



**ARTICLE**

Genetics and Genomics

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

Sonja Neumeier et al.

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

**METHODS:** We used a Mendelian randomisation approach to test for a causal effect of age at menarche and age at menopause as surrogates for endogenous oestrogen 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.

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

*British Journal of Cancer* (2018) 118:1639–1647; <https://doi.org/10.1038/s41416-018-0108-8>

**INTRODUCTION**

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

Epidemiologic studies examining reproductive factors such as age at menarche and age at menopause with CRC risk have reported inconsistent results.<sup>5–7</sup> A meta-analysis summarising evidence based on 22 studies published by Li et al.<sup>8</sup> did not find a significant association between age at menarche and CRC risk. 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.<sup>9</sup> 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 randomisation (MR) approach<sup>10</sup> 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.<sup>11, 12</sup> 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 oestrogen reduces risk of CRC.<sup>3, 4, 13</sup> Estrogen exerts its effects in colon cells predominantly through the nuclear receptor oestrogen-receptor  $\beta$  (ER $\beta$ ),<sup>14, 15</sup> which has mainly anti-proliferative effects<sup>16</sup> and its expression is inversely related to cancer stage, tumour extent, and mortality.<sup>17, 18</sup> Longer use of MHT appears to be associated with high ER $\beta$  expression in tumours.<sup>19</sup> If similar mechanisms hold then longer endogenous oestrogen exposure (through earlier age at menarche and/or later age at menopause) would reduce CRC risk as well. To test this hypothesis, we conducted an 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.

Correspondence: Jenny Chang-Claude (j.chang-claude@dkfz-heidelberg.de)  
Extended author information available on the last page of the article

Received: 7 December 2017 Revised: 9 April 2018 Accepted: 11 April 2018  
Published online: 24 May 2018

## 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)<sup>20</sup> (5386 cases and 5696 controls), the Colon Cancer Family Registry (CCFR)<sup>21</sup> (1678 cases and 1188 controls) and the Colorectal Cancer Transdisciplinary (CORECT) Consortium<sup>22</sup> (5880 cases and 3857 controls) (see Supplementary Tables 1 and 2). Different centres of CCFR participated as individual studies in GECCO and/or CORECT, and therefore were analysed 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 harmonisation was done centrally as previously described.<sup>23</sup> Self-reported data for age at menarche and age at menopause have not been harmonised 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 ( $P < 1 \times 10^{-4}$ ), or low minor allele frequency ( $\leq 1\%$ ). 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<sup>24</sup> and 54 SNPs associated with age at natural menopause<sup>25</sup> at a genome-wide significance level (i.e.,  $P < 5 \times 10^{-8}$ ). 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 harmonised 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 ( $\beta$ ) for association with age at menarche and age at menopause.<sup>24, 25</sup>

Age at menarche:

$$GRS_K = \sum_{n=1}^{358} \beta_n G_{Kn}.$$

Age at menopause:

$$GRS_K = \sum_{n=1}^{51} \beta_n G_{Kn}.$$

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 (1492 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 oestrogen was calculated as the sum of the scaled risk scores for age at menarche and age at menopause.

$$GRS_{\text{time period}} = GRS_{\text{age at menarche}} + GRS_{\text{age at menopause}}.$$

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 oestrogen exposure reflects higher risk for longer exposure to endogenous oestrogen. 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,<sup>24, 25</sup>  $K$  indicates the number of instrumental variables, and  $n$  stands for the sample size.<sup>26</sup> An  $F$ -statistic  $> 10$  suggests that the genetic instrument is sufficiently strong.<sup>27</sup> 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), MHT (oestrogen/progestin combined and oestrogen alone), using linear regression for continuous variables and logistic/multinomial logistic regression for categorical variables in a subset of the studies with available data ( $n = 6285$  controls). To address the third assumption of MR, i.e. to assess the presence of pleiotropy, we applied the MR-Egger method.<sup>28</sup> 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).<sup>29, 30</sup> 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,<sup>9, 31</sup> stratified analyses by MHT were performed (only in GECCO where harmonised data were available). Additional stratified analyses were performed according to menopausal status (for age at menarche), combined menopausal oestrogen or progesterone therapy, oestrogen therapy alone, and BMI categories in kg/m<sup>2</sup> (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,<sup>32, 33</sup> we also conducted site-specific analyses for 4037 female colon cancer cases and 1184 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 at menopause, which corresponds to the variance in age at menarche/age at menopause explained by the SNPs used for this analyses.<sup>34</sup>

