Human Molecular Genetics, 2014, Vol. 23, No. 14 **3898–3905**
doi:10.1093/hmg/ddu087
Advance Access published on February 21, 2014

Estimating the heritability of colorectal cancer

Shuo Jiao[1], Ulrike Peters[1], Sonja Berndt[2], Hermann Brenner[15], Katja Butterbach[15], Bette J. Caan[3], Christopher S. Carlson[1,8], Andrew T. Chan[4,5], Jenny Chang-Claude[6], Stephen Chanock[2], Keith R. Curtis[1], David Duggan[7], Jian Gong[1], Tabitha A. Harrison[1], Richard B. Hayes[9], Brian E. Henderson[10], Michael Hoffmeister[15], Laurence N. Kolonel[11], Loic Le Marchand[13], John D. Potter[1,8,14], Anja Rudolph[3], Robert E. Schoen[15], Daniela Seminara[3], Martha L. Slattery[16], Emily White[1] and Li Hsu[1,*]

[1]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA, [2]Division of Cancer Epidemiology and Genetics and [3]Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD, USA, [4]Division of Clinical Epidemiology and Aging Research and [5]Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany, [6]Division of Research, Kaiser Permanente Medical Care Program, Broadway, Oakland, CA, USA, [7]School of Public Health, University of Washington, Seattle, WA, USA, [8]Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA, [9]Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA, [10]Translational Genomics Research Institute, Phoenix, AZ, USA, [11]Division of Epidemiology, New York University School of Medicine, New York, NY, USA, [12]Keck School of Medicine, University of Southern California, Los Angeles, CA, USA, [13]Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI, USA, [14]Centre for Public Health Research, Massey University, Wellington, New Zealand, [15]Department of Medicine and Epidemiology, University of Pittsburgh Medical Center, Pittsburgh, PA, USA and [16]Department of Internal Medicine, University of Utah Health Sciences Center, Salt Lake City, UT, USA

Received November 14, 2013; Revised January 20, 2014; Accepted February 19, 2014

A sizable fraction of colorectal cancer (CRC) is expected to be explained by heritable factors, with heritability estimates ranging from 12 to 35% twin and family studies. Genome-wide association studies (GWAS) have successfully identified a number of common single-nucleotide polymorphisms (SNPs) associated with CRC risk. Although it has been shown that these CRC susceptibility SNPs only explain a small proportion of the genetic risk, it is not clear how much of the heritability these SNPs explain and how much is left to be detected by other, yet to be identified, common SNPs. Therefore, we estimated the heritability of CRC under different scenarios using Genome-Wide Complex Trait Analysis in the Genetics and Epidemiology of Colorectal Cancer Consortium including 8025 cases and 10 814 controls. We estimated that the heritability explained by known common CRC SNPs identified in GWAS was 0.65% (95% CI:0.3–1%; $P = 1.11 \times 10-16$), whereas the heritability explained by all common SNPs was at least 7.42% (95% CI: 4.71–10.12%; $P = 8.13 \times 10^{-8}$), suggesting that many common variants associated with CRC risk remain to be detected. Comparing the heritability explained by the common variants with that from twin and family studies, a fraction of the heritability may be explained by other genetic variants, such as rare variants. In addition, our analysis showed that the gene × smoking interaction explained a significant proportion of the CRC variance ($P = 1.26 \times 10^{-2}$). In summary, our results suggest that known CRC SNPs only explain a small proportion of the heritability and more common SNPs have yet to be identified.

*To whom correspondence should be addressed at: Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, M2-B500, Seattle, WA 98109, USA. Tel: +1 2066672854; Fax: +1 2066677004; Email: lih@fhcrc.org

© *The Author 2014. Published by Oxford University Press. All rights reserved.*
For Permissions, please email: journals.permissions@oup.com

Downloaded from https://academic.oup.com/hmg/article/23/14/3898/557628 by guest on 20 August 2022

INTRODUCTION

Colorectal cancer (CRC) is one of the most common cancers in developed countries. In the United States, the lifetime risk of CRC is estimated to be 5.2% for men and 4.8% for women, and CRC is the second leading cause of cancer death (1,2).

The heritability of CRC has been studied in twin and family studies. The largest twin study comparing monozygotic to dizygotic twins estimated that the heritability of CRC is 35%; however, while this significant estimate had a wide confidence interval (95% CI: 10–48%) (3). Using the nationwide Swedish Family-Cancer Database including 9.6 million subjects, another study estimated that the heritability of colon cancer is 13% (95% CI: 12–18%) and the heritability of rectal cancer is 12% (95% CI: 8–13%) (4). Of note is that the aforementioned estimates are for additive (or narrow sense) heritability as opposed to broad sense heritability, which includes all non-additive effects such as gene–gene interaction, dominant effects, and copy number variations in addition to additive effects. In this work, we will focus on estimating the additive heritability using the same scale as the twin and the family studies to put our findings in the context of these widely cited heritability estimates.

Genome-wide association studies (GWAS) have successfully identified many common single-nucleotide polymorphisms (SNPs) associated with CRC risk (5–19). However, it is not clear how much heritability the identified CRC susceptibility loci can explain. In addition, as the SNPs found by GWAS have to pass a very stringent significance threshold, there could be many SNPs with weak effect sizes that do not pass the threshold but still contribute to the heritability. Therefore, it is of great interest to find out the magnitude of the heritability explained by those potential susceptibility SNPs, which will help us make informed decisions about whether to pursue larger sample sizes to detect those yet-to-be detected common variants in GWAS. In addition, it will tell us whether less common and rare variants are important in explaining the CRC heritability. Similarly, estimating the variance explained by genome-wide gene–environment interaction (G × E) can also provide useful guidance in the search for G × E.

Yang *et al.* (20) proposed a method in Genome-wide Complex Trait Analysis (GCTA) to estimate heritability based on GWAS SNPs, no matter whether they pass a certain significance threshold or not (20). GCTA calculates the genetic similarity between subjects using all genotyped SNPs and uses the restricted maximum likelihood approach to estimate the heritability. In this paper, we aim to explore the CRC heritability explained by common SNPs using GCTA in the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO).

