



OPEN Mendelian randomization study of sleep traits and risk of colorectal cancer

Olympia Dimopoulou^{1,2}✉, Harriett Fuller³, Rebecca C. Richmond^{1,2}, Emmanouil Bouras⁴, Bryony Hayes^{1,2}, Niki Dimou⁵, Neil Murphy⁵, Hermann Brenner^{6,7,8}, Andrea Gsur⁹, Loic Le Marchand¹⁰, Victor Moreno^{11,12,13,14}, Rish K. Pai¹⁵, Amanda I. Phipps^{3,16}, Caroline Y. Um¹⁷, Franzel J. B. van Duijnhoven¹⁸, Pavel Vodicka^{19,20,21}, Richard M. Martin^{1,2,22}, Elizabeth A. Platz^{23,24}, Marc J. Gunter⁵, Ulrike Peters^{3,16}, Sarah J. Lewis¹, Yin Cao^{25,26,27} & Konstantinos K. Tsilidis^{4,28}

A potential association of endogenous circadian rhythm disruption with risk of cancer development has been suggested, however, epidemiological evidence for the association of sleep traits with colorectal cancer (CRC) is limited and often contradictory. Here we investigated whether genetically predicted chronotype, insomnia and sleep duration are associated with CRC risk in males, females and overall and according to CRC anatomical subsites using Mendelian randomization (MR). The two-sample inverse variance weighted (IVW) method was applied using summary-level data in up to 58,221 CRC cases and 67,694 controls and genome-wide association data of genetic variants for self-reported sleep traits. Secondary analyses using alternative instruments and sensitivity analyses assessing potential violations of MR assumptions were conducted. Genetically predicted morning preference was associated with 13% lower risk of CRC in men ($OR_{IVW} = 0.87$, 95% CI = 0.78, 0.97, $P = 0.01$), but not in women or in both sexes combined. This association remained consistent in some, but not all, sensitivity analyses and was very similar for colon and rectal cancer. There was no evidence of an association for any other sleep trait. Overall, this study provides little to no evidence of an association between genetically predicted sleep traits and CRC risk.

¹Department of Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK. ²MRC Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK. ³Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. ⁴Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece. ⁵Nutrition and Metabolism Branch, International Agency for Research On Cancer-WHO, Lyon, France. ⁶Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁷Division of Preventive Oncology, German Cancer Research Center (DKFZ), National Center for Tumor Diseases (NCT), Heidelberg, Germany. ⁸German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany. ⁹Center for Cancer Research, Medical University of Vienna, Vienna, Austria. ¹⁰University of Hawaii Cancer Center, Honolulu, HI, USA. ¹¹Oncology Data Analytics Program, Catalan Institute of Oncology-IDIBELL, L'Hospitalet de Llobregat, Barcelona, Spain. ¹²CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain. ¹³Department of Clinical Sciences, Faculty of Medicine, University of Barcelona, Barcelona, Spain. ¹⁴ONCOBEL Program, Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, Barcelona, Spain. ¹⁵Department of Laboratory Medicine and Pathology, Mayo Clinic Arizona, Scottsdale, AZ, USA. ¹⁶Department of Epidemiology, University of Washington, Seattle, WA, USA. ¹⁷Population Science Department, American Cancer Society, Atlanta, GA, USA. ¹⁸Division of Human Nutrition, Wageningen University and Research, Wageningen, The Netherlands. ¹⁹Department of Molecular Biology of Cancer, Institute of Experimental Medicine of the Czech Academy of Sciences, Prague, Czech Republic. ²⁰Institute of Biology and Medical Genetics, First Faculty of Medicine, Charles University, Prague, Czech Republic. ²¹Faculty of Medicine and Biomedical Center in Pilsen, Charles University, Pilsen, Czech Republic. ²²University Hospitals Bristol and Weston NHS Foundation Trust and the University of Bristol National Institute for Health Research Bristol Biomedical Research Centre, University of Bristol, Bristol, UK. ²³Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. ²⁴Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, MD, USA. ²⁵Division of Public Health Sciences, Department of Surgery, Washington University in St. Louis, St. Louis, MO, USA. ²⁶Alvin J. Siteman Cancer Center, Washington University School of Medicine, St. Louis, MO, USA. ²⁷Division of Gastroenterology, Department of Medicine, Washington

University School of Medicine, St. Louis, MO, USA. ²⁸Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK. ✉email: olymphia.dimopoulou@bristol.ac.uk

Colorectal cancer (CRC) is the third most commonly diagnosed cancer and the second leading cause of cancer death worldwide, accounting for up to 10% of new cancer cases and 9% of cancer deaths in 2020¹. Night shift work was classified as a probable risk factor (Group 2A carcinogen) to humans according to the 2019 report by the International Agency for Research on Cancer². Beyond previously studied associations between night shift work and CRC, epidemiological evidence linking other metrics of circadian disruption with risk of CRC in the general population is emerging. A dose–response meta-analysis of six studies in the United States showed that long but not short sleep duration relative to 7–8 h of sleep was associated with a 21% increased risk of developing CRC³. In a more recent nested case–control study from Taiwan, participants (166 cases) with sleep disorders (having three or more outpatient diagnoses according to ICD-9) were found to have a 29% increased risk of developing CRC compared to those without⁴.

Genome-wide association studies (GWASs) have identified a large number of genetic variants robustly associated with sleep traits. According to recently published GWASs, 13.7% of chronotype (morning or evening preference) variance was explained by additive genetic variants⁵. Likewise, the heritability of insomnia was estimated at 16.7%⁶ and independently at 2.6%⁷ and of sleep duration at 9.8%⁸. Although data linking chronotype and insomnia with CRC risk are limited, Mendelian randomization (MR), a common approach to address inherent biases in observational studies and investigate potential causal associations provided that certain assumptions are met, has also been widely adopted to study the association of sleep-related traits with several outcomes, including breast^{9,10} and prostate cancer¹⁰, as well as known risk factors of CRC, such as adiposity¹¹ and type 2 diabetes mellitus (T2DM)¹². MR studies have reported that morning preference was associated with increased T2DM risk¹³, alcohol consumption¹⁴, educational attainment¹⁵, and reduced risks of breast^{9,10} and prostate cancer¹⁰; insomnia was associated with an increased T2DM risk, higher body mass index (BMI) and a decrease in educational attainment⁷; increased sleep duration was associated with an increased risk of breast cancer⁹ and a higher BMI in children⁸ and, lastly, short sleep duration was associated with an increased CRC risk using data from UK Biobank (UKB) based on 5,486 CRC cases¹⁶. To our knowledge, there is no other MR study that has investigated the potential association between sleep-related traits and CRC risk.

The aim of the current study was to assess whether genetically predicted chronotype, insomnia and sleep duration are associated with CRC risk in males, females and overall and according to CRC anatomical subsites using two-sample MR, leveraging data from the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO), the Colorectal Cancer Transdisciplinary Study (CORECT) and the Colon Cancer Family Registry (CCFR). We also investigated whether genetic variants associated with short sleep, long sleep and accelerometer-derived measures of sleep traits support the primary associations of self-reported sleep traits on CRC risk. We accounted for potential pleiotropy due to BMI, alcohol use, educational attainment, smoking and T2DM, phenotypes with a high frequency of associations with genome-wide significant sleep trait variants according to PhenoScanner database. Furthermore, we used genetic variants identified in sex-specific sleep trait GWASs to assess the extent to which findings differed by sex, as it has been suggested by observational findings^{17–21} on the effect of sleep traits on CRC outcomes and also by evidence from molecular studies reporting a differential timing of clock gene expression among men and women²².

Methods

Sleep trait GWASs

A GWAS Catalog search (<https://www.ebi.ac.uk/gwas>) was conducted in December 2020 for published GWASs on sleep traits. Genetic variants reported to be independently associated with self-reported chronotype⁵, insomnia symptoms^{6,7} and sleep duration⁸ at genome wide significance ($P < 5 \times 10^{-8}$) in participants of European ancestry were selected from the largest GWASs with adequate summary association data.

Self-reported chronotype was assessed at recruitment in UKB²³ and 23andMe²⁴ via a single question: “Do you consider yourself to be?” and “Are you naturally a night person or a morning person?”, respectively, and several possible answers, as described in more detail elsewhere⁵. To maximise statistical power, Jones et al. dichotomized chronotype into morning or evening preference and used results from both UKB and 23andMe in the GWAS meta-analysis ($n = 697,828$). Summary statistics were adjusted for age, sex, genotyping platform (UKB, 23andMe), study centre (UKB) and principal components (23andMe). Overall, 351 independent genetic variants at $P < 5 \times 10^{-8}$ were identified⁵.

Self-reported insomnia symptoms were assessed in UKB²³ in response to the question: “Do you have trouble falling asleep at night, or do you wake up in the middle of the night?”, and similar questions in the Nord-Trøndelag Health Study (HUNT)²⁵ and the Partners HealthCare Biobank (Partners Biobank)²⁶. Lane et al. performed two GWAS meta-analyses for frequent insomnia symptoms ($n = 254,767$) and for any insomnia symptoms ($n = 532,378$). The models were adjusted for age, sex, genetic ancestry, and genotyping array⁶, and a total of 57 independent variants at $P < 5 \times 10^{-8}$ were identified. Summary statistics for insomnia complaints measured with similar questionnaires were also obtained from a large-scale GWAS meta-analysis of up to 1,331,010 European-ancestry individuals from UKB ($n = 386,533$) and 23andMe ($n = 944,477$) and overall 250 lead variants at $P < 5 \times 10^{-8}$ were identified⁷.

Self-reported sleep duration was assessed at recruitment in UKB²³ with the question: “About how many hours sleep do you get in every 24 h? (please include naps)” and was treated as a continuous variable. Dashti et al. performed a GWAS among UKB participants ($n = 446,118$) and identified 78 independent variants after adjustment for age, sex, 10 principal components, genotyping array, and genetic correlation matrix. In addition, a fixed-effects meta-analysis of the UKB and the Cohorts for Heart and Aging Research in Genomics Epidemiology

(CHARGE) ($n=493,298$)²⁷ identified 13 additional new variants. Overall, 91 independent variants at $P < 5 \times 10^{-8}$ emerged from this study⁸.