It is known that BMI in childhood is strongly associated with age at menarche<sup>35</sup> and that some SNPs associated with age at menarche could also have pleiotropic effects and therefore are related to BMI as well.<sup>24</sup> To address this issue and thereby also account for violations of the third MR assumption, we conducted further BMI-specific sensitivity analyses. We adjusted for BMI in the logistic regression analysis using a weighted GRS for BMI comprising 77 SNPs previously reported to be associated with BMI at a genome-wide significance level in European subjects.<sup>36</sup> For the second sensitivity analysis, we identified age at menarche SNPs showing pleiotropy by testing the association of these SNPs with BMI in a subset of our sample ( $n = 5832$  cases/6285 controls) and found 29 SNPs to be associated with BMI at nominal significance ( $P$ -value < 0.05). Two further SNPs overlapped with reported BMI-SNPs<sup>36</sup> and 11 more SNPs were in high linkage disequilibrium ( $R^2 > 0.1$ ) with BMI SNPs. A restricted GRS for age at menarche excluding the 42 BMI-associated SNPs ( $n = 316$  SNPs) was constructed and then assessed for association with CRC risk. For the analysis of lifetime oestrogen exposure we also generated a restricted risk score excluding the same 42 BMI-associated SNPs.

**Two-sample MR analyses.** We performed two-sample MR analyses as sensitivity analyses using published summary statistics for SNP–exposure associations (age at menarche<sup>24</sup>/age at menopause<sup>25</sup>); SNP outcome associations were estimated in GECCO/CORECT (Supplementary Tables 8 and 9 for age at menarche/menopause-SNPs, respectively). We applied the weighted median estimator approach, which is robust against violations due to pleiotropic SNPs even when up to 50% of the genetic instruments are invalid.<sup>37</sup> For this approach, we used SNP–exposure and SNP–outcome associations to build ratio estimates for each SNP. These estimates were ordered and weighted by the inverse of their variance. Bootstrapped standard errors were 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.<sup>28</sup>

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 Randomisation”<sup>38</sup> 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 = 1755,  $R^2 = 0.069$ <sup>24</sup>;  $F$ -statistic for age at menopause = 1431,  $R^2 = 0.057$ <sup>25</sup>). 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, oestrogen/progestin therapy, oestrogen 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 oestrogen 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 oestrogen 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 oestrogen 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: 0.93–1.01) for lifetime oestrogen exposure (details of results per GECCO and CORECT consortium in Supplementary Table 7). The intercept term from MR-Egger regression was centred at the origin for age at menarche (intercept term  $-0.0008$ , 95% CI:  $-0.007$  to  $0.005$ ,  $P$ -value 0.80) and age at menopause (intercept term  $-0.009$ , 95% CI:  $-0.023$  to  $0.006$ ,  $P$ -value 0.23), suggesting absence of strong directional pleiotropy (Supplementary Figures 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 Figures 3 and 4). Table 2 shows stratified analyses for the association of GRS for age at menarche and CRC risk by menopausal status, combined oestrogen/progesterone therapy, and oestrogen 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 oestrogen/progesterone therapy, oestrogen 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

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

Variable (per year)	Cases/Controls	GRS-based analyses <sup>b</sup>						2-sample MR <sup>f</sup>			
		MR estimate <sup>a</sup>		Adjusted by GRS-BMI <sup>a,c</sup>		Restricted risk score <sup>a,d</sup>		MR-Egger <sup>b,e</sup>		Weighted median estimator <sup>b</sup>	
		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

*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 randomisation, *NA* not applicable, *OR* odds ratio per year, *SNPs* single-nucleotide polymorphisms, *se* standard error. <sup>a</sup>Logistic regression model adjusted for age, study, and principal components of genetic ancestry. <sup>b</sup>Meta-analysed estimate of GECCO/CORECT datasets (CCFR centres participated in GECCO or CORECT and were analysed as such). <sup>c</sup>Additionally adjusted for a GRS for BMI out of 77 reported SNPs for BMI<sup>36</sup>. <sup>d</sup>42 BMI-associated SNPs were excluded from the age at menarche and time period of oestrogen exposure—risk scores. <sup>e</sup>Estimate derived from the slope of MR-Egger. <sup>f</sup>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 <sup>a</sup>	95% CI	P-value	Heterogeneity P-value <sup>b</sup>
All	5832/6285	0.97	0.92–1.02	0.20	
<b>Menopausal status</b>					
Premenopausal	545/621	0.95	0.80–1.12	0.54	0.94
Postmenopausal	5263/5647	0.97	0.92–1.02	0.23	
<b>Menopausal hormone therapy combined</b>					
No	3266/3490	0.96	0.90–1.03	0.24	0.69
Yes	593/807	0.98	0.85–1.14	0.82	
<b>Oestrogen monotherapy</b>					
No	3120/3214	0.94	0.88–1.01	0.11	0.17
Yes	716/1088	1.04	0.91–1.19	0.55	
<b>Site</b>					
Colon	4037/6285	0.97	0.91–1.02	0.26	0.25 <sup>c</sup>
Rectum	1184/6285	1.04	0.95–1.14	0.36	

