Article:
Tian, Y, Kim, AE, Bien, SA et al. (74 more authors) (2022) Genome-Wide Interaction Analysis of Genetic Variants with Menopausal Hormone Therapy for Colorectal Cancer Risk. Journal of the National Cancer Institute. ISSN 0027-8874

https://doi.org/10.1093/jnci/djac094

Genome-Wide Interaction Analysis of Genetic Variants with Menopausal Hormone Therapy for Colorectal Cancer Risk

Authors:

Yu Tian, PhD[1,2=]; Andre E Kim, PhD [3,=]; Stephanie A Bien, PhD [4]; Yi Lin, MS[4]; Conghui Qu, MS[4]; Tabitha Harrison, MPH[4]; Robert Carreras-Torres, PhD[5,6,7]; Virginia Díez-Obrero, PhD[5,6,7]; Niki Dimou, PhD[8]; David A. Drew, PhD [9,10]; Akihisa Hidaka, MD, PhD[4]; Jeroen R Huyghe, PhD[4]; Kristina M. Jordahl , PhD[4,11]; John Morrison, MS[3], Neil Murphy, PhD[8]; Mireia Obón-Santacana, PhD[5,6,7]; Cornelia M. Ulrich, PhD[12,13]; Jennifer Ose, PhD[12,13]; Anita R. Peoples, PhD, MPH[12,13]; Edward A. Ruiz-Narvaez, Sc.D.[14]; Anna Shcherbina, PhD[15]; Mariana Stern, PhD[16]; Yu-Ru Su, PhD[4,17]; Franzel JB van Duijnhoven, PhD[18]; Volker Arndt, MD, MPH[19]; James Baurley, PhD[20,21]; Sonja I Berndt, PharmD, PhD[22]; D Timothy Bishop, PhD[23]; Hermann Brenner, MD, MPH[19,24,25]; Daniel D Buchanan, PhD[26,27,28]; Andrew T Chan, MD, MPH[9,10,29,30,31,32]; Jane C Figueiredo, PhD[33,34]; Steven Gallinger, MD[35]; Stephen B Gruber, MD, MPH, PhD[16], Sophia Harlid, PhD[36]; Michael Hoffmeister, PhD[19]; Mark A Jenkins, PhD[37]; Amit D Joshi, PhD[31]; Temitope O Keku, PhD, MSPH[38]; Susanna C Larsson, PhD[39]; Loic Le Marchand, MD, PhD[40]; Li Li, MD, PhD[41]; Graham G Giles, PhD[37,42,43]; Roger L Milne, PhD[37,42,43]; Hongmei Nan, MD, PhD[44,45], Rami Nassir, PhD[46]; Shuji Ogino, MD, PhD[31,47,48]; Arif Budiarto, MS[20], Elizabeth A Platz, Sc.D., MPH[49]; John D Potter, MD, PhD[4,50], Ross L Prentice, PhD[4,51]; Gad Rennert, MD, PhD[52,53,54]; Lori C Sakoda, PhD[4,55]; Robert E Schoen, MD, MPH[56]; Martha L Slattery, PhD, MPH[57]; Stephen N Thibodeau, PhD[58]; Bethany Van Guelpen, PhD[37,59]; Kala Visvanathan, MD, MHS[49]; Emily White, PhD[4,11]; Alicja Wolk, PhD[39]; Michael O Woods, PhD[60]; Anna H Wu, PhD[34]; Peter T Campbell, PhD[61]; Graham Casey, PhD[62]; David V Conti, PhD[16]; Marc J Gunter, PhD[8]; Anshul Kundaje, PhD[63,64]; Juan Pablo Lewinger, PhD[3]; Victor Moreno, PhD[5,6,7,65];

Polly A Newcomb, PhD[4,11]; Bens Pardamean, PhD[20]; Duncan C Thomas, PhD[16]; Konstantinos K Tsilidis, PhD[66,67]; Ulrike Peters, PhD[4,11,*]; W. James Gauderman, PhD[16,*]; Li Hsu, PhD[4,51,*]; Jenny Chang-Claude, PhD[1,68,*]

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

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

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

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

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

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

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

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

[9]Clinical and Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA

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

[11]Department of Epidemiology, University of Washington School of Public Health, Seattle, Washington, USA

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

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

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

[15]Biomedical Informatics Program, Dept. of Biomedical Data Sciences, Stanford University, Stanford, California, USA

[16] Division of Biostatistics, Department of Population and Public Health

Sciences & USC Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, California, USA

[17]Kaiser Permanente Washington Health Research Institute, Seattle, Washington, USA

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

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

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

[21]BioRealm LLC, Walnut, California, USA

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

[23]Leeds Institute of Cancer and Pathology, University of Leeds, Leeds, UK

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

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

[26]Colorectal Oncogenomics Group, Department of Clinical Pathology, The University of Melbourne, Parkville, Victoria 3010 Australia

[27]University of Melbourne Centre for Cancer Research, Victorian Comprehensive Cancer Centre, Parkville, Victoria 3010 Australia

[28]Genomic Medicine and Family Cancer Clinic, The Royal Melbourne Hospital, Parkville, Victoria, Australia

[29]Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA

[30]Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA

[31]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Harvard University, Boston, Massachusetts, USA

[32]Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Harvard University, Boston, Massachusetts, USA

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

[34]Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, California, USA

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

[36]*Department of Radiation Sciences, Oncology Unit, Umeå University, Umeå, Sweden*

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

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

[39]*Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden*

[40]*University of Hawaii Cancer Center, Honolulu, Hawaii, USA*

[41]*Department of Family Medicine, University of Virginia, Charlottesville, Virginia, USA*

[42]*Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, Victoria, Australia*

[43]*Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria, Australia*

[44]*Department of Global Health, Richard M. Fairbanks School of Public Health, Indianapolis, Indiana, USA*

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

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

[47]*Program in MPE Molecular Pathological Epidemiology, Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA*

[48]*Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA*

[49]*Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA*

[50]*Research Centre for Hauora and Health, Massey University, Wellington, NEW ZEALAND*

[51]*Department of Biostatistics, University of Washington, Seattle, Washington, USA*

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

[53]*Ruth and Bruce Rappaport Faculty of Medicine, Technion-Israel Institute of Technology, Haifa, Israel*

[54]*Clalit National Cancer Control Center, Haifa, Israel*

[55]*Division of Research, Kaiser Permanente Northern California, Oakland,*

California, USA

[56]*Department of Medicine and Epidemiology, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania, USA*

[57]*Department of Internal Medicine, University of Utah, Salt Lake City, Utah, USA*

[58]*Division of Laboratory Genetics, Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota, USA*

[59]*Wallenberg Centre for Molecular Medicine, Umeå University, Umeå, Sweden*

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

[61]*Department of Population Science, American Cancer Society, Atlanta, Georgia, USA*

[62]*Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia, USA*

[63]*Department of Genetics, Stanford University, Stanford, California, USA*

[64]*Department of Computer Science, Stanford University, Stanford, California, USA*

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

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

[67]*Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece*

[68]*University Cancer Centre Hamburg (UCCH), University Medical Centre Hamburg-Eppendorf, Hamburg, Germany*

[=] These authors contributed equally to this work.

[*] These authors jointly supervised this work.

Correspondence to:

Prof. Dr. Jenny Chang-Claude

Division of Cancer Epidemiology

German Cancer Research Center (DKFZ)

Im Neuenheimer Feld 581

69120, Heidelberg

Germany

Tel.: +49 6221 42 2373

Fax: +49 6221 42 2203

Email: j.chang-claude@dkfz.de

Abstract

Background The use of menopausal hormone therapy (MHT) may interact with genetic variants to influence colorectal cancer (CRC) risk. **Methods** We conducted a genome-wide gene-environment interaction between single nucleotide polymorphisms and the use of any MHT, estrogen-only, and combined estrogen-progestogen therapy with CRC risk, among 28,486 postmenopausal women (11,519 cases and 16,967 controls) from 38 studies, using logistic regression, two-step method, and 2- or 3-degree-of-freedom (d.f.) joint test. A set-based score test was applied for rare genetic variants. **Results** The use of any MHT, estrogen-only and estrogen-progestogen were associated with a reduced CRC risk [odds ratio (OR) with 95% confidence interval (95% CI) of 0.71 (0.64-0.78), 0.65 (0.53-0.79), and 0.73 (0.59-0.90), respectively]. The two-step method identified a statistically significant interaction between a *GRIN2B* variant rs117868593 and MHT use, whereby MHT-associated CRC risk was significantly reduced in women with the GG genotype [0.68 (0.64-0.72)] but not within strata of GC or CC genotypes. A statistically significant interaction between a *DCBLD1* intronic variant at 6q22.1 (rs10782186) and MHT use was identified by the 2-d.f. joint test. The MHT-associated CRC risk was reduced with increasing number of rs10782186-C alleles, showing ORs of 0.78 (0.70-0.87) for TT, 0.68 (0.63-0.73) for TC, and 0.66 (0.60-0.74) for CC genotypes. In addition, five genes in rare variant analysis showed suggestive interactions with MHT (two-sided

P<1.2x10^{-4}). **Conclusion** Genetic variants that modify the association between MHT and CRC risk were identified, offering new insights into pathways of CRC carcinogenesis and potential mechanisms involved.

