Mendelian randomisation implicates hyperlipidaemia as a risk factor for colorectal cancer

Short title: Hyperlipidaemia as a risk factor for colorectal cancer

Henry Rodriguez-Broadbent[1*], Philip J. Law[1*], Amit Sud[1*], Kimmo Palin[2,3], Sari Tuupanen[2,3], Alexandra Gylfe[2,3], Ulrika A. Hänninen[2,3], Tatiana Cajuso[2,3], Tomas Tanskanen[2,3], Johanna Kondelin[2,3], Eevi Kaasinen[2,3], Antti-Pekka Sarin[4], Samuli Ripatti[4,5,6], Johan G. Eriksson[7,8,9], Harri Rissanen[7], Paul Knekt[7], Eero Pukkala[10,11], Pekka Jousilahti[7], Veikko Salomaa[7], Aarno Palotie[4,12,13,14], Laura Renkonen-Sinisalo[15], Anna Lepistö[15], Jan Böhm[16], Jukka-Pekka Mecklin[17], Nada A. Al-Tassan[18], Claire Palles[19], Lynn Martin[19], Ella Barclay[19], Susan M. Farrington[20], Maria N. Timofeeva[20], Brian F. Meyer[18], Salma M. Wakil[18], Harry Campbell[21], Christopher G. Smith[22], Shelley Idziaszczyk[22], Timothy S. Maughan[23], Richard Kaplan[24], Rachel Kerr[25], David Kerr[26], Michael N. Passarelli[27], Jane C. Figueiredo[28,29], Daniel D. Buchanan[30,31], Aung K. Win[31], John L. Hopper[31], Mark A. Jenkins[31], Noralane M. Lindor[32], Polly A. Newcomb[33], Steven Gallinger[34], David Conti[35], Fred Schumacher[35], Graham Casey[36], Lauri A. Aaltonen[2,3], Jeremy P. Cheadle[22], Ian P. Tomlinson[19], Malcolm G. Dunlop[20], Richard S. Houlston[1‡]

1 Division of Genetics and Epidemiology, The Institute of Cancer Research, London, SW7 3RP, UK
2 Genome-Scale Biology Research Program, Research Programs Unit, University of Helsinki, Helsinki, 00014, Finland
3 Department of Medical and Clinical Genetics, Medicum, University of Helsinki, Helsinki, 00014, Finland
4 Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, 00014, Finland
5 Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, United Kingdom
6 Department of Public Health, University of Helsinki, Helsinki, 00014, Finland
7 National Institute for Health and Welfare, Helsinki, 00271, Finland
8 Folkhälsan Research Centre, Helsinki, 00250, Finland
9 Unit of General Practice and Primary Health Care, University of Helsinki and Helsinki University Hospital, Helsinki, 00014, Finland
10 Finnish Cancer Registry, Institute for Statistical and Epidemiological Cancer Research, Helsinki, 00130, Finland
11 School of Health Sciences, University of Tampere, Tampere, 33014, Finland
12 Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA, 02114, USA

[13] Program in Medical and Population Genetics, The Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA
[14] Department of Neurology, Massachusetts General Hospital, Boston, MA, 02114, USA
[15] Abdominal Center, Department of Surgery, Helsinki University Hospital, Helsinki, 00029, Finland
[16] Department of Pathology, Central Finland Central Hospital, Jyväskylä, 40620, Finland
[17] Department of Surgery, Jyväskylä Central Hospital, University of Eastern Finland, Jyväskylä, 40620, Finland
[18] Department of Genetics, King Faisal Specialist Hospital and Research Center, Riyadh, 12713, Saudi Arabia
[19] Wellcome Trust Centre for Human Genetics and NIHR Comprehensive Biomedical Research Centre, Oxford, OX3 7BN, UK
[20] Colon Cancer Genetics Group, University of Edinburgh and MRC Human Genetics Unit, Western General Hospital, Edinburgh, EH4 2XU, UK
[21] Centre for Population Health Sciences, University of Edinburgh, Edinburgh, EH8 9AG, UK
[22] Division of Cancer and Genetics, School of Medicine, Cardiff University, Cardiff, CF14 4XN, UK
[23] CRUK/MRC Oxford Institute for Radiation Oncology, University of Oxford, Oxford, OX3 7DQ, UK
[24] MRC Clinical Trials Unit, Aviation House, London, WC2B 6NH, UK
[25] Oxford Cancer Centre, Department of Oncology, University of Oxford, Churchill Hospital, Oxford, OX3 7LE, UK
[26] Nuffield Department of Clinical Laboratory Sciences, University of Oxford, John Radcliffe Hospital, Oxford, OX3 9DU, UK
[27] Department of Epidemiology, Geisel School of Medicine at Dartmouth, Dartmouth-Hitchcock Medical Center, Lebanon, NH, 03756, USA
[28] Samuel Oschin Comprehensive Cancer Center, Cedars-Sinai Medical Center, Los Angeles CA, USA
[29] Keck School of Medicine, University of Southern California, Los Angeles, CA, USA
[30] Colorectal Oncogenomics Group, Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Victoria, 3010, Australia
[31] Centre for Epidemiology and Biostatistics, The University of Melbourne, Victoria, 3010, Australia
[32] Department of Health Sciences Research, Mayo Clinic, Scottsdale, AZ, 85259, USA
[33] Cancer Prevention Program, Fred Hutchinson Cancer Research Center, Seattle, WA, 98109, USA
[34] Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, ON, M5G 1X5, Canada
[35] Department of Preventive Medicine, University of Southern California, Los Angeles, CA, 90033, USA

