Mendelian randomization study of age at menarche and age at menopause and the risk of colorectal cancer

Running title: Age at menarche/menopause and colorectal cancer

Sonja Neumeyer[1], Barbara L. Banbury[2], Volker Arndt[3], Sonja I. Berndt[4], Stephane Bezieau[5], Stephanie A. Bien[2], Dan D Buchanan[6,7], Katja Butterbach[1], Bette J. Caan[8], Peter T. Campbell[9], Graham Casey[10], Andrew T. Chan[11,12,13], Stephen J. Chanock[14], James Y Dai[2], Steven Gallinger[15], Edward L. Giovannucci[12,16,17,18], Graham G. Giles[7,19], William M. Grady[20,21], Jochen Hampe[22], Michael Hoffmeister[3], John L. Hopper[7], Li Hsu[2], Mark A. Jenkins[7], Amit Joshi[13,16], Susanna C. Larsson[23], Loic Le Marchand[24], Annika Lindblom[25,26], Victor Moreno[27,28], Mathieu Lemire[29], Li Li[30], Yi Lin[2], Kenneth Offit[31], Polly A. Newcomb[2], Paul D. Pharaoh[32], John D. Potter[2], Lihong Qi[33], Gad Rennert[34,35,36], Clemens Schafmayer[37], Robert E. Schoen[38], Martha L. Slattery[39], Mingyang Song[11,13,17], Cornelia M. Ulrich[40], Aung K. Win[7,41], Emily White[2], Alicja Wolk[23,42], Michael O. Woods[43], Anna H. Wu[44], Stephen B. Gruber[45], Hermann Brenner[3,46,47], Ulrike Peters[2], Jenny Chang-Claude[1,48*]

[1] Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany; [2]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98124, USA; [3]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. [4]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892-9776, USA; [5]Centre Hospitalier Universitaire (CHU) Nantes, Service de Génétique Médicale, Nantes 44093, France; [6]Colorectal Oncogenomics Group, Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Victoria 3010, Australia; [7]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Parkville, Victoria 3010, Australia; [8]Division of Research, Kaiser Permanente Medical Care Program of Northern California, Oakland 94612, CA, USA; [9]Epidemiology Research Program, American Cancer Society, Atlanta, GA 30329-4251, USA;[10]Center for Public Health Genomics, University of Virginia, Charlottesville, VA 22908, USA; [11] Clinical and Translational Epidemiology Unit and Division of Gastroenterology, Massachusetts General Hospital, Boston, MA 02115, USA; [12]Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA.; [13]Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02115, USA, [14] Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892-9776, USA; [15] Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada; [16] Department of Epidemiology, Harvard

School of Public Health, Boston, MA 02115, USA; [17] Department of Nutrition, Harvard School of Public Health, Boston, MA 02115, USA; [18]Department of Medicine, Harvard Medical School, Boston, MA 02115, USA; [19] Cancer Epidemiology & Intelligence Division, Cancer Council Victoria, Melbourne, Victoria 3010, Australia.; [20]Department of Medicine, Division of Gastroenterology, University of Washington School of Medicine, Seattle, Washington 98195, USA., [21]Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA [22]Medical Department 1, University Hospital Dresden, TU Dresden, 01307 Dresden, Germany., [23]Institute of Environmental Medicine, Karolinska Institutet Solna, SE-171 77 Stockholm, Sweden, [24]Epidemiology Program, University of Hawaii Cancer Center, Honolulu 96822, HI, USA;[25]Department of Clinical Genetics, Karolinska University Hospital Solna, SE-171 77 Stockholm, Sweden. [26] Department of Molecular Medicine and Surgery, Karolinska Institutet Solna, SE-171 77 Stockholm, Sweden, [27]Catalan Institute of Oncology, Bellvitge Biomedical Research Institute (IDIBELL), 08028 Barcelona, Spain; [28]CIBER Epidemiología y Salud Pública (CIBERESP), 28029 Madrid, Spain; University of Barcelona, Barcelona 08007, Spain. [29] Ontario Institute for Cancer Research, Toronto ON M5G 0A3, Canada; [30]Department of Family Medicine and Community Health, Case Western Reserve University, Cleveland, Ohio 44106, USA., [31] Department of Medicine, Clinical Genetics Service, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA., [32]Centre for Cancer Genetic Epidemiology, Department of Public Health & Primary Care, University of Cambridge, Cambridge CB2 8AR, UK, [33]Department of Public Health Sciences, University of California, Davis, CA 95817, USA; [34]Bruce Rappaport Faculty of Medicine, Technion-Israel Institute of Technology, Haifa, Israel. [35]Clalit Health Services National Israeli Cancer Control Center, Haifa 34361, Israel.; [36]Department of Community Medicine and Epidemiology, Carmel Medical Center, Haifa 34361, Israel, [37]Department of General and Thoracic Surgery, University Hospital Schleswig-Holstein, Campus Kiel, 24118,Kiel, Germany; [38]Department of Medicine and Epidemiology, University of Pittsburgh Medical Center, Pittsburgh 15213, PA, USA; [39]Department of Internal Medicine, University of Utah Health Sciences Center, Salt Lake City, UT 84108, USA; [40]Huntsman Cancer Institute and Department of Population Health Sciences, University of Utah, Salt Lake City, Utah 84112, USA, [41]Genetic Medicine and Familial Cancer Centre, Royal Melbourne Hospital, Parkville, Victoria 3050, Australia; [42]Department of Surgical Sciences, Uppsala University, Uppsala SE-171 77, Sweden, [43]Faculty of Medicine, Memorial University of Newfoundland, St. John's, Newfoundland NL A1C 5S7, Canada, [44]Department of Preventive Medicine, University of Southern California Keck School of Medicine, Los Angeles 90033, California; [45]Department of Medicine, Keck School of Medicine, University of Southern California, Los Angeles 90033, CA, USA, [46]Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National

Center for Tumor Diseases (NCT), 69120 Heidelberg, Germany; [47]German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany; [48]Genetic Tumour Epidemiology Group, University Medical Center Hamburg-Eppendorf, University Cancer Center Hamburg, 20246 Hamburg, Germany.

***Corresponding author**

Correspondence:
Prof. Dr. Jenny Chang-Claude,
Division of Cancer Epidemiology,
German Cancer Research Center (DKFZ),
Im Neuenheimer Feld 581, Heidelberg, 69120, Germany.
Phone: +49 6221 42 2373;
E-mail: j.chang-claude@dkfz-heidelberg.de

Abbreviations:

BMI, body mass index; CCFR, Colon Cancer Family Registry; CI, confidence interval; CORECT, Colorectal Cancer Transdisciplinary Consortium; CRC, colorectal cancer; ERβ, estrogen-receptor β; GECCO, Genetics and Epidemiology of Colorectal Cancer Consortium; GRS, genetic risk scores; GWAS, genome-wide association study; LD, linkage disequilibrium; MHT, menopausal hormone therapy; MR, Mendelian randomization; OR, odds ratio, SNP, single nucleotide polymorphism.

Word count:

Abstract: 200

Text: 4370

Abstract

Background: Substantial evidence supports an association between use of menopausal hormone therapy and increased colorectal cancer (CRC) risk, indicating a role of exogenous sex hormones in CRC development. However, findings on endogenous estrogen exposure and CRC are inconsistent.

