Modifiable pathways for colorectal cancer: A Mendelian randomisation analysis

Alex J. Cornish (Ph.D.)[1], Philip J. Law (Ph.D.)[1], Maria Timofeeva (Ph.D.)[2], Kimmo Palin (Ph.D.)[3], Susan M. Farrington (Ph.D.)[2], Claire Palles (Ph.D.)[4], Mark A. Jenkins (Ph.D.; Professor)[5], Graham Casey (Ph.D.; Professor)[6], Hermann Brenner (Ph.D.; Professor)[7,8,9], Jenny Chang-Claude (Ph.D.; Professor)[10,11], Michael Hoffmeister (Ph.D.)[7], Iva Kirac (M.D.)[12], Tim Maughan (Ph.D.; Professor)[13], Stefanie Brezina (Ph.D.)[14], Andrea Gsur (Ph.D.; Professor)[14], Jeremy P. Cheadle (Ph.D.; Professor)[15], Lauri A Aaltonen (M.D.; Professor)[3], Ian Tomlinson (Ph.D.; Professor)[16], Malcolm G. Dunlop (M.D.; Professor)[2], Richard S. Houlston (M.D.; Professor)[1]

1. Division of Genetics and Epidemiology, The Institute of Cancer Research, London, SM2 5NG, UK.
2. Cancer Research UK Edinburgh Centre and MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, UK.
3. Department of Medical and Clinical Genetics, Medicum and Genome-Scale Biology Research Program, Research Programs Unit, University of Helsinki, Helsinki, Finland.
4. Gastrointestinal Cancer Genetics Laboratory, Institute of Cancer and Genomic Sciences, University of Birmingham, Vincent Drive, Edgbaston, Birmingham, B15 2TT, UK.
5. Centre for Epidemiology and Biostatistics, The University of Melbourne, Melbourne, Vic. 3010, Australia.
6. Center for Public Health Genomics, University of Virginia, Virginia, VA 22903, USA.
7. Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany.
8. German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany.
9. Division of Preventive Oncology, German Cancer Research Center (DKFZ), National Center for Tumor Diseases (NCT), Heidelberg, Germany.

10. Unit of Genetic Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany.
11. Cancer Epidemiology Group, University Medical Center Hamburg-Eppendorf, University Cancer Center Hamburg, Hamburg, Germany.
12. Department of Surgical Oncology, University Hospital for Tumours, Sestre milosrdnice University Hospital Centre, Zagreb, Croatia.
13. Department of Oncology, Old Road Campus Research Building, University of Oxford, Oxford, OX3 7LE, UK.
14. Institute of Cancer Research, Department of Medicine I, Medical University of Vienna, Borschkegasse 8a, 1090 Vienna, Austria.
15. Institute of Medical Genetics, School of Medicine, Cardiff University, Heath Park, Cardiff, CF14 4XN, UK.
16. Edinburgh Cancer Research Centre, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, UK.

‡ Correspondence to: Alex J. Cornish, Division of Genetics and Epidemiology, The Institute of Cancer Research, 15 Cotswold Road, London, SM2 5NG, UK. Tel: +44 (0) 20 8722 4568; E-mail: alex.cornish@icr.ac.uk.

SUMMARY

Background: Epidemiological studies have linked lifestyle, cardiometabolic, reproductive, developmental and inflammatory factors with colorectal cancer (CRC) risk. However, it is unclear which specific factors influence risk and the strength of effects.

Methods: Under a random-effects model we examined the relationship between 39 potentially modifiable risk factors and CRC using genetic variants as instruments using two-sample Mendelian randomisation (MR), thereby limiting bias from confounding and reverse causation. Using genetic data on 26,397 CRC patients and 41,481 controls, we calculated odds ratios of CRC risk per genetically predicted standard deviation unit increase in each putative risk factor (OR_{SD}). Evidence of MR assumption violation was sought using MR-Egger regression. A Bonferroni-corrected threshold of $P=1.3x10^{-3}$ was considered significant, and $P<0.05$ considered suggestive of an association.

Findings: No putative risk factors were significantly associated with CRC risk after correction for multiple testing. Suggestive associations were however seen between genetically predicted body fat percentage (OR_{SD}=1.14, 95% confidence interval [CI]=1.03-1.25, *P*=0.0086), BMI (OR_{SD}=1.09, 95% CI=1.01-1.17, *P*=0.023), waist circumference (OR_{SD}=1.13, 95% CI=1.02-1.26, *P*=0.018) and basal metabolic rate (OR_{SD}=1.10, 95% CI=1.03-1.18, *P*=0.0079) with higher CRC risk. Low-density lipoprotein cholesterol level (OR_{SD}=1.14, 95% CI=1.04-1.25, *P*=0.0056) and circulating serum iron (OR_{SD}=1.17, 95% CI=1.00-1.36, *P*=0.049) also showed suggestive associations with increased CRC risk. A suggestive association was observed between serum vitamin B12 concentration and increased CRC risk (OR_{SD}=1.21, 95% CI=1.04–1.42, *P*=0.016), although potential pleiotropy amongst genetic variants used as instruments for this factor constrains the finding. Low blood selenium concentration also showed suggestive association with CRC (OR_{SD}=0.85, 95% CI=0.75-0.96, *P*=0.0078), albeit based on a single variant. CRC risk was not associated with any

reproductive factor, serum calcium or circulating 25-hydroxyvitamin D concentrations.

