HHS Public Access

Author manuscript

Cancer Epidemiol Biomarkers Prev. Author manuscript; available in PMC 2023 September 06.

Published in final edited form as:

Cancer Epidemiol Biomarkers Prev. 2023 March 06; 32(3): 315–328.

doi:10.1158/1055-9965.EPI-22-0763.

Genome-wide interaction study with smoking for colorectal cancer risk identifies novel genetic loci related to tumor suppression, inflammation and immune response

A full list of authors and affiliations appears at the end of the article.

Abstract

Background: Tobacco smoking is an established risk factor for colorectal cancer (CRC). However, genetically-defined population subgroups may have increased susceptibility to smoking-related effects on CRC.

Methods: A genome-wide interaction scan was performed including 33,756 CRC cases and 44,346 controls from three genetic consortia.

Results: Evidence of an interaction was observed between smoking status (ever vs never smokers) and a locus on 3p12.1 (rs9880919, $p=4.58x10^{-8}$), with higher associated risk in subjects carrying the GG genotype (OR 1.25, 95%CI 1.20-1.30) compared with the other genotypes (OR <1.17 for GA and AA). Among ever smokers, we observed interactions between smoking intensity (increase in 10 cigarettes smoked per day) and two loci on 6p21.33 (rs4151657, $p=1.72x10^{-8}$) and 8q24.23 (rs7005722, $p=2.88x10^{-8}$). Subjects carrying the rs4151657 TT genotype showed higher risk (OR 1.12, 95%CI 1.09-1.16) compared with the other genotypes (OR <1.06 for TC and CC). Similarly, higher risk was observed among subjects carrying the rs7005722 AA genotype (OR 1.17, 95%CI 1.07-1.28) compared with the other genotypes (OR <1.13 for AC and CC). Functional annotation revealed that SNPs in 3p12.1 and 6p21.33 loci were located in regulatory regions, and were associated with expression levels of nearby genes. Genetic models predicting gene expression revealed that smoking parameters were associated with lower CRC risk with higher expression levels of *CADM2* (3p12.1) and *ATF6B* (6p21.33).

Conclusions: Our study identified novel genetic loci that may modulate the risk for CRC of smoking status and intensity, linked to tumor suppression and immune response.

Impact: These findings can guide potential prevention treatments.

Correspondence to: Prof. Dr. W. James Gauderman, Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, SSB 202G 1845 N. Soto Street, Health Sciences Campus, 90032, Los Angeles, CA, USA, jimg@usc.edu; **Prof. Dr. Victor Moreno;** Oncology Data Analytics Program, Catalan Institute of Oncology 08908, L'Hospitalet de Llobregat, Barcelona, Spain., Tel.: +34 93 260 77 75, v.moreno@iconcologia.net; **Prof. Dr. Ulrike Peter** Public Health Sciences Division, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. N., 19024, Seattle, WA, USA, Tel.: +1 206 667 2450, Fax: +1 206 667 7850, upeters@fredhutch.org.

[=]These authors contributed equally to this work

The authors declare no potential conflicts of interest.

Introduction

Colorectal cancer (CRC) is one of the most commonly diagnosed malignancies globally, being the 2nd and 3rd most frequent cancer among women and men, respectively (1).

Germ-line genetics play an important role in the etiology of this complex disease. Genome-wide association studies (GWAS) to date have identified approximately 150 independent loci associated with CRC risk (2-6). However, the identified GWAS risk loci explain less than 20% of the heritability of CRC, leaving a large fraction unexplained (7). Gene-environment (GxE) interactions have been postulated to identify novel genetic risk loci and to explain some of this missing heritability (8).

Among modifiable cancer risk factors, tobacco smoking is the single strongest cause of cancer worldwide. It has been observed that between 8 to 35% of cancer deaths are attributable to smoking in the United States (9); while recent cohort and literature-based studies of wide-world populations estimated that around 10% of CRC cases can be attributable to cigarette smoking (10,11). There is sufficient evidence to consider cigarette smoking as a causal risk factor for CRC, as reported in the last Monograph of the International Agency for Research on Cancer (IARC) (12). Ever smokers showed a risk increase of 18-20% for CRC as compared with never smokers. This risk was higher for rectal cancer, and for CRC tumors with CpG island methylator phenotype (CIMP), microsatellite instability (MSI), and presence of *BRAF* somatic mutations, features often seen in tumors derived through the serrated pathway (13,14). Additionally, CRC risk increases with smoking intensity and duration in a dose-dependent manner (14,15). However, a recent study using an instrumental approach found no causal association between ever having smoked regularly and CRC risk; while positive association between lifetime amount of smoking and CRC risk was observed (16). The dichotomization of the continuous phenotype (amount of smoking) to the binary exposure (ever having smoked) was done at 100 cigarettes over the course of life. In an instrumental setting, the causal estimates for a binary exposure assume that the causal effect is a stepwise function at the point of dichotomization (17), which could not reflect the relationship between smoking and CRC risk.

Tobacco use has been observed to interact with germ-line genetics, providing novel risk loci for some cancers, such as lung (18), pancreatic (19) and bladder cancers (20). Furthermore, our group estimated that a proportion of CRC risk heritability is explained by interactions between smoking and common genetic variants (21). However, our previous genome-wide interaction study (GWIS) of smoking and CRC risk did not identify any interaction that reached genome-wide significance (22). At that time, the study included 11,219 cases and 11,382 controls, and the null results were possibly due to the limited sample size.

The standard interaction test generally needs a sample size at least 4 times larger than that required to detect a main effect of comparable magnitude (23). However, novel statistical approaches enable an increase in power to detect GxE interactions. For instance, joint tests can detect novel susceptibility loci by accounting for both main and interaction effects,

while two-step methods prioritize SNPs to decrease multiple testing burden by applying a filtering step prior to GxE testing (24).

In this study, we applied a set of interaction tests in a combined large dataset from three CRC consortia to identify novel genetic risk loci among: a) nearly 35,000 cases and 45,000 controls with smoking status data, and b) over 13,000 cases and 16,000 controls among current or former smokers with smoking intensity and duration data. Significant results were stratified by colon subsite (proximal colon, distal colon and rectum) and, in a subset of cases, by tumor molecular markers including CIMP, MSI, and *BRAF* and *KRAS* mutations. Additionally, significant lead and correlated genetic variants were evaluated for chromatin accessibility and associations with gene expression of nearby genes. For the identified genes, genetic models predicting expression were developed, and interaction between predicted gene expression and smoking was tested on the CRC consortia datasets.

Materials and Methods.

Study participants and smoking habits

A total of 34 studies (cohort and case-control) comprising individuals of European ancestry were included in this study from three CRC genetic consortia: the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO), the Colorectal Cancer Transdisciplinary Study (CORECT) and the Colon Cancer Family Registry (CCFR) (2). For cohort studies, nested case-control sets were assembled via risk-set sampling, while population-based controls were used for case-control studies. Controls were matched on age, sex, race, and enrollment date/trial group, when applicable. Cases were defined as colorectal adenocarcinoma or advanced adenomas, and were confirmed by medical records, pathological reports, or death certificate information. For the small subset of advanced adenoma cases, matched controls displayed polyp-free sigmoidoscopy or colonoscopy at the time of adenoma selection. All participants gave written informed consent and studies were approved by their respective Institutional Review Boards.

Lifestyle and environmental risk factors were collected by telephone or in-person interviews and structured self-administrated questionnaires. Harmonized quality-control checks were performed in each study following similar criteria (25,26). Participants with any available present or past smoking habits comprised a total of 33,756 CRC cases (14,975 never vs 18,781 ever cigarette smokers) and 44,346 controls (21,976 never vs 22,370 ever cigarette smokers). Among ever smokers, 13,320 cases and 16,176 controls also had indicators of smoking intensity (cigarettes smoked per day), and 12,531 cases and 15,271 controls had data on duration of cigarette smoking (packs of cigarettes smoked in years of habit; pack-years) (Supplementary Table S1). For the analysis of ever versus never smoking variable, we excluded 184 participants from the Alpha-Tocopherol, Beta Carotene Cancer Prevention Study (ATBC) since study participants only included ever smokers. Never smokers were excluded from analyses of smoking intensity and duration. To improve the interpretation of the results, smoking intensity and duration parameters were rescaled to units of 10 cigarettes smoked per day and 20 pack-years, respectively, in order to approximate one unit of the rescaled measures to one standard deviation of the raw measures.

Genotyping and imputation

Details on quality control and genotyping were previously published (2,27), and the genotyping arrays used are summarized in Table S1. Briefly, exclusion criteria included: single nucleotide polymorphisms (SNPs) with missing call rate >2-5%, departure from Hardy-Weinberg equilibrium (HWE) ($p<1x10^{-4}$), inconsistencies between self-reported and genotypic sex, and discordant genotype calls within duplicate samples. Genotypes were imputed to the Haplotype Reference Consortium (HRC) panel (28) (39.1 million variants) using the University of Michigan Imputation Server (29), and converted into a binary format for data management and analyses using the *BinaryDosage* R package (https://cran.r-project.org/web/packages/BinaryDosage). Imputed SNPs were restricted based on a pooled minor allele frequency (MAF) ≥1% and imputation accuracy ($R^2>0.8$). After imputation and quality control, a total of over 7.2 million SNPs were selected for subsequent analyses. Principal component analysis (PCA) for population stratification assessment was performed using PLINK1.9 on 30,000 randomly sampled imputed SNPs with MAF>5% and $R^2>0.99$ of imputation quality score. We observed that the first three PCs explained the 72% of the variation and that the additional PCs explained a small and decreasing amount of the variation.

Statistical analyses

The association of smoking exposure variables with CRC risk was assessed by meta-analysis of study-specific estimates, adjusted by sex, age, study, and three principal genetic components to account for population stratification. Between-study heterogeneity was investigated using the Chi-squared test, and inconsistency was measured using the I^2 statistic, which represents the proportion of total variation attributable to between-study variance. No outlier studies were identified, by estimating the posterior probability (>0.99) of having outliers based on mixture random effects using the "outlierProbs" function of *metaplus* R package. We also assessed smoking habits with CRC risk stratified by sex and by tumor site. Between-sex and between-site heterogeneity were investigated using the chi-square test, and inconsistency was measured using the I^2 statistic.

