# Causal Effects of Lifetime Smoking on Breast and Colorectal Cancer Risk: Mendelian Randomization Study

Niki Dimou[1], James Yarmolinsky[2,3], Emmanouil Bouras[4,5], Konstantinos K. Tsilidis[5,6], Richard M. Martin[2,3,7], Sarah J. Lewis[2,3], Inger T. Gram[8], Marije F. Bakker[9], Hermann Brenner[10,11,12], Jane C. Figueiredo[13,14], Renée T. Fortner[15], Stephen B. Gruber[16], Bethany van Guelpen[17,18], Li Hsu[19,20], Rudolf Kaaks[15,21], Sun-Seog Kweon[22,23], Yi Lin[19], Noralane M. Lindor[24], Polly A. Newcomb[19,25], Maria-Jose Sánchez[26,27,28,29], Gianluca Severi[30,31], Hilary A. Tindle[32], Rosario Tumino[33], Elisabete Weiderpass[34], Marc J. Gunter[1], and Neil Murphy[1]

## ABSTRACT

**Background:** Observational evidence has shown that smoking is a risk factor for breast and colorectal cancer. We used Mendelian randomization (MR) to examine causal associations between smoking and risks of breast and colorectal cancer.

**Methods:** Genome-Wide Association Study summary data were used to identify genetic variants associated with lifetime amount of smoking ($n = 126$ variants) and ever having smoked regularly ($n = 112$ variants). Using two-sample MR, we examined these variants in relation to incident breast (122,977 cases/105,974 controls) and colorectal cancer (52,775 cases/45,940 controls).

**Results:** In inverse-variance weighted models, a genetic predisposition to higher lifetime amount of smoking was positively associated with breast cancer risk [OR per 1-SD increment: 1.13; 95% confidence interval (CI): 1.00–1.26; $P = 0.04$]; although heterogeneity was observed. Similar associations were found for estrogen receptor–positive and estrogen receptor–negative tumors. Higher lifetime amount of smoking was positively associated with colorectal cancer (OR per 1-SD increment, 1.21; 95% CI, 1.04–1.40; $P = 0.01$), colon cancer (OR, 1.31; 95% CI, 1.11–1.55; $P < 0.01$), and rectal cancer (OR, 1.36; 95% CI, 1.07–1.73; $P = 0.01$). Ever having smoked regularly was not associated with risks of breast (OR, 1.01; 95% CI, 0.90–1.14; $P = 0.85$) or colorectal cancer (OR, 0.97; 95% CI, 0.86–1.10; $P = 0.68$).

**Conclusions:** These findings are consistent with prior observational evidence and support a causal role of higher lifetime smoking amount in the development of breast and colorectal cancer.

**Impact:** The results from this comprehensive MR analysis indicate that lifetime smoking is a causal risk factor for these common malignancies.

[1]Section of Nutrition and Metabolism, International Agency for Research on Cancer, Lyon, France. [2]Department of Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom. [3]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom. [4]Laboratory of Hygiene, Social & Preventive Medicine and Medical Statistics, Department of Medicine, School of Health Sciences, Aristotle University of Thessaloniki, Thessaloniki, Greece. [5]Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece. [6]Department of Epidemiology and Biostatistics, The School of Public Health, Imperial College London, London, United Kingdom. [7]National Institute for Health Research (NIHR) Bristol Biomedical Research Centre, University Hospitals Bristol NHS Foundation Trust and University of Bristol, Bristol, United Kingdom. [8]Department of Community Medicine, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway. [9]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, University Utrecht, Utrecht, the Netherlands. [10]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany. [11]Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany. [12]German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany. [13]Department of Medicine, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, California. [14]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California. [15]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany. [16]Center for Precision Medicine, City of Hope National Medical Center, Duarte, California. [17]Department of Radiation Sciences, Oncology Unit, Umea University, Umea, Sweden. [18]Wallenberg Centre for Molecular Medicine, Umeå University, Umeå Sweden. [19]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington. [20]Department of Biostatistics, University of Washington, Seattle, Washington. [21]Member of the German Center for Lung Research (DZL), Translational Lung Research Center (TLRC), Heidelberg, Germany. [22]Department of Preventive Medicine, Chonnam National University Medical School, Gwangju, Korea. [23]Jeonnam Regional Cancer Center, Chonnam National University Hwasun Hospital, Hwasun, Korea. [24]Department of Health Science Research, Mayo Clinic, Scottsdale, Arizona. [25]School of Public Health, University of Washington, Seattle, Washington. [26]Escuela Andaluza de Salud Pública (EASP), Granada, Spain. [27]Instituto de Investigación Biosanitaria ibs.GRANADA, Granada, Spain. [28]Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain. [29]Department of Preventive Medicine and Public Health, University of Granada, Granada, Spain. [30]CESP UMR1018, Université Paris-Saclay, UVSQ, Inserm, Gustave Roussy, Villejuif, France. [31]Department of Statistics, Computer Science and Applications "G. Parenti", University of Florence, Florence, Italy. [32]General Internal Medicine, Vanderbilt University Medical Center, Vanderbilt University, Nashville, Tennessee. [33]Cancer Registry and Histopathology Department, Provincial Health Authority (ASP 7) Ragusa, Ragusa, Italy. [34]Office of the Director, International Agency for Research on Cancer, World Health Organization, Lyon, France.

**Note:** Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (http://cebp.aacrjournals.org/).

**Corresponding Author:** Niki Dimou, Section of Nutrition and Metabolism, International Agency for Research on Cancer, 150 Cours Albert Thomas, Lyon 69732, France. Phone: 33-472-73-81-88; Fax; 33-472-73-86-33; E-mail: DimouN@fellows.iarc.fr

Cancer Epidemiol Biomarkers Prev 2021;30:953–64

**doi:** 10.1158/1055-9965.EPI-20-1218

## Introduction

Breast and colorectal cancer are two of the most common cancers globally with a combined estimated number of 4 million new cases and 1.5 million deaths in 2018 (1). Smoking is the most common cancer risk factor globally, with an estimated 18% of all cancers attributable to smoking (2). Tobacco smoke is the main source of exposure to aromatic amines (i.e., aniline, and o-anisidine), with strong evidence of carcinogenic effects in animals (3). Epidemiologic studies have generally found smoking to be associated with higher risks of breast and colorectal cancer (4, 5). However, the potential causality in humans of these relationships is uncertain due to the modest strength of the reported relative risks (often in the order of 1.07 to 1.20 for current vs. never smokers; refs. 4, 5) and the possibility of residual confounding from other established risk factors for these malignancies, such as elevated adults body mass index (BMI) and alcohol consumption.

Mendelian randomization (MR) is an increasingly used analytic method to investigate causal inference between a given exposure and outcome. MR uses genetic variants robustly associated with the exposure of interest as instrumental variables to facilitate the estimation of causal effect of an exposure on an outcome (6). MR analyses should be largely free of the sources of confounding and reverse causality that are characteristic of conventional observational epidemiologic studies, as genetic variants are randomly assigned, and fixed, at conception.

We used a two-sample MR framework to examine the association of smoking with breast and colorectal cancer. We constructed two genetic instruments for smoking: (i) lifetime amount of smoking (7) and, (ii) ever having smoked regularly (8). We then investigated the associations of these genetic instruments with risks of breast cancer (122,977 cases and 105,974 controls) and colorectal cancer (52,775 cases and 45,940 controls).

## Methods

### Data on lifetime amount of smoking and ever having smoked regularly

We selected genetic variants for the MR analysis on the basis of a genome-wide significant association (i.e., $P$ value threshold for inclusion at $<5 \times 10^{-8}$) with lifetime amount of smoking (capturing information on smoking duration, intensity, and cessation) and, ever having smoked regularly (binary phenotype with ever vs. never smoking status). Variants associated with lifetime amount of smoking were selected from a genome-wide association study (GWAS) of 462,690 individuals in the UK Biobank, from which a score was constructed on the basis of combined information on smoking intensity (number of cigarettes per day), smoking duration, and ever/never regular smoking status (7). Information on the calculation of lifetime amount of smoking score, quality control, and selection of the participants was detailed in the published GWAS (7). Briefly, this score is zero in nonsmokers, uses age at initiation and age at cessation to quantify smoking duration in current and former smokers, number of cigarettes smoked per day to measure smoking intensity, and a half-life constant that captures the exponentially decreasing effect of smoking at a given time on health outcomes. For the ever having smoked regularly instrument, we extracted variants from a GWAS meta-analysis of the GSCAN consortium for a total of 1,232,091 individuals of European ancestry (8). It is a binary phenotype, coding participants reporting ever being a regular smoker in their life (current or former) as ever having smoked regularly, while any participant who reported never being a regular smoker in their life was coded as never having smoked regularly. More specifically, ever smokers regularly were classified as those who had smoked over 100 cigarettes over the course of their life, and/or had smoked every day for at least a month, and/or had ever smoked regularly. Information about pipes/cigar/chew, or other noncigarette forms of tobacco use is not included. The selected genetic variants were pruned based on a $R^2$ linkage disequilibrium (LD) <0.001, resulting in 126 independent variants for lifetime amount of smoking and 112 for ever having smoked regularly. The genetic instruments explained ∼0.4% and ∼2.0% of the variance in lifetime amount of smoking and ever having smoked regularly, respectively.