Interpretation: This analysis highlights a number of modifiable targets for primary prevention of CRC, including lifestyle, obesity and cardiometabolic factors that should inform public health policy.

RESEARCH IN CONTEXT

Evidence before this study: We searched PubMed to identify dietary, lifestyle, obesity-related, inflammatory, reproductive and developmental factors that had been assessed in observational epidemiological studies potentially influencing colorectal cancer (CRC) risk. Studies provide strong evidence for body mass index (BMI) and hypercholesterolaemia being associated with increased CRC risk. For most other factors there is inconclusive evidence from conventional observational studies to reliably establish specific associations.

Added value of this study: Mendelian randomisation exploits germline genetic variants as instrumental variables for putative risk factors. Because these genetic variants are randomly assorted at conception they are not influenced by reverse causation and so can provide evidence for causal relationships. We used genetic variants for 39 potentially modifiable CRC risk factors in 26,397 CRC patients and 41,481 controls. There was suggestive evidence for associations of serum vitamin B12, iron and selenium concentrations with CRC. In addition to providing suggestive evidence for a causal relationship between higher BMI and increased CRC risk, we found evidence for an association between genetically predicted low-density lipoprotein with risk of CRC. No associations with CRC risk were identified for any reproductive factor.

Implications of all the available evidence: These data provide two main findings: Firstly, genetic corroboration of causal relationships between higher BMI and hypercholesterolaemia and elevated CRC risk. Secondly, findings support the assertion that vitamin B12 supplementation should be limited to individuals with a known indication, such as proven deficiency. Our analysis highlights important targets for primary prevention of CRC, including lifestyle, obesity and cardiometabolic factors.

INTRODUCTION

Colorectal cancer (CRC) is the third most common diagnosed malignancy and the second leading cause of cancer-related death in the world, accounting for around 1.8 million new cases and 860,000 deaths in 2018(1). Based on current demographic trajectories, it is projected that the global burden of CRC will increase by 72% to over 3 million new cases and by 82% to 1.6 million cancer deaths annually by 2040(1). Differences in CRC incidence between countries and migration studies have implicated dietary and other lifestyle factors in CRC development(2). In view of this there is increasing interest in developing public health programs to reduce CRC incidence by targeting modifiable risk factors.

The World Cancer Research Fund (WCRF) and The American Institute for Cancer Research (AICR) have concluded that there is convincing evidence for body mass index (BMI) and alcohol intake being causally associated with increased CRC risk, and physical activity being causally associated with reduced CRC risk(3). Furthermore, it is probable that red meat intake is causally associated with increased CRC risk, whereas dietary fibre, dairy products and calcium supplements are causally associated with a lower risk(3). For most other factors there is inconclusive evidence from these conventional observational studies to reliably establish associations(3).

Much of the available evidence for a causal relationship between potentially modifiable factors and CRC risk is derived from observational studies(3), which are susceptible to confounding bias and reverse causation(4). Moreover, data from randomised trials tend to be scarce and often inconclusive(5, 6). Finally, establishing which specific components of risk factors such as diet are important is notoriously problematic in conventional observational epidemiological studies(7).

Mendelian randomisation (MR) is an analytical approach, whereby germline genetic variants are used as proxies, or instrumental variables, for putative risk factors(8). Because these genetic variants are randomly assorted at conception they are not influenced by reverse causation, and in the absence of pleiotropy (*i.e.* genetic

variants being associated with the disease through alternative pathways) they can provide unconfounded estimates of disease risk(8). Since MR-based studies can circumvent many limitations of conventional observational studies the methodology is increasingly being employed as an effective strategy to examine the potential impact of interventions on disease risk.

We have investigated potentially causal and modifiable CRC risk factors using a two-sample MR framework (**Supplementary Figure 1**) whereby genetic variants associated with relevant risk factors as instrumental variables were first identified from genome-wide association studies (GWAS). We then evaluated the association of these instrumental variables with CRC in a large GWAS comprising 26,397 cases of CRC and 41,481 control subjects(9).

METHODS

Identification of potentially modifiable risk factors

As well as evaluating dietary, lifestyle, obesity-related, inflammatory, reproductive and developmental factors that had been the subject of the report by the WCRF and AICR(3), we also searched PubMed to identify additional modifiable CRC risk factors that have been reviewed in published epidemiological meta-analyses or MR analyses (**Supplementary Table 1**; **Supplementary Information**).