Genetic variant selection

Out of the 351⁵ genetic variants for chronotype, 57⁶ and 250⁷ for insomnia symptoms and 91⁸ for continuous sleep duration, we excluded those variants which were not available in the CRC datasets, were palindromic with minor allele frequency ≥ 0.3 , were in linkage disequilibrium (LD) ($R^2 \geq 0.01$) and had an F-statistic $< 10^{28}$ to account for weak instrument bias based on the strength of their association with the sleep trait. The final number of variants used in the primary analysis were 290 for chronotype, 38 and 190 for insomnia and 76 for sleep duration, with 1% genetic variant overlap observed between chronotype and insomnia by Jansen et al. (2019), 2% between insomnia by Lane et al. (2019) and insomnia by Jansen et al. (2019), and no overlap in genetic variants between the rest of the exposures. Genetic variant selection is shown in Supplementary Figure S1.

Colorectal cancer GWASs

The association of the genetic instruments for chronotype, insomnia and sleep duration with CRC risk was obtained from the GECCO, CORECT and CCRF consortia^{29–31}. Further information on the characteristics of the contributing studies is provided in Supplementary Table S1. All CRC cases were defined as colorectal neoplasia and confirmed by pathological reports, medical records or death certificate information. In analysis of both sexes combined data on 58,221 CRC cases and 67,694 controls were available for 16,900,397 autosomal variants. GWAS results were also available by sex, with data on 59,889 females and 66,026 males overall, and by tumour anatomical subsite, with summary association data on 31,083 colon, 13,857 proximal colon, 15,306 distal colon and 15,775 rectal cancer cases.

Statistical power

An online tool available at <http://cnsgenomics.com/shiny/mRnd/> was employed to perform statistical power calculations³². For an odds ratio (OR) of 1.10 (or its reciprocal 0.91) per 1 standard deviation (SD) change in the exposure and a type 1 error of 5%, the statistical power (Supplementary Table S2) and the minimum detectable OR at 80% power were calculated (Supplementary Table S3).

Primary MR analyses

Prior to MR analysis, it is essential to harmonize the data so that all variants are standardized to associate with the exposure in the same direction and that the effect and other alleles are aligned between the instrument-exposure (GX) and the instrument-outcome (GY) datasets ensuring that they are both coded from the same strand. Pearson's correlation coefficient of the effect allele frequencies (EAFs) among datasets prior and post harmonization was estimated as a quality control measure of the harmonization process³³ (Supplementary Tables S6, S15, S22, S43). Two-sample MR was conducted using summary association data from published GWASs to assess whether genetically predicted sleep traits (chronotype, insomnia and sleep duration) are associated with CRC risk overall and according to sex and anatomical subsites. Meta-analysis of the Wald estimates for each variant was performed by implementing an inverse-variance weighted (IVW) method. The fixed-effects IVW model assumes absence of horizontal pleiotropy and will return an unbiased estimate provided that all variants are valid instrumental variables (IVs)³⁴. Correspondingly, the IVW random-effects model will return an unbiased estimate in the presence of balanced horizontal pleiotropy³⁵.

Sensitivity analyses for the assessment of MR assumptions

To evaluate instrument strength and potential violation of the first MR assumption, i.e., the genetic variants must be associated with the sleep trait of interest^{36,37}, the F-statistic for each variant-trait association was calculated³⁸. According to the second and third MR assumptions, the genetic variants should not be associated with any confounder of the sleep trait-CRC relationship and should not influence CRC via some other pathway than that of the sleep trait^{36,37}. To account for potential horizontal pleiotropy, another five MR methods were performed, each of them providing a valid MR estimate under different combinations of assumptions. MR-Egger provides an unbiased causal effect estimate even if the third MR assumption is violated and all the variants are invalid IVs, provided that the InSIDE assumption of independency between the horizontal pleiotropic effects and the variants-exposure effects is met³⁹. Instrument strength and variability of the genetic instrument-exposure association due to potential violation of the NO Measurement Error (NOME) assumption were also inspected via the I^2_{GX} statistic⁴⁰. The weighted median returns a valid estimate when at least 50% of the weight of the genetic variants comes from valid IVs⁴¹. The weighted mode is valid under the Zero Modal Pleiotropy Assumption (ZEMPA) according to which out of all clusters of variants with similar effects, the largest is the group of valid IVs⁴². The contamination mixture is a likelihood-based method that identifies groups of variants with similar estimates, an indicator of the existence of distinct mechanisms in the exposure-outcome relationship, and provides robust estimates under the ZEMPA assumption⁴³. To detect potential outlying genetic variants, we implemented the MR pleiotropy residual sum and outlier test (MR-PRESSO), a method that identifies and excludes outliers, applying a random-effects IVW model⁴⁴. The Cochran's Q statistic for heterogeneity was also computed³⁴. To further graphically assess the presence of pleiotropy, we generated diagnostic scatter, funnel and forest plots. To explore potential associations between the genetic instruments with secondary phenotypic traits that could indicate possible violations in the second or third MR assumptions, we queried the PhenoScanner database (<http://www.phenoscanter.medschl.cam.ac.uk/>) using a P threshold of 1×10^{-5} and tabulated the resulting associations per variant. All analyses were conducted in the statistical software R version 4.2.0 using the packages MendelianRandomization⁴⁵ and MRCIEU/TwoSampleMR³⁶.

Secondary analyses using alternative exposures/instruments

To account for potential non-linear associations of sleep duration with CRC risk, we performed two-sample MR using 27 genome-wide significant variants for short sleep (< 7 h vs. 7–8 h; $n = 106,192$ cases and 305,742 controls) and 8 variants for long sleep (> 8 h vs. 7–8 h; $n = 34,184$ cases and 305,742 controls) in participants of European ancestry from UKB⁸.

Given that all sleep traits were based on self-reported information, we assessed the robustness of our findings using accelerometer-derived equivalents of sleep measures in up to 85,670 participants from UKB. We used genetic variants for seven accelerometer-derived sleep traits representative of chronotype [L5 timing defined as the midpoint (number of hours since the previous midnight) of the least active five hours of each day, reflects whether a person goes to bed earlier or later in the day (6 variants); M10 timing defined as the midpoint of the most-active 10 h, reflects whether a person is most active earlier or later in the day (1 variant); sleep midpoint (1 variant)], insomnia [number of sleep episodes (21 variants)] and sleep duration [actigraphy nocturnal sleep duration (11 variants); sleep efficiency (5 variants); diurnal inactivity (2 variants)] as described elsewhere⁴⁶. The genetic correlation of self-reported and accelerometer-derived measures has been investigated in a previous study, according to which the correlation coefficient is equal to 0.903 between chronotype and L5 timing, 0.095 between any insomnia and the number of sleep episodes, and 0.430 between sleep duration and actigraphy nocturnal sleep duration⁹. The observed overlap in genetic variants was approximately 1% between L5 timing and chronotype, and 1%, 0.5%, 1% and 0.5% between insomnia by Jansen et al. (2019) and L5 timing, number of sleep episodes, actigraphy nocturnal sleep duration and sleep efficiency, respectively. There was no overlap between the rest of accelerometer-derived measures and any self-reported sleep trait.

Traits with a high frequency of associations with genome-wide significant variants for sleep traits using the PhenoScanner database, i.e., BMI, alcohol use, educational attainment, lifetime smoking and T2DM were followed up in Multivariable Mendelian randomization (MVMR) analysis to assess direct effects of sleep traits independent of potential confounders and to evaluate potential pleiotropy arising from those factors⁴⁷. In order for the MVMR estimates to be valid, the genetic variants should be robustly associated with at least one of the exposures in the model (relevance), be independent of all confounders of all the exposure-outcome associations (exchangeability) and be independent of the outcome conditional on the exposures and confounders (exclusion restriction). Details on the GWASs used in MVMR analysis are available in Supplementary Table S28 and summary association data for the corresponding genetic instruments are included in Supplementary Tables S29–S33. To test for instrument strength in the MVMR setting, the conditional F-statistic was calculated with a threshold of less than 10 suggesting probable weak instrument bias⁴⁸. To further test instrument strength, a modified version of the Cochran's Q statistic was calculated under the null hypothesis that IVs do not sufficiently account for any of the variation in the exposures. Also, to test for pleiotropy, an adjusted version of the Cochran's Q statistic (Q_A), under the null hypothesis of no heterogeneity in the genetic variants (Supplementary Table S34), as well as the MVMR MR-Egger regression⁴⁹ (Supplementary Table S35) were used. MVMR analysis was conducted using the R packages MendelianRandomization⁴⁵ and MRPracticals⁴⁹, R version 4.2.0. MVMR findings were followed up in a bi-directional MR framework to decipher the direction of the association of sleep traits with factors of interest from the MVMR analyses (Supplementary Tables S36–S41).

The genetic instruments used in all previous primary and secondary analyses were identified in the corresponding GWAS meta-analyses for chronotype, insomnia and sleep duration, in which results for females and males were combined. We also carried out MR analyses using instruments identified in sex-specific GWASs from UKB. Following LD-clumping, 32 variants were associated at $P < 5 \times 10^{-8}$ with chronotype in males, 65 with chronotype in females, 10 with frequent insomnia symptoms in males, 15 with frequent insomnia symptoms in females, 7 with sleep duration in males, and 16 with sleep duration in females. After relaxing the threshold at $P < 5 \times 10^{-6}$ in supplementary analysis to increase the number of instruments, 199, 126, 90, 58, 80 and 72 variants were identified, respectively.

Results

An outline of our study design is presented in Figure 1. Associations of the genetic instruments with the exposures/outcomes are shown in Supplementary Tables S4, S5 (primary analyses) and Supplementary Tables S13, S14, S21, S29–S33, S36, S42 (secondary analyses). Figure 2 shows all primary MR associations. Table 1 depicts all primary IVW MR associations and two sensitivity analyses for the assessment of MR assumptions (Weighted median, MR-Egger). The random-effects IVW estimate is reported throughout the manuscript (OR_{IVW}), unless otherwise specified where there is under-dispersion in the association estimates ($OR_{IVW-fixed}$). All two-sample MR and corresponding sensitivity analyses are detailed in Supplementary Tables S7–S12 (primary analyses) and Supplementary Tables S16–S20, S23–S27, S34, S35, S37–S41, S44–S48 (secondary analyses). Diagnostic scatter, funnel and forest plots for the primary analyses are shown in supplementary figures (Supplementary Fig. S2–S148).