### GWAS data for breast and colorectal cancer

The association of smoking-related genetic variants with breast cancer risk was obtained using data from the Breast Cancer Association Consortium (BCAC), involving 122,977 cases (105,974 controls), of which, 69,501 were estrogen receptor–positive ($ER^+$) cases and 21,468 were estrogen receptor–negative ($ER^-$) cases (Supplementary Table S1; ref. 9). For colorectal cancer, we used summary data from a meta-analysis of 98,715 participants (52,775 colorectal cancer cases and 45,940 controls) within the ColoRectal Transdisciplinary Study (CORECT), the Colon Cancer Family Registry (CCFR), and the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO; ref. 10; Supplementary Table S1). All smoking-related genetic variants had an imputation score above 0.6 and 0.3 in breast and colorectal cancer GWAS, respectively.

### Statistical analysis

For our main analysis, we used a random-effects inverse-variance weighted (IVW) method (11, 12). A series of statistical tests were conducted to investigate potential violation of the MR assumptions (13, 14), and to assess the possible influence of horizontal pleiotropy on the causal estimates, including MR-Egger regression (15) and the estimator from the weighted median approach (16). We calculated the Cochran $Q$ statistic that quantifies the heterogeneity in effect sizes attributed to the selected genetic variants. We also estimated the intercept term from the MR–Egger regression, with a deviation from zero being indicative of directional (nonbalanced horizontal) pleiotropy (15). The MR pleiotropy residual sum and outlier test (MR-PRESSO) was performed to identify, and exclude, any outlying variants ($P$ value threshold set at < 0.05; ref. 17). Leave-one-variant out analyses were conducted to assess the influence of individual variants on the observed associations. We also conducted multivariable MR analyses, adjusting for BMI and alcohol consumption, established confounders of the smoking-cancer association (4, 5). For BMI, we used data from a recent GWAS of the GIANT consortium and the UK Biobank including approximately 700,000 individuals of European ancestry (18). Sex-specific estimates for BMI were obtained from a previous GWAS within the GIANT consortium including 171,977 women and 152,893 men (19). For alcohol consumption, GWAS data on drinks per week for 537,349 participants were used (8). Further multivariable MR analyses were performed including GWAS data for educational attainment to control for socioeconomic status (20). For breast cancer, we also examined whether the effect of smoking was independent of age at menarche in a multivariable MR setting using GWAS data in up to approximately 370,000 women (21). For multivariable MR analyses, genome-wide significant signals of the exposures assessed (i.e., BMI, alcohol consumption, educational attainment, and age at menarche) independent of smoking-related variants ($R^2$ LD < 0.001) were also included in the models (Supplementary Tables S2 and S3). As a positive control analysis, we examined

the association between ever having smoked regularly and lung cancer using data from the International Lung Cancer Consortium (ILCCO) GWAS, which comprises 11,348 cases and 15,861 controls (22). The genetic instrument for lifetime amount of smoking has previously been shown to be strongly and robustly associated with lung cancer risk [OR per 1-SD increment: 4.21 (95% confidence interval (CI): 2.98–5.96); $P < 0.01$; Supplementary Table S4; ref. 7)]. Missing variants in any of the exposures used in the multivariable MR analyses and in the outcomes assessed were replaced by a suitable proxy (minimum LD $R^2 = 0.8$) where available.

The statistical analyses were implemented in the MR R package (23) and MR-PRESSO version 1.0 in the R environment (17).

### Data availability statement

The data underlying this article will be shared on reasonable request to the corresponding author.

## Results

### MR estimates for lifetime amount of smoking

In the IVW model, a genetic predisposition to higher lifetime amount of smoking was positively associated with risk of breast cancer (OR per 1-SD increment, 1.13; 95% CI: 1.00–1.26; $P = 0.04$), with similar effect estimates found for $ER^+$ (OR, 1.11; 95% CI, 0.97–1.26; $P = 0.12$) and $ER^-$ breast cancer (OR, 1.15; 95% CI, 0.96–1.37; $P = 0.14$; **Table 1**). Genetic predisposition to lifetime amount of smoking was also positively associated with risks of colorectal cancer (OR, 1.21; 95% CI, 1.04–1.40; $P = 0.01$], colon cancer (OR, 1.31; 95% CI, 1.11–1.55; $P < 0.01$), and rectal cancer (OR, 1.36; 95% CI, 1.07–1.73; $P = 0.01$; **Table 1**). Similar magnitude effect estimates were found for distal colon cancer (OR, 1.50; 95% CI, 1.21–1.87; $P < 0.01$) and proximal colon cancer (OR, 1.19; 95% CI, 0.98–1.46; $P = 0.09$; $P_{heterogeneity} = 0.13$). A positive effect of lifetime amount of smoking on colorectal cancer risk was found for men (OR, 1.32; 95% CI, 1.11–1.59; $P < 0.01$), but not women (OR, 1.11; 95% CI, 0.90–1.37; $P = 0.34$), although the evidence for a difference by sex was weak ($P_{heterogeneity} = 0.21$).

Evidence of effect heterogeneity (Cochran Q $P$ values <0.01) was found for most models. Little evidence of directional pleiotropy was found for all models (nearly zero MR-Egger intercept values), with similar magnitude effect estimates generally found for the weighted median method (**Table 1**). The MR-Egger analyses also yielded similar effect estimates to the IVW models for all endpoints, with the exception of ER- breast cancer, for which an inverse effect estimate was found (**Table 1**). The MR-PRESSO and the analysis removing one variant at a time, yielded similar results in all analyses performed (**Table 1**; Supplementary Tables S5A and S6A).

Multivariable MR analyses, adjusting for BMI and alcohol consumption, resulted in raised effect estimates for breast cancer (OR, 1.28; 95% CI, 1.09–1.49; $P < 0.01$), $ER^+$ breast cancer (OR, 1.25; 95% CI, 1.05–1.49; $P = 0.01$), and $ER^-$ breast cancer (OR, 1.45; 95% CI, 1.14–1.85; $P < 0.01$), with this strengthening of association driven mostly by adjustment for BMI (**Table 1**). Conversely, the positive associations between genetic predisposition to lifetime amount of smoking and colorectal cancer were slightly attenuated in the multivariable model (OR, 1.18; 95% CI, 0.98–1.43; $P = 0.09$; **Table 1**), with adjustment for BMI being most influential (OR, 1.15; 95% CI, 0.97–1.36; $P = 0.12$). However, the positive associations for colon cancer (OR, 1.25; 95% CI, 1.02–1.53; $P = 0.03$), distal colon cancer (OR, 1.39; 95% CI, 1.07–1.80; $P = 0.01$] and rectal cancer (OR, 1.33; 95% CI, 0.99–1.79; $P = 0.05$) remained after BMI and alcohol consumption multivariable adjustments (**Fig. 1A**). In the multivariable MR models accounting for educational attainment, a genetic predisposition to higher lifetime amount of smoking was positively associated with risk of breast cancer (OR, 1.18; 95% CI, 1.00–1.38; $P = 0.05$) but attenuated toward the null for colorectal cancer (OR, 1.11; 95% CI, 0.93–1.32; $P = 0.24$). A positive association remained between lifetime smoking and distal colon cancer after accounting for educational attainment in multivariable MR models (OR, 1.45; 95% CI, 1.10–1.91; $P = 0.01$). Multivariable MR analyses, adjusting for age at menarche, resulted in similar effect estimates to the IVW models for breast cancer (OR, 1.12; 95% CI, 0.99–1.26; $P = 0.06$), $ER^+$ breast cancer (OR, 1.11; 95% CI, 0.97–1.26; $P = 0.13$), and $ER^-$ breast cancer (OR, 1.13; 95% CI, 0.94–1.36; $P = 0.18$; Supplementary Table S7).

