancer prevention (3, 17). A study by Romeguera et. al. used these recommendations within the EPIC Cohort to construct a lifestyle score that ranged from 0-6 for men and 0-7 for women; increasing score representing a better concordance with the WCRF/AICR recommendations (17). The score compiled risk factors with suggestive evidence of CRC risk reduction – of which included body fatness, physical activity, alcohol intake, red/processed meat intake, fruit/vegetable intake and foods promoting weight gain. WCRF/AICR scores constructed for women also included breastfeeding. The individual components were given a score based on whether they fully met the criteria (1 point), half met the criteria (0.5 points) or did not meet the criteria (0 points). Scoring criteria for plant food intake and foods promoting weight gain contained multiple components within its score and therefore an average was calculated - making their plausible values 0, 0.25, 0.5, 0.75 and 1 (17). Supplemental table 1 explains WCRF/AICR score construction criteria in greater detail.

The study then used Cox proportional hazards tests to approximate deaths secondary to CRC. The study showed an inverse association between high WCRF/AICR concordance and CRC mortality, as well as all-cause mortality (17). Additionally, a 1-point increase in WCRF score was associated with a 10% reduction in CRC mortality. These results are in agreement with current scientific literature that suggests there are several modifiable factors that have preventative effects on cancer occurrence and recurrence. Since publication, this score has been implemented in studies of other cancer types, such as prostate cancer as in Thederan et al. This retrospective study of 2227 men with prostate cancer ("PCa") scheduled to receive prostatectomy showed that the study sample was disproportionately discordant to the recommendations. 67.3% of individuals did not meet BMI standards, 33.5% reported no physical activity, 49.6% were current smokers, 75.4% did not meet recommended meat intake criteria and 88.8% did not meet fruit/vegetable intake criteria (18). The researchers also report that a self-questionnaire was used to obtain all health and dietary information, suggesting the results are likely underreported from the truth (18). This study concluded that a very small proportion of the individuals with PCa were adhering to lifestyle suggestions, further strengthening the role of WCRF/AICR concordance with cancer prevention (18).

Gene-Environment Interaction

According to a journal by the Institute of Medicine, the complexity behind cancer prevention and treatment lies in the underlying mechanisms at play. Cancer is widely believed to be a culmination of several molecular abnormalities that have some crossover in function, but ultimately lead to uncontrolled cell growth that eventually becomes a cancerous tumor (20). The article also notes that the individual aspects of colorectal

cancer formation are somewhat understood, but the complexity arises in how these separate mechanisms play off of one another (20). Healthy lifestyle choices, such as moderate physical activity and diets favoring micronutrient intake are widely suggested to have a preventative association with CRC risk; current literature has focused on micronutrients such as vitamin D, calcium, fiber and folate (20). Genetic mechanisms are a bit more complex. A study of aspirin use for CRC treatment showed moderate association to reducing tumor formation, suggesting that its mechanism of action, COX inhibition, may have a role in mitigating carcinogenesis (20). Other mechanisms such as epigenetics – which are molecular changes such as methylation and histone modification that occur over our lifetime – may also be involved in tumor formation. The article notes that some of the most successful gene-environment studies to date have incorporated genetics, environment and demographic into their models of risk assessment (20). Gene-Environment interaction studies have become more robust in approximating colorectal cancer formation with advancing measurement strategies such as GWAS analysis, meta-analyses and even newer methods such as Gene-environment-wide interaction studies (GEWIS). These advances have solved some of the problems that existed in GxE interaction studies prior to (such as publication bias, marginal associations, small sample size and insufficient threshold for significance (21)), but scientists continue to work and improve the methods used and the underlying assumptions they carry. Evidence of such problems with these studies are obvious when we look at the GxE studies of breast cancer published prior to May 2011. 307 of 407 (75%) of the studies reported statistically significant GxE interaction – suggesting that we may have false-positives occurring because of these underlying assumptions between the two factors that we don't quite

understand (21). Still, unconditional logistic regression is a standard, and robust, tool for GxE interaction and will be our method of analysis within our study. Several gene-environment studies of colorectal cancer have been published. Hutter et al. conducted a meta-analysis of approximately 4,200 Caucasians, in two independent populations, and attempted to observe estimates of association by SNP (22). The SNP's suggested a similar association as that of previous literature but, when measuring for interaction between the individual SNP's and risk factors of CRC, only one combination of factors (rs10808555 and BMI) showed a significant difference in CRC risk (22). All other factors analyzed showed no such association. A more recent study by Balavarca et al. constructed a 55-SNP genetic score, as well as an environmental risk score that utilized several previously suggested risk factors of CRC. Advanced colorectal neoplasms, which include both cancerous and noncancerous polyps of the colon, were the outcome of interest (24). The study suggested that a combination of the two factors can more greatly modify your risk of colorectal neoplasm than just genetics or environment individually (24).

Summary

The purpose of this study is to examine the association of genetic susceptibility to CRC and lifestyle habits, separately and jointly, on CRC risk. Previous studies have examined the association of WCRF/AICR concordance and Genetic Risk Scores on CRC individually, but there is limited literature examining both factors concurrently to evaluate how lifestyle habits may modify CRC risk within individuals of various genetic risk.

Methods

Study population and data collection

The rationale and methods of the EPIC study, including information on dietary assessment methods, blood collection protocols, and follow-up procedures have been reviewed previously (26, 27). The European Prospective Investigation into Cancer and Nutrition (EPIC) cohort is a multi-center prospective cohort consisting of approximately 519,978 individuals (366,521 women and 153,457 men) aged 35-70 years old distributed across 23 medical centers and 10 countries in Europe (26). Altogether, this dataset represents the largest single resource for investigations into cancer etiology and mortality. Individuals who were eligible for the study were selected from the general population of a specific geographical area, town, or province. Exceptions included the French sub-cohort, which is based on members of the health insurance system or state-school employees, and the Utrecht (Netherlands) sub-cohort, which is based on women who underwent screening for breast cancer. Between 1992 and 1998, standardized lifestyle and personal history questionnaires, anthropometric data, and blood samples were collected from most participants at recruitment, before disease onset or diagnosis. Diet over the previous 12 months was assessed at recruitment by validated country-specific questionnaires designed to ensure high compliance and improved measures of local dietary habits. 24 Blood samples were stored at the International Agency for Research on Cancer (Lyon, France; -196°C, liquid nitrogen) for all countries except Denmark (-150°C, nitrogen vapor) and Sweden (-80°C freezers). Values for dietary intake of total energy, vitamin D, calcium, and retinol were computed using country-specific food composition tables.

Nested case-control design and participant selection

Case ascertainment and selection

Colon cancers were defined as tumors in the cecum, appendix, ascending colon, hepatic flexure, transverse colon, splenic flexure, descending and sigmoid colon (C18.0-C18.7 as per the 10th Revision of the International Statistical Classification of Diseases, Injury and Causes of Death), and overlapping or unspecified origin tumors (C18.8 and C18.9). Rectal cancers were defined as tumors occurring at the rectosigmoid junction (C19) or rectum (C20). Anal canal cancers were excluded. Colorectal cancer is the combination of the colon and rectal cancer cases (26, 27). After excluding all other cancer-types, 1,462 eligible cases were included (colon cancer=929; rectal cancer=533). 1,322 of the eligible cases had a complete WCRF/AICR and were eligible to use in statistical analysis; the 140 cases with incomplete WCRF/AICR score are reported in table 2 but have been excluded from all unconditional logistic regression models. Cases were not selected from Norway (blood samples only recently collected; few colorectal cancers diagnosed after blood donation) and the Malmö center of Sweden.