Methods: We used a Mendelian randomization approach to test for a causal effect of age at menarche and age at menopause as surrogates for endogenous estrogen exposure on CRC risk. Weighted genetic risk scores based on 358 single nucleotide polymorphisms associated with age at menarche and 51 single nucleotide polymorphisms associated with age at menopause were used to estimate the association with CRC risk using logistic regression in 12 944 women diagnosed with CRC and 10 741 women without CRC from three consortia. Sensitivity analyses were conducted to address pleiotropy and possible confounding by body mass index.

Results: Genetic risk scores for age at menarche (odds ratio per year 0.98, 95% confidence interval: 0.95-1.02) and age at menopause (odds ratio 0.98, 95% confidence interval: 0.94-1.01) were not significantly associated with CRC risk. The sensitivity analyses yielded similar results.

Conclusion: Our study does not support a causal relationship between genetic risk scores for age at menarche and age at menopause and CRC risk.

Keywords: BMI, egger regression, endogenous estrogen, hormonal factors, menstrual, polygenic risk score, reproductive factor, sex hormone, weighted median

BACKGROUND

Colorectal cancer (CRC) is the third most common cancer worldwide and incidence rates are higher in men than in women.[1] The sex-specific difference might be partly attributed to differential exposure to sex-hormones, especially estrogen.[2] This hypothesis is partially supported by epidemiologic studies as well as a recent meta-analysis out of four randomized controlled trials, eight cohort and eight case-control studies, which have shown that use of exogenous sex hormones in the form of combined estrogen-progestogen menopausal hormone therapy (MHT) is inversely associated with the risk of CRC.[3, 4]

Epidemiologic studies examining reproductive factors such as age at menarche and age at menopause with CRC risk have reported inconsistent results.[5-7] A meta-analysis summarizing evidence based on 22 studies published by Li et al. did not find a significant association between age at menarche and CRC risk.[8] In a large prospective cohort of the NIH American Association of Retired Persons (AARP) Diet and Health Study with more than 214,000 postmenopausal women, an inverse association between age at menarche and CRC risk was observed for women without a history of MHT use, whereas increasing age at menopause was associated with higher CRC risk.[9] It is possible that the inconsistency in results is due to recall bias or to improper adjustment for confounders, which are inherent limitations of observational studies. The Mendelian randomization (MR) approach[10] uses genetic variants as instrumental variables to test for the causal effect of an exposure risk factor on an outcome. Since the genetic variants in offspring are randomly distributed at conception independent of environmental factors given parental genotypes, confounding and reverse causation are less likely to occur in MR analyses. For genetic variants to function as valid instrumental variables in MR analyses, three assumptions have to be met: 1) the genetic variants have to be associated with the exposure risk factor; 2) the variants are not associated with any confounding variables of the exposure-outcome association; and 3) the variants are unrelated to the outcome except through the risk factor of interest.[11, 12] Using variants associated with age at menarche and age at menopause instead of self-reported measure of

the risk factors themselves, this approach can help to avoid issues of confounding, recall bias and reverse causation. Evidence from MHT use suggests that more exposure to estrogen reduces risk of CRC.[3, 4, 13] Estrogen exerts its effects in colon cells predominantly through the nuclear receptor estrogen-receptor β (ERβ),[14, 15] which has mainly anti-proliferative effects[16] and its expression is inversely related to cancer stage, tumor extent and mortality.[17, 18] Longer use of MHT appears to be associated with high ERβ expression in tumors.[19] If similar mechanisms hold then longer endogenous estrogen exposure (through earlier age at menarche and/or later age at menopause) would reduce CRC risk as well. To test this hypothesis, we conducted a MR analysis using summary data for single nucleotide polymorphisms (SNPs) known to be associated with age at menarche and age at menopause from prior studies as well as using individual level data of three consortia.

MATERIALS AND METHODS

Study population. Epidemiological and genetic data were derived from 26 studies participating in three large consortia of CRC, the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO)[20] (5 386 cases and 5 696 controls), the Colon Cancer Family Registry (CCFR)[21] (1 678 cases and 1 188 controls) and the Colorectal Cancer Transdisciplinary (CORECT) Consortium[22] (5 880 cases and 3 857 controls) (see Supplementary Tables 1 and 2). Different centers of CCFR participated as individual studies in GECCO and/or CORECT, and therefore were analyzed as such. Any participant overlap between the three consortia was excluded. In total, 12 944 female colorectal cancer cases and 10 741 female controls, both of European ancestry were included. All participants provided written, informed consent and studies were approved by their respective institutional review boards. Women with incident invasive colorectal adenocarcinoma (International Classification of Disease Code, 9th revision (ICD-9), codes 153-154) were included as cases. Data on demographic factors and lifestyle were collected using in-person interviews or self-completed questionnaires. Data harmonization was done centrally as previously described.[23]

Self-reported data for age at menarche and age at menopause have not been harmonized and therefore could not be used.

Genotype data and imputation. Genotype information was available for all included studies. Details on genotyping, quality assurance and imputation are included in the Supplementary Information. In short, SNPs were excluded based on call rate (<98% GECCO; <95% in CORECT), Hardy-Weinberg equilibrium in controls ($P<1x10^{-4}$) or low minor allele frequency (≤1%). Participants received a value of 0, 1, or 2 for carrying 0 (wild-type homozygous), 1 (heterozygous) or 2 (homozygous for the risk allele) alleles associated with higher age at menarche/age at menopause for each SNP. For imputed SNPs, participants were assigned continuous values between 0 and 2.

Calculation of genetic risk scores. Two recent genome-wide association studies conducted by the REPROGEN consortium (Reproductive Genetics Consortium) involving up to ~370 000 women for the study on age at menarche and up to 69 360 women for the study on age at menopause identified 389 genetic variants associated with age at menarche[24] and 54 SNPs associated with age at natural menopause[25] at a genome-wide significance level (i.e., $P<5x10^{-8}$). Of the reported 389 genetic variants for age at menarche, 12 variants on sex chromosomes (which were not available in our datasets) were excluded. For further 42 missing variants, we used proxy SNPs in strong linkage disequilibrium ($R^2>0.8$, median $R^2=0.985$, R^2 range: 0.83-1.00) with effect alleles harmonized to reflect increase in age at menarche. Seven variants were excluded because no proxy was found. We checked for correlations between individual SNPs and excluded 12 correlated SNPs ($R^2>0.01$) (always the SNP with higher *P*-value for association with age at menarche was dropped). So data were available for 358 SNPs for this analysis (mean imputation quality score=0.97) (Supplementary Table 5). For age at menopause, three correlated SNPs ($R^2>0.01$) were excluded. All remaining 51 SNPs for age at menopause were available in the datasets (mean imputation quality score=0.98) (Supplementary Table 6). The genotype data described were

used to construct genetic risk scores (GRS) as instrumental variables for age at menarche and age at menopause, respectively.

The GRS for the k^{th} women is calculated by the sum of the number of risk increasing alleles carried (G) (imputed allele doses) for each SNP weighted by the reported beta-coefficient (β) for association with age at menarche and age at menopause.[24, 25]

Age at menarche

$$\mathrm{GRS}_k = \sum_{n=1}^{358} \beta_n G_{kn}$$

Age at menopause

$$\mathrm{GRS}_k = \sum_{n=1}^{51} \beta_n G_{kn}$$

As risk scores themselves do not have meaningful units, we scaled the risk scores in terms of age in years. In this way resulting odds ratios (OR) can be interpreted as the relative change in CRC risk per year older age at menarche/age at menopause. Scaling was done by dividing the GRS by regression coefficients of a linear regression of GRS on self-reported age at menarche (beta= 0.57) or age at menopause (beta=0.98). These regression coefficients were obtained using the DACHS (Darmkrebs: Chancen der Verhütung durch Screening) study (853 control women) and the WHI (Womens' Health Initiative) study (1 492 control women) for which self-reported data on age at menarche were available and regression coefficients of the two studies were combined using meta-analysis.

