ORIGINAL ARTICLE

Pleiotropic effects of genetic risk variants for other cancers on colorectal cancer risk: PAGE, GECCO and CCFR consortia


ABSTRACT

Objective Genome-wide association studies have identified a large number of single nucleotide polymorphisms (SNPs) associated with a wide array of cancer sites. Several of these variants demonstrate associations with multiple cancers, suggesting pleiotropic effects and shared biological mechanisms across some cancers. We hypothesised that SNPs previously associated with other cancers may additionally be associated with colorectal cancer. In a large-scale study, we examined 171 SNPs previously associated with 18 different cancers for their associations with colorectal cancer.

Design We examined 13 338 colorectal cancer cases and 40 967 controls from three consortia: Population Architecture using Genomics and Epidemiology (PAGE), Genetic Epidemiology of Colorectal Cancer (GECCO), and the Colon Cancer Family Registry (CCFR). Study-specific logistic regression results, adjusted for age, sex, principal components of genetic ancestry, and/or study specific factors (as relevant) were combined using fixed-effect meta-analyses to evaluate the association between each SNP and colorectal cancer risk. A Bonferroni-corrected p value of 2.92×10⁻⁴ was used to determine statistical significance of the associations.

Results Two correlated SNPs—rs10090154 and rs4242382—in Region 1 of chromosome 8q24, a prostate cancer susceptibility region, demonstrated statistically significant associations with colorectal cancer risk. The most significant association was observed with rs4242382 (meta-analysis OR=1.12; 95% CI 1.07 to 1.16).

Significance of this study

What is already known on this subject?

▸ Several hundred common genetic variants have been associated with a wide array of cancer types.
▸ Only a small proportion of the heritability of colorectal cancer can be explained by the currently identified risk loci from genome-wide association studies of colorectal cancer.
▸ Identifying shared genetic associations between diseases (ie, pleiotropy) is a useful approach to identify new risk loci, and may elucidate common etiologies and help in risk prediction.

What are the new findings?

▸ This study clearly shows that two genetic variants in Region 1 of the 8q24 locus, a prostate cancer risk region, are also associated with colorectal cancer risk.
▸ Furthermore, this study provides additional evidence that the telomerase reverse transcriptase locus is associated with colorectal cancer.

How might it impact on clinical practice in the foreseeable future?

▸ Colorectal risk variants may be used as part of a risk prediction model to define high-risk populations for targeted screening regimens and, possibly, inform clinical decision making.

INTRODUCTION
Since the first series of genome-wide association studies (GWAS) for cancer was published in 2007, several hundred common genetic variants have been associated with a wide array of cancer sites. As GWAS continue to identify variants associated with cancer, patterns of pleiotropic associations have emerged that highlight key loci and shared pathways that affect multiple cancer sites. For instance, genetic variants at chromosome 8q24 have been associated with cancers of the prostate, colorectum, breast, bladder and other sites. Similarly, genetic variants in and near the telomerase reverse transcriptase (TERT) gene, which encodes for telomerase activity, have been associated with glioma, lung, prostate, colorectal and other cancers, emphasising the importance of cellular ageing in cancer development.

Pleiotropy occurs when a genetic locus is associated with multiple phenotypic traits. Accordingly, any genetic difference at a pleiotropic locus may have wide-ranging effects across different cell types. Evidence of pleiotropic associations can improve our understanding of disease aetiology by identifying shared molecular components underlying disease risk and by validating the pathogenicity of variants at a locus. To illustrate, a recent study of the genetic overlap between systematic lupus erythematosus and other autoimmune diseases found novel pleiotropic associations that support a role for T cell and innate immune response pathways, providing valuable evidence for dissecting the biological mechanisms that underlie their shared aetiologies.

Previous analyses of shared genetic variants across cancers have focused primarily on hereditary disorders, such as the Lynch and Li-Fraumeni syndromes. Although multiple cancer types are known to cluster within families, studies of shared genetic factors across various non-familial cancers have been limited. Given the numerous associations reported by GWAS of cancer, we now have an opportunity to assess pleiotropy across different cancers. These pleiotropic associations may have been missed in prior GWAS of colorectal cancer (CRC) due to smaller sample sizes, and the stringent threshold of significance of testing hundreds of thousands to millions of single nucleotide polymorphisms (SNPs) in GWAS. For this study, we tested GWAS-identified risk variants of 18 other cancers for pleiotropic associations with CRC risk in a large-scale collaboration, including multiple racial/ethnic groups. Specifically, we conducted a meta-analysis study of 13,338 CRC cases and 40,967 controls from 16 studies of three consortia: Population Architecture using Genomics and Epidemiology (PAGE); Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO); and the Colon Cancer Family Registry (CCFR).