Background

The use of menopausal hormone therapy (MHT) has been identified to be associated with a reduced risk of colorectal cancer (CRC) (1-4). In a meta-analysis including 20 studies, both ever use of estrogen-only MHT (RR 0.79, 95% CI 0.69-0.91) and ever use of combined estrogen-progestogen MHT (RR 0.74, 95% CI 0.68-0.81) were associated with a reduced CRC risk (1).

Previous gene-environment (GxE) interaction studies that investigated the association of MHT use with CRC risk according to genetic variants (5-10) have reported a few potential genetic modifiers of CRC risk associated with the use of MHT, however, these studies were based on limited candidate genes/pathways or limited sample size. We conducted a comprehensive genome-wide GxE analysis of both common and rare genetic variants, using the largest known study sample to date, on the one hand, to identify novel genetic variants that may modify the beneficial influence of MHT on CRC risk in order to obtain insight into potential mechanisms behind the association between MHT and CRC risk. On the other hand, the analysis can yield novel genetic susceptibility alleles for CRC risk, which may not be identified without accounting for the GxE component.

Methods

Study participants

We included 38 studies from North America, Australia, and Europe participating in the multi-centered Colon Cancer Family Registry (CCFR), the

Colorectal Transdisciplinary Study (CORECT), the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO), and the United Kingdom Biobank (UKBiobank), which were included in genome-wide association studies as described previously (11-13). Study details and descriptions can be found in the Supplementary Materials (available online). All studies were approved by their respective Institutional Review Boards, and study participants provided informed consent.

Exposure assessment

Information on demographics and environmental risk factors were collected by interviews and/or structured questionnaires. We carried out a multi-step data-harmonization procedure at the GECCO coordinating center (Fred Hutchinson Cancer Research Center) as described previously (10, 14, 15).

Postmenopausal status was defined by using (I) menopausal status derived from studies, if available; or (II) self-reported menopausal status, if study-derived was not available; or (III) age>55, if neither study derived nor self-report were available (Supplementary table 1). MHT use was considered as any MHT use (i.e., either use of estrogen-only or estrogen-progestogen) or estrogen-only use or estrogen-progestogen use at or up to the reference time. Non-users of any MHT at or up to reference time were used as the reference group.

Genotyping, quality control, and imputation

Details on genotyping, imputation and quality control have been reported previously (16). In brief, genotyped single nucleotide polymorphisms (SNPs) were excluded on the basis of call rate (<98%), evidence of departure from Hardy-Weinberg equilibrium (HWE) in controls ($P<1x10^{-4}$). All autosomal SNPs in all studies were imputed to the Haplotype Reference Consortium r1.1 (2016) reference panel via the Michigan Imputation Server (17) and converted into a binary format for data management and analyses using R package BinaryDosage (18). Imputed common SNPs were restricted based on a pooled MAF≥1% and imputation accuracy ($R^2>0.8$). After imputation and quality control analyses, a total of over 7.2 million common SNPs were included. All analyses were restricted to samples clustering with the Utah residents of Northern and Western European ancestry from the CEU population in principal component analysis.

Statistical Methods

Statistical analyses of all data were conducted centrally on individual-level data. All tests of statistical significance were two-sided. Unless otherwise indicated, we adjusted for age at the reference time, study center, and the first three principal components (Plink2) to account for potential population substructure. SNPs were treated as continuous variables (i.e., log-additive effects). To evaluate MHT main effects, each study was analyzed separately using logistic regression models, and study-specific results were combined using fixed- and random-effects meta-analysis methods to obtain summary

odds ratios (ORs) and 95% confidence intervals (CIs) across studies. We calculated the heterogeneity p-values using Cochran's Q statistics (19). Quantile-quantile (Q-Q) plots were used to assess whether the distribution of the p-values was consistent with the null distribution (except for the extreme tail).

Genome-wide interaction scans of common markers were conducted using R package GxEScanR (20), which implements several interaction testing methods. To test for multiplicative statistical interactions between each SNP and environmental risk factors (MHT / estrogen-only / estrogen-progestogen), we primarily used conventional case-control logistic regression analysis, and two-step methods (21-23) to test the GxE interaction term. Additionally, we also used a 2-degree of freedom (2-d.f.) joint test (24) and 3-degree of freedom (3-d.f.) joint test (25) to test GxE interaction in the context of simultaneously testing for SNP-CRC and SNP-E (MHT / estrogen-only / estrogen-progestogen) associations. For the 2- and 3-d.f. test we do not report on known loci (16). For all novel findings we examined the ORs of MHT / estrogen-only / estrogen-progestogen stratified by genotypes of significant SNPs. More details in these testing methods can be found in the Supplementary Methods (available online).

For interaction analysis of rare genetic risk variants (MAF<1%) and MHT, we conducted the Mixed effects Score Tests for interaction (MiSTi) (26), a set-based statistical framework providing mixed effects score tests for GxE

interaction and addressing issues of power and low effect sizes, to discover genes that interact with MHT in relation to CRC risk (see the Supplementary Methods, available online). Since more than 20,000 genes were tested (22,476 genes for any MHT use, 20,609 for estrogen-only, and 20,360 for estrogen-progestogen), interactions with $P<2.5x10^{-6}$ were considered statistically significant, while those with $P<1.2x10^{-4}$ were considered as suggestive.

Functional annotation

We performed bioinformatic follow-up for genome-wide interaction study (GWIS) variants that were deemed statistically significant for downstream analysis (for more details, see the Supplementary Methods, available online). Relevant regional plots were generated using the command line version (Standalone) of LocusZoom v1.3 (27). Measures of linkage disequilibrium (LD) were estimated using study population controls.

Results

Detailed descriptive characteristics of the cases and controls are shown in Table 1. MHT use was associated with reduced CRC risk both in cohort studies and case-control studies (Figure 1-3).

Genome-wide MHT-interaction scans for CRC risk

Statistical interaction results for genetic variants are summarized in Table 2. While conventional case-control logistic regression models with a Bonferroni correction for multiple testing did not identify any statistically

significant interactions between the use of any MHT / estrogen-only / estrogen-progestogen and genetic variants (data not shown), we identified two interactions with common genetic variants reaching statistical significance for the two-step method and 2-d.f. joint test. The two-step method (with G|E in step1) identified a statistically significant interaction for any MHT use with SNP rs117868593 located 20kb downstream of *GRIN2B* variant at 12p13.1 (p-observed: 3.42×10^{-3}, p-threshold: 5×10^{-3}, Supplementary Figure 1, 2). The 2-d.f. joint test identified a further statistically significant interaction for any MHT use with a *DCBLD1* intronic variant at 6q22.1 (rs10782186; joint p-observed: 4.23×10^{-8}, p-threshold: 5×10^{-8}, Supplementary Figure 3, 4). Several *DCBLD1* intronic variants at 6q22.1 (rs4945586, rs9320604, rs4946260), which were in LD with rs10782186, also yielded low p-values using the 2-d.f. joint test although not genome-wide significant (5.28×10^{-8}, 5.60×10^{-8} and 5.70×10^{-8}; Supplementary Figure 3, 4). We did not identify any genome-wide statistically significant interactions between estrogen-only use or estrogen-progestogen use and common genetic variants for CRC risk. Common variants that reached the suggestive interaction level ($P<5 \times 10^{-6}$) with MHT use for CRC risk are shown in Supplementary table 2, 3, and 4, which included 87 SNPs with any MHT use, 80 with estrogen-only use and 137 with estrogen-progestogen use. We also performed GWIS stratified by colon and rectal cancer, but the common variant analysis did not yield any statistically significant interactions for the MHT variables, respectively (data not shown).

Table 3 presents associations of MHT use with CRC risk by the genotype of the two SNPs that were found to be significant. For rs117868593, there was a significant protective effect of any MHT use only among women with the *GG* homozygotes (OR: 0.68; 95%CI: 0.64-0.72; P=4.3x10^{-37}), but not in women with the *GC* genotype (OR: 0.91; 95%CI: 0.77-1.09; P=0.31), or with the *CC* genotype (OR: 0.64; 95%CI: 0.22-1.85; P=0.41). When stratified by MTH use, there was a significant per-minor allele association with CRC risk in users of any MHT (OR: 1.20; 95%CI: 1.05-1.37), but not in non-users (OR: 0.93; 95%CI: 0.83-1.03). For rs10782186, the protective effect of any MHT use compared with women not using any MHT was increasingly stronger for women with an increasing number of C alleles: *TT* (OR: 0.78; 95%CI: 0.70-0.87; P=4.3x10^{-6}), *TC* (OR: 0.68; 95%CI: 0.63-0.73; P=1.4x10^{-22}) and *CC* (OR: 0.66; 95%CI: 0.60-0.74; P=5.7x10^{-14}). When rs10782186 was investigated in relation to CRC risk among strata of MTH use, the per-minor allele OR for CRC risk was attenuated in users of any MHT (OR: 1.05; 95%CI: 0.99-1.11) compared to non-users (OR:1.14; 95%CI: 1.09-1.19).