[36] Center for Public Health Genomics, University of Virginia, Charlottesville, VA, 22908, USA

* These authors contributed equally to this work

‡ Correspondence to: Richard S Houlston, Division of Genetics and Epidemiology, The Institute of Cancer Research, 15 Cotswold Road, London, SM2 5NG. Tel: +44 (0) 208 722 4175; Fax: +44 (0) 722 4365; E-mail: richard.houlston@icr.ac.uk

Key words

Mendelian randomisation, hyperlipidaemia, cholesterol, colorectal cancer, risk

Abbreviations

CI: confidence interval

CRC: colorectal cancer

HDL: high-density lipoprotein

IV: instrumental variables

LDL: low-density lipoprotein

MR: Mendelian randomisation

OR: odds ratios

SNP: single nucleotide polymorphisms

TC: total cholesterol

TG: triglyceride

Novelty and Impact

While observational studies have suggested an association between blood cholesterol levels and colorectal cancer (CRC), they do not establish causality and may be influenced by confounding factors. Here we use Mendelian randomisation using genetic instrumental variables to provide evidence for a causal link between blood cholesterol levels and colorectal cancer. Thus, reducing hyperlipidaemia is an important target for primary prevention of CRC in the population.

ABSTRACT

While elevated blood cholesterol has been associated with an increased risk of colorectal cancer (CRC) in observational studies, causality is uncertain. Here we apply a Mendelian randomisation (MR) analysis to examine the potential causal relationship between lipid traits and CRC risk. We used single nucleotide polymorphisms (SNPs) associated with blood levels of total cholesterol (TC), triglyceride (TG), low-density lipoprotein (LDL), and high-density lipoprotein (HDL) as instrumental variables (IV). We calculated MR estimates for each risk factor with CRC using SNP-CRC associations from 9,254 cases and 18,386 controls. Genetically predicted higher TC was associated with an elevated risk of CRC (odds ratios (OR) per unit SD increase = 1.46, 95% confidence interval [CI]: 1.20-1.79, P=1.68x10^{-4}). The pooled ORs for LDL, HDL, and TG were 1.05 (95% CI: 0.92-1.18, P=0.49), 0.94 (95% CI: 0.84-1.05, P= 0.27), and 0.98 (95% CI: 0.85-1.12, P=0.75) respectively. A genetic risk score for 3-hydoxy-3-methylglutaryl-coenzyme A reductase (*HMGCR*) to mimic the effects of statin therapy was associated with a reduced CRC risk (OR=0.69, 95% CI: 0.49-0.99, P=0.046). This study supports a causal relationship between higher levels of TC with CRC risk, and a further rationale for implementing public health strategies to reduce the prevalence of hyperlipidaemia.

INTRODUCTION

Colorectal cancer (CRC) is the third most common cancer diagnosed in economically developed countries[1]. The mortality rate from CRC has been declining over the last twenty years as a consequence of improved medical care and probably through the introduction of population screening programs for the early detection of tumours[2-4]. Despite this improvement in patient outcome, it is still important to understand the risk factors for CRC in order to inform public health policy.

A number of factors influenced by lifestyle have been reported to be associated with the development of CRC in epidemiological observational studies, including a positive correlation with circulating levels of plasma cholesterol and other components of the lipid profile[5, 6]. It is, however, unclear from these studies if findings reflect a causal relationship or are simply a consequence of confounding by factors common to the aetiology of both CRC and hyperlipidaemia (*e.g.* common dietary factors) or reverse causality. Because lipid levels can be modified by lifestyle and treatment with statins, deciphering the basis for the association should be informative in formulating and optimizing prevention programs for CRC.

Evidence that statin use will effect a reduction in CRC is highly controversial[7, 8]. Although an analysis of The Health Improvement Network (THIN) database found that statin usage was associated with reduced CRC (long term usage: odds ratio [OR] = 0.95, 95% confidence interval [CI]: 0.91-0.99; short term usage: OR= 0.92, 95% CI: 0.85-0.99); no difference was shown between continued versus discontinued therapy, suggesting indication bias[8]. Moreover a recent meta-analysis of data from eight randomized controlled trials (RCTs) failed to demonstrate a beneficial effect which was statistically significant (relative risk = 0.89, 95% CI: 0.74-1.07)[9]. Each of these RCTs, however have the same limitations of short follow-up time, few CRC cases, and ascertainment of CRC as a secondary outcome.

Mendelian randomisation (MR) provides a useful complement to the traditional epidemiological study[10]. This strategy makes use of genetic variants that are robustly associated with traits of interest, in this case lipid traits - total cholesterol (TC), low-density

lipoprotein (LDL), high-density lipoprotein (HDL), and triglyceride (TG) - as instrumental variables (IV) to infer whether associations between exposure and disease are causal. The use of genetic variants as IV to proxy modifiable exposure therefore avoids confounding by environmental factors, can be reflective of life-long exposure (propensity), and is not be subject to reverse causality. The strength of the IV in MR is important for power, but weak instruments can also lead to inconsistent instrumental variables estimators. Hence using a genetic score derived from a combination of single nucleotide polymorphisms (SNPs), which collectively explains more of the variance in the risk factor, mitigates against weak instrument bias thereby increasing study power.