METHODS
Study participants
Three consortia contributed data to this meta-analysis study: PAGE15 GECCO11,16 and CCFR. This collaboration comprised 13,338 CRC cases and 40,967 controls from 16 studies (see online supplementary table S1). Briefly, PAGE studies included: Atherosclerosis Risk in Communities (ARIC), which is part of Causal Variants Across the Life Course (CALiCo); Epidemiologic Architecture for Genes Linked to Environment, which accesses the Vanderbilt University biorepository (EAGLE-BioVU);19 Multiethnic Cohort (MEC);20 and Women’s Health Initiative (WHI). GECCO studies included: French Association Study Evaluating RISK for sporadic CRC (ASTERISK);21 Hawaii Colorectal Cancer Studies 2 & 3 (Colo2&3);22 Darmkrebs: Chancen der Verhütung durch Screening (DACHS);23 Diet, Activity, and Lifestyle Study (DALS);24 Health Professionals Follow-up Study (HPFS);24 Nurses’ Health Study (NHS); Ontario Familial Colorectal Cancer Registry (OFCCR);25 Physicians’ Health Study (PHS);27 Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO);28 Post-Menopausal Hormones Supplemental Study to the CCFR (PMH-CCFR);10 Vitamins And Lifestyle (VITAL) study;31 and WHI.32 33 While WHI participates in both PAGE and GECCO, only WHI data as a part of GECCO was used. CCFR17 included a population-based case-control subset.

SNP selection and genotyping
A total of 171 SNPs previously associated with 18 cancers other than CRC were selected by PAGE researchers (see online supplementary table S3). These SNPs were identified to be associated with cancer, as of January 2010, from the National Human Genome Research Institute (NHGRI) GWAS catalogue (http://www.genome.gov/26525384) as well as review of the cancer GWAS and fine-mapping literature.15 References for each selected SNP are provided in online supplementary table S3. The majority of studies used incident CRC cases; controls had no diagnosis of CRC. Six GECCO studies (DACHS, DALS, HPFS, NHS, PLCO and WHI) contained study-specific subsets that were genotyped and analysed individually due to differences in sample collection, year of ascertainment, or controls used for each subset (see online supplementary material; supplementary table S2). This led to a total of 22 analytic subsets from the 16 studies. Supplementary figure S1 shows the participating studies and overall study design. Institutional review board approval was obtained for all studies.

To control for potential bias due to population stratification (ie, confounding due to racial/ethnic differences in allele frequencies and disease risk), 128 ancestry informative markers that capture the major continental genetic diversity were genotyped in all PAGE studies with the exception of ARIC. In ARIC, GECCO and CCFR, genotype data were abstracted from previously generated GWAS data.

Control for potential bias due to population stratification (ie, confounding due to racial/ethnic differences in allele frequencies and disease risk), 128 ancestry informative markers that capture the major continental genetic diversity were genotyped in all PAGE studies with the exception of ARIC. In ARIC, GECCO and CCFR, genotype data were abstracted from previously generated GWAS data.
estimate genotypes for untyped SNPs based on shared haplotypes and correlation with genotyped SNPs. Standard quality-assurance and quality-control measures were used to ensure genotyping quality. Further details are provided in the online supplementary material. The majority of the 171 SNPs of interest were available across studies (97% SNPs were genotyped or imputed in all 22 analytic study sets; see online supplementary table S3).

### Statistical analyses

For each study, the association between each SNP and CRC was estimated using unconditional logistic regression. SNPs were coded additively with 0, 1, 2 referring to the number of risk alleles (or the allele dosage for imputed SNPs). Primary models were adjusted for age, sex and the most relevant principal components of genetic ancestry to account for relevant population substructure for each study. A few studies were additionally adjusted for study centre (CCFR, DALS, PLCO and DACHS), study component (WHI), smoking (PHS), or batch effects (ASTERISK). To examine patterns of associations across race/ethnicity, each study with at least 80 CRC cases per race/ethnicity conducted analyses stratified by racial/ethnic population. Polytomous unconditional logistic regression was also performed in each study to examine associations across anatomical subsite (colon vs rectum). This method allowed us to simultaneously examine the associations for colon and rectal cancer in a single regression model, providing an efficient approach and the ability to test for heterogeneity in effects by anatomical subsite.