The GxE interactions between rs117868593 or rs10782186 and any MHT were not heterogeneous across studies overall (P=0.98, 0.56, respectively) or stratified by study regions (North America, Australia, and Europe). The corresponding forest plots are shown in Supplementary Figures 5 and 6.

Rare variants for CRC risk

The rare variant analysis did not yield any statistically significant interactions ($P<2.5x10^{-6}$) for the MHT variables. However, several genes were found to reach the suggestive level for interaction ($P<1.2x10^{-4}$) for CRC risk, i.e., *PREX1* with any MHT use ($P=5.02x10^{-5}$), *SOS2* with estrogen-only therapy ($P=9.23x10^{-5}$), as well as *TMEM189-UBE2V1* ($P=2.46x10^{-5}$), *FAM149A* ($P=9.67x10^{-5}$), and *RPS13* ($P=1.02x10^{-5}$) with estrogen-progestogen therapy (Table 4, QQ-plots shown in Supplementary Figures 7-9).

Functional annotations of genetic loci

We performed bioinformatic analysis of the two loci showing significant interactions with MHT use (rs117868593 located 20kb downstream of *GRIN2B* variant at 12p13.1 and a *DCBLD1* intronic variant rs10782186 at 6q22.1). Annotation was performed for all variants tagged by the most significant SNPs ($r^2>0.5$) using our novel functional annotation analyses. The *GRIN2B* rs117868593 locus is in LD with rs17822202 (D'=0.93 and $r^2=0.85$ in 1000 Genomes Project CEU), which is downstream of the GRIN2B gene. We noted that this SNP was associated with more pronounced enhancer activity in colon tumor and cancer cell lines than in normal colon tissues (Supplementary Figure 10). The *DCBLD1* rs10782186 is in high LD with rs9320604 (D'=0.99 and $r^2=0.98$ in 1000 Genomes Project CEU), a SNP overlapping histone methylation patterns with enhancer activity in normal colon tissues, colon tumor, and cancer cell lines, and associated with strong

DNase hypersensitivity in tumor tissues (Supplementary Figure 11).

Based on BarcUVa-Seq eQTL analysis (Supplementary Methods available online), we identified four genes i.e., *EMP1*, *RPL13AP20*, *FAM234B* and *CDKN1B,* whose expression in normal colon tissue was significantly associated with the SNP rs117868593 or the SNPs in LD (R^2>0.5) (P<0.05) (Supplementary Table 5, Supplementary Figure 12), as well as two genes, i.e., *ROS1* and *GOPC,* with the SNP rs10782186 or the SNPs in LD (P<0.05) (Supplementary Table 6, Supplementary Figure 13). These eQTL effects persisted when restricting the sample to postmenopausal women although significant for rs10782186_ROS1, rs117868593_RPL13AP20 and rs1806217_FAM234B.

Discussion

We identified novel GxE interactions between the use of any MHT and common variants at two loci for CRC risk among postmenopausal women. The putative target genes underlying these interactions include *EMP1*, *RPL13AP20*, *FAM234B*, *CDKN1B, ROS1* and *GOPC.* In addition, we found suggestive interactions between the use of MHT and rare variants in *PREX1*, *SOS2*, *TMEM189-UBE2V1*, *FAM149A* and *RPS13*. Using independent samples in the current study, the previously found SNPs for GxE interactions (Supplementary Table 7) (7, 10) did not show statistically significant interaction with MHT with respect to CRC risk. These earlier studies used a candidate gene approach, different covariable adjustment, or different

exposure/non-exposure definitions compared with our GWAS study. Additionally, power could be further reduced by variations in the underlying distribution of MHT as new studies were introduced to the larger cohort.

Currently, the underlying etiologic mechanisms by which MHT affects CRC are not yet well understood. It is likely that protective cellular effects of estrogen and progesterone in the development of CRC are mediated through ESR1 (estrogen receptor α), ESR2 (estrogen receptor β), and PGR (progesterone receptor) (28-30). Estrogen and progestin may play a role in the pathway leading to DNA hyper-methylation (31, 32), which regulates gene expression including that of tumor suppressor genes and thereby play a crucial role in tumorigenesis of CRC. Estrogen has also been found to have an impact on a large number of serum proteome which plays a role in mucosal protection and repair in the gastrointestinal tract (33) as well as colon transcriptome (34). In addition, estrogen may contribute to maintaining the genomic stability in colonic epithelial cells by upregulation of mismatch repair genes (35). MHT use has also been reported to have growth inhibiting effects on colon cancer cells through upregulating cell cycle regulators, e.g., TP53 (36). Consortium efforts that are powered to explore the relationships of MHT with specific subtypes of CRC may yield further insights to GxE interactions with respect to hormonal contributions to the pathogenesis of CRC (37).

The SNP rs117868593 located about 20kb downstream from *GRIN2B* (glutamate receptor, ionotropic, N-methyl D-aspartate 2B) was not found to be

associated with expression of the nearest gene *GRIN2B* but with *EMP1*, *RPL13AP20*, *FAM234B* and *CDKN1B*. Expression of *EMP1* (epithelial membrane protein 1) has been found to be lower in human CRC than normal adjacent colorectal tissues (38) and overexpression of *EMP1* was observed to reduce proliferation and induce apoptosis of CRC cells (39), which are consistent with our findings, i.e., lower expression of *EMP1* and higher risk of CRC associated with G allele of rs117868593. We found the MHT users with GG have a stronger significant reduction of CRC risk, suggesting that EMP1 may function as an oncogene in hormone-dependent epithelium, which has been observed for EMP2, a paralog of EMP1 (40). Downregulation of CDKN1B (cyclin-dependent kinase inhibitor-1B, p27, kip1), which mainly results from increased ubiquitin-mediated proteasomal degradation, has been associated with tumor progression in CRC (41), and CDKN1B could be induced through ESR2 (estrogen receptor β)-mediated repression of the F-box protein p45 (SKP2) which has been identified as the substrate recognition component that targets and binds CDKN1B for ubiquitination and subsequent degradation (41-43). The link between CDKN1B and ESR2 might explain the observed interaction of *CDKN1B* with MHT. Potential mechanisms through which *RPL13AP20* and *FAM234B* act in modifying MHT associated CRC risk are unknown.

The region in which *DCBLD1* (encoding the discoidin, CUB, and LCCL domain-containing 1 protein) is located, chromosome 6q22.1, has been

reported as one of the suggestive susceptibility regions ($P=3.20x10^{-6}$) in a GWAS meta-analysis on CRC risk (12). Association estimates for the index SNP rs10782186 and correlated SNPs (rs4945586, rs9320604, and rs4946260) were reported in the supplementary tables of that above-mentioned GWAS paper. The significance ($P=4.23x10^{-8}$) of the interaction in our GWIS using the 2-d.f. joint test was mainly driven by the genetic association ($P=6.79x10^{-8}$) and was further strengthened by the GxE product term ($P=2.79x10^{-2}$). Thus, incorporating the GxE component helped to uncover genetic susceptibility variants for CRC risk, which did not reach genome-wide significance level in GWAS. Analyses of associated gene expression indicated the involvement of *ROS1* and *GOPC.* ROS1 (c-ros oncogene 1) is a transmembrane receptor tyrosine kinase that often shows genetic rearrangements in colorectal tumor tissue, such as intrachromosomal fusion with GOPC due to microdeletions at 6q22.1, which is highly prevalent in CRC (44, 45). GOPC-ROS fusion proteins have been shown to activate the downstream signaling pathway, signal transducers and activators of transcription-3 (STAT3) that play a significant role in progression of CRC (45, 46). The transcription factor of STAT3 in epithelial cells is activated by IL6, promoting CRC tumorigenesis (29, 47), whereas ESR2 mediates the downregulation of the inflammatory cytokine IL6 network (48), which may explain the observed interaction with MHT.