*CCFR* Colon Cancer Family Registry, *CI* confidence interval, *GRS* genetic risk score, *OR* odds ratio per year. <sup>a</sup>All analyses adjusted for age, sex, study, and principal components of genetic ancestry. <sup>b</sup>P-value calculated using likelihood ratio tests comparing the model with and without interaction term. <sup>c</sup>P-value for heterogeneity was obtained in case-only analysis of colon vs. rectal cancer

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 oestrogen exposure. We investigated the association of GRS for age at menarche and age at menopause as surrogates for endogenous oestrogen exposure on CRC risk. Our

**Table 3.** Association of genetically predicted age at menopause and CRC risk according to subgroups, CCFR, GECCO Consortium

	N (cases/controls)	OR <sup>a</sup>	95% CI	P-value	Heterogeneity P-value <sup>b</sup>
All	5832/6285	0.99	0.95–1.03	0.56	
<b>Combined oestrogen–progesterone therapy</b>					
No	3266/3490	0.99	0.93–1.05	0.78	0.75
Yes	593/807	0.96	0.85–1.10	0.58	
<b>Oestrogen monotherapy</b>					
No	3120/3214	0.99	0.93–1.05	0.27	0.76
Yes	716/1088	1.01	0.89–1.14	0.90	
<b>BMI</b>					
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	
<b>Site</b>					
Colon	4037/6285	0.98	0.94–1.03	0.53	0.61 <sup>c</sup>
Rectum	1184/6285	0.98	0.91–1.06	0.65	

*BMI* body mass index, *CCFR* Colon Cancer Family Registry, *CI* confidence interval, *GRS* genetic risk score, *OR* odds ratio per year. <sup>a</sup>All analyses adjusted for age, sex, study, and principal components of genetic ancestry. <sup>b</sup>P-value calculated using likelihood ratio tests comparing the model with and without interaction term. <sup>c</sup>P-value for heterogeneity was obtained in case-only analysis of colon vs. rectal cancer

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<sup>6, 8, 39</sup> between self-reported age at menarche and CRC risk, although some studies found an inverse association.<sup>5, 9</sup> 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.<sup>9</sup> 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 and Murphy et al.<sup>31</sup> found a hazard ratio of 0.72 (95% CI: 0.54–0.96) for age at menarche (>15 vs. 11–12 years). We therefore assessed the association of GRS for age at menarche with CRC risk stratified by ever/never use of MHT, separately for combined oestrogen–progesterone therapy and for oestrogen monotherapy. No difference in the association according to either combined oestrogen–progesterone therapy or oestrogen-alone therapy was found.

Higher BMI in childhood is associated with earlier age at menarche<sup>24, 35</sup> and also with a higher risk for CRC.<sup>40</sup> 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<sup>41</sup> and also a high concordance between adolescent and adult BMI-SNPs,<sup>42</sup> 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.<sup>24</sup> found age at menarche variants that appear unrelated to BMI at a nominal level in their sample to be still BMI associated collectively ( $P = 4.2 \times 10^{-9}$ ). 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 ( $\geq 55$  vs.  $< 40$ ; HR 1.50; 95% CI: 1.23–1.83),<sup>9</sup> whereas most other studies reported null associations.<sup>6, 31, 39</sup> 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 oestrogen 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 ( $\geq 36$  years vs.  $\leq 30$  years) and CRC risk in a prospective observational study conducted in Japan<sup>43</sup>. The GRS for lifetime for endogenous oestrogen exposure, which we constructed, does not account for other factors like parity or breast feeding, which influence overall oestrogen exposure; however, these factors have not been associated with CRC risk.