To identify novel CRC interaction loci, we performed genome-wide scans using the *GxEScanR* R package (https://cran.r-project.org/web/packages/GxEScanR), which implements several interaction testing methods. Imputed allelic dosages were modelled as continuous variables. All the analyses were adjusted for sex, age, study, and three principal genetic components to account for population stratification. These covariates did not lead to any missing values. No single GxE test is universally most powerful and with that in mind, we applied three different testing procedures to maximize the chance of discovering novel loci related to CRC. These include the standard one-degree-of-freedom (1df) test of GxE interaction, a 3df joint test, and an efficient two-step test.

All three of the abovementioned tests utilize a logistic regression model of the form: $logit(Pr(D = 1|G)) = \beta_0 + \beta_G G + \beta_E E + \beta_{GxE} GxE + \beta_C C$, where D is CRC, G is a particular SNP, E is smoking, and C are the adjustment covariates. The standard 1df GxE test is based on a likelihood ratio test of the null hypothesis $H0$: $\beta_{GxE} = 0$. The 3df test (30) extends the well-known 2df joint test (31) of $H0$: $\beta_G = \beta_{GxE} = 0$ to test $H0$: $\beta_G = \beta_{GxE} = \delta_G$

= 0, where δ_G represents the association between G and E in the combined case-control sample (32). The component sources of information for the 3df test have been shown to be statistically independent (33), thus guaranteeing that the overall Type I error rate for this test is preserved. The two-step procedure weights GxE interaction tests (step 2, based on β_{GxE}) using ranks of a filtering statistic (step 1, based on β_G and δ_G) (34) under the weighted hypothesis testing framework (35,36) (Supplementary Fig. S1). The two-step procedure can decrease multiple testing burden and improve power to detect interaction loci. A full description of this two-step methodology can be found in Supplementary data.

Only SNPs not previously associated with CRC risk or smoking phenotypes, and not in linkage disequilibrium (LD) with variants associated with these phenotypes were considered. The likelihood of confounding is substantially smaller when testing for interactions as a confounder needs to be correlated with both the main effects and the interaction (37). However, since it has been observed that obesity is causally influencing smoking habits using an instrumental setting (38), SNPs that reached genome-wide significance ($5x10^{-8}$) in the scans were further adjusted by body-mass index. In addition, significant SNPs were further assessed by stratifying by sex, and cases by tumor location (proximal colon, distal colon and rectum) or tumor molecular markers, when available, including CIMP, MSI or microsatellite stable (MSS), and *BRAF* and *KRAS* somatic mutation status.

Functional follow up

Regional plots for all significant findings were generated, which enables inspection of strength of association, the extent of association signal and LD, and position of findings relative to genes in the region. Plots were generated using the software LocusZoom v1.3 (39). Measures of LD were estimated using European populations of the 1000 Genomes Project.

The putative functional role of these SNPs and those in LD (R^2>0.5) at 500 kb flanking regions were investigated with relation to their potential contribution to regulate gene expression in two ways: first, by their physical location in regions of chromatin accessibility or histone modifications, and second, through their direct association with expression of nearby genes (expression quantitative trait loci: eQTLs). To evaluate the physical location we obtained regions containing active enhancer elements in tissue from healthy colon and from tumor colon tissue samples from previously analysed ATAC-seq, DNaseI Hypersensitivity (DHS)-seq, and H3K27ac histone ChIP-seq datasets (40). To identify eQTLs, we used the following databases: colon transverse tissue sample from GTEx v8 dataset (41), and Colon Transcriptome Explorer (CoTrEx 2.0; https://barcuvaseq.org/cotrex/, accessed on May 2021) of the University of Barcelona and University of Virginia genotyping and RNA sequencing (BarcUVa-Seq) project dataset, which is comprised of 445 epithelium-enriched healthy colon biopsies from ascending, transverse and descending colon (42). We reported on all genes for which expression was associated with the lead and correlated SNPs. In addition, for loci at which gene expression was associated with lead and correlated SNPs, eQTL models predicting gene expression were developed through an elastic net regularized regression (43) using SNPs with MAF>0.01 and imputation R^2>0.7

of the BarcUVa-Seq project dataset, and validated in the colon transverse tissue sample from GTEx v8 dataset (34). These predicted gene expression levels were converted into quartiles and tested for interaction with smoking on CRC risk using the previously described logistic regression models.

Data availability

The summary statistics data that support the findings of this study are available on request from the corresponding authors.

Results

Smoking habits and CRC risk

Ever smoking was more prevalent among CRC cases (56%) than controls (50%, $p<0.001$) (Supplementary Table S2). Among ever smokers, CRC cases were heavier smokers than controls (18.6 ±11.5 vs 18.0 ±11.4 cigarettes smoked per day $p<0.001$; and 26.8 ±22.4 vs 24.6 ±21.6 pack-years, $p<0.001$) (Supplementary Table S1).

We observed that ever being a cigarette smoker was associated with increased CRC risk (odds ratio (OR) 1.25, 95% confidence interval (CI) 1.20-1.30; $p_{het}=0.07$; $I^2=23\%$) (Fig. 1). Sensitivity analyses showed a slightly lower estimate for cohort-based studies (OR 1.22, 95%CI 1.16-1.29) compared with case-control studies (OR 1.28, 95%CI 1.19-1.37). No evidence for heterogeneity was observed among cohort-based studies ($p_{het}=0.33$; $I^2=9\%$), while results among case-control studies showed moderate heterogeneity ($p_{het}=0.05$; $I^2=34\%$) (Supplementary Table S3). Analyses stratified by sex showed higher risk for men (OR 1.32, 95%CI 1.24-1.40) than for women (OR 1.19, 95%CI 1.12-1.26) ($p_{sex\ difference}=0.01$) (Fig. 1). The association between smoking and CRC varied by tumor site, with the strongest association observed in rectum (OR 1.34, 95%CI 1.24-1.45), compared with distal colon (OR 1.23, 95%CI 1.15-1.31) and proximal colon (OR 1.18, 95%CI 1.13-1.24) ($p_{site\ difference}=0.03$) (Fig. 1). Among smokers, smoking intensity and duration parameters also showed a positive association with CRC risk (OR 1.06, 95%CI 1.03-1.09, per an increase in 10 cigarettes smoked per day; and OR 1.11, 95%CI 1.08-1.15, per an increase in 20 pack-years). Similar patterns for smoking intensity and duration were observed in analyses stratified by sex and tumor site (Fig. 1). Given that these results showed CRC-smoking associations overall and across all subsets, we focused our genome-wide GxE testing on the overall study population to maximize power. However, to explore variation in GxE effect estimates, we also conducted stratified analyses for novel findings.

Genome-wide smoking-interaction scans for CRC risk

Genomic control inflation and quantile-quantile (QQ) plots for the SNP-smoking interactions for risk of CRC at the genome-wide level did not show evidence for residual population stratification (Supplementary Fig. S2).

The initial conventional test did not identify any genome-wide significant loci interacting with smoking status on the risk of CRC (Fig. 2). However, based on the 3-df test, we identified a novel susceptibility locus at chromosome 3p12.1 ($p<=4.58x10^{-8}$) (Table 1).

The most significant SNP in this locus is rs9880919, located in an intron of cell adhesion molecule 2 (*CADM2,* 3p12.1) gene (Supplementary Table S4; Supplementary Fig. S3). This SNP showed evidence of an interaction with smoking status on CRC risk, as well as direct associations with both CRC and smoking status (Table 1). When stratified by genotype for this SNP, smoking status showed higher risk in those carrying the GG genotype (OR 1.25, 95%CI 1.20-1.30, $p=4.80x10^{-27}$) compared with those carrying the GA genotype (OR 1.17, 95%CI 1.11-1.23, $p=2.40x10^{-9}$), and the AA genotype (OR 1.12, 95%CI 0.99-1.27, p=0.08) (Fig. 3; Supplementary Table S5). We observed similar interaction effects when analyses were stratified for study type, sex or tumor site (Supplementary Table S4). The two-step approach did not identify any significant interactions (Supplementary Fig. S4).

Among smokers, we identified two genome-wide significant loci interacting with smoking intensity (cigarettes smoked per day) on the risk of CRC (Fig. 2). The first locus was on chromosome 6p21.33, with rs4151657 being the most significant interacting SNP (1df GxE $p=1.72x10^{-8}$, 3df test $p=3.52x10^{-8}$, Table 1) located in the intron of complement factor B (*CFB*) gene (Supplementary Table S4; Supplementary Fig. S5). When risk for CRC of smoking intensity was stratified by rs4151657 genotype, we observed that the association between smoking intensity and CRC risk was stronger in those carrying the TT genotype (OR 1.12, 95%CI 1.09-1.16, $p=1.30x10^{-12}$, per an increase in 10 cigarettes smoked per day) compared with those carrying the TC genotype (OR 1.06, 95%CI 1.03-1.10, $p=1.10x10^{-4}$) or the CC genotype (OR 0.94, 95%CI 0.89-0.99, p=0.03) (Fig. 3; Supplementary Table S6). The second locus was on chromosome 8q24.23 with rs7005722 being the most significant interaction SNP (1df GxE $p=2.88x10^{-8}$, Table 1) located in an intergenic region between an uncharacterized non-coding RNA *LOC101927915* and family with sequence similarity 135 member B (*FAM135B*) gene (Supplementary Table S4 and Supplementary Fig. S5). When stratified by genotype, we observed that smoking intensity was only associated with CRC risk in those with genotypes carrying the A allele (OR 1.17, 95%CI 1.07-1.28, $p=6.70x10^{-5}$; and OR 1.13, 95%CI 1.10-1.16, $p=3.40x10^{-13}$; among individuals with AA and AC genotypes, respectively, for each 10 extra cigarettes smoked per day) but not in those carrying CC genotype (OR 1.01, 95%CI 0.98-1.04, p=0.48) (Fig. 3; Supplementary Table S6). Similar interaction patterns were observed when analyses were stratified by study type, sex or tumor colon site for both rs4151657 and rs7005722 (Supplementary Table S4). We did not identify novel susceptibility loci interacting with cigarettes per day using 3-df joint test or two-step test (Supplementary Fig. S4) and did not observe interactions with pack-years by any method (Fig. 2, and Supplementary Fig. S4).

Further adjustment of interaction scans with body-mass index did not modify observed interactions between genetics and smoking parameters on CRC risk (Supplementary Table S7).