RESULTS

To increase the precision of the heritability estimation, we grouped the sample set by genotyping platform (300K, 550K and 730K platform). The details of the study populations are described by platform in Table 1. There are 248 977, 508 952 and 622 887 SNPs with MAF $\geq$ 0.01 on the 300K, 550K and 730K platforms, respectively. Of the 31 known CRC susceptibility SNPs (Supplementary Material, Table S1), 18, 30 and 26 of the SNPs themselves or proxies ($r^2 > 0.8$) were available on the 300K, 550K and 730K platform, respectively.

GCTA was used to estimate the CRC heritability and the results are summarized in Table 2. CRC heritability explained by known CRC susceptibility SNPs (or their proxies) was 0.64% in the 550K platform and 0.67% in the 730K platform. As only 18 out of 31 known CRC SNPs were represented on the 300K platform, the corresponding heritability was not used in the meta-analysis. The combined (across the 550K and 730K platform) heritability estimate explained by known loci

Table 1. Studies within GECCO used for estimating CRC heritability

Study	Platform	Case[a]	Control[a]	Female		Colon		Age (years)	
				No.	%	No.	%	Mean	Range
300K platform		$N = 4312$	$N = 4356$						
Colo2&3	Illumina 300K	87	125	95	44.8	59	67.8	65.2	38–86
DACHS Set 1	Illumina 300K	1710	1708	1395	40.8	1037	60.6	68.6	33–98
DALS Set 2	Illumina 300K	410	464	414	47.4	410	100	65.4	30–79
MEC	Illumina 300K	328	346	313	46.4	241	73.5	63.0	45–76
PLCO Set 2	Illumina 300K	486	415	383	42.5	320	65.8	63.6	55–75
VITAL	Illumina 300K	285	288	273	47.6	215	75.4	66.5	50–76
WHI Set 2	Illumina 300K	1006	1010	2016	100	703	69.9	65.8	50–79
550K platform		$N = 1709$	$N = 4214$						
DALS Set 1	Illumina 550K & 610K	706	710	615	43.4	702	99.4	65.0	30–79
PLCO Set 1[¶]	Illumina 300 & 240S, 610K	533	1976	667	26.6	516	96.8	64.1	55–74
WHI Set 1 + hip fracture	Illumina 550K, 550K duo, 610K	470	1528	1998	100	454	96.6	69.0	50–79
730K platform		$N = 2004$	$N = 2244$						
DACHS Set 2	Illumina 730K	666	498	435	37.4	385	57.8	69.1	30–99
HPFS Set 1	Illumina 730K	227	230	0	0	158	69.6	66.4	48–82
HPFS Set 2	Illumina 730K	176	172	0	0	111	63.1	63.7	48–83
NHS Set 1	Illumina 730K	394	774	1168	100	307	77.9	60.0	44–69
NHS Set 2	Illumina 730K	159	181	340	100	113	76.4	59.0	44–69
PHS Set 1 + 2	Illumina 730K	382	389	0	0	257	76.5	59.6	40–85

[a]Sample sizes given only for subjects clustering with HapMap CEU population in PCA (for data that has undergone QC), includes participants with data downloaded from dbGaP prostate and lung studies.

Downloaded from https://academic.oup.com/hmg/article/23/14/3898/557628 by guest on 20 August 2022

Table 2. CRC heritability estimates

	300K, heritability (s.e.) *P*	550K, heritability (s.e.) *P*	730K, heritability (s.e.) *P*	Meta, heritability (s.e.) *P*
Known SNP/proxy		0.0064(0.0025) 2.00e−10	0.0067(0.0027) 2.00e−08	0.0065(0.0018)[a] 1.11e−16
Known region		0.0118(0.0037) 1.00e−05	0.0110(0.0042) 4.00e−04	0.0114(0.0028)[a] 8.13e−08
Genome-wide SNP	0.0678(0.0165) 7.00e−06	0.0758(0.0330) 1.00e−02	0.1064(0.0382) 3.00e−04	0.0742(0.0138) 6.88e−09
Genome-wide SNP (proximal)	0.1092(0.0278) 7.00e−06	0.0997(0.0513) 3.00e−02	0.0833(0.0694) 7.00e−02	0.1044(0.0231) 2.67e−06
Genome-wide SNP (distal + rectal)	0.0885(0.0228) 3.00e−05	0.1550(0.0656) 1.00e−02	0.0779(0.0492) 3.00e−02	0.0928(0.0197) 1.72e−06

[a]Meta-analysis does not include estimate from 300K.

Table 3. Proportion of variance explained by G × E

	300K, proportion (s.e.) *P*	550K, proportion (s.e.) *P*	730K, proportion (s.e.) *P*	Meta, proportion (s.e.) *P*
Genome-wide SNPxSmoking	0.0683(0.0315) 1.00e−02	0.0820(0.0675) 1.00e−01	0.0586(0.0825) 3.00e−01	0.0694(0.0270) 1.26e−02
Genome-wide SNPxNSAIDs	0.0172(0.0308) 3.00e−01	0.0000(0.0626) 5.00e−01	0.0557(0.0760) 2.00e−01	0.0187(0.0260) 3.20e−01
Genome-wide SNPxT2D	0.0000(0.0450) 5.00e−01	0.0000(0.1507) 5.00e−01	0.0936(0.1135) 2.00e−01	0.0118(0.0403) 4.24e−01
Genome-wide SNPxProcMeat	0.0936(0.0391) 9.00e−03	0.0000(0.0722) 5.00e−01	0.0088(0.0834) 5.00e−01	0.0632(0.0318) 5.78e−02
Genome-wide SNPxVegetable	0.0026(0.0351) 5.00e−01	0.0841(0.0936) 2.00e−01	0.0000(0.0872) 5.00e−01	0.0111(0.0308) 4.24e−01
Genome-wide SNPxHeight	0.0008(0.0311) 5.00e−01	0.0271(0.0661) 3.00e−01	0.0937(0.0768) 1.00e−01	0.0160(0.0264) 2.10e−01
Genome-wide SNPxBMI	0.0466(0.0318) 6.00e−02	0.0000(0.0721) 5.00e−01	0.0000(0.0796) 5.00e−01	0.0344(0.0273) 2.10e−01
Genome-wide SNPxHRT	0.0000(0.0595) 5.00e−01	0.0000(0.1409) 5.00e−01	0.3699(0.2231) 9.00e−02	0.0211(0.0533) 2.70e−01
Genome-wide SNPxAlcohol	0.0000(0.0342) 5.00e−01	0.0674(0.0810) 2.00e−01	0.1300(0.0811) 5.00e−02	0.0258(0.0293) 1.02e−01
Genome-wide SNPxCalcium	0.0265(0.0385) 3.00e−01	0.0010(0.0908) 5.00e−01	0.1367(0.0925) 7.00e−02	0.0372(0.0331) 1.67e−01
Genome-wide SNPxFolate	0.0000(0.0620) 5.00e−01	0.0654(0.0897) 2.00e−01	0.2465(0.1287) 2.00e−02	0.0517(0.0474) 5.31e−02
Genome-wide SNPxFruit	0.0000(0.0349) 5.00e−01	0.0000(0.0875) 5.00e−01	0.0903(0.0863) 2.00e−01	0.0112(0.0303) 4.24e−01
Genome-wide SNPxFiber	0.0100(0.0600) 4.00e−01	0.0868(0.0932) 2.00e−01	0.0619(0.1643) 3.00e−01	0.0351(0.0482) 2.80e−01
Genome-wide SNPxExcercise	0.0117(0.0397) 4.00e−01	0.0220(0.0914) 4.00e−01	0.0993(0.1555) 3.00e−01	0.0178(0.0354) 4.15e−01