An additional GRS as a surrogate for total lifetime exposure to endogenous estrogen was calculated as the sum of the scaled risk scores for age at menarche and age at menopause.

$$\mathrm{GRS}_{time\ period} = \mathrm{GRS}_{age\ at\ menarc\ he} + \mathrm{GRS}_{age\ at\ menopause}$$

For this analysis, risk scores for age at menarche were calculated as the sum of the number of alleles associated with decreasing age at menarche and the number of alleles associated with increasing age at menopause. So the GRS for time period of estrogen exposure reflects higher risk for longer exposure to endogenous estrogen. For this score, four SNPs (rs3136269, rs11031040, rs537244, rs4303811) were excluded due to high linkage disequilibrium (R^2>0.01) between age at menarche and age at menopause SNPs.

Statistical analysis.

Validating MR assumptions

The first assumption of MR regarding instrumental variable strength (i.e. association between the genetic variants and the exposure risk factor) was verified by calculating the F-statistic out of $F= (R^2(n-K-1))/((1-R^2)K))$, where R^2 refers to the variance explained by the instrumental variable,[24, 25] K indicates the number of instrumental variables and n stands for the sample size.[26] An F-statistic >10 suggests that the genetic instrument is sufficiently strong.[27] To evaluate the second assumption of MR (i.e. no association between genetic variants and potential confounders), we tested associations between GRS for age at menarche/age at menopause and the following risk factors for CRC: smoking status (ever/never), family history of CRC, education/educational level, ever aspirin/NSAID use (at least once per month for more than one year), body mass index (BMI) (continuous), menopausal hormone therapy (estrogen/progestin combined and estrogen alone), using linear regression for continuous variables and logistic/multinomial logistic regression for categorical variables in a subset of the studies with available data (n=6 285 controls). To address the third assumption of MR i.e. to assess the presence of pleiotropy, we applied the MR-Egger method.[28] MR-Egger relies on the InSIDE assumption (Instrument Strength Independent of Direct Effect), which is the assumption that the pleiotropic effects of the genetic variants are not correlated with the effects of genetic variants on the risk factors. MR-Egger uses an inverse-variance weighted estimator and by plotting the SNP`s effect on

the exposure against its effect on the outcome, the intercept term of MR-Egger provides a test for directional pleiotropy, i.e. the average effect of pleiotropy is non-zero, across all genetic variants used. If the average pleiotropic effect of all variants is zero and the InSide assumption is satisfied, pleiotropy is "balanced" and will not be detected. If the intercept differs from zero, it suggests horizontal pleiotropy, which means that some genetic variants affect the outcome through a pathway different from the exposure of interest. For visual inspection of pleiotropy, we used funnel plots of each SNPs ratio estimate against its precision (1/standard error of the ratio estimate).[29, 30] Any deviation from symmetry would suggest pleiotropy.

Estimation of Causal Effect

GRS based analyses

We examined the association between GRS and CRC risk using logistic regression models adjusted for study, age as well as principal components (PCs) of genetic ancestry (three PCs were used for GECCO and 10 PCs were used for CORECT), to account for potential population stratification. Summary results for GECCO and CORECT were derived using fixed-effects meta-analysis assuming that the included studies share a common effect size. As there are indications for differential associations of age at menarche with CRC risk according to menopausal hormone use,[9, 31] stratified analyses by menopausal hormone therapy were performed (only in GECCO where harmonized data were available). Additional stratified analyses were performed according to menopausal status (for age at menarche), combined menopausal estrogen or progesterone therapy, estrogen therapy alone, and BMI categories in kg/m² (BMI<18.5: underweight, 18.5–24.9: normal weight, 25-30: overweight, >30: obese) (only for age at menopause). Due to reported differences in risk between colon and rectal cancer associated with hormone use,[32, 33] we also conducted site-specific analyses for 4 037 female colon cancer cases and 1 184 rectal cancer cases in GECCO. Power calculations were conducted to estimate the magnitude of effects detectable with our study size assuming 5% alpha level and an R^2 of 0.069 for age at menarche and R^2 of 0.057 for age

281 at menopause, which corresponds to the variance in age at menarche/age at menopause
282 explained by the SNPs used for this analyses.[34]

283 It is known that BMI in childhood is strongly associated with age at menarche[35] and that
284 some SNPs associated with age at menarche could also have pleiotropic effects e.g. are
285 related to BMI as well.[24] To address this issue and thereby also account for violations of the
286 third MR assumption, we conducted further BMI-specific sensitivity analyses. We adjusted
287 for BMI in the logistic regression analysis using a weighted GRS for BMI comprising 77
288 SNPs previously reported to be associated with BMI at a genome-wide significance level in
289 European subjects.[36] For the second sensitivity analysis, we identified age at menarche SNPs
290 showing pleiotropy by testing the association of these SNPs with BMI in a subset of our
291 sample (n=5 832 cases/6 285 controls) and found 29 SNPs to be associated with BMI at
292 nominal significance (P value<0.05). Two further SNPs overlapped with reported BMI-
293 SNPs[36] and eleven more SNPs were in high linkage disequilibrium (R^2>0.1) with BMI SNPs.
294 A restricted GRS for age at menarche excluding the 42 BMI-associated SNPs (n=316 SNPs)
295 was constructed and then assessed for association with CRC risk. For the analysis of lifetime
296 estrogen-exposure we also generated a restricted risk score excluding the same 42 BMI-
297 associated SNPs.

298 Two-sample MR analyses

299 We performed two-sample MR analyses as sensitivity analyses using published summary
300 statistics for SNP-exposure associations (age at menarche[24]/age at menopause[25]), SNP
301 outcome associations were estimated in GECCO/CORECT (Supplementary Tables 8 and 9
302 for age at menarche/menopause-SNPs respectively). We applied the weighted median
303 estimator approach, which is robust against violations due to pleiotropic SNPs even when up
304 to 50% of the genetic instruments are invalid.[37] For this approach, we used SNP-exposure
305 and SNP-outcome associations to build ratio estimates for each SNP. These estimates were
306 ordered and weighted by the inverse of their variance. Bootstrapped standard errors were
307 calculated and used for construction of 95% confidence intervals (CI). Furthermore, we

assessed the slope of MR-Egger regression (see section on validation of MR assumptions) which yields a pleiotropy-adjusted estimate of the true causal effect.[28]

All analyses were conducted using R version 3.2.2 (R Foundation for Statistical Computing, Vienna, Austria). For supplementary figures, we used the R packages "Mendelian Randomization"[38] and "ggplot2".