### MR estimates for ever having smoked regularly

In the IVW model, no association was found between genetic predisposition to ever having smoked regularly and risks of breast cancer (OR, 1.01; 95% CI, 0.90–1.14; $P = 0.85$), $ER^+$ breast cancer (OR, 1.01; 95% CI, 0.88–1.15; $P = 0.88$), and $ER^-$ breast cancer (OR, 1.03; 95% CI, 0.87–1.21; $P = 0.74$; **Table 2**). For colorectal cancer, we found no association between genetic predisposition to ever having smoked regularly and risks of colorectal cancer (OR, 0.97; 95% CI, 0.86–1.10; $P = 0.68$), colon cancer (OR, 0.99; 95% CI, 0.86–1.13; $P = 0.84$) and rectal cancer (OR, 1.00; 95% CI, 0.83–1.20; $P = 0.99$; **Table 2**). Similar null results were found for distal colon cancer (OR, 1.00; 95% CI, 0.85–1.18; $P = 0.99$), proximal colon cancer (OR, 0.97; 95% CI, 0.82–1.15; $P = 0.73$), and for colorectal cancer in women (OR, 0.94; 95% CI, 0.80–1.09; $P = 0.40$) and men (OR, 1.02; 95% CI, 0.88–1.18; $P = 0.80$).

For all colorectal and breast cancer models, estimates from MR-Egger and the weighed median approach did not provide any evidence of a causal effect. In all analyses, there was no evidence of aggregated directional pleiotropy using MR-Egger ($P_{pleiotropy} > 0.05$). We detected heterogeneity among the causal estimates of the 112 index variants (all $P_{heterogeneity} < 0.01$ for colorectal and breast cancer). In the sensitivity analyses using MR-PRESSO or removing one variant at a time, the results for all analyses were near identical (**Table 2**; Supplementary Tables S5B and S6B). The multivariable analyses, adjusting for BMI and alcohol consumption, showed no association for ever having smoked regularly with breast cancer (OR, 1.12; 95% CI, 0.94–1.33; $P = 0.20$), $ER^+$ breast cancer (OR, 1.08; 95% CI, 0.89–1.31; $P = 0.43$), and $ER^-$ breast cancer (OR, 1.16; 95% CI, 0.92–1.46; $P = 0.21$; **Fig. 1B**). No association was observed for ever having smoked regularly and risk of breast cancer after multivariable adjustment for educational attainment and age at menarche (Supplementary Table S8). Similarly, null findings were observed for ever having smoked regularly and risk of colorectal cancer after multivariable adjustment for BMI and alcohol, with no apparent differential effects of these two exposures. Ever having smoked regularly was not related to risk of colorectal cancer after multivariable adjustment for educational attainment (Supplementary Table S8).

In the positive control analysis, assessing the association between genetic predisposition to ever having smoked regularly and lung cancer, we found a strong positive effect estimate (OR, 1.90; 95% CI, 1.56–2.31; $P < 0.01$), demonstrating the validity of the genetic instrument (Supplementary Table S4).

## Discussion

In this MR analysis, genetic predisposition to higher lifetime amount of smoking was associated with elevated risks of breast cancer and colorectal cancer. In contrast, no association was found between

**Table 1.** Mendelian randomization estimates for the causal effect of lifetime amount of smoking on cancer risk.

| Methods | OR (95% CI) | *P* | *P* value for pleiotropy or heterogeneity[a] |
|---|---|---|---|
| Breast cancer | | | |
| IVW | 1.13 (1.00–1.26) | 0.04 | <0.01 |
| MR-Egger | 1.42 (0.90–2.24) | 0.13 | 0.30 |
| MR-Egger intercept | −0.004 (−0.010–0.003) | | |
| Weighted median | 1.16 (1.02–1.34) | 0.03 | NA |
| MR PRESSO (rs2867112)[b] | 1.15 (1.03–1.28) | 0.01 | <0.01 |
| Multivariable IVW (BMI) | 1.25 (1.08–1.45) | <0.01 | <0.01 |
| Multivariable IVW (alcohol consumption) | 1.12 (0.97–1.29) | 0.13 | <0.01 |
| Multivariable IVW (BMI, alcohol consumption) | 1.28 (1.09–1.49) | <0.01 | <0.01 |
| Breast cancer $ER^+$ subset | | | |
| IVW | 1.11 (0.97–1.26) | 0.12 | <0.01 |
| MR-Egger | 1.67 (1.00–2.80) | 0.05 | 0.11 |
| MR-Egger intercept | −0.006 (−0.014–0.001) | | |
| Weighted median | 1.12 (0.95–1.31) | 0.17 | NA |
| MR PRESSO (rs11783093)[b] | 1.08 (0.96–1.23) | 0.21 | <0.01 |
| Multivariable IVW (BMI) | 1.23 (1.05–1.44) | 0.01 | <0.01 |
| Multivariable IVW (alcohol consumption) | 1.11 (0.95–1.31) | 0.20 | <0.01 |
| Multivariable IVW (BMI, alcohol consumption) | 1.25 (1.05–1.49) | 0.01 | <0.01 |
| Breast cancer $ER^-$ subset | | | |
| IVW | 1.15 (0.96–1.37) | 0.14 | <0.01 |
| MR-Egger | 0.70 (0.34–1.45) | 0.33 | 0.17 |
| MR-Egger intercept | 0.008 (−0.003–0.018) | | |
| Weighted median | 1.21 (0.95–1.54) | 0.12 | NA |
| MR PRESSO (rs2867112)[b] | 1.20 (1.01–1.42) | 0.04 | 0.05 |
| Multivariable IVW (BMI) | 1.33 (1.06–1.68) | 0.01 | <0.01 |
| Multivariable IVW (alcohol consumption) | 1.17 (0.96–1.43) | 0.13 | <0.01 |
| Multivariable IVW (BMI, alcohol consumption) | 1.45 (1.14–1.85) | <0.01 | <0.01 |
| Colorectal cancer | | | |
| IVW | 1.21 (1.04–1.40) | 0.01 | 0.01 |
| MR-Egger | 1.58 (0.87–2.89) | 0.13 | 0.36 |
| MR-Egger intercept | −0.004 (−0.013–0.005) | | |
| Weighted median | 1.23 (1.01–1.50) | 0.04 | NA |
| MR PRESSO[b] | 1.21 (1.04–1.40) | 0.01 | 0.01 |
| Multivariable IVW (BMI) | 1.15 (0.97–1.36) | 0.12 | <0.01 |
| Multivariable IVW (alcohol consumption) | 1.23 (1.04–1.45) | 0.02 | 0.01 |
| Multivariable IVW (BMI, alcohol consumption) | 1.18 (0.98–1.43) | 0.09 | <0.01 |
| Colorectal cancer (women) | | | |
| IVW | 1.11 (0.90–1.37) | 0.34 | <0.01 |
| MR-Egger | 1.96 (0.83–4.66) | 0.12 | 0.18 |
| MR-Egger intercept | −0.009 (−0.022–0.004) | | |
| Weighted median | 1.08 (0.82–1.44) | 0.59 | NA |
| MR PRESSO (rs62175972)[b] | 1.08 (0.88–1.32) | 0.47 | 0.02 |
| Multivariable IVW (BMI) | 1.00 (0.79–1.27) | 0.99 | 0.01 |
| Multivariable IVW (alcohol consumption) | 1.11 (0.87–1.41) | 0.39 | <0.01 |
| Multivariable IVW (BMI, alcohol consumption) | 1.00 (0.78–1.30) | 0.98 | 0.01 |
| Colorectal cancer (men) | | | |
| IVW | 1.32 (1.11–1.59) | <0.01 | 0.52 |
| MR-Egger | 1.17 (0.56–2.44) | 0.67 | 0.74 |
| MR-Egger intercept | 0.002 (−0.009–0.013) | | |
| Weighted median | 1.36 (1.03–1.79) | 0.03 | NA |
| MR PRESSO[b] | 1.32 (1.11–1.59) | <0.01 | 0.52 |
| Multivariable IVW (BMI) | 1.27 (1.02–1.58) | 0.03 | 0.18 |
| Multivariable IVW (alcohol consumption) | 1.35 (1.10–1.65) | <0.01 | 0.50 |
| Multivariable IVW (BMI, alcohol consumption) | 1.26 (0.99–1.60) | 0.06 | 0.15 |
| Colon cancer | | | |
| IVW | 1.31 (1.11–1.55) | <0.01 | 0.08 |
| MR-Egger | 1.57 (0.80–3.09) | 0.19 | 0.59 |
| MR-Egger intercept | −0.003 (−0.013–0.007) | | |
| Weighted median | 1.37 (1.09–1.72) | 0.01 | NA |
| MR PRESSO[b] | 1.31 (1.11–1.55) | <0.01 | 0.08 |
| Multivariable IVW (BMI) | 1.23 (1.02–1.47) | 0.03 | 0.09 |
| Multivariable IVW (alcohol consumption) | 1.30 (1.07–1.57) | 0.01 | 0.02 |
| Multivariable IVW (BMI, alcohol consumption) | 1.25 (1.02–1.53) | 0.03 | 0.04 |