Results of prospective studies that investigated the association between serum levels of endogenous oestrogens and CRC risk have also been inconsistent. One study reported a positive association between circulating estradiol and CRC risk,<sup>44</sup> another study observed an inverse relationship<sup>45</sup> while most other studies found no associations.<sup>46–48</sup> There are reports that earlier age at menarche is associated with higher oestrogen levels.<sup>49, 50</sup> So oestrogen 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 oestrogen exposure by MHT mainly in the form of combined oestrogen–progestogen was associated with a reduced risk for CRC.<sup>13, 51</sup> The Womens's Health Initiative Clinical Trial (WHI-CT) reported no effect of oestrogen-alone therapy,<sup>52, 53</sup> and a

significant risk reduction for the association of oestrogen plus progestin vs. placebo and CRC risk,<sup>33</sup> which was suggested to have resulted from diagnostic delay instead of true risk reduction<sup>54</sup>. However, a recent meta-analysis which summarised results of four clinical trials including WHI-CT and 16 observational studies concluded that there is consistent evidence to support a protective effect of MHT on CRC risk.<sup>4</sup>

Thus, it appears that exogenous and endogenous oestrogens, which also vary in absolute amount of oestrogen, may play different roles in the development of CRC, presumably by different mechanisms, which are not well understood. Oestrogen acts in colon cells predominantly through ER $\beta$ ,<sup>14, 15</sup> which exerts proapoptotic and anti-proliferative effects in the colon<sup>16</sup> and its expression is reduced in tumour tissue.<sup>17, 18</sup> According to an *in vivo* study, oestrogen treatment was associated with an increase in expression of ER $\beta$  in colon tissue,<sup>55</sup> supporting a mechanism by which MHT may affect CRC risk. There is also some evidence that the protective effect of MHT on CRC risk may vary by the expression status of ER $\beta$ . Two studies found that the magnitude of risk reduction by MHT was different between colorectal tumours with higher and with lower expression of ER $\beta$ .<sup>56, 57</sup> It is possible that the effect of endogenous oestrogens on CRC risk may be modulated by ER $\beta$  expression as well. Therefore, larger studies with data on expression of ER $\beta$  in colon tissue are warranted to assess whether the association of age at menarche/age at menopause and CRC risk differs by ER $\beta$  expression status.

In this MR study, we aimed to use proxies for start and endpoint of endogenous oestrogen exposure in women, specifically age at menarche/age at menopause, which themselves are complex traits influenced by many variants with only small effects on the trait. Although we did not see large pleiotropic effects using Egger regression, there might have been residual pleiotropy, which is difficult to exclude. Previous studies of age at menarche performed LD score regression using 123 SNPs associated with age at menarche and found, amongst others, genetic correlations with BMI, adult height, or type 2 diabetes. Residual pleiotropy related to adult height or type 2 diabetes, which have been reported to be associated with higher risk for CRC as well, cannot be fully excluded.<sup>58</sup> In addition, Day et al. reported genetic correlations for the 54 age at menopause SNPs with adult obesity and other growth-related traits. The top menopause-SNPs were also associated with fasting glucose and were enriched in DNA repair pathways, yielding further sources of residual pleiotropy.<sup>25</sup> Residual pleiotropy is a general limitation of MR, especially when exploring complex traits. When considering the recently published hypothesis of an omnigenic model of complex traits, coined “network pleiotropy”, essentially any regulatory variant in a trait-relevant cell-type can have some effect on the trait.<sup>59</sup> This is because specific cell-types have specific regulatory networks, where any single variant could affect trait relevant genes, “core genes”, mediated through the same regulatory networks. So also for GWAS findings, it is highly likely that some genetic variants exhibit horizontal pleiotropy.<sup>60</sup> As the selected age at menarche/age at menopause SNPs might also contribute tiny effects on further traits through network-pleiotropy, we cannot fully rule out pleiotropy. These limitations should be kept in mind and methods to explore the impact of such effects should be developed. That said, our sensitivity analyses especially Egger regression, did not indicate large pleiotropic effects.

The second assumption of MR is that the IV is not associated with confounding factors of the observational association between age at menarche/menopause and CRC risk. We were able to exclude most risk factors for CRC (smoking, family history of CRC, education, aspirin use, MHT) as confounding variables except for BMI, which was accounted for using several sensitivity analyses. In addition, substantial overlap between datasets used for estimating SNP-exposure and SNP-outcome associations

would bias results in the direction of the observational estimate. There was some overlap between the studies (see Supplementary Information on participant overlap) but unlikely to have substantially influenced the results.

Strengths of our study include the large sample size, the availability of centrally harmonised data, and the robustness of the instrumental variables. Power calculations showed that our 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.