RESULTS

The assessment of the MR assumptions indicated that our instrumental variables for age at menarche and for age at menopause were both strong instruments (F-statistic for age at menarche=1 755, R^2=0.069[24]; F-statistic for age at menopause=1 431, R^2=0.057[25]. Secondly, we did not find significant associations between the GRS for age at menarche and CRC risk factors, including smoking, family history of cancer, education, aspirin/NSAID use, estrogen/progestin therapy, estrogen alone therapy, with the exception of BMI, which showed a significant association (Supplementary Table 3). Similarly, there was no association of the GRS for age at menopause with any of the tested risk factors (Supplementary Table 3). Table 1 shows the results of the MR analyses for age at menarche, age at menopause and lifetime estrogen exposure with CRC risk. Yearly increment in GRS for age at menarche was associated with CRC risk with an OR of 0.98 (95% CI: 0.95-1.02). Sensitivity analyses adjusting for BMI using a GRS yielded similar results for age at menarche (OR 0.99 per year, 95% CI: 0.95-1.02). Further sensitivity analysis using restricted risk scores for age at menarche (by excluding 42 BMI-associated SNPs) showed similar effect sizes for the association between age at menarche and CRC risk (OR 0.99 per year, 95% CI: 0.95-1.03). We also did not find evidence to support a causal association with risk of CRC for age at menopause (OR 0.98 per year, 95% CI: 0.94-1.01) or for lifetime estrogen exposure (OR 0.99 per year, 95% CI: 0.97-1.02) using GRS-based analyses.

Results of the weighted median estimator approach also did not indicate causal association between age at menarche (OR per year 1.00, 95% CI: 0.90-1.11), age at menopause (OR per

year 1.00, 95% CI: 0.95-1.05) or lifetime estrogen-exposure (OR per year 1.00, 95% CI: 0.95-1.05) with CRC risk.

The pleiotropy adjusted MR estimate (OR) derived from the slope of Egger regression was 0.99 (95% CI: 0.83-1.17) for age at menarche, 1.02 (95% CI: 0.94-1.10) for age at menopause and 0.97 (95% CI: 0.91-1.01) for lifetime estrogen-exposure (details of results per GECCO and CORECT consortium in Supplementary Table 7). The intercept term from MR-Egger regression was centered at the origin for age at menarche (intercept term -0.0008, 95% CI: -0.007-0.005, *P* value 0.80) and age at menopause (intercept term -0.009, 95% CI: -0.023-0.006, *P* value 0.23) suggesting absence of strong directional pleiotropy (Supplementary Figure 1 and 2). Also the funnel plots for age at menarche and age at menopause appear to be generally symmetrical and therefore do not suggest presence of pleiotropy (Supplementary Figure 3 and 4). Table 2 shows stratified analyses for the association of GRS for age at menarche and CRC risk by menopausal status, combined estrogen/progesterone therapy and estrogen alone therapy as well as cancer site based on GECCO data. None of these factors modified substantially the association between age at menarche and CRC risk. The site specific analysis showed no evidence for a difference in association between colon cancer (OR 0.97 per year, 95% CI: 0.91-1.02) and rectal cancer (OR 1.04 per year, 95% CI: 0.95-1.14). The analyses for age at menopause stratified by combined estrogen/progesterone therapy, estrogen alone therapy, BMI and cancer site did not yield evidence for effect heterogeneity (Table 3). Power calculation shows that our study had > 80 % power to detect an OR of 0.85 per standard deviation change in exposure variable (1.5 years for age at menarche, 4.8 years for age at menopause) but only around 50% power for an OR of 0.90 (Supplementary Table 4).

DISCUSSION

In this large MR study we aimed to clarify the inconsistent findings from observational studies regarding the association of age at menarche and age at menopause with CRC risk and thereby the role of endogenous estrogen exposure. We investigated the association of

GRS for age at menarche and age at menopause as surrogates for endogenous estrogen exposure on CRC risk. Our results do not support an association between GRS for age at menarche and age at menopause and CRC risk, which under MR assumptions can be interpreted as absence of a causal effect.

In line with the findings of our study, several prospective studies and one meta-analysis reported no association[6, 8, 39] between self-reported age at menarche and CRC risk, although some studies found an inverse association.[5, 9] Two more recent prospective studies reported inverse associations with CRC risk for age at menarche only among never users of any MHT. In never users of hormone therapy, Zervoudakis et al. reported a hazard ratio of 0.73 (95% CI: 0.57-0.94) for age at menarche (>15 vs. 11-12 years) in association with risk of CRC[9] and Murphy et al. found a hazard ratio of 0.72 (95% CI: 0.54-0.96) for age at menarche (>15 vs. 11-12 years).[31] We therefore assessed the association of GRS for age at menarche with CRC risk stratified by ever/never use of MHT, separately for combined estrogen-progesterone therapy and for estrogen monotherapy. No difference in the association according to either combined estrogen-progesterone therapy or estrogen-alone therapy was found.

Higher BMI in childhood is associated with earlier age at menarche[24, 35] and also with a higher risk for CRC.[40] Therefore we conducted BMI-specific sensitivity analyses to account for violations of the MR assumptions by confounding and pleiotropy. Due to a strong association between childhood/adolescent BMI and adult BMI[41] and also a high concordance between adolescent and adult BMI-SNPs,[42] we accounted for adult BMI in sensitivity analyses. The effect sizes observed in the sensitivity analysis (by excluding BMI-associated SNPs and by adjustment using a GRS for BMI) were slightly smaller compared to results of the main analysis, which suggests that some of the effect was confounded by BMI. Our restricted risk-score might not be totally BMI-unrelated, considering that Day et al.[24] found age at menarche variants that appear unrelated to BMI at a nominal level in their sample to be still BMI associated collectively ($P=4.2x10^{-9}$). Because of this strong interrelationship between age at menarche and BMI it is difficult to separate the SNPs into BMI-related and

BMI-unrelated variants. So our adjustment of the analysis by GRS-BMI might more effectively control for BMI. Thus, any observed inverse relationship between age at menarche and CRC risk in previous observational studies could have been due to inadequate control of confounding by higher BMI in childhood.

For age at menopause, results of observational studies on the association with CRC risk have also been inconclusive. The large NIH-AARP study observed a statistically significant elevated risk for higher age at menopause in postmenopausal women (≥55 vs < 40; HR 1.50, 95% CI: 1.23, 1.83),[9] whereas most other studies reported null associations.[6, 31, 39] Corresponding to the lack of association with age at menarche and age at menopause, the GRS for the reproductive period as indicator for the lifetime estrogen exposure was also not significantly associated to CRC risk. Additional adjustment of that analysis by education, family history of CRC, ever regular aspirin use, MHT usage, BMI and smoking did not substantially change the results (Supplementary Table 10). This is compatible with the observation of no association between the reproductive period (≥ 36 years vs. ≤ 30 years) and CRC risk in a prospective observational study conducted in Japan[43]. The GRS for lifetime for endogenous estrogen exposure, which we constructed, does not account for other factors like parity or breast feeding, which influence overall estrogen exposure, however, these factors have not been associated with CRC risk.

Results of prospective studies that investigated the association between serum levels of endogenous estrogens and CRC risk have also been inconsistent. One study reported a positive association between circulating estradiol and CRC risk,[44] another study observed an inverse relationship[45] while most other studies found no associations.[46-48] There are reports that earlier age at menarche is associated with higher estrogen levels.[49, 50] So estrogen levels could also be a possible link between age at menarche and CRC risk. Due to the inconsistent results of these reports, further studies are needed to clarify these associations.