Control selection

Controls were selected by incidence density sampling from all cohort members alive and free of cancer. After exclusion, 1,482 eligible controls were included. Of these 1,482 controls, 1,342 had a complete WCRF/AICR score and were eligible for use in statistical analysis. The 140 controls with incomplete WCRF/AICR scores are within table 2 but have been excluded from all unconditional logistic regression models.

SNP Selection and Genotyping

Polymorphisms of interest were selected from previous GWAS studies related to CRC susceptibility loci. 25 SNPs were selected that have been previously associated with CRC within these studies; however, one SNP (rs4444235) was excluded due to a call rate > 0.05. The remaining 24 SNPs met Hardy-Weinberg criteria, had <10% missing data, and were included within the score. Further information on the SNPs selected can be found in Table 1**.**

The SNPs were genotyped as part of a custom GoldenGate assay designed using the Illumina online Assay Design Tool in May 2012. The Genetics Laboratory at Imperial College London performed Genotyping. All genotyping underwent standard quality control including concordance checks for blinded duplicates and examination of sample and SNP call rates. The lowest reproducibility frequency across 62 replicate samples was 0.98. The call rate was 95% for all samples and 95% for all SNPs (27).

GRS Construction

In effort to build a GRS that may evaluate the cumulative impact of CRC susceptibility loci, we constructed a score from 0 to 24 that corresponds to the 24 SNPs previously mentioned. Subjects were then assigned either a zero or one for each variant, with a one representing the presence of at least one risk allele for that SNP. The risk alleles chosen for each of the 24 SNPs are based on observations made in previous literature. Of the 2,944 initial subjects, 2,104 (cases and controls) had complete genotypic information for all 24 SNPs and were eligible for further analysis.

WCRF/AICR Construction

The WCRF/AICR score used within this study is adapted from that used by Romaguera et al; their scoring criteria per recommendation is provided in supplemental table 1.

Unlike the score used by Romaguera et al, we have chosen to remove breastfeeding as a component of the score due to observations of a null association between breastfeeding and CRC (17). The WCRF/AICR score used in this study has a range of 0-6 for both sexes, with a higher value being suggestive of higher concordance with cancer prevention recommendations. As mentioned previously, 2,104 of the 2,944 subjects had complete WCRF/AICR scores assembled and were eligible for further analysis.

Statistical Analysis

Differences in baseline characteristics between cases and controls were assessed using the mean (continuous variables) or percentages (categorical variables).

We used unconditional logistic regression analysis to assess the association of individual SNPs with CRC risk, adjusting for age (continuous), sex, and coordinating center. In a nested case-control study where controls are selected using incidence density sampling, this procedure estimates the incidence rate ratio (RR), which, given the rarity of the disease, is approximately equal to the OR. Results were similar when we used conditional logistic regression on 1,328 complete case-matched sets. We assumed a log-additive genetic model, but also tested dominant and recessive models, as the underlying genetic model for these SNPs is unknown. ORs and 95% confidence intervals of all individual SNPs, by model, can be found in supplemental table 2. Further adjustment for body mass index (BMI; continuous), physical activity (active, moderately active, moderately inactive, and inactive), red and processed meats (continuous), smoking status (never, former, current smokers, missing), fruit and vegetables (continuous), alcohol intake (continuous), and total energy intake (continuous) did not substantially change the results, and thus these variables were not included in the final statistical model.

Subgroup analyses by sex and tumor location (colon *vs.* rectum) did not substantially change our results and were not included in the final statistical model.

GRS and WCRF/AICR scores are stratified into three-level categorical variables in effort to ensure reasonable sample size within each category (GRS categories: 0-14 SNPs, 15-17, and ≥18; WCRF/AICR categories: 0-2.99, 3-3.749, ≥3.75). The cutoff values of each category are based on the distribution of these scores within all controls in effort to analyze our cases against a sample representative of the population. Sensitivity analysis on the categorical parameters were tested using unconditional logistic regression models on various parameter cutoff points but no appreciable change in association with CRC risk was observed.

Stratified analysis of the GRS and WCRF/AICR, and its association with CRC risk were assessed using unconditional logistic regression models to estimate the OR and 95% for each tertile of genetic risk and WCRF/AICR adherence compared to the lowest risk group of stratification. Joint analysis of GRS and WCRF/AICR used similar models as stated above but all measurements of association were compared to subjects with the lowest level of genetic risk and highest level of WCRF/AICR adherence (GRS≤14; WCRF≥3.75). We used traditional methods to assess potential interactions between GRS and WCRF/AICR adherence.

All statistical methods were two-sided with P-values <0.05 considered statistically significant (SAS software, version 9.4, SAS Institute, Cary, NC).

Results

Study Population

Table 1 includes baseline characteristics of our study population. The study contains 2,944 individuals, 732 (49.4%) of whom are women. Participants with CRC, on average, had a higher BMI, total energy intake, alcohol intake, and red/processed meat consumption when compared to those without CRC; individuals with CRC were also less physically active and consumed less fruit and vegetables compared to those without CRC. However, differences with respect to most single risk factors were relatively small.

Genetic Risk Score

The risk allele frequencies of the 24 SNPs used in our sample closely match those published in previous literature. Out of the 24 SNPs selected, 22 displayed a consistent association with CRC (OR>1). The other two SNPs (rs10411210 and rs4939827) suggested a protective association to CRC; however, prior literature is not agreement on the risk allele for these SNPs. The OR and 95% of each additional SNP was calculated using <14 SNPs as the reference category; all but two (SNP count = 14, 18) reported an OR>1 and confidence interval that did not cross the null (OR=1). These measurements of association to CRC risk can be found in figure 1. The GRS, when analyzed as a categorical variable, had a similar association with CRC as those measuring the cumulative effect. When the two higher categories of genetic risk were compared to the lowest category, both were statistically significantly associated with higher CRC risk (OR=1.28, 95%: 1.02,1.61; OR= 1.36, 95%: 1.09, 1.69). Similar measurements of association were appreciable when CRC was split into colon and rectal cancer (table 3).

WCRF/AICR Score

The association between WCRF/AICR score on CRC was assessed using the lowest tertile (WCRF<2.99) as the reference category. Both a moderate and high level of adherence to the WCRF/AICR recommendations are statistically significant in suggesting an inverse association to CRC risk when compared to the lowest level of adherence (OR=0.71, 95%: 0.57, 0.88; OR=0.79, 95%: 0.63, 0.98). These results support previous knowledge that certain modifiable factors have a preventative association with CRC.

Interaction between Genetic Score and WCRF/AICR Score

Higher genetic risk was associated with a higher risk of CRC compared to low genetic risk, but the associations were lowered as WCRF/AICR concordance increased. Moderate and high genetic risk and low adherence to the WCRF/AICR recommendations were statistically significantly associated with higher CRC risk compared to low genetic risk and low adherence to the WCRF/AICR recommendations. Our final analysis attempted to compare the "healthiest" individual to all other categories by setting the lowest genetic risk and highest adherence with WCRF/AICR recommendations as the reference category. This method suggested an increased association with CRC as the genetic risk increased, and became statistically significant when moderate and high genetic risk but low WCRF/AICR adherence were compared to individuals with high adherence to WCRF/AICR recommendations and low genetic risk (OR=1.71, 95%: 1.14, 2.55; OR=1.63, 95%: 1.12, 2.38).