SNP-smoking interactions by CRC tumor molecular markers

To investigate if the above observed interactions differed by tumor subtypes defined by CIMP, MSI, *BRAF* and *KRAS* mutation status, we estimated the interaction in a subset with available data on these molecular characteristics (Supplementary Table S8). Similar to results using the whole dataset, the identified SNPs showed some evidence of interaction in

all tumor marker subtypes, with the exception of the SNP in 3p12.1. However, the strength and significance of the interaction were different across the subtypes (Supplementary Table S9). For both rs4151657 (6p21.33) and rs7005722 (8q24.23), the interaction was observed for CIMP-, MSS, *BRAF* non-mutated, and *KRAS* both mutated and non-mutated tumors (Supplementary Table S9).

Functional annotations of the genetic loci

Regarding functional annotation, SNPs correlated with rs9880919 (3p12.1) and rs4151657 (6p21.33) were located in accessible chromatin regions, and H3K27ac and DHS peaks in non-malignant colon samples, and in H3K27ac peaks in colon tumor samples (Table 2). This suggests a potential regulatory role of these genetic regions on transcription of nearby genes in non-malignant and tumor samples. The SNP rs7005722 (8q24.23) was not correlated with other SNPs, nor was it identified as a variant enhancer locus or eQTL of nearby genes (Table 2). Lead and correlated SNPs in the 3p12.1 region were associated with expression levels of *CADM2*, in both BarcUVa-Seq and GTEx colon transverse datasets (Table 2). For the SNPs in the 6p21.33 region, the long LD extension of this region provided association with 22 genes; of which activating transcription factor 6 beta (*ATF6B*) and epidermal growth factor like domain multiple 8 (*EGFL8*) genes were associated in both datasets (BarcUVa-Seq and GTEx colon transverse, Table 2).

Genetic models predicting gene expression were developed for *CADM2* (3p12.1), and *ATF6B* and *EGFL8* (6p21.33) genes (Supplementary Table S10). The lead SNPs for loci 3p12.1 (rs9880919), and 6p21.33 (rs4151657) were not selected in the genetic models, but they were strongly associated with predicted gene expression levels of corresponding nearby genes in the full case-control sample ($p<1x10^{-16}$). In turn, a statistical interaction was observed between expression levels of *CADM2* with smoking status on risk for CRC (p=0.034) (Supplementary Table S11). When stratified by quartiles of *CADM2* expression levels, smoking status showed higher risk in those individuals in the first quartile (OR 1.28, 95%CI 1.20-1.36, $p=3.67x10^{-14}$) compared with those in upper quartiles (4^{th} quartile, OR 1.16, 95%CI 1.09-1.23, $p=3.00x10^{-6}$) (Fig. 4). In addition, an interaction was observed between expression levels of *ATF6B* with smoking intensity (cigarettes smoked per day) on risk for CRC ($p=5x10^{-5}$) (Supplementary Table S10). When stratified by quartiles of *ATF6B* expression levels, we observed that the association between smoking intensity and CRC risk was stronger in those in the first quartile (OR 1.11, 95%CI 1.06-1.16, $p=2.65x10^{-6}$, per an increase in 10 cigarettes smoked per day) compared with those in upper quartiles (4^{th} quartile, OR 0.99, 95%CI 0.95-1.03, p=0.5) (Fig. 4). Finally, the interaction between expression of *EGFL8* and smoking intensity on CRC risk was not significant (p=0.052, Supplementary Table S11).

Discussion

Combining genome-wide genetic data with harmonized smoking habit data across 34 studies enabled us to conduct the largest genome-wide exploration of GxSmoking interactions for CRC risk to date. We discovered one novel CRC locus in our genome-wide GxE scan of smoking status (3p12.1), and two SNPs in our scan of smoking intensity among

smokers. None of the three genetic loci we discovered have been associated with CRC or smoking previously. Additionally, two of the identified GxE interactions (3p12.1 and 6p21.33) showed differences in strength by CRC tumor marker subtypes as defined based on MSI status, CIMP, and *BRAF* and *KRAS* mutation status. In addition, for these two loci, functional annotations indicated that they are located in enhancer regions and are linked to differential gene expression, and the putative gene for each was identified by demonstrating interactions between smoking and predicted expression affecting CRC risk.

Consistent with previous studies, we observed a strong positive association between smoking habits with risk of CRC with similar magnitudes to those previously observed. As expected, results showed higher risk estimates in men compared with women, and in rectal compared with colon cancer. The different associations across colon subsites have been related to tumor molecular markers of the serrated pathway associated with cigarette smoking and with colon subsites in the same fashion (13).

An enhancer locus in the *CADM2* gene was found interacting with smoking status on CRC risk. In addition, genetically predicted *CADM2* gene expression was observed to decrease CRC risk among ever smokers, but not among never smokers. This is the first time that this locus and *CADM2* expression have been associated with CRC risk. However, *CADM2* expression has been reported to reduce risk in prostate cancer, ovarian cancer, liver cancer, kidney cancer, lymphoma, melanoma, and glioma (44-50). These functional data could indicate a tumor suppression role of *CADM2* gene. Other functional roles could be related to this gene. This locus has been associated with several risk-taking traits, where smoking was included among other phenotypes; however risk-taking proneness was found mediating these associations (51,52). *CADM2* risk-taking risk variants were associated with increased *CADM2* expression levels in brain tissues.(53) These reported results could explain both the marginal evidence of association with smoking status in this study, and its interaction with smoking status on CRC risk as tumor suppressor gene.

For smoking intensity, we observed an interaction with rs4151657 at 6p21.33 located in the human leukocyte antigen (HLA) super-locus that encodes for many genes key for immune regulation and cellular processes. Genetic variants in this gene rich region have been linked to several different diseases including CRC. The SNP rs4151657 falls within an enhancer region in the intron of the *CFB* gene. The complement system plays a pivotal role in the innate immune response to pathogens. It also has an important homeostatic role in recognising damaged or altered “self” components (such as apoptotic and necrotic cells) to promote their elimination. Complement factor B amplifies the turnover of C3, the most abundant complement protein, activating the terminal pathway (54,55). Insufficient or excessive complement activation can promote infections, immune-related disease, or tissue damage. The C allele of rs4151657 SNP (which homozygous genotype cancels out the risk effect of smoking intensity on CRC) has been reported to be associated to increased risk for ulcerative colitis (56,57) and IgA nephropathy (58). We did not observe an association between SNPs in this locus and expression of *CFB.* When we expanded our eQTL analysis to nearby genes we observed associations with expression of *ATF6B* and *EGFL8*. In addition, predicted *ATF6B* gene expression was observed to increase CRC risk among light smokers, but not among heavy smokers. The protein encoded by *ATF6B* is a

transcription factor in the unfolded protein response (UPR) pathway during endoplasmic reticulum (ER) stress (59), and genetic variants in this gene region have been associated with levels of complement C4 protein (60) as well as with inflammatory lung diseases (61,62). Downregulation of *EGFL8* expression has been observed in colon tumor tissue (63). It has also been associated with a less favourable immune component in tumor tissue and with poor prognosis for patients with CRC, indicating that *EGFL8* may act as a tumor suppressor (64). In summary, the observed interaction between smoking and genetic variants in the HLA region point to as strong candidate risk factor the differential expression of *ATF6B* gene linked to inflammation and immune function. Interestingly, smoking promotes cellular damage (65,66) and bacterial CRC-related infections (67,68), and cigarette smoking has been associated with higher CRC risk particularly in tumor samples with lower number of infiltrated T cells and macrophages (69,70). Therefore, the identified germline genetic variants related to inflammation and immune response, may modify or reduce the risk of the effects of tobacco smoking on CRC.

Finally, we observed an interaction between smoking intensity and the intergenic region at 8q24.23 upstream of *LOC101927915* and downstream of *FAM135B*. In contrast with the other identified regions, genetic markers of this region did not appear associated with enhancing activity or gene expression of nearby genes. *FAM135B* has been reported to be associated with risk for esophageal squamous cell carcinoma (71). Accordingly, the functional support for this genetic locus interacting with smoking on colorectal cancer risk is limited. In our analysis of GxE interaction incorporating gene expression data from the BarcUVa sample, tested genes (*CADM2, ATF6B* and *EGFL8*) did not include the respective lead SNP discovered in the primary genome-wide GxE scan in the predictive model. However, each lead SNP was strongly associated with predicted gene expression of corresponding nearby genes in our full case-control sample. This suggests that these lead SNPs are likely not causal variants themselves, but rather that they are pointing to genes for which expression is modifying the effect of smoking on CRC risk.

This study generated results using a large dataset from three CRC consortia for gene-cigarette smoking interaction on CRC risk, including analyses stratified by tumor molecular markers, and functional annotations related to gene regulation. However, molecular marker analysis was limited on sample size, and we could not provide robust results on the pathways where the identified potential mechanisms are involved. In the findings for interaction with smoking intensity, genetic variants seem to contribute to CRC risk in the tumor molecular subtypes characteristic of the adenoma pathway. Thus, cigarette smoking would be related to CRC risk mainly in the cases derived from the serrated pathway, but interacting with genetic variants to increase CRC risk in the cases derived from the adenoma pathway. Further analyses are needed to evaluate this dual relationship between cigarette smoking and CRC risk.

Finally, the map of common genetic variants interacting with smoking habits and associated with CRC risk is not still saturated. Novel functional genetic regions can be predicted integrating strong functional data, for instance genomic single cell data, and thereby reducing the multiple comparison. In addition, further efforts are needed to enlarge the sample size of the analysis including wide-world populations, since most common germline

genetic risk loci are have been replicated across racial and ethnic groups (72,73), reducing inequalities in research, and to provide independent datasets to estimate the heritability contribution.