*(Continued on the following page)*

**Table 1.** Mendelian randomization estimates for the causal effect of lifetime amount of smoking on cancer risk. (Cont'd)

| Methods | OR (95% CI) | *P* | *P* value for pleiotropy or heterogeneity[a] |
|---|---|---|---|
| Distal colon cancer | | | |
| IVW | 1.50 (1.21–1.87) | <0.01 | 0.05 |
| MR-Egger | 1.92 (0.79–4.70) | 0.15 | 0.57 |
| MR-Egger intercept | −0.004 (−0.017–0.010) | | |
| Weighted median | 1.42 (1.04–1.93) | 0.03 | NA |
| MR PRESSO[b] | 1.50 (1.21–1.87) | <0.01 | 0.05 |
| Multivariable IVW (BMI) | 1.39 (1.10–1.75) | <0.01 | 0.17 |
| Multivariable IVW (alcohol consumption) | 1.44 (1.12–1.85) | <0.01 | 0.01 |
| Multivariable IVW (BMI, alcohol consumption) | 1.39 (1.07–1.80) | 0.01 | 0.12 |
| Proximal colon cancer | | | |
| IVW | 1.19 (0.98–1.46) | 0.09 | 0.25 |
| MR-Egger | 1.06 (0.47–2.41) | 0.88 | 0.77 |
| MR-Egger intercept | 0.002 (−0.010–0.014) | | |
| Weighted median | 1.23 (0.92–1.65) | 0.17 | NA |
| MR PRESSO[b] | 1.19 (0.98–1.46) | 0.09 | 0.25 |
| Multivariable IVW (BMI) | 1.11 (0.89–1.40) | 0.35 | 0.11 |
| Multivariable IVW (alcohol consumption) | 1.24 (0.98–1.56) | 0.07 | 0.13 |
| Multivariable IVW (BMI, alcohol consumption) | 1.17 (0.91–1.51) | 0.22 | 0.07 |
| Rectal cancer | | | |
| IVW | 1.36 (1.07–1.73) | 0.01 | <0.01 |
| MR-Egger | 1.11 (0.42–2.93) | 0.84 | 0.67 |
| MR-Egger intercept | 0.003 (−0.011–0.018) | | |
| Weighted median | 1.31 (0.95–1.81) | 0.10 | NA |
| MR PRESSO[b] | 1.36 (1.07–1.73) | 0.01 | <0.01 |
| Multivariable IVW (BMI) | 1.38 (1.06–1.79) | 0.02 | <0.01 |
| Multivariable IVW (alcohol consumption) | 1.40 (1.07–1.82) | 0.01 | <0.01 |
| Multivariable IVW (BMI, alcohol consumption) | 1.33 (0.99–1.79) | 0.05 | <0.01 |

Abbreviations: BMI, body mass index; CI, confidence interval; ER, estrogen receptor; IVW, Inverse-variance weighted; OR, odds ratio.
[a]*P* value for pleiotropy in MR-Egger regression; P-value for heterogeneity in inverse-variance weighted analysis.
[b]Variants in the parenthesis of the MR PRESSO method were identified as outlying and excluded.

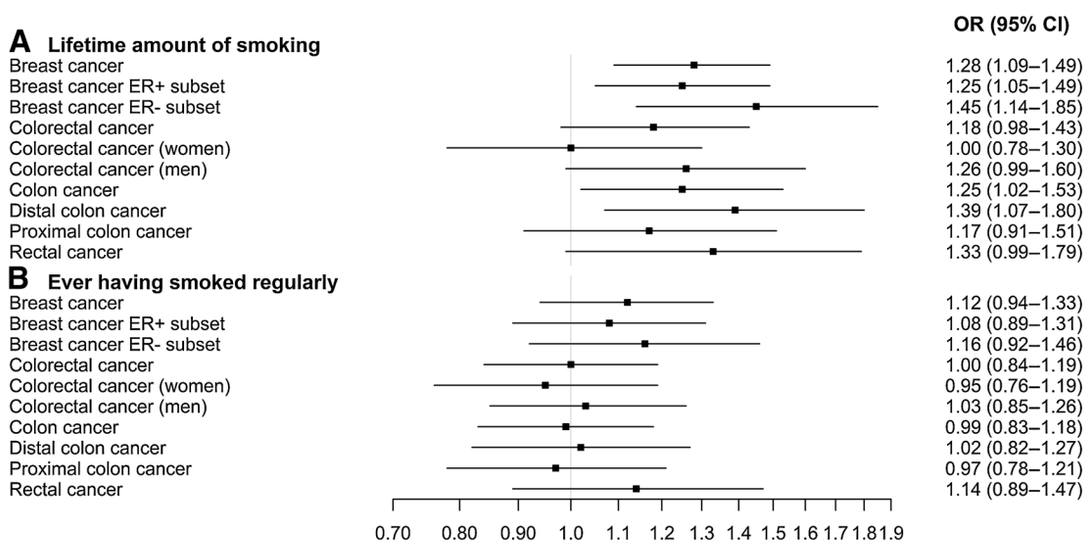


**Figure 1.**
Multivariable MR estimates, adjusted for body mass index and alcohol consumption, for the causal effect of lifetime amount of smoking (**A**) and ever having smoked regularly (**B**) on cancer risk. BMI, body mass index; ER, estrogen receptor.

**Table 2.** Mendelian randomization estimates for the causal effect of ever having smoked regularly on cancer risk.

| Methods | OR (95% CI) | *P* | *P* value for pleiotropy or heterogeneity[a] |
|---|---|---|---|
| Breast cancer | | | |
| IVW | 1.01 (0.90–1.14) | 0.85 | <0.01 |
| MR-Egger | 1.31 (0.82–2.10) | 0.25 | 0.26 |
| MR-Egger intercept | −0.006 (−0.016–0.004) | | |
| Weighted median | 1.04 (0.94–1.15) | 0.50 | NA |
| MR PRESSO (rs6731872, rs62246017, rs7072776, and rs8103660) | 1.02 (0.94–1.10) | 0.64 | <0.01 |
| Multivariable IVW (BMI) | 1.05 (0.90–1.22) | 0.52 | <0.01 |
| Multivariable IVW (alcohol consumption) | 1.05 (0.91–1.21) | 0.48 | <0.01 |
| Multivariable IVW (BMI, alcohol consumption) | 1.12 (0.94–1.33) | 0.20 | <0.01 |
| Breast cancer $ER^+$ subset | | | |
| IVW | 1.01 (0.88–1.15) | 0.88 | <0.01 |
| MR-Egger | 1.58 (0.94–2.66) | 0.08 | 0.08 |
| MR-Egger intercept | −0.010 (−0.021–0.001) | | |
| Weighted median | 1.00 (0.89–1.13) | 0.95 | NA |
| MR PRESSO (rs62246017, rs11783093, rs7072776, and rs8103660) | 0.96 (0.88–1.05) | 0.40 | 0.01 |
| Multivariable IVW (BMI) | 1.04 (0.88–1.23) | 0.66 | <0.01 |
| Multivariable IVW (alcohol consumption) | 1.05 (0.89–1.23) | 0.56 | <0.01 |
| Multivariable IVW (BMI, alcohol consumption) | 1.08 (0.89–1.31) | 0.43 | <0.01 |
| Breast cancer $ER^-$ subset | | | |
| IVW | 1.03 (0.87–1.21) | 0.74 | <0.01 |
| MR-Egger | 0.76 (0.39–1.47) | 0.41 | 0.35 |
| MR-Egger intercept | 0.007 (−0.007–0.021) | | |
| Weighted median | 1.05 (0.88–1.26) | 0.59 | NA |
| MR PRESSO (rs12739243, rs6731872, rs1549979, rs7696257, and rs8103660) | 1.08 (0.94–1.24) | 0.27 | <0.01 |
| Multivariable IVW (BMI) | 1.07 (0.88–1.30) | 0.51 | <0.01 |
| Multivariable IVW (alcohol consumption) | 1.06 (0.87–1.29) | 0.59 | <0.01 |
| Multivariable IVW (BMI, alcohol consumption) | 1.16 (0.92–1.46) | 0.21 | <0.01 |
| Colorectal cancer | | | |
| IVW | 0.97 (0.86–1.10) | 0.68 | <0.01 |
| MR-Egger | 1.10 (0.67–1.79) | 0.71 | 0.63 |
| MR-Egger intercept | −0.003 (−0.013–0.008) | | |
| Weighted median | 0.97 (0.83–1.13) | 0.67 | NA |
| MR PRESSO (rs77215829) | 0.99 (0.88–1.12) | 0.87 | <0.01 |
| Multivariable IVW (BMI) | 1.00 (0.86–1.15) | 0.96 | <0.01 |
| Multivariable IVW (alcohol consumption) | 0.98 (0.85–1.14) | 0.82 | <0.01 |
| Multivariable IVW (BMI, alcohol consumption) | 1.00 (0.84–1.19) | 0.98 | <0.01 |
| Colorectal cancer (women) | | | |
| IVW | 0.94 (0.80–1.09) | 0.40 | <0.01 |
| MR-Egger | 1.30 (0.70–2.43) | 0.40 | 0.28 |
| MR-Egger intercept | −0.007 (−0.021–0.006) | | |
| Weighted median | 0.94 (0.77–1.15) | 0.56 | NA |
| MR PRESSO | 0.94 (0.80–1.09) | 0.40 | <0.01 |
| Multivariable IVW (BMI) | 0.92 (0.76–1.12) | 0.42 | <0.01 |
| Multivariable IVW (alcohol consumption) | 0.94 (0.78–1.14) | 0.53 | <0.01 |
| Multivariable IVW (BMI, alcohol consumption) | 0.95 (0.76–1.19) | 0.64 | <0.01 |
| Colorectal cancer (men) | | | |
| IVW | 1.02 (0.88–1.18) | 0.80 | 0.05 |
| MR-Egger | 0.93 (0.51–1.67) | 0.79 | 0.74 |
| MR-Egger intercept | 0.002 (−0.010–0.015) | | |
| Weighted median | 1.04 (0.84–1.28) | 0.71 | NA |
| MR PRESSO | 1.02 (0.88–1.18) | 0.80 | 0.05 |
| Multivariable IVW (BMI) | 1.05 (0.89–1.24) | 0.56 | 0.06 |
| Multivariable IVW (alcohol consumption) | 1.03 (0.86–1.23) | 0.76 | 0.05 |
| Multivariable IVW (BMI, alcohol consumption) | 1.03 (0.85–1.26) | 0.74 | 0.05 |
| Colon cancer | | | |
| IVW | 0.99 (0.86–1.13) | 0.84 | <0.01 |
| MR-Egger | 1.03 (0.60–1.75) | 0.92 | 0.88 |
| MR-Egger intercept | −0.001 (−0.012–0.011) | | |
| Weighted median | 1.02 (0.85–1.22) | 0.86 | NA |
| MR PRESSO (rs77215829) | 1.00 (0.88–1.14) | 0.95 | 0.02 |
| Multivariable IVW (BMI) | 1.00 (0.86–1.16) | 0.99 | <0.01 |
| Multivariable IVW (alcohol consumption) | 1.00 (0.85–1.18) | 0.99 | <0.01 |
| Multivariable IVW (BMI, alcohol consumption) | 0.99 (0.83–1.18) | 0.90 | <0.01 |

