medRxiv preprint doi: https://doi.org/10.1101/2023.03.10.23287084; this version posted November 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Identifying metabolic features of colorectal cancer liability using Mendelian randomization

Caroline J. Bull[1,2,3*], Emma Hazelwood[1,2*], Joshua A. Bell[1,2], Vanessa Y. Tan[1,2], Andrei-Emil Constantinescu[1,2], Maria Carolina Borges[1,2], Danny N. Legge[3], Kimberly Burrows[1,2], Jeroen R. Huyghe[4], Hermann Brenner[5,6,7], Sergi Castellví-Bel[8], Andrew T Chan[9,10,11,12,13,14], Sun-Seog Kweon[15,16], Loic Le Marchand[17], Li Li[18], Iona Cheng[19,20], Rish K. Pai[21], Jane C. Figueiredo[22], Neil Murphy[23], Marc J. Gunter[23,24], Nicholas J. Timpson[1,2], Emma E. Vincent[1,2,3].

[1] MRC Integrative Epidemiology Unit at the University of Bristol, Bristol, UK

[2] Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

[3] Translational Health Sciences, Bristol Medical School, University of Bristol, UK

[4] Public Health Sciences Division, Fred Hutchinson Cancer Center, Seattle, Washington, USA

[5] Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany

[6] Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany

[7] German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany

[8] Gastroenterology Department, Hospital Clínic, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), University of Barcelona, Barcelona, Spain

[9] Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA

[10] Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA

[11] Clinical and Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA

[12] Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA

[13] Department of Epidemiology, Harvard T.H. Chan School of Public Health, Harvard University, Boston, Massachusetts, USA

[14] Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Harvard University, Boston, Massachusetts, USA

[15] Department of Preventive Medicine, Chonnam National University Medical School, Gwangju, Korea

[16] Jeonnam Regional Cancer Center, Chonnam National University Hwasun Hospital, Hwasun, Korea

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2023.03.10.23287084; this version posted November 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

[17] University of Hawaii Cancer Center, Honolulu, Hawaii, USA

[18] Department of Family Medicine, University of Virginia, Charlottesville, Virginia, USA

[19] Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, California, USA

[20] University of California, San Francisco Helen Diller Family Comprehensive Cancer Center, San Francisco, San Francisco, California, USA

[21] Department of Pathology and Laboratory Medicine, Mayo Clinic, Arizona, Scottsdale, Arizona, USA

[22] Department of Medicine, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA

[23] Nutrition and Metabolism Branch, International Agency for Research on Cancer, Lyon, France

[24] Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, United Kingdom

* These authors contributed equally and are joint first authors.

Corresponding author: Emma E. Vincent, emma.vincent@bristol.ac.uk

Abstract

Background: Recognizing the early signs of cancer risk is vital for informing prevention, early detection, and survival.

Methods: To investigate whether changes in circulating metabolites characterise the early stages of colorectal cancer (CRC) development, we examined associations between a genetic risk score (GRS) associated with CRC liability (72 single nucleotide polymorphisms) and 231 circulating metabolites measured by nuclear magnetic resonance spectroscopy in the Avon Longitudinal Study of Parents and Children (N=6,221). Linear regression models were applied to examine associations between genetic liability to colorectal cancer and circulating metabolites measured in the same individuals at age 8, 16, 18 and 25 years.

Results: The GRS for CRC was associated with up to 28% of the circulating metabolites at FDR-P<0.05 across all time points, particularly with higher fatty acids and very-low- and low-density lipoprotein subclass lipids. Two-sample reverse Mendelian randomization (MR) analyses investigating CRC liability (52,775 cases, 45,940 controls) and metabolites measured in a random subset of UK Biobank participants (N=118,466, median age 58y) revealed broadly consistent effect estimates with the GRS analysis. In conventional (forward)

medRxiv preprint doi: https://doi.org/10.1101/2023.03.10.23287084; this version posted November 9, 2023. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

MR analyses, genetically predicted polyunsaturated fatty acid concentrations were most strongly associated with higher CRC risk.

Conclusions: These analyses suggest that higher genetic liability to CRC can cause early alterations in systemic metabolism, and suggest that fatty acids may play an important role in CRC development.