was 0.65% with 95% CI 0.3–1% and *P*-value of 1.11×10^{-16}. As fine-mapping of GWAS SNPs has often found stronger signals for putative causal variants tagged by GWAS SNPs and has also identified secondary independent signals we estimated the heritability explained by the region surrounding the GWAS SNPs (250 kb up and downstream). The meta-analysis heritability estimate for these GWAS regions was 1.14% (95% CI: 0.59–1.69%; $P = 8.13 \times 10^{-8}$). When we expanded the SNPs to all genome-wide common SNPs (MAF $\geq$ 0.01), the estimated heritability from the meta-analysis was 7.42% (95% CI: 4.71–10.12%; $P = 6.88 \times 10^{-9}$). When stratified by cancer sites (proximal versus distal + rectal), the estimated heritability for proximal CRC was 10.44% (95% CI: 5.91–14.97%; $P = 2.67 \times 10^{-6}$). For distal and rectal CRC, the estimated heritability was 9.28% (95% CI: 5.42–13.14%; $P = 1.72 \times 10^{-6}$). We also used the bivariate analysis of GCTA to estimate the genetic correlation between proximal and distal + rectal CRC using samples from the 300K platform since it is the only platform that has large enough sample size to yield meaningful estimates. The genetic correlation was estimated to be 0.42 (s.e. = 0.24). We also examined the variance explained by gene–environment interaction for numerous environmental variables such as smoking, NSAIDs, type 2 diabetes (T2D), height, body mass index (BMI), hormone replacement therapy (HRT; estimated in female), exercise and intakes of alcohol, calcium, folate, fruit, vegetable, fiber and processed meat (ProcMeat). From Table 3, it can be seen that the proportion of variance explained by the gene × smoking interaction was 6.94% (95% CI: 1.65–12.23%; $P = 1.26 \times 10^{-2}$) in the meta-analysis. We also observed marginally significant contributions of the gene × ProcMeat and gene × Folate interactions. We did not observe any variance significantly different from 0 that was explained by interactions from other environment variables.

DISCUSSION

In this paper, we estimated the heritability of CRC under different settings using GCTA. We found that the known CRC susceptibility SNPs identified so far by GWAS explain only a small proportion of the CRC heritability. In contrast, the explained heritability was much larger when considering all common SNPs together. We also found evidence that gene–environment interaction contributes to the CRC variance.

We estimated the heritability explained by the known CRC susceptibility SNPs to be 0.65% and the explained heritability increases to 1.14% after adding the SNPs within 250 kb of the known CRC SNPs. This finding indicates the existence of additional or stronger signals tagged by the GWAS SNP in the known CRC regions, which can be further explored in fine-mapping studies. In fact, fine-mapping studies have found variants with stronger signals for other complex traits such as T2D, BMI and prostate cancer (13,21,22). To find out whether the heritability explained by known regions is indeed enriched with CRC association signals, we randomly selected independent SNPs on the 550K platform that not in linkage disequilibrium (LD) with the known CRC SNPs but have similar MAF's. The regions were defined in the same way as the known regions ($\pm$250 kb of the randomly selected SNPs). Then we estimated

Downloaded from https://academic.oup.com/hmg/article/23/14/3898/557628 by guest on 20 August 2022

the heritability explained by the randomly sampled regions. We replicated this process 10 times and the resulting heritability estimates ranges from 0.0001 to 0.6% with an average of 0.15%. Compared with the heritability estimates 1.18% for known CRC regions on the 550K platform, this result demonstrates that the previous GWAS have (as can be expected) successfully highlighted numerous 'hotspots' for CRC risk.

Numerous previous studies have also estimated the variance explained by known CRC SNPs. Tomlinson *et al*. (10) estimated the GWAS loci accounted for 3–4% of the excess familial CRC risk and Houlston *et al*. (6) estimated the loci they identified explained ~1.5% of the excess familial risk (6, 10). It can be seen from our estimates that the proportion of heritability explained by known loci ranges from 5.4% (0.65%/12%) to 1.9% (0.65%/35%), which is comparable with previous estimates. Another previous publication estimated the heritability explained by 10 CRC susceptibility genes to be 1.26% on liability scale using the method developed by Wray *et al*. (23,24), which is based on MAF and effect sizes of the known SNPs. Using the same method, we estimated the heritability explained by 31 known loci to be 1.92%, which is >0.65% given by GCTA. Further investigations are needed to elucidate the reason for the discrepancy between the two methods.