Genetic instruments for putative risk factors

Single nucleotide polymorphisms (SNPs) associated with putative risk factor traits suitable for use in MR analysis were identified from the largest GWAS or meta-analysis of each trait conducted to date (**Table 1; Supplementary Table 2**; **Supplementary Information**). Traits were only considered if the proportion of variance explained (PVE) by the associated SNPs was >0.1%. PVE estimates were either obtained from the publication or computed directly from the association statistics (**Table 1**)(10). Suitable genetic instruments were not available for many risk factors, such as physical activity, dietary patterns and vitamin C intake, precluding their inclusion in this study (**Supplementary Table 1**). We considered only continuous traits, as analysis of binary traits (such as disease status) with binary outcomes in two-sample MR frameworks can result in inaccurate causal estimates(11). Only SNPs associated with each trait at $P<5\times10^{-8}$ in GWAS of European populations with a minor allele frequency >0.01 were considered as potential instruments. To mitigate against co-linearity between SNPs, which can bias causal effect estimates, we used MR-Base to exclude correlated SNPs at a linkage disequilibrium threshold of $r^2>0.01$, retaining those SNPs with the strongest effect on the associated trait(12).

Colorectal cancer genotyping data

To examine the association of each genetic instrument with CRC risk, we used summary CRC effect estimates and corresponding standard errors (SEs) from a recent meta-analysis of 15 CRC GWAS(9). After imputation, this meta-analysis

related >10 million genetic variants to CRC in individuals of European ancestry. UK BioBank data were used to obtain genetic instruments for age at menarche, basal metabolic rate, birth weight, body fat percentage and waist circumference, as well as in one of the CRC GWAS meta-analysed by Law *et al.*(9). To avoid sample overlap biasing this two-sample MR analysis(13) we therefore excluded the UK BioBank CRC GWAS and recomputed association statistics using the remaining 14 CRC GWAS (**Supplementary Table 3**) with an inverse variance weighted (IVW) fixed-effects model, as described by Law *et al.*(9). After exclusion of the UK BioBank CRC GWAS, the meta-analysis comprised 26,397 patients and 41,481 controls. SNPs with poor imputation quality (*i.e*. info score <0.8) were not considered in the MR analysis. As some potentially modifiable reproductive risk factors are female-specific, where sex data were available we further computed CRC association statistics using only 7,952 female cases and 11,680 female controls. We used MR-Base to harmonize SNPs to ensure that the effect estimates of each SNP on each trait and CRC risk corresponded to the same allele(12). Effect estimates for the association of each trait SNP with CRC risk are shown in **Supplementary Table 2**. For vitamins, positive beta values indicate that the effect allele is associated with increased serum concentration.

Statistical analysis

The MR methodology is predicated on the assumption that genetic variants used as instruments for a risk factor are associated with the risk factor and not with a confounder or alternative causal pathway (**Figure 1**). Additionally, to accurately estimate the size of the causal effect, the associations depicted in **Figure 1** must be linear and unaffected by statistical interactions(14). We estimated causal effects for each SNP using the Wald ratio (**Supplementary Figure 2**). For traits with multiple SNPs available as instruments, causal effects were estimated using the random-effects maximum likelihood estimation (MLE-RE) method(15). To assess the robustness of our findings, we also obtained weighted median estimates (WME)(16) and mode-based estimates (MBE)(17). We used the MR-Egger regression approach to evaluate the extent to which directional pleiotropy may affect the causal estimates(18). Finally, we conducted leave-one-out analysis using the multiplicative

random-effects inverse variance weighted method(12) to examine the impact of outlying and pleiotropic SNPs on causal estimates (**Supplementary Table 4**). I^2 statistics were computed to estimate the proportion of variance across SNPs due to heterogeneity (**Figure 2; Supplementary Table 5**)**.** Results are reported as odds ratios (OR_{SD}) and 95% confidence intervals (CIs) per genetically predicted standard deviation (SD) unit increase in each putative risk factor. To address the issue of multiple testing, we applied a Bonferroni-corrected significance threshold computed as 0.0013 (*i.e*. 0.05/39 putative risk factors). $0.0013<P<0.05$ was considered as suggestive of a potential association. The power of MR to demonstrate a causal effect depends on the proportion of variance in the risk factor explained by the genetic variants used as instruments, and we therefore estimated study power at an alpha of 0.05 for each risk factor *a priori* (**Table 1**)(19). Statistical analyses were performed using R v3.4.0 and MR analyses were performed using MR-Base(12).

Role of the funding sources

Funders had no role in study design, in the collection, analysis and interpretation of data, or in writing the report. The corresponding author had full access to all of the data and the final responsibility to submit for publication.