On the other hand, observational studies reported that exogenous estrogen exposure by MHT mainly in the form of combined estrogen-progestogen was associated with a reduced risk for CRC.[13, 51] The Womens`s Health Initiative Clinical Trial (WHI-CT) reported no effect of

417 estrogen-alone therapy,[52, 53] and a significant risk reduction for the association of estrogen
418 plus progestin vs. placebo and CRC risk,[33] which was suggested to have resulted from
419 diagnostic delay instead of true risk reduction[54]. However, a recent meta-analysis which
420 summarized results of four clinical trials including WHI-CT and 16 observational studies
421 concluded that there is consistent evidence to support a protective effect of MHT on CRC
422 risk.[4]

423 Thus, it appears that exogenous and endogenous estrogens, which also vary in absolute
424 amount of estrogen, may play different roles in the development of CRC, presumably by
425 different mechanisms, which are not well understood. Estrogen acts in colon cells
426 predominantly through ERβ,[14, 15] which exerts proapoptotic and anti-proliferative effects in
427 the colon[16] and its expression is reduced in tumor tissue.[17, 18] According to an in vivo study,
428 estrogen treatment was associated with an increase in expression of ERβ in colon tissue,[55]
429 supporting a mechanism by which MHT may affect CRC risk. There is also some evidence
430 that the protective effect of MHT on CRC risk may vary by the expression status of ERβ.
431 Two studies found that the magnitude of risk reduction by MHT was different between
432 colorectal tumors with higher and with lower expression of ERβ.[56, 57] It is possible that the
433 effect of endogenous estrogens on CRC risk may be modulated by ERβ expression as well.
434 Therefore, larger studies with data on expression of ERβ in colon tissue are warranted to
435 assess whether the association of age at menarche/age at menopause and CRC risk differs by
436 ERβ expression status.

437 In this MR study, we aimed to use proxies for start and endpoint of endogenous estrogen
438 exposure in women, specifically age at menarche/age at menopause, which themselves are
439 complex traits influenced by many variants with only small effects on the trait. Although we
440 did not see large pleiotropic effects using Egger regression, there might have been residual
441 pleiotropy, which is difficult to exclude. Previous studies of age at menarche performed LD
442 score regression using 123 SNPs associated with age at menarche and found, amongst others,
443 genetic correlations with BMI, adult height or Type 2 diabetes. Residual pleiotropy related to

444 adult height or Type 2 diabetes, which have been reported to be associated with higher risk
445 for CRC as well, cannot be fully excluded.[58] In addition, Day et al. reported genetic
446 correlations for the 54 age at menopause SNPs with adult obesity and other growth-related
447 traits. The top menopause-SNPs were also associated with fasting glucose and were enriched
448 in DNA repair pathways, yielding further sources of residual pleiotropy.[25] Residual
449 pleiotropy is a general limitation of MR, especially when exploring complex traits. When
450 considering the recently published hypothesis of an omnigenic model of complex traits,
451 coined „network pleiotropy“, essentially any regulatory variant in a trait-relevant cell-type
452 can have some effect on the trait.[59] This is because specific cell-types have specific
453 regulatory networks, where any single variant could affect trait relevant genes, „core genes“,
454 mediated through the same regulatory networks. So also for GWAS findings, it is highly
455 likely that some genetic variants exhibit horizontal pleiotropy.[60] As the selected age at
456 menarche/age at menopause SNPs might also contribute tiny effects on further traits through
457 network-pleiotropy, we cannot fully rule out pleiotropy. These limitations should be kept in
458 mind and methods to explore the impact of such effects should be developed. That said, our
459 sensitivity analyses especially Egger regression, did not indicate large pleiotropic effects.
460 The second assumption of MR is that the IV is not associated with confounding factors of
461 the observational association between age at menarche/menopause and CRC risk. We were
462 able to exclude most risk factors for CRC (smoking, family history of CRC, education,
463 aspirin use, MHT) as confounding variables except for BMI, which was accounted for using
464 several sensitivity analyses. In addition, substantial overlap between datasets used for
465 estimating SNP-exposure and SNP-outcome associations would bias results in the direction
466 of the observational estimate. There was some overlap between the studies (see
467 Supplementary Information on participant overlap) but unlikely to have substantially
468 influenced the results.
469 Strengths of our study include the large sample size, the availability of centrally harmonized
470 data and the robustness of the instrumental variables. Power calculations showed that our
471 study has limited power to detect weak effects. Therefore, we cannot exclude a weak

association of CRC with age at menarche or age at menopause. In summary, in our large MR study, evidence is limited for causal associations between age at menarche/age at menopause and CRC risk.

Additional Information

Ethics approval and consent to participate

All participants provided written, informed consent and studies were approved by their respective institutional review boards.

Availability of data and materials

Genotyping data of the GECCO studies will soon be available at the database of Genotypes and Phenotypes (dbGaP) for download at the accession number: phs001078.v1.p1.

Conflict of interest:

The authors declare no conflict of interest.

Funding

GECCO: National Cancer Institute, National Institutes of Health, U.S. Department of Health and Human Services (U01 CA137088; R01 CA059045)

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).

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

CCFR: This work was supported by grant UM1 CA167551 from the National Cancer Institute and through cooperative agreements with the following CCFR centers:

Australasian Colorectal Cancer Family Registry (U01 CA074778 and U01/U24 CA097735)

Mayo Clinic Cooperative Family Registry for Colon Cancer Studies (U01/U24 CA074800)

Ontario Familial Colorectal Cancer Registry (U01/U24 CA074783)

Seattle Colorectal Cancer Family Registry (U01/U24 CA074794)

University of Hawaii Colorectal Cancer Family Registry (U01/U24 CA074806)

USC Consortium Colorectal Cancer Family Registry U01/U24 CA074799)

The Colon CFR GWAS was supported by funding from the National Cancer Institute, National Institutes of Health (U01 CA122839 and R01 CA143237 to Graham Casey). The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute 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 or the CCFR.

ColoCare: This work was supported by the National Institutes of Health (grant numbers R01 CA189184, U01 CA206110, 2P30CA015704-40 (Gilliland), U01CA152756), the Matthias

Lackas-Foundation, the German Consortium for Translational Cancer Research, and the EU TRANSCAN initiative.

CORECT: National Cancer Institute, National Institutes of Health under RFA # CA-09-002 (U19 CA148107). The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in CORECT, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government or CORECT. This work was also supported by the National Institutes of Health (P30 CA014089 to S. B. Gruber).

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.

DACHS: German Research Council (Deutsche Forschungsgemeinschaft, BR 1704/6-1, BR 1704/6-3, BR 1704/6-4 and CH 117/1-1), and the German Federal Ministry of Education and Research (01KH0404 and 01ER0814).

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

ESTHER / VERDI was supported by grants from Baden-Württemberg State Ministry of Science, Research and Education and German Cancer Aid.

NHS is supported by the National Institutes of Health (R01 CA137178, P01 CA087969, UM1 CA186107, R01 CA151993, R35 CA197735, K07 CA190673, and P50 CA127003).

MEC: National Institutes of Health (R37 CA54281, P01 CA033619, and R01 CA063464).

MECC: The Molecular Epidemiology of Colorectal Cancer study was funded by the National Institutes of Health (R01 CA81488 and R01 CA197350 to S.B. Gruber).

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

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

OFCCR: National Institutes of Health, through funding allocated to the Ontario Registry for Studies of Familial Colorectal Cancer (U01 CA074783); 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.