## 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 Benschaid, 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 harmonising 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 Lejbkovicz, 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%20Write%20a%20Paper/WHI%20Investigator%20Short%20List.pdf>. This work was funded by 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 G.C.). 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.G.). 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). DAL: National Institutes of Health (R01 CA48998 to M.L.S.); 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. G.). 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 was 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):645–649), 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):986–990, and Petersen, G.M. 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):224–228), and the Lung Cancer and Smoking study (Landi M.T., 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):679–691). 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.N.). 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.

## AUTHOR CONTRIBUTIONS

S.N. conducted the literature review, performed parts of the data analysis, interpreted the data, and drafted the manuscript. B.L.B. analysed the data, K.B. and S.A.B. contributed to writing the manuscript and provided detailed information on the participating studies. J.Y.D. contributed to analytical methods of the study. V.A., S.I.B., S.B., D.D.B., B.J.C., P.T.C., G.C., A.T.C., J.C.C., S.J.C., S.G., E.L.G., G.G.G., W.M.G., J.H., M.H., J.L.H., L.H., M.A.J., A.J., S.C.L., L.L.M., A.L., V.M., M.L., L.L., Y.L., K.O., P.A.N., P.D.P., J.D.P., L.Q., G.R., C.S., R.E.S., M.L.S., M.S., C.M.U., A.K.W., E.W., A.W., M.O.W., A.H.W., S.B.G., H.B. designed and implemented the different studies included in this analysis. U.P. coordinated the cooperation of the studies and data harmonisation. J.C.C. 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.

## ADDITIONAL INFORMATION

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41416-018-0108-8>.

**Competing interests:** The authors declare no competing interests.

**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.

**Ethics approval and consent to participate:** All participants provided written, informed consent and studies were approved by their respective institutional review boards.

**Note:** This work is published under the standard license to publish agreement. After 12 months the work will become freely available and the license terms will switch to a Creative Commons Attribution 4.0 International (CC BY 4.0).