Funding: This work was supported by the Elizabeth Blackwell Institute for Health Research, University of Bristol, the Wellcome Trust, the Medical Research Council, Diabetes UK, the University of Bristol NIHR Biomedical Research Centre, and Cancer Research UK. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. This work used the computational facilities of the Advanced Computing Research Centre, University of Bristol - http://www.bristol.ac.uk/acrc/.

Keywords: Colorectal cancer; Metabolism; Genetics; NMR; Epidemiology; Mendelian randomization; ALSPAC; UK Biobank; GECCO; CCTS; CCFR

Introduction

Colorectal cancer (CRC) is the third most frequently diagnosed cancer worldwide and the fourth most common cause of death from cancer.[1,2] There is a genetic component to risk of the disease, which is thought to explain up to 35% of variability in CRC risk.[3–5] In addition, modifiable lifestyle factors, including obesity, consumption of processed meat, and alcohol are thought to increase CRC risk.[2,6–9] However, the underlying biological pathways remain unclear, which limits targeted prevention strategies. Whilst CRC has higher mortality rates when diagnosed at later stages, early-stage CRC or precancerous lesions are largely treatable, meaning colorectal cancer screening programmes have the potential to be highly effective.[10,11] Due to the lack of known predictive biomarkers for CRC, wide-scale screening (if implemented at all) is expensive and often targeted crudely by age range. Identifying biomarkers predictive of CRC, or with causal roles in disease development, is therefore vital.

One potential source of biomarkers for CRC risk is the circulating metabolome, which offers a dynamic insight into cellular processes and disease states. It is increasingly clear from mechanistic studies that both systemic and intracellular tumour metabolism play an important role in CRC development and progression.[12,13] Interestingly, several major risk factors for CRC are known to have profound effects on metabolism.[14] For instance, obesity has been shown via conventional observational and Mendelian randomization (MR) analyses to strongly alter circulating metabolite levels.[9,15–17] This suggests that the circulating metabolome may play a mediating role in the relationship between at least some common risk factors, such as obesity, and CRC – or at least might be a useful biomarker for disease or intermediates thereof. In particular, previous work has highlighted polyunsaturated fatty acids (PUFA) as potentially having a role in colorectal cancer development. The term PUFA includes omega-3 and -6 fatty acids. Recent MR work has highlighted a possible link between PUFAs, in particular omega 6 PUFAs, and colorectal cancer risk.[18] Further investigating the relationship between CRC and circulating metabolites may therefore provide powerful insights into the causal pathways underlying disease risk, or alternatively may be valuable in prediction and early diagnosis.

MR is a genetic epidemiological approach used to evaluate causal relationships between traits.[19,20] This method uses genetic variation as a proxy measure for traits in an instrumental variable framework to assess the causal relevance of the traits in disease development. As germline genetic variants are theoretically randomised between generations and fixed at

conception, this approach should be less prone to bias and confounding than conventional analyses undertaken in an observational context. Conventionally, MR is used to investigate the effect of an exposure on a disease outcome. In reverse MR, genetic instruments proxy the association between liability to a disease and other traits.[21] This approach can identify biomarkers which cause the disease, are predictive for the disease, or have diagnostic potential.[21] Given the suspected importance of the circulating metabolome in CRC development, employing both reverse MR and conventional forward MR for metabolites in the same study may be an efficient approach for revealing causal and predictive biomarkers for CRC. Although previous observational studies have investigated associations between the circulating metabolome and CRC risk, these studies may have been influenced by confounding bias which should be less relevant to MR analyses.[22–31] Additionally, these studies focussed on adults, who commonly take medications which may confound metabolite associations, further complicating interpretations.

Here, we applied a reverse MR framework to identify circulating metabolites which are associated with CRC liability across different stages of the early life course (spanning childhood to young adulthood, when use of medications and CRC are both rare) using data from a birth cohort study. We then attempted to replicate these results using reverse two-sample MR in an independent cohort of middle-aged adults (UK Biobank). We then performed conventional 'forward' MR of metabolites onto CRC risk using large-scale cancer consortia data to identify metabolites which may have a causal role in CRC development.