The estimated heritability explained by the genome-wide common SNPs is 7.42%. This finding showed that known CRC SNPs only explained a small fraction of the heritability suggesting that many more common SNPs associated with CRC risk are yet to be discovered, possibly due to small effect sizes. Our findings agree with the previous studies of other complex traits, which also found that the heritability explained by all common SNPs on commercially available genotype chips is usually much larger than that by GWAS findings for a number of traits. For example, the heritability for height explained by all common SNPs is 45% compared with 10.5% explained by GWAS findings (25); for endometriosis, Alzheimer's disease and multiple sclerosis, the numbers are 26 versus <1%, 24 versus 18%, 30 versus 6%, respectively (26); for Parkinson's disease, it is 27 versus 3% (27).

The previous CRC heritability estimated from twin or family studies ranged from 12 to 35% (3,4). Our estimated heritability explained by all common SNPs is 7.42%. It can be seen that part of the CRC heritability is likely explained by other types of heritable factors, such as less common and rare variants. However, it has been pointed out before that the heritability estimated from a pedigree design can be biased due to non-additive genetic effects or incorrect assumption about shared environment. So the results should be interpreted with caution.

We did not observe significant difference in heritability estimates between proximal and distal + rectal CRC. We observed a genetic correlation of 0.42 between proximal and distal + rectal. The correlation is both different from 0, the independence ($P = 0.03$) and from 1, the complete correlation ($P = 0.03$). This finding, if replicated in independent datasets, implies that proximal and distal + rectal CRC have correlated yet distinctive genetic components.

Speed *et al* (2012) has observed that GCTA could produce inaccurate heritability estimates if there is uneven LD between SNPs so they proposed a method called Linkage Disequilibrium Adjusted Kinships (LDAK) to correct the bias (28). We re-estimated the CRC heritability using LDAK (Supplementary Material, Table S2) except for the known SNPs because they are independent with each other and LDAK will produce the same results as GCTA. It can be seen that the results from GCTA and LDAK were mostly similar. The only exceptions were the heritability estimates for the 730K platforms, where the LDAK estimates tend to be smaller than the GCTA estimates. For example, for the known regions, GCTA estimates the heritability to be 1.1% ($P = 4e-4$) for the 730K platform, which is close to the GCTA heritability estimates for the 550K platform (1.2%; $P = 1e-5$). However, the LDAK estimates for the 730K platform is much smaller (0.6%) and the *P*-value is not significant ($P = 0.06$), whereas the LDAK estimates for the 550K platform (1.3%; $P = 5e-5$) is still similar as that of GCTA. As the 730K platform is the densest among the three platforms used in the study, we suspect LDAK may have overcorrected the LD and resulted in underestimation of heritability. Further investigations are needed to study this peculiar behavior of LDAK on 730K platform. We also used LDAK to repeat the experiment of randomly sampling regions on the 550K platform to compare with the known regions and got similar conclusions. The LDAK heritability estimates for the randomly sampled regions range from 0.0001 to 0.06% with an average of 0.05%, which is much <1.3% explained by the known regions.

It has long been hypothesized that the interplay between genetic and environmental factors plays an important role in complex diseases such as CRC. Our results provide support for genome-wide gene × smoking interaction. One should note that the proportion of variance explained by G × E is a result of the interplay between G and E so it includes contributions from both G and E and cannot be directly called heritability. The rule of thumb is that it takes four times as many samples to detect an interaction as the number needed to detect a main genetic effect of comparable effect size (29). Therefore, it is harder to detect significant G × E contributions. Overall, the estimates from G × E were less significant compared with the additive heritability and not as consistent across platforms. Thus the results should be interpreted with caution and certainly merit replication in independent studies. If replicated, our results show that G × E may contribute a sizable fraction of CRC variability for some of the environmental variables and that genome-wide G × E searches should be conducted in studies with very large sample sizes.

For the heritability estimation for binary traits, one common problem is that estimates from different methods are usually not directly comparable due to the ascertainment (30). In GCTA, a prevalence parameter needs to be specified in order to transform the estimated heritability from observed scale to liability scale (31). In the literature, both the disease prevalence (27) and the lifetime risk (26,32) have been used for the purpose. In previous studies where the CRC heritability was estimated using twins or family data, the liability threshold was computed from the disease prevalence (3,4). As we aimed to put our results in context with the estimates from the family or twin studies, we based our estimates also on the disease prevalence. For comparison, we also provided the heritability estimates based on lifetime risk in Supplementary Material, Tables S3 and S4.

In addition to the known CRC SNPs identified from GWAS, a number of CRC familial genes have also been found (33,34). It was estimated that 2–5% of the CRC cases can be attributed to

Downloaded from https://academic.oup.com/hmg/article/23/14/3898/557628 by guest on 20 August 2022

those high penetrance genes leading to Lynch syndrome, familial adenomatous polyposis and other CRC syndromes (35–37). Note that 2–5% is not an heritability estimate. In fact, as the causal variants within CRC familial genes are very rare, the corresponding explained heritability is hard to estimate using standard tools even if their effect sizes are very large.

As different platforms have different genomic coverage, the expected explained heritability can also be different. To explore the impact of different coverage has on the heritability estimate, we down-sampled the SNPs on the 730K platform to 201 147 overlapping SNPs between 730K and 300K platforms. Using the same samples on 730K platform, the heritability explained by the 201 147 down-sampled SNPs was 9.5%, which shows very little decrease compared with the 10.6% explained by the full SNP set in 730K platform. This finding suggests that even the least dense platform (300K) may have covered most CRC-related common SNPs so that heritability estimates in the three different platforms are largely comparable.

A major strength of our study is the large study population with genetic data, which is essential for calculating relatively accurate heritability estimates. Another particular strength is the availability of a wide range of harmonized environmental variables on a large number of samples across our studies, which makes it possible to estimate the contribution of gene–environment interaction to CRC variability. However, there are also limitations that should be noted. We did not use imputed data to unify the SNP set across different platforms for two reasons: first, even though imputation provides a uniform SNP set, the imputation quality can still vary across different platforms with higher-coverage platform having better imputation quality, which may also cause incomparable heritability estimates across platforms; second, it has been observed in previous studies that heritability estimates are very close using imputed and genotyped SNPs (27,38).