*(Continued on the following page)*

**Table 2.** Mendelian randomization estimates for the causal effect of ever having smoked regularly on cancer risk. (Cont'd)

| Methods | OR (95% CI) | *P* | *P* value for pleiotropy or heterogeneity[a] |
|---|---|---|---|
| Distal colon cancer | | | |
| IVW | 1.00 (0.85–1.18) | 0.99 | 0.03 |
| MR-Egger | 0.83 (0.43–1.60) | 0.58 | 0.57 |
| MR-Egger intercept | 0.004 (−0.010–0.018) | | |
| Weighted median | 0.97 (0.78–1.22) | 0.82 | NA |
| MR PRESSO (rs9627272)[b] | 0.98 (0.84–1.15) | 0.81 | 0.13 |
| Multivariable IVW (BMI) | 1.02 (0.85–1.22) | 0.87 | 0.04 |
| Multivariable IVW (alcohol consumption) | 1.01 (0.83–1.23) | 0.94 | 0.03 |
| Multivariable IVW (BMI, alcohol consumption) | 1.02 (0.82–1.27) | 0.84 | 0.04 |
| Proximal colon cancer | | | |
| IVW | 0.97 (0.82–1.15) | 0.73 | <0.01 |
| MR-Egger | 1.22 (0.63–2.38) | 0.55 | 0.48 |
| MR-Egger intercept | −0.005 (−0.019–0.009) | | |
| Weighted median | 0.98 (0.78–1.23) | 0.88 | NA |
| MR PRESSO[b] | 0.97 (0.82–1.15) | 0.73 | <0.01 |
| Multivariable IVW (BMI) | 0.97 (0.81–1.17) | 0.77 | <0.01 |
| Multivariable IVW (alcohol consumption) | 1.00 (0.82–1.22) | 0.99 | <0.01 |
| Multivariable IVW (BMI, alcohol consumption) | 0.97 (0.78–1.21) | 0.79 | <0.01 |
| Rectal cancer | | | |
| IVW | 1.00 (0.83–1.20) | 0.99 | <0.01 |
| MR-Egger | 1.00 (0.48–2.11) | 0.99 | 0.99 |
| MR-Egger intercept | 0.000 (−0.016–0.016) | | |
| Weighted median | 1.03 (0.81–1.30) | 0.82 | NA |
| MR PRESSO[b] | 1.00 (0.83–1.20) | 0.99 | <0.01 |
| Multivariable IVW (BMI) | 1.04 (0.84–1.29) | 0.70 | <0.01 |
| Multivariable IVW (alcohol consumption) | 1.08 (0.86–1.34) | 0.51 | <0.01 |
| Multivariable IVW (BMI, alcohol consumption) | 1.14 (0.89–1.47) | 0.29 | <0.01 |

Abbreviations: BMI, Body mass index; ER, estrogen receptor; IVW, Inverse-variance weighted.
[a]*P* value for pleiotropy in MR-Egger regression; *P* value for heterogeneity in inverse-variance weighted analysis.
[a]Variants in the parenthesis of the MR PRESSO method were identified as outlying and excluded.

genetic predisposition to ever having smoked regularly and risks of breast and colorectal cancer.

The association between smoking and breast cancer risk is unclear. Experimental evidence has shown that chemical compounds found in tobacco smoke can induce mammary cancers in rodents (24–26), and smoking-related DNA adducts and p53 mutation have been identified in breast tissue (27, 28). Results from recent pooled analyses of cohort studies supported these experimental studies, with positive associations found between smoking and breast cancer risk, especially for women who initiated smoking before first childbirth (4, 29–31). However, causal inference of the smoking and breast cancer relationship is hampered by the epidemiologic association being relatively weak [e.g., the relative risk for current vs. never smokers was estimated to range from 1.07 to 1.12 in recent pooled analyses (4, 30)] and the possibility that confounding by other breast cancer risk factors, such as alcohol consumption and BMI, may induce bias (4). MR analyses are less susceptible to conventional confounding (6) and our use of novel multivariable MR methods meant we were able to estimate the direct effect of genetic predisposition to ever having smoked regularly and lifetime amount of smoking on breast cancer risk conditional on alcohol consumption and BMI (as it is likely that genetic variants associated with smoking exert pleiotropic effects with other addictive behaviors; ref. 32). Importantly, for breast cancer, the positive associations for lifetime amount of smoking were strongest in our multivariable MR models where we adjusted for BMI and alcohol intake. In addition, the positive associations for lifetime amount of smoking and risk of breast cancer were robust in the multivariable models accounting for possible pleiotropic effects of educational attainment, an indicator of socioeconomic status, and age at menarche.

We found similar magnitude positive effect estimates between being genetically predisposed to lifetime amount of smoking and both $ER^+$ and $ER^-$ breast cancer. Recent prospective studies have generally found that smoking-related exposures were more consistently and strongly associated with $ER^+$ than $ER^-$ breast cancer risk (4).

Tobacco smoke contains many carcinogens, that can damage DNA and the colorectal mucosa (33, 34). Findings from an experimental study of mice found that a tobacco-specific nitrosamine derived from nicotine acts as an initiator of colorectal cancer development (35). The strongest positive effect estimates for genetic predisposition to lifetime amount of smoking we found were for distal colon cancer and rectal cancer. For both anatomic subsites, the positive associations remained stable in multivariable MR models that conditioned for BMI and alcohol consumption, two established colorectal cancer risk factors (36). There is conflicting observational evidence for the association between smoking and risk of colorectal cancer by anatomic subsite. An earlier meta-analysis of case–control and cohort studies reported stronger positive relationships for rectal cancer and proximal colon cancer (37). However, a recent joint Nurses' Health Study and Health Professionals Follow-up Study, found no heterogeneity for the positive association between pack-years of smoking and risks of distal colon, proximal colon, and rectal cancer (38).

The positive effect estimate we found between lifetime amount of smoking and colorectal cancer risk was most apparent for men. Evidence for heterogeneity by sex for the association between smoking

and colorectal cancer is mixed. It has previously been suggested that the effect of smoking is either limited or stronger in men than for women (39), partly attributed to interactions with endogenous sex steroid hormone levels and/or adiposity (5). Two older meta-analyses of prospective cohort studies have shown that risks for current smoking are higher in men than women (5, 40). However, subsequent studies have reported positive associations between smoking and colorectal cancer among women (41–45).