Discussion

In this large European prospective case-control study nested within the EPIC cohort, we investigated whether genetic variants previously identified in the GWAS studies to be associated with higher risk for CRC are associated with a cumulative increase in CRC risk, and whether this association is modified by lifestyle factors as assessed by the adherence to the WCRF/AIRC recommendations.

Our study found that higher genetic risk is statistically significantly associated with higher CRC risk compared to low genetic risk. It also showed that a higher concordance with the WCRF/AICR recommendations is associated with an inverse association with CRC risk when compared to those with low adherence. Furthermore, this study is one of the first to investigate how the two factors together modify CRC risk. We showed that, regardless of genetic risk, higher concordance with WCRF/AICR recommendations is associated with lower risk for CRC.

This study is in agreement with those four previous studies (7, 9, 10, 11) suggesting that GWAS-identified CRC susceptibility loci are not only associated with CRC individually, but that these loci accumulate cumulative effect on CRC risk. The study also reaffirms previous literature that suggests a higher concordance with the dietary and lifestyle recommendations put forth by the WCRF/AICR is inversely associated with CRC risk (17, 18, 19).

This study is one of the first to analyze the association of susceptibility loci of CRC and lifestyle recommendations on CRC risk together, recognizing that CRC has both a genetic component as well as a strong environmental aspect and may modify our risk of

CRC in a different manner than the two factors alone. One such study recently published predicts the risk of advanced colorectal neoplasms using a 55-SNP GRS and self-constructed environmental risk score (24). The study was similar to ours in supporting that the two factors can modify CRC risk in a manner that is not suggested when measuring them individually. Our study differs from Balavarca et al in that we are only interested in CRC, not the combination of cancerous and noncancerous polyp formation.

Strengths

The main strength of our study was that it was based on a prospective cohort that was followed over an 11-year period, and obtained multiple measures necessary for cancer etiology studies. Information pertaining to dietary patterns was collected, and then validated in a subset of the population to ensure accuracy. Additionally, the cohort contains a very large sample size of approximately 520,000 individuals of which we were able to use 2,104 to conduct our analysis. The utilization of WCRF/AICR parameters for health assessment is valuable in that all recommendations within the score have been well studied in previous literature, and suggest a strong association between the individual recommendation and CRC risk.

Limitations

Although this study has many strengths secondary to a large sample size and multiple measuring factors already associated, albeit individually, to CRC risk – we do have some weaknesses within our study that would be beneficial to address in future studies. Many of our weaknesses are secondary to the EPIC dataset being locked, and therefore incapable of providing updated data. Family History of CRC, or any cancer, was not

ascertained during the follow-up period but has been associated with higher CRC risk in past studies (10, 12, 24). Additionally, approximately 800 subjects had to be excluded from our analysis because of missing or incomplete data; as was the case with both our GRS and WCRF/AICR score. In regards to our SNPs selected, we were limited to the SNPs that had been genotyped prior to receiving the dataset. GWAS studies of susceptibility loci of CRC published since this dataset has identified several additional SNPs significantly associated with CRC, suggesting that we may not have constructed a GRS that is most representative of common variants within the population. Discrepancies in risk alleles for each SNP also present an area of confusion. GWAS studies of the SNPs we incorporated into our GRS commonly listed measurements of association that crossed the null when compared to one another, or identified similar measurements of association but assigned a different risk allele. Our study attempted to measure association to CRC risk using the risk allele that has been most agreed upon in multiple studies, but we cannot be certain that our decision is the most accurate without further research.

Conclusions and Future Directions

Our results for the individual association of GRS and WCRF/AICR to CRC risk are consistent with previous literature; however, our study is one of the first to look at the interaction between genetics and lifestyle habits and how they modify the association to CRC risk when analyzed together. Our study demonstrated that common genetic variants of CRC have a marginal impact on CRC risk individually, but can have a higher association to CRC as the number of genetic variants increases. With respect to modifiable factors, our WCRF/AICR score suggests that certain lifestyle choices such as body fatness, plant food intake and physical activity are capable of lowering CRC risk, regardless of genetic risk. The study further suggests that our lifestyle factors play a critical role in cancer prevention and health maintenance. While our genetic makeup is determined and (mostly) unchangeable, our lifestyle habits represent our everyday choices and can either lower or further amplify our risk of disease; with the change in association depending on each individual's combination of the two factors. Future directions of this study should attempt to capture a larger GRS that incorporates more recently identified CRC susceptibility loci, as well as loci that have risk alleles strongly agreed upon by prior studies. The study further strengthens the use of primary prevention in cancer prevention.

Citations

1. American Cancer Society. Cancer Facts & Figures 2019. Atlanta: American Cancer Society; 2019.

2. American Cancer Society. Colorectal Cancer Facts & Figures 2017-2019. Atlanta: American Cancer Society; 2017.

3. World Cancer Research Fund/American Institute for Cancer Research. *Diet, Nutrition, Physical Activity and Cancer: a Global Perspective*. Continuous Update Project Expert Report 2018.

4. U.S. Cancer Statistics Working Group. U.S. Cancer Statistics Data Visualizations Tool, based on November 2017 submission data (1999-2015): U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute; www.cdc.gov/cancer/dataviz, June 2018.

5. H. Brenner, M. Kloor, C.P. Pox. Colorectal cancer. Lancet (Lond., Engl.), 383 (9927) (2014), pp. 1490-1502

6. PY Kwok, X Chen. Detection of single nucleotide polymorphisms. Curr Issues Mol Biol, 5 (2003), pp. 43-60

7. Peters, U., Hutter, C. M., Hsu, L., Schumacher, F. R., Conti, D. V., Carlson, C. S., … Casey, G. (2011). Meta-analysis of new genome-wide association studies of colorectal cancer risk. *Human genetics*, *131*(2), 217–234. doi:10.1007/s00439-011-1055-0

8. Thean LF, Li HH, Teo YY, Koh W-P, Yuan J-M, et al. (2012) Association of Caucasian-Identified Variants with Colorectal Cancer Risk in Singapore Chinese. PLoS ONE 7(8): e42407. doi:10.1371/journal.pone.0042407

9. Lubbe SJ, Di Bernardo MC, Broderick P, et al. Comprehensive evaluation of the impact of 14 genetic variants on colorectal cancer phenotype and risk, Am J Epidemiol , 2012, vol. 175 1(pg. 1-10)

10. Dunlop, M. G., Tenesa, A., Farrington, S. M., Ballereau, S., Brewster, D. H., Koessler, T., … Houlston, R. S. (2012). Cumulative impact of common genetic variants and other risk factors on colorectal cancer risk in 42,103 individuals. *Gut*, *62*(6), 871–881. doi:10.1136/gutjnl-2011-300537

11. Burnett-Hartman, A. N., Passarelli, M. N., Adams, S. V., Upton, M. P., Zhu, L. C., Potter, J. D., & Newcomb, P. A. (2013). Differences in epidemiologic risk factors for colorectal adenomas and serrated polyps by lesion severity and anatomical site. *American journal of epidemiology*, *177*(7), 625–637. doi:10.1093/aje/kws282

12. Zhang, B., Shrubsole, M. J., Li, G., Cai, Q., Edwards, T., Smalley, W. E., … Zheng, W. (2012). Association of genetic variants for colorectal cancer differs by subtypes of polyps in the colorectum. *Carcinogenesis*, *33*(12), 2417–2423. doi:10.1093/carcin/bgs308