RESULTS

Diet and lifestyle factors

Under a MLE-RE model, a suggestive association was seen between genetically predicted serum vitamin B12 concentration and higher CRC risk (OR_{SD}=1.21, 95% CI=1.04–1.42, *P*=0.016), however substantial heterogeneity exists between the SNPs used as IVs (I^2=79.1). Leave-one-out analysis identified SNP rs602662 at a known CRC risk locus as having a strong influence on the causal estimate (**Supplementary Table 4**)(9). Expression quantitative trait loci analysis has indicated that variation at this potentially pleiotropic locus may influence CRC risk through *FUT2* and interactions with intestinal bacteria and viruses(9). There was a suggestive association between genetically predicted greater serum iron concentration and higher CRC risk (OR_{SD}=1.17, 95% CI=1.00-1.36, *P*=0.049), with no outlying genetic variant identified (**Supplementary Figure 2**). There was also a suggestive association between higher serum selenium concentration and lower CRC risk (OR_{SD}=0.85, 95% CI=0.75-0.96, *P*=0.0078) albeit based on only one SNP. Genetically predicted alcohol and coffee consumption, and blood methionine, zinc, 25-hydroxyvitamin D, carotenoids, calcium and vitamins A (retinol), B6 and E concentrations, showed no evidence for association with CRC risk (**Figure 2**). Causal effect estimates for serum vitamin B12 concentration was similar in sensitivity analyses using the WME and MBE methods (**Supplementary Table 5**). MR-Egger regression showed no evidence of directional pleiotropy in the analyses of vitamin B12 or serum iron concentration (**Supplementary Table 6**). The causal effects estimated by MR-Egger were non-significant for vitamin B12 (**Supplementary Table 5**), possibly as a result of the reduced power of MR-Egger to detect causal effects when compared to other MR methodologies(18).

Fatty acid profile and metabolism

Fatty acid (FA) metabolism involves sequential enzymatic conversions (**Supplementary Figure 3**), and SNPs influencing the metabolism of one FA are therefore often associated with circulating concentrations of multiple FAs(20). Additionally, many genes involved in FA desaturation and elongation form parts of

numerous FA pathways, and hence influence circulating concentrations of multiple classes of FA (**Supplementary Figure 3**). To limit bias introduced by such vertical and horizontal pleiotropy, we restricted our analysis to classes of FAs (such as omega-6 polyunsaturated FAs [PUFAs] and monounsaturated FAs [MUFAs]), rather than individual fatty acids, and excluded SNPs known to be associated with multiple FA classes (**Supplementary Table 7**). In this restricted analysis, no association were observed for omega-6 PUFA or MUFA concentrations, or for blood levels of the fatty acid transport molecule carnitine (**Figure 2**). After removal of potentially pleiotropic SNPs, only a single SNP was suitable for use as an instrumental variable for MUFA concentration, prohibiting sensitivity analysis using WME and MBE approaches.

Cardiometabolic and inflammatory factors

Using information on all genetic variants associated with cardiometabolic factors, we observed that measures of obesity and hyperlipidaemia were suggestively associated with CRC (**Figure 2**). Specifically, suggestive associations were seen between genetically predicted basal metabolic rate (OR_{SD}=1.10, 95% CI=1.03-1.18, *P*=0.0079), body fat percentage (OR_{SD}=1.14, 95% CI=1.03-1.25, *P*=0.0086), BMI (OR_{SD}=1.09, 95% CI=1.01-1.17, *P*=0.023) and waist circumference (OR_{SD}=1.13, 95% CI=1.02-1.26, *P*=0.018), and higher odds of CRC. No association between birth weight or adiponectin levels and CRC risk was seen (**Figure 2**). Causal estimates for basal metabolic rate, BMI and waist circumference were broadly concordant in sensitivity analyses using the WME and MBE methods (**Supplementary Table 5**). Conversely, the effect estimate for body fat percentage from the MBE approach (OR_{SD}=0.98, 95% CI=0.72-1.33, *P*=0.90) differed in direction to the estimates from other MR implementations (**Supplementary Table 5**), suggesting that some of the instruments used to assess the causal effects of body fat percentage may be invalid. MR-Egger regression did not identify evidence of horizontal pleiotropy for body fat percentage or any other obesity-related trait (**Supplementary Table 6**).

Genetically predicted low-density lipoprotein (LDL) cholesterol (OR_{SD}=1.14, 95% CI=1.04-1.25, *P*=0.0056) and total cholesterol (OR_{SD}=1.09, 95% CI=1.01-1.18, *P*=0.025) showed suggestive associations with higher odds of CRC. No association

between high-density lipoprotein (HDL) cholesterol or total triglyceride levels was seen (**Figure 2**). Similarly, genetically predicted metrics of glycaemia - fasting glucose, fasting proinsulin, and HbA1c - were not associated with CRC risk (**Figure 2**).

Based on a single SNP, a suggestive association was observed between plasma levels of interleukin 6 (IL-6) receptor subunit alpha and lower CRC risk (OR_{SD}=0.98, 95% CI=0.98-1.00, *P*=0.035). Associations between circulating C-reactive protein and serum immunoglobulin E and CRC risk were not however demonstrated (**Figure 2**).

Sex hormones and reproduction

It has been hypothesised that sex-specific differences in CRC incidence may be partly attributable to differential sex hormone exposure(21). However, we observed no association between age at menarche, a surrogate for endogenous estrogen exposure, and CRC risk (OR_{SD}=0.99, 95% CI=0.84-1.18, *P*=0.92) using CRC data from females only. Similarly, we did not observe associations between plasma estradiol and progesterone and CRC risk in sex-specific analyses (**Figure 2**). The genetic variants used as instruments for these traits explain only a small proportion of their variance (**Table 1**) and we are therefore unable to exclude a small to moderate effect of sex hormone exposure on CRC risk. MR-Egger regression analysis of genetic instruments for age at menopause provided evidence of horizontal pleiotropy (*P*=0.01; **Supplementary Table 6**) and we therefore did not consider this trait in our MR analysis.