We found no association between ever having smoked regularly and risks of both breast and colorectal cancer. The genetic instrument we used for ever smoking was a binary phenotype indicating if an individual ever smoked regularly. Similar null associations for ever having smoked regularly and risks of breast and colorectal cancers were reported in a recent MR study that used UK Biobank and BCAC data (46). Our positive control analysis showed the validity of the genetic instrument for ever having smoking regularly with an expected strong and robust positive association found for lung cancer risk. However, compared with lung cancer, smoking is more weakly associated with risks of breast and colorectal cancer (4, 5, 30), and it is possible that the null results we found for genetic disposition to ever smoking regularly with these cancers are a consequence of this phenotype crudely capturing exposure to cigarette smoke, with intensity and duration not accounted for. In particular, observational evidence suggests that ever smokers who started to smoke more than 1 year after the first childbirth had no increased risk for breast cancer, while those who initiated smoking more than 10 years before their first childbirth had a 60% increased risk of breast cancer, compared with never smokers (29). For colorectal cancer, it has been found that pack-years or duration among ever smokers are the smoking variables most consistently associated with higher disease risk (47). In contrast, the genetic instrument for lifetime amount of smoking, which was positively associated with both breast and colorectal cancer risks in our analyses, combined information in a score on smoking duration, smoking initiation, and number of cigarettes per day. Because this index captures smoking duration (age at initiation and age at cessation in current and former smokers) and heaviness (number of cigarettes smoked per day), we did not perform separate MR analyses for age of initiation of regular smoking, smoking cessation, and number of cigarettes per day.

To our knowledge, this is the first MR study to report causal associations for a smoking phenotype with breast and colorectal cancers. We conducted multiple sensitivity analyses to assess, and adjust for, the influence of horizontal pleiotropy on our results and, overall, our results were robust. However, MR-Egger results for breast cancer are not consistent with those in the IVW models and Weighted median approach and are the least reliable due to low $I^2_{GX}$ statistic (50% for lifetime amount of smoking and 71% for of ever having smoked regularly), which tests the suitability of this method (48). Furthermore, the agnostic use of all smoking-related variants irrespective of their biological function could explain the heterogeneity observed in MR results; however, multivariable MR methodology was employed to account for possible pleiotropic effects. A limitation of our study was that our use of summary-level data meant we were unable to assess the associations by menopausal status (although ∼85% of breast cancer cases were post-menopausal), subgroups of other risk factors (e.g., BMI, exogenous hormone use), and conduct analyses for localized and advanced breast and colorectal cancers.

In summary, using a comprehensive MR analytic framework, we estimated potential causal associations between smoking and risks of breast cancer and colorectal cancer. Our findings are consistent with prior observational evidence and provide support for higher lifetime amount of smoking being a causal risk factor for these common malignancies.

## Authors' Disclosures

R.M. Martin reports grants from Cancer Research UK during the conduct of the study. R.T. Fortner reports that grants from German Cancer Aid and from German Ministry of Education and Research supported the conduct of EPIC Heidelberg. S.B. Gruber reports other from Brogent International LLC outside the submitted work. B. van Guelpen reports grants from Swedish Research Council, Swedish Cancer Society, Knut and Alice Wallenberg Foundation, Lion's Cancer Research Foundation at Umeå University, and Cancer Research Foundation in Northern Sweden during the conduct of the study. No disclosures were reported by the other authors.

## Disclaimers

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.

The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the UK National Institute for Health Research (NIHR) or the Department of Health and Social Care.

## Authors' Contributions

**N. Dimou:** Conceptualization, formal analysis, methodology, writing–original draft, writing–review and editing. **J. Yarmolinsky:** Methodology, writing–original draft, writing–review and editing. **E. Bouras:** Writing–review and editing. **K.K. Tsilidis:** Writing–review and editing. **R.M. Martin:** Writing–review and editing. **S.J. Lewis:** Writing–review and editing. **I.T. Gram:** Writing–review and editing. **M.F. Bakker:** Writing–review and editing. **H. Brenner:** Writing–review and editing. **J.C. Figueiredo:** Writing–review and editing. **R.T. Fortner:** Writing–review and editing. **S.B. Gruber:** Writing–review and editing. **B. van Guelpen:** Writing–review and editing. **L. Hsu:** Writing–review and editing. **R. Kaaks:** Writing–review and editing. **S.-S. Kweon:** Writing–review and editing. **Y. Lin:** Writing–review and editing. **N.M. Lindor:** Writing–review and editing. **P.A. Newcomb:** Writing–review and editing. **M.-J. Sánchez:** Writing–review and editing. **G. Severi:** Writing–review and editing. **H.A. Tindle:** Writing–review and editing. **R. Tumino:** Writing–review and editing. **E. Weiderpass:** Writing–review and editing. **M.J. Gunter:** Conceptualization, resources, supervision, funding acquisition, methodology, writing–original draft, writing–review and editing. **N. Murphy:** Conceptualization, resources, supervision, funding acquisition, validation, methodology, writing–original draft, writing–review and editing.

## Acknowledgments

ASTERISK: 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.

CLUE: We appreciate the continued efforts of the staff members at the Johns Hopkins George W. Comstock Center for Public Health Research and Prevention in the conduct of the CLUE II study. We thank the participants in CLUE. Cancer incidence data for CLUE were provided by the Maryland Cancer Registry, Center for Cancer Surveillance and Control, Maryland Department of Health, 201 W. Preston Street, Room 400, Baltimore, MD 21201, http://phpa.dhmh.maryland.gov/cancer, 410-767-4055. We acknowledge the State of Maryland, the Maryland Cigarette Restitution Fund, and the National Program of Cancer Registries of the Centers for Disease Control and Prevention for the funds that support the collection and availability of the cancer registry data.

COLON and NQplus: the authors would like to thank the COLON and NQplus investigators at Wageningen University & Research and the involved clinicians in the participating hospitals.

CORSA: We kindly thank all those who contributed to the screening project Burgenland against CRC. Furthermore, we are grateful to Doris Mejri and Monika Hunjadi for laboratory assistance.

CPS-II: 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.

Czech Republic CCS: 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: We thank all participants and cooperating clinicians, and Ute Handte-Daub, Utz Benscheid, MuhabbetCelik and Ursula Eilber for excellent technical assistance.

EDRN: We acknowledge all the following contributors to the development of the resource: University of Pittsburgh School of Medicine, Department of Gastroenterology, Hepatology and Nutrition: Lynda Dzubinski; University of Pittsburgh School of Medicine, Department of Pathology: Michelle Bisceglia; and University of Pittsburgh School of Medicine, Department of Biomedical Informatics.

EPIC: 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.

EPICOLON: We are sincerely grateful to all patients participating in this study who were recruited as part of the EPICOLON project. We acknowledge the Spanish National DNA Bank, Biobank of Hospital Clínic–IDIBAPS and Biobanco Vasco for the availability of the samples. The work was carried out (in part) at the Esther Koplowitz Centre, Barcelona.

Harvard cohorts (HPFS, NHS, PHS): 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.

Kentucky: We would like to acknowledge the staff at the Kentucky Cancer Registry.

LCCS: We acknowledge the contributions of Jennifer Barrett, Robin Waxman, Gillian Smith and Emma Northwood in conducting this study.

NCCCS I & II: We would like to thank the study participants, and the NC Colorectal Cancer Study staff.

NSHDS investigators thank the Biobank Research Unit at Umeå University, the Västerbotten Intervention Programme, the Northern Sweden MONICA study and Region Västerbotten for providing data and samples and acknowledge the contribution from Biobank Sweden, supported by the Swedish Research Council (VR 2017-00650).

PLCO: 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.

SCCFR: The authors would like to thank the study participants and staff of the Hormones and Colon Cancer and Seattle Cancer Family Registry studies (CORE Studies).

SEARCH: We thank the SEARCH team.

SELECT: 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.

WHI: The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: http://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Short%20List.pdf

We acknowledge the National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands, for their contribution and ongoing support to the EPIC Study.

Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO): National Cancer Institute, National Institutes of Health, U.S. Department of Health and Human Services (U01 CA137088, R01 CA059045, R01CA201407). This research was funded in part through the NIH/NCI Cancer Center Support Grant P30 CA015704.

ASTERISK: a Hospital Clinical Research Program (PHRC-BRD09/C) from the University Hospital Center of Nantes (CHU de Nantes) and supported by the Regional Council of Pays de la Loire, the Groupement des EntreprisesFrançaises dans la Luttecontre le Cancer (GEFLUC), the Association Anne de Bretagne Génétique and the Ligue RégionaleContre le Cancer (LRCC).

The ATBC Study is supported by the Intramural Research Program of the U.S. National Cancer Institute, National Institutes of Health, and by U.S. Public Health Service contract HHSN261201500005C from the National Cancer Institute, Department of Health and Human Services.