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. Additionally, a subset of control samples were genotyped as part of the Cancer Genetic Markers of Susceptibility (CGEMS) Prostate Cancer GWAS (Yeager, M et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nat Genet 2007 May;39(5):645-9), CGEMS pancreatic cancer scan (PanScan) (Amundadottir, L et al. Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. Nat Genet. 2009 Sep;41(9):986-90, and Petersen, GM et al. A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. Nat Genet. 2010 Mar;42(3):224-8), and the Lung Cancer and Smoking study (Landi MT, et al. A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. Am J Hum Genet. 2009 Nov;85(5):679-91). The prostate and PanScan study datasets were accessed with appropriate approval through the dbGaP online resource (http://cgems.cancer.gov/data/) accession numbers phs000207.v1.p1 and phs000206.v3.p2, respectively, and the lung datasets were accessed from the dbGaP website (http://www.ncbi.nlm.nih.gov/gap) through accession number phs000093.v2.p2. Funding for the Lung Cancer and Smoking study 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. For the lung study, the GENEVA Coordinating Center provided assistance with genotype cleaning and general study coordination, and the Johns Hopkins University Center for Inherited Disease Research conducted genotyping.

PMH: National Institutes of Health (R01 CA076366 to P.A. Newcomb).

SEARCH: Cancer Research UK (C490/A16561).

The 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), Catalan Government DURSI (grant 2014SGR647), and Junta de Castilla y León (grant LE22A10-2).

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).

SMC: The Swedish Mammography Cohort was funded by the Swedish Cancer Society and the Swedish Research Council/Council for Research Infrastructures.

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, HHSN268201600018C, HHSN268201600001C, HHSN268201600002C, HHSN268201600003C, and HHSN268201600004C."

Authors' contributions

SN conducted the literature review, performed parts of the data analysis, interpreted the data and drafted the manuscript. BLB analyzed the data, KB and SAB contributed to writing the manuscript and provided detailed information on the participating studies. JYD contributed

to analytical methods of the study. VA, SIB, SB, DDB, BJC, PTC,GC, ATC, JCC, SJC, SG, ELG, GGG, WMG, JH, MH, JLH, LH, MAJ, AJ, SCL, LLM, AL, VM, ML, LL, YL, KO, PAN, PDP, JDP, LQ, GR, CS, RES, MLS, MS, CMU, AKW, EW, AW, MOW, AHW, SB G, HB designed and implemented the different studies included in this analysis. UP coordinated the cooperation of the studies and data harmonization. JCC conceived the study, provided input to interpreting the data and writing the manuscript. All authors discussed the results, provided critical feedback and approved the final manuscript.

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.

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.

CCFR: We graciously thank the generous contributions of our study participants, the dedication of study staff, and the financial support from the U.S. National Cancer Institute, for without each of these this important registry would not exist.

DACHS: We thank all participants and cooperating clinicians, and Ute Handte-Daub, Utz Benscheid, Muhabbet Celik and Ursula Eilber for excellent technical assistance.

GECCO: The authors would like to thank all those at the GECCO Coordinating Center for helping bring together the data and people that made this project possible. The authors also acknowledge Deanna Stelling, Mark Thornquist, Greg Warnick, Carolyn Hutter, and team members at COMPASS (Comprehensive Center for the Advancement of Scientific Strategies) at the Fred Hutchinson Cancer Research Center for their work harmonizing the GECCO epidemiological data set. The authors acknowledge Dave Duggan and team members at TGEN (Translational Genomics Research Institute), the Broad Institute, and the Génome Québec Innovation Center for genotyping DNA samples of cases and controls, and for scientific input for GECCO.

NHS: We would like to acknowledge Patrice Soule and Hardeep Ranu of the Dana Farber Harvard Cancer Center High-Throughput Polymorphism Core who assisted in the genotyping for NHS under the supervision of Dr. Immaculata Devivo and Dr. David Hunter, Qin (Carolyn) Guo and Lixue Zhu who assisted in programming for NHS. We would like to thank the participants and staff of the Nurses' Health Study, for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. The authors assume full responsibility for analyses and interpretation of these data.

MECC: The authors thank the women who participated in the Molecular Epidemiology of Colorectal Cancer Study, as well as the study coordinators, interviewers, and study staff,

especially Hedy Rennert, Mila Pinchex, Flavio Lejbkowicz, Joseph D. Bonner, Joel K. Greenson, Shu-Chen Huang, Marilena Melas, Chenxu Qu, and Stephanie Schmit,

PLCO : The authors thank Drs. Christine Berg and Philip Prorok, Division of Cancer Prevention, National Cancer Institute, the Screening Center investigators and staff or the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial, Mr. Tom Riley and staff, Information Management Services, Inc., Ms. Barbara O'Brien and staff, Westat, Inc., and Drs. Bill Kopp and staff, SAIC-Frederick. Most importantly, we acknowledge the study participants for their contributions to making this study possible. The statements contained herein are solely those of the authors and do not represent or imply concurrence or endorsement by NCI.

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

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

Supplementary information is available at the British Journal of Cancer's website.

References:

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International journal of cancer Journal international du cancer.* 2015; **136**: E359-386.
2. Koo JH, Leong RW. Sex differences in epidemiological, clinical and pathological characteristics of colorectal cancer. *J Gastroenterol Hepatol.* 2010; **25**: 33-42.
3. Hoffmeister M, Raum E, Krtschil A, Chang-Claude J, Brenner H. No evidence for variation in colorectal cancer risk associated with different types of postmenopausal hormone therapy. *Clinical pharmacology and therapeutics.* 2009; **86**: 416-424.
4. Lin KJ, Cheung WY, Lai JY, Giovannucci EL. The effect of estrogen vs. combined estrogen-progestogen therapy on the risk of colorectal cancer. *International journal of cancer Journal international du cancer.* 2012; **130**: 419-430.
5. Martinez ME, Grodstein F, Giovannucci E, Colditz GA, Speizer FE, Hennekens C et al. A prospective study of reproductive factors, oral contraceptive use, and risk of colorectal cancer. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology.* 1997; **6**: 1-5.
6. Tsilidis KK, Allen NE, Key TJ, Bakken K, Lund E, Berrino F et al. Oral contraceptives, reproductive history and risk of colorectal cancer in the European Prospective Investigation into Cancer and Nutrition. *British journal of cancer.* 2010; **103**: 1755-1759.
7. La Vecchia C, Franceschi S. Reproductive factors and colorectal cancer. *Cancer causes & control : CCC.* 1991; **2**: 193-200.
8. Li CY, Song B, Wang YY, Meng H, Guo SB, Liu LN et al. Age at menarche and risk of colorectal cancer: a meta-analysis. *PLoS One.* 2013; **8**: e65645.
9. Zervoudakis A, Strickler HD, Park Y, Xue X, Hollenbeck A, Schatzkin A et al. Reproductive history and risk of colorectal cancer in postmenopausal women. *Journal of the National Cancer Institute.* 2011; **103**: 826-834.
10. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *International journal of epidemiology.* 2003; **32**: 1-22.
11. Burgess S, Bowden J, Fall T, Ingelsson E, Thompson SG. Sensitivity Analyses for Robust Causal Inference from Mendelian Randomization Analyses with Multiple Genetic Variants. *Epidemiology (Cambridge, Mass).* 2017; **28**: 30-42.
12. Evans DM, Davey Smith G. Mendelian Randomization: New Applications in the Coming Age of Hypothesis-Free Causality. *Annual review of genomics and human genetics.* 2015; **16**: 327-350.
13. Grodstein F, Newcomb PA, Stampfer MJ. Postmenopausal hormone therapy and the risk of colorectal cancer: a review and meta-analysis. *The American journal of medicine.* 1999; **106**: 574-582.
14. Elbanna HG, Ebrahim MA, Abbas AM, Zalata K, Hashim MA. Potential value of estrogen receptor beta expression in colorectal carcinoma: interaction with apoptotic index. *Journal of gastrointestinal cancer.* 2012; **43**: 56-62.
15. Kennelly R, Kavanagh DO, Hogan AM, Winter DC. Oestrogen and the colon: potential mechanisms for cancer prevention. *The Lancet Oncology.* 2008; **9**: 385-391.
16. Caiazza F, Ryan EJ, Doherty G, Winter DC, Sheahan K. Estrogen receptors and their implications in colorectal carcinogenesis. *Frontiers in oncology.* 2015; **5**: 19.
17. Konstantinopoulos PA, Kominea A, Vandoros G, Sykiotis GP, Andricopoulos P, Varakis I et al. Oestrogen receptor beta (ERbeta) is abundantly expressed in normal colonic mucosa, but declines in colon adenocarcinoma paralleling the tumour's dedifferentiation. *European journal of cancer (Oxford, England : 1990).* 2003; **39**: 1251-1258.