Developmental and growth factors

Whilst height is not modifiable once stabilised in adulthood, it is influenced by developmental factors and growth processes, which may themselves be modifiable. In concordance with evidence reviewed by the WCRF and AICR(3), we observed a suggestive association between greater genetically predicted adult height and increased odds of CRC (OR_{SD}=1.04, 95% CI=1.00-1.08, *P*=0.032), further supporting the notion that factors during childhood may influence CRC risk. We observed no association between plasma insulin-like growth factor 1 (IGF-1) and CRC risk (**Figure 2**), although this analysis was conducted using a single genetic variant explaining

only a small proportion of IGF-1 variance, and therefore had limited power to detect an effect (**Table 1**).

DISCUSSION

With genetic variants as proxies for the putative risk factors, this MR study provides suggestive evidence for associations between higher body fat percentage, BMI, waist circumference and basal metabolic rate and increased CRC risk. We also found suggestive evidence for associations between genetically predicted LDL and total cholesterol and risk of CRC, but no evidence of associations with HDL or total triglyceride levels. The suggestive association between genetically determined higher serum vitamin B12 levels and increased CRC risk is intriguing. There was also suggestive evidence for possible associations of genetically predicted serum iron and selenium concentrations.

Strengths of this study include examination of multiple factors in relation to CRC risk, by exploiting data from large GWAS of risk factors and CRC. Many of the putative risk factors considered in this study have not previously been assessed using MR frameworks (**Supplementary Table 8**). Of those factors for which suggestive associations were seen (**Figure 2**), body fat percentage, waist circumference, basal metabolic rate, iron status, and blood selenium, serum vitamin B12 and plasma IL-6 subunit alpha concentrations have not previously been considered in MR analyses of CRC risk (**Supplementary Table 8**). For those CRC risk factors that have previously been considered in MR analyses(22) the number of CRC cases and controls we consider here affords us greater power to detect causal relationships and allows us to more accurately estimate effect magnitudes. For example, while Rodriguez-Broadbent *et al.(23)* reported a non-significant association between LDL cholesterol and risk of CRC (OR_{SD}=1.05, 95% CI=0.92-1.18, *P*=0.49), herein a suggestive relationship was identified (OR_{SD}=1.14, 95% CI=1.04-1.25, *P*=0.0056), possibly due to increased power of the present analysis. By comparing the results of this study to those of previous MR analyses of CRC risk we are also able to identify previously reported causal relationships that may represent false positives, such as an association between genetically predicted C-reactive protein concentrations and CRC risk(24) (**Figure 2; Supplementary Table 8**).

Although F-statistics were high (>10) for all considered traits (**Table 1**), we cannot exclude the possibility that some of our findings may have been affected by weak instrument bias. For 19 of the traits for which we identified no association with CRC risk, our study had <80% power to identify OR_{SD}<0.91 or >1.10 (**Table 1**), and we are therefore unable to exclude the possibility that these traits have a small effect on CRC risk.

As with all MR studies, excluding pleiotropy or an alternative direct causal pathway as the basis of association is a challenge. High I^2 statistics for many traits indicate the presence of such pleiotropy in this analysis (**Figure 2**). To address this issue we implemented the WME and MBE methods, which can provide unbiased causal effect estimates even when many genetic variants used represent invalid instruments(16, 17). For the majority of traits with either a significant or suggestive association with CRC risk, the effects estimated were similar using MLE-RE, WME and MBE methods (**Supplementary Table 5**), supporting causal relationships with CRC. It is important to note that there exists overlap between the CRC cases and controls considered in this study, and those considered in some previous MR analyses(22), and that results from this study therefore cannot be considered independent replication.

Our study provides no evidence for an association between genetically predicted fasting glucose and proinsulin and risk of CRC, suggesting that metabolic syndrome may not influence CRC risk through these factors. However, due to the limited power of this analysis, we cannot preclude these factors having small effects on CRC risk (**Table 1**).

Our estimate that an SD increment in adult height increases CRC risk by 4% is concordant with many observational studies(3), with greater exposure to growth hormones and insulin-like growth factors during childhood being posited as potential mechanisms for this association(26). Whilst we observed no significant association between plasma IGF-1 and CRC risk (**Figure 2**), the limited power of this analysis means that we are unable to exclude small to moderate effect sizes (**Table 1**). Taller

adults tend to have larger colons, and so greater at-risk cell populations might also explain the apparent causal inference.

Of the nutritional factors analysed, a relationship between genetically predicted vitamin B12 levels and CRC risk was shown (**Supplementary Table 5**). Our findings are concordant with a randomized trial that found vitamin B12 supplementation increases CRC risk(29). Although less convincing, we also found suggestive evidence to support high selenium levels having a beneficial effect and greater iron status being detrimental (**Supplementary Table 5**).