In conclusion, we identified novel genetic loci that modulates the association between cigarette smoking status and intensity and CRC risk. The potential mechanism behind these associations could be linked, in part, to tumor suppression, inflammation and immune response to the effects of tobacco smoking. These findings can guide potential prevention treatments in addition to quitting smoking. Additional functional studies are needed to verify the role of the identified loci interacting with smoking for CRC risk.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Robert Carreras-Torres[1,2,3,=], Andre E Kim[4,=], Yi Lin[5], Virginia Diez-Obrero[1,2,6], Stephanie A Bien[5], Conghui Qu[5], Jun Wang[4], Niki Dimou[7], Elom K Aglago[7], Demetrius Albanes[8], Volker Arndt[9], James W Baurley[10], Sonja I Berndt[8], Stéphane Bézieau[11], D Timothy Bishop[12], Emmanouil Bouras[13], Hermann Brenner[9,14,15], Arif Budiarto[10], Peter T Campbell[16], Graham Casey[17], Andrew T Chan[18], Jenny Chang-Claude[19], Xuechen Chen[9], David V Conti[4], Christopher H Dampier[20], Matthew AM Devall[17], David A Drew[21], Jane C Figueiredo[22], Steven Gallinger[23], Graham G Giles[24], Stephen B Gruber[25], Andrea Gsur[26], Marc J Gunter[7], Tabitha A Harrison[5], Akihisa Hidaka[5], Michael Hoffmeister[9], Jeroen R Huyghe[5], Mark A Jenkins[27], Kristina M Jordahl[5], Eric Kawaguchi[4], Temitope O Keku[28], Anshul Kundaje[29], Loic Le Marchand[30], Juan Pablo Lewinger[4], Li Li[31], Bharuno Mahesworo[10], John L Morrison[4], Neil Murphy[7], Hongmei Nan[32], Rami Nassir[33], Polly A Newcomb[5], Mireia Obón-Santacana[1,2,6], Shuji Ogino[34], Jennifer Ose[35,36], Rish K Pai[37], Julie R Palmer[38], Nikos Papadimitriou[7], Bens Pardamean[10], Anita R Peoples[35], Paul D P Pharoah[39], Elizabeth A Platz[40], Gad Rennert[41], Edward Ruiz-Narvaez[42], Lori C Sakoda[43], Peter C Scacheri[44], Stephanie L Schmit[45,46], Robert E Schoen[47], Anna Shcherbina[48], Martha L Slattery[49], Mariana C Stern[4], Yu-Ru Su[5], Catherine M Tangen[50], Duncan C Thomas[4], Yu Tian[19,51], Konstantinos K Tsilidis[52,53], Cornelia M Ulrich[35,36], Franzel JB van Duijnhoven[54], Bethany Van Guelpen[55,56], Kala Visvanathan[40], Pavel Vodicka[57], Tjeng Wawan Cenggoro[10], Stephanie J Weinstein[8], Emily White[58], Alicja Wolk[59], Michael O Woods[60], Li Hsu[5], Ulrike Peters[5,61], Victor Moreno[1,2,6,62], W James Gauderman[4]

Affiliations

[1]Colorectal Cancer Group, ONCOBELL Program, Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, Barcelona, Spain.

[2]Oncology Data Analytics Program, Catalan Institute of Oncology, L'Hospitalet de Llobregat, Barcelona, Spain.

[3]Digestive Diseases and Microbiota Group, Girona Biomedical Research Institute (IDIBGI), Salt, 17190, Girona, Spain.

[4]Division of Biostatistics, Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, California, USA.

[5]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA.

[6]Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), Madrid, Spain

[7]Nutrition and Metabolism Branch, International Agency for Research on Cancer, Lyon, France.

[8]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA.

[9]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany.

[10]Bioinformatics and Data Science Research Center, Bina Nusantara University, Jakarta, Indonesia.

[11]Service de Génétique Médicale, Centre Hospitalier Universitaire (CHU) Nantes, Nantes, France.

[12]Leeds Institute of Cancer and Pathology, University of Leeds, Leeds, UK.

[13]Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece.

[14]Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany

[15]German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany

[16]Behavioral and Epidemiology Research Group, American Cancer Society, Atlanta, Georgia, USA.

[17]Center for Public Health Genomics, Department of Public Health Sciences, University of Virginia, Charlottesville, Virginia, USA.

[18]Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA.

[19]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany.

[20]Department of General Surgery, University of Virginia School of Medicine, Charlottesville, Virginia, USA.

[21]Clinical & Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA.

[22]Department of Medicine, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA.

[23]Lunenfeld Tanenbaum Research Institute, Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada.

[24]Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, Victoria, Australia.

[25]Department of Medical Oncology & Therapeutics Research, City of Hope National Medical Center, Duarte, CA, USA.

[26]Institute of Cancer Research, Department of Medicine I, Medical University Vienna, Vienna, Austria.

[27]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia.

[28]Center for Gastrointestinal Biology and Disease, University of North Carolina, Chapel Hill, North Carolina, USA.

[29]Department of Genetics, Department of Computer Science, Stanford University, Stanford, California, USA.

[30]University of Hawaii Cancer Center, Honolulu, Hawaii, USA.

[31]Department of Family Medicine, University of Virginia, Charlottesville, Virginia, USA.

[32]Department of Epidemiology, Richard M. Fairbanks School of Public Health, Indianapolis, Indiana, USA.

[33]Department of Pathology, School of Medicine, Umm Al-Qura'a University, Saudi Arabia

[34]Program in MPE Molecular Pathological Epidemiology, Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA; Department of Oncologic Pathology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA; Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA; Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.

[35]Huntsman Cancer Institute, Salt Lake City, Utah, USA

[36]Department of Population Health Sciences, University of Utah, Salt Lake City, Utah, USA.

[37]Department of Laboratory Medicine and Pathology, Mayo Clinic Arizona, Scottsdale, Arizona, USA.

[38]Slone Epidemiology Center at Boston University, Boston, MA, USA.

[39]Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.

[40]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA.

[41]Department of Community Medicine and Epidemiology, Lady Davis Carmel Medical Center, Haifa, Israel.

[42]Department of Nutritional Sciences, University of Michigan School of Public Health, Ann Arbor, Michigan, USA.

[43]Division of Research, Kaiser Permanente Northern California, Oakland, California, USA.

[44]Department of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, Ohio, USA.

[45]Genomic Medicine Institute, Cleveland Clinic, Cleveland, Ohio, USA.

[46]Population and Cancer Prevention Program, Case Comprehensive Cancer Center, Cleveland, Ohio, USA

[47]Department of Medicine and Epidemiology, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania, USA.

[48]Biomedical Informatics Program, Dept. of Biomedical Data Sciences, Stanford University

[49]Department of Internal Medicine, University of Utah, Salt Lake City, Utah, USA.

[50]SWOG Statistical Center, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA.

[51]School of Public Health, Capital Medical University, Beijing, China

[52]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK

[53]Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece.

[54]Division of Human Nutrition and Health, Wageningen University & Research, Wageningen, The Netherlands.

[55]Department of Radiation Sciences, Oncology Unit, Umeå University, Umeå, Sweden.

[56]Wallenberg Centre for Molecular Medicine, Umeå University, Umeå, Sweden.

[57]Department of Molecular Biology of Cancer, Institute of Experimental Medicine of the Czech Academy of Sciences, Prague, and Biomedical Center, Medical Faculty, Pilsen, Czech Republic.

[58]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA.

[59]Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden.

[60]Memorial University of Newfoundland, Discipline of Genetics, St. John's, Canada.

[61]School of Public Health, University of Washington, Seattle, Washington, USA

[62]Department of Clinical Sciences, Faculty of Medicine, University of Barcelona, Barcelona, Spain

Acknowledgements

Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO): This research and all authors were funded through the National Cancer Institute, National Institutes of Health, U.S. Department of Health and Human Services (U01 CA137088, R01 CA059045, U01 CA164930, R21 CA191312, R01201407, R01CA488857). Genotyping/Sequencing services were provided by the Center for Inherited Disease Research (CIDR) contract number HHSN2682017000006I and HHSN268201200008I. This research and all authors were funded in part through the NIH/NCI Cancer Center Support Grant P30 CA015704. Scientific Computing Infrastructure at Fred Hutch funded by ORIP grant S10OD028685.

ASTERISK: This research and all authors were supported by a Hospital Clinical Research Program (PHRC-BRD09/C) from the University Hospital Center of Nantes (CHU de Nantes) and by the Regional Council of Pays de la Loire, the Groupement des Entreprises Françaises dans la Lutte contre le Cancer (GEFLUC), the Association Anne de Bretagne Génétique and the Ligue Régionale Contre le Cancer (LRCC). We are very grateful to Dr. Bruno Buecher without whom this project would not have existed. We also thank all those who agreed to participate in this study, including the patients and the healthy control persons, as well as all the physicians, technicians and students.

The ATBC Study and all authors were supported by the Intramural Research Program of the U.S. National Cancer Institute, National Institutes of Health, Department of Health and Human Services.

The Colon Cancer Family Registry (CCFR, www.coloncfr.org) and all authors were supported in part by funding from the National Cancer Institute (NCI), National Institutes of Health (NIH) (award U01 CA167551). Support for case ascertainment was provided in part from the Surveillance, Epidemiology, and End Results (SEER) Program and the following U.S. state cancer registries: AZ, CO, MN, NC, NH; and by the Victoria Cancer Registry (Australia) and Ontario Cancer Registry (Canada). The CCFR Set-1 (Illumina 1M/1M-Duo) and Set-2 (Illumina Omni1-Quad) scans were supported by NIH awards U01 CA122839 and R01 CA143247 (to G. Casey). The CCFR Set-3 (Affymetrix Axiom CORECT Set array) was supported by NIH award U19 CA148107 and R01 CA81488 (to S.B. Gruber). The CCFR Set-4 (Illumina OncoArray 600K SNP array) was supported by NIH award U19 CA148107 (to S.B. Gruber) and by the Center for Inherited Disease Research (CIDR), which is funded by the NIH to the Johns Hopkins University, contract number HHSN268201200008I. Additional funding for the OFCCR/ARCTIC was through award GL201-043 from the Ontario Research Fund (to B.W. Zanke), award 112746 from the Canadian Institutes of Health Research (to T.J. Hudson), through a Cancer Risk Evaluation (CaRE) Program grant from the Canadian Cancer Society (to S. Gallinger), and through generous support from the Ontario Ministry of Research and Innovation. The SFCCR Illumina HumanCytoSNP array was supported in part through NCI/NIH awards U01/U24 CA074794 and R01 CA076366 (to P.A. Newcomb). The content of this manuscript does not necessarily reflect the views or policies of the NCI, NIH or any of the collaborating centers in the Colon Cancer Family Registry (CCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government, any cancer registry, or the CCFR. The Colon CFR graciously thanks the generous contributions of their study participants, dedication of study staff, and the financial support from the U.S. National Cancer Institute, without which this important registry would not exist. The authors would like to thank the study participants and staff of the Seattle Colon Cancer Family Registry and the Hormones and Colon Cancer study (CORE Studies).