There are still considerable challenges in investigating GxE interaction

of rare genetic variants because of the scarcity of subjects with data on both these variants and the relevant environmental / lifestyle exposures. Therefore, the role that rare predisposition alleles play in modifying the association between environmental factors and CRC risk remains poorly understood. Our study used the set-based test (MiSTi) to tackle the challenge for GxE interaction analysis of rare variants, which strengthened statistical power to robustly uncover potential rare variant GxE association signals. Through this method, we found suggestive interaction for MHT use with rare variants in five genes for CRC risk. Despite their as yet unknown mechanisms in modifying CRC risk associated with MHT use, our application of GxE interaction analysis for CRC risk to rare variants alongside common variants represents a novel and rigorous approach. GxE interaction studies of rare genetic variants that incorporate functional genomic information ideally accounting for MHT effects and studies with larger sample sizes and hence with greater statistical power may contribute to understanding any missing heritability of cancer that remains unexplained by common variants.

Our study has several strengths. First, our large sample size, including more than 28,000 case and control participants, facilitated the most powerful scan for gene-MHT interaction to date. Second, we used recently developed statistical approaches that can provide greater statistical power than conventional case-control logistic regression (49). Since no single approach provides the best power across all possible patterns of GxE interaction, we

utilized a combination of approaches to maximize the chance of identifying novel loci in this discovery analysis. A novel statistical set-based score test for interaction, MiSTi, used for rare variant analysis, helped identify suggestive associations with CRC risk through interaction with MHT for five genes that warrant further follow-up. Third, we carefully harmonized environmental data on MHT use and other covariates across studies to minimize between-study heterogeneity bias as previously described (11). We acknowledge, however, that our analysis was limited to populations of European ancestry thus the results might not be generalizable to other race/ethnicity groups. Measurement error of the primarily self-reported exposure assessment might also have contributed to reduced power, however, previous studies have found the high validity for self-reported MHT use when compared with population-based prescription databases as references (50) and a high concordance between self-reported MHT use and that of physicians' reports (51). Despite our sizable sample size and use of advanced statistical methods we acknowledge that statistical power remains limited to detect small-to-modest sized interaction effects in a genome-wide scan setting. This might explain the relatively small number of novel findings. To overcome these issues, it will be critical to expand sample sizes of well characterized studies as well as incorporating functional genomic data relevant to CRC and MHT use, such as multi-omics data of normal and tumor colon tissue exposed and unexposed to MHT.

From a comprehensive genome-wide GxE interaction investigation, we identified two common loci, which were significantly associated with CRC risk in conjunction with MHT use, as well as five genes, which showed suggestive evidence of GxE interaction through rare variant set analysis. The putative target genes of the two identified loci, *EMP1*, *RPL13AP20, FAM234B*, *CDKN1B*, *ROS1* and *GOPC*, may explain the GxE interactions with MHT and offer new insights into CRC etiological mechanisms and pathways of CRC carcinogenesis. Further downstream follow-up studies for exploring potential genetic functions are warranted to confirm the involvement of these genetic variants or genes in CRC risk associated with MHT use.

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 CA137088, R01 CA059045, U01 CA164930, R01201407). Genotyping/ services were provided by the Center for Inherited Disease Research (CIDR) contract number HHSN268201200008I. This research was funded in part through the NIH/NCI Cancer Center Support Grant P30 CA015704. Scientific Computing Infrastructure at Fred Hutch funded by ORIP grant S10OD028685.

CLUE II: 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.

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). Support for case ascertainment was provided in part from the Surveillance, Epidemiology, and End Results (SEER) Program and the following U.S. state cancer registries: AZ, CO, MN, NC, NH; and by the Victoria Cancer Registry (Australia) and Ontario Cancer Registry (Canada). The CCFR Set-1 (Illumina 1M/1M-Duo) scan was 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. Additional funding for the OFCCR/ARCTIC was through award GL201-043 from the Ontario Research Fund (to BWZ), award 112746 from the Canadian Institutes of Health Research (to TJH), through a Cancer Risk Evaluation (CaRE) Program grant from the Canadian

Cancer Society (to SG), and through generous support from the Ontario Ministry of Research and Innovation. The SFCCR Illumina HumanCytoSNP array was supported in part through NCI/NIH awards U01/U24 CA074794 and R01 CA076366 (to PAN). 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.

COLO2&3: National Institutes of Health (R01 CA60987).

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

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), Plataforma Biobancos PT13/0010/0013 and ICOBIOBANC, sponsored by the Catalan Institute of Oncology.

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

DALS: National Institutes of Health (R01 CA48998 to M. L. Slattery).

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 Potsdam- Rehbruecke (DIfE), Federal Ministry of

Education and Research (BMBF) (Germany); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy, Compagnia di SanPaolo and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands); Health Research Fund (FIS) - Instituto de Salud Carlos III (ISCIII), Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, and the Catalan Institute of Oncology - ICO (Spain); Swedish Cancer Society, Swedish Research Council and 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).

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

Harvard cohort (NHS): National Institutes of Health (R01 CA137178, P01 CA087969, UM1 CA186107, K24 DK098311, R01 CA151993, and R35 CA197735).

Kentucky: 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: The 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: National Institutes of Health (R37 CA054281, P01 CA033619, and R01 CA063464).

MECC: National Institutes of Health, U.S. Department of Health and Human Services (R01 CA081488, R01 CA197350).

NCCCS I & II: 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 Institutes of Health, U.S. Department of Health and Human Serivces (U01 CA74783); and National Cancer Institute of Canada grants (18223 and 18226). The authors wish to acknowledge the contribution of Alexandre Belisle and the genotyping team of the McGill University and Génome Québec Innovation Centre, Montréal, Canada, for genotyping the

Sequenom panel in the NFCCR samples. Funding was provided to Michael O. Woods by the Canadian Cancer Society Research Institute.

NSHDS: The research was supported by Biobank Sweden through funding from the Swedish Research Council (VR 2017-00650, VR 2017-01737), the Swedish Cancer Society (CAN 2017/581), Region Västerbotten (VLL-841671, VLL-833291), Knut and Alice Wallenberg Foundation (VLL-765961), and the Lion's Cancer Research Foundation (several grants) and Insamlingsstiftelsen, both at Umeå University.

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.

REACH: 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.)

SMC_COSM: 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.

UK Biobank: This research has been conducted using the UK Biobank

Resource under Application Number 8614

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.

Notes

Role of the funder: The funders had no role in the design of the study; the collection, analysis, and interpretation of the data; the writing of the manuscript; and the decision to submit the manuscript for publication.

Disclosures: All authors have no conflicts of interest to disclose. KV, who is a *JNCI* Associate Editor and co-author on this paper, was not involved in the editorial review or decision to publish the manuscript.

Author contributions: Writing- original draft: YT, AEK. Analysis- statistical methods: YT, AEK, YL, CQ, VDO. Supervising: UP, WJG, LH, JCC. Writing- reviewing & editing: YT, AEK, SAB, YL, CQ, TH, RCT, VDO, ND, DAD, AH, JRH, KMJ, JM, NM, MOS, CMU, JO, ARP, EAR, AS, MS, YS, FJD, VA, JB, SIB, DTB, HB, DDB, ATC, JCF, SG, SBG, SH, MH, MAJ, ADJ, TOK, SCL, LLM, LL, GGG, RLM, HN, RN, SO, AB, EAP, JDP, RLP, GR, LCS, RES, MLS, SNT, BVG, KV, EW, AW, MOW, AHW, PTC, GC, DVC, MJG, AK, JPL, VM,

PAN, BP, DCT, KKT, UP, WJG, LH, JCC.

Prior presentations

The content in the manuscript was partly reported as poster presentation in the 2nd International DKFZ Conference on Cancer Prevention on September 17-18, 2020.

Disclaimer

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.

Acknowledgments

We would like to thank Ferran Moratalla Navarro [Oncology Data Analytics Program, Catalan Institute of Oncology (ICO). L'Hospitalet de Llobregat, Barcelona, Spain. Colorectal Cancer Group, ONCOBELL Program, Bellvitge Biomedical Research Institute (IDIBELL). L'Hospitalet de Llobregat, Barcelona, Spain. Consortium for Biomedical Research in Epidemiology and Public Health, (CIBERESP), Spain. Department of Clinical Sciences, Faculty of Medicine, University of Barcelona, Barcelona, Spain] for providing a further eQTL analysis for the revision of our manuscript.

CCFR: We graciously thanks the generous contributions of CCFR study participants, dedication of study staff, and the financial support from the

U.S. National Cancer Institute, without which this important registry would not exist. The authors would like to thank the study participants and staff of the Seattle Colon Cancer Family Registry and the Hormones and Colon Cancer study (CORE Studies).

CLUE II: We thank the participants of Clue II and appreciate the continued efforts of the staff at the Johns Hopkins George W. Comstock Center for Public Health Research and Prevention in the conduct of the Clue II Cohort Study. Cancer data were provided by the Maryland Cancer Registry, Center for Cancer Prevention and Control, Maryland Department of Health, with funding from the State of Maryland and the Maryland Cigarette Restitution Fund. The collection and availability of cancer registry data is also supported by the Cooperative Agreement NU58DP006333, funded by the Centers for Disease Control and Prevention. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention or the Department of Health and Human Services.