Primary MR analyses of sleep traits and CRC and assessment of MR assumptions

The F-statistics for the eligible genetic variants did not indicate the presence of weak instruments (F-statistic range = 30–430 for chronotype, 16–217 for insomnia, 30–221 for sleep duration) (Supplementary Table S4).

Two-sample MR suggested a 13% lower risk of CRC in men with morning preference in comparison with evening preference chronotype [$OR_{IVW} = 0.87$, 95% confidence interval (CI) = 0.78, 0.97, $P = 0.01$]. This finding was consistent in site-specific analysis of colon cancer ($OR_{IVW} = 0.86$, 95% CI = 0.76, 0.98, $P = 0.02$) and rectal cancer ($OR_{IVW-fixed} = 0.87$, 95% CI = 0.75, 1.00, $P = 0.04$) in males. However, there was no evidence of an association for chronotype with CRC risk in women ($OR_{IVW} = 0.99$, 95% CI = 0.89, 1.09, $P = 0.78$), with an 83.2% of the variation in the effect estimates between men and women being attributed to heterogeneity rather than chance (Cochrane's $Q = 5.97$, Cochrane's Q test $P = 0.02$). Similarly, there was no evidence of an association in both

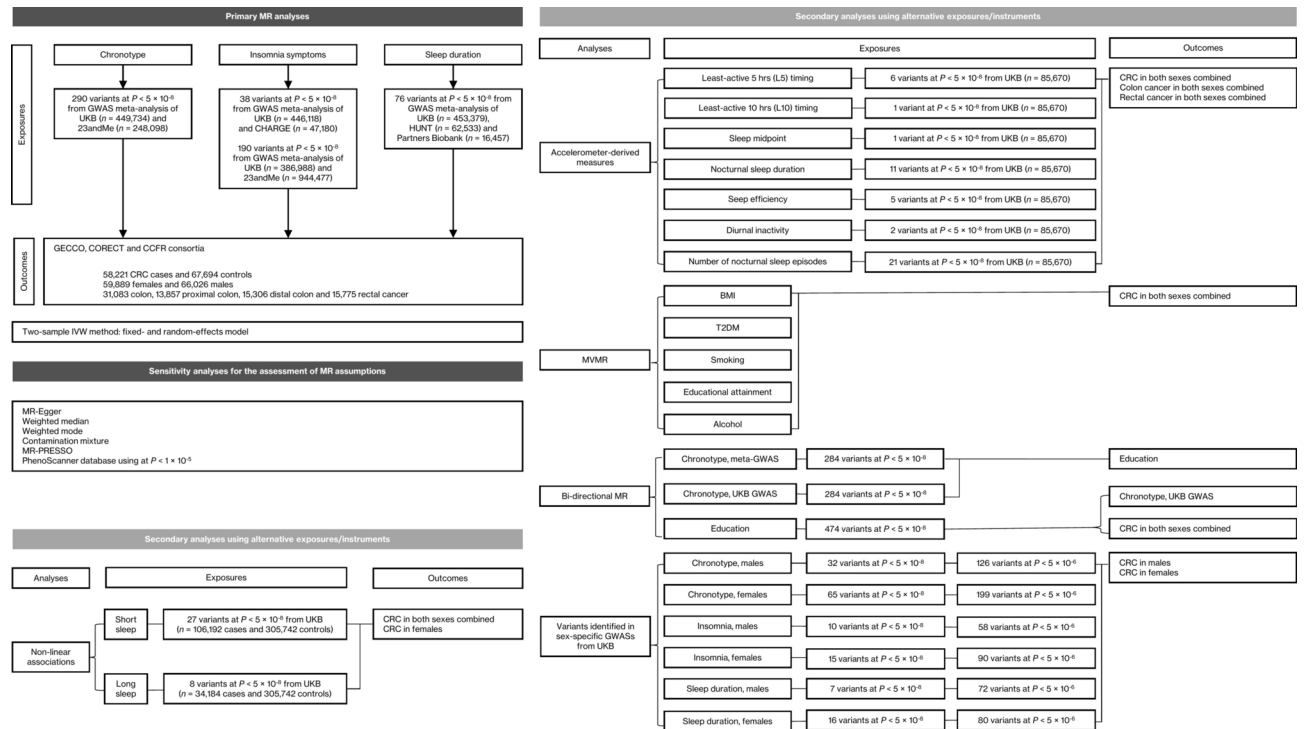


Fig. 1. Outline of the study design and data sources for the primary and secondary analyses performed. BMI: body mass index; CCFR: Colon Cancer Family Registry; CHARGE: Cohorts for Heart and Aging Research in Genomics Epidemiology; CORECT: Colorectal Cancer Transdisciplinary Study; CRC: colorectal cancer; GECCO: Genetics and Epidemiology of Colorectal Cancer Consortium; GWAS: Genome-wide association study; HUNT: Nord-Trøndelag Health Study; L5: midpoint of the least active five hours of each day; M10: midpoint of the most-active 10 h; MR: Mendelian randomization; MR-PRESSO: MR pleiotropy residual sum and outlier test; MVMR: Multivariable Mendelian randomization; Partners Biobank: Partners HealthCare Biobank; T2DM: type 2 diabetes mellitus; UKB: UK Biobank.

sexes combined ($OR_{IVW} = 0.97$, 95% CI = 0.92, 1.02, $P = 0.19$) or in site-specific analyses (Table 1, Supplementary Table S7 and Fig. 2). There was strong evidence for heterogeneity that cannot be explained by sampling variation alone among the causal estimates, suggesting potential violation of the MR assumptions (Supplementary Table S9; smallest Cochran's Q test $P = 6.80 \times 10^{-11}$). Some indication of horizontal pleiotropy based on the MR-Egger intercept test was evident in the association of morning preference with CRC in females (intercept $P = 0.04$, $OR_{MR-Egger} = 0.77$, 95% CI = 0.60, 0.99, $P = 0.04$) (Table 1, Supplementary Tables S7, S8), however, the association was not supported by any other primary or sensitivity analyses. Results were consistent in terms of direction in most of the applied sensitivity analyses (Table 1, Supplementary Table S7), although the association of chronotype with CRC, colon and rectal cancer in males attenuated in the weighted median analysis. The MR-PRESSO analysis identified two outliers for chronotype (i.e., rs45597035, rs6007594) (Supplementary Table S10). A PhenoScanner search revealed that the variants used as genetic instruments for MR analyses (Supplementary Table S12) and the outlying genetic instruments from MR-PRESSO (Supplementary Table S11) were associated at $P < 5 \times 10^{-5}$ with several secondary phenotypic traits relevant to CRC. However, after removal of the outlying variants, the IVW effect estimates for the association of chronotype with CRC (MR-PRESSO $OR_{IVW} = 0.89$, 95% CI = 0.78, 0.99, $P = 0.02$) and with colon cancer in males (MR-PRESSO $OR_{IVW} = 0.87$, 95% CI = 0.74, 1.00, $P = 0.03$) remained largely unchanged (Supplementary Table S10).

There was no evidence of an association for genetically predicted frequent/any insomnia symptoms with CRC or its subsites in both sexes combined and in sex-stratified analyses (Table 1, Supplementary Table S7 and Fig. 2). The results were largely consistent in sensitivity analyses (Table 1, Supplementary Tables S7-S10). Some indication of horizontal pleiotropy based on the MR-Egger intercept test was evident in the association of insomnia with proximal cancer in both sexes combined (intercept $P = 0.01$, $OR_{MR-Egger} = 0.46$, 95% CI = 0.26, 0.84, $P = 0.01$) and with rectal cancer in females (intercept $P = 0.03$, $OR_{MR-Egger} = 1.69$, 95% CI = 1.09, 2.61, $P = 0.02$) (Table 1, Supplementary Tables S7, S8), however, these associations were not consistent among other sensitivity analysis methods.

Two-sample IVW MR analysis found some evidence of an association of sleep duration with rectal cancer in males ($OR_{IVW} = 0.70$, 95% CI = 0.50, 0.98 per hour increase, $P = 0.04$), however, this finding was not supported by sensitivity analyses. There was no evidence of an association for genetically predicted sleep duration with CRC or its subsites in both sexes combined and in other sex-stratified analyses (Table 1, Supplementary Table S7 and Fig. 2).