CLUE funding was from 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. COLO2&3: NIH (R01 CA60987).

ColoCare: This work was supported by the National Institutes of Health [grant numbers R01 CA189184 (Li/Ulrich), U01 CA206110 (Ulrich/Li/Siegel/Figueiredo/Colditz), 2P30CA015704- 40 (Gilliland), R01 CA207371 (Ulrich/Li)], the Matthias Lackas-Foundation, the German Consortium for Translational Cancer Research, and the EU TRANSCAN initiative.

The Colon Cancer Family Registry (CCFR, www.coloncfr.org) is supported in part by funding from the National Cancer Institute (NCI), National Institutes of Health (NIH) (award U01 CA167551). 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 GC). The CCFR Set-3 (Affymetrix Axiom CORECT Set array) was supported by NIH award U19 CA148107 and R01 CA81488 (to SBG). The CCFR Set-4 (Illumina OncoArray 600K SNP array) was supported by NIH award U19 CA148107 (to SBG) and by the Center for Inherited Disease Research (CIDR), which is funded by the NIH to the Johns Hopkins University, contract number HHSN268201200008I. 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.

COLON: The COLON study is sponsored by WereldKankerOnderzoek Fonds, including funds from grant 2014/1179 as part of the World Cancer Research Fund International Regular Grant Programme, by Alped'Huzes and the Dutch Cancer Society (UM 2012–5653, UW 2013-5927, UW2015-7946), and by TRANSCAN (JTC2012-MetaboCCC, JTC2013-FOCUS). The Nqplus study is sponsored by a ZonMW investment grant (98-10030); by PREVIEW, the project PREVention of diabetes through lifestyle intervention and population studies in Europe and around the World (PREVIEW) project which received funding from the European Union Seventh Framework Programme (FP7/2007–2013) under grant no. 312057; by funds from TI Food and Nutrition (cardiovascular health theme), a public–private partnership on precompetitive research in food and nutrition; and by FOODBALL, the Food Biomarker Alliance, a project from JPI Healthy Diet for a Healthy Life.

Colorectal Cancer Transdisciplinary (CORECT) Study: The CORECT Study was supported by the 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, NIH (grant number T32 ES013678).

CORSA: "ÖsterreichischeNationalbankJubiläumsfondsprojekt" (12511) and Austrian Research Funding Agency (FFG) grant 829675.

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.

CRCGEN: Colorectal Cancer Genetics & Genomics, Spanish study was 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), PlataformaBiobancos PT13/0010/0013 and ICO-BIOBANC, sponsored by the Catalan Institute of Oncology.

Czech Republic CCS: This work was supported by the Grant Agency of the Czech Republic (grants CZ GA CR: GAP304/10/1286 and 1585) and by the Grant Agency of the Ministry of Health of the Czech Republic (grants AZV 15-27580A and AZV 17-30920A).

DACHS: This work was 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 BR1704/17-1), 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).

DALS: NIH (R01 CA48998, to M.L. Slattery).

EDRN: This work is funded and supported by the NCI, EDRN grant (U01 CA 84968-06).

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 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 PotsdamRehbruecke (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 County Councils 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).

EPICOLON: This work was supported by grants from Fondo de Investigación Sanitaria/FEDER (PI08/0024, PI08/1276, PS09/02368, P111/00219, PI11/00681, PI14/00173, PI14/00230, PI17/00509, 17/00878, Acción Transversal de Cáncer), Xunta de Galicia (PGIDIT07PXIB9101209PR), Ministerio de Economia y Competitividad (SAF07-64873, SAF 2010-19273, SAF2014-54453R), Fundación Científica de la Asociación Española contra el Cáncer (GCB13131592CAST), Beca Grupo de Trabajo "Oncología" AEG (Asociación Española de Gastroenterología), Fundación Privada Olga Torres, FP7 CHIBCHA Consortium, Agència de Gestiód'AjutsUniversitaris i de Recerca (AGAUR, Generalitat de Catalunya, 2014SGR135, 2014SGR255, 2017SGR21, 2017SGR653), Catalan Tumour Bank Network (Pla Director d'Oncologia, Generalitat de Catalunya), PERIS (SLT002/16/00398, Generalitat de Catalunya), CERCA Programme (Generalitat de Catalunya) and COST Action BM1206 and CA17118. CIBERehd is funded by the Instituto de Salud Carlos III.

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

Harvard cohorts (HPFS, NHS, PHS): HPFS is supported by the National Institutes of Health (P01 CA055075, UM1 CA167552, U01 CA167552, R01 CA137178, R01 CA151993, R35CA197735, K07 CA190673, and P50 CA127003), NHS by the National Institutes of Health (R01 CA137178, P01 CA087969, UM1 CA186107, R01 CA151993, R35 CA197735, K07CA190673, and P50 CA127003) and PHS by the National Institutes of Health (R01 CA042182).

Hawaii Adenoma Study: NCI grants R01 CA72520.

HCES-CRC: the Hwasun Cancer Epidemiology Study–Colon and Rectum Cancer (HCES-CRC; grants from Chonnam National University Hwasun Hospital, HCRI15011-1).

Kentucky: This work was supported by the following grant support: Clinical Investigator Award from Damon Runyon Cancer Research Foundation (CI-8); NCI R01CA136726.

LCCS: The Leeds Colorectal Cancer Study was funded by the Food Standards Agency and Cancer Research UK Programme Award (C588/A19167).

MCCS cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further supported by Australian NHMRC grants 509348, 209057, 251553 and 504711 and by infrastructure provided by Cancer Council Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry (VCR) and the Australian Institute of Health and Welfare (AIHW), including the National Death Index and the Australian Cancer Database.

MEC: NIH (R37 CA54281, P01 CA033619, and R01 CA063464).

MECC: This work was supported by the NIH, U.S. Department of Health and Human Services (R01 CA81488 to SBG and GR).

MSKCC: The work at Sloan Kettering in New York was supported by the Robert and Kate Niehaus Center for Inherited Cancer Genomics and the Romeo Milio Foundation. Moffitt: This work was supported by funding from the NIH (grant numbers R01 CA189184, P30 CA076292), Florida Department of Health Bankhead-Coley Grant 09BN-13, and the University of South Florida Oehler Foundation. Moffitt contributions were supported in part by the Total Cancer Care Initiative, Collaborative Data Services Core, and Tissue Core at the H. Lee Moffitt Cancer Center & Research Institute, a NCI-designated Comprehensive Cancer Center (grant number P30 CA076292).

NCCCS I & II: We acknowledge funding support for this project from the NIH, R01 CA66635 and P30 DK034987.

NFCCR: This work was supported by an Interdisciplinary Health Research Team award from the Canadian Institutes of Health Research (CRT 43821); the NIH, U.S. Department of Health and Human Services (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 Michael O. Woods by the Canadian Cancer Society Research Institute.

NSHDS: Swedish Research Council; Swedish Cancer Society; Cutting-Edge Research Grant and other grants from Region Västerbotten; Knut and Alice Wallenberg Foundation; Lion's Cancer Research Foundation at Umeå University; the Cancer Research Foundation in Northern Sweden; and the Faculty of Medicine, Umeå University, Umeå, Sweden.

OFCCR: The Ontario Familial Colorectal Cancer Registry was supported in part by the NCI of the NIH under award U01 CA167551 and award U01/U24 CA074783 (to S.B. Gruber). Additional funding for the OFCCR and ARCTIC testing and genetic analysis was through and a Canadian Cancer Society CaRE (Cancer Risk Evaluation) program grant and Ontario Research Fund award GL201-043 (to BWZ), through the Canadian Institutes of Health Research award 112746 (to TJH), and through generous support from the Ontario Ministry of Research and Innovation.OSUMC: OCCPI funding was provided by Pelotonia and HNPCC funding was provided by the NCI (CA16058 and CA67941).

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 by NIH, Genes, Environment and Health Initiative (GEI) Z01 CP 010200, NIH U01 HG004446, and NIH GEI U01 HG 004438.

SCCFR: The Seattle Colon Cancer Family Registry was supported in part by the NCI of the NIH under awards U01 CA167551, U01 CA074794 (to JDP), and awards U24 CA074794 and R01 CA076366 (to PAN).

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.

SELECT: Research reported in this publication was supported in part by the NCI of the NIH under Award Numbers U10 CA37429 (to C.D. Blanke), and UM1 CA182883 (to C.M. Tangen/I.M. Thompson). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

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

The Swedish Low-risk Colorectal Cancer Study: The study was supported by grants from the Swedish research council; K2015-55X-22674-01-4, K2008-55X-20157-03-3, K2006-72X-20157-01-2 and the Stockholm County Council (ALF project).

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 Alicja Wolk.

VITAL: NIH (K05 CA154337).