Methods

Study populations

This study uses data from 2 cohort studies: the Avon Longitudinal Study of Parents and Children (ALSPAC) offspring (generation 1) cohort (individual-level data) and the UK Biobank cohort (summary-level data); plus summary-level data from a genome-wide association study (GWAS) meta-analysis of CRC comprising the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO), Colorectal Transdisciplinary Study (CORECT), and Colon Cancer Family Registry (CCFR).

ALSPAC is a population-based birth cohort study in which 14,541 pregnant women with an expected delivery date between 1 April 1991 and 31 December 1992 were recruited from the former Avon County of southwest England.[32] Since then, 13,988 offspring alive at one year have been followed repeatedly with questionnaire- and clinic-based assessments.[33,34] Sufficient information was available on 6,221 of these individuals to be included in our analysis, as metabolomics was not performed for all individuals in the ALSPAC study. Study data were collected and managed using REDCap electronic data capture tools hosted at the University of Bristol.[35] REDCap (Research Electronic Data Capture) is a secure, web-based software platform designed to support data capture for research studies. Offspring genotype was assessed using the Illumina HumanHap550 quad chip platform. Quality control measures included exclusion of participants with sex mismatch, minimal or excessive heterozygosity, disproportionately missing data, insufficient sample replication, cryptic relatedness, and non-European ancestry. Imputation was performed using the Haplotype Reference Consortium (HRC) panel. Offspring were considered for the current analyses if they had no older siblings in ALSPAC (203 excluded) and were of white ethnicity (based on reports by parents, 604 excluded) to reduce the potential for confounding by genotype. The study website contains details of all available data through a fully searchable data dictionary and variable search tool (http://www.bristol.ac.uk/alspac/researchers/our-data/).

UK Biobank is a population-based cohort study based in 22 centres across the UK.[36] The cohort is made up of around 500,000 adults aged 40-80 years old, who were enrolled between 2006 and 2010. Genotyping data is available for 488,377 participants.[37] Participants were genotyped using one of two arrays – either the Applied Biosystems UK BiLEVE Axiom Array by Affymetrix (now part of Thermo Fisher Scientific), or the closely related Applied Biosystems UK Biobank Axiom Array. Approaches based on Principal Component Analysis (PCA) were used to account for population structure. Individuals were excluded: if reported sex differed from inferred sex based on genotyping data; if they had sex chromosome karyotypes which were not XX or XY; if they were outliers in terms of heterozygosity and missing rates; or if they had high relatedness to another participant. Multiallelic SNPs or those with a minor allele frequency of below 1% were removed. Imputation was performed using the UK10K haplotype and HRC reference panels.

The GWAS meta-analysis for CRC included up to 52,775 cases and 45,940 controls.[38,39] This sample excluded cases and controls from UK Biobank to avoid potential bias due to sample

overlap which may be problematic in MR analyses.[40] Cases were diagnosed by a physician and recorded overall and by site (colon, 28,736 cases; proximal colon, 14,416 cases; distal colon, 12,879 cases; and rectal, 14,150 cases). Colon cancer included proximal colon (any primary tumour arising in the cecum, ascending colon, hepatic flexure, or transverse colon), distal colon (any primary tumour arising in the pleenic flexure, descending colon or sigmoid colon), and colon cases with unspecified site. Rectal cancer included any primary tumour arising in the rectum or rectosigmoid junction.[38] Approximately 92% of participants in the overall CRC GWAS were white-European (~8% were East Asian). All participants included in site-specific CRC analyses were of European ancestry. Imputation was performed using the Michigan imputation server and HRC r1.0 reference panel. Regression models were further adjusted for age, sex, genotyping platform, and genomic principal components as described previously.[38]