Genetics scores derived from multiple SNPs for lipid traits have been used in MR studies to investigate associations between blood lipids and coronary heart disease[11], and most recently prostate cancer[12]. Here we have employed MR to examine the impact of lipid traits on the risk of developing CRC.

METHODS

Colorectal cancer datasets

We investigated the relationship between genetic risk scores for lipid traits and CRC risk using data from seven previously reported genome-wide association studies (GWAS) of CRC[13] (**Table 1**). Briefly, these GWAS were all based on individuals with European ancestry and comprise: CCFR1, CCFR2, COIN, FINLAND, UK1, Scotland1 and VQ58. All studies were approved by their respective institutional review boards and conducted with appropriate ethical criteria in each country and in accordance with the Declaration of Helsinki. Comprehensive details on the cases and controls are available in previously published work[13-16].

Genotyping data

Details of the genotyping and quality control of the seven CRC GWAS have been previously published[13]. Briefly, we excluded SNPs with a minor allele frequency of <1%, low call rate <95%, SNPs violating Hardy-Weinberg equilibrium, and individuals with non-European ancestry as assessed using HapMap v2 reference data[17]. Imputation of untyped SNP genotypes was performed using IMPUTEv2 software[18] using a merged reference panel consisting of Sequencing Initiative Suomi (for the FINLAND data) or UK10K (for the remaining data) in addition to 1000 Genomes Project data. Poorly imputed SNPs (*i.e.* INFO score of <0.8) were excluded. Summary statistics from the seven GWAS were used to calculate the ORs for lipid-related SNPs.

Gene variants used to construct genetic risk scores

Genetic risk scores as IVs for circulating lipid fractions were developed from SNPs previously identified by the Global Lipids Genetics Consortium (GLGC)[19]. Median and range of standard deviations of lipid trait measurements in European cohorts of the Global Lipids Genetics Consortium are shown in **Supplementary Table 1.** We considered only SNPs associated at genome-wide significance (*i.e.* $P \leq 5.0 \times 10^{-8}$) and restricted to individuals with European Ancestry. To avoid co-linearity between SNPs, we excluded SNPs that were correlated (*i.e*. r^2 value ≥ 0.01), only considering the SNP with the strongest effect on the lipid trait for inclusion in genetic risk scores. Pairwise r^2 values were calculated using PLINK v1.90 utilising samples

of European ancestry from the 1000 Genomes and UK10K sequencing projects (**Supplementary Data**). This resulted in 58 SNPs for HDL, 29 SNPs for LDL, 26 SNPs for TG, and 38 SNPs for TC (**Supplementary Table 2**). Because lipid traits share common genetic variants, in addition to calculating an 'unrestricted allele score' that included all SNPs associated with the lipid trait, we also calculated a 'restricted allele score' as *per* Holmes *et al* [11] based on SNPs exclusively associated with HDL (n=43), LDL (n=9), or TG (n=14) to make them as specific as possible (**Supplementary Table 3**). Risk alleles were those that were positively associated with TC, LDL and TG or negatively associated with HDL levels. For all identified SNPs, we recovered the chromosome positions, the risk alleles, association estimates and standard errors.

Statistical analysis

We performed MR analysis to assess the association between TC, LDL, HDL, TG and CRC using summary statistics as described Burgess *et al.* (2015) [20]. The combined ratio estimate ($\hat{\beta}$) of all SNPs associated with each lipid trait on CRC was calculated under a fixed-effects model:

$$\hat{\beta} = \frac{\sum_k X_k Y_k \sigma_{Y_k}^{-2}}{\sum_k X_k^2 \sigma_{Y_k}^{-2}} \quad .$$

X_k corresponds to the association between SNP *k* with the lipid trait and Y_k is the association between SNP *k* and CRC risk with standard error σ_{Y_k}. The standard error of the combined ratio estimate is given by:

$$\mathrm{se}(\hat{\beta}) = \sqrt{\frac{1}{\sum_k X_k^2 \sigma_{Y_k}^{-2}}} \quad .$$

With the statistics generated by following these calculations on the seven different cohorts in the CRC data, we performed a meta-analysis under a fixed-effects model to derive the final ORs and confidence intervals.

A key assumption for this MR analysis is there is no pleiotropism (*i.e.* a gene influencing multiple traits) between the genes influencing CRC and the lipid traits under study. Therefore, before performing the MR analysis, we performed LD regression to test for global evidence of pleiotropy as *per* Bulik-Sullivan *et al.* (2015) [21, 22], and subsequently implemented an MR-Egger regression to examine for violation of the standard IV assumptions in our analysis [23].

For each statistical test we considered a global significance level of $P \leq 0.05$ as being satisfactory to derive conclusions. To assess the robustness of our conclusions, we imposed a conservative Bonferroni-corrected significance threshold of 0.0125 (*i.e*. 0.05/4 lipid traits). We deemed a *P*-value > 0.05 as non-significant (*i.e*. no association), a *P*-value ≤0.05 as evidence for a potential causal association, and a *P*-value ≤0.0125 as significant evidence for a causal association. All statistical analyses were undertaken using R software (Version 2.14.1).

The power of a MR investigation depends greatly on the proportion of variance in the risk factor that is explained by the IV. We estimated study power using the methodology of Burgess (2014) [24], utilizing published estimates of the heritability of lipid trait associated IV SNPs [19] and the reported effect of each trait on CRC risk in epidemiological studies [8].