## REFERENCES

- Ferlay, J. et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* **136**, E359–E386 (2015).
- Koo, J. H. & Leong, R. W. Sex differences in epidemiological, clinical and pathological characteristics of colorectal cancer. *J. Gastroenterol. Hepatol.* **25**, 33–42 (2010).
- 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. *Clin. Pharmacol. Ther.* **86**, 416–424 (2009).
- Lin, K. J., Cheung, W. Y., Lai, J. Y. & Giovannucci, E. L. The effect of oestrogen vs. combined oestrogen-progestogen therapy on the risk of colorectal cancer. *Int. J. Cancer* **130**, 419–430 (2012).
- Martinez, M. E. et al. A prospective study of reproductive factors, oral contraceptive use, and risk of colorectal cancer. *Cancer Epidemiol. Biomarkers Prev.* **6**, 1–5 (1997).
- Tsilidis, K. K. et al. Oral contraceptives, reproductive history and risk of colorectal cancer in the European Prospective Investigation into Cancer and Nutrition. *Br. J. Cancer* **103**, 1755–1759 (2010).
- La Vecchia, C. & Franceschi, S. Reproductive factors and colorectal cancer. *Cancer Causes Control* **2**, 193–200 (1991).
- Li, C. Y. et al. Age at menarche and risk of colorectal cancer: a meta-analysis. *PLoS ONE* **8**, e65645 (2013).
- Zervoudakis, A. et al. Reproductive history and risk of colorectal cancer in postmenopausal women. *J. Natl. Cancer Inst.* **103**, 826–834 (2011).
- Smith, G. D. & Ebrahim, S. 'Mendelian randomisation': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* **32**, 1–22 (2003).
- Burgess, S., Bowden, J., Fall, T., Ingelsson, E. & Thompson, S. G. Sensitivity analyses for robust causal inference from Mendelian randomisation analyses with multiple genetic variants. *Epidemiology* **28**, 30–42 (2017).
- Evans, D. M. & Davey Smith, G. Mendelian randomisation: new applications in the coming age of hypothesis-free causality. *Annu. Rev. Genom. Hum. Genet.* **16**, 327–350 (2015).
- Grodstein, F., Newcomb, P. A. & Stampfer, M. J. Postmenopausal hormone therapy and the risk of colorectal cancer: a review and meta-analysis. *Am. J. Med.* **106**, 574–582 (1999).
- Elbanna, H. G., Ebrahim, M. A., Abbas, A. M., Zalata, K. & Hashim, M. A. Potential value of oestrogen receptor beta expression in colorectal carcinoma: interaction with apoptotic index. *J. Gastrointest. Cancer* **43**, 56–62 (2012).
- Kennelly, R., Kavanagh, D. O., Hogan, A. M. & Winter, D. C. Oestrogen and the colon: potential mechanisms for cancer prevention. *Lancet Oncol.* **9**, 385–391 (2008).
- Caiazza, F., Ryan, E. J., Doherty, G., Winter, D. C. & Sheahan, K. Estrogen receptors and their implications in colorectal carcinogenesis. *Front. Oncol.* **5**, 19 (2015).
- Konstantinopoulos, P. A. et al. Oestrogen receptor beta (ERbeta) is abundantly expressed in normal colonic mucosa, but declines in colon adenocarcinoma paralleling the tumour's dedifferentiation. *Eur. J. Cancer* **39**, 1251–1258 (2003).
- Niv, Y. Estrogen receptor beta expression and colorectal cancer: a systematic review and meta-analysis. *Eur. J. Gastroenterol. Hepatol.* **27**, 1438–1442 (2015).
- Topi, G. et al. Association of the oestrogen receptor beta with hormone status and prognosis in a cohort of female patients with colorectal cancer. *Eur. J. Cancer* **83**, 279–289 (2017).
- Peters, U. et al. Identification of genetic susceptibility loci for colorectal tumours in a genome-wide meta-analysis. *Gastroenterology* **144**, 799–807 e724 (2013).
- Newcomb, P. A. et al. Colon Cancer Family Registry: an international resource for studies of the genetic epidemiology of colon cancer. *Cancer Epidemiol. Biomarkers Prev.* **16**, 2331–2343 (2007).
- Schumacher, F. R. et al. Genome-wide association study of colorectal cancer identifies six new susceptibility loci. *Nat. Commun.* **6**, 7138 (2015).
- Hutter, C. M. et al. Characterization of gene-environment interactions for colorectal cancer susceptibility loci. *Cancer Res.* **72**, 2036–2044 (2012).
- Day F. R. 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.* **49**, 834–841 (2017).
- Day, F. R. et al. Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. *Nat. Genet.* **47**, 1294–1303 (2015).
- Pierce, B. L., Ahsan, H. & Vanderweele, T. J. Power and instrument strength requirements for Mendelian randomisation studies using multiple genetic variants. *Int. J. Epidemiol.* **40**, 740–752 (2011).
- Stock, J. & Staiger, D. Instrumental variables regression with weak instruments. *Econometrica* **65**, 557–586 (1997).
- Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomisation 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 randomisation analyses using MR-Egger regression: the role of the I<sup>2</sup> statistic. *Int. J. Epidemiol.* **45**, 1961–1974 (2016).
- Haycock, P. C. et al. Best (but oft-forgotten) practices: the design, analysis, and interpretation of Mendelian randomisation studies. *Am. J. Clin. Nutr.* **103**, 965–978 (2016).
- Murphy, N., et al. Reproductive and menstrual factors and colorectal cancer incidence in the Women's Health Initiative Observational Study. *Br. J. Cancer* **116**, 117–125 (2016).
- Morch, L. S., Lidsgaard, O., Keiding, N., Lokkegaard, E. & Kjaer, S. K. The influence of hormone therapies on colon and rectal cancer. *Eur. J. Epidemiol.* **31**, 481–489 (2016).
- Chlebowski, R. T. et al. Estrogen plus progestin and colorectal cancer in postmenopausal women. *N. Engl. J. Med.* **350**, 991–1004 (2004).
- Burgess, S. Sample size and power calculations in Mendelian randomisation with a single instrumental variable and a binary outcome. *Int. J. Epidemiol.* **43**, 922–929 (2014).
- Davison, K. K., Susman, E. J. & Birch, L. L. Percent body fat at age 5 predicts earlier pubertal development among girls at age 9. *Pediatrics* **111**, 815–821 (2003).