CLUE II: This study and all authors were funded by the National Cancer Institute (U01 CA86308, Early Detection Research Network; P30 CA006973), National Institute on Aging (U01 AG18033), and the American Institute for Cancer Research. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US government. Maryland Cancer Registry (MCR) Cancer data was provided by the Maryland Cancer Registry, Center for Cancer Prevention and Control, Maryland Department of Health, with funding from the State of Maryland and the Maryland Cigarette Restitution Fund. The collection and availability of cancer registry data is also supported by the Cooperative Agreement NU58DP006333, funded by the Centers for Disease Control and Prevention. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention or the Department of Health and Human Services. We thank the participants of Clue II and appreciate the continued efforts of the staff

at the Johns Hopkins George W. Comstock Center for Public Health Research and Prevention in the conduct of the Clue II Cohort Study.

COLO2&3: This study and all authors were funded by the National Institutes of Health (R01 CA60987).

Colorectal Cancer Transdisciplinary (CORECT) Study: The CORECT Study and all authors were supported by the National Cancer Institute, National Institutes of Health (NCI/NIH), U.S. Department of Health and Human Services (grant numbers U19 CA148107, R01 CA81488, P30 CA014089, R01 CA197350; P01 CA196569; R01 CA201407) and National Institutes of Environmental Health Sciences, National Institutes of Health (grant number T32 ES013678).

CORSA: The CORSA study was funded by Austrian Research Funding Agency (FFG) BRIDGE (grant 829675, to A. Gsur), the “Herzfelder’sche Familienstiftung” (grant to A. Gsur) and was supported by COST Action BM1206. We kindly thank all individuals who agreed to participate in the CORSA study. Furthermore, we thank all cooperating physicians and students and the Biobank Graz of the Medical University of Graz.

CPS-II: The American Cancer Society funds the creation, maintenance, and updating of the Cancer Prevention Study-II (CPS-II) cohort. This study was conducted with Institutional Review Board approval. The authors thank the CPS-II participants and Study Management Group for their invaluable contributions to this research. The authors would also like to acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention National Program of Cancer Registries, and cancer registries supported by the National Cancer Institute Surveillance Epidemiology and End Results program.

CRCGEN: Colorectal Cancer Genetics & Genomics, Spanish study and all authors were supported by Instituto de Salud Carlos III, co-funded by FEDER funds –a way to build Europe– (grants PI14-613 and PI09-1286), Agency for Management of University and Research Grants (AGAUR) of the Catalan Government (grant 2017SGR723), and Junta de Castilla y León (grant LE22A10-2). Sample collection of this work was supported by the Xarxa de Bancs de Tumors de Catalunya sponsored by Pla Director d’Oncología de Catalunya (XBTC), Plataforma Biobancos PT13/0010/0013 and ICOBIOBANC, sponsored by the Catalan Institute of Oncology. Spanish Association Against Cancer (AECC) Scientific Foundation grant GCTRA18022MORE.

Czech Republic CCS: This work and all authors were supported by the Grant Agency of the Czech Republic (21-27902S, 20-03997S), by the Grant Agency of the Ministry of Health of the Czech Republic (grants AZV NV18/03/00199 and AZV NV19-09-00237), and Charles University grants Unce/Med/006 and Progress Q28/LF1. We are thankful to all clinicians in major hospitals in the Czech Republic, without whom the study would not be practicable. We are also sincerely grateful to all patients participating in this study.

DACHS: This work and all authors were supported by the German Research Council (BR 1704/6-1, BR 1704/6-3, BR 1704/6-4, CH 117/1-1, HO 5117/2-1, HE 5998/2-1, KL 2354/3-1, RO 2270/8-1 and BR 1704/17-1), the Interdisciplinary Research Program of the National Center for Tumor Diseases (NCT), Germany, and the German Federal Ministry of Education and Research (01KH0404, 01ER0814, 01ER0815, 01ER1505A and 01ER1505B). We thank all participants and cooperating clinicians, and everyone who provided excellent technical assistance.

DALS: This study was supported by the National Institutes of Health (R01 CA48998 to M. L. Slattery).

EDRN: This work and all authors were funded and supported by the NCI, EDRN Grant (U01 CA 84968-06). We acknowledge all contributors to the development of the resource at University of Pittsburgh School of Medicine, Department of Gastroenterology, Department of Pathology, Hepatology and Nutrition and Biomedical Informatics.

EPIC: The coordination of EPIC is financially supported by International Agency for Research on Cancer (IARC) and also by the Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London which has additional infrastructure support provided by the NIHR Imperial Biomedical Research Centre (BRC). The national cohorts and all authors are supported by: Danish Cancer Society (Denmark); Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l’Education Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research Center (DKFZ), German Institute of Human Nutrition Potsdam- Rehbruecke (DIfE), Federal Ministry of Education and Research (BMBF) (Germany); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy, Compagnia di SanPaolo and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands); Health Research Fund (FIS) - Instituto de Salud Carlos III (ISCIII), Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, and the Catalan Institute of Oncology - ICO (Spain); Swedish Cancer Society, Swedish Research Council and Regions of Skåne and Västerbotten (Sweden); Cancer Research UK (14136 to EPIC-Norfolk; C8221/A29017 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk; MR/M012190/1 to EPIC-Oxford). (United Kingdom). Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not

necessarily represent the decisions, policy or views of the International Agency for Research on Cancer/World Health Organization.

ESTHER/VERDI. This work and all authors were supported by grants from the Baden-Württemberg Ministry of Science, Research and Arts and the German Cancer Aid.

Harvard cohorts (HPFS, NHS, PHS): HPFS and all authors are supported by the National Institutes of Health (P01 CA055075, UM1 CA167552, U01 CA167552, R01 CA137178, R01 CA151993, and R35 CA197735), NHS by the National Institutes of Health (R01 CA137178, P01 CA087969, UM1 CA186107, R01 CA151993, and R35 CA197735) and PHS by the National Institutes of Health (R01 CA042182). The study protocol was approved by the institutional review boards of the Brigham and Women's Hospital and Harvard T.H. Chan School of Public Health, and those of participating registries as required. We acknowledge Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital as home of the NHS. We would like to thank the participants and staff of the HPFS, NHS and PHS for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. The authors assume full responsibility for analyses and interpretation of these data.

Hawaii Adenoma Study: This study was supported by the NCI grants R01 CA072520 (to L. Le Marchand).

Kentucky: This work and all authors were supported by the following grant support: Clinical Investigator Award from Damon Runyon Cancer Research Foundation (CI-8); NCI R01CA136726. We would like to acknowledge the staff at the Kentucky Cancer Registry.

LCCS: The Leeds Colorectal Cancer Study and all authors were funded by the Food Standards Agency and Cancer Research UK Programme Award (C588/A19167). We acknowledge the contributions of Jennifer Barrett, Robin Waxman, Gillian Smith and Emma Northwood in conducting this study.

Melbourne Collaborative Cohort Study (MCCS) cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS and all authors were further augmented by Australian National Health and Medical Research Council grants 209057, 396414 and 1074383 and by infrastructure provided by Cancer Council Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry and the Australian Institute of Health and Welfare, including the National Death Index and the Australian Cancer DatabaseMEC: National Institutes of Health (R37 CA054281, P01 CA033619, and R01 CA063464).

NCCCS I & II: We acknowledge funding support for this project and to all authors from the National Institutes of Health, R01 CA66635 and P30 DK034987. We would like to thank the study participants, and the NC Colorectal Cancer Study staff.

NFCCR: This work was supported by an Interdisciplinary Health Research Team award from the Canadian Institutes of Health Research (CRT 43821); the National Institutes of Health, U.S. Department of Health and Human Serivces (U01 CA74783); and National Cancer Institute of Canada grants (18223 and 18226). The authors wish to acknowledge the contribution of Alexandre Belisle and the genotyping team of the McGill University and Génome Québec Innovation Centre, Montréal, Canada, for genotyping the Sequenom panel in the NFCCR samples. Funding was provided to M. O. Woods by the Canadian Cancer Society Research Institute.

PLCO: Intramural Research Program of the Division of Cancer Epidemiology and Genetics and supported by contracts from the Division of Cancer Prevention, National Cancer Institute, NIH, DHHS. Funding was provided to the study and all authors by National Institutes of Health (NIH), Genes, Environment and Health Initiative (GEI) Z01 CP 010200, NIH U01 HG004446, and NIH GEI U01 HG 004438. The authors thank the PLCO Cancer Screening Trial screening center investigators and the staff from Information Management Services Inc and Westat Inc. Most importantly, we thank the study participants for their contributions that made this study possible. Cancer incidence data have been provided by the District of Columbia Cancer Registry, Georgia Cancer Registry, Hawaii Cancer Registry, Minnesota Cancer Surveillance System, Missouri Cancer Registry, Nevada Central Cancer Registry, Pennsylvania Cancer Registry, Texas Cancer Registry, Virginia Cancer Registry, and Wisconsin Cancer Reporting System. All are supported in part by funds from the Center for Disease Control and Prevention, National Program for Central Registries, local states or by the National Cancer Institute, Surveillance, Epidemiology, and End Results program. The results reported here and the conclusions derived are the sole responsibility of the authors.

SEARCH: The University of Cambridge has received salary support in respect of PDPP from the NHS in the East of England through the Clinical Academic Reserve. Cancer Research UK (C490/A16561); the UK National Institute for Health Research Biomedical Research Centres at the University of Cambridge. We thank the SEARCH team

SELECT: Research reported in this publication was supported in part by the National Cancer Institute of the National Institutes of Health under Award Numbers U10 CA37429 (to C.D. Blanke), and UM1 CA182883 (to C.M. Tangen and I.M. Thompson). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We thank the research and clinical staff at the sites that participated on SELECT study, without whom the trial would not have been successful. We are also grateful to the 35,533 dedicated men who participated in SELECT.

SMS and REACH: This work was supported by the National Cancer Institute (grant P01 CA074184 to J.D. Potter and P.A. Newcomb, grants R01 CA097325, R03 CA153323, and K05 CA152715 to P.A. Newcomb, and the National Center for Advancing Translational Sciences at the National Institutes of Health (grant KL2 TR000421 to A.N. Burnett-Hartman)

Swedish Mammography Cohort and Cohort of Swedish Men: This work is supported by the Swedish Research Council /Infrastructure grant, the Swedish Cancer Foundation, and the Karolinska Institute's Distinguished Professor Award to A. Wolk.