In a subsidiary analysis we constructed a genetic risk score for 3-hydroxy-3-methylglutaryl-CoA reductase (HMGCR) using rs12916, rs17238484, rs5909, rs2303152, rs10066707 and rs2006760. These specific SNPs have previously been used to mimic statin intervention to estimate a causal association of statin use and coronary heart disease and diabetes [25].

RESULTS

Using LD regression, we found no evidence for global pleiotropism (*i.e.* shared genetic components) between CRC and any of the lipid traits under investigation (**Table 2**). Following on from these observations we performed MR-Egger regression tests to explicitly examine for infringement of the standard instrumental variable assumptions in our MR analysis. We did not find evidence of any violation in respect to TC, LDL, HDL or TG (**Table 2, Supplementary Figure 1**). In view of the totality of these findings we were reassured of the validity of our MR-based analysis to infer whether the relation between exposures and CRC were likely to be causal.

The associations of each unrestricted allele score for respective target lipid traits are shown in **Figure 1**. A positive correlation between variants associated with higher risk levels of TC and CRC was observed. The pooled OR meta-analysis for CRC by TC, estimated in IV analysis using the allele score was 1.46 per genetically instrumented SD increase in TC (95% CI: 1.20-1.79, $P = 1.68 \times 10^{-4}$, test for heterogeneity between studies $I^2 = 6\%$, $P_{het} = 0.38$).

The strongest reported SNP association for TC levels was provided by rs10401969 (CILP2) and rs12916 (HMGCR)[19]. To examine if the correlation between TC and CRC risk was primarily driven by these variants, we performed a sensitivity analysis excluding rs10401969 and rs12916. Omission of these two SNPs from the MR analysis did not appreciably affect our MR findings with results remaining significant (OR = 1.69, 95% CI: 1.25-2.28, $P = 6.76 \times 10^{-4}$). Albeit not significant, there was some support for a positive association with LDL (OR = 1.05, 95% CI: 0.92-1.18, $P = 0.49$) and CRC risk, and a negative association between HDL (OR = 0.94, 95% CI: 0.84-1.05, $P = 0.27$) and CRC risk.

Following on from these analyses, we performed a MR based analysis of LDL, HDL and TG using genetic scores derived from restricted sets of SNPs. As with the unrestricted analysis, no significant causal effect for each of these lipid traits was observed (**Supplementary Figure 2**).

Finally, genetically predicted lowered TC using the HMGCR genetic risk score was associated with 43% reduction in CRC (OR=0.69, 95% CI: 0.49-0.99, P=0.046, P_{het}= I^2=56%).

DISCUSSION

The present study strengthens a causal inference between circulating levels of TC and risk of developing CRC that is independent of known confounding effects. The positive correlation between the IV for TC and CRC risk, remained significant even after imposing a Bonferroni-correction to account for multiple testing. It is noteworthy that none of the IV SNPs for TC also represent IVs for obesity[26], supporting an independent relationship between TC and CRC. As illustrated here and in previously studies of obesity and CRC [27, 28], insulin levels and uterine cancer [29], and lipid levels and coronary heart disease [30] MR provides an attractive means of establishing causal associations. In addition to demonstrating an association between TC and CRC risk we found that genetic variants that mimic the effect of HMGCR inhibition were associated with a reduced CRC risk, supporting findings from observational epidemiological studies that statins have beneficial effect on the population burden of CRC.

Studies in mice have shown that knocking out the cell surface cholesterol-sensing receptor gene *NPC1L1*, which plays a critical role in the absorption of intestinal cholesterol, reduces CRC risk[31]. However, the biological mechanism by which cholesterol may affect CRC risk remains to be established. Cholesterol is thought to have multiple carcinogenic/cancer promoting effects at the cellular level and several mechanisms have been variously suggested, including the cholesterol-mediated activation of the NLRP3 inflammasome[32]. Since statins are largely retained by hepatocytes, their effect on CRC will be indirect, via HMGCR inhibition. Intriguingly, recent data suggests that any impact of statin therapy on CRC is by prevention of progression of adenomas to frank cancers rather than their development *per se* [33]. Further research on the biological relationship between cholesterol and CRC is needed to address such a proposition.

A major strength of our MR analysis is that it does not suffer from the influence of recall bias and confounding that affects traditional observational studies. Nevertheless, a primary assumption in MR is that the variants used to generate genetic scores are indeed associated with the exposure being examined. To ensure this was the case, we only made use of variants associated with each lipid trait at genome-wide significance from hypothesis-free GWAS. A

second assumption is that variants are associated with CRC only through the exposure and are not confounded by shared genetic (*i.e.* pleiotropy). This would be revealed as an increasing linear relationship between SNPs and their effect size for any lipid trait and CRC risk; we did not observe such a relationship. Although it is not possible to exclude confounding by unknown confounders, the use of multiple independent variants acting through different pathways reduces the likelihood of confounded IV-associations. Moreover by using LD regression, we have been able to exclude pleiotropism on a global basis[21]. Finally, we only made use of data from individuals of European descent in the GWAS SNPs to limit potential bias from population stratification influencing study findings.

