

SUPPLEMENTARY TEXT

Colorectal tumor anatomic sublocation definitions

We defined proximal colon cancer as any primary tumor arising in the cecum, ascending colon, hepatic flexure, or transverse colon (ICD-9 codes: 153.4, 153.6, 153.0, or 153.1, respectively); distal colon cancer as any primary tumor arising in the splenic flexure, descending colon, or sigmoid colon (ICD-9 codes: 153.7, 153.2, or 153.3, respectively); and rectal cancer as any primary tumor arising in the rectum or rectosigmoid junction (ICD-9 codes: 154.1, or 154.0, respectively). To examine the effect of including rectosigmoid junction in the rectal cancer analyses, we performed a sensitivity analysis in which we excluded rectosigmoid junction cases. Since we found that association signals were attenuated in the analyses without the rectosigmoid junction, we included it in the definition of rectal cancer. For the GWAS discovery analyses, we analyzed five case subgroups based on primary tumor sublocation. In addition to the three aforementioned mutually exclusive case sets proximal colon, distal colon, and rectal cancer, we defined colon cancer and distal/left-sided colorectal cancer case sets. Colon cancer cases comprised combined proximal colon and distal colon cancer cases, and additional colon cases with ICD-9 code 153.9. In the distal/left-sided colorectal cancer cases analysis, we combined distal colon and rectal cancer cases based on the different embryonic origins of the proximal colon versus the distal colon and rectum. Supplementary figure 1 and supplementary table 1 summarize distributions of age of diagnosis by sex and primary tumor site. Our data recapitulate the reported higher percentage of female proximal colon cancer cases, a male-to-female ratio that increases progressively from the proximal colon to the rectum,[1] and differences in age of onset by primary tumor site, with an earlier age of onset for rectal cancer[2] (supplementary figure 1 and supplementary table 1).

Assessing population stratification in the GWAS meta-analyses

To examine confounding due to residual population stratification, we inspected quantile-quantile (QQ) plots of test statistics for the five GWAS meta-analyses (supplementary figure 2) and genomic control

inflation factors (λ_{GC}). λ_{GC} were 1.073, 1.077, 1.089, 1.090, and 1.101 for the GWAS meta-analyses for CRC case subgroups defined by proximal colon, distal colon, rectum, colon, and left-sided tumors subsites, respectively. The corresponding λ_{1000} values (λ_{GC} rescaled to a sample of 1,000 cases and 1,000 controls) were 1.007, 1.008, 1.007, 1.005, and 1.006. Given the highly polygenic genetic architecture of CRC, these estimates likely reflect the aggregate small effects of a large number of risk variants, although a small contribution of population stratification or bias cannot be excluded.

Genomic annotation of new GWAS loci and gene prioritization

To gain insight into the molecular mechanisms underlying the associations, and to nominate candidate causal variants and the most likely target gene(s) at each locus, we annotated all new risk loci with five types of functional and regulatory genomic annotations: (i) cell-type-specific regulatory annotations for histone modifications and open chromatin, (ii) nonsynonymous coding variation, (iii) evidence of transcription factor binding, (iv) predicted functional impact across different databases for non-coding and coding variants, (v) co-localization with expression quantitative trait loci (eQTL) signal.

We operationally defined a locus as the lead variant and variants in linkage disequilibrium ($r^2 \geq 0.8$) with the lead variant. We intersected variants with previously published functional and regulatory genomic annotations for normal colonic crypt epithelium and colonic mucosa tissue, and diverse CRC cell lines or tissue. Specifically, we assessed overlap with active enhancer regions identified by histone mark H3K27ac, and active regulatory regions identified by accessible chromatin identified through DNase I hypersensitive sites (DHSs) and ATAC-seq.[3,4] Additionally, to assess tissue-specificity we intersected with a compendium of ten groups of tissue-specific regulatory annotations for histone modifications (H3K4me1, H3K4me3, H3K9ac, and H3K27ac) obtained from various resources,[5,6] and previously derived by Finucane et al.[7] To assess gene-centric functional consequences of variants, we used the Ensembl Variant Effect Predictor tool to annotate variants relative to GENCODE and reference genome GRCh37.p13.[8] To determine whether variants potentially disrupt transcription factor (TF) binding, we

examined overlap with predicted transcription factor binding site (TFBS) motifs using the Haploreg v4.1 database,[9] and the chromatin immuno-precipitation sequence (ChIP-seq) binding sites for 161 transcription factors from the ENCODE Project.[10] We assessed colocalization with eQTL signals across tissues using GTEx V8 data.[11] To examine predicted functional impact, we annotated all locus variants with the CADD score (Phred scores >20 predicted as deleterious),[12] and we also annotated coding variants with the DANN score (scores >0.9 predicted as deleterious),[13] and PolyPhen-2 (benign, possibly damaging, probably damaging).[14] To evaluate expression in colorectal tissue, we interrogated the Human Protein Atlas[15] and GTEx V8 data.[11] Genes were prioritized based on biological relevance, expression in colorectal tissue, the presence of associated non-synonymous coding variants predicted to be deleterious, evidence from laboratory-based functional studies, somatic alterations, or familial syndromes linking them to CRC or cancer pathogenesis.