To examine whether the top associations found for the prostate cancer risk variants at Region 1 of chromosome 8q24 were independent from Region 3, an established colorectal risk region at 8q24, rs6983267 (a Region 3 CRC risk variant; meta-analysis OR=1.14; p=5×10\(^{-14}\)) was included in the regression model with each Region 1 prostate cancer risk variant.

Log odds regression estimates were combined across studies using inverse-variance weighted, fixed-effect meta-analysis in METAL\(^{16}\) for overall and stratified analyses. Heterogeneity p values were estimated based on Cochran’s Q statistic. SNP associations demonstrating heterogeneity in associations across studies at p<0.05 were additionally examined using random-effects meta-analysis (see online supplementary table S4). A Bonferroni-corrected p=2.92×10\(^{-4}\) (nominal α/number of SNPs tested=0.05/171) was used to determine the statistical significance of the overall association for each SNP with CRC.

### RESULTS

The main characteristics of the 54 305 subjects (13 338 cases; 40 967 controls) are presented in online supplementary table S1. The PAGE studies consisted of six different racial/ethnic populations, whereas the GECCO and CCFR consisted of European ancestry populations. In sum, the majority of the subjects were of European ancestry (80.6%), with the remainder comprising 7.0% African-American, 4.5% Hispanic, 6.4% Asian and 1.4% Pacific Islander or Native American ancestry. Most studies ascertained men and women (51.1% women overall), with the exception of WHI and NHS (women only) and HPFS and PHS (men only). Age varied across studies: ARIC ascertained younger adults (mean age of cases=55.8, controls=54.0), whereas the MEC ascertained older adults (mean age of cases=70.0, controls=68.4). Disease stage and anatomical subsite also varied across studies: EAGLE-BioVU, a clinic-based collection of patients, had the largest proportions of advanced stage disease (59.2%) and rectal tumours (42%).

A total of 171 risk variants for 18 cancers other than CRC, representing 100 unique chromosomal regions, were tested in 13 338 cases and 40 967 controls from 16 studies across three consortia. Of the 171 risk variants, 16 variants were nominally associated with CRC at p<0.05 (see online supplementary table S3, figure 1), which was more than the \(~9\) associations expected by

![Figure 1](gut.bmj.com)  Manhatan plot of the meta-analysis association between risk variants of 18 other cancers and colorectal cancer. The solid line is the Bonferroni-corrected significance threshold. Each association is coloured according to the cancer for which the single nucleotide polymorphism was originally reported, and positioned on the x-axis according to its genome position.
chance (171 SNPs \times 0.05 = 8.55). These 16 risk variants consisted of 1 basal cell carcinoma SNP, 1 breast cancer SNP, 1 glioma SNP, 1 leukemia SNP, 2 lung cancer SNPs, 1 non-Hodgkin’s Lymphoma SNP, 8 prostate cancer SNPs, and 1 testicular cancer SNP (figure 1, see online supplementary table S3). Four of these 16 variants are correlated (8q24 Region 1 variants; \( r^2 > 0.88 \) in HapMap CEU) and may not represent independent results.