As with any MR analysis, there are potential limitations to our findings, including the limited trait variance explained by genetic variants, restricting statistical power. This is especially relevant for null findings, since wide confidence intervals leave uncertainty over the presence of a causal effect. It is estimated that the SNPs from the Global Lipids Genetics Consortium GWAS explain approximately 8-11% of the total variation in each lipid trait[19]. Recent analyses of observational studies found higher impact on CRC for TC than LDL or TG; respective ORs and 95% CIs – 1.49 (1.32-1.69), 1.37 (1.11-1.69), and 1.16 (1.06-1.27) [8]. Based on these data our MR study was well-powered to demonstrate a causal relation for TC (≈80%, stipulating a *P*-value of 0.05), but we had limited power to identify associations for other lipid traits, particularly TG and HDL (respective power estimates for TG, LDL and HDL being 13%, 68% and 31%). Hence while the ORs for CRC with LDL and TG are congruous with observational studies [34] larger studies are required to formally establish a relationship using MR.

There are differences in the genomic landscapes of colonic and rectal cancers which presumably may reflect differences in aetiology. Unfortunately, these data were not uniformly collected across datasets, and we therefore did not investigate the possibility of differential effects of cholesterol on risk by anatomical location within the colorectum[35].

In conclusion, this study provides evidence for a causal role of higher TC levels in the aetiology of CRC. Hence our findings encouragingly support the overall findings of past observational studies. Our limited power to further refine the relationship between lipid profile and CRC provides a motivational for larger MR studies, which will benefit from enhanced statistical

power to demonstrate relationships for the spectrum of colorectal neoplasia. Irrespective of the exact functional basis of the association between TC and CRC risk, reducing hyperlipidaemia is an important target for primary prevention of CRC in the population. Our analysis therefore supports the hypothesis that the increasing use of statins in the population for prevention of cardiovascular disease will have the added bonus of reducing the burden of CRC.

FUNDING

H.R-B was supported by a summer student grant from the Wellcome Trust (202639/Z/16/Z). At the Institute of Cancer Research, this work was supported by Cancer Research UK (C1298/A8362 - Bobby Moore Fund for Cancer Research UK). Additional support was provided by the National Cancer Research Network. A.S. is supported by a clinical fellowship from Cancer Research UK. In Edinburgh the work was supported by Programme Grant funding from Cancer Research UK (C348/A12076). In Oxford additional funding was provided by the Oxford Comprehensive Biomedical Research Centre and the EU FP7 CHIBCHA grant. Core infrastructure support to the Wellcome Trust Centre for Human Genetics, Oxford was provided by grant (090532/Z/09/Z). We are grateful to many colleagues within UK Clinical Genetics Departments (for CORGI) and to many collaborators who participated in the VICTOR and QUASAR2 trials. We also thank colleagues from the UK National Cancer Research Network (for NSCCG). Support from the European Union (FP7/207-2013, grant 258236) and FP7 collaborative project SYSCOL and COST Action in the UK is also acknowledged (BM1206). The COIN and COIN-B trials were funded by Cancer Research UK and the Medical Research Council and were conducted with the support of the National Institute of Health Research Cancer Research Network. COIN and COIN-B translational studies were supported by the Bobby Moore Fund from Cancer Research UK, Tenovus, the Kidani Trust, Cancer Research Wales and the National Institute for Social Care and Health Research Cancer Genetics Biomedical Research Unit (2011-2014). N.A.A., B.F.M. and S.M.W. were funded and supported by KFSHRC. In Finland, this work was supported by grants from the Academy of Finland (Finnish Center of Excellence Program 2012–2017, 250345), the Jane and Aatos Erkko Foundation, the Finnish Cancer Society (K.P.), the European Research Council [ERC; 268648], the Sigrid Juselius Foundation, SYSCOL, the Nordic Information for Action eScience Center (NIASC), the Nordic

Center of Excellence financed by NordForsk (project 62721, K.P.) and State Research Funding of Kuopio University Hospital (B1401). We acknowledge the computational resources provided by the ELIXIR node, hosted at the CSC–IT Center for Science, Finland, and funded by the Academy of Finland (grants 271642 and 263164), the Ministry of Education and Culture, Finland. V.S. was supported by the Finnish Academy (grant number 139635). Sample collection and genotyping in the Finnish Twin Cohort has been supported by the Wellcome Trust Sanger Institute, ENGAGE – European Network for Genetic and Genomic Epidemiology, FP7-HEALTH-F4-2007 (201413), the National Institute of Alcohol Abuse and Alcoholism [grants AA-12502 and AA-00145 to Richard J Rose and K02AA018755 to Danielle M Dick] and the Academy of Finland (100499, 205585, 265240 and 263278 to J.K.). The work of the Colon Cancer Family Registry (CCFR) was supported by from the National Cancer Institute (UM1 CA167551), National Institutes of Health and through cooperative agreements with the following CCFR centers: Australasian Colorectal Cancer Family Registry (U01 CA074778, U01/U24 CA097735), USC Consortium Colorectal Cancer Family Registry (U01/U24 CA074799), Mayo Clinic Cooperative Familial Registry for Colon Cancer Studies (U01/U24 CA074800), Ontario Familial Colorectal Cancer Registry (U01/U24 CA074783), Seattle Colorectal Cancer Family Registry (U01/U24 CA074794), and University of Hawaii Colorectal Cancer Family Registry (U01/U24 CA074806). The CCFR Illumina GWAS was supported by funding from the National Cancer Institute, National Institutes of Health (U01 CA122839 and R01 CA143237 to G.C.). Additional support was provided from the Surveillance, Epidemiology and End Results (SEER) Program of the National Cancer Institute to Fred Hutchinson Cancer Research Center (Control Nos. N01-CN-67009, N01-PC-35142 and Contract No. HHSN2612013000121), the Hawai'i Department of Health (Control Nos. N01-PC-67001, N01-PC-35137, and Contract No. HHSN26120100037C), and the California Department of Public Health (contract HHSN261201000035C) awarded to the University of Southern California. The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute, any SEER program or any of the collaborating centers in the CCFR, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government, SEER or the CCFR.