Data availability

All genotype data analyzed in this study have been previously published and have been deposited in the database of Genotypes and Phenotypes (dbGaP), which is hosted by NCBI, under accession numbers phs001415.v1.p1, phs001315.v1.p1, and phs001078.v1.p1. The UK Biobank resource was accessed through application number 8614. CRC-relevant epigenome data were retrieved from the NCBI Gene Expression Omnibus (GEO) database under accession numbers GSE77737 and GSE36401.

Supplementary references

- 1 Murphy G, Devesa SS, Cross AJ, *et al.* Sex disparities in colorectal cancer incidence by anatomic subsite, race and age. *Int J Cancer* 2011;**128**:1668–75. doi:10.1002/ijc.25481
- 2 Murphy CC, Singal AG. Establishing a research agenda for early-onset colorectal cancer. *PLoS Med* 2018;**15**:e1002577. doi:10.1371/journal.pmed.1002577
- 3 Cohen AJ, Saiakhova A, Corradin O, *et al.* Hotspots of aberrant enhancer activity punctuate the colorectal cancer epigenome. *Nat Commun* 2017;**8**:14400. doi:10.1038/ncomms14400
- 4 Huyghe JR, Bien SA, Harrison TA, *et al.* Discovery of common and rare genetic risk variants for colorectal cancer. *Nat Genet* 2019;**51**:76–87. doi:10.1038/s41588-018-0286-6
- 5 Trynka G, Sandor C, Han B, *et al.* Chromatin marks identify critical cell types for fine mapping

- complex trait variants. *Nat Genet* 2013;**45**:124–30. doi:10.1038/ng.2504
- 6 Hnisz D, Abraham BJ, Lee TI, *et al.* Super-enhancers in the control of cell identity and disease. *Cell* 2013;**155**:934–47. doi:10.1016/j.cell.2013.09.053
 - 7 Finucane HK, Bulik-Sullivan B, Gusev A, *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* 2015;**47**:1228–35. doi:10.1038/ng.3404
 - 8 McLaren W, Gil L, Hunt SE, *et al.* The ensembl variant effect predictor. *Genome Biol* 2016;**17**:122. doi:10.1186/s13059-016-0974-4
 - 9 Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* 2012;**40**:D930-4. doi:10.1093/nar/gkr917
 - 10 ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;**489**:57–74. doi:10.1038/nature11247
 - 11 GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 2013;**45**:580–5. doi:10.1038/ng.2653
 - 12 Kircher M, Witten DM, Jain P, *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014;**46**:310–5. doi:10.1038/ng.2892
 - 13 Quang D, Chen Y, Xie X. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics* 2015;**31**:761–3. doi:10.1093/bioinformatics/btu703
 - 14 Adzhubei IA, Schmidt S, Peshkin L, *et al.* A method and server for predicting damaging missense mutations. *Nat Methods* 2010;**7**:248–9. doi:10.1038/nmeth0410-248
 - 15 Uhlén M, Fagerberg L, Hallström BM, *et al.* Proteomics. Tissue-based map of the human proteome. *Science* 2015;**347**:1260419. doi:10.1126/science.1260419

FUNDING AND ACKNOWLEDGEMENTS

Funding:

Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO): National Cancer Institute, National Institutes of Health, U.S. Department of Health and Human Services (U01 CA164930, U01 CA137088, R01 CA059045, R21 CA191312, R01 CA201407).

Genotyping/Sequencing services were provided by the Center for Inherited Disease Research (CIDR) (X01-HG008596 and X01-HG007585). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268201200008I. This research was funded in part through the NIH/NCI Cancer Center Support Grant P30 CA015704. PCS is supported by the National Institutes of Health (R01 CA160356, R01 CA193677, R01 CA204279 and R01 CA143237).

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 Entreprises Françaises dans la Lutte contre le Cancer (GEFLUC), the Association Anne de Bretagne Génétique and the Ligue Régionale Contre 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: National Institutes of Health (R01 CA60987).