18. Niv Y. Estrogen receptor beta expression and colorectal cancer: a systematic review and meta-analysis. *European journal of gastroenterology & hepatology.* 2015; **27**: 1438-1442.
19. Topi G, Ehrnstrom R, Jirstrom K, Palmquist I, Lydrup ML, Sjolander A. Association of the oestrogen receptor beta with hormone status and prognosis in a cohort of female patients with colorectal cancer. *European journal of cancer (Oxford, England : 1990).* 2017; **83**: 279-289.
20. Peters U, Jiao S, Schumacher FR, Hutter CM, Aragaki AK, Baron JA et al. Identification of Genetic Susceptibility Loci for Colorectal Tumors in a Genome-Wide Meta-analysis. *Gastroenterology.* 2013; **144**: 799-807 e724.
21. Newcomb PA, Baron J, Cotterchio M, Gallinger S, Grove J, Haile R et al. Colon Cancer Family Registry: an international resource for studies of the genetic epidemiology of colon cancer. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology.* 2007; **16**: 2331-2343.
22. Schumacher FR, Schmit SL, Jiao S, Edlund CK, Wang H, Zhang B et al. Genome-wide association study of colorectal cancer identifies six new susceptibility loci. *Nature communications.* 2015; **6**: 7138.
23. 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**: 2036-2044.
24. Day FR, Thompson DJ, Helgason H, Chasman DI, Finucane H, Sulem P et al. Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat Genet.* 2017.
25. Day FR, Ruth KS, Thompson DJ, Lunetta KL, Pervjakova N, Chasman DI et al. Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. *Nat Genet.* 2015; **47**: 1294-1303.
26. Pierce BL, Ahsan H, Vanderweele TJ. Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. *International journal of epidemiology.* 2011; **40**: 740-752.
27. Stock J, Staiger D. Instrumental Variables Regression with Weak Instruments. *Econometrica.* 1997; **65**: 557-586.
28. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *International journal of epidemiology.* 2015; **44**: 512-525.
29. Bowden J, Del Greco MF, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic. *International journal of epidemiology.* 2016; **45**: 1961-1974.
30. Haycock PC, Burgess S, Wade KH, Bowden J, Relton C, Davey Smith G. Best (but oft-forgotten) practices: the design, analysis, and interpretation of Mendelian randomization studies. *The American journal of clinical nutrition.* 2016; **103**: 965-978.
31. Murphy N, Xu L, Zervoudakis A, Xue X, Kabat G, Rohan TE et al. Reproductive and menstrual factors and colorectal cancer incidence in the Women's Health Initiative Observational Study. *British journal of cancer.* 2016.
32. Morch LS, Lidegaard O, Keiding N, Lokkegaard E, Kjaer SK. The influence of hormone therapies on colon and rectal cancer. *European journal of epidemiology.* 2016; **31**: 481-489.
33. Chlebowski RT, Wactawski-Wende J, Ritenbaugh C, Hubbell FA, Ascensao J, Rodabough RJ et al. Estrogen plus progestin and colorectal cancer in postmenopausal women. *The New England journal of medicine.* 2004; **350**: 991-1004.
34. Burgess S. Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. *International journal of epidemiology.* 2014; **43**: 922-929.

35. Davison KK, Susman EJ, Birch LL. Percent body fat at age 5 predicts earlier pubertal development among girls at age 9. *Pediatrics.* 2003; **111**: 815-821.
36. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature.* 2015; **518**: 197-206.
37. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genetic epidemiology.* 2016; **40**: 304-314.
38. Yavorska OO, Burgess S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *International journal of epidemiology.* 2017.
39. Buron Pust A, Alison R, Blanks R, Pirie K, Gaitskell K, Barnes I et al. Heterogeneity of colorectal cancer risk by tumour characteristics: Large prospective study of UK women. *International journal of cancer Journal international du cancer.* 2017; **140**: 1082-1090.
40. Zhang X, Wu K, Giovannucci EL, Ma J, Colditz GA, Fuchs CS et al. Early life body fatness and risk of colorectal cancer in u.s. Women and men-results from two large cohort studies. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology.* 2015; **24**: 690-697.
41. Simmonds M, Llewellyn A, Owen CG, Woolacott N. Predicting adult obesity from childhood obesity: a systematic review and meta-analysis. *Obesity reviews : an official journal of the International Association for the Study of Obesity.* 2016; **17**: 95-107.
42. Graff M, Ngwa JS, Workalemahu T, Homuth G, Schipf S, Teumer A et al. Genome-wide analysis of BMI in adolescents and young adults reveals additional insight into the effects of genetic loci over the life course. *Human molecular genetics.* 2013; **22**: 3597-3607.
43. Akhter M, Inoue M, Kurahashi N, Iwasaki M, Sasazuki S, Tsugane S. Reproductive factors, exogenous female hormone use and colorectal cancer risk: the Japan Public Health Center-based Prospective Study. *European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation (ECP).* 2008; **17**: 515-524.
44. Gunter MJ, Hoover DR, Yu H, Wassertheil-Smoller S, Rohan TE, Manson JE et al. Insulin, insulin-like growth factor-I, endogenous estradiol, and risk of colorectal cancer in postmenopausal women. *Cancer research.* 2008; **68**: 329-337.
45. Murphy N, Strickler HD, Stanczyk FZ, Xue X, Wassertheil-Smoller S, Rohan TE et al. A Prospective Evaluation of Endogenous Sex Hormone Levels and Colorectal Cancer Risk in Postmenopausal Women. *Journal of the National Cancer Institute.* 2015; **107**.
46. Clendenen TV, Koenig KL, Shore RE, Levitz M, Arslan AA, Zeleniuch-Jacquotte A. Postmenopausal levels of endogenous sex hormones and risk of colorectal cancer. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology.* 2009; **18**: 275-281.
47. Lin JH, Zhang SM, Rexrode KM, Manson JE, Chan AT, Wu K et al. Association between sex hormones and colorectal cancer risk in men and women. *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association.* 2013; **11**: 419-424 e411.
48. Falk RT, Dallal CM, Lacey JV, Jr., Bauer DC, Buist DS, Cauley JA et al. Estrogen Metabolites Are Not Associated with Colorectal Cancer Risk in Postmenopausal Women. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology.* 2015; **24**: 1419-1422.
49. Apter D, Reinila M, Vihko R. Some endocrine characteristics of early menarche, a risk factor for breast cancer, are preserved into adulthood. *International journal of cancer Journal international du cancer.* 1989; **44**: 783-787.
50. Emaus A, Espetvedt S, Veierod MB, Ballard-Barbash R, Furberg AS, Ellison PT et al. 17-beta-estradiol in relation to age at menarche and adult obesity in premenopausal women. *Human reproduction (Oxford, England).* 2008; **23**: 919-927.