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.

DACHS: We thank all participants and cooperating clinicians, and everyone who provided excellent technical assistance.

EPIC: We acknowledge the contributions of EPIC investigators, staff, and all participants.

Harvard cohorts (NHS): 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.

NSHDS: We thank the Västerbotten Intervention Programme, the Northern Sweden MONICA study, the Biobank Research Unit at Umeå

University and Biobanken Norr at Region Västerbotten for providing data and samples and acknowledge the contribution from Biobank Sweden, supported by the Swedish Research Council.

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.

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

Data Availability

Genotype data as well as limited phenotype data have been deposited at dbGaP under accession numbers phs001415.v1.p1, phs001315.v1.p1, phs001078.v1.p1, phs001499.v1.p1, phs001903.v1.p1, phs001856.v1.p1, phs001045.v1.p1, and phs001499.v1.p1. Further information is available from the corresponding authors upon request.

References

1. Lin KJ, Cheung WY, Lai JY-C, Giovannucci EL. The effect of estrogen vs. combined estrogen-progestogen therapy on the risk of colorectal cancer. 2012;130(2):419-30.

2. Botteri E, Støer NC, Sakshaug S, Graff-Iversen S, Vangen S, Hofvind S, et al. Menopausal hormone therapy and colorectal cancer: a linkage between nationwide registries in Norway. 2017;7(11):e017639.

3. Mørch LS, Lidegaard Ø, Keiding N, Løkkegaard E, Kjær SK. The influence of hormone therapies on colon and rectal cancer. European journal of epidemiology. 2016;31(5):481-9.

4. Rennert G, Rennert HS, Pinchev M, Lavie O, Gruber SB. Use of hormone replacement therapy and the risk of colorectal cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2009;27(27):4542-7.

5. Lin J, Zee RYL, Liu K-Y, Zhang SM, Lee IM, Manson JE, et al. Genetic variation in sex-steroid receptors and synthesizing enzymes and colorectal cancer risk in women. Cancer Causes & Control. 2010;21(6):897-908.

6. Lin JH, Manson JE, Kraft P, Cochrane BB, Gunter MJ, Chlebowski RT, et al. Estrogen and

progesterone-related gene variants and colorectal cancer risk in women. BMC medical genetics. 2011;12(1):78.

7. Rudolph A, Sainz J, Hein R, Hoffmeister M, Frank B, Försti A, et al. Modification of menopausal hormone therapy-associated colorectal cancer risk by polymorphisms in sex steroid signaling, metabolism and transport related genes. Endocrine-related cancer. 2011;18(3):371-84.

8. Slattery ML, Lundgreen A, Herrick JS, Kadlubar S, Caan BJ, Potter JD, et al. Variation in the CYP19A1 gene and risk of colon and rectal cancer. Cancer Causes & Control. 2011;22(7):955-63.

9. Kantor ED, Hutter CM, Minnier J, Berndt SI, Brenner H, Caan BJ, et al. Gene–Environment Interaction Involving Recently Identified Colorectal Cancer Susceptibility Loci. 2014;23(9):1824-33.

10. Garcia-Albeniz X, Rudolph A, Hutter C, White E, Lin Y, Rosse SA, et al. CYP24A1 variant modifies the association between use of oestrogen plus progestogen therapy and colorectal cancer risk. British journal of cancer. 2016;114(2):221-9.

11. Hutter CM, Chang-Claude J, Slattery ML, Pflugeisen BM, Lin Y, Duggan D, et al. Characterization of gene-environment interactions for colorectal cancer susceptibility loci. Cancer research. 2012;72(8):2036-44.

12. Peters U, Jiao S, Schumacher FR, Hutter CM, Aragaki AK, Baron JA, et al. Identification of Genetic Susceptibility Loci for Colorectal Tumors in a Genome-Wide Meta-analysis. Gastroenterology. 2013;144(4):799-807.e24.

13. Schmit SL, Edlund CK, Schumacher FR, Gong J, Harrison TA, Huyghe JR, et al. Novel

Common Genetic Susceptibility Loci for Colorectal Cancer. J Natl Cancer Inst. 2019;111(2):146-57.

14. Gong J, Hutter CM, Newcomb PA, Ulrich CM, Bien SA, Campbell PT, et al. Genome-Wide Interaction Analyses between Genetic Variants and Alcohol Consumption and Smoking for Risk of Colorectal Cancer. PLOS Genetics. 2016;12(10):e1006296.

15. Jeon J, Du M, Schoen RE, Hoffmeister M, Newcomb PA, Berndt SI, et al. Determining Risk of Colorectal Cancer and Starting Age of Screening Based on Lifestyle, Environmental, and Genetic Factors. Gastroenterology. 2018;154(8):2152-64.e19.

16. 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. Nature genetics. 2019;51(1):76-87.

17. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nature genetics. 2016;48(10):1284-7.

18. John Morrison (2020). BinaryDosage: Creates, Merges, and Reads Binary Dosage Files. R package version 1.0.0. https://CRAN.R-project.org/package=BinaryDosage.

19. Cochran WG. The Combination of Estimates from Different Experiments. Biometrics. 1954;10(1):101-29.

20. John Morrison (2020). GxEScanR: Run GWAS/GWEIS Scans Using Binary Dosage Files. R package version 2.0.2. https://CRAN.R-project.org/package=GxEScanR.

21. Kooperberg C, Leblanc M. Increasing the power of identifying gene x gene interactions in genome-wide association studies. Genet Epidemiol. 2008;32(3):255-63.

22. Murcray CE, Lewinger JP, Gauderman WJ. Gene-environment interaction in genome-wide

association studies. American journal of epidemiology. 2009;169(2):219-26.

23. Gauderman WJ, Zhang P, Morrison JL, Lewinger JP. Finding novel genes by testing G × E interactions in a genome-wide association study. Genet Epidemiol. 2013;37(6):603-13.

24. Kraft P, Yen YC, Stram DO, Morrison J, Gauderman WJ. Exploiting gene-environment interaction to detect genetic associations. Human heredity. 2007;63(2):111-9.

25. Gauderman WJ, Kim A, Conti DV, Morrison J, Thomas DC, Vora H, et al. A Unified Model for the Analysis of Gene-Environment Interaction. American journal of epidemiology. 2019;188(4):760-7.

26. Su YR, Di CZ, Hsu L. A unified powerful set-based test for sequencing data analysis of GxE interactions. Biostatistics (Oxford, England). 2017;18(1):119-31.

27. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010;26(18):2336-7.

28. Barzi A, Lenz AM, Labonte MJ, Lenz H-J. Molecular Pathways: Estrogen Pathway in Colorectal Cancer. 2013;19(21):5842-8.

29. Caiazza F, Ryan EJ, Doherty G, Winter DC, Sheahan K. Estrogen receptors and their implications in colorectal carcinogenesis. Frontiers in oncology. 2015;5:19.

30. Williams C, DiLeo A, Niv Y, Gustafsson J-Å. Estrogen receptor beta as target for colorectal cancer prevention. Cancer Letters. 2016;372(1):48-56.

31. Newcomb PA, Zheng Y, Chia VM, Morimoto LM, Doria-Rose VP, Templeton A, et al. Estrogen Plus Progestin Use, Microsatellite Instability, and the Risk of Colorectal Cancer in Women. 2007;67(15):7534-9.

32. Neumeyer S, Popanda O, Butterbach K, Edelmann D, Bläker H, Toth C, et al. DNA methylation profiling to explore colorectal tumor differences according to menopausal hormone therapy use in women. 2019;11(16):1765-78.

33. Katayama H, Paczesny S, Prentice R, Aragaki A, Faca VM, Pitteri SJ, et al. Application of serum proteomics to the Women's Health Initiative conjugated equine estrogens trial reveals a multitude of effects relevant to clinical findings. Genome Medicine. 2009;1(4):47.

34. Hases L, Archer A, Indukuri R, Birgersson M, Savva C, Korach-André M, et al. High-fat diet and estrogen impacts the colon and its transcriptome in a sex-dependent manner. Scientific Reports. 2020;10(1):16160.

35. Slattery ML, Potter JD, Curtin K, Edwards S, Ma KN, Anderson K, et al. Estrogens reduce and withdrawal of estrogens increase risk of microsatellite instability-positive colon cancer. Cancer research. 2001;61(1):126-30.

36. Hsu HH, Cheng SF, Wu CC, Chu CH, Weng YJ, Lin CS, et al. Apoptotic effects of over-expressed estrogen receptor-beta on LoVo colon cancer cell is mediated by p53 signalings in a ligand-dependent manner. Chin J Physiol. 2006;49(2):110-6.

37. Ogino S, Chan AT, Fuchs CS, Giovannucci E. Molecular pathological epidemiology of colorectal neoplasia: an emerging transdisciplinary and interdisciplinary field. Gut. 2011;60(3):397-411.