ColoCare: This work was supported by the National Institutes of Health (grant numbers R01 CA189184 (Li/Ulrich), U01 CA206110 (Ulrich/Li/Siegel/Figueireido/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 (CFR) Illumina GWAS was supported by funding from the National Cancer Institute, National Institutes of Health (grant numbers U01 CA122839, R01 CA143247 to G Casey). The Colon CFR/CORECT Affymetrix Axiom GWAS and OncoArray GWAS were supported by funding from National Cancer Institute, National Institutes of Health (grant number U19 CA148107 to S Gruber). The Colon CFR participant recruitment and collection of data and biospecimens used in this study were supported by the National Cancer Institute, National Institutes of Health (grant number U01 CA167551) and through cooperative agreements with the following Colon CFR centers: Australasian Colorectal Cancer Family Registry (NCI/NIH grant numbers U01 CA074778 and U01/U24 CA097735), USC Consortium Colorectal Cancer Family Registry (NCI/NIH grant numbers U01/U24 CA074799), Mayo Clinic Cooperative Family Registry for Colon Cancer Studies (NCI/NIH grant number U01/U24 CA074800), Ontario Familial Colorectal Cancer Registry (NCI/NIH grant number U01/U24 CA074783), Seattle Colorectal Cancer Family Registry (NCI/NIH grant number U01/U24 CA074794), and University of Hawaii Colorectal Cancer Family Registry (NCI/NIH grant number U01/U24 CA074806), Additional support for case ascertainment was provided from the Surveillance, Epidemiology and End Results (SEER) Program of the National Cancer Institute to Fred Hutchinson Cancer Research Center (Control Nos. N01-CN-67009 and N01-PC-35142, and Contract No. HHSN2612013000121), the Hawai'i Department of Health (Control Nos. N01-PC-67001 and N01-PC-35137, and Contract No. HHSN26120100037C, and the California Department of Public Health (contracts HHSN261201000035C awarded to the University of Southern California, and the following state cancer registries: AZ, CO, MN, NC, NH, and by the Victoria Cancer Registry and Ontario Cancer Registry.

COLON: The COLON study is sponsored by Wereld Kanker Onderzoek Fonds, including funds from grant 2014/1179 as part of the World Cancer Research Fund International Regular Grant Programme, by Alpe d'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 pre-competitive 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 National Cancer Institute, National Institutes of Health (NCI/NIH), U.S. Department of Health and Human Services (grant numbers U19 CA148107, R01 CA81488, P30 CA014089, R01 CA197350, P01 CA196569, R01 CA201407, and R01 CA143237) and National Institutes of Environmental Health Sciences, National Institutes of Health (grant number T32 ES013678).

CORSA: This study was funded by FFG BRIDGE (grant 829675 to Andrea Gsur), the “Herzfelder’sche Familienstiftung” (grant to Andrea Gsur) and was supported by COST Action BM1206.

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’Oncologia de Catalunya (XBTC), Plataforma Biobancos PT13/0010/0013 and ICOBIOBANC, sponsored by the Catalan Institute of Oncology.

Czech Republic CCS: This project was supported by Grant Agency of the Ministry of Health of the Czech Republic (NV19-09-00237) and by Charles University Research Centre program UNCE/MED/006 “University Center of Clinical and Experimental Liver Surgery” and National Sustainability Program I (NPU I) Nr. LO1503 provided by the Ministry of Education Youth and Sports of the Czech Republic.

DACHS: This work was supported by the German Research Council (BR 1704/6-1, BR 1704/6-3, BR 1704/6-4, BR 1704/6-6, CH 117/1-1, HO 5117/2-1, HE 5998/2-1, KL 2354/3-1, RO 2270/8-1 and BR 1704/17-1), the Interdisciplinary Research Program of the National Center for Tumor Diseases (NCT), Germany, and the German Federal Ministry of Education and Research (01KH0404, 01ER0814, 01ER0815, 01ER1505A and 01ER1505B).