Two correlated prostate cancer risk variants (rs10090154 and rs4242382; \( r^2 = 0.79 \) in CEU) in Region 1 of chromosome 8q24 (125.6–129.4 Mb) demonstrated statistically significant associations with CRC, reaching a conservative Bonferroni-corrected criterion of significance (\( p < 2.92 \times 10^{-4} \)). For the most statistically significant association, rs4242382, we observed a 12\% increased risk of CRC among CRC cases in comparison to controls (overall meta-analysis OR = 1.12, 95\% CI 1.07 to 1.18; \( p = 1.74 \times 10^{-5} \); figure 2), and no evidence of heterogeneity across studies (\( p_{\text{het}} = 0.07 \)). Notably, with rs10090154 and rs4242382 remained statistically significant when adjusting for rs6983267, a CRC risk variant in Region 3 of 8q24 (Region 3 adjusted meta-analysis OR = 1.11; \( p = 5.0 \times 10^{-5} \) and OR = 1.11; \( p = 5.7 \times 10^{-5} \)). Two additional prostate cancer risk variants in Region 1 of 8q24 (rs7837688, rs1447295) and one in Region 3 (rs7000448) were also associated with CRC. (\( p = 3.32 \times 10^{-4} - 4.85 \times 10^{-5} \)) though they did not reach our conservative threshold of statistical significance. These five prostate cancer SNPs demonstrated similar positive associations with CRC for the corresponding prostate cancer risk alleles. These SNPs are located upstream of MYC at chromosome 8q24, spanning ~98 kb, and are in various amounts of linkage disequilibrium among HapMap Europeans. The Region 1 variants appear correlated with each other (\( r^2 > 0.88 \)) but not with the Region 3 variant (\( r^2 \leq 0.02 \); HapMap release 22 CEU).

Outside of chromosome 8q24, we observed a marginally significant association with rs2736100, a glioma risk variant at the TERT locus at 5p15, and CRC (meta-analysis for the G allele OR = 0.94; 95\% CI 0.91 to 0.97; \( p = 6.57 \times 10^{-4} \); \( p_{\text{het}} \) studies = 0.31; see online supplementary table S3). This inverse association with CRC was in the opposite direction to that observed with the glioma G risk allele of this SNP (figure 3). Another potentially interesting inverse association was observed with the A risk allele of rs981782, a breast cancer variant at the HCN1 locus at 5p12 (meta-analysis OR = 0.96; 95\% CI 0.93 to 0.99; \( p = 0.009 \); \( p_{\text{het}} \) studies = 0.79; see online supplementary table S3).

Next, we evaluated the 16 associations at \( p < 0.05 \) for patterns of associations across race/ethnicity and anatomical subsite (see online supplementary tables S5 and S6). We observed no evidence of heterogeneity in associations by race/ethnicity, with the exception of a potentially nominal association with rs7837688 (\( p_{\text{het}} = 0.049 \)). For the most statistically significant overall association, rs4242382, we observed consistent positive associations at \( p < 0.05 \) for African–American (OR = 1.22; 95\% CI 1.03 to 1.45; \( p = 0.024 \)), Asian (OR = 1.28; 95\% CI 1.09 to 1.51; \( p = 3.06 \times 10^{-3} \)), and European ancestry populations (OR = 1.10; 95\% CI 1.04 to 1.17; \( p = 1.91 \times 10^{-3} \)). Additionally, we observed generally similar directions of association in colon and rectal tumours (see online supplementary table S6). Nominal evidence of heterogeneity in associations by anatomical subsite was observed for rs11155133 at chromosome 6q24 (\( p_{\text{het}} = 0.03 \)), where a stronger inverse association was observed for rectal cancer (meta-analysis OR = 0.60; \( p = 5.58 \times 10^{-3} \)) than colon cancer (meta-analysis OR = 0.87; \( p = 0.039 \)).

Figure 2  Forest plot of the association between rs4242382 at Region 1 of chromosome 8q24 and colorectal cancer risk. Study specific and meta-analysis associations are plotted, modelling the A risk allele for prostate cancer.
DISCUSSION

In this large meta-analysis of 54,305 CRC cases and controls, we examined GWAS-identified risk variants of other cancers for their effects on CRC risk. To our knowledge, this is the first systematic analysis of pleiotropic associations of risk variants for other cancers with CRC. We identified two correlated SNPs—rs10090154 and rs4242382—at Region 1 of chromosome 8q24, a well-established prostate cancer susceptibility locus that demonstrated robust associations with CRC and reached a conservative criterion of statistical significance. We also observed a notable association at TERT, a key susceptibility locus for several cancers.

Chromosome 8q24 has been identified as an important risk locus for multiple cancers,\(^2\)–\(^6\) \(^39\)–\(^43\) including CRC.\(^38\)–\(^48\) Distinct regions within this locus defined by their linkage disequilibrium structure have been associated with various risk variants. SNPs within Region 3, initially identified as a 60 kb region from 128.48 to 128.54 Mb at 8q24,\(^38\) have been consistently associated with CRC in GWAS\(^44\)–\(^48\) and subsequent follow-up studies.\(^11\) \(^49\)–\(^53\) Although associations between Region 3 of chromosome 8q24 and CRC risk are well established, our findings appear to be the first demonstration of highly significant associations with CRC and reached a conservative criterion of statistical significance. We also observed a notable association at TERT, a key susceptibility locus for several cancers.