Inevitably, further research is required to decipher the biological pathways underpinning associations. However, irrespective of the exact functional basis of associations using a genetic approach, our analysis highlights important targets for primary prevention of CRC in the population. Firstly, between obesity and CRC risk, the strong corroboration for obesity being a major risk factor for CRC supports reducing the population incidence of obesity a priority. Secondly, our findings are consistent with hypercholesterolemia being causally linked to risk and therefore support the hypothesis that the increasing use of statins in the population for prevention of cardiovascular disease will have the added bonus of reducing CRC burden. The limited power of this study to refine robustly the relationship between some putative risk factors provides motivation for larger MR studies to demonstrate relationships for the spectrum of colorectal neoplasia. Such work may shed additional light on other potentially modifiable factors to reduce the overall burden of CRC.

DECLARATIONS

Contributors

RSH managed the project. AJC and PJL performed statistical analyses. MAJ and GC acquired and analysed CCFR1 and CCFR2 data. JPC and TM acquired and analysed COIN data. SB and AG acquired and analysed CORSA data. IK acquired and analysed Croatia data. HB, JC-C and MH acquired and analysed DACHS data. KP and LAA acquired and analysed FIN data. PJL and RSH acquired and analysed NSCCG-OncoArray and VQ58 data. CP and IT acquired and analysed SCOT data. MT, SMF and MGD acquired and analysed Scotland1, SOCCS/GS and SOCCS/LBC data. AJC, PJL and RSH drafted the manuscript. All authors reviewed, read and approved the final manuscript.

Declaration of interests

The authors declare that they have no competing interests

Acknowledgements

Support for this work was provided by grants from Cancer Research UK (C1298/A25514, C348/A12076, C6199/A16459, C348/A18927) and by a Project Leader grant within the MRC Human Genetics Unit Centre Grant, Edinburgh (U127527198). We also acknowledge support from the European Union COST Action (BM1206). This work was supported by grant UM1 CA167551 from the National Cancer Institute and through cooperative agreements with the following Colon Cancer Family Registry sites: Australasian Colorectal Cancer Family Registry (U01 CA074778 and U01/U24 CA097735); Ontario Familial Colorectal Cancer Registry (U01/U24 CA074783); and Seattle Colorectal Cancer Family Registry (U01/U24 CA074794).

Ethics approval and consent to participate

Two-sample MR was undertaken using GWAS data. Ethical approval was not sought for this specific project because all data came from the summary statistics of published GWAS, and no individual-level data were used.

Availability of data and material

Genetic instruments can be obtained through MR-Base (http://www.mrbase.org/; last accessed 5 August 2019) or from the individual referenced papers. Meta-analysed CRC GWAS data were obtained from the study by Law *et al.*(9).

FIGURES

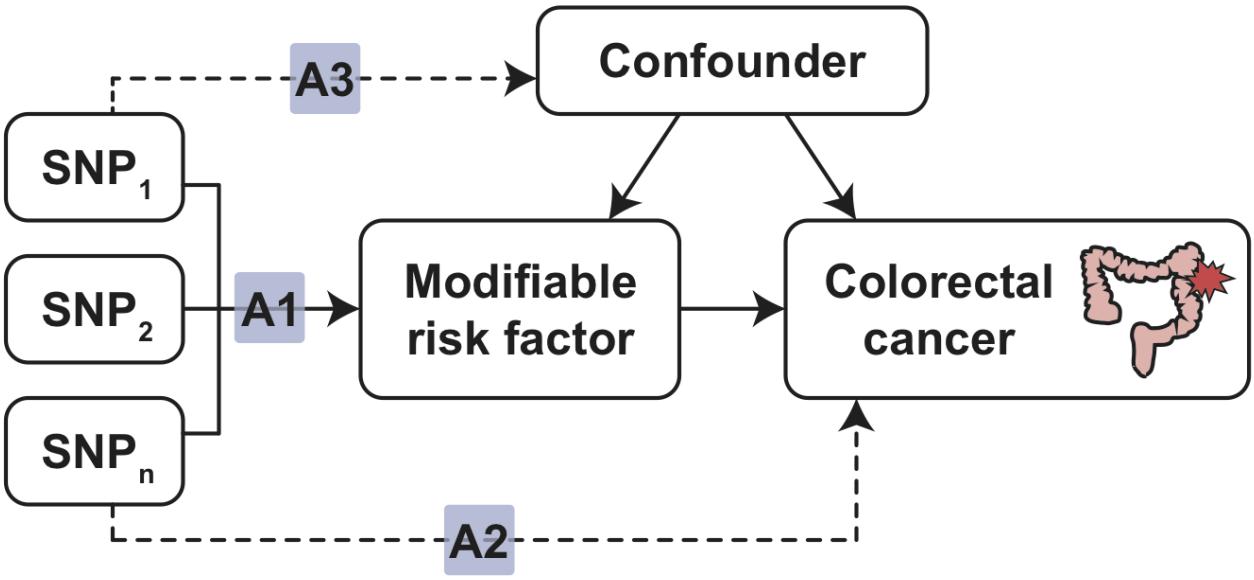

Figure 1: Principles of Mendelian randomisation and the assumptions that need to be satisfied to derive unbiased causal effect estimates. Dashed lines represent direct causal and potential pleiotropic effects that would violate Mendelian randomisation assumptions. A1: Genetic variants used as instrumental variables are associated with the risk factor; A2: Genetic variants influence the risk of colorectal cancer only through the risk factor; A3: Genetic variants are not associated with any measured or unmeasured confounders. SNP: single nucleotide polymorphism.