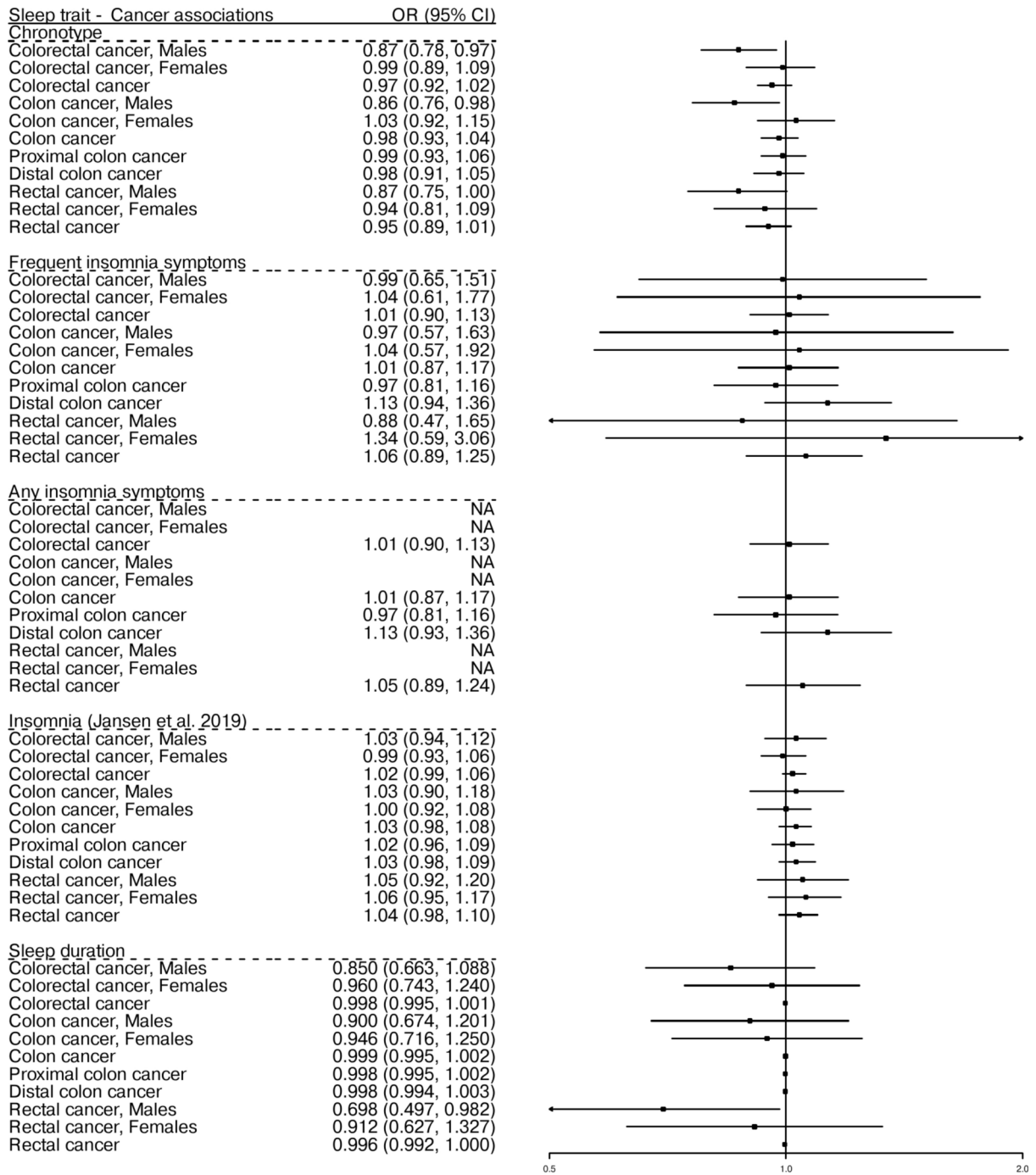


Fig. 2. Mendelian randomization (MR) estimates for the association of sleep traits with colorectal cancer outcomes in primary MR analyses. Forest plot of two-sample MR random-effects inverse-variance weighted estimates for the association between sleep traits and risk of colorectal, colon, proximal colon, distal colon, and rectal cancer in both sexes combined and in sex-stratified analyses. CI: confidence interval; NA: Non Applicable; OR: odds ratio.

Secondary analyses

Two-sample MR showed no clear evidence that either short ($OR_{IVW} = 1.02$, 95% CI = 0.75, 1.38 per doubling of genetic liability for short sleep duration, $P = 0.91$) or long sleep duration ($OR_{IVW-fixed} = 1.03$, 95% CI = 0.91, 1.17 per doubling of genetic liability for long sleep duration, $P = 0.62$) compared to 7–8 h of sleep were associated with CRC in both sexes combined (F-statistic range = 16–57 for short sleep duration, 16–46 for long sleep duration). Likewise, there was no evidence of an association for either short or long sleep duration and CRC in females (Supplementary Table S16). Corresponding data in males were not available. The results were consistent in sensitivity analyses (Supplementary Tables S16–S20).

Exposure Outcome	Method	Chromotopy			Frequent insomnia symptoms			Any insomnia symptoms			Insomnia—Jansen et al. (2019)			Sleep duration		
		GVs	OR (95% CI)	P	GVs	OR (95% CI)	P	GVs	OR (95% CI)	P	GVs	OR (95% CI)	P	GVs	OR (95% CI)	P
Colorectal cancer, Males	IVW random-effects model	289	0.87 (0.78, 0.97)	0.01	38	0.99 (0.65, 1.51)	0.96	NA	NA	NA	1.03 (0.94, 1.12)	0.53	76	0.850 (0.663, 1.088)	0.20	
	Weighted median method	289	0.91 (0.79, 1.05)	0.20	38	1.04 (0.58, 1.89)	0.89	NA	NA	NA	0.94 (0.83, 1.07)	0.37	76	1.003 (0.728, 1.381)	0.99	
	MR-Egger method	289	0.92 (0.69, 1.23)	0.56	38	0.34 (0.07, 1.66)	0.18	NA	NA	NA	0.81 (0.61, 1.07)	0.14	76	0.760 (0.336, 1.721)	0.51	
Colorectal cancer, Females	IVW random-effects model	290	0.99 (0.89, 1.09)	0.78	38	1.04 (0.61, 1.77)	0.89	NA	NA	NA	0.99 (0.93, 1.06)	0.85	76	0.960 (0.743, 1.240)	0.75	
	Weighted median method	290	0.95 (0.82, 1.09)	0.44	38	0.71 (0.36, 1.37)	0.31	NA	NA	NA	0.94 (0.86, 1.04)	0.24	76	1.246 (0.915, 1.697)	0.16	
	MR-Egger method	290	0.77 (0.60, 0.99)	0.04	38	0.34 (0.05, 2.24)	0.26	NA	NA	NA	1.12 (0.86, 1.47)	0.40	76	0.842 (0.359, 1.973)	0.69	
Colorectal cancer	IVW random-effects model	290	0.97 (0.92, 1.02)	0.19	38	1.01 (0.90, 1.13)	0.87	38	1.01 (0.90, 1.13)	0.88	1.02 (0.99, 1.06)	0.25	76	0.998 (0.995, 1.001)	0.27	
	Weighted median method	290	0.96 (0.90, 1.02)	0.16	38	1.02 (0.87, 1.19)	0.85	38	1.01 (0.86, 1.19)	0.89	1.02 (0.97, 1.08)	0.38	76	0.998 (0.994, 1.002)	0.29	
	MR-Egger method	290	0.95 (0.83, 1.08)	0.42	38	0.81 (0.54, 1.20)	0.29	38	0.81 (0.54, 1.20)	0.29	1.13 (0.98, 1.31)	0.09	76	0.995 (0.983, 1.007)	0.41	
Colon cancer, Males	IVW random-effects model	289	0.86 (0.76, 0.98)	0.02	38	0.97 (0.57, 1.63)	0.90	NA	NA	NA	1.03 (0.90, 1.18)	0.63	76	0.900 (0.674, 1.201)	0.47	
	Weighted median method	289	0.95 (0.80, 1.14)	0.61	38	1.01 (0.48, 2.14)	0.98	NA	NA	NA	0.93 (0.79, 1.08)	0.33	76	1.133 (0.757, 1.695)	0.54	
	MR-Egger method	289	0.97 (0.69, 1.37)	0.87	38	0.41 (0.06, 2.91)	0.37	NA	NA	NA	0.75 (0.49, 1.13)	0.17	76	0.872 (0.336, 2.260)	0.78	
Colon cancer, Females	IVW random-effects model	290	1.03 (0.92, 1.15)	0.62	38	1.04 (0.57, 1.92)	0.89	NA	NA	NA	1.00 (0.92, 1.08)	0.93	76	0.946 (0.716, 1.250)	0.70	
	Weighted median method	290	1.07 (0.91, 1.26)	0.43	38	1.23 (0.57, 2.65)	0.61	NA	NA	NA	0.94 (0.84, 1.05)	0.31	76	0.972 (0.680, 1.388)	0.88	
	MR-Egger method	290	0.80 (0.60, 1.08)	0.15	38	0.32 (0.04, 2.75)	0.30	NA	NA	NA	1.08 (0.78, 1.49)	0.63	76	0.691 (0.276, 1.731)	0.43	
Colon cancer	IVW random-effects model	290	0.98 (0.93, 1.04)	0.50	38	1.01 (0.87, 1.17)	0.92	38	1.01 (0.87, 1.17)	0.93	1.03 (0.98, 1.08)	0.24	76	0.999 (0.995, 1.002)	0.45	
	Weighted median method	290	1.00 (0.93, 1.08)	0.91	38	1.07 (0.88, 1.29)	0.51	38	1.06 (0.87, 1.28)	0.58	1.04 (0.98, 1.11)	0.18	76	1.001 (0.996, 1.005)	0.82	
	MR-Egger method	290	0.99 (0.85, 1.15)	0.91	38	0.75 (0.45, 1.25)	0.27	38	0.75 (0.45, 1.23)	0.26	1.11 (0.93, 1.34)	0.25	76	0.993 (0.980, 1.006)	0.28	
Proximal colon cancer	IVW random-effects model	290	0.99 (0.93, 1.06)	0.82	38	0.97 (0.81, 1.16)	0.77	38	0.97 (0.81, 1.16)	0.75	1.02 (0.96, 1.09)	0.50	76	0.998 (0.995, 1.002)	0.34	
	Weighted median method	290	1.01 (0.92, 1.11)	0.76	38	1.02 (0.80, 1.30)	0.87	38	1.01 (0.80, 1.29)	0.92	1.06 (0.98, 1.15)	0.13	76	0.999 (0.993, 1.004)	0.63	
	MR-Egger method	290	1.01 (0.83, 1.21)	0.95	38	0.46 (0.26, 0.84)	0.01	38	0.47 (0.26, 0.84)	0.01	1.06 (0.83, 1.35)	0.63	76	0.993 (0.979, 1.007)	0.30	
Distal colon cancer	IVW random-effects model	290	0.98 (0.91, 1.05)	0.60	38	1.13 (0.94, 1.36)	0.20	38	1.13 (0.93, 1.36)	0.21	1.03 (0.98, 1.09)	0.27	76	0.998 (0.994, 1.003)	0.50	
	Weighted median method	290	1.00 (0.91, 1.10)	0.98	38	1.30 (1.01, 1.66)	0.04	38	1.26 (0.98, 1.62)	0.07	1.04 (0.96, 1.13)	0.35	76	0.998 (0.993, 1.004)	0.58	
	MR-Egger method	290	0.97 (0.80, 1.19)	0.79	38	1.35 (0.70, 2.63)	0.37	38	1.29 (0.67, 2.49)	0.45	1.12 (0.90, 1.39)	0.32	76	0.992 (0.975, 1.008)	0.32	
Rectal cancer, Males	IVW random-effects model	289	0.87 (0.75, 1.00)	0.06	38	0.88 (0.47, 1.65)	0.70	NA	NA	NA	1.05 (0.92, 1.20)	0.44	76	0.698 (0.497, 0.982)	0.04	
	Weighted median method	289	0.84 (0.68, 1.04)	0.12	38	0.94 (0.39, 2.26)	0.89	NA	NA	NA	0.99 (0.83, 1.18)	0.90	76	0.799 (0.497, 1.284)	0.35	
	MR-Egger method	289	0.81 (0.55, 1.20)	0.30	38	1.05 (0.10, 11.03)	0.97	NA	NA	NA	0.96 (0.63, 1.47)	0.87	76	0.895 (0.291, 2.759)	0.85	
Rectal cancer, Females	IVW random-effects model	290	0.94 (0.81, 1.09)	0.43	38	1.34 (0.59, 3.06)	0.48	NA	NA	NA	1.06 (0.95, 1.17)	0.31	76	0.912 (0.627, 1.327)	0.63	
	Weighted median method	290	0.87 (0.70, 1.10)	0.24	38	1.07 (0.34, 3.32)	0.91	NA	NA	NA	1.08 (0.92, 1.26)	0.33	76	0.884 (0.528, 1.480)	0.64	
	MR-Egger method	290	0.91 (0.62, 1.35)	0.64	38	0.64 (0.03, 12.25)	0.77	NA	NA	NA	1.69 (1.09, 2.61)	0.02	76	1.099 (0.317, 3.814)	0.88	
Rectal cancer	IVW random-effects model	290	0.95 (0.89, 1.01)	0.13	38	1.06 (0.89, 1.25)	0.52	38	1.05 (0.89, 1.24)	0.58	1.04 (0.98, 1.10)	0.18	76	0.996 (0.992, 1.000)	0.06	
	Weighted median method	290	0.98 (0.89, 1.08)	0.72	38	1.19 (0.93, 1.51)	0.16	38	1.18 (0.92, 1.50)	0.18	1.03 (0.95, 1.11)	0.45	76	0.996 (0.991, 1.002)	0.20	
	MR-Egger method	290	0.93 (0.77, 1.12)	0.43	38	1.40 (0.78, 2.52)	0.26	38	1.27 (0.71, 2.27)	0.42	1.23 (0.99, 1.52)	0.06	76	0.999 (0.983, 1.015)	0.89	