In summary, we performed a comprehensive exploration of CRC heritability using large sample sizes. Our findings suggest that the previously identified CRC-associated SNPs explain a small fraction of the heritability and the heritability explained by the undetected common SNPs is >10-fold higher than the heritability explained by the known CRC SNPs. We also found that a fraction of the CRC heritability is unlikely to be explained by common SNPs, supporting the ideas of expanding the search to other factors, such as structural and rare variants using whole-exome and whole-genome sequencing, and epigenetic factors, among others. In addition, we found evidence that the gene × smoking interaction explained a significant proportion of CRC variance, which supports the potential important role of gene−environment interactions in CRC.

MATERIALS AND METHODS

Study participants

Each study is described in detail in the Supplementary Material. In brief, CRC cases were defined as colorectal adenocarcinomas confirmed by medical records, pathologic reports or death certificate. All participants gave written informed consent and studies were approved by their respective Institutional Review Boards.

Genotyping and QC

Detailed information on genotyping and quality control procedures has been described before (5) and is available in Supplementary Material. In brief, DNA was extracted from blood samples or for a small subset of samples, from buccal cells, using conventional methods. All studies included 1–6% blinded duplicates to monitor quality of the genotyping. For studies used for estimating CRC heritability, the genotyping was done on Illumina 300K, Illumina 550K, combined Illumina 300K&240K, Illumina 610K or Illumina 730K chips. Samples were excluded based on call rate, heterozygosity, unexpected duplicates, gender discrepancy and unexpectedly high identity-by-descent or unexpected genotype concordance (>65%) with another individual. All analyses were restricted to samples clustering with the Utah residents with Northern and Western European ancestry from the CEPH collection (CEU) population in principal component analysis, including the HapMap II populations as reference. SNPs were excluded if they were triallelic, not assigned an rs number, or were reported or observed as not performing consistently across platforms. Additionally, genotyped SNPs were excluded based on call rate (<98%), lack of Hardy Weinberg equilibrium in controls (HWE; $P < 1 \times 10^{-4}$) and minor allele frequency.

Harmonization of environmental data

Information on basic demographics and environmental risk factors was collected by using in-person interviews and/or structured questionnaires, as detailed previously (39–48). We carried out a multi-step data harmonization procedure, reconciling each study's unique protocols and data-collection instruments at the GECCO coordinating center (Fred Hutchinson Cancer Research Center). First, we defined common data elements (CDEs). We examined the questionnaires and data dictionaries for each study to identify study-specific data elements that could be mapped to the CDEs. Through an iterative process, we communicated with each data contributor to obtain relevant data and coding information. The data elements were written to a common data platform, transformed via an SQL programming script, and combined into a single dataset with common definitions, standardized permissible values and standardized coding. The mapping and resulting data were reviewed for quality assurance, and range and logic checks were performed to assess data distributions within and between studies. Outlying samples were truncated to the minimum or maximum value of established range for each variable. The reference time for cohort studies was time of enrollment (WHI and PLCO) or blood draw (HPFS, NHS and PHS). Data harmonization were performed using SAS and T-SQL. The associations between environmental risk factors and CRC risk were highly statistically significant in the predicted direction.

Statistical method

We used GCTA to estimate CRC heritability under various scenarios (20). First, a genetic relationship matrix (GRM) of all pairs of samples within each platform was computed based on all SNPs. The GRMs were then used as input for the restricted maximum likelihood analysis to estimate the heritability explained by the selected set of SNPs, such as GWAS SNPs or

Downloaded from https://academic.oup.com/hmg/article/23/14/3898/557628 by guest on 20 August 2022

all SNPs on the platform. We adjusted for age, gender and study. The CRC prevalence was estimated to be 0.004 according to SEER Cancer Statistics (1). GCTA used the provided disease prevalence to transform the estimated heritability to the liability scale.

As large sample sizes are required for accurate heritability estimates (20,30), we grouped studies by their genotyping platforms (Table 1). We estimated the heritability for the following scenarios:

(1) Heritability explained by known GWAS SNPs for CRC. We estimated the heritability explained by 31 autosomal SNPs that have previously been shown to be associated with CRC risk in GWAS of European ancestry individuals (Supplementary Material, Table S1). If a known GWAS SNP was not on a platform, we identified a proxy SNP on the platform based on the highest correlation (r^2) with the known GWAS SNP. If no proxy SNP was found with $r^2 > 0.8$, the known SNP was not be included in the estimation.
(2) Heritability explained by known GWAS regions for CRC. We estimated the heritability explained by all variants spanning genomic regions 250 kb upstream or downstream of the 31 known GWAS SNPs for CRC.
(3) Heritability explained by genome-wide common SNPs. We estimated the heritability explained by all SNPs with MAF $\geq$ 0.01 separately for each platform.
(4) Variance explained by gene–environment interactions. GCTA also allows estimating the variance explained by gene–environment interactions with categorized environmental factors by including a vector of genotype–environment interaction effects in the model, so that for the pairs of individuals in the same environment the GRM for interaction is the same as that for genotype and for the pairs of individuals in different environments the GRM is 0. This method has been applied to estimate the variance explained by genotype–sex interaction for height, weight, BMI, vWF and QTi (49). For smoking, NSAIDs, T2D, exercise and HRT, dichotomous variables were used; for processed meat, vegetable, calcium, folate, fruit and fiber intake, study- and gender-specific quartiles were used; gender-specific quartiles were used for height; BMI was categorized into $<$18.5, 18.5–24.9, 25–30 and $>$30; alcohol intake was categorized into three categories nondrinker, 1–28 and $>$28 g/day. We examined the variance explained by the potential interaction between these environmental variables and genome-wide common SNPs.

We combined the heritability estimates from the three different platform groups using inverse variance weighting meta-analysis. The meta-analysis *P*-value was calculated using Fisher's method.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGMENTS

GECCO: The authors thank all those at the GECCO Coordinating Center for helping bring together the data and people that made this project possible. DACHS: We thank all participants and cooperating clinicians, and Ute Handte-Daub, Renate Hettler-Jensen, Utz Benscheid, Muhabbet Celik and Ursula Eilber for excellent technical assistance. HPFS, NHS and PHS: We acknowledge Patrice Soule and Hardeep Ranu of the Dana Farber Harvard Cancer Center High-Throughput Polymorphism Core who assisted in the genotyping for NHS, HPFS and PHS under the supervision of Dr Immaculata Devivo and Dr David Hunter, Qin (Carolyn) Guo and Lixue Zhu who assisted in programming for NHS and HPFS, and Haiyan Zhang who assisted in programming for the PHS. We thank the participants and staff of the Nurses' Health Study and the Health Professionals Follow-Up 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. In addition, this study was approved by the Connecticut Department of Public Health (DPH) Human Investigations Committee. Certain data used in this publication were obtained from the DPH. The authors assume full responsibility for analyses and interpretation of these data. 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, Wen Shao and staff, SAIC-Frederick. Most importantly, we acknowledge the study participants for their contributions to making this study possible. 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: https://cleo.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Short%20List.pdf

Conflict of Interest statement. None decalred.