UK Biobank: This research has been conducted using the UK Biobank Resource under Application Number 8614

VITAL: This study was funded by the National Institutes of Health (K05 CA154337 to E. White).

WHI: The WHI program and all authors are funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C. The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: http://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Short%20List.pdf

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2018;68:394–424. [PubMed: 30207593]
2. Huyghe JR, Bien SA, Harrison TA, Kang HM, Abecasis GR, Nickerson DA, et al. Discovery of common and rare genetic risk variants for colorectal cancer. Nat Genet. 2019;51:76–87. [PubMed: 30510241]
3. Huyghe JR, Harrison TA, Bien SA, Hampel H, Figueiredo JC, Schmit SL, et al. Genetic architectures of proximal and distal colorectal cancer are partly distinct. Gut. 2021;0:1–10.
4. Law PJ, Timofeeva M, Fernandez-Rozadilla C, Broderick P, Studd J, Fernandez-Tajes J, et al. Association analyses identify 31 new risk loci for colorectal cancer susceptibility. Nat Commun. 2019;10:2154. [PubMed: 31089142]
5. Lu Y, Kweon SS, Tanikawa C, Jia WH, Xiang YB, Cai Q, et al. Large-Scale Genome-Wide Association Study of East Asians Identifies Loci Associated With Risk for Colorectal Cancer. Gastroenterology [Internet]. Elsevier, Inc; 2019;156:1455–66. Available from: 10.1053/j.gastro.2018.11.066 [PubMed: 30529582]
6. Lu Y, Kweon S, Cai Q, Tanikawa C, Shu X, Jia W, et al. Identification of Novel Loci and New Risk Variant in Known Loci for Colorectal Cancer Risk in East Asians. Cancer Epidemiol Biomarkers Prev. 2020;29:477–86. [PubMed: 31826910]
7. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. Environmental and heritable factors in the causation of cancer - Analyses of Cohorts of Twins from Sweden, Denmark, and Finland. N Engl J Med. 2000;343:78–85. [PubMed: 10891514]
8. Génin E Missing heritability of complex diseases: case solved? Hum Genet [Internet]. Springer Berlin Heidelberg; 2020;139:103–13. Available from: 10.1007/s00439-019-02034-4 [PubMed: 31165258]
9. Islami F, Bandi P, Sahar L, Ma J, Drope J, Jemal A. Cancer deaths attributable to cigarette smoking in 152 U.S. metropolitan or micropolitan statistical areas , 2013 – 2017. Cancer Causes Control [Internet]. Springer International Publishing; 2021;32:311–6. Available from: 10.1007/s10552-020-01385-y [PubMed: 33496899]

10. Kim H, Wang K, Song M, Giovannucci EL. A comparison of methods in estimating population attributable risk for colorectal cancer in the United States. Int J Cancer. 2021;148:2947–53. [PubMed: 33527363]
11. Wang S, Yuan Z, Wang Y, Zhao X, Gao W, Li H, et al. Modifiable lifestyle factors have a larger contribution to colorectal neoplasms than family history. BMC Cancer [Internet]. BioMed Central; 2022;22:1051. Available from: 10.1186/s12885-022-10141-1 [PubMed: 36207694]
12. Secretan B, Straif K, Baan R, Grosse Y, Ghissassi F El, Bouvard V, et al. A review of human carcinogens — Part E: tobacco, areca nut, alcohol, coal smoke, and salted fish. Lancet Oncol [Internet]. 2009;10:1033–4. Available from: 10.1016/S1470-2045(09)70326-2 [PubMed: 19891056]
13. Murphy N, Ward HA, Jenab M, Rothwell JA, Carbonnel F, Kvaskoff M, et al. Heterogeneity of Colorectal Cancer Risk Factors by Anatomical Subsite in 10 European Countries: A Multinational Cohort Study. Clin Gastroenterol Hepatol [Internet]. Elsevier, Inc; 2019;17:1323–1331.e6. Available from: 10.1016/j.cgh.2018.07.030 [PubMed: 30056182]
14. Botteri E, Borroni E, Sloan EK, Bagnardi V, Bosetti C, Peveri G, et al. Smoking and Colorectal Cancer Risk, Overall and by Molecular Subtypes: A Meta-Analysis. Am J Gastroenterol. 2020;115:1940–9. [PubMed: 32773458]
15. Liang PS, Chen T, Giovannucci E. Cigarette smoking and colorectal cancer incidence and mortality: Systematic review and meta-analysis. Int J Cancer. 2009;124:2406–15. [PubMed: 19142968]
16. Dimou N, Yarmolinsky J, Bouras E, Tsilidis KK, Martin RM, Lewis SJ, et al. Causal effects of lifetime smoking on breast and colorectal cancer risk: Mendelian randomization study. Cancer Epidemiol Biomarkers Prev. 2021;30:953–64. [PubMed: 33653810]
17. Burgess S, Labrecque JA. Mendelian randomization with a binary exposure variable: interpretation and presentation of causal estimates. Eur J Epidemiol [Internet]. Springer Netherlands; 2018;33:947–52. Available from: 10.1007/s10654-018-0424-6 [PubMed: 30039250]
18. Li Y, Xiao X, Han Y, Gorlova O, Qian D, Leighl N, et al. Genome-wide interaction study of smoking behavior and non-small cell lung cancer risk in Caucasian population. Carcinogenesis. 2018;39:336–46. [PubMed: 29059373]
19. Kundu P, Mocci E, Wheeler W, Amundadottir LT, Li D, Jacobs EJ, et al. Genome-wide interaction scan identifies gene by smoking interaction at 2q21.3 for pancreatic cancer risk. Cancer Res. 2020;
20. Figueroa JD, Han SS, Baris D, Jacobs EJ, Kogevinas M, Schwenn M, et al. Genome-wide interaction study of smoking and bladder cancer risk. Carcinogenesis. 2014;35:1737–44. [PubMed: 24662972]
21. Jiao S, Peters U, Berndt S, Brenner H, Butterbach K, Caan BJ, et al. Estimating the heritability of colorectal cancer. Hum Mol Genet. 2014;23:3898–905. [PubMed: 24562164]
22. Gong J, Hutter CM, Newcomb PA, Ulrich CM, Bien A, Campbell PT, et al. Genome-Wide Interaction Analyses between Genetic Variants and Alcohol Consumption and Smoking for Risk of Colorectal Cancer. PLoS Genet. 2016;12:e1006296. [PubMed: 27723779]
23. Smith PG, Day NE. The Design of Case-Control Studies: The Influence of Confounding and Interaction Effects. Int J Epidemiol. 1984;13:356–365. [PubMed: 6386716]
24. Gauderman WJ, Mukherjee B, Aschard H, Hsu L, Lewinger JP, Patel CJ, et al. Special Article Update on the State of the Science for Analytical Methods for Gene-Environment Interactions. Am J Epidemiol. 2017;186:762–70. [PubMed: 28978192]
25. Jeon J, Du M, Schoen RE, Hoffmeister M, Newcomb PA, Berndt SI, et al. Determining Risk of Colorectal Cancer and Starting Age of Screening Based on Lifestyle, Environmental, and Genetic Factors. Gastroenterology. 2018;154:2152–64. [PubMed: 29458155]
26. Wang X, Connell KO, Jeon J, Song M, Hunter D, Hoffmeister M, et al. Combined effect of modifiable and non-modifiable risk factors for colorectal cancer risk in a pooled analysis of 11 based studies. BMJ Open Gastroentorology. 2019;6:e000339.
27. Peters U, Jiao S, Schumacher FR, Hutter CM, Aragaki AK, Baron JA, et al. Identification of Genetic Susceptibility Loci for Colorectal Tumors in a Genome-Wide Meta-analysis. Gastroenterology. 2013;144:799–807. [PubMed: 23266556]

28. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet. 2016;48:1279–83. [PubMed: 27548312]
29. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet. 2016;48:1284–7. [PubMed: 27571263]
30. Gauderman WJ, Kim A, Conti DV, Morrison J, Thomas DC, Vora H, et al. A Unified Model for the Analysis of Gene-Environment Interaction. Am Lournal Epidemiol. 2019;188:760–7.
31. Kraft P, Yen Y, Stram O, Morrison J, Gauderman WJ. Exploiting Gene-Environment Interaction to Detect Genetic Associations. Hum Hered. 2007;63:111–9. [PubMed: 17283440]
32. Murcray CE, Lewinger JP, Gauderman WJ. Gene-Environment Interaction in Genome-Wide Association Studies. Am J Epidemiol. 2009;169:219–26. [PubMed: 19022827]
33. Dai JY, Kooperberg C, Leblanc M, Prentice RL. Two-stage testing procedures with independent filtering for genome-wide gene-environment interaction. Biometrika. 2012;99:929–44. [PubMed: 23843674]
34. Gauderman WJ, Zhang P, Morrison JL, Lewinger JP. Finding Novel Genes by Testing G×E Interactions in a Genome-Wide Association Study. Genet Epidemiol. 2013;37:603–613,. [PubMed: 23873611]
35. Ionita-Laza I, McQueen MB, Laird NM, Lange C. Genomewide Weighted Hypothesis Testing in Family-Based Association Studies, with an Application to a 100K Scan. Am J Hum Genet. 2007;81:607–14. [PubMed: 17701906]
36. Kawaguchi ES, Kim AE, Lewinger JP, Gauderman WJ. Improved two-step testing of genome-wide gene-environment interactions. bioRxiv [Internet]. 2022; Available from: https://www.biorxiv.org/content/10.1101/2022.06.14.496154v2
37. Vander Weele TJ, Ko YA, Mukherjee B. Environmental confounding in gene-environment interaction studies. Am J Epidemiol. 2013;178:144–52. [PubMed: 23821317]
38. Carreras-Torres R, Johansson M, Haycock PC, Relton CL, Davey Smith G, Brennan P, et al. Role of obesity in smoking behaviour: Mendelian randomisation study in UK Biobank. BMJ. 2018;361:k1767. [PubMed: 29769355]
39. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010;26:2336–7. [PubMed: 20634204]
40. Cohen AJ, Saiakhova A, Corradin O, Luppino JM, Lovrenert K, Bartels CF, et al. Hotspots of aberrant enhancer activity punctuate the colorectal cancer epigenome. Nat Commun. Nature Publishing Group; 2017;8:14400. [PubMed: 28169291]
41. The GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science (80-). 2020;369:1318–1330.
42. Díez-Obrero V, Dampier CH, Moratalla-Navarro F, Devall M, Plummer SJ, Díez-Villanueva A, et al. Genetic Effects on Transcriptome Profiles in Colon Epithelium Provide Functional Insights for Genetic Risk Loci. Cell Mol Gastroenterol Hepatol [Internet]. Elsevier Inc; 2021;12:181–97. Available from: 10.1016/j.jcmgh.2021.02.003 [PubMed: 33601062]
43. Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-michaels K, Carroll RJ, et al. A gene-based association method for mapping traits using reference transcriptome data. Nat Genet. 2015;47.
44. Chang G, Xu S, Dhir R, Chandran U, Keefe DSO, Greenberg NM, et al. Hypoexpression and Epigenetic Regulation of Candidate Tumor Suppressor Gene CADM-2 in Human Prostate Cancer. Clin Cancer Res. 2010;16:5390–402. [PubMed: 21062931]
45. Chen YB, Gao L, Zhang JD, Guo J, You PH, Tang LY, et al. Weighted Gene Coexpression Network Analysis to Construct Competitive Endogenous RNA Network in Chromogenic Renal Cell Carcinoma. Biomed Res Int. 2021;2021.
46. Cody NAL, Shen Z, Ripeau J, Provencher DM, Chevrette M, Tonin PN. Characterization of the 3p12.3-pcen Region Associated With Tumor Suppression in a Novel Ovarian Cancer Cell Line Model Genetically Modified by Chromosome 3 Fragment Transfer. Mol Carcinog. 2009;48:1077–92. [PubMed: 19347865]