ACKNOWLEDGEMENTS

We are grateful to all individuals who participated in the various studies. This study made use of genotyping data from the 1958 Birth Cohort, kindly made available by the Wellcome Trust Case Control Consortium 2. A full list of the investigators who contributed to the generation of the data is available at http://www.wtccc.org.uk/.

Conflict of interest statement: None declared.

REFERENCES

1. Forman D, Bray F, Brewster DH, Gombe Mbalawa C, Kohler B, Piñeros M, Steliarova-Foucher E, Swaminathan R, Ferlay J. Cancer incidence in five continents IARC Scientific publications 2014;X:917.

2. Cancer Research UK. vol. 2016, 2016.

3. Hewitson P, Glasziou P, Watson E, Towler B, Irwig L. Cochrane systematic review of colorectal cancer screening using the fecal occult blood test (hemoccult): an update. Am J Gastroenterol 2008;103:1541-9.

4. Shaukat A, Mongin SJ, Geisser MS, Lederle FA, Bond JH, Mandel JS, Church TR. Long-term mortality after screening for colorectal cancer. N Engl J Med 2013;369:1106-14.

5. Yao X, Tian Z. Dyslipidemia and colorectal cancer risk: a meta-analysis of prospective studies. Cancer causes & control : CCC 2015;26:257-68.

6. Tian Y, Wang K, Li J, Wang J, Wang Z, Fan Y, Ye Y, Ji G, Li Y. The association between serum lipids and colorectal neoplasm: a systemic review and meta-analysis. Public health nutrition 2015;18:3355-70.

7. Bardou M, Barkun A, Martel M. Effect of statin therapy on colorectal cancer. Gut 2010;59:1572-85.

8. Mamtani R, Lewis JD, Scott FI, Ahmad T, Goldberg DS, Datta J, Yang YX, Boursi B. Disentangling the Association between Statins, Cholesterol, and Colorectal Cancer: A Nested Case-Control Study. PLoS medicine 2016;13:e1002007.

9. Lytras T, Nikolopoulos G, Bonovas S. Statins and the risk of colorectal cancer: an updated systematic review and meta-analysis of 40 studies. World journal of gastroenterology 2014;20:1858-70.

10. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? International journal of epidemiology 2003;32:1-22.

11. Holmes MV, Asselbergs FW, Palmer TM, Drenos F, Lanktree MB, Nelson CP, Dale CE, Padmanabhan S, Finan C, Swerdlow DI, Tragante V, van Iperen EP, et al. Mendelian randomization of blood lipids for coronary heart disease. European heart journal 2015;36:539-50.

12. Bull CJ, Bonilla C, Holly JM, Perks CM, Davies N, Haycock P, Yu OH, Richards JB, Eeles R, Easton D, Kote-Jarai Z, Amin Al Olama A, et al. Blood lipids and prostate cancer: a Mendelian randomization analysis. Cancer medicine 2016;5:1125-36.

13. Orlando G, Law PJ, Palin K, Tuupanen S, Gylfe A, Hänninen U, Cajuso T, Tanskanen T, Kondelin J, Kaasinen E, Sarin A-P, Kaprio J, et al. Variation at 2q35 (PNKD and TMBIM1) influences colorectal cancer risk and identifies a pleiotropic effect with inflammatory bowel disease. Human Molecular Genetics 2016.

14. Houlston RS, members of C. COGENT (COlorectal cancer GENeTics) revisited. Mutagenesis 2012;27:143-51.

15. Whiffin N, Hosking FJ, Farrington SM, Palles C, Dobbins SE, Zgaga L, Lloyd A, Kinnersley B, Gorman M, Tenesa A, Broderick P, Wang Y, et al. Identification of susceptibility loci for colorectal cancer in a genome-wide meta-analysis. Human molecular genetics 2014;23:4729-37.

16. Al-Tassan NA, Whiffin N, Hosking FJ, Palles C, Farrington SM, Dobbins SE, Harris R, Gorman M, Tenesa A, Meyer BF, Wakil SM, Kinnersley B, et al. A new GWAS and meta-analysis with 1000Genomes imputation identifies novel risk variants for colorectal cancer. Sci Rep 2015;5:10442.

17. International HapMap C. The International HapMap Project. Nature 2003;426:789-96.

18. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 2009;5:e1000529.

19. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J, Buchkovich ML, Mora S, Beckmann JS, Bragg-Gresham JL, et al. Discovery and refinement of loci associated with lipid levels. Nat Genet 2013;45:1274-83.

20. Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG, Consortium E-I. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. European journal of epidemiology 2015;30:543-52.

21. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, Daly MJ, Price AL, Neale BM. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 2015;47:291-5.

22. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, Duncan L, Perry JR, Patterson N, Robinson EB, Daly MJ, Price AL, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet 2015;47:1236-41.

23. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. International Journal of Epidemiology 2015;44:512-25.

24. Burgess S. Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. International journal of epidemiology 2014;43:922-9.