13. Carvajal-Carmona, L. G., Cazier, J. B., Jones, A. M., Howarth, K., Broderick, P., Pittman, A., … Tomlinson, I. (2011). Fine-mapping of colorectal cancer susceptibility loci at 8q23.3, 16q22.1 and 19q13.11: refinement of association signals and use of in silico analysis to suggest functional variation and unexpected candidate target genes. *Human molecular genetics*, *20*(14), 2879–2888. doi:10.1093/hmg/ddr190

14. COGENT Study, Houlston, R. S., Webb, E., Broderick, P., Pittman, A. M., Di Bernardo, M. C., … Dunlop, M. G. (2008). Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nature genetics*, *40*(12), 1426–1435. doi:10.1038/ng.262

15. Tomlinson, I. P. M., Webb, E., Carvajal-Carmona, L., Broderick, P., Howarth, K., Pittman, A. M., ... Houlston, R. S. (2008). A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nature Genetics*, *40*(5), 623-630. https://doi.org/10.1038/ng.111

16. Pittman, A. M., Naranjo, S., Jalava, S. E., Twiss, P., Ma, Y., Olver, B., … Houlston, R. S. (2010). Allelic variation at the 8q23.3 colorectal cancer risk locus functions as a cis-acting regulator of EIF3H. *PLoS genetics*, *6*(9), e1001126. doi:10.1371/journal.pgen.1001126

17. Romaguera, D., Vergnaud, A.-C., Peeters, P. H., van Gils, C. H., Chan, D. S. M., Ferrari, P., … Norat, T. (2012). Is concordance with World Cancer Research Fund/American Institute for Cancer Research guidelines for cancer prevention related to subsequent risk of cancer? : Results from the EPIC study. American Journal of Clinical Nutrition, 96(1), 150–163. https://doi.org/10.3945/ajcn.111.031674

18. Poor Adherence to International Cancer Prevention Recommendations Among Patients With Prostate Cancer: First Results From the MARTINI-Lifestyle Cohort. Thederan, Imke et al. European Urology Focus , Volume 0 , Issue 0 ,

19. Hastert, T. A., Beresford, S. A., Patterson, R. E., Kristal, A. R., & White, E. (2013). Adherence to WCRF/AICR cancer prevention recommendations and risk of postmenopausal breast cancer. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology, 22(9), 1498–1508. doi:10.1158/1055-9965.EPI-13-0210

20. Institute of Medicine (US) Roundtable on Environmental Health Sciences, Research, and Medicine; Wilson S, Jones L, Couseens C, et al., editors. Cancer and the Environment: Gene-Enviroment Interaction. Washington (DC): National Academies Press (US); 2002. 5, Gene–Environment Interaction in Site-Specific Cancers. Available from: https://www.ncbi.nlm.nih.gov/books/NBK220887

21. Hutter, C. M., Mechanic, L. E., Chatterjee, N., Kraft, P., Gillanders, E. M., & NCI Gene-Environment Think Tank (2013). Gene-environment interactions in cancer epidemiology: a National Cancer Institute Think Tank report. *Genetic epidemiology*, *37*(7), 643–657. doi:10.1002/gepi.21756

22. Hutter, C. M., Slattery, M. L., Duggan, D. J., Muehling, J., Curtin, K., Hsu, L., … Peters, U. (2010). Characterization of the association between 8q24 and colon cancer: gene-environment exploration and meta-analysis. *BMC cancer*, *10*, 670. doi:10.1186/1471-2407-10-670

23. Gong J, Hutter CM, Newcomb PA, Ulrich CM, Bien SA, Campbell PT, et al. (2016) Genome-Wide Interaction Analyses between Genetic Variants and Alcohol Consumption and Smoking for Risk of Colorectal Cancer. PLoS Genet 12(10): e1006296. https://doi.org/10.1371/journal.pgen.1006296

24. Balavarca, Y. , Weigl, K. , Thomsen, H. and Brenner, H. (2019), Performance of individual and joint risk stratification by an environmental risk score and a genetic risk score in a colorectal cancer screening setting. Int. J. Cancer. doi:10.1002/ijc.32272

25. Yang, T. , Li, X. , Montazeri, Z. , Little, J. , Farrington, S. M., Ioannidis, J. P., Dunlop, M. G., Campbell, H. , Timofeeva, M. and Theodoratou, E. (2019), Gene–environment interactions and colorectal cancer risk: An umbrella review of systematic reviews and meta-analyses of observational studies. Int. J. Cancer. doi:10.1002/ijc.32057

26. Riboli, E., Hunt, K., Slimani, N., Ferrari, P., Norat, T., Fahey, M.,…Saracci, R. (2002). European Prospective Investigation Into Cancer and Nutrition (EPIC): Study populations and data collection. Public Health Nutrition, 5(6b), 1113-1124. Doi: 10.1079/PHN2002394

27. Jenab M, McKay J, Bueno-de-Mesquita HB , et al. Vitamin D receptor and calcium sensing receptor polymorphisms and the risk of colorectal Cancer in european populations. Cancer Epidemiol Biomarkers Prev 2009;18:248591. doi:10.1158/1055-9965.EPI-09-0319

Tables/Figures

Table 1. SNPs identified in GWAS to be significantly associated with CRC and considered for inclusion in the genetic risk score (GRS), EPIC Study, 1992-2003

SNP	Gene	Chr	Position	Risk Allele	MAF EUR	MAF among controls	% Missing Data	Included in GRS	Reference
rs2373859	SLC8A1	2	40390680	T	0.34	0.36	6.08	y	Peters, 2012
rs275454	LOC442132/POLS	5	6815900	A	0.39	0.35	4.08	y	Peters, 2012
rs2853668	TERT/CLPTM1L	5	1299910	G	0.27	0.26	4.62	y	Peters, 2012
rs1525461	TPK1/CNTNAP2	7	145031198	C	0.20	0.19	5.98	y	Peters, 2012
rs11986063	EIF3H/TRPS1	8	116628076	T	0.10	0.10	4.08	y	Houlston, 2008
rs16888522	EIF3H/TRPS1	8	116561675	T	0.06	0.07	3.87	y	Peters, 2012
rs16888589	EIF3H/TRPS1	8	116623363	A	0.09	0.09	5.98	y^{++}	Pittman,2010
rs16892766	TRPS1/EIF3H	8	116618444	C	0.09	0.09	3.67	y	Tomlinson, 2008
rs6983267	POU5F1B/FAM84B	8	127401060	G	0.50	0.50	6.35	y	Zhang, 2013
rs7837208	RAD21/UTP23	8	116785633	G	0.11	0.10	3.77	y	Caravaja-Carmona,2011
rs10795668	GATA3/SFTA1P	10	8659256	G	0.32	0.33	15.35	n^{+}	Lubbe,2011
rs3802842	LOC120376	11	111300984	C	0.27	0.27	5.33	y	Lubbe,2011
rs7315438	MED13L/TBX3	12	115453598	T	0.45	0.42	4.45	y	Peters, 2012
rs4779584	SCG5/GREM1	15	32702555	T	0.20	0.18	39.71	n^{+}	Zhang, 2012
rs2059254	CDH1	16	68783536	C	0.29	0.30	13.76	n^{+}	Caravaja-Carmona,2011
rs2113200	CDH1	16	68781045	T	0.29	0.29	3.97	y^{++}	Caravaja-Carmona,2011
rs8056538	CDH1	16	68768379	G	0.29	0.29	7.85	y^{++}	Caravaja-Carmona,2011
rs9925923	CDH1	16	68785711	C	0.29	0.29	3.87	y	Caravaja-Carmona,2011
rs9929218	CDH1	16	68787043	G	0.29	0.30	4.14	y^{++}	Lubbe,2011
rs4939827	SMAD7	18	48927093	T	0.47	0.50	7.81	y	Lubbe,2011
rs10411210	RHPN2	19	33041394	C	0.10	0.10	3.63	y	Lubbe,2011
rs4813802	BMP2/FERMT1	20	6718948	G	0.32	0.33	4.72	y	Peters, 2012
rs4925386	LAMA5	20	62345988	C	0.33	0.29	6.42	y	Zhang, 2012
rs961253	BMP2/FERMT1	20	6423634	A	0.36	0.35	6.28	y	Houlston, 2008