47. Lake SL, Coupland SE, Taktak AFG, Damato BE. Whole-Genome Microarray Detects Deletions and Loss of Heterozygosity of Chromosome 3 Occurring Exclusively in Metastasizing Uveal Melanoma. Anat Pathol. 2010;51:4884–91.
48. Li D, Zhang Y, Zhang H, Zhan C, Li X, Ba T, et al. CADM2, as a new target of miR-10b, promotes tumor metastasis through FAK/AKT pathway in hepatocellular carcinoma. J Exp Clin Cancer Res. 2018;37:1–11. [PubMed: 29301578]
49. Liu N, Yang C, Bai W, Wang Z, Wang X, Johnson M, et al. CADM2 inhibits human glioma proliferation, migration and invasion. Oncol Rep. 2019;41:2273–80. [PubMed: 30816549]
50. Roy D, Sin S, Damania B, Dittmer DP. Tumor suppressor genes FHIT and WWOX are deleted in primary effusion lymphoma (PEL) cell lines. Blood. 2011;118:32–9.
51. Sanchez-Roige S, Fontanillas P, Elson SL, Gray JC, De Wit H, MacKillop J, et al. Genome-wide association studies of impulsive personality traits (BIS-11 and UPPS-P) and drug experimentation in up to 22,861 adult research participants identify loci in the CACNA1I and CADM2 genes. J Neurosci. 2019;39:2562–72. [PubMed: 30718321]
52. Arends RM, Pasman JA, Verweij KJH, Derks EM, Gordon SD, Hickie I, et al. Associations between the CADM2 gene, substance use, risky sexual behavior, and self-control: A phenome-wide association study. Addict Biol. 2021;26:1–13.
53. Strawbridge RJ, Ward J, Cullen B, Tunbridge EM, Hartz S, Bierut L, et al. Genome-wide analysis of self-reported risk-taking behaviour and cross-disorder genetic correlations in the UK Biobank cohort. Transl Psychiatry. Springer US; 2018;8:1–11.
54. Gros P, Milder FJ, Janssen BJC. Complement driven by conformational changes. Nat Rev Immunol. 2008;8:48–58. [PubMed: 18064050]
55. Carroll MV, Sim RB. Complement in health and disease. Adv Drug Deliv Rev [Internet]. Elsevier B.V.; 2011;63:965–75. Available from: 10.1016/j.addr.2011.06.005 [PubMed: 21704094]
56. Juyal G, Negi S, Sood A, Gupta A, Prasad P, Senapati S, et al. Genome-wide association scan in north Indians reveals three novel HLA-independent risk loci for ulcerative colitis. Gut. 2015;64:571–9. [PubMed: 24837172]
57. Gupta A, Juyal G, Sood A, Midha V, Yamazaki K, Vila AV, et al. A cross-ethnic survey of CFB and SLC44A4 , Indian ulcerative colitis GWAS hits , underscores their potential role in disease susceptibility. Eur J Hum Genet [Internet]. Nature Publishing Group; 2017;25:111–22. Available from: 10.1038/ejhg.2016.131
58. Shi D, Zhong Z, Wang M, Cai L, Fu D, Peng Y, et al. Identification of susceptibility locus shared by IgA nephropathy and inflammatory bowel disease in a Chinese Han population. J Hum Genet. 2020;65:241–9. [PubMed: 31857673]
59. Ron D, Walter P. Signal integration in the endoplasmic reticulum unfolded protein response. Nat Rev Mol Cell Biol. 2007;8:519–29. [PubMed: 17565364]
60. Suhre K, Arnold M, Bhagwat AM, Cotton RJ, Engelke R, Raffler J, et al. Connecting genetic risk to disease end points through the human blood plasma proteome. Nat Commun. 2017;8:14357. [PubMed: 28240269]
61. Rivera NV, Ronninger M, Shchetynsky K, Franke A, Nöthen MM, Ller-Quernheim JM, et al. High-Density Genetic Mapping Identifies New Susceptibility Variants in Sarcoidosis Phenotypes and Shows Genomic-driven Phenotypic Differences. Am J Respir Crit Care Med. 2016;193:1008–22. [PubMed: 26651848]
62. Kim W, Prokopenko D, Sakornsakolpat P, Hobbs BD, Lutz SM, Hokanson JE, et al. Genome-Wide Gene-by-Smoking Interaction Study of Chronic Obstructive. Am J Epidemiol. 2020;190:875–85.
63. Wu F, Shirahata A, Sakuraba K, Kitamura Y, Goto T, Saito M, et al. Down-regulation of EGFL8: A Novel Biomarker for Advanced Gastric Cancer. Anticancer Res. 2011;31:3377–80. [PubMed: 21965749]
64. Shi S, Ma T, Xi Y. A Pan-Cancer Study of Epidermal Growth Factor-Like Domains 6/7/8 as Therapeutic Targets in Cancer. Front Genet. 2020;11:598743. [PubMed: 33391349]
65. Savin Z, Kivity S, Yonath H, Yehuda S. Smoking and the intestinal microbiome. Arch Microbiol [Internet]. Springer Berlin Heidelberg; 2018;200:677–84. Available from: 10.1007/s00203-018-1506-2 [PubMed: 29626219]

66. Huang C, Shi G. Smoking and microbiome in oral, airway, gut and some systemic diseases. J Transl Med [Internet]. BioMed Central; 2019;17:1–15. Available from: 10.1186/s12967-019-1971-7 [PubMed: 30602370]

67. Wong SH, Yu J. Gut microbiota in colorectal cancer: mechanisms of action and clinical applications. Nat Rev Gastroenterol Hepatol [Internet]. Springer US; 2019;16:690–704. Available from: 10.1038/s41575-019-0209-8 [PubMed: 31554963]

68. Huybrechts I, Zouiouich S, Loobuyck A, Vandenbulcke Z, Vogtmann E, Pisanu S, et al. The Human Microbiome in Relation to Cancer Risk: A Systematic Review of Epidemiologic Studies. Cancer Epidemiol Biomarkers Prev. 2020;29:1856–69. [PubMed: 32727720]

69. Hamada T, Nowak JA, Masugi Y, Drew DA, Song M, Cao Y, et al. Smoking and risk of colorectal cancer sub-classified by tumor-infiltrating T cells. J Natl Cancer Inst. 2019;111:42–51. [PubMed: 30312431]

70. Ugai T, Väyrynen JP, Haruki K, Akimoto N, Lau MC, Zhong R, et al. Smoking and Incidence of Colorectal Cancer Subclassified by Tumor-Associated Macrophage Infiltrates. JNCI J Natl Cancer Inst. 2021;00:1–10.

71. Wang T, Lv X, Jiang S, Han S, Wang Y. Expression of ADAM29 and FAM135B in the pathological evolution from normal esophageal epithelium to esophageal cancer: Their differences and clinical significance. Oncol Lett. 2020;19:1727–34. [PubMed: 32194665]

72. Wojcik GL, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR, et al. Genetic analyses of diverse populations improves discovery for complex traits. Nature [Internet]. Springer US; 2019;570:514–8. Available from: 10.1038/s41586-019-1310-4

73. Hou K, Ding Y, Xu Z, Wu Y, Bhattacharya A, Mester R, et al. Causal effects on complex traits are similar across segments of different continental ancestries within admixed individuals. MedRxiv. 2022;

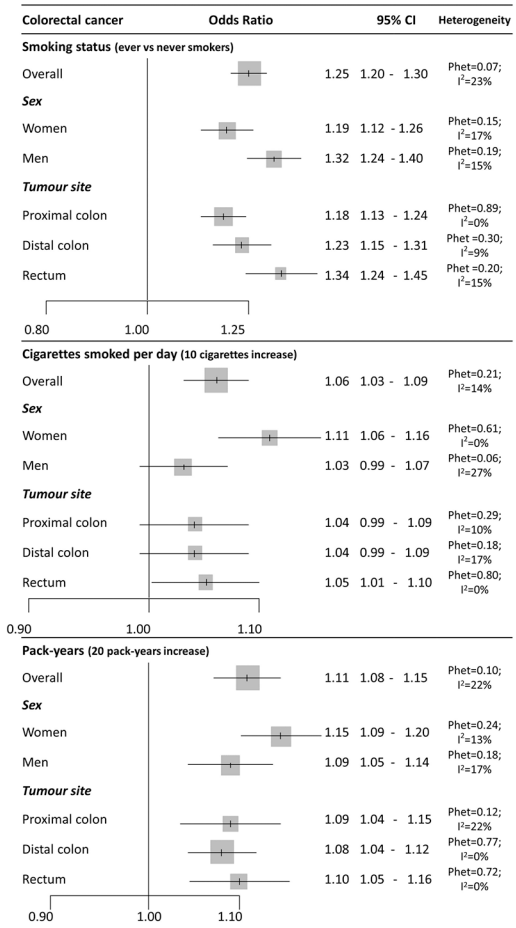

Figure 1-

Association parameters of smoking habits for colorectal cancer risk in the overall sample and stratified by sex and tumour site.