Table 1. Two-sample Mendelian randomization estimates for the association between sleep traits and risk of colorectal, colon, proximal colon, distal colon, and rectal cancer in both sexes combined and in sex-stratified analyses. CI: Confidence Interval; GV: Genetic Variants; IVW: Inverse-Variance Weighted; MR: Mendelian Randomization; NA: Non Applicable; OR: Odds Ratio

In analyses using variants associated with accelerometer-derived equivalents of sleep measures (F-statistic range = 30–155), there was no evidence of an association for L5 timing ($OR_{IVW-fixed} = 1.04$, 95% CI = 0.86, 1.27 per hour elapsed since previous midpoint, $P = 0.67$), M10 timing ($OR_{IVW-fixed} = 1.23$, 95% CI = 0.68, 2.23 per hour elapsed since previous midpoint, $P = 0.49$), sleep midpoint (Wald ratio estimator = 1.39, 95% CI = 0.70, 2.76 per hour increase, $P = 0.34$), actigraphy nocturnal sleep duration ($OR_{IVW-fixed} = 0.95$, 95% CI = 0.81, 1.12 per hour increase, $P = 0.53$), sleep efficiency ($OR_{IVW-fixed} = 0.97$, 95% CI = 0.76, 1.25 per 1% increase, $P = 0.83$), diurnal inactivity (Wald ratio estimator = 1.32, 95% CI = 0.89, 1.97 per hour increase, $P = 0.17$) and number of sleep episodes ($OR_{IVW-fixed} = 0.96$, 95% CI = 0.84, 1.09 per sleep episode, $P = 0.51$) and CRC using both the IVW method and in sensitivity analyses. We further assessed the association of accelerometer-derived measures with colon and rectal cancer, but the results did not change (Supplementary Tables S23–S27). Although the findings for L5, M10 and sleep midpoint have wide CIs, they are consistent with the association of evening preference with an increased CRC risk.

Results from the MVMR analysis on CRC in both sexes combined were largely unchanged in terms of direction and increased in magnitude when accounting for BMI, alcohol, smoking or T2DM compared to the unadjusted association estimates. After adjusting for educational attainment, a direct effect of chronotype on CRC was observed ($OR_{educational\ attainment-adj} = 0.80$, 95% CI = 0.65, 0.98, $P = 0.03$) (Supplementary Table S34). In bi-directional MR analysis, a SD increase in educational attainment was associated with a decrease in morning preference chronotype and CRC risk ($OR_{IVW} = 0.98$, 95% CI = 0.97, 1.00, $P = 0.01$ and $OR_{IVW} = 0.75$, 95% CI = 0.69, 0.82, $P = 3.88 \times 10^{-10}$, respectively), and morning compared to evening preference was associated with a decrease in years of education by 0.98 SD (Supplementary Table S37). The bi-directional association between chronotype and educational attainment supports that morning preference is associated with a lower educational attainment and vice versa, however, it does not preclude a potential role of educational attainment as a confounder in the association of chronotype with CRC.

Finally, the association of morning preference with a decrease in CRC risk in males was consistent (but with wider CIs) in MR analyses using instruments identified in sex-specific GWASs from UKB (Supplementary Table S44).

Discussion

In the current study, we employed genetic variants as IVs in a two-sample MR framework to investigate potential associations of three modifiable sleep traits, including chronotype, the endogenous circadian rhythm, insomnia symptoms and sleep duration with CRC risk. Morning compared with an evening preference was suggestively associated with lower CRC and colon cancer risk among men. This finding was consistent in some, but not all, sensitivity analysis methods. Site-specific analyses and secondary analyses using alternative exposures/instruments predominantly supported this association, although with wider CIs. There was no evidence of an association between chronotype and CRC risk in women or in both sexes combined. There was also no evidence of an association between insomnia and sleep duration and CRC or its subsites.

The epidemiological evidence for an association between sleep traits and CRC risk is limited and, in most cases, inconsistent. The suggestive association of morning preference with a decreased risk of CRC and colon cancer in men is supported by previous findings from a population-based case-control study, according to which colon cancer risk was elevated among men who ever worked at night compared to men who never worked at night¹⁷, and a cohort of Korean male professional emergency responders (ERs), whose standardized incidence ratios showed an increase with reference to the general population¹⁸. The adverse effect of light-at-night exposure, inherent to night-shift work, on carcinogenesis is aligned with animal study observations⁵⁰. In addition, the lack of an association between chronotype and CRC risk in women is in accordance with an earlier female cohort of radio and telegraph operators working evening or night shifts¹⁹, and with two female cohorts of nurses working rotating night shifts, the Nurses' Health Study (NHS)²⁰ and the NHS2^{20,21}. In a previous NHS analysis, a positive association between night shift work for over 15 years and CRC risk was documented⁵¹; however, in a most recent analysis of the NHS cohort by Papantoniou et al. (2018), which expanded both the follow-up time and sample size, the former relationship was attenuated²⁰. It is worth noting that circadian misalignment and its potential adverse effects caused by night shift work might vary depending on diurnal preference, with Vetter et al. (2015) reporting an increased risk in T2DM in women with timing discrepancy between their chronotype and work schedule⁵². Moreover, according to the Bureau of Labor Statistics of the U.S. Department of Labor, 15% of full-time wage and salary employees worked on alternative shifts in 2004, but only 11.5% of them worked a non-day schedule out of personal preference⁵³. In this context and given the shortage of epidemiological evidence evaluating the effect of chronotype per se on carcinogenesis and the insufficient evaluation in existing epidemiological studies of the participants' socioeconomic status impact on the potential interconnection between shift work and colorectal malignancies, shift work should be used with prudence as a proxy for chronotype for lack of a more suitable measure, considering also that chronotype could potentially be acting as an effect modifier for shift work i.e., the magnitude of the effect of shift work on cancer risk might differ depending on an individual's chronotype. The speculated difference in the effect of chronotype by sex is augmented by molecular studies on clock gene expression, according to which daily rhythms of *PER2*, *PER3* and *ARNTL1* genes are delayed in men compared to women²² as well as by evidence supporting an earlier pineal melatonin secretion in women compared to men relative to their sleep time⁵⁴. Thus, the observed association of morning preference with CRC and colon cancer risk in men could potentially be attributed to inherent sex differences and not to sample variation alone.

We observed a lack of an association between insomnia and risk of CRC in the current study using both self- and accelerometer-derived measures of insomnia. The existing literature is limited to examining insomnia symptoms as a disease aftermath in cancer patients or survivors and is accordingly short on exposure data prior

to the onset of the malignancy. Therefore, there is a need for the accumulation of exposure data long before cancer diagnosis to assess whether insomnia is a potential risk factor for CRC in an observational setting.

No evidence of an association between sleep duration and CRC risk was observed both in this comprehensive MR study and in the literature. Specifically, two cohorts in women^{21,55}, one cohort in both sexes combined⁵⁶, and an MR study using data from the UKB¹⁶ reported no evidence of an association between sleep duration and CRC. Our secondary analyses using accelerometer-derived measures for sleep duration showed no evidence for an association of sleep duration equivalents, with CRC, colon and rectal cancer and our analysis assessing potential non-linear effects of sleep duration also indicated that neither short nor long sleep duration associated with CRC risk relative to 7–8 h of sleep. These findings partly support the results from a recent MR study of 367,586 UKB participants reporting no evidence of an effect for long sleep ($P=0.66$) and 48% higher odds of CRC ($P=0.01$) for short sleep duration. However, this study included only 5,486 CRC cases from UKB and its results were not supported in replication analysis using data from the FinnGen consortium¹⁶. In contrast to our findings, some observational studies reported an adverse effect of sleep duration. Among those, three are case–control studies^{57–59}, prone to recall bias and further limitations due to their design, and two are cohorts but with data on specific population groups (postmenopausal women receiving hormone replacement therapy and overweight individuals who snorted) and therefore ambiguous in terms of generalizability^{60,61}. Moreover, according to a previous study, there is a moderate correlation between self-reported and measured sleep duration⁶², introducing potential information bias in the arising role of sleep duration as a potential risk factor for CRC. Consequently, the observed variation in the literature could be partially attributed to the potentially high levels of participants' misclassification, and hence any future studies should focus on the accurate characterization of sleep traits.