Assessment of CRC genetic liability

Genetic liability to CRC was based on single nucleotide polymorphisms (SNPs) associated with CRC case status at genome-wide significance ($P<5\times10^{-8}$). 108 independent SNPs reported by two major GWAS meta-analyses were eligible for inclusion in a CRC genetic risk score (GRS).[38,41] The set of SNPs was filtered, excluding 36 SNPs that were in linkage disequilibrium based on $R^2>0.001$ using the TwoSampleMR package (SNPs with the lowest P-values were retained).[42] This left 72 SNPs independently associated with CRC (**Supplementary File 1a**), 65 of which were available in imputed ALSPAC genotype data post quality control. As GWAS of site-specific CRC have identified marked heterogeneity,[43] GRS describing site-specific CRCs were constructed for sensitivity analyses using the same process outlined above. The GRS for colon cancer, rectal cancer, proximal colon cancer and distal colon cancer were comprised of 38, 25, 20 and 24 variants, respectively (**Supplementary File 1a**). For overall CRC and site-specific CRC analyses, sensitivity analyses excluding any SNPs in the FADS cluster (i.e. within the gene regions of *FADS1*, *FADS2*, or *FADS3*) (**Supplementary File 1a**) were performed given a likely role for these SNPs in influencing circulating metabolite levels directly, in particular via lipid metabolism (i.e., not primarily due to CRC).[44–50]

Assessment of circulating metabolites

Circulating metabolite measures were drawn from ALSPAC and UK Biobank using the same targeted metabolomics platform. In ALSPAC, participants provided non-fasting blood samples

during a clinic visit while aged approximately 8y, and fasting blood samples from clinic visits while aged approximately 16y, 18y, and 25y. Proton nuclear magnetic resonance (^{1}H-NMR) spectroscopy was performed on Ethylenediaminetetraacetic acid (EDTA) plasma (stored at or below -70 degrees Celsius pre-processing) to quantify a maximum of 231 metabolites.[51] Quantified metabolites included the cholesterol and triglyceride content of lipoprotein particles; the concentrations and diameter/size of these particles; apolipoprotein B and apolipoprotein A-1 concentrations; as well as fatty acids and their ratios to total fatty acid concentration, branched chain and aromatic amino acids, glucose and pre-glycaemic factors including lactate and citrate, fluid balance factors including albumin and creatinine, and the inflammatory marker glycoprotein acetyls (GlycA). This metabolomics platform has limited coverage of fatty acids. In UK Biobank, EDTA plasma samples from 117,121 participants, a random subset of the original ~500,000 who provided samples at assessment centres between 2006 and 2013, were analysed between 2019 and 2020 for levels of 249 metabolic traits (168 concentrations plus 81 ratios) using the same high-throughput ^{1}H-NMR platform. Data pre-processing and QC steps are described previously.[51–53] To allow comparability between MR and GRS estimates all metabolite measures were standardised and normalised using rank-based inverse normal transformation. For descriptive purposes in ALSPAC, body mass index (BMI) was calculated at each time point as weight (kg) divided by squared height (m^2) based on clinic measures of weight to the nearest 0.1 kg using a Tanita scale and height measured in light clothing without shoes to the nearest 0.1 cm using a Harpenden stadiometer.

CRC liability variants were combined into a GRS using PLINK 1.9, specifying the effect (risk raising) allele and coefficient (logOR) with estimates from the CRC GWAS used as external weights.[38,41] GRSs were calculated as the number of effect alleles (or dosages if imputed) at each SNP (0, 1, or 2) multiplied by its weighting, summing these, and dividing by the total number of SNPs used. Z-scores of GRS variables were calculated to standardize scoring.

Statistical approach

An overview of the study design is presented in **Figure 1**. To estimate the effect of increased genetic liability to CRC on circulating metabolites we conducted a GRS analysis in ALSPAC and reverse two-sample MR analyses in UK Biobank. Estimates were interpreted within a 'reverse MR' framework,[54] wherein results are taken to reflect 'metabolic features' of CRC liability which could capture causal or predictive metabolite-disease associations. To clarify the direction of

metabolite-CRC associations, we additionally performed conventional 'forward' two-sample MR analyses to estimate the effect of circulating metabolites on CRC risk using large-scale GWAS data on metabolites and CRC.