WHI: The WHI program is funded by the National Heart, Lung, and Blood Institute, NIH, U.S. Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C.

R.M. Martin is supported in part by the National Institute for Health Research Bristol Biomedical Research Centre. The CAP trial is funded by Cancer Research UK and the UK Department of Health (C11043/A4286, C18281/A8145, C18281/A11326, C18281/A15064, and C18281/A24432).

R.M. Martin and S.J. Lewis were supported by a Cancer Research UK (C18281/A29019) programme grant (the Integrative Cancer Epidemiology Programme).

J. Yarmolinsky is supported by a Cancer Research UK Population Research Postdoctoral Fellowship (C68933/A28534).

This work was supported by Cancer Research UK (18281/A29019).

Received August 14, 2020; revised November 2, 2020; accepted February 23, 2021; published first March 2, 2021.

## 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. Cancer J Clin 2018;68:394–424.
2. Brown KF, Rumgay H, Dunlop C, Ryan M, Quartly F, Cox A, et al. The fraction of cancer attributable to modifiable risk factors in England, Wales, Scotland, Northern Ireland, and the United Kingdom in 2015. Br J Cancer 2018;118: 1130–41.
3. IARC Monographs Vol 127 group. Carcinogenicity of some aromatic amines and related compounds. Lancet Oncol 2020;S1470–2045:30375–2.
4. Gaudet MM, Carter BD, Brinton LA, Falk RT, Gram IT, Luo J, et al. Pooled analysis of active cigarette smoking and invasive breast cancer risk in 14 cohort studies. Int J Epidemiol 2017;46:881–93.
5. Tsoi KK, Pau CY, Wu WK, Chan FK, Griffiths S, Sung JJ. Cigarette smoking and the risk of colorectal cancer: a meta-analysis of prospective cohort studies. Clin Gastroenterol Hepatol 2009;7:682–8.
6. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol 2003;32:1–22.
7. Wootton RE, Richmond RC, Stuijfzand BG, Lawn RB, Sallis HM, Taylor GMJ, et al. Evidence for causal effects of lifetime smoking on risk for depression and schizophrenia: a Mendelian randomisation study. Psychol Med 2020;50:2435–43.
8. Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. Nat Genet 2019;51:237–44.
9. Michailidou K, Lindstrom S, Dennis J, Beesley J, Hui S, Kar S, et al. Association analysis identifies 65 new breast cancer risk loci. Nature 2017;551:92–4.
10. Huyghe JR, Bien SA, Harrison TA, Kang HM, Chen S, Schmit SL, et al. Discovery of common and rare genetic risk variants for colorectal cancer. Nat Genet 2019; 51:76–87.
11. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol 2013;37: 658–65.
12. Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG., Consortium E-I. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. Eur J Epidemiol 2015;30:543–52.
13. Greenland S. An introduction to instrumental variables for epidemiologists. Int J Epidemiol 2018;47:358.
14. Martens EP, Pestman WR, de Boer A, Belitser SV, Klungel OH. Instrumental variables: application and limitations. Epidemiology 2006;17:260–7.
15. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol 2015;44:512–25.
16. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol 2016;40:304–14.
17. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet 2018;50:693–8.
18. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, et al. Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. Hum Mol Genet 2018; 27:3641–9.
19. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature 2015;518:197–206.
20. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet 2018;50:1112–21.
21. Day FR, Thompson DJ, Helgason H, Chasman DI, Finucane H, Sulem P, et al. Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. Nat Genet 2017;49:834–41.
22. Wang Y, McKay JD, Rafnar T, Wang Z, Timofeeva MN, Broderick P, et al. Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. Nat Genet 2014;46:736–41.
23. Yavorska OO, Burgess S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. Int J Epidemiol 2017;46:1734–9.
24. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Tobacco smoke and involuntary smoking. IARC Monogr Eval Carcinog Risks Hum 2004;83:1–1438.
25. US Department of Health and Human Services. The Health Consequences of Smoking: A Report of the Surgeon General. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health; 2004.
26. Hecht SS. Tobacco smoke carcinogens and breast cancer. Environ Molr Mutagen 2002;39:119–26.
27. Conway K, Edmiston SN, Cui L, Drouin SS, Pang J, He M, et al. Prevalence and spectrum of p53 mutations associated with smoking in breast cancer. Cancer Res 2002;62:1987–95.
28. Li D, Zhang W, Sahin AA, Hittelman WN. DNA adducts in normal tissue adjacent to breast cancer: a review. Cancer Detect Prev 1999;23: 454–62.
29. Bjerkaas E, Parajuli R, Weiderpass E, Engeland A, Maskarinec G, Selmer R, et al. Smoking duration before first childbirth: an emerging risk factor for breast cancer? Results from 302,865 Norwegian women. Cancer Causes Control 2013; 24:1347–56.
30. Gaudet MM, Gapstur SM, Sun J, Diver WR, Hannan LM, Thun MJ. Active smoking and breast cancer risk: original cohort data and meta-analysis. J Natl Cancer Inst 2013;105:515–25.
31. Gram IT, Braaten T, Terry PD, Sasco AJ, Adami HO, Lund E, et al. Breast cancer risk among women who start smoking as teenagers. Cancer Epidemiol Biomarkers Prev 2005;14:61–6.
32. Burgess S, Thompson SG. Multivariable mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. Am J Epidemiol 2015;181: 251–60.
33. Hoffmann D, Djordjevic MV, Hoffmann I. The changing cigarette. Prev Med 1997;26:427–34.
34. Huang K, Sandler RS, Millikan RC, Schroeder JC, North KE, Hu J. GSTM1 and GSTT1 polymorphisms, cigarette smoking, and risk of colon cancer: a population-based case-control study in North Carolina (United States). Cancer Causes Control 2006;17:385–94.
35. Kim M, Miyamoto S, Sugie S, Yasui Y, Ishigamori-Suzuki R, Murakami A, et al. A tobacco-specific carcinogen, NNK, enhances AOM/DSS-induced colon carcinogenesis in male A/J mice. In Vivo 2008;22:557–63.

36. World Cancer Research Fund/American Insitute for Cancer Research. Continuous update project expert report 2018. Diet, nutrition, physical activity and colorectal cancer. Available from: dietandcancerreport.org.
37. Botteri E, Iodice S, Bagnardi V, Raimondi S, Lowenfels AB, Maisonneuve P. Smoking and colorectal cancer: a meta-analysis. JAMA 2008;300:2765–78.
38. Wang L, Lo CH, He X, Hang D, Wang M, Wu K, et al. Risk factor profiles differ for cancers of different regions of the colorectum. Gastroenterology 2020;159: 241–56.
39. Giovannucci E. An updated review of the epidemiological evidence that cigarette smoking increases risk of colorectal cancer. Cancer Epidemiol Biomarkers Prev 2001;10:725–31.
40. Huxley RR, Ansary-Moghaddam A, Clifton P, Czernichow S, Parr CL, Woodward M. The impact of dietary and lifestyle risk factors on risk of colorectal cancer: a quantitative overview of the epidemiological evidence. Int J Cancer 2009;125:171–80.
41. Gram IT, Braaten T, Lund E, Le Marchand L, Weiderpass E. Cigarette smoking and risk of colorectal cancer among Norwegian women. Cancer Causes Control 2009;20:895–903.
42. Hannan LM, Jacobs EJ, Thun MJ. The association between cigarette smoking and risk of colorectal cancer in a large prospective cohort from the United States. Cancer Epidemiol Biomarkers Prev 2009;18:3362–7.
43. Hurley S, Goldberg D, Nelson DO, Lu Y, Henderson K, Bernstein L, et al. Risk of colorectal cancer associated with active smoking among female teachers. Cancer Causes Control 2013;24:1291–304.
44. Leufkens AM, Van Duijnhoven FJ, Siersema PD, Boshuizen HC, Vrieling A, Agudo A, et al. Cigarette smoking and colorectal cancer risk in the European prospective investigation into cancer and nutrition study. Clin Gastroenterol Hepatol 2011;9:137–44.
45. Paskett ED, Reeves KW, Rohan TE, Allison MA, Williams CD, Messina CR, et al. Association between cigarette smoking and colorectal cancer in the women's health initiative. J Natl Cancer Inst 2007;99:1729–35.
46. Larsson SC, Carter P, Kar S, Vithayathil M, Mason AM, Michaelsson K, et al. Smoking, alcohol consumption, and cancer: a mendelian randomisation study in UK Biobank and international genetic consortia participants. PLoS Med 2020;17: e1003178.
47. Gram IT, Park SY, Wilkens LR, Haiman CA, Le Marchand L. Smoking-related risks of colorectal cancer by anatomical subsite and sex. Am J Epidemiol 2020; 189:543–53.
48. Bowden J, Del Greco MF, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for two-sample mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic. Int J Epidemiol 2016;45:1961–74.