38. Sun GG, Wang YD, Cui DW, Cheng YJ, Hu WN. Epithelial membrane protein 1 negatively regulates cell growth and metastasis in colorectal carcinoma. World journal of gastroenterology. 2014;20(14):4001-10.

39. Ahmat Amin MKB, Shimizu A, Ogita H. The Pivotal Roles of the Epithelial Membrane

Protein Family in Cancer Invasiveness and Metastasis. Cancers (Basel). 2019;11(11):1620.

40. Wang YW, Cheng HL, Ding YR, Chou LH, Chow NH. EMP1, EMP 2, and EMP3 as novel therapeutic targets in human cancer. Biochim Biophys Acta Rev Cancer. 2017;1868(1):199-211.

41. Ogino S, Kawasaki T, Kirkner GJ, Yamaji T, Loda M, Fuchs CS. Loss of nuclear p27 (CDKN1B/KIP1) in colorectal cancer is correlated with microsatellite instability and CIMP. Modern Pathology. 2007;20(1):15-22.

42. Kondakova IV, Shashova EE, Sidenko EA, Astakhova TM, Zakharova LA, Sharova NP. Estrogen Receptors and Ubiquitin Proteasome System: Mutual Regulation. Biomolecules. 2020;10(4):500.

43. Hartman J, Edvardsson K, Lindberg K, Zhao C, Williams C, Ström A, et al. Tumor Repressive Functions of Estrogen Receptor β in SW480 Colon Cancer Cells. Cancer research. 2009;69(15):6100.

44. Consortium APG. AACR Project GENIE: Powering Precision Medicine through an International Consortium. Cancer Discov. 2017;7(8):818-31.

45. Davies KD, Doebele RC. Molecular pathways: ROS1 fusion proteins in cancer. Clin Cancer Res. 2013;19(15):4040-5.

46. Huang Y, Wang J, Cao F, Jiang H, Li A, Li J, et al. SHP2 associates with nuclear localization of STAT3: significance in progression and prognosis of colorectal cancer. Scientific reports. 2017;7(1):17597-.

47. Grivennikov S, Karin E, Terzic J, Mucida D, Yu G-Y, Vallabhapurapu S, et al. IL-6 and Stat3 are required for survival of intestinal epithelial cells and development of colitis-associated cancer. Cancer Cell. 2009;15(2):103-13.

48. Edvardsson K, Ström A, Jonsson P, Gustafsson J, Williams C. Estrogen receptor β induces antiinflammatory and antitumorigenic networks in colon cancer cells. Molecular endocrinology (Baltimore, Md). 2011;25(6):969-79.

49. Gauderman WJ, Mukherjee B, Aschard H, Hsu L, Lewinger JP, Patel CJ, et al. Update on the State of the Science for Analytical Methods for Gene-Environment Interactions. American journal of epidemiology. 2017;186(7):762-70.

50. Løkkegaard EL, Johnsen SP, Heitmann BL, Stahlberg C, Pedersen AT, Obel EB, et al. The validity of self-reported use of hormone replacement therapy among Danish nurses. Acta Obstetricia et Gynecologica Scandinavica. 2004;83(5):476-81.

51. Kropp S, Terboven T, Hedicke J, Mutschelknauss E, Slanger T, Braendle W, et al. Good agreement between physician and self-reported hormone therapy data in a case-control study. J Clin Epidemiol. 2007;60(12):1280-7.

Table 1. Descriptive characteristics of study participants included in the genome-wide interaction analysis between common variants and menopausal hormone therapy for risk of colorectal cancer

Study	Case participants						Control participants					
	N	No use of MHT n (%)	Any MHT n (%)	E-only n	E+P n	Age at diagnosis mean (SD)	N	No use of MHT n (%)	Any MHT n (%)	E-only n	E+P n	Age at enrollment mean (SD)
CCFR Set 1	259	183 (70.7)	76 (29.3)	35	34	58.5 (9.9)	372	206 (55.4)	166 (44.6)	93	65	61.7 (7.9)
CCFR Set 3	427	292 (68.4)	135 (31.6)	68	42	61.8 (7.6)	250	155 (62.0)	95 (38.0)	54	27	62.7 (7.4)
CCFR Set 4	383	259 (67.6)	124 (32.4)	78	41	62.1 (9.3)	118	77 (65.3)	41 (34.7)	16	18	61.7 (9.3)
CLUEII	114	98 (86)	16 (14.0)	-	-	74.9 (9.5)	108	97 (89.8)	11 (10.2)	-	-	65.0 (9.4)
Colo 2&3	37	18 (48.6)	19 (51.4)	-	-	66.5 (11.4)	44	17 (38.6)	27 (61.4)	-	-	67.3 (9.2)
CPSII_1	263	176 (66.9)	87 (33.1)	50	37	74.8 (5.9)	255	142 (55.7)	113 (44.3)	83	30	74.3 (5.8)
CPSII_2	172	116 (67.4)	56 (32.6)	35	21	79.4 (6.1)	177	101 (57.1)	76 (42.9)	46	30	79.0 (6.0)
CRCGEN	274	266 (97.1)	8 (2.9)	-	-	68.8 (9.9)	394	377 (95.7)	17 (4.3)	-	-	65.9 (9.2)
DACHS_1	630	416 (66.0)	214 (34.0)	-	-	71.1 (9.5)	630	294 (46.7)	336 (53.3)	-	-	70.4 (8.7)
DACHS_2	229	161 (70.3)	68 (29.7)	-	-	72.2 (9.8)	162	88 (54.3)	74 (45.7)	-	-	72.3 (9.0)
DACHS_3	420	297 (70.7)	123 (29.3)	-	-	71.4 (9.5)	195	113 (57.9)	82 (42.1)	-	-	70.3 (10.1)
DALS_1	267	204 (76.4)	63 (23.6)	-	-	68.0 (7.7)	270	189 (70.0)	81 (30.0)	-	-	67.7 (7.9)
DALS_2	159	127 (79.9)	32 (20.1)	-	-	67.5 (7.5)	194	137 (70.6)	57 (29.4)	-	-	67.5 (8.2)
EPIC	771	544 (70.6)	227 (29.4)	-	-	67.2 (6.6)	865	619 (71.6)	246 (28.4)	-	-	72.5 (5.9)
ESTHER_VERDI	70	52 (74.3)	18 (25.7)	-	-	68.4 (6.8)	70	49 (70.0)	21 (30.0)	-	-	65.8 (6.7)
Kentucky	397	184 (46.3)	213 (53.7)	100	56	64.4 (8.9)	525	150 (28.6)	375 (71.4)	166	86	66.7 (6.6)
LCCS	116	90 (77.6)	26 (22.4)	-	-	66.1 (6.9)	108	88 (81.5)	20 (18.5)	-	-	65.7 (5.5)
MCCS_1	211	159 (75.4)	52 (24.6)	-	-	72.0 (7.1)	184	132 (71.7)	52 (28.3)	-	-	71.1 (7.2)
MCCS_2	85	65 (76.5)	20 (23.5)	-	-	74.3 (8.4)	86	65 (75.6)	21 (24.4)	-	-	73.8 (8.0)
MEC_1	99	55 (55.6)	44 (44.4)	27	-	70.3 (7.9)	115	42 (36.5)	73 (63.5)	37	-	70.3 (7.6)
MEC_2	15	2 (13.3)	13 (86.7)	5	-	80.1 (6.2)	30	4 (13.3)	26 (86.7)	12	-	74.6 (6.1)
MECC_3	309	260 (84.1)	49 (15.9)	-	-	69.5 (10.3)	367	290 (79.0)	77 (21.0)	-	-	73.0 (10.0)
NCCCSII	219	128 (58.4)	91 (41.6)	-	-	63.8 (9.8)	221	89 (40.3)	132 (59.7)	-	-	65.4 (9.4)
NFCCR_2	60	51 (85.0)	9 (15.0)	-	-	61.1 (7.9)	130	104 (80.0)	26 (20.0)	-	-	60.2 (7.2)
NHS_1_2	328	174 (53.0)	154 (47.0)	23	7	68.0 (7.4)	673	321 (47.7)	352 (52.3)	42	7	68.5 (6.9)
NHS_3_AD	410	187 (45.6)	223 (54.4)	21	10	68.1 (6.7)	335	133 (39.7)	202 (60.3)	15	7	67.9 (6.7)