36. Locke, A. E. et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
37. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent estimation in Mendelian randomisation with some invalid instruments using a weighted Median estimator. *Genet. Epidemiol.* **40**, 304–314 (2016).
38. Yavorska O. O., Burgess S. Mendelian randomisation: an R package for performing Mendelian randomisation analyses using summarised data. *Int. J. Epidemiol.* **46**, 1734–1739 (2017).
39. Buron Pust, A. et al. Heterogeneity of colorectal cancer risk by tumour characteristics: large prospective study of UK women. *Int. J. Cancer* **140**, 1082–1090 (2017).
40. Zhang, X. et al. Early life body fatness and risk of colorectal cancer in U.S. women and men—results from two large cohort studies. *Cancer Epidemiol. Biomarkers Prev.* **24**, 690–697 (2015).
41. Simmonds, M., Llewellyn, A., Owen, C. G. & Woolcott, N. Predicting adult obesity from childhood obesity: a systematic review and meta-analysis. *Obes. Rev.* **17**, 95–107 (2016).
42. Graff, M. 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. *Hum. Mol. Genet.* **22**, 3597–3607 (2013).
43. Akhter, M. et al. Reproductive factors, exogenous female hormone use and colorectal cancer risk: the Japan Public Health Center-based Prospective Study. *Eur. J. Cancer Prev.* **17**, 515–524 (2008).
44. Gunter, M. J. et al. Insulin, insulin-like growth factor-I, endogenous estradiol, and risk of colorectal cancer in postmenopausal women. *Cancer Res.* **68**, 329–337 (2008).
45. Murphy N., et al. A prospective evaluation of endogenous sex hormone levels and colorectal cancer risk in postmenopausal women. *J. Natl. Cancer. Inst.* **107**, djv210–djv210 (2015).
46. Clendenen, T. V. et al. Postmenopausal levels of endogenous sex hormones and risk of colorectal cancer. *Cancer Epidemiol., Biomarkers Prev.* **18**, 275–281 (2009).
47. Lin, J. H. et al. Association between sex hormones and colorectal cancer risk in men and women. *Clin. Gastroenterol. Hepatol.* **11**, 419–424 e411 (2013).
48. Falk, R. T. et al. Estrogen metabolites are not associated with colorectal cancer risk in postmenopausal women. *Cancer Epidemiol. Biomarkers Prev.* **24**, 1419–1422 (2015).
49. Apter, D., Reinila, M. & Vihko, R. Some endocrine characteristics of early menarche, a risk factor for breast cancer, are preserved into adulthood. *Int. J. Cancer* **44**, 783–787 (1989).
50. Emaus, A. et al. 17-beta-estradiol in relation to age at menarche and adult obesity in premenopausal women. *Hum. Reprod.* **23**, 919–927 (2008).
51. Limsui, D. et al. Postmenopausal hormone therapy and colorectal cancer risk by molecularly defined subtypes among older women. *Gut* **61**, 1299–1305 (2012).
52. Anderson, G. L. et al. Effects of conjugated equine oestrogen in postmenopausal women with hysterectomy: the Women's Health Initiative randomised controlled trial. *JAMA* **291**, 1701–1712 (2004).
53. Lavasani, S. et al. Estrogen and colorectal cancer incidence and mortality. *Cancer* **121**, 3261–3271 (2015).
54. Simon, M. S. et al. Estrogen plus progestin and colorectal cancer incidence and mortality. *J. Clin. Oncol.* **30**, 3983–3990 (2012).
55. Weyant, M. J. et al. Reciprocal expression of ERalpha and ERbeta is associated with oestrogen-mediated modulation of intestinal tumourigenesis. *Cancer Res.* **61**, 2547–2551 (2001).
56. Tillmans, L. S. et al. Associations between environmental exposures and incident colorectal cancer by ESR2 protein expression level in a population-based cohort of older women. *Cancer Epidemiol. Biomarkers Prev.* **24**, 713–719 (2015).
57. Rudolph, A. et al. Colorectal cancer risk associated with hormone use varies by expression of oestrogen receptor-beta. *Cancer Res.* **73**, 3306–3315 (2013).
58. Day, F. R. et al. Shared genetic aetiology of puberty timing between sexes and with health-related outcomes. *Nat. Commun.* **6**, 8842 (2015).
59. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An expanded view of complex traits: from polygenic to omnigenic. *Cell* **169**, 1177–1186 (2017).
60. Hemani G. et al. Automating Mendelian randomisation 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).