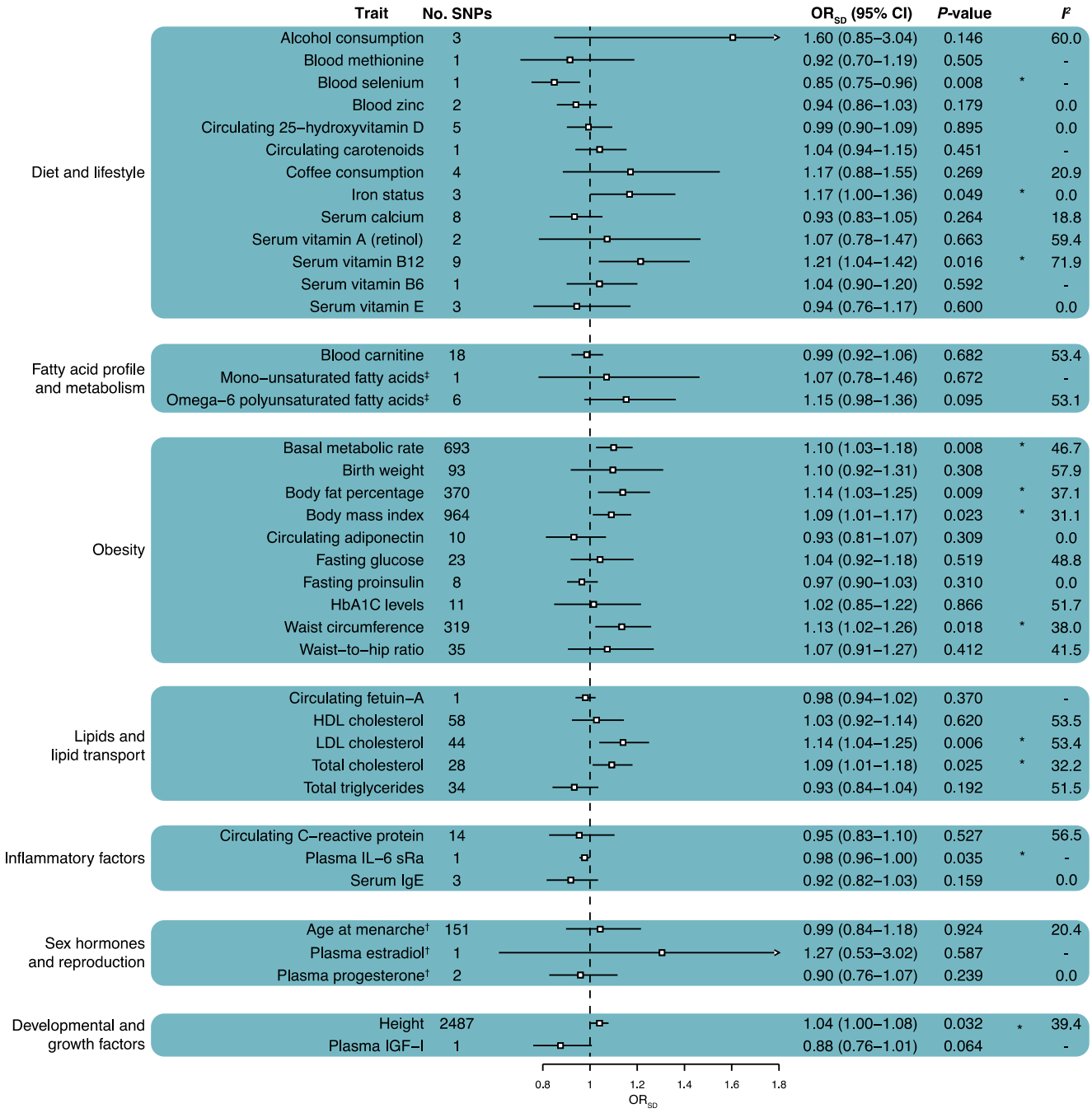

Figure 2: Odds ratios for associations between genetically predicted risk factors and colorectal cancer. Results reported as odds ratios (OR_{SD}) and 95% confidence intervals (CIs) per genetically predicted standard deviation (SD) unit increase in the risk factor. A maximum likelihood estimate random-effects (MLE-RE) method was used to summarize Wald ratio estimates from individual single nucleotide polymorphisms (SNPs). IL-6 sRa: interleukin 6 receptor subunit alpha; IGF: insulin-like growth factor. * $P < 0.05$; ‡ OR_{SD} from restricted analysis, which excludes SNPs known to be associated with other classes of fatty acid. † OR_{SD} computed using CRC data from female cases and controls.