Chromosome 8q24 has been identified as an important risk locus for multiple cancers,\(^2\)–\(^6\) \(^39\)–\(^43\) including CRC.\(^38\)–\(^48\) Distinct regions within this locus defined by their linkage disequilibrium structure have been associated with various risk variants. SNPs within Region 3, initially identified as a 60 kb region from 128.48 to 128.54 Mb at 8q24,\(^38\) have been consistently associated with CRC in GWAS\(^44\)–\(^48\) and subsequent follow-up studies.\(^11\) \(^49\)–\(^53\) Although associations between Region 3 of chromosome 8q24 and CRC risk are well established, our findings appear to be the first demonstration of highly significant associations with CRC and reached a conservative criterion of statistical significance. We also observed a notable association at TERT, a key susceptibility locus for several cancers.

Chromosome 8q24 has been identified as an important risk locus for multiple cancers,\(^2\)–\(^6\) \(^39\)–\(^43\) including CRC.\(^38\)–\(^48\) Distinct regions within this locus defined by their linkage disequilibrium structure have been associated with various risk variants. SNPs within Region 3, initially identified as a 60 kb region from 128.48 to 128.54 Mb at 8q24,\(^38\) have been consistently associated with CRC in GWAS\(^44\)–\(^48\) and subsequent follow-up studies.\(^11\) \(^49\)–\(^53\) Although associations between Region 3 of chromosome 8q24 and CRC risk are well established, our findings appear to be the first demonstration of highly significant associations with CRC and reached a conservative criterion of statistical significance. We also observed a notable association at TERT, a key susceptibility locus for several cancers.

Chromosome 8q24 has been identified as an important risk locus for multiple cancers,\(^2\)–\(^6\) \(^39\)–\(^43\) including CRC.\(^38\)–\(^48\) Distinct regions within this locus defined by their linkage disequilibrium structure have been associated with various risk variants. SNPs within Region 3, initially identified as a 60 kb region from 128.48 to 128.54 Mb at 8q24,\(^38\) have been consistently associated with CRC in GWAS\(^44\)–\(^48\) and subsequent follow-up studies.\(^11\) \(^49\)–\(^53\) Although associations between Region 3 of chromosome 8q24 and CRC risk are well established, our findings appear to be the first demonstration of highly significant associations with CRC and reached a conservative criterion of statistical significance. We also observed a notable association at TERT, a key susceptibility locus for several cancers.
CRC without the CCFR: a similar nominally significant positive association was observed (meta-analysis OR for the T allele=1.05; 95% CI 1.01 to 1.09; p=0.007). This provides further data for the involvement of TERT in CRC susceptibility. Additionally, an overall meta-analysis between our findings and those of Knoppers et al resulted in a more significant association between rs2736100 and CRC (meta-analysis OR for the T allele=1.06; 95% CI 1.04 to 1.09; p=7.99×10^-7).

The numerous risk loci identified by GWAS of cancer provide a valuable opportunity to assess similarities in the genetic susceptibility of different malignancies. Pleiotropic associations can underscore established etiologic links, as well as uncover novel connections that provide new clues to shared molecular pathways.12 Although cancer is a complex and heterogeneous disease with more than 200 different types, our findings identify shared genetic susceptibility variants between CRC and other cancers of the prostate, lung, breast, testis and glioma. While the magnitudes of these associations are small, the cumulative effect of many such CRC risk variants may help explain the heritability of CRC.13 Furthermore, these pleiotropic associations may indicate the biological importance of such shared genetic regions, and suggest they should be prioritised for future functional and fine-mapping efforts. Specifically, our findings provide additional evidence for Region 1 of chromosome 8q24 and TERT as two such priority regions.

Our study is strengthened by the large number of subjects from well-designed CRC studies and the inclusion of multiple racial/ethnic populations. Limitations of this study include limited study power for 6 SNPs that were not available across all studies. Additionally, the smaller number of non-European ancestry participants limits our ability to fully explore generalisability across race/ethnicity. Finally, as more recent GWAS have reduced study power for 6 SNPs that were not available across several hundred new cancer risk loci, these variants remain to be evaluated for their pleiotropic effects with CRC.