DALS: National Institutes of Health (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 the School of Public Health, Imperial College London and the International Agency for Research on Cancer. The national cohorts are supported by Danish Cancer Society (Denmark); Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Éducation Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research Center (DKFZ), Federal Ministry of Education and Research (BMBF), Deutsche Krebshilfe, Deutsches Krebsforschungszentrum and Federal Ministry of Education and Research (Germany); the Hellenic Health Foundation (Greece); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy 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); ERC-2009-AdG 232997 and Nordforsk, Nordic Centre of Excellence programme on Food, Nutrition and Health (Norway); Health Research Fund (FIS), PI13/00061 to Granada, PI13/01162 to EPIC-Murcia, Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, ISCIII RETIC (RD06/0020) (Spain); Swedish Cancer Society, Swedish Research Council and County Councils of Skåne and Västerbotten (Sweden); Cancer Research UK (14136 to EPIC-Norfolk; C570/A16491 and C8221/A19170 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 (PGDIT07PXIB9101209PR), Ministerio de Economía 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 Gestió d'Ajuts Universitaris 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 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, R35 CA197735, K07 CA190673, and P50 CA127003), NHS by the National Institutes of Health (R01 CA137178, P01 CA087969, UM1 CA186107, R01 CA151993, R35 CA197735, K07 CA190673, and P50 CA127003) and PHS by the National Institutes of Health (R01 CA042182).

Hawaii Adenoma Study: NCI grants R01 CA72520.

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 209057, 396414 and 1074383, 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: National Institutes of Health (R37 CA54281, P01 CA033619, R01 CA063464, and U01 CA164973).

MECC: This work was supported by the National Institutes of Health, U.S. Department of Health and Human Services (R01 CA81488 to SBG, R01 CA197350 to SBG, U19 CA148107 to SBG, N01 CN043302 assigned to SBG, 5P01 CA196569 to WG, P30 CA014089 to SBG, R01 CA144040 to SDM, and P50 CA150964 to SDM), as well as funding from the Ravitz Foundation, the Irving Weinstein Foundation, the Anton B. Burg Foundation, the Jane and Kris Popovich Chair in Cancer Research, and a generous gift from Daniel and Maryann Fong.

MSKCC: The work at Memorial Sloan Kettering Cancer Center in New York was supported by the Robert and Kate Niehaus Center for Inherited Cancer Genomics and the Romeo Milio Foundation. It is also supported by the National Cancer Institute-designated Comprehensive Cancer Center (grant number P30 CA008748).

NCCCS I & II: We acknowledge funding support for this project from the National Institutes of Health, 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 National Cancer Institute, US National Institutes of Health (grant number U01 CA74783 to S Gallinger); and National Cancer Institute of Canada grants (18223 and 18226). Funding was provided to Michael O. Woods by the Canadian Cancer Society Research Institute.

NSHDS: This work was funded by the Swedish Research Council, Swedish Cancer Society; Cancer Research Foundation and Lion's Cancer Research Foundation in Northern Sweden; J C Kempe Memorial Fund; Faculty of Medicine, Umeå University, Umeå, Sweden; and the Västerbotten County Council, Sweden.

OFCCR: National Institutes of Health, through funding allocated to the Ontario Registry for Studies of Familial Colorectal Cancer (U01 CA74783 to S Gallinger); see CCFR section above. Additional funding toward genetic analyses of OFCCR includes the Ontario Research Fund, the Canadian Institutes of Health Research, and the Ontario Institute for Cancer Research, through generous support from the Ontario Ministry of Research and Innovation.

OSUWMC: 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 National Institutes of Health (NIH), Genes, Environment and Health Initiative (GEI) Z01 CP 010200, NIH U01 HG004446, and NIH GEI U01 HG 004438.

PMH-SCCFR: National Cancer Institute, National Institutes of Health (grant numbers U01 CA074794 to J Potter and R01 CA076366 to P.A. Newcomb).

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: The Selenium and Vitamin E Cancer Prevention Trial (SELECT) was supported by the National Cancer Institute of the National Institutes of Health under Award Numbers UM1CA182883 and U10CA37429.

SMS: 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 National Institutes of Health (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 grants from the Swedish Cancer Foundation, the Swedish Research Council for the Swedish Infrastructure for Medical Population-based Life-course Environmental Research (SIMPLER) and Karolinska Institute's Distinguished Professor Award to Alicja Wolk.

VITAL: National Institutes of Health (K05 CA154337).

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

Acknowledgements:

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 also thank all individuals who agreed to participate in the CORSA study, all physicians and students.

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, Ansgar Brandhorst, Utz Benschaid, Muhabbet Celik 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.

EPICOLON: We are sincerely grateful to all patients, the Spanish National DNA Bank (BNADN) and Biobank of Hospital Clínic-IDIBAPS.

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 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.

NFCCR: 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.

NSHDS: We thank all participants in the NSHDS cohorts and the staff at the Department of Biobank Research, Umeå University, as well as Kerstin Näslund, formerly of the Department of Medical Biosciences, and Inger Cullman, Department of Chemistry, both at Umeå University for excellent technical assistance.

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.

PMH: The authors would like to thank the study participants and staff of the Hormones and Colon Cancer study.

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>