To our knowledge, this study constitutes the first comprehensive MR investigation of the association of five self-reported sleep traits and seven accelerometer-derived equivalents with the risk of CRC and its subsites in both sex-combined and in sex-stratified analyses. Among its principal strengths is the implementation of MR, a methodology which counteracts by design some of the main limitations inherent to observational studies, namely exposure measurement error, reverse causation, and residual confounding, provided that certain assumptions are met^{36,37}. Although it is not possible to eliminate potential violations of the MR assumptions, we used several sensitivity analyses, with differences in statistical power and assumptions to be satisfied, to address this issue, and the interpretation of our findings was based on a thorough co-assessment of these methods. To account for potential heterogeneity, we implemented additional stratified analyses by sex and CRC anatomical subsites. We further conducted analyses to explore potential pleiotropic effects and detect outlying variants using MR-PRESSO. MVMR was implemented to assess the direct effects of sleep traits on CRC and to account for potential horizontal pleiotropy due to suspected confounders arising from our PhenoScanner search. Another key strength is the use of a large number of genetic variants robustly associated with sleep traits and identified in meta-analysis GWASs of well-established epidemiological cohorts with the largest available sample size to date, as well as the inclusion of colorectal cancer data from the GECCO, CORECT and CCFR consortia. In the case of insomnia, two meta-analytic summary association datasets were used to account for winner's curse, i.e., the effect sizes for the genetic variants being larger in the discovery sample compared to the general population, the published GWAS by Jansen et al. (2019)⁷ and the second largest available dataset by Lane et al. (2019)⁶. However, there is an overlap of UKB participants in these datasets. As genetic variants for all analyses were identified in GWASs of both sexes combined, we also examined whether our findings were supported by genetic variants identified in sex-specific sleep trait GWASs. To assist the reproducibility of results and the evaluation of the harmonization process, all post harmonization datasets are available as supplementary material.

One potential limitation of the current study is the use of self-reported exposure data in the conducted GWASs. To mitigate this source of bias, we conducted secondary analyses using accelerometer-derived equivalents to the self-reported traits as more objective measures. However, given the limited number of variants associated with accelerometer-derived sleep traits, the smaller datasets and the lack of sex-specific data, the resulting effect estimates were of low precision. In addition, the use of summary data poses a restriction to the examination of potential non-linear effects and the stratification by other potential risk factors of the exposure–outcome association. However, we did try to account for potential deviation from linearity using instruments robustly associated with short and long sleep duration. Also, the use of smaller datasets to assess the site-specific effects of sleep traits might be limiting the precision of the corresponding analyses. Our study was also restricted to participants of European ancestry, so further studies are required to explore the associations of sleep traits in non-European ancestry groups. Another limitation related to the selection of participants is a small sample overlap between the exposure and outcome datasets. However according to simulation studies this bias should be considered negligible especially in the case of no weak genetic instruments⁶³. It should be noted that although we did not explicitly correct for multiple testing, we interpreted our findings in a conservative way to account for the large number of exposures and outcomes evaluated in this study. Furthermore, results from MVMR and bi-directional analyses suggested a potential role of educational attainment in the chronotype–CRC association, which remains to be investigated in future studies.

This study provides little to no evidence of an association between genetically predicted sleep traits and CRC risk. The suggestive association observed between genetically predicted chronotype and CRC in men was consistent in terms of direction in some sensitivity and secondary analyses, however, due to their lack in precision and the aforementioned limitations of our study, any observations should be interpreted tentatively. Considering the ever-increasing circadian rhythm dysregulation imposed by modern lifestyle and the emerging evidence supporting its role in carcinogenesis, it is important to expand upon our findings and investigate potential underlying biological mechanisms that might inform on future interventions for CRC prevention.

Data availability

All data generated or analysed during this study are included in this published article (and its supplementary information files). Further information is available available from the corresponding author on reasonable request.

Received: 9 December 2022; Accepted: 16 December 2024

Published online: 18 April 2025

References

- Sung, H. et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* **71**, 209–249 (2021).
- IARC, I. A. of R. on C.-. *NIGHT SHIFT WORK. IARC Monographs on the Identification of Carcinogenic Hazards to Humans Volume 124*. vol. 124 (2020).
- Chen, Y. et al. Sleep duration and the risk of cancer: A systematic review and meta-analysis including dose-response relationship. *BMC Cancer* **18**, 1–13 (2018).
- Lin, C. L., Liu, T. C., Wang, Y. N., Chung, C. H. & Chien, W. C. The association between sleep disorders and the risk of colorectal cancer in patients: A population-based nested case-control study. *In Vivo (Brooklyn)*. **33**, 573–579 (2019).
- Jones, S. E. et al. Genome-wide association analyses of chronotype in 697,828 individuals provides insights into circadian rhythms. *Nat. Commun.* **10**, 343 (2019).
- Lane, J. M. et al. Biological and clinical insights from genetics of insomnia symptoms. *Nat. Genet.* **19**, 387–393 (2019).
- Jansen, P. R. et al. Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nat. Genet.* **51**, 394–403 (2019).
- Dashti, H. S. et al. Genome-wide association study identifies genetic loci for self-reported habitual sleep duration supported by accelerometer-derived estimates. *Nat. Commun.* **10**, 1–12 (2019).
- Richmond, R. C. et al. Investigating causal relations between sleep traits and risk of breast cancer in women: Mendelian randomisation study. *BMJ* **365**, 1–12 (2019).
- Hayes, B. L. et al. Do sex hormones confound or mediate the effect of chronotype on breast and prostate cancer? A Mendelian randomization study. *PLOS Genet.* **18**, e1009887 (2022).
- Kyrgiou, M. et al. Adiposity and cancer at major anatomical sites: umbrella review of the literature. *BMJ* <https://doi.org/10.1136/bmj.j477> (2017).
- Pearson-Stuttard, J. et al. Type 2 diabetes and cancer: An umbrella review of observational and mendelian randomization studies. *Cancer Epidemiol. Biomark. Prev.* **30**, 1218–1228 (2021).
- Ma, Y. et al. Integrative identification of genetic loci jointly influencing diabetes-related traits and sleep traits of insomnia, sleep duration, and chronotypes. *Biomedicines* **10**, 368 (2022).
- Williams, J. A. et al. A causal web between chronotype and metabolic health traits. *Genes (Basel)*. **12**, 1029 (2021).
- Lane, J. M. et al. Genome-wide association analysis identifies novel loci for chronotype in 100,420 individuals from the UK Biobank. *Nat. Commun.* **7**, 1–10 (2016).
- Titova, O. E. et al. Sleep duration and risk of overall and 22 site-specific cancers: A mendelian randomization study. *Int. J. Cancer* **148**, 914–920 (2021).
- Parent, M. É., El-Zein, M., Rousseau, M. C., Pintos, J. & Siemiatycki, J. Night work and the risk of cancer among men. *Am. J. Epidemiol.* **176**, 751–759 (2012).
- Ahn, Y. S., Jeong, K. S. & Kim, K. S. Cancer morbidity of professional emergency responders in Korea. *Am. J. Ind. Med.* **55**, 768–778 (2012).
- Tynes, T., Hannevik, M., Andersen, A. & Vistnes, A. I. Incidence female radio of breast and operators telegraph. *Cancer Causes Control* **7**, 197–204 (1996).
- Papantoniou, K. et al. Rotating night shift work and colorectal cancer risk in the nurses' health studies. *Int. J. Cancer* **143**, 2709–2717 (2018).
- Devore, E. E. et al. Rotating night shift work, sleep, and colorectal adenoma in women. *Int. J. Colorectal Dis.* **32**, 1013–1018 (2017).
- Lim, A. S. P. et al. Sex difference in daily rhythms of clock gene expression in the aged human cerebral cortex. *J. Biol. Rhythms* **28**, 117–129 (2013).
- Sudlow, C. et al. UK Biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, 1–10 (2015).
- Hu, Y. et al. GWAS of 89,283 individuals identifies genetic variants associated with self-reporting of being a morning person. *Nat. Commun.* **7**, 1–9 (2016).
- Krokstad, S. et al. Cohort profile: The HUNT study. *Norway. Int. J. Epidemiol.* **42**, 968–977 (2013).
- Gainer, V. S. et al. The biobank portal for partners personalized medicine: A query tool for working with consented biobank samples, genotypes, and phenotypes using i2b2. *J. Pers. Med.* **6**, 11 (2016).
- Gottlieb, D. J. et al. Novel loci associated with usual sleep duration: The CHARGE consortium genome-wide association study. *Mol. Psychiatr.* **20**, 1232–1239 (2015).
- Burgess, S. & Thompson, S. G. Avoiding bias from weak instruments in mendelian randomization studies. *Int. J. Epidemiol.* **40**, 755–764 (2011).
- Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
- Delaneau, O., Howie, B., Cox, A. J., Zagury, J. F. & Marchini, J. Haplotype estimation using sequencing reads. *Am. J. Hum. Genet.* **93**, 687–696 (2013).
- McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
- Brion, M. J. A., Shakhbazov, K. & Visscher, P. M. Calculating statistical power in mendelian randomization studies. *Int. J. Epidemiol.* **42**, 1497–1501 (2013).
- Hartwig, F. P., Davies, N. M., Hemani, G. & Davey Smith, G. Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique. *Int. J. Epidemiol.* **45**, 1717–1726 (2016).
- Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37**, 658–665 (2013).
- Bowden, J. et al. A framework for the investigation of pleiotropy in two-sample summary data mendelian randomization. *Stat. Med.* **36**, 1783–1802 (2017).
- Hemani, G. et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7**, 1–29 (2018).
- Lawlor, D. A., Harbord, R. M., Sterne, J. A. C., Timpson, N. & Davey Smith, G. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Stat. Med.* **27**, 1133–1163 (2008).
- Staiger, D. & Stock, J. H. Instrumental variables regression with weak instruments. *Econometrica* **65**, 557 (1997).
- Bowden, J., Smith, G. D. & Burgess, S. Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).
- Bowden, J. et al. Assessing the suitability of summary data for two-sample mendelian randomization analyses using MR-Egger regression: The role of the I² statistic. *Int. J. Epidemiol.* **45**, 1961–1974 (2016).

41. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet. Epidemiol.* **40**, 304–314 (2016).
42. Hartwig, F. P., Smith, G. D. & Bowden, J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int. J. Epidemiol.* **46**, 1985–1998 (2017).
43. Burgess, S., Foley, C. N., Allara, E., Staley, J. R. & Howson, J. M. A robust and efficient method for Mendelian randomization with hundreds of genetic variants: unravelling mechanisms linking HDL-cholesterol and coronary heart disease. *BioRxiv* <https://doi.org/10.1101/566851> (2019).
44. Verbanck, M., Chen, C. Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* **50**, 693–698 (2018).
45. Yavorska, O. O. & Burgess, S. MendelianRandomization: An R package for performing mendelian randomization analyses using summarized data. *Int. J. Epidemiol.* **46**, 1734–1739 (2017).
46. Jones, S. E. et al. Genetic studies of accelerometer-based sleep measures yield new insights into human sleep behaviour. *Nat. Commun.* **10**, 1–12 (2019).
47. Burgess, S. & Thompson, S. G. Multivariable mendelian randomization: The use of pleiotropic genetic variants to estimate causal effects. *Am. J. Epidemiol.* **181**, 251–260 (2015).
48. Sanderson, E., Spiller, W. & Bowden, J. Testing and correcting for weak and pleiotropic instruments in two-sample multivariable mendelian randomisation. *BioRxiv* <https://doi.org/10.1101/2020.04.02.021980> (2020).
49. Sanderson, E., Davey Smith, G., Windmeijer, F. & Bowden, J. An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. *Int. J. Epidemiol.* **48**, 713–727 (2019).
50. Straif, K. et al. Carcinogenicity of shift-work, painting, and fire-fighting. *Lancet Oncol.* **8**, 1065–1066 (2007).
51. Schernhammer, E. S. et al. Night-shift work and risk of colorectal cancer in the Nurses' Health Study. *J. Natl. Cancer Inst.* **95**, 825–828 (2003).
52. Vetter, C. et al. Mismatch of sleep and work timing and risk of type 2 diabetes. *Diabetes Care* **38**, 1707–1713 (2015).
53. United States Department of Labor. Workers on flexible and shift schedule in May 2004. 1–16 (2005).
54. Cain, S. W. et al. Sex differences in phase angle of entrainment and melatonin amplitude in humans. *J. Biol. Rhythms* **25**, 288–296 (2010).
55. Hurley, S., Goldberg, D., Bernstein, L. & Reynolds, P. Sleep duration and cancer risk in women. *Cancer Causes Control* **26**, 1037–1045 (2015).
56. Gu, F. et al. Sleep duration and cancer in the NIH-AARP diet and health study cohort. *PLoS One* **11**, 1–14 (2016).
57. Lin, Y. et al. Associations of dinner-to-bed time, post-dinner walk and sleep duration with colorectal cancer A case-control study. *Med. (United States)* **97**, 1–7 (2018).
58. Thompson, C. L. et al. Short duration of sleep increases risk of colorectal adenoma. *Cancer* **117**, 841–847 (2011).
59. Papantoniou, K. et al. Sleep duration and napping in relation to colorectal and gastric cancer in the MCC-Spain study. *Sci. Rep.* **11**, 1–14 (2021).
60. Jiao, L. et al. Sleep duration and incidence of colorectal cancer in postmenopausal women. *Br. J. Cancer* **108**, 213–221 (2013).
61. Zhang, X. et al. Associations of Self-Reported Sleep Duration and Snoring with Colorectal Cancer Risk in Men and Women. *Sleep* **36**, 681–688 (2013).
62. Lauderdale, D. S., Knutson, K. L., Yan, L. L., Liu, K. & Rathouz, P. J. Self-reported and measured sleep duration: How similar are they?. *Epidemiology* **19**, 838–845 (2008).
63. Burgess, S., Davies, N. M. & Thompson, S. G. Bias due to participant overlap in two-sample mendelian randomization. *Genet. Epidemiol.* **40**, 597–608 (2016).

Acknowledgments

ASTERISK: We are very grateful to Dr. Bruno Buecher without whom this project would not have existed. We also thank all those who agreed to participate in this study, including the patients and the healthy control persons, as well as all the physicians, technicians and students. CCFR: The Colon CFR graciously thanks the generous contributions of their 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. COLON and NQplus: the authors would like to thank the COLON and NQplus investigators at Wageningen University & Research and the involved clinicians in the participating hospitals. CORSA: We kindly thank all individuals who agreed to participate in the CORSA study. Furthermore, we thank all cooperating physicians and students and the Biobank Graz of the Medical University of Graz. CPS-II: The authors thank the CPS-II participants and Study Management Group for their invaluable contributions to this research. The authors would also like to acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention National Program of Cancer Registries, and cancer registries supported by the National Cancer Institute Surveillance Epidemiology and End Results program. Czech Republic CCS: We are thankful to all clinicians in major hospitals in the Czech Republic, without whom the study would not be practicable. We are also sincerely grateful to all patients participating in this study.

DACHS: We thank all participants and cooperating clinicians, and everyone who provided excellent technical assistance. EDRN: We acknowledge all contributors to the development of the resource at University of Pittsburgh School of Medicine, Department of Gastroenterology, Department of Pathology, Hepatology and Nutrition and Biomedical Informatics. EPIC: Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer/World Health Organization. EPICOLON: We are sincerely grateful to all patients participating in this study who were recruited as part of the EPICOLON project. We acknowledge the Spanish National DNA Bank, Biobank of Hospital Clínic-IDIBAPS and Biobanco Vasco for the availability of the samples. The work was carried out (in part) at the Esther Koplowitz Centre, Barcelona. Harvard cohorts (HPFS, NHS, PHS): The study protocol was approved by the institutional review boards of the Brigham and Women's Hospital and Harvard T.H. Chan School of Public Health, and those of participating registries as required. We acknowledge Channing Division of Network Medicine, Department of Medicine, Brigham and

Women's Hospital as home of the NHS. The authors would like to acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention's National Program of Cancer Registries (NPCR) and/or the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program. Central registries may also be supported by state agencies, universities, and cancer centers. Participating central cancer registries include the following: Alabama, Alaska, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, Florida, Georgia, Hawaii, Idaho, Indiana, Iowa, Kentucky, Louisiana, Massachusetts, Maine, Maryland, Michigan, Mississippi, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Puerto Rico, Rhode Island, Seattle SEER Registry, South Carolina, Tennessee, Texas, Utah, Virginia, West Virginia, Wyoming. The authors assume full responsibility for analyses and interpretation of these data. Kentucky: We would like to acknowledge the staff at the Kentucky Cancer Registry. LCCS: We acknowledge the contributions of Jennifer Barrett, Robin Waxman, Gillian Smith and Emma Northwood in conducting this study. NCCCS I & II: We would like to thank the study participants, and the NC Colorectal Cancer Study staff. NSHDS investigators thank the 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. SEARCH: We thank the SEARCH team. SELECT: We thank the research and clinical staff at the sites that participated on SELECT study, without whom the trial would not have been successful. We are also grateful to the 35,533 dedicated men who participated in SELECT. WHI: The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: <http://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Short%20List.pdf>

Author contributions

K.K.T., Y.C. contributed to the study conception. K.K.T., Y.C. and R.C.R. designed the study. O.D. and H.F. carried out the data extraction, manipulation and analysis. E.B., R.C.R. and B.H. provided access to data for the association of genetic variants with the exposures of interest. H.B., A.G., L.L.M., V.M., R.K.P., A.I.P., C.Y.U., F.J.B.V.D., P.V. and U.P. provided access to colorectal cancer data from the GECCO, CORECT and CCFR consortia. O.D. drafted the manuscript under the supervision of K.K.T. All authors critically reviewed previous versions of the manuscript and assisted with the interpretation of data. All authors read and approved the content and presentation of the final article and agreed to be accountable for all aspects of this work. The work reported in the paper has been performed by the authors, unless clearly specified in the text.

Funding

This work was supported by Cancer Research UK (grant number C18281/A29019). RMM is a National Institute for Health Research Senior Investigator (NIHR202411). RMM is supported by a Cancer Research UK 25 (C18281/A29019) programme grant (the Integrative Cancer Epidemiology Programme). RMM is also supported by the NIHR Bristol Biomedical Research Centre which is funded by the NIHR (BRC-1215-20,011) and is a partnership between University Hospitals Bristol and Weston NHS Foundation Trust and the University of Bristol. RMM is affiliated with the Medical Research Council Integrative Epidemiology Unit at the University of Bristol which is supported by the Medical Research Council (MC_UU_00011/1, MC_UU_00011/3, MC_UU_00011/6, and MC_UU_00011/4) and the University of Bristol. Department of Health and Social Care disclaimer: The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. 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. 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, R21 CA191312, R01201407). Genotyping/Sequencing services were provided by the Center for Inherited Disease Research (CIDR) contract number HHSN268201700006I and 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. ASTERISK: a Hospital Clinical Research Program (PHRC-BRD09/C) from the University Hospital Center of Nantes (CHU de Nantes) and supported by the Regional Council of Pays de la Loire, the Groupement des Entreprises Françaises dans la Lutte contre le Cancer (GEFLUC), the Association Anne de Bretagne Génétique and the Ligue Régionale Contre le Cancer (LRCC). The ATBC Study is supported by the Intramural Research Program of the U.S. National Cancer Institute, National Institutes of Health, Department of Health and Human Services. CLUE II funding was from the National Cancer Institute (U01 CA86308, Early Detection Research Network; P30 CA006973), National Institute on Aging (U01 AG18033), and the American