25. Ference BA, Robinson JG, Brook RD, Catapano AL, Chapman MJ, Neff DR, Voros S, Giugliano RP, Davey Smith G, Fazio S, Sabatine MS. Variation in PCSK9 and HMGCR and Risk of Cardiovascular Disease and Diabetes. New England Journal of Medicine 2016;375:2144-53.

26. Berndt SI, Gustafsson S, Magi R, Ganna A, Wheeler E, Feitosa MF, Justice AE, Monda KL, Croteau-Chonka DC, Day FR, Esko T, Fall T, et al. Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. Nat Genet 2013;45:501-12.

27. Jarvis D, Mitchell JS, Law PJ, Palin K, Tuupanen S, Gylfe A, Hanninen UA, Cajuso T, Tanskanen T, Kondelin J, Kaasinen E, Sarin AP, et al. Mendelian randomisation analysis strongly implicates adiposity with risk of developing colorectal cancer. Br J Cancer 2016;115:266-72.

28. Thrift AP, Gong J, Peters U, Chang-Claude J, Rudolph A, Slattery ML, Chan AT, Locke AE, Kahali B, Justice AE, Pers TH, Gallinger S, et al. Mendelian Randomization Study of Body Mass Index and Colorectal Cancer Risk. Cancer Epidemiol Biomarkers Prev 2015;24:1024-31.

29. Nead KT, Sharp SJ, Thompson DJ, Painter JN, Savage DB, Semple RK, Barker A, Australian National Endometrial Cancer Study G, Perry JR, Attia J, Dunning AM, Easton DF, et al. Evidence of a Causal Association Between Insulinemia and Endometrial Cancer: A Mendelian Randomization Analysis. J Natl Cancer Inst 2015;107.

30. Burgess S, Harshfield E. Mendelian randomization to assess causal effects of blood lipids on coronary heart disease: lessons from the past and applications to the future. Curr Opin Endocrinol Diabetes Obes 2016;23:124-30.

31. He J, Shin H, Wei X, Kadegowda AK, Chen R, Xie SK. NPC1L1 knockout protects against colitis-associated tumorigenesis in mice. BMC Cancer 2015;15:189.

32. Du Q, Wang Q, Fan H, Wang J, Liu X, Wang H, Wang Y, Hu R. Dietary cholesterol promotes AOM-induced colorectal cancer through activating the NLRP3 inflammasome. Biochem Pharmacol 2016;105:42-54.

33. Jung YS, Park CH, Eun CS, Park DI, Han DS. Statin Use and the Risk of Colorectal Adenoma: A Meta-Analysis. Journal of gastroenterology and hepatology 2016.

34. Murai T. Cholesterol lowering: role in cancer prevention and treatment. Biol Chem 2015;396:1-11.

35. Frattini M, Balestra D, Suardi S, Oggionni M, Alberici P, Radice P, Costa A, Daidone MG, Leo E, Pilotti S, Bertario L, Pierotti MA. Different genetic features associated with colon and rectal carcinogenesis. Clin Cancer Res 2004;10:4015-21.

FIGURES and TABLES

Figure 1: Meta-analysis odds ratios (OR) for colorectal cancer per unit increase in genetic risk score (SD trait) for each lipid trait. TC: Total cholesterol, TG: Triglyceride, LDL: low density lipoprotein, HDL: high density lipoprotein; Horizontal lines: 95% Confidence Intervals (95% CI). P_{het}: *P*-value for heterogeneity; I^2: proportion of the total variation due to heterogeneity. Box: OR point estimate; its area is proportional to the weight of the study. Diamond: overall summary estimate, with confidence interval given by its width. Vertical line: null value (OR = 1.0).