In summary, our study indicates that several risk variants identified for other cancers also contribute to CRC risk. For the first time, these findings clearly demonstrate the importance of Region 1 at chromosome 8q24 in CRC susceptibility, and further bolster the evidence of this region as a multicancer risk locus. Further replication and future research into the biological mechanisms by which inherited differences in shared cancer risk loci influence CRC will expand our understanding of the key contributors to CRC development.

Author affiliations

1Cancer Prevention Institute of California, Fremont, California, USA
2Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada
3Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, Tennessee, USA
4Center for Human Genetics Research, Vanderbilt University, Nashville, Tennessee, USA
5Department of Health Science Research, Mayo Clinic Arizona, Arizona, USA
6Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany
7Department of Epidemiology, University of North Carolina, Chapel Hill, North Carolina, USA
8Epidemiology Program, University of Hawaii Cancer Center, University of Hawaii, Honolulu, Hawaii, USA
9Department of Internal Medicine, University of Utah Health Sciences Center, Salt Lake City, Utah, USA
10Channing Division of Network Medicine, Department of Medicine, Brigham and Women’s Hospital and Harvard, Boston, Massachusetts, USA
11Gastrointestinal Unit, Massachusetts General Hospital, Boston, Massachusetts, USA
12Department of Medicine, University of North Carolina School of Medicine, Chapel Hill, North Carolina, USA
13Office of Population Genomics, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA
14Department of Preventive Medicine, Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, California, USA
15Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany
16Program in Molecular and Genetic Epidemiology, Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA
17Division of Genetic Basis of Human Disease, Translational Genomics Research Institute, Phoenix, Arizona, USA
18Department of Preventive Medicine, Northwestern University, Chicago, Illinois, USA
19Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts, USA
20Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA
21Centre for Molecular, Environmental, Genetic & Analytic Epidemiology, School of Population Health, The University of Melbourne, Melbourne, Australia
22Stanford Cancer Institute, Palo Alto, California, USA
23Epidemiology Research Program, American Cancer Society, Atlanta, Georgia, USA
24Centre for Public Health Research, Massey University, Wellington, New Zealand
25Division of Research, Kaiser Permanente Medical Care Program, Oakland, California, USA
26Department of Medicine and Epidemiology, University of Pittsburgh Medical Center, Pennsylvania, USA
27Division of Epidemiology, Department of Environmental Medicine, New York University School of Medicine, New York, New York, USA
28Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA
29Service de Genétique Médicale, CHU Nantes, Nantes, France
30Information Sciences Institute, University of Southern California, Marina del Rey, California, USA
31Vanderbilt Technologies for Advanced Genomics, Vanderbilt University, Nashville, Tennessee, USA
32Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada
33Ontario Institute for Cancer Research, Toronto, Ontario, Canada
34Department of Surgery, Mount Sinai Hospital, Toronto, Ontario, Canada
35Department of Biostatistics, University of Washington, Seattle, Washington, USA
36Division of Cancer Control and Population Sciences, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA
37Division of Epidemiology, University of Washington School of Public Health, Seattle, Washington, USA
38Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada
39Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada
40Germany German Cancer Consortium (DKTK), Heidelberg, Germany
41Department of Biomedical Informatics, Vanderbilt University, Nashville, Tennessee, USA

Acknowledgements The authors would like to thank all participants, staff, physicians and investigators for making this project possible. We thank Dr Bruno Buecher of ASTERISK; Ute Handte-Daube, Muhabbet Celik and Ursula Eliber of DACHS; Patrice Soule, Hardeep Ranu, Immaculata Devivo, David Hunter, Qin (Carolyn) Guo, Luxue Zhu and Haiyan Zhang of HPFS, NHS and PHS; Christine Berg and Philip Prokop of PLCO; Tom Riley and staff of Information Management Services Inc; Barbara O’Brien and staff of Westat Inc; Bill Kopp, Wen Shao and staff of SAIC-Frederick; WHI investigators (see https://clinicaltrial.gov/researchers/SitePages/Write%20a%20Paper.aspx) and the GECCO Coordinating Center. We would like to thank Awapuhi Lee and Kristine Winters for their technical assistance. We would additionally like to thank the participants and staff of the Nurses’ Health Study and the Health Professionals Follow-up Study for their valuable contributions, as well as the following state cancer registries for their help: 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.