51. Limsui D, Vierkant RA, Tillmans LS, Wang AH, Weisenberger DJ, Laird PW et al. Postmenopausal hormone therapy and colorectal cancer risk by molecularly defined subtypes among older women. *Gut.* 2012; **61**: 1299-1305.
52. Anderson GL, Limacher M, Assaf AR, Bassford T, Beresford SA, Black H et al. Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: the Women's Health Initiative randomized controlled trial. *Jama.* 2004; **291**: 1701-1712.
53. Lavasani S, Chlebowski RT, Prentice RL, Kato I, Wactawski-Wende J, Johnson KC et al. Estrogen and colorectal cancer incidence and mortality. *Cancer.* 2015; **121**: 3261-3271.
54. Simon MS, Chlebowski RT, Wactawski-Wende J, Johnson KC, Muskovitz A, Kato I et al. Estrogen plus progestin and colorectal cancer incidence and mortality. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* 2012; **30**: 3983-3990.
55. Weyant MJ, Carothers AM, Mahmoud NN, Bradlow HL, Remotti H, Bilinski RT et al. Reciprocal expression of ERalpha and ERbeta is associated with estrogen-mediated modulation of intestinal tumorigenesis. *Cancer research.* 2001; **61**: 2547-2551.
56. Tillmans LS, Vierkant RA, Wang AH, Samadder NJ, Lynch CF, Anderson KE et al. Associations between Environmental Exposures and Incident Colorectal Cancer by ESR2 Protein Expression Level in a Population-Based Cohort of Older Women. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology.* 2015; **24**: 713-719.
57. Rudolph A, Toth C, Hoffmeister M, Roth W, Herpel E, Schirmacher P et al. Colorectal cancer risk associated with hormone use varies by expression of estrogen receptor-beta. *Cancer research.* 2013; **73**: 3306-3315.
58. Day FR, Bulik-Sullivan B, Hinds DA, Finucane HK, Murabito JM, Tung JY et al. Shared genetic aetiology of puberty timing between sexes and with health-related outcomes. *Nature communications.* 2015; **6**: 8842.
59. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell.* 2017; **169**: 1177-1186.
60. Hemani G, Bowden J, Haycock PC, Zheng J, Davis O, Flach P et al. Automating Mendelian randomization through machine learning to construct a putative causal map of the human phenome. Preprint at: https://www.biorxiv.org/content/biorxiv/early/2017/08/23/173682.full.pdf. (2017).

Table 1: Association between age at menarche/age at menopause and CRC risk using MR analyses and sensitivity analyses

		GRS-based analyses						2-sample MR[f]			
		MR estimate[a,b]		Adjusted by GRS-BMI[a,c]		Restricted risk score[a,d]		MR-Egger[b,e]		Weighted Median Estimator[b]	
Variable (per year)	**Cases/Controls**	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**
Age at menarche	12 944/10 741	0.98	0.95-1.02	0.99	0.95-1.02	0.99	0.95-1.03	0.99	0.83-1.17	1.00	0.90-1.11
Age at menopause	12 944/10 741	0.98	0.94-1.01	NA	NA	NA	NA	1.02	0.94-1.10	1.00	0.95-1.05
Time period of hormone exposure	12 944/10 741	0.99	0.97-1.02	0.99	0.97-1.02	0.99	0.97-1.01	0.97	0.93-1.01	1.00	0.95-1.05

Abbreviations: BMI, body mass index; CCFR, colon cancer family registry; CI, confidence interval; CORECT, Colorectal Transdisciplinary study; GECCO, Genetics and Epidemiology of Colorectal Cancer Consortium; GRS, genetic risk score; MR, Mendelian randomization; NA, not applicable; OR, odds ratio per year, SNPs, single nucleotide polymorphisms; se, standard error.

[a] Logistic regression model adjusted for age, study and principal components of genetic ancestry;

[b] Meta-analyzed estimate of GECCO/CORECT datasets (CCFR centers participated in GECCO or CORECT and were analyzed as such);

[c] additionally adjusted for a GRS for BMI out of 77 reported SNPs for BMI [33];

[d] 42 BMI-associated SNPs were excluded from the age at menarche and time period of estrogen -exposure – risk scores;

[e] estimate derived from the slope of MR-Egger;

[f] estimates derived using summary statistics; se for calculation of CI obtained via bootstrapping.

Table 2. Association of genetically predicted age at menarche with CRC risk according to subgroups, CCFR, GECCO Consortium					
Subgroup	**N (cases/controls)**	**OR[a]**	**95% CI**	***P* value**	**Heterogeneity *P* value[b]**
All	5832/6285	0.97	0.92-1.02	0.20	
Menopausal status					
Premenopausal	545/621	0.95	0.80-1.12	0.54	0.94
Postmenopausal	5263/5647	0.97	0.92-1.02	0.23	
Menopausal hormone therapy combined					
No	3266/3490	0.96	0.90-1.03	0.24	0.69
yes	593/807	0.98	0.85-1.14	0.82	
Estrogen monotherapy					
No	3120/3214	0.94	0.88-1.01	0.11	0.17
yes	716/1088	1.04	0.91-1.19	0.55	
Site					
Colon	4037/6285	0.97	0.91-1.02	0.26	0.25[c]
Rectum	1184/6285	1.04	0.95-1.14	0.36	

Abbreviations: CCFR, colon cancer family registry; CI, confidence interval; GRS, genetic risk score; OR, odds ratio per year.

[a] all analyses adjusted for age, sex, study and principal components of genetic ancestry;

[b] *P* value calculated using likelihood ratio tests comparing the model with and without interaction term;

[c] *P* value for heterogeneity was obtained in case-only analysis of colon vs. rectal cancer.

Table 3. Association of genetically predicted age at menopause and CRC risk according to subgroups, CCFR, GECCO Consortium					
	N (cases/controls)	OR[a]	95% CI	*P* value	Heterogeneity *P* value [b]
All	5832/6285	0.99	0.95-1.03	0.56	
Combined estrogen-progesterone therapy					
No	3266/3490	0.99	0.93-1.05	0.78	0.75
Yes	593/807	0.96	0.85-1.10	0.58	
Estrogen monotherapy					
No	3120/3214	0.99	0.93-1.05	0.27	0.76
Yes	716/1088	1.01	0.89-1.14	0.90	
BMI					
Normal weight	2146/2665	1.00	0.93-1.07	0.97	0.50
Overweight	1866/1918	0.99	0.91-1.07	0.77	
Obese	1284/1115	1.00	0.91-1.11	0.93	
Underweight	74/80	0.63	0.40-0.98	0.04	
Site					
Colon	4037/6285	0.98	0.94-1.03	0.53	0.61[c]
Rectum	1184/6285	0.98	0.91-1.06	0.65	

Abbreviations: BMI, body mass index; CCFR, colon cancer family registry; CI, confidence interval; GRS, genetic risk score; OR, odds ratio per year.

[a] all analyses adjusted for age, sex, study and principal components of genetic ancestry

[b] *P* value calculated using likelihood ratio tests comparing the model with and without interaction term

[c] *P* value for heterogeneity was obtained in case-only analysis of colon vs. rectal cancer