## AFFILIATIONS

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

<sup>1</sup>Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany; <sup>2</sup>Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98124, USA; <sup>3</sup>Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany; <sup>4</sup>Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892-9776, USA; <sup>5</sup>Centre Hospitalier Universitaire (CHU), Nantes Service de Génétique Médicale, 44093 Nantes, France; <sup>6</sup>Colorectal Oncogenomics Group, Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Melbourne, VIC 3010, Australia; <sup>7</sup>Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Parkville, VIC 3010, Australia; <sup>8</sup>Division of Research, Kaiser Permanente Medical Care Program of Northern California, Oakland 94612 CA, USA; <sup>9</sup>Epidemiology Research Program, American Cancer Society, Atlanta, GA 30329-4251, USA; <sup>10</sup>Center for Public Health Genomics, University of Virginia, Charlottesville, VA 22908, USA; <sup>11</sup>Clinical and Translational Epidemiology Unit and Division of Gastroenterology, Massachusetts General Hospital, Boston, MA 02115, USA; <sup>12</sup>Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA; <sup>13</sup>Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02115, USA; <sup>14</sup>Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892-9776, USA; <sup>15</sup>Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, ON M5G 1X5, Canada; <sup>16</sup>Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115, USA; <sup>17</sup>Department of Nutrition, Harvard School of Public Health, Boston, MA 02115, USA; <sup>18</sup>Department of Medicine, Harvard Medical School, Boston, MA 02115, USA; <sup>19</sup>Cancer Epidemiology & Intelligence Division, Cancer Council Victoria, Melbourne, VIC 3010, Australia; <sup>20</sup>Department of Medicine, Division of Gastroenterology, University of Washington School of Medicine, Seattle, WA 98195, USA; <sup>21</sup>Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA; <sup>22</sup>Medical Department 1, University Hospital Dresden, TU Dresden, 01307 Dresden, Germany; <sup>23</sup>Institute of Environmental Medicine, Karolinska Institutet Solna, SE-171 77 Stockholm, Sweden; <sup>24</sup>Epidemiology Program, University of Hawaii Cancer Center, Honolulu 96822 HI, USA; <sup>25</sup>Department of Clinical Genetics, Karolinska University Hospital Solna, SE-171 77 Stockholm, Sweden; <sup>26</sup>Department of Molecular Medicine and Surgery, Karolinska Institutet Solna, SE-171 77 Stockholm, Sweden; <sup>27</sup>Catalan Institute of Oncology, Bellvitge Biomedical Research Institute (IDIBELL), 08028 Barcelona, Spain; <sup>28</sup>CIBER Epidemiología y Salud Pública (CIBERESP), 28029 Madrid, Spain and University of Barcelona, Barcelona 08007, Spain; <sup>29</sup>Ontario Institute for Cancer Research, Toronto, ON M5G 0A3, Canada; <sup>30</sup>Department of Family Medicine and Community Health, Case Western Reserve University, Cleveland, OH 44106, USA; <sup>31</sup>Department of Medicine, Clinical Genetics



Service, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA; <sup>32</sup>Centre for Cancer Genetic Epidemiology, Department of Public Health & Primary Care, University of Cambridge, Cambridge CB2 8AR, UK; <sup>33</sup>Department of Public Health Sciences, University of California, Davis, CA 95817, USA; <sup>34</sup>Bruce Rappaport Faculty of Medicine, Technion-Israel Institute of Technology, Haifa, Israel; <sup>35</sup>Clalit Health Services National Israeli Cancer Control Center, Haifa 34361, Israel; <sup>36</sup>Department of Community Medicine and Epidemiology, Carmel Medical Center, Haifa 34361, Israel; <sup>37</sup>Department of General and Thoracic Surgery, University Hospital Schleswig-Holstein, Campus Kiel, 24118 Kiel, Germany; <sup>38</sup>Department of Medicine and Epidemiology, University of Pittsburgh Medical Center, Pittsburgh, PA 15213, USA; <sup>39</sup>Department of Internal Medicine, University of Utah Health Sciences Center, Salt Lake City, UT 84108, USA; <sup>40</sup>Huntsman Cancer Institute and Department of Population Health Sciences, University of Utah, Salt Lake City, UT 84112, USA; <sup>41</sup>Genetic Medicine and Familial Cancer Centre, Royal Melbourne Hospital, Parkville, VIC 3050, Australia; <sup>42</sup>Department of Surgical Sciences, Uppsala University, Uppsala SE-171 77, Sweden; <sup>43</sup>Faculty of Medicine, Memorial University of Newfoundland, St. John's, NL A1C 5S7, Canada; <sup>44</sup>Department of Preventive Medicine, University of Southern California Keck School of Medicine, Los Angeles, CA 90033, USA; <sup>45</sup>Department of Medicine, University of Southern California Keck School of Medicine, Los Angeles, CA 90033, USA; <sup>46</sup>Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), 69120 Heidelberg, Germany; <sup>47</sup>German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany and <sup>48</sup>Genetic Tumour Epidemiology Group, University Medical Center Hamburg-Eppendorf, University Cancer Center Hamburg, 20246 Hamburg, Germany