SNPs were excluded from the genetic score based on missingness and linkage disequilibrium (Rsquared >.8): $^{+}$ SNPs with > 10% missing data; $^{++}$ rs16888589 serves as a proxy for rs16892766, rs2113200 serves as a proxy for rs9925923, rs9929218 serves as a proxy for rs9925923 and rs2113200, rs8056538 serves as a proxy for rs9925923, rs2113200, and rs9929218.

SNP = Single Nucleotide Polymorphism; Chr = Chromosome; MAF=Minor Allele Frequency

Table 2. Selected baseline characteristics of incident colon and rectal cancer cases and their controls, the EPIC study, 1992-2003

Characteristic*	Colorectal cancer		Colon cancer	Rectal cancer
	Cases	Controls	Cases	Cases
N	1462	1482	929	533
Women, n(%)	725 (49.6)	732 (49.4)	491 (52.9)	234 (43.9)
Mean age at blood collection, yrs	58.5	58.6	58.8	58.1
Mean years between blood collection and diagnosis	4.5	--	4.4	4.6
Educational attainment, n(%)				
Primary	505 (34.9)	546 (37.1)	324 (35.2)	181 (34.4)
Technical/professional school	348 (24)	397 (27)	208 (22.6)	140 (26.6)
Secondary	221 (15.3)	165 (11.2)	155 (16.8)	66 (12.5)
Degree	263 (18.2)	261 (17.7)	157 (17.1)	106 (20.1)
Smoking status, n(%)				
Never smoker	601 (41.1)	627 (42.3)	397 (42.7)	204 (38.3)
Former smoker	489 (33.5)	479 (32.3)	309 (33.3)	180 (33.8)
Current smoker	352 (24.1)	360 (24.3)	210 (22.6)	142 (26.6)
Physical activity, n(%)				
Inactive	238 (16.3)	204 (13.8)	148 (15.9)	90 (16.9)
Moderately inactive	418 (28.6)	405 (27.3)	266 (28.6)	152 (28.5)
Moderately active	570 (39)	617 (41.6)	375 (40.4)	195 (36.6)
Active	126 (8.6)	140 (9.5)	72 (7.8)	54 (10.1)
BMI, kg/m^2, mean±SD	26.8±4.2	26.3±3.8	26.9±4.4	26.8±4.0
Mean baseline dietary intakes, g/d				
Total energy, kcal/d	2142.1	2114.7	2111.7	2195.2
Alcohol	16.6	14.9	15.1	19.2
Calcium	1.0	1.0	1.0	1.0
Fiber	22.8	23.2	22.5	23.3
Folate, ug/d	301.4	302.4	299.6	304.5
Fruit and vegetables	409.9	423.2	418.5	394.9
Red and processed meats	88.1	83.3	84.2	95.0
WCRF/AICR Score, Categorized, Tertiles, n(%)				
Category 0 (0-2.749)	517 (39.1)	450 (33.5)	316 (37.6)	201 (41.8)
Category 1 (2.749-3.749)	380 (28.7)	438 (32.6)	249 (29.6)	131 (27.2)
Category 2 (3.75+)	425 (32.2)	454 (33.8)	276 (32.8)	149 (31)

*Number Missing/Unknown within CRC: smoking=52 , physical activity=226, education=24, BMI=17, Baseline dietary intakes=3, WCRF=280

Table 3: Unconditional Logistic Regression Analysis of Tertiles of Genetic Risk Score and WCRF/AICR Score and CRC risk, EPIC Study, 1992 - 2003

Tertiles	Sample Size		Colorectal Cancer	Colon Cancer	Rectal Cancer
	Cases	**Controls**	**OR (95%)**	**OR (95%)**	**OR (95%)**
Genetic Risk Score					
0-14	242	301	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)
15-17	348	341	1.28 (1.02, 1.61)	1.38 1.06, 1.79)	1.13 (0.83, 1.54)
18+	456	416	1.36 (1.09, 1.69)	1.41 (1.09, 1.81)	1.33 (0.99, 1.78)
WCRF/AICR Score					
0-2.99	416	350	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)
3-3.749	292	346	0.71 (0.57, 0.88)	0.72 (0.56, 0.92)	0.68 (0.51, 0.91)
3.75+	338	362	0.79 (0.63, 0.98)	0.76 (0.59, 0.98)	0.81 (0.61, 1.09)

*Adjusted for sex,age at recruitment and coordinating center

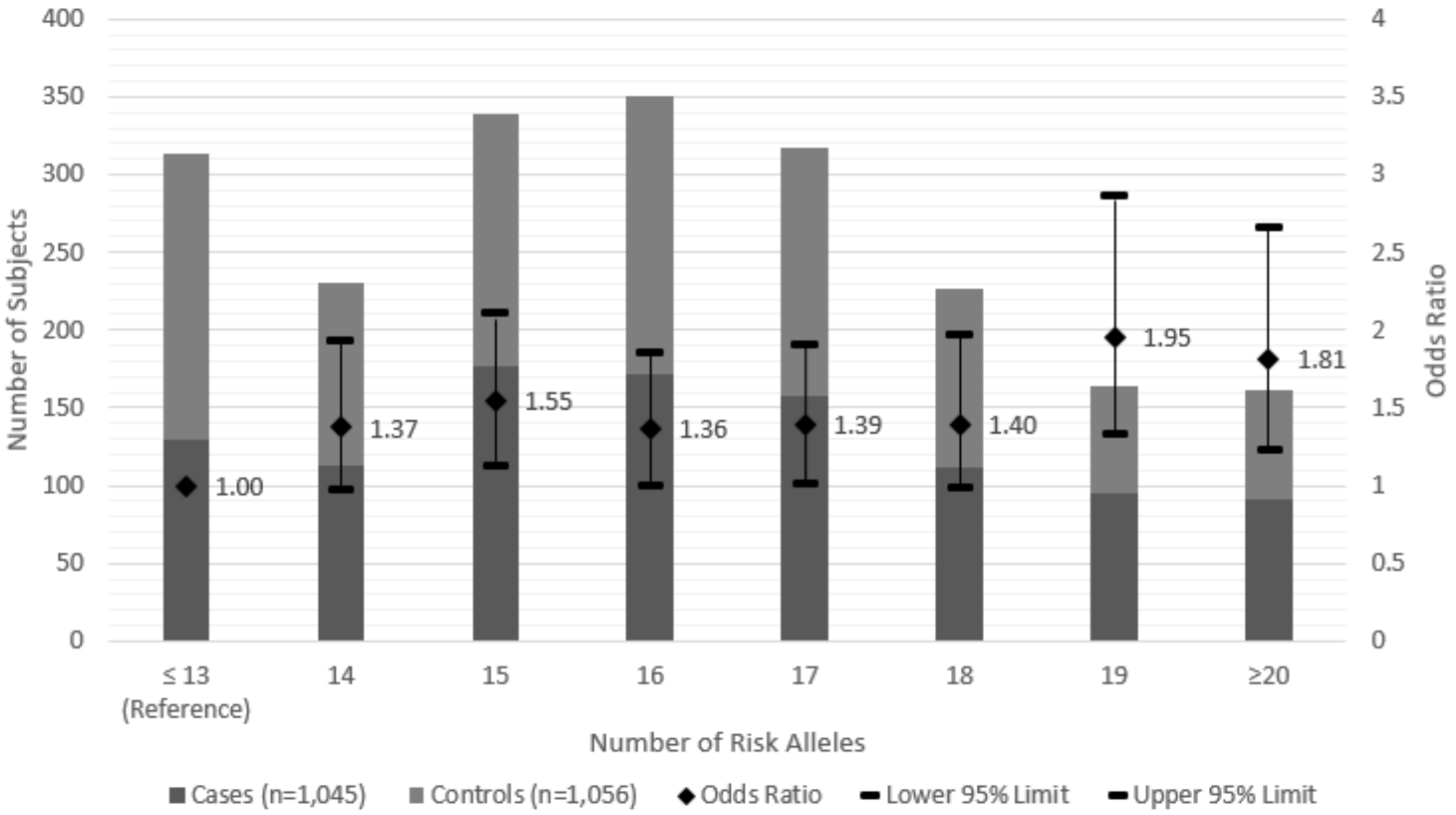
Table 4: Unconditional Logistic Regression Analysis Stratified by Tertiles of Genetic Risk Score and/or WCRF/AICR Score and CRC risk, EPIC Study, 1992 - 2003