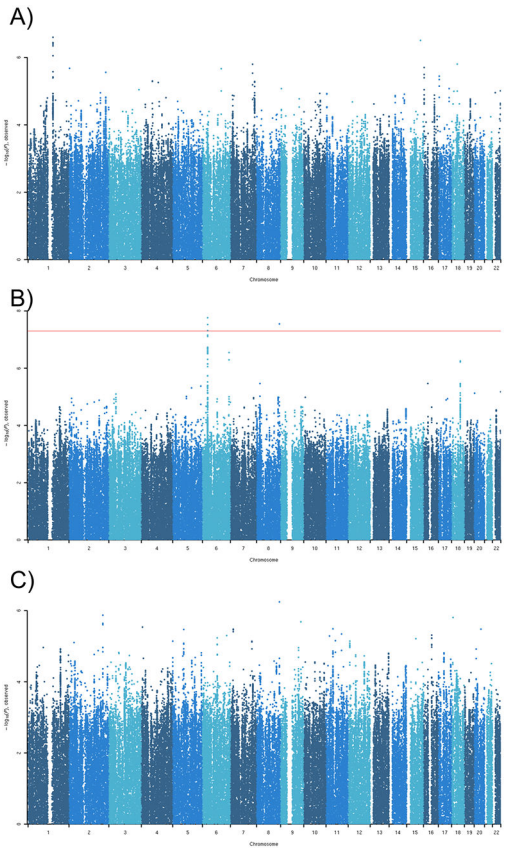

Figure 2-

Manhattan plots for the standard interaction tests of smoking habits for colorectal cancer risk. A: Smoking status (ever vs never smokers). B: Cigarettes smoked per day. C: Pack-years.

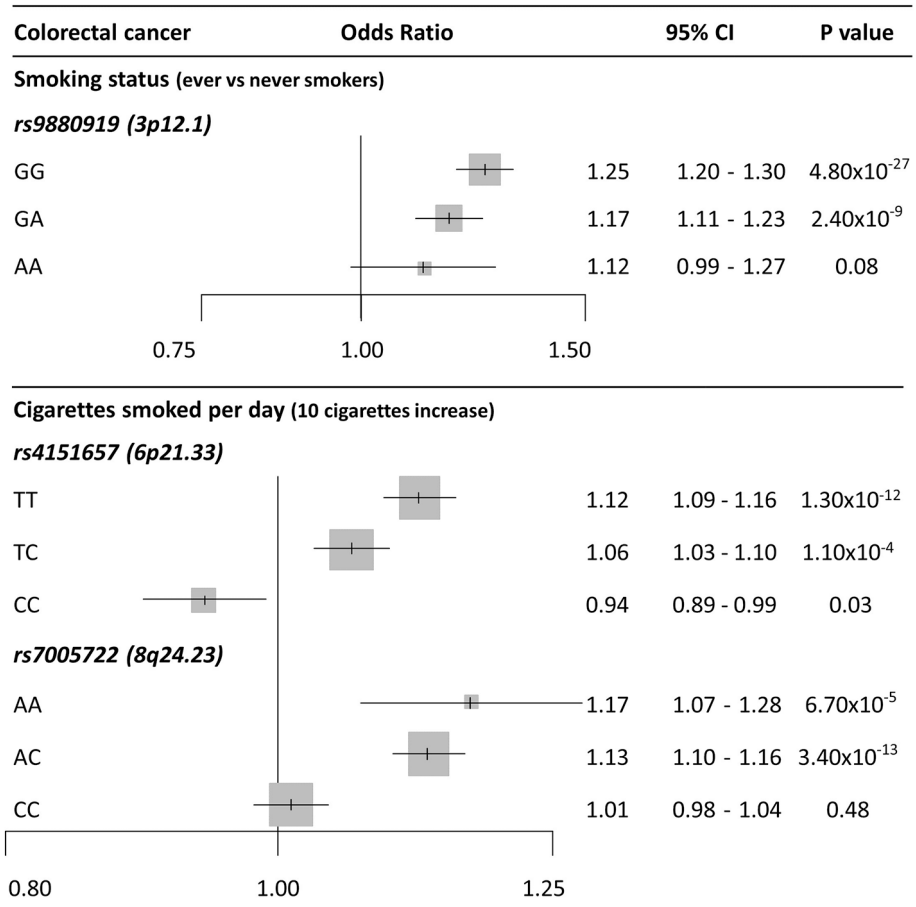

Figure 3-

Association parameters of smoking habits for colorectal cancer risk stratified by genotypes.

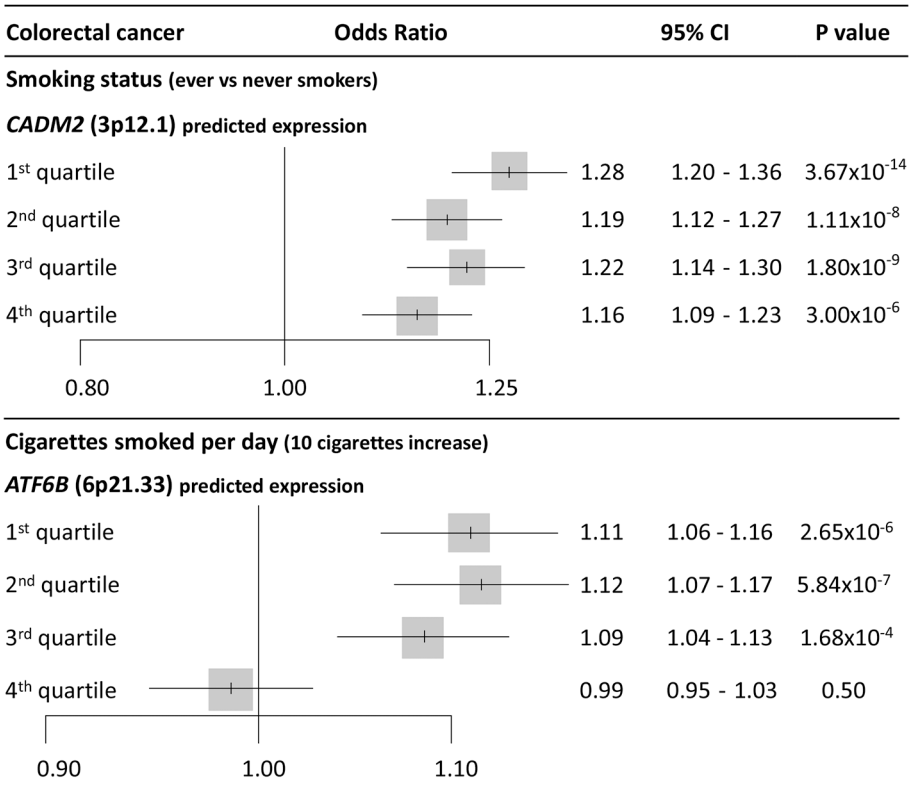

Figure 4-

Association parameters of genetically predicted gene expression for colorectal cancer risk stratified by smoking status.

Table 1-

Association parameters of identified genetic variants interacting with smoking habits on colorectal cancer risk.

rs number: single nucleotide polymorphism (SNP) code, Chr: chromosome, Bp: base pair position in hg19, Ref Al: reference allele, Eff Al: effect allele, Eff Al Freq: reference allele frequency, P1: P value of genetic component for CRC risk (Ho: βG=0), P2: P value of interaction component for CRC risk (Ho: βGxE=0), P3: P value of genetic component for the exposure (Ho:δG=0), P3df; P value of 3 df test (Ho:βG=βGxE=δG=0). Allele frequencies calculated in 1000G EUR population.

Smoking habit	rs number	Chr	Bp	Locus	Ref Al	Eff Al	Eff Al Freq	Gene	P1 (β_G)	P2 (β_{GxE})	P3 (δ_G)	P3df
Status (ever vs never)	rs9880919	3	85,461,302	3p12.1	G	A	0.243	*CADM2*	$1.86x10^{-6}$	0.01	$4.20x10^{-3}$	$4.58x10^{-8}$
Cigarettes smoked per day	rs4151657	6	31,917,540	6p21.33	T	C	0.365	*CFB*	0.21	$1.72x10^{-8}$	0.04	$3.52x10^{-8}$
	rs7005722	8	138,788,813	8q24.23	A	C	0.743	*Intergenic*	0.51	$2.84x10^{-8}$	0.66	$6.81x10^{-7}$

Table 2-

Functional annotation of lead and correlated SNPs as variant enhancer loci (VEL) or expression quantitative trait loci (eQTL).

Characters in bold have the lead SNP as VEL or eQTL. eQTLs were assessed in the BarcUVA-Seq project data and GTEx colon transverse.

				N VELs for regulating elements					**N eQTLs for genes**
				Non-malignant tissue			**Tumor tissue**		
Smoking habit	**Locus**	**Lead SNP**	**SNPs in LD (R^2>0.5)**	**ATAC-Seq**	**H3K27ac**	**DHS**	**H3K27ac**	**DHS**	**Genes (N in BarcUVa-Seq; N in GTEx colon transverse)**
Status (ever vs never)	3p12.1	rs9880919	588	13	1	5	3	0	**CADM2(460**;43)
Cigarettes smoked per day	6p21.33	rs4151657	44	0	4	3	13	0	VARS2 (4;8); CCHCR1 (0;9); PSORS1C3 (0;4); HCG27 (0,25); *SNORA38 (36;0); BAG6 (0;25); CSNK2B (36;0)*; LY6G5B (0;1); VARS1 (8;0); SKIV2L (2;0); C4A (1;0); **CYP21A1P (0;36); TNXA (0;5)**; C4B (5;0); **ATF6B (38;36)**; PPT2 (5;0); **EGFL8 (2;27)**; RNF5 (1;1); PBX2 (3;0); **HLA-DRB6 (34;0)**; HLA-DQA2 (12;6); HLA-DQB2 (5;0)
	8q24.23	rs7005722	0	0	0	0	0	0	-