Contributors CRC, YL, JLA, GK, DMR, RJG, HHG, PB, LH and SJ were involved in data analysis. WSB was involved in data analysis, data interpretation, and manuscript writing. S-AL was involved in data analysis and manuscript writing. DS and CSC were involved in data interpretation. LD, NML, CLA, ATC, JAB, LAH, SLP, FR5, DJD, DCL, DVC, DCC, CAH, CMH and EW were involved in data interpretation and manuscript writing. TJH and HB were involved in data interpretation, monitoring and quality control, provision of study subjects, and manuscript writing. AKM, LH, and RES were involved in manuscript writing. PK, KRM, SM, JM, ELG and TAH were involved in monitoring collected data. CF, BIC, MAJ, SK, BWZ, ML, and SG were involved in monitoring data collected and provision of study subjects. MJH was involved in monitoring data collected, provision of study subjects, and manuscript writing. PAN, RBH, SIC, SIB were involved in the provision of study subjects. SB was involved in provision of study subjects and manuscript writing. JLM, RHG, J-C, PTC, CK, DCC, GH, JDP, MLS and GC were involved in the provision of study subjects, data interpretation, and manuscript writing. LLM and UP were involved in study...
design and conception, provision of study subjects, data interpretation, and manuscript writing. IC, JMK and LRW were involved in study design, data analysis, data interpretation, and manuscript writing.

**Funding** PAGE: (a) The Population Architecture Using Genomics and Epidemiology (PAGE) program is funded by the National Human Genome Research Institute (NHGRI), supported by U01HG004803 (CALICO), U01HG004798 (EAGLE), U01HG004802 (MEC), U01HG004796 (NHS) and U01HG004801 (Coordinating Center), and their respective NHGRI ARRA supplements. The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. The complete list of PAGE members can be found at http://www.pagestudy.org. (b) The data and materials included in this report result from collaboration between the following studies: The Epidemiologic Architecture for Genes Linked to Environment (EAGLE) is funded through the NHGRI PAGE program (U01HG004798-01 and its NHGRI ARRA supplement). The dataset(s) used for the analyses described were obtained from Vanderbilt University Medical Center’s BioVU which is supported by institutional funding and by the Vanderbilt CTSA grant UL1 TR000445 from NCATS/NHLBI. The Vanderbilt University Center for Human Genetics Research, Computational Genomics Core provided computational and/or analytical support for this work. The Multibehavior Cohort study (MEC) characterization of epidemiological architecture is funded through the NHGRI PAGE program (U01HG004802 and its NHGRI ARRA supplement). The MEC study is funded through the National Cancer Institute (R37CA54281, R01 CA63464, P01CA33619, U01CA136792 and U01CA98758). Funding support for the ‘Epidemiology of putative genetic variants: The Women’s Health Initiative’ study is provided through the NHGRI PAGE program (U01HG004790 and its NHGRI ARRA supplement). The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, US Department of Health and Human Services through contracts HS/N268211000046C, HS/N268211000010C, HS/N268211000023C, HS/N268211000020C and HS/N2712101100004C. The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full list of all WHI investigators can be found at: http://www.whiscience.org/publications/WHI_investigators_shortlist.pdf. Funding support for the Genetic Epidemiology of Causal Variants Across the Life Course (CALICO) program was provided through the NHGRI PAGE program (U01HG004803 and its NHGRI ARRA supplement). The following study contributed to this manuscript and is funded by the following agencies: The Atherosclerosis Risk in Communities (ARIC) Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HS/N268211000005C, HS/N268211000006C, HS/N268211000007C, HS/N268211000008C, HS/N268211000009C, HS/N268211000100C, HS/N268211000111C and HS/N26821100012C). The National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HS/N26820625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025055, a component of the National Institutes of Health and NIH Roadmap for Medical Research. Assistance with phenotype harmonisation, SNP selection and annotation, data cleaning, data management, integration and dissemination, and general study coordination was provided by the PAGE Coordinating Center (U01HG004801-01 and its NHGRI ARRA supplement). The National Institutes of Mental Health also contributes to the support for the Coordinating Center. The PAGE consortium thanks the staff and participants of all PAGE studies for their important contributions.

**REFERENCES**