WCRF Score	Genetic Risk Score								
	Tertile 1 (0-14)			Tertile 2 (15-16)			Tertile 3 (≥18)		
	Cases	Controls	OR(95%)	Cases	Controls	OR(95%)	Cases	Controls	OR(95%)
Stratified by Genetic Risk Score									
Tertile 1 (0-2.99)	95	109	1.00 (Ref.)	134	99	1.00 (Ref.)	187	142	1.00 (Ref.)
Tertile 2(3-3.749)	67	93	0.84 (0.55, 1.29)	105	125	0.62 (0.42, 0.90)	120	128	0.72 (0.51, 1.01)
Tertile 3(3.75+)	80	99	0.88 (0.57, 1.36)	109	117	0.67 (0.45, 0.99)	149	146	0.81 (0.58, 1.14)
Stratified by WCRF Score									
Tertile 1 (0-2.99)	95	109	1.00 (Ref.)	134	99	1.57 (1.07, 2.30)*	187	142	1.51 (1.05, 2.16)*
Tertile 2(3-3.749)	67	93	1.00 (Ref.)	105	125	1.18 (0.78, 1.79)	120	128	1.35 (0.89, 2.03)
Tertile 3(3.75+)	80	99	1.00 (Ref.)	109	117	1.16 (0.78, 1.73)	149	146	1.28 (0.87, 1.87)
Joint Genetic Risk Score x WCRF									
Tertile 1 (0-2.99)	95	109	1.08 (0.72, 1.63)	134	99	1.71 (1.14, 2.55)	187	142	1.63 (1.12, 2.38)
Tertile 2(3-3.749)	67	93	0.89 (0.58, 1.38)	105	125	1.04 (0.70, 1.55)	120	128	1.17 (0.79, 1.73)
Tertile 3(3.75+)	80	99	1.00 (Ref.)	109	117	1.17 (0.79, 1.74)	149	146	1.27 (0.87, 1.85)

Tertile parameters are based on the distribution of the variable within controls

*P interaction = 0.95

Figure 1. Associations of Genetic Risk Score with CRC, the EPIC study, 1992-2003.



Appendix

Supplemental Table 1: Parameter criteria for the WCRF/AICR score construction (adapted from Romeguera et al), EPIC study, 1992-2003.

WCRF/AICR Recommendation	Personal recommendations	Operationalisation	Scoring
1 - BODY FATNESS Be as lean as possible without becoming underweight	1a - Ensure that body weight through childhood and adolescent growth projects towards the lower end of the normal BMI range at age 21	*Insufficient data available*	**n.a.**
	1b - Maintain body weight within the normal range from age 21	BMI 18.5 – 24.9 kg/m^2	**1**
		BMI 25 – 29.9 kg/m^2	**0.5**
		BMI<18.5 or BMI >30 kg/m^2	**0**
	1c - Avoid weight gain and increases in waist circumference throughout adulthood	*Insufficient data available*	**n.a.**
2 - PHYSICAL ACTIVITY[1] Be physically active as part of your everyday life	2a - Be moderately physically active, equivalent to brisk walking, for at least 30 minutes every day	Manual/heavy manual job, or >2 h/w of vigorous PA, or >30 min/d of cycling/sports[2]	**1**
		15 – 30 min/d of cycling/sports	**0.5**
		<15 min/d of cycling/sports	**0**
	2b - As fitness improves, aim for 60 minutes or more of moderate or for 30 minutes or more of vigorous, physical activity every day	*Insufficient data available*	**n.a.**
	2c - Limit sedentary habits such as watching television	*Insufficient data available*	**n.a.**
3 - FOODS AND DRINKS THAT PROMOTE WEIGHT GAIN[2,3] Limit consumption of energy-dense foods; avoid sugary drinks	3a - Consume energy-dense foods sparingly	**ED**[4] ≤125 kcal/100g/d	**1**
		ED >125 - <175kcal/100g/d	**0.5**
		ED >175 kcal/100g/d	**0**
	3b - Avoid sugary drinks	Sugary drinks intake[5] = 0 g/d	**1**

		Sugary drinks intake ≤250 g/d	**0.5**
		Sugary drinks intake>250 g/d	**0**
	3c - Consume fast foods sparingly, if at all	*Insufficient data available*	**n.a.**
4 - PLANT FOODS[2,3] Eat mostly foods of plant origin	4a - Eat at least five portions / servings (at least 400 g) of a variety of non-starchy vegetables and of fruits every day	F&V intake ≥400 g/d	**1**
		F&V intake 200 - <400 g/d	**0.5**
		F&V intake <200 g/d	**0**
	4b - Eat relatively unprocessed cereals (grains) and/or pulses (legumes) with every meal	Dietary fibre intake ≥25 g/d	**1**
		Dietary fibre intake 12.5- <25 g/d	**0.5**
		Dietary fibre intake <12.5 g/d	**0**
	4c - Limit refined starchy foods	*Insufficient data available*	**n.a.**
	4d - People who consume starchy roots or tubers as staples should also ensure sufficient intake of non-starchy vegetables, fruits, and pulses (legumes)	*Not applicable to this population*	**n.a.**
5 - ANIMAL FOODS[3] Limit intake of red meat and avoid processed meat	5a - People who eat red meat to consume less than 500 g a week, very little if any to be processed	Red and processed meat <500 g/w and processed meat intake <3 g/d	**1**
		Red and processed meat <500 g/w and processed meat intake 3-<50 g/d	**0.5**
		Red and processed meat ≥500 g/w or processed meat intake ≥50 g/d	**0**
6 - ALCOHOLIC DRINKS Limit alcoholic drinks	6a - If alcoholic drinks are consumed, limit consumption to no more than two drinks a day for men and one drink a day for women	Ethanol intake ≤20 g/d (♂) Ethanol intake ≤10 g/d (♀)	**1**
		Ethanol intake >20-30 g/d (♂) Ethanol intake >10-20 g/d (♀)	**0.5**

Supplementary Table 2. Individual SNP associations with CRC risk using unconditional logistic regression with adjustments for age at recruitment, sex, and center, EPIC Study, 1992-2003.