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394-424.
2. Flood DM, Weiss NS, Cook LS, Emerson JC, Schwartz SM, Potter JD. Colorectal cancer incidence in Asian migrants to the United States and their descendants. Cancer Causes Control. 2000;11(5):403-11.
3. WCRF, AICR. Diet, nutrition, physical activity and colorectal cancer. 2018.
4. Fewell Z, Davey Smith G, Sterne JA. The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study. Am J Epidemiol. 2007;166(6):646-55.
5. Cole BF, Logan RF, Halabi S, Benamouzig R, Sandler RS, Grainge MJ, et al. Aspirin for the chemoprevention of colorectal adenomas: meta-analysis of the randomized trials. J Natl Cancer Inst. 2009;101(4):256-66.
6. Baron JA, Beach M, Mandel JS, van Stolk RU, Haile RW, Sandler RS, et al. Calcium supplements for the prevention of colorectal adenomas. Calcium Polyp Prevention Study Group. N Engl J Med. 1999;340(2):101-7.
7. Tarasuk VS, Brooker AS. Interpreting epidemiologic studies of diet-disease relationships. J Nutr. 1997;127(9):1847-52.
8. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet. 2014;23(R1):R89-98.
9. Law PJ, Timofeeva M, Fernandez-Rozadilla C, Broderick P, Studd J, Fernandez-Tajes J, et al. Association analyses identify 31 new risk loci for colorectal cancer susceptibility. Nat Commun. 2019;10(1):2154.
10. Shim H, Chasman DI, Smith JD, Mora S, Ridker PM, Nickerson DA, et al. A multivariate genome-wide association analysis of 10 LDL subfractions, and their response to statin treatment, in 1868 Caucasians. PLoS One. 2015;10(4):e0120758.
11. Disney-Hogg L, Cornish AJ, Sud A, Law PJ, Kinnersley B, Jacobs DI, et al. Impact of atopy on risk of glioma: a Mendelian randomisation study. BMC Med. 2018;16(1):42.

12. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife. 2018;7.

13. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. Genet Epidemiol. 2016;40(7):597-608.

14. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med. 2008;27(8):1133-63.

15. Yavorska OO, Burgess S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. Int J Epidemiol. 2017;46(6):1734-9.

16. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol. 2016;40(4):304-14.

17. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. Int J Epidemiol. 2017;46(6):1985-98.

18. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015;44(2):512-25.

19. Brion MJ, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. Int J Epidemiol. 2013;42(5):1497-501.

20. Wu JH, Lemaitre RN, Manichaikul A, Guan W, Tanaka T, Foy M, et al. Genome-wide association study identifies novel loci associated with concentrations of four plasma phospholipid fatty acids in the de novo lipogenesis pathway: results from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium. Circ Cardiovasc Genet. 2013;6(2):171-83.

21. Koo JH, Leong RW. Sex differences in epidemiological, clinical and pathological characteristics of colorectal cancer. J Gastroenterol Hepatol. 2010;25(1):33-42.

22. Cornish AJ, Tomlinson IPM, Houlston RS. Mendelian randomisation: A powerful and inexpensive method for identifying and excluding non-genetic risk factors for colorectal cancer. Mol Aspects Med. 2019.

23. Rodriguez-Broadbent H, Law PJ, Sud A, Palin K, Tuupanen S, Gylfe A, et al. Mendelian randomisation implicates hyperlipidaemia as a risk factor for colorectal cancer. Int J Cancer. 2017;140(12):2701-8.

24. Nimptsch K, Aleksandrova K, Boeing H, Janke J, Lee YA, Jenab M, et al. Association of CRP genetic variants with blood concentrations of C-reactive protein and colorectal cancer risk. Int J Cancer. 2015;136(5):1181-92.

25. Burgess S, Bowden J, Fall T, Ingelsson E, Thompson SG. Sensitivity Analyses for Robust Causal Inference from Mendelian Randomization Analyses with Multiple Genetic Variants. Epidemiology. 2017;28(1):30-42.

26. Gunnell D, Okasha M, Smith GD, Oliver SE, Sandhu J, Holly JM. Height, leg length, and cancer risk: a systematic review. Epidemiol Rev. 2001;23(2):313-42.

27. Bartsch H, Nair J, Owen RW. Dietary polyunsaturated fatty acids and cancers of the breast and colorectum: emerging evidence for their role as risk modifiers. Carcinogenesis. 1999;20(12):2209-18.

28. Wang D, DuBois RN. The role of anti-inflammatory drugs in colorectal cancer. Annu Rev Med. 2013;64:131-44.

29. Oliai Araghi S, Kiefte-de Jong JC, van Dijk SC, Swart KMA, van Laarhoven HW, van Schoor NM, et al. Folic Acid and Vitamin B12 Supplementation and the Risk of Cancer: Long-term Follow-up of the B Vitamins for the Prevention of Osteoporotic Fractures (B-PROOF) Trial. Cancer Epidemiol Biomarkers Prev. 2019;28(2):275-82.