FUNDING

The work was supported by: GECCO: National Cancer Institute, National Institutes of Health, U.S. Department of Health and Human Services (U01 CA137088; R01 CA059045; U01 CA164930). COLO2&3: National Institutes of Health (R01 CA60987). DACHS: German Research Council (Deutsche Forschungsgemeinschaft, BR 1704/6-1, BR 1704/6-3, BR 1704/6-4 and CH 117/1-1), and the German Federal Ministry of Education and Research (01KH0404 and 01ER0814). DALS: National Institutes of Health (R01 CA48998 to M.L.S.); HPFS is supported by the National Institutes of Health (P01 CA 055075, UM1 CA167552, R01 137178 and P50 CA 127003), NHS by the National Institutes of Health (R01 CA137178, P01 CA 087969 and P50 CA 127003) and PHS by the National Institutes of Health (CA42182). MEC: National Institutes of Health (R37 CA54281, P01 CA033619 and R01 CA63464). PLCO: Intramural Research Program of the Division of Cancer Epidemiology and Genetics and supported by contracts from the Division of Cancer Prevention, National Cancer Institute, NIH, DHHS. Additionally, a subset of control samples were

Downloaded from https://academic.oup.com/hmg/article/23/14/3898/557628 by guest on 20 August 2022

genotyped as part of the Cancer Genetic Markers of Susceptibility (CGEMS) Prostate Cancer GWAS (50), Colon CGEMS pancreatic cancer scan (PanScan) (51,52) and the Lung Cancer and Smoking study. The prostate and PanScan study datasets were accessed with appropriate approval through the dbGaP online resource (http://cgems.cancer.gov/data/) accession numbers phs000207v.1p1 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. 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 HHSN268 201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C and HHSN27 1201100004C.