			Colorectal Cancer[+]			
	SNP	Genotype	Case	Control	OR (95% CI)	P
	rs10411210					
Model 1	0	CC	1146	1137	1 (ref)	
T allele is minor allele	1	CT	255	265	0.95 (0.79-1.15)	
	2	TT	18	16	1.11 (0.56-2.20)	0.831
Model 2		Additive*(CC = 0, CT = 1, TT = 2)	1419	1418	0.97 (0.82-1.16)	0.499
Model 3		Dominant(CT + TT vs. CC)	1419	1418	0.96 (0.8-1.16)	0.676
Model 4		Recessive(TT vs. CC + CT)	1419	1418	1.12 (0.57-2.22)	0.734
	rs10795668					
Model 1	0	GG	584	588	1 (ref)	
A allele is minor allele	1	GA	530	511	1.05 (0.89-1.24)	
	2	AA	128	151	0.86 (0.66-1.12)	0.331
Model 2		Additive*(GG = 0, GA = 1, AA = 2)	1242	1250	0.97 (0.86-1.09)	0.499
Model 3		Dominant(GA + AA vs. GG)	1242	1250	1.01 (0.86-1.18)	0.935
Model 4		Recessive(AA vs. GG + GA)	1242	1250	0.84 (0.65-1.08)	0.170
	rs11986063					
Model 1	0	CC	1160	1143	1 (ref)	
T allele is minor allele	1	CT	234	266	0.86 (0.71-1.05)	
	2	TT	14	7	1.93 (0.77-4.81)	0.116
Model 2		Additive*(CC = 0, CT = 1, TT = 2)	1408	1416	0.93 (0.78-1.11)	0.499
Model 3		Dominant(CT + TT vs. CC)	1408	1416	0.89 (0.74-1.08)	0.242
Model 4		Recessive(TT vs. CC + CT)	1408	1416	1.98 (0.8-4.94)	0.142
	rs1525461					
Model 1	0	TT	886	893	1 (ref)	
C allele is minor allele	1	TC	451	417	1.09 (0.92-1.28)	
	2	CC	56	65	0.87 (0.6-1.26)	0.404
Model 2		Additive*(TT = 0, TC = 1, CC = 2)	1393	1375	1.02 (0.89-1.16)	0.499
Model 3		Dominant(TC + CC vs. TT)	1393	1375	1.06 (0.9-1.24)	0.481
Model 4		Recessive(CC vs. TT + TC)	1393	1375	0.85 (0.59-1.22)	0.370
	rs16888522					
Model 1	0	CC	1228	1224	1 (ref)	
T allele is minor allele	1	CT	180	191	0.94 (0.76-1.17)	
	2	TT	6	1	6.03 (0.72-50.43)	0.214
Model 2		Additive*(CC = 0, CT = 1, TT = 2)	1414	1416	1.00 (0.81-1.23)	0.499
Model 3		Dominant(CT + TT vs. CC)	1414	1416	0.97 (0.78-1.20)	0.756
Model 4		Recessive(TT vs. CC + CT)	1414	1416	6.08 (0.73-50.84)	0.096
	rs16888589					
Model 1	0	AA	1159	1125	1 (ref)	
G allele is minor allele	1	AG	221	247	0.87 (0.71-1.06)	
	2	GG	8	8	0.97 (0.36-2.59)	0.382
Model 2		Additive*(AA = 0, AG = 1, GG = 2)	1388	1380	0.88 (0.73-1.06)	0.499
Model 3		Dominant(AG + GG vs. AA)	1388	1380	0.87 (0.72-1.06)	0.170
Model 4		Recessive(GG vs. AA + AG)	1388	1380	0.99 (0.37-2.65)	0.981
	rs16892766					
Model 1	0	AA	1191	1172	1 (ref)	
C allele is minor allele	1	AC	215	237	0.89 (0.73-1.09)	
	2	CC	11	10	1.07 (0.45-2.54)	0.537
Model 2		Additive*(AA = 0, AC = 1, CC = 2)	1417	1419	0.92 (0.76-1.10)	0.499
Model 3		Dominant(AC + CC vs. AA)	1417	1419	0.90 (0.74-1.10)	0.300
Model 4		Recessive(CC vs. AA + AC)	1417	1419	1.09 (0.46-2.59)	0.839
	rs2059254					
Model 1	0	CC	590	613	1 (ref)	
T allele is minor allele	1	CT	544	551	1.03 (0.87-1.21)	
	2	TT	122	119	1.06 (0.8-1.41)	0.892
Model 2		Additive*(CC = 0, CT = 1, TT = 2)	1256	1283	1.03 (0.91-1.16)	0.499
Model 3		Dominant(CT + TT vs. CC)	1256	1283	1.03 (0.88-1.21)	0.681
Model 4		Recessive(TT vs. CC + CT)	1256	1283	1.05 (0.8-1.37)	0.720

Supplementary Table 2. Individual SNP associations with CRC risk using unconditional logistic regression with adjustments for age at recruitment, sex, and center, EPIC Study, 1992-2003. (Cont.)