Institute for Cancer Research. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US government. Maryland Cancer Registry (MCR) Cancer data was 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. ColoCare: This work was supported by the National Institutes of Health (grant numbers R01 CA189184 (Li/Ulrich), U01 CA206110 (Ulrich/Li/Siegel/Figueiredo/Colditz, 2P30CA015704-40 (Gilliland), R01 CA207371 (Ulrich/Li)), the Matthias Lackas-Foundation, the German Consortium for Translational Cancer Research, and the EU TRANSCAN initiative. The Colon Cancer Family Registry (CCFR, www.coloncfr.org) is supported in part by funding from the National Cancer Institute (NCI), National Institutes of Health (NIH) (award U01 CA167551). 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 1 M/1 M-Duo) and Set-2 (Illumina Omni1-Quad) scans were supported by NIH awards U01 CA122839 and R01 CA143247 (to GC). The CCFR Set-3 (Affymetrix Axiom CORECT Set array) was supported by NIH award U19 CA148107 and R01 CA81488 (to SBG). The CCFR Set-4 (Illumina OncoArray 600 K 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 112,746 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 Human-CytoSNP 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. COLON: The COLON study is sponsored by Wereld Kanker Onderzoek Fonds, including funds from grant 2014/1179 as part of the World Cancer Research Fund International Regular Grant Programme, by Alpe d'Huzes and the Dutch Cancer Society (UM 2012–5653, UW 2013-5927, UW2015-7946), and by TRANSCAN (JTC2012-MetaboCCC, JTC2013-FOCUS). The Nqplus study is sponsored by a ZonMW investment grant (98-10,030); by PREVIEW, the project PREvention of diabetes through lifestyle intervention and population studies in Europe and around the World (PREVIEW) project which received funding from the European Union Seventh Framework Programme (FP7/2007–2013) under grant no. 312057; by funds from TI Food and Nutrition (cardiovascular health theme), a public-private partnership on precompetitive research in food and nutrition; and by FOODBALL, the Food Biomarker Alliance, a project from JPI Healthy Diet for a Healthy Life. Colorectal Cancer Transdisciplinary (CORECT) Study: The CORECT Study was supported by the National Cancer Institute, National Institutes of Health (NCI/NIH), U.S. Department of Health and Human Services (grant numbers U19 CA148107, R01 CA081488, P30 CA014089, R01 CA197350; P01 CA196569; R01 CA201407; R01 CA242218), National Institutes of Environmental Health Sciences, National Institutes of Health (grant number T32 ES013678) and a generous gift from Daniel and Maryann Fong. CORSA: The CORSA study was funded by Austrian Research Funding Agency (FFG) BRIDGE (grant 829,675, to Andrea Gsur), the “Herzfelder’sche Familienstiftung” (grant to Andrea Gsur) and was supported by COST Action BM1206. CPS-II: The American Cancer Society funds the creation, maintenance, and updating of the Cancer Prevention Study-II (CPS-II) cohort. This study was conducted with Institutional Review Board approval. CRCGEN: Colorectal Cancer Genetics & Genomics, Spanish study was supported by Instituto de Salud Carlos III, co-funded by FEDER funds –a way to build Europe– (grants PI14-613 and PI09-1286), Agency for Management of University and Research Grants (AGAUR) of the Catalan Government (grant 2017SGR723), Junta de Castilla y León (grant LE22A10-2), the Spanish Association Against Cancer (AECC) Scientific Foundation grant GCTRA18022MORE and the Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), action Genrisk. Sample collection of this work was supported by the Xarxa de Bancs de Tumors de Catalunya sponsored by Pla Director d’Oncologia de Catalunya (XBTC), Plataforma Biobancos PT13/0010/0013 and ICOBIOBANC, sponsored by the Catalan Institute of Oncology. We thank CERCA Programme, Generalitat de Catalunya for institutional support. Czech Republic CCS: This work was supported by the Grant Agency of the Czech Republic (18-09709S, 20-03997S), by the Grant Agency of the Ministry of Health of the Czech Republic (grants AZV NV18/03/00,199 and AZV NV19-09-00,237), and Charles University grants Unce/Med/006 and Progress Q28/LF1. DACHS: This work was supported by the German Research Council (BR 1704/6-1, BR 1704/6-3, BR 1704/6-4, CH 117/1-1, HO 5117/2-1, HE 5998/2-1, KL 2354/3-1, RO 2270/8-1 and BR 1704/17-1), the Interdisciplinary Research Program of the National Center for Tumor Diseases (NCT), Germany, and the German Federal Ministry of Education and Research (01KH0404, 01ER0814, 01ER0815, 01ER1505A and 01ER1505B). DALs: National Institutes of Health (R01 CA48998 to M. L. Slattery). EDRN: This work is funded and supported by the NCI, EDRN Grant (U01 CA 84,968-06). EPIC: The coordination of EPIC is financially supported by International Agency for Research on Cancer (IARC) and also by the Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London which has additional infrastructure support provided by the NIHR Imperial Biomedical Research Centre (BRC). The national cohorts are supported by: Danish Cancer Society (Denmark); Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l’Éducation 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 (14,136 to EPIC-Norfolk; C8221/A29017 to EPIC-Oxford), Medical Research Council (1,000,143 to EPIC-Norfolk; MR/M012190/1 to EPIC-Oxford). (United Kingdom). EPICOLON: This work was supported by grants from Fondo de Investigación Sanitaria/FEDER (PI08/0024, PI08/1276, PS09/02,368, P111/00,219, P111/00,681, P114/00,173, P114/00,230, P117/00,509, 17/00,878, PI20/00,113, PI20/00,226, Acción Transversal de Cáncer), Xunta de Galicia (PGIDIT07PXIB9101209PR), Ministerio de Economía y Competitividad (SAF07-64,873, SAF 2010-19,273, SAF2014-54453R), Fundación Científica de la Asociación Española contra el Cáncer (GCB13131592CAST), Beca Grupo de Trabajo “Oncología” AEG (Asociación Española de Gastroenterología), Fundación Privada Olga Torres, FP7 CHIBCHA Consortium, Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR, Generalitat de Catalunya, 2014SGR135, 2014SGR255, 2017SGR21, 2017SGR653), Catalan Tumour Bank Network (Pla Director d'Oncologia, Generalitat de Catalunya), PERIS (SLT002/16/00,398, Generalitat de Catalunya), CERCA Programme (Generalitat de Catalunya) and COST Action BM1206 and CA17118. CIBERehd is funded by the Instituto de Salud Carlos III. ESTHER/VERDI. This work was supported by grants from the Baden-Württemberg Ministry of Science, Research and Arts and the German Cancer Aid. Harvard cohorts: HPFS is supported by the National Institutes of Health (P01 CA055075, UM1 CA167552, U01 CA167552, R01 CA137178, R01 CA151993, and R35 CA197735), NHS by the National Institutes of Health (P01 CA087969, UM1 CA186107, R01 CA137178, R01 CA151993, and R35 CA197735), and PHS by the National Institutes of Health (R01 CA042182). Hawaii Adenoma Study: NCI grants R01 CA72520. HCES-CRC: the Hwasun Cancer Epidemiology Study—Colon and Rectum Cancer (HCES-CRC; grants from Chonnam National University Hwasun Hospital, HCRI15011-1). Kentucky: This work was supported by the following grant support: Clinical Investigator Award from Damon Runyon Cancer Research Foundation (CI-8); NCI R01CA136726. LCCS: The Leeds Colorectal Cancer Study was funded by the Food Standards Agency and Cancer Research UK Programme Award (C588/A19167). MCCC cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCC was further supported by Australian NHMRC grants 509,348, 209,057, 251,553 and 504,711 and by infrastructure provided by Cancer Council Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry (VCR) and the Australian Institute of Health and Welfare (AIHW), including the National Death Index and the Australian Cancer Database. MEC: National Institutes of Health (R37 CA54281, P01 CA033619, and R01 CA063464). MECC: This work was supported by the National Institutes of Health, U.S. Department of Health and Human Services (R01 CA081488, R01 CA197350, U19 CA148107, R01 CA242218, and a generous gift from Daniel and Maryann Fong. 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). NCCCS I & II: We acknowledge funding support for this project from the National Institutes of Health, R01 CA66635 and P30 DK034987. NFCCR: This work was supported by an Interdisciplinary Health Research Team award from the Canadian Institutes of Health Research (CRT 43,821); the National Institutes of Health, U.S. Department of Health and Human Services (U01 CA74783); and National Cancer Institute of Canada grants (18,223 and 18,226). 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 2017-00,650, VR 2017-01,737), 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. OSUMC: OC-CPI funding was provided by Pelotonia and HNPCC funding was provided by the NCI (CA16058 and CA67941). PLCO: Intramural Research Program of the Division of Cancer Epidemiology and Genetics and supported by contracts from the Division of Cancer Prevention, National Cancer Institute, NIH, DHHS. Funding was provided by National Institutes of Health (NIH), Genes, Environment and Health Initiative (GEI) Z01 CP 010,200, NIH U01 HG004446, and NIH GEI U01 HG 004,438. SEARCH: The University of Cambridge has received salary support in respect of PDPP from the NHS in the East of England through the Clinical Academic Reserve. Cancer Research UK (C490/A16561); the UK National Institute for Health Research Biomedical Research Centres at the University of Cambridge. SELECT: Research reported in this publication was supported in part by the National Cancer Institute of the National Institutes of Health under Award Numbers U10 CA37429 (CD Blanke), and UM1 CA182883 (CM Tangen/IM Thompson). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. SMS and REACH: This work was supported by the National Cancer Institute (grant P01 CA074184 to J.D.P. and P.A.N., grants R01 CA097325, R03 CA153323, and K05 CA152715 to P.A.N., and the National Center for Advancing Translational Sciences at the National Institutes of Health (grant KL2 TR000421 to A.N.B.-H.). The Swedish Low-risk Colorectal Cancer Study: The study was supported by grants from the Swedish research council; K2015-55X-22674-01-4, K2008-

55X-20157-03-3, K2006-72X-20157-01-2 and the Stockholm County Council (ALF project). Swedish Mammography Cohort and Cohort of Swedish Men: This work is supported by 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.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-83693-w>.

Correspondence and requests for materials should be addressed to O.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024