REFERENCES

1. Howlader, N., Noone, A., Krapcho, M., Garshell, J., Neyman, N., Altekruse, S., Kosary, C., Yu, M., Ruhl, J., Tatalovich, Z. *et al.* (2013) *SEER Cancer Statistics Review, 1975–2010*. National Cancer Institute, Bethesda, MD.
2. Jemal, A., Siegel, R., Xu, J. and Ward, E. (2010) Cancer statistics, 2010. *CA Cancer J. Clin.*, **60**, 277–300.
3. Lichtenstein, P., Holm, N.V., Verkasalo, P.K., Iliadou, A., Kaprio, J., Koskenvuo, M., Pukkala, E., Skytthe, A. and Hemminki, K. (2000) Environmental and heritable factors in the causation of cancer – analyses of cohorts of twins from Sweden, Denmark, and Finland. *N. Engl. J. Med.*, **343**, 78–85.
4. Czene, K., Lichtenstein, P. and Hemminki, K. (2002) Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database. *Int. J. Cancer*, **99**, 260–266.
5. Peters, U., Jiao, S., Schumacher, F.R., Hutter, C.M., Aragaki, A.K., Baron, J.A., Berndt, S.I., Bézieau, S., Brenner, H., Butterbach, K. *et al.* (2012) Identification of genetic susceptibility loci for colorectal tumors in a genome-wide meta-analysis. *Gastroenterology*, 10.1053/j.gastro.2012.12.020.
6. Houlston, R.S., Cheadle, J., Dobbins, S.E., Tenesa, A., Jones, A.M., Howarth, K., Spain, S.L., Broderick, P., Domingo, E., Farrington, S. *et al.* (2010) Meta-analysis of three genome-wide association studies identifies susceptibility loci for colorectal cancer at 1q41, 3q26.2, 12q13.13 and 20q13.33. *Nat. Genet.*, **42**, 973–977.
7. Gong, J., Hutter, C., Baron, J.A., Berndt, S., Caan, B., Campbell, P.T., Casey, G., Chan, A.T., Cotterchio, M., Fuchs, C.S. *et al.* (2012) A pooled analysis of smoking and colorectal cancer: timing of exposure and interactions with environmental factors. *Cancer Epidemiol. Biomarkers Prev.*, **21**, 1974–1985.
8. Jia, W.-H., Zhang, B., Matsuo, K., Shin, A., Xiang, Y.-B., Jee, S.H., Kim, D.-H., Ren, Z., Cai, Q., Long, J. *et al.* (2013) Genome-wide association analyses in East Asians identify new susceptibility loci for colorectal cancer. *Nat. Genet.*, **45**, 191–196.
9. Dunlop, M.G., Dobbins, S.E., Farrington, S.M., Jones, A.M., Palles, C., Whiffin, N., Tenesa, A., Spain, S., Broderick, P., Ooi, L.-Y. *et al.* (2012) Common variation near CDKN1A, POLD3 and SHROOM2 influences colorectal cancer risk. *Nat. Genet.*, **44**, 770–776.
10. Tomlinson, I.P.M., Webb, E., Carvajal-Carmona, L., Broderick, P., Howarth, K., Pittman, A.M., Spain, S., Lubbe, S., Walther, A., Sullivan, K. *et al.* (2008) A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat. Genet.*, **40**, 623–630.
11. Tomlinson, I., Webb, E., Carvajal-Carmona, L., Broderick, P., Kemp, Z., Spain, S., Penegar, S., Chandler, I., Gorman, M., Wood, W. *et al.* (2007) A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat. Genet.*, **39**, 984–988.
12. Zanke, B.W., Greenwood, C.M.T., Rangrej, J., Kustra, R., Tenesa, A., Farrington, S.M., Prendergast, J., Olschwang, S., Chiang, T., Crowdy, E. *et al.* (2007) Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat. Genet.*, **39**, 989–994.
13. Peters, U., North, K.E., Sethupathy, P., Buyske, S., Haessler, J., Jiao, S., Fesinmeyer, M.D., Jackson, R.D., Kuller, L.H., Rajkovic, A. *et al.* (2013) A systematic mapping approach of 16q12.2/FTO and BMI in more than 20,000 African Americans Narrows in on the underlying functional variation: results from the Population Architecture using Genomics And Epidemiology (PAGE) Study. *PLoS Genet.*, **9**, e1003171.
14. Broderick, P., Carvajal-Carmona, L., Pittman, A.M., Webb, E., Howarth, K., Rowan, A., Lubbe, S., Spain, S., Sullivan, K., Fielding, S. *et al.* (2007) A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. *Nat. Genet.*, **39**, 1315–1317.
15. Tenesa, A., Farrington, S.M., Prendergast, J.G.D., Porteous, M.E., Walker, M., Haq, N., Barnetson, R.A., Theodoratou, E., Cetnarskyj, R., Cartwright, N. *et al.* (2008) Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat. Genet.*, **40**, 631–637.
16. Jaeger, E., Webb, E., Howarth, K., Carvajal-Carmona, L., Rowan, A., Broderick, P., Walther, A., Spain, S., Pittman, A., Kemp, Z. *et al.* (2008) Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. *Nat. Genet.*, **40**, 26–28.
17. Houlston, R.S., Webb, E., Broderick, P., Pittman, A.M., Di Bernardo, M.C., Lubbe, S., Chandler, I., Vijayakrishnan, J., Sullivan, K., Penegar, S. *et al.* (2008) Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nat. Genet.*, **40**, 1426–1435.
18. Tomlinson, I.P.M., Carvajal-Carmona, L.G., Dobbins, S.E., Tenesa, A., Jones, A.M., Howarth, K., Palles, C., Broderick, P., Jaeger, E.E.M., Farrington, S. *et al.* (2011) Multiple common susceptibility variants near BMP pathway loci GREM1, BMP4, and BMP2 explain part of the missing heritability of colorectal cancer. *PLoS Genet.*, **7**, e1002105.
19. Peters, U., Hutter, C.M., Hsu, L., Schumacher, F.R., Conti, D.V., Carlson, C.S., Edlund, C.K., Haile, R.W., Gallinger, S., Zanke, B.W. *et al.* (2011) Meta-analysis of new genome-wide association studies of colorectal cancer risk. *Hum. Genet.*, **131**, 217–234.
20. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W. *et al.* (2010) Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.*, **42**, 565–569.
21. Morris, A.P., Voight, B.F., Teslovich, T.M., Ferreira, T., Segrè, A.V., Steinthorsdottir, V., Strawbridge, R.J., Khan, H., Grallert, H., Mahajan, A. *et al.* (2012) Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.*, **44**, 981–990.
22. Kote-Jarai, Z., Saunders, E.J., Leongamornlert, D.A., Tymrakiewicz, M., Dadaev, T., Jugurnauth-Little, S., Ross-Adams, H., Al Olama, A.A., Benlloch, S., Halim, S. *et al.* (2013) Fine-mapping identifies multiple prostate cancer risk loci at 5p15, one of which associates with TERT expression. *Hum. Mol. Genet.*, **22**, 2520–2528.
23. Tenesa, A. and Dunlop, M.G. (2009) New insights into the aetiology of colorectal cancer from genome-wide association studies. *Nat. Rev. Genet.*, **10**, 353–358.
24. Wray, N.R., Goddard, M.E. and Visscher, P.M. (2007) Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res.*, **17**, 1520–1528.
25. Lango Allen, H., Estrada, K., Lettre, G., Berndt, S.I., Weedon, M.N., Rivadeneira, F., Willer, C.J., Jackson, A.U., Vedantam, S., Raychaudhuri, S. *et al.* (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*, **467**, 832–838.
26. Lee, S.H., Harold, D., Nyholt, D.R., Goddard, M.E., Zondervan, K.T., Williams, J., Montgomery, G.W., Wray, N.R. and Visscher, P.M. (2013) Estimation and partitioning of polygenic variation captured by common SNPs for Alzheimer's disease, multiple sclerosis and endometriosis. *Hum. Mol. Genet.*, **22**, 832–841.
27. Keller, M.F., Saad, M., Bras, J., Bettella, F., Nicolaou, N., Simón-Sánchez, J., Mittag, F., Büchel, F., Sharma, M., Gibbs, J.R. *et al.* (2012) Using

Downloaded from https://academic.oup.com/hmg/article/23/14/3898/557628 by guest on 20 August 2022