PLCO_1_Rematch	216	125 (57.9)	91 (42.1)	-	-	68.8 (6.0)		123	61 (49.6)	62 (50.4)	-	-	67.5 (6.2)
PLCO_2	196	110 (56.1)	86 (43.9)	-	-	70.6 (6.6)		163	90 (55.2)	73 (44.8)	-	-	70.6 (6.3)
PLCO_3	295	157 (53.2)	138 (46.8)	-	-	67.0 (7.2)		1964	900 (45.8)	1064 (54.2)	-	-	62.1 (5.3)
PLCO_4_AD	434	241 (55.5)	193 (44.5)	-	-	64.0 (5.9)		587	274 (46.7)	313 (53.3)	-	-	61.9 (5.3)
REACH_AD	9	7 (77.8)	2 (22.2)	-	-	62.9 (4.0)		75	47 (62.7)	28 (37.3)	-	-	62.3 (5.6)
SMC_COSM	179	90 (50.3)	89 (49.7)	-	-	69.7 (9.7)		330	145 (43.9)	185 (56.1)	-	-	64.6 (7.8)
UKB_1	1073	996 (92.8)	77 (7.2)	-	-	65.4 (5.3)		4254	3928 (92.3)	326 (7.7)	-	-	65.4 (5.3)
USC_HRT_CRC	296	127 (42.9)	169 (57.1)	75	67	66.3 (5.5)		400	150 (37.5)	250 (62.5)	116	82	65.0 (6.8)
VITAL	114	61 (53.5)	53 (46.5)	-	-	70.5 (6.4)		126	60 (47.6)	66 (52.4)	-	-	71.5 (6.6)
WHI_1	450	297 (66.0)	153 (34.0)	95	58	71.0 (7.1)		519	282 (54.3)	237 (45.7)	137	100	71.2 (7.0)
WHI_2	977	576 (59.0)	401 (41.0)	202	199	72.2 (7.4)		990	512 (51.7)	478 (48.3)	260	217	72.0 (7.2)
WHI_3	556	313 (56.3)	243 (43.7)	117	126	78.6 (6.9)		558	267 (47.8)	291 (52.2)	148	142	78.5 (6.9)
Total	11519	7664 (66.5)	3855 (33.5)	931	698	-		16967	10795 (63.6)	6172 (36.4)	1225	811	-

N, number; MHT, menopausal hormone therapy; E-only, estrogen only; E+P, combined estrogen-progestogen; SD, standard deviation; CCFR, Colon Cancer Family Registry; CLUEII, Campaign against Cancer and Heart Disease II; Colo 2&3, Hawaii Colorectal Cancer Studies 2 & 3; CPSII, Cancer Prevention Study-II; CRCGEN, Colorectal Cancer Genetics & Genomics; DACHS, Darmkrebs: Chancen der Verhütung durch Screening; DALS, Diet, Activity, and Lifestyle Study; EPIC, European Prospective Investigation into Cancer; ESTHER_VERDI, Epidemiologische Studie zu Chancen der Verhütung, Früherkennung und optimierten THerapie chronischer ERkrankungen in der älteren Bevölkerung; Kentucky, Kentucky Case-Control Study; LCCS, Leeds Colorectal Cancer Study; MCCS, Melbourne Collaborative Cohort Study; MEC, Multiethnic Cohort Study; MECC, Molecular Epidemiology of Colorectal Cancer Study; NCCCSII, The North Carolina Colon Cancer Study II; NFCCR, Newfoundland Case-Control Study; NHS, Nurses' Health Study; PLCO, Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial; REACH, Colon Cancer Pathways: Hyperplastic Polyps and Adenomas; SMC_COSM, Swedish Mammography Cohort and Swedish Men Cohort; UKB, UK Biobank; USC_HRT_CRC, University of Southern California Hormone Replacement Therapy Colorectal Cancer Study; VITAL, Cancer Screening Trial VITamins And Lifestyle cohort; WHI, The Women's Health Initiative

Table 2. Results of genome-wide interaction analyses with menopausal hormone therapy for colorectal cancer risk among postmenopausal women

MHT Type	SNP	Chr	BP Position	Locus	Gene	Count Allele	Count Allele frequency	Statistical method used to detect the GxMHT interaction	P-value threshold for GxMHT interaction	P-value observed for GxMHT interaction	P-value for heterogeneity	No. of studies included
Any MHT	rs117868593	12	13670508	12p13.1	*GRIN2B*	C	0.05	Two-step method (by G|E in step 1)	$5x10^{-3}$	$3.42x10^{-3}$	0.98	38
Any MHT	rs10782186	6	117823508	6q22.1	*DCBLD1*	C	0.50	2-d.f. joint test	$5x10^{-8}$	$4.23x10^{-8}$	0.56	38

MHT, menopausal hormone therapy; SNP, single nucleotide polymorphism; Chr, chromosome; BP Position, base pair position based on NCBI Build37; G|E, associations between G and E in the combined case-control population; 2-d.f., 2-degree of freedom. Notes: Directly genotyped SNPs were coded as 0, 1, or 2 copies of the count allele. Imputed SNPs were coded as expected gene dosage. Multiplicative interaction terms were modelled as the product of MHT and each SNP of interest. All statistical tests were two-sided.

Table 3. Associations with colorectal cancer risk stratified by use of any menopausal hormone therapy and genotypes of SNPs of interest

SNP	MHT use	Genotype of SNP										
		Homozygous non-carriers			Heterozygous			Homozygous carries of the minor allele			Per minor allele within strata of MHT use	
		N Ca/Co	OR (95% CI)	*P*	N Ca/Co	OR (95% CI)	*P*	N Ca/Co	OR (95% CI)	P	OR (95% CI)	*P*
		GG			GC			CC			per C allele within strata of MHT use	
rs117868593	No	6991.4/9745.2	1.00 (Ref.)	-	652.6/1026.1	0.91 (0.81-1.03)	0.13	20/23.7	1.09 (0.53-2.22)	0.81	0.93 (0.83-1.03)	0.17
	Yes	3390.5/5537.1	0.68 (0.64-0.72)	$4.3x10^{-37}$	450.7/614.9	0.83 (0.72-0.96)	0.011	13.8/20.1	0.69 (0.31-1.54)	0.37	1.20 (1.05-1.37)	$7.7x10^{-3}$
	OR (95% CI)	-	0.68 (0.64-0.72)	$4.3x10^{-37}$	-	0.91 (0.77-1.09)	0.31	-	0.64 (0.22-1.85)	0.41		
		TT			TC			CC			per C allele within strata of MHT use	
rs10782186	No	1861.2/2936.4	1.00 (Ref.)	-	3806.9/5361.8	1.15 (1.07-1.24)	$3.5x10^{-4}$	1995.9/2496.9	1.29 (1.18-1.41)	$1.9x10^{-8}$	1.14 (1.09-1.19)	$1.8x10^{-8}$
	Yes	993/1624.8	0.78 (0.70-0.87)	$4.3x10^{-6}$	1861.9/3068.3	0.78 (0.71-0.85)	$3.8x10^{-8}$	1000/1478.9	0.85 (0.77-0.95)	$3.8x10^{-3}$	1.05 (0.99-1.11)	0.14
	OR (95% CI)	-	0.78 (0.70-0.87)	$4.3x10^{-6}$	-	0.68 (0.63-0.73)	$1.4x10^{-22}$	-	0.66 (0.60-0.74)	$5.7x10^{-14}$		

SNP, single nucleotide polymorphism; MHT, menopausal hormone therapy; N, number; Ca/Co, case/control; OR, odds ratio; CI, confidence interval; *P*, probability value. Case/control counts were calculated by imputed genotype probabilities.

Table 4. Suggestive association ($P<1.2x10^{-4}$) of genes from rare variants analyses of GxE with menopausal hormone therapy for colorectal cancer risk among postmenopausal women

MHT type	Gene	Gene name	Chr	N of SNPs	*P* value
Any MHT	ENSG00000124126	*PREX1*	20	45	$5.02x10^{-5}$
E-only	ENSG00000100485	*SOS2*	14	15	$9.23 x10^{-5}$
E+P	ENSG00000124208	*TMEM189-UBE2V1*	20	57	$2.46x10^{-5}$
E+P	ENSG00000109794	*FAM149A*	4	8	$9.67x10^{-5}$
E+P	ENSG00000110700	*RPS13*	11	5	$1.02x10^{-4}$

MHT, menopausal hormone therapy; E-only, estrogen only; E+P, combined estrogen-progestogen; Chr, chromosome; N of SNPs, number of SNPs in gene; P value, Fisher's P value by the set-based score (MiSTi) test.

Figure legends

Figure 1. Association of any menopausal hormone therapy use with the risk of colorectal cancer. N, number; OR, odds ratio; CI, confidence interval; P, probability value.

Figure 2. Association of use of estrogen only with the risk of colorectal cancer. N, number; OR, odds ratio; CI, confidence interval; P, probability value.

Figure 3. Association of use of combined estrogen-progestogen with the risk of colorectal cancer. N, number; OR, odds ratio; CI, confidence interval; P, probability value.