	rs2113200					
Model 1	0	TT	717	693	1 (ref)	
A allele is minor allele	1	TA	576	616	0.90 (0.77-1.05)	
	2	AA	118	107	1.06 (0.8-1.41)	0.325
Model 2		Additive*(TT = 0, TA = 1, AA = 2)	1411	1416	0.97 (0.86-1.09)	0.499
Model 3		Dominant(TA + AA vs. TT)	1411	1416	0.93 (0.8-1.07)	0.311
Model 4		Recessive(AA vs. TT + TA)	1411	1416	1.11 (0.85-1.46)	0.443
	rs2373859					
Model 1	0	TT	537	570	1 (ref)	
C allele is minor allele	1	TC	629	626	1.07 (0.91-1.26)	
	2	CC	219	184	1.27 (1.01-1.60)	0.120
Model 2		Additive*(TT = 0, TC = 1, CC = 2)	1385	1380	1.11 (1-1.24)	0.499
Model 3		Dominant(TC + CC vs. TT)	1385	1380	1.12 (0.96-1.30)	0.156
Model 4		Recessive(CC vs. TT + TC)	1385	1380	1.23 (0.99-1.52)	0.060
	rs275454					
Model 1	0	GG	579	602	1 (ref)	
A allele is minor allele	1	GA	652	642	1.06 (0.9-1.24)	
	2	AA	177	172	1.07 (0.84-1.36)	0.750
Model 2		Additive*(GG = 0, GA = 1, AA = 2)	1408	1416	1.04 (0.93-1.16)	0.499
Model 3		Dominant(GA + AA vs. GG)	1408	1416	1.06 (0.91-1.23)	0.454
Model 4		Recessive(AA vs. GG + GA)	1408	1416	1.04 (0.83-1.30)	0.731
	rs2853668					
Model 1	0	GG	779	773	1 (ref)	
T allele is minor allele	1	GT	547	530	1.03 (0.88-1.20)	
	2	TT	77	102	0.75 (0.55-1.03)	0.156
Model 2		Additive*(GG = 0, GT = 1, TT = 2)	1403	1405	0.94 (0.83-1.06)	0.499
Model 3		Dominant(GT + TT vs. GG)	1403	1405	0.98 (0.84-1.14)	0.809
Model 4		Recessive(TT vs. GG + GT)	1403	1405	0.74 (0.55-1.01)	0.057
	rs3802842					
Model 1	0	AA	684	773	1 (ref)	
C allele is minor allele	1	AC	583	513	1.29 (1.1-1.50)	
	2	CC	118	116	1.15 (0.87-1.51)	0.007
Model 2		Additive*(AA = 0, AC = 1, CC = 2)	1385	1402	1.16 (1.03-1.30)	0.499
Model 3		Dominant(AC + CC vs. AA)	1385	1402	1.26 (1.09-1.46)	0.002
Model 4		Recessive(CC vs. AA + AC)	1385	1402	1.03 (0.79-1.35)	0.824
	rs4779584					
Model 1	0	CC	541	643	1 (ref)	
T allele is minor allele	1	CT	237	270	1.04 (0.84-1.29)	
	2	TT	48	36	1.60 (1.02-2.51)	0.123
Model 2		Additive*(CC = 0, CT = 1, TT = 2)	826	949	1.14 (0.97-1.34)	0.499
Model 3		Dominant(CT + TT vs. CC)	826	949	1.11 (0.91-1.35)	0.321
Model 4		Recessive(TT vs. CC + CT)	826	949	1.58 (1.01-2.47)	0.044
	rs4813802					
Model 1	0	TT	604	637	1 (ref)	
G allele is minor allele	1	TG	629	611	1.09 (0.93-1.27)	
	2	GG	169	155	1.15 (0.9-1.47)	0.406
Model 2		Additive*(TT = 0, TG = 1, GG = 2)	1402	1403	1.08 (0.97-1.20)	0.499
Model 3		Dominant(TG + GG vs. TT)	1402	1403	1.1 (0.95-1.28)	0.207
Model 4		Recessive(GG vs. TT + TG)	1402	1403	1.1 (0.88-1.39)	0.402
	rs4925386					
Model 1	0	CC	723	685	1 (ref)	
T allele is minor allele	1	CT	526	583	0.86 (0.73-1.00)	
	2	TT	127	111	1.09 (0.83-1.44)	0.081
Model 2		Additive*(CC = 0, CT = 1, TT = 2)	1376	1379	0.96 (0.86-1.08)	0.499
Model 3		Dominant(CT + TT vs. CC)	1376	1379	0.89 (0.77-1.04)	0.138
Model 4		Recessive(TT vs. CC + CT)	1376	1379	1.17 (0.89-1.53)	0.256

Supplementary Table 2. Individual SNP associations with CRC risk using unconditional logistic regression with adjustments for age at recruitment, sex, and center, EPIC Study, 1992-2003. (Cont.)

	rs4939827					
Model 1	0	TT	433	378	1 (ref)	
C allele is minor allele	1	CT	664	634	0.92 (0.77-1.09)	
	2	CC	248	357	0.61 (0.49-0.75)	<.0001
Model 2		Additive*(TT = 0, CT = 1, CC = 2)	1345	1369	0.79 (0.71-0.88)	0.499
Model 3		Dominant(CT + CC vs. TT)	1345	1369	0.81 (0.68-0.95)	0.010
Model 4		Recessive(CC vs. TT + CT)	1345	1369	0.64 (0.53-0.77)	<.0001
	rs6983267					
Model 1	0	TT	289	343	1 (ref)	
G allele is minor allele	1	GT	647	687	1.14 (0.95-1.38)	
	2	GG	428	363	1.41 (1.14-1.74)	0.005
Model 2		Additive*(TT = 0, GT = 1, GG = 2)	1364	1393	1.19 (1.07-1.32)	0.499
Model 3		Dominant(GT + GG vs. TT)	1364	1393	1.23 (1.03-1.47)	0.023
Model 4		Recessive(GG vs. TT + GT)	1364	1393	1.28 (1.09-1.51)	0.003
	rs7315438					
Model 1	0	TT	489	494	1 (ref)	
C allele is minor allele	1	TC	662	658	1.01 (0.86-1.20)	
	2	CC	251	259	0.97 (0.78-1.21)	0.930
Model 2		Additive*(TT = 0, TC = 1, CC = 2)	1402	1411	0.99 (0.89-1.10)	0.499
Model 3		Dominant(TC + CC vs. TT)	1402	1411	1.00 (0.86-1.17)	0.985
Model 4		Recessive(CC vs. TT + TC)	1402	1411	0.97 (0.8-1.17)	0.725
	rs7837208					
Model 1	0	GG	1148	1147	1 (ref)	
A allele is minor allele	1	GA	250	257	0.97 (0.8-1.18)	
	2	AA	19	12	1.56 (0.75-3.24)	0.455
Model 2		Additive*(GG = 0, GA = 1, AA = 2)	1417	1416	1.02 (0.86-1.22)	0.499
Model 3		Dominant(GA + AA vs. GG)	1417	1416	1.00 (0.83-1.20)	0.983
Model 4		Recessive(AA vs. GG + GA)	1417	1416	1.57 (0.76-3.26)	0.223
	rs8056538					
Model 1	0	GG	686	683	1 (ref)	
A allele is minor allele	1	GA	552	589	0.93 (0.80-1.09)	
	2	AA	107	96	1.1 (0.82-1.48)	0.467
Model 2		Additive*(GG = 0, GA = 1, AA = 2)	1345	1368	1.00 (0.88-1.12)	0.499
Model 3		Dominant(GA + AA vs. GG)	1345	1368	0.96 (0.82-1.11)	0.569
Model 4		Recessive(AA vs. GG + GA)	1345	1368	1.14 (0.85-1.52)	0.375
	rs961253					
Model 1	0	CC	549	581	1 (ref)	
A allele is minor allele	1	CA	640	633	1.07 (0.91-1.26)	
	2	AA	193	163	1.26 (0.99-1.6)	0.167
Model 2		Additive*(CC = 0, CA = 1, AA = 2)	1382	1377	1.11 (0.99-1.24)	0.499
Model 3		Dominant(CA + AA vs. CC)	1382	1377	1.11 (0.95-1.29)	0.181
Model 4		Recessive(AA vs. CC + CA)	1382	1377	1.21 (0.97-1.52)	0.090
	rs9925923					
Model 1	0	CC	720	697	1 (ref)	
T allele is minor allele	1	CT	578	609	0.92 (0.79-1.07)	
	2	TT	117	109	1.03 (0.78-1.37)	0.483
Model 2		Additive*(CC = 0, CT = 1, TT = 2)	1415	1415	0.97 (0.87-1.09)	0.499
Model 3		Dominant(CT + TT vs. CC)	1415	1415	0.94 (0.81-1.08)	0.376
Model 4		Recessive(TT vs. CC + CT)	1415	1415	1.08 (0.82-1.41)	0.598
	rs9929218					
Model 1	0	GG	710	692	1 (ref)	
A allele is minor allele	1	GA	579	610	0.92 (0.79-1.08)	
	2	AA	121	110	1.07 (0.8-1.41)	0.465
Model 2		Additive*(GG = 0, GA = 1, AA = 2)	1410	1412	0.98 (0.88-1.11)	0.499
Model 3		Dominant(GA + AA vs. GG)	1410	1412	0.95 (0.81-1.1)	0.455
Model 4		Recessive(AA vs. GG + GA)	1410	1412	1.11 (0.84-1.45)	0.468

* Additive models impose a structure in which each additional copy of the variant allele increases the response(log odds ratio) by the same amount.