genome-wide complex trait analysis to quantify "missing heritability" in Parkinson's disease. *Hum. Mol. Genet.*, **21**, 4996–5009.
28. Speed, D., Hemani, G., Johnson, M.R. and Balding, D.J. (2012) Improved heritability estimation from genome-wide SNPs. *Am. J. Hum. Genet.*, **91**, 1011–1021.
29. Smith, P.G. and Day, N.E. (1984) The design of case-control studies: the influence of confounding and interaction effects. *Int. J. Epidemiol.*, **13**, 356–365.
30. Zaitlen, N. and Kraft, P. (2012) Heritability in the genome-wide association era. *Hum. Genet.*, **131**, 1655–1664.
31. Lee, S.H., Wray, N.R., Goddard, M.E. and Visscher, P.M. (2011) Estimating missing heritability for disease from genome-wide association studies. *Am. J. Hum. Genet.*, **88**, 294–305.
32. Ek, W.E., Levine, D.M., D'Amato, M., Pedersen, N.L., Magnusson, P.K.E., Bresso, F., Onstad, L.E., Schmidt, P.T., Törnblom, H., Nordenstedt, H. *et al.* (2013) Germline genetic contributions to risk for esophageal adenocarcinoma, Barrett's Esophagus, and gastroesophageal reflux. *J. Natl. Cancer Inst.*, 10.1093/jnci/djt303.
33. Pearson, P., Francomano, C., Foster, P., Bocchini, C., Li, P. and McKusick, V. (1994) The status of online Mendelian inheritance in man (OMIM) medio 1994. *Nucleic Acids Res.*, **22**, 3470–3473.
34. Palles, C., Cazier, J.-B., Howarth, K.M., Domingo, E., Jones, A.M., Broderick, P., Kemp, Z., Spain, S.L., Guarino, E., Guarino Almeida, E. *et al.* (2013) Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat. Genet.*, **45**, 136–144.
35. Dunlop, M.G., Tenesa, A., Farrington, S.M., Ballereau, S., Brewster, D.H., Koessler, T., Pharoah, P., Schafmayer, C., Hampe, J., Völzke, H. *et al.* (2013) Cumulative impact of common genetic variants and other risk factors on colorectal cancer risk in 42,103 individuals. *Gut*, **62**, 871–881.
36. Jasperson, K.W., Tuohy, T.M., Neklason, D.W. and Burt, R.W. (2010) Hereditary and familial colon cancer. *Gastroenterology*, **138**, 2044–2058.
37. Hampel, H., Frankel, W.L., Martin, E., Arnold, M., Khanduja, K., Kuebler, P., Clendenning, M., Sotamaa, K., Prior, T., Westman, J.A. *et al.* (2008) Feasibility of screening for Lynch syndrome among patients with colorectal cancer. *J. Clin. Oncol.*, **26**, 5783–5788.
38. Yang, J., Lee, T., Kim, J., Cho, M.-C., Han, B.-G., Lee, J.-Y., Lee, H.-J., Cho, S. and Kim, H. (2013) Ubiquitous polygenicity of human complex traits: genome-wide analysis of 49 traits in Koreans. *PLoS Genet.*, **9**, e1003355.
39. Küry, S., Buecher, B., Robiou-du-Pont, S., Scoul, C., Sébille, V., Colman, H., Le Houérou, C., Le Neel, T., Bourdon, J., Faroux, R. *et al.* (2007) Combinations of cytochrome P450 gene polymorphisms enhancing the risk for sporadic colorectal cancer related to red meat consumption. *Cancer Epidemiol. Biomarkers Prev.*, **16**, 1460–1467.
40. Brenner, H., Chang-Claude, J., Seiler, C.M., Rickert, A. and Hoffmeister, M. (2011) Protection from colorectal cancer after colonoscopy: a population-based, case-control study. *Ann. Intern. Med.*, **154**, 22–30.
41. Slattery, M.L., Potter, J., Caan, B., Edwards, S., Coates, A., Ma, K.N. and Berry, T.D. (1997) Energy balance and colon cancer – beyond physical activity. *Cancer Res.*, **57**, 75–80.
42. Christen, W.G., Gaziano, J.M. and Hennekens, C.H. (2000) Design of Physicians' Health Study II – a randomized trial of beta-carotene, vitamins E and C, and multivitamins, in prevention of cancer, cardiovascular disease, and eye disease, and review of results of completed trials. *Ann. Epidemiol.*, **10**, 125–134.
43. Prorok, P.C., Andriole, G.L., Bresalier, R.S., Buys, S.S., Chia, D., Crawford, E.D., Fogel, R., Gelmann, E.P., Gilbert, F., Hasson, M.A. *et al.* (2000) Design of the prostate, lung, colorectal and ovarian (PLCO) cancer screening trial. *Control. Clin. Trials*, **21**, 273S–309S.
44. Anderson, G.L., Manson, J., Wallace, R., Lund, B., Hall, D., Davis, S., Shumaker, S., Wang, C.-Y., Stein, E. and Prentice, R.L. (2003) Implementation of the Women's Health Initiative Study design. *Ann. Epidemiol.*, **13**, S5–17.
45. Newcomb, P.A., Baron, J., Cotterchio, M., Gallinger, S., Grove, J., Haile, R., Hall, D., Hopper, J.L., Jass, J., Le Marchand, L. *et al.* (2007) Colon cancer family registry: an international resource for studies of the genetic epidemiology of colon cancer. *Cancer Epidemiol. Biomarkers Prev.*, **16**, 2331–2343.
46. Hoffmeister, M., Raum, E., Krtschil, A., Chang-Claude, J. and Brenner, H. (2009) No evidence for variation in colorectal cancer risk associated with different types of postmenopausal hormone therapy. *Clin. Pharmacol. Ther.*, **86**, 416–424.
47. Colditz, G.A. and Hankinson, S.E. (2005) The Nurses' Health Study: lifestyle and health among women. *Nat. Rev. Cancer*, **5**, 388–396.
48. Giovannucci, E., Rimm, E.B., Stampfer, M.J., Colditz, G.A., Ascherio, A. and Willett, W.C. (1994) Aspirin use and the risk for colorectal cancer and adenoma in male health professionals. *Ann. Intern. Med.*, **121**, 241–246.
49. Yang, J., Manolio, T.A., Pasquale, L.R., Boerwinkle, E., Caporaso, N., Cunningham, J.M., de Andrade, M., Feenstra, B., Feingold, E., Hayes, M.G. *et al.* (2011) Genome partitioning of genetic variation for complex traits using common SNPs. *Nat. Genet.*, **43**, 519–525.
50. Yeager, M., Orr, N., Hayes, R.B., Jacobs, K.B., Kraft, P., Wacholder, S., Minichiello, M.J., Fearnhead, P., Yu, K., Chatterjee, N. *et al.* (2007) Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat. Genet.*, **39**, 645–649.
51. Amundadottir, L., Kraft, P., Stolzenberg-Solomon, R.Z., Fuchs, C.S., Petersen, G.M., Arslan, A.A., Bueno-de-Mesquita, H.B., Gross, M., Helzlsouer, K., Jacobs, E.J. *et al.* (2009) Genome-wide association study identifies variants in the ABO locus associated with susceptibility to pancreatic cancer. *Nat. Genet.*, **41**, 986–990.
52. Petersen, G.M., Amundadottir, L., Fuchs, C.S., Kraft, P., Stolzenberg-Solomon, R.Z., Jacobs, K.B., Arslan, A.A., Bueno-de-Mesquita, H.B., Gallinger, S., Gross, M. *et al.* (2010) A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. *Nat. Genet.*, **42**, 224–228.

Downloaded from https://academic.oup.com/hmg/article/23/14/3898/557628 by guest on 20 August 2022