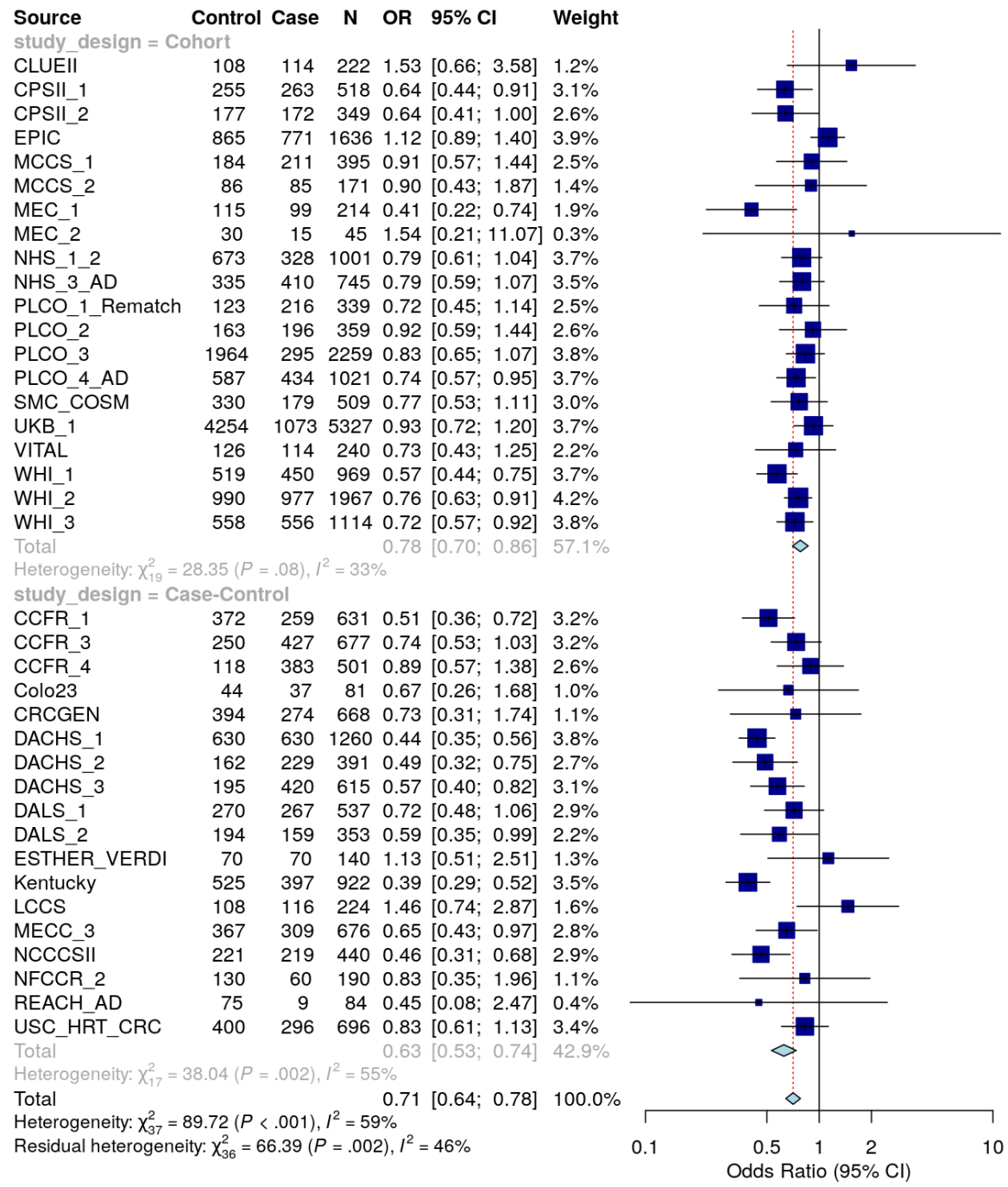

Figure 1

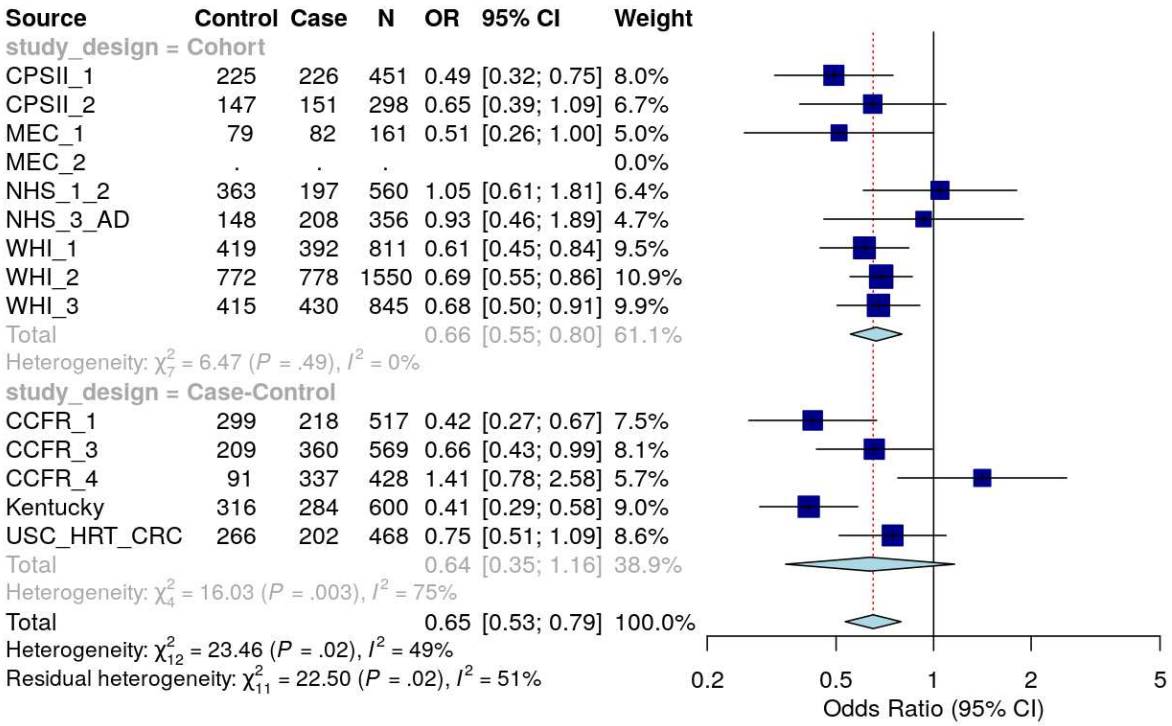

Figure 2

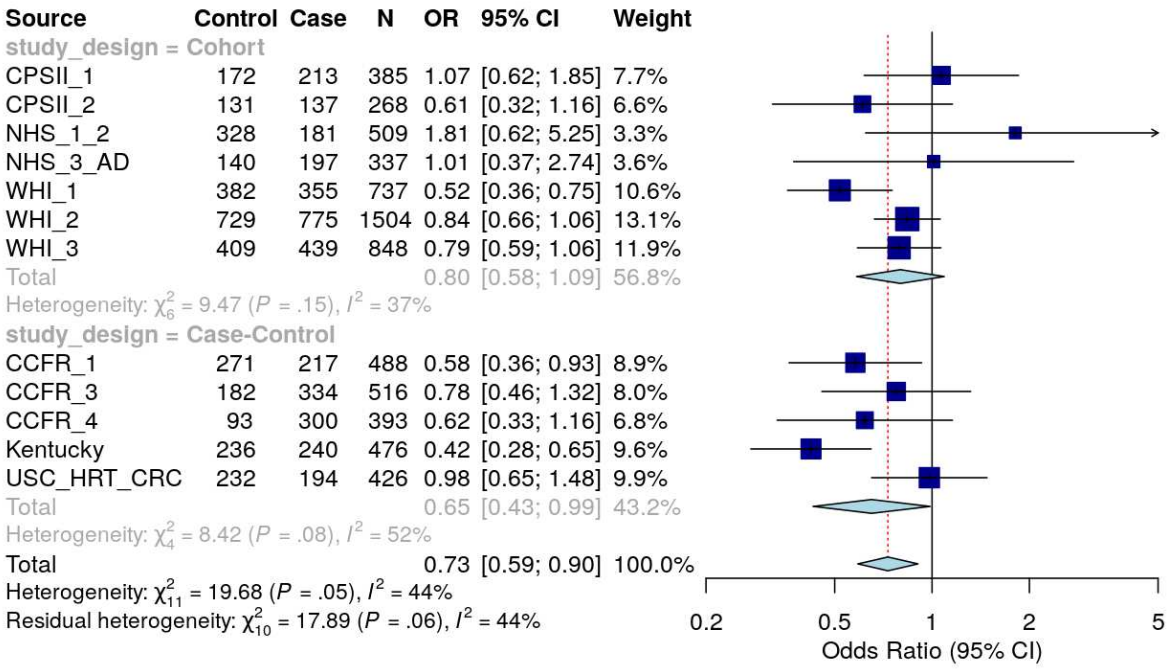

Figure 3