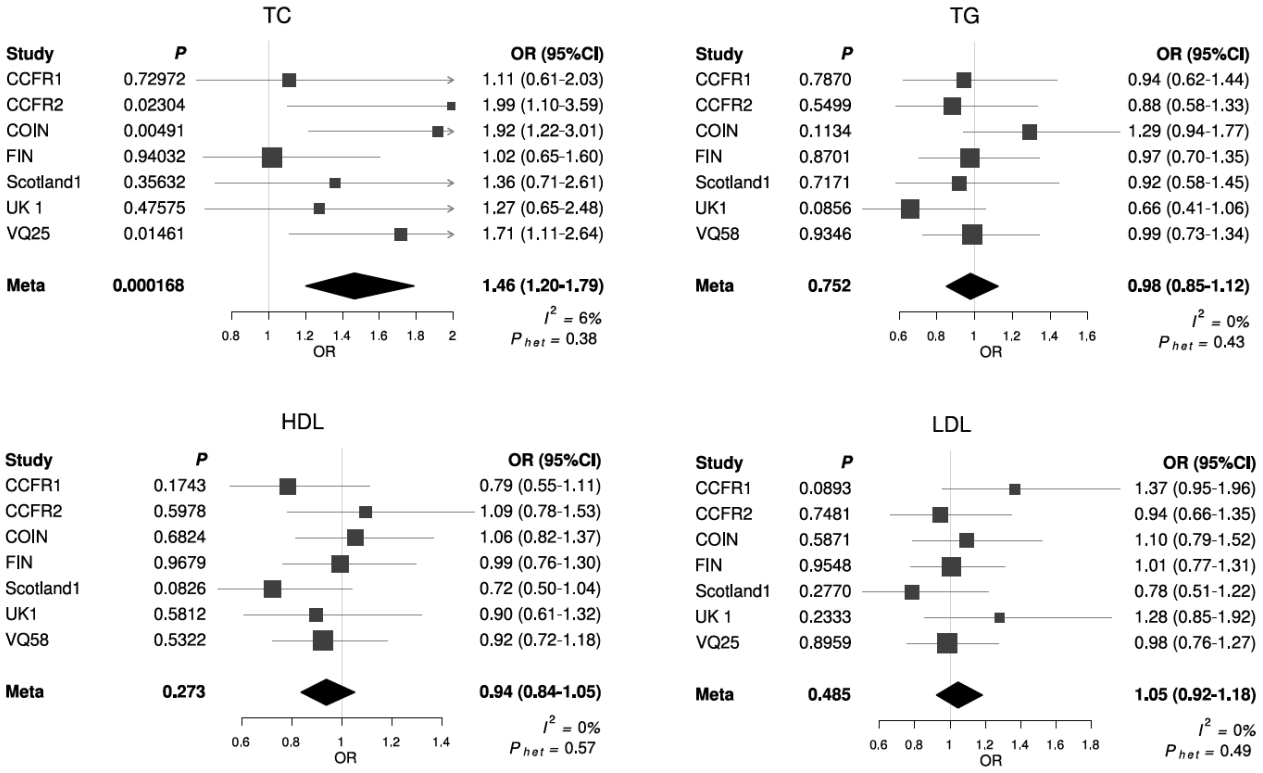

Table 1: Summary of the seven genome-wide association studies of colorectal cancer (9,254 cases and 18,386)

Series	Study setting	Study centre	Sampling	No. cases	No. controls
CCFR1	Colon Cancer Family Registry	University of Southern California	Recently diagnosed cases reported to population-based cancer registries in the USA (Seattle Familial Colorectal Cancer Registry). Canada (Ontario Familial Cancer Registry) and Australia (Australasian Colorectal Cancer Family Study). Population-based controls.	1,290	1,055
CCFR2	Colon Cancer Family Registry	University of Southern California	Recently diagnosed cases reported to population-based cancer registries in the USA (Seattle Familial Colorectal Cancer Registry, Mayo Clinic Cooperative Family Registry for Colon Cancer Studies, USC Consortium Colorectal Cancer Family Registry, University of Hawaii Colorectal Cancer Family Registry). Canada (Ontario Familial Cancer Registry) Australia (Australasian Colorectal Cancer Family Study). Unaffected family controls.	796	2,236
COIN	COIN trial: Multicentre study of cetuximab and other therapies in metastatic CRC. Controls were unselected blood donors	Cardiff University	Cases recruited as a clinical-based series and controls as population-based series.	2,244	2,162
FINLAND	Finnish Colorectal Cancer Predisposition Study	Helsinki University	Cases requited through Finnish Hospitals and Finnish Cancer Registry. Population-based controls from FINRISK, Health 2000, Finnish Twin Cohort and Helsinki Birth Cohort Studies.	1,172	8,266
UK1	CORGI (colorectal Tumour Gene Identification Consortium)	Oxford University	Cases enriched for family history of CRC, ascertained through UK clinical genetics clinics. Spouse controls with no personal history or family history of CRC.	940	965
Scotland1	COGS (Colorectal Cancer Susceptibility Study)	Edinburgh University	Scottish population-based incidence cases aged <55 at diagnosis. Population-based controls frequency matched by area of residence. Scotland	1,012	1,012
VQ58	Cases: VICTOR, post treatment stager of a phase III, randomised trial of rofecoxib (VIOXX) in patients after potentially curative therapy. QUASAR2, multi-centre study of capectibine±bevacizumb as adjuvant treatment. 1958 Birth Cohort controls	Oxford University	Cases recruited as a clinical-based series and controls as population-based series.	1,800	2,690

Table 2: Testing for global and instrumental-specific pleiotropism. Point estimates, confidence intervals, and *P*-values from linkage disequilibrium (LD) regression analysis, and MR-Egger methods. For MR-Egger, the intercept represents the average pleiotropic effect; an intercept significantly different from zero implies directional pleiotropy.

LD regression results

Trait	Heritability estimate	Genetic correlation	Standard error	*P*-value
TC	0.2408	0.049	0.0635	0.4402
TG	0.2939	0.0322	0.0639	0.6143
LDL	0.2122	0.0729	0.066	0.2696
HDL	0.2499	-0.0603	0.563	0.2834

MR-Egger regression results

Trait		Estimate	Corrected standard error	CI lower	CI upper	*P*-value
TC	intercept	1.11×10^{-2}	1.25×10^{-2}	-1.42×10^{-2}	3.64×10^{-2}	0.38
	slope	0.16	0.33	-0.51	0.83	0.64
TG	intercept	-1.13×10^{-2}	1.10×10^{-2}	-3.38×10^{-2}	1.12×10^{-2}	0.31
	slope	5.65×10^{-2}	0.17	-0.30	0.42	0.75
LDL	intercept	-3.41×10^{-3}	7.67×10^{-3}	-1.91×10^{-2}	1.23×10^{-2}	0.66
	slope	0.10	0.11	-0.11	0.32	0.34
HDL	intercept	2.23×10^{-3}	5.58×10^{-3}	-8.94×10^{-3}	1.34×10^{-2}	0.69
	slope	-0.11	0.11	-0.32	0.11	0.31