1. Associations of CRC liability with circulating metabolites in early life

Separate linear regression models with robust standard errors were used to estimate coefficients and 95% confidence intervals for associations of GRSs with each metabolite as a dependent variable measured on the same individuals at age 8y, 16y, 18y, and 25y, adjusted for sex and age at the time of metabolite assessment. To aid interpretations, estimates were multiplied by 0.693 ($\log_e 2$) to reflect SD-unit differences in metabolites per doubling of genetic liability to CRC.[55] The Benjamini-Hochberg method was used to adjust P-values for multiple testing and an adjusted P-value of <0.05 was used as a heuristic for evidence for association given current sample sizes.[56]

2. Reverse MR of the effects of CRC liability on circulating metabolites in middle adulthood

"Reverse" MR analyses[54] were conducted using UK Biobank for outcome datasets in two sample MR to examine the effect of CRC liability on circulating metabolites. SNP-outcome (metabolite) estimates were obtained from a GWAS of metabolites in UK Biobank.[57,58] Prior to GWAS, all metabolite measures were standardised and normalised using rank-based inverse normal transformation. Genetic association data for metabolites were retrieved using the MRC IEU UK Biobank GWAS pipeline.[59] Full summary statistics are available via the IEU Open GWAS project.[54,60] Up to 3 statistical methods were used to generate reverse MR estimates of the effect of CRC liability on circulating metabolites using the TwoSampleMR package[61]: random-effects inverse variance weighted (IVW), weighted-median, and weighted-mode, which each make differing assumptions about directional pleiotropy and SNP heterogeneity.[62–64] The IVW MR model will produce biased effect estimates in the presence of horizontal pleiotropy, i.e. where one or more genetic variant(s) included in the instrument affect the outcome by a pathway other than through the exposure. In the weighted median model, each genetic variant is weighted according to its distance from the median effect of all genetic variants. Thus, the weighted median model will provide an unbiased estimate when at least 50% of the information in an instrument comes from genetic variants that are not horizontally pleiotropic. The weighted mode model uses a

similar approach but weights genetic instruments according to the mean effect. In this model, over 50% of the weight of the genetic instrument can be contributed to by genetic variants which are horizontally pleiotropic, but the most common amount of pleiotropy must be zero (known as the Zero Modal Pleiotropy Assumption (ZEMPA))[65]. As above, estimates were multiplied by 0.693 ($\log_e 2$) to reflect SD-unit differences in metabolites per doubling of genetic liability to CRC.[66]

3. Forward MR of the effects of metabolites on CRC

Forward MR analyses were conducted using summary statistics from UK Biobank for the same NMR-measured metabolites (SNP-exposure) and from GECCO/CORECT/CCFR as outlined above (SNP-outcome). We identified SNPs that were independently associated (R^2<0.001 and $P<5x10^{-8}$) with metabolites from a GWAS of 249 metabolites in UK Biobank described above. As before, we used up to 3 statistical methods to generate MR estimates of the effect of circulating metabolites on CRC risk (overall and site-specific): random-effects IVW, weighted-median, and weighted-mode. The Benjamini-Hochberg method was used to adjust P-values for multiple testing and an adjusted P-value of <0.05 was used as a heuristic for nominal evidence for a causal effect.[56] MR outputs are beta coefficients representing the logOR for CRC per SD higher metabolite, exponentiated to reflect the OR for CRC per SD metabolite.

MR analyses were performed in R version 4.0.3.[67] and GRS analyses in Stata 16.1 (StataCorp, College Station, Texas, USA). The ggforestplot R package was used to generate results visualisations.[68]

Ethics

Written informed consent was obtained for all study participants. Ethical approval was obtained from the ALSPAC Law and Ethics Committee and the local research ethics committee (proposal B3538). Consent for biological samples has been collected in accordance with the Human Tissue Act (2004). Informed consent for the use of data collected via questionnaires and clinics was obtained from participants following the recommendations of the ALSPAC Ethics and Law Committee at the time. Ethics for the CRC GWAS were approved by respective institutional review boards.

Results

Associations of CRC liability with circulating metabolites in early life

At the time the ALSPAC blood samples were taken, the mean age of participants was 7.5y (N=4,767), 15.5y (N=2,930), 17.8y (N=2,613), and 24.5y (N=2,559) for the childhood, early adolescence, late adolescence and young adulthood time points respectively. The proportion of participants which were male were 50.5%, 47.4%, 44.5%, and 39.1% and mean BMI was 16.2, 21.4, 22.7, and 24.8 kg/m^2 for each time point respectively. The socio-demographic profile of ALSPAC offspring participants has been reported previously.[69] Mean and standard deviation (SD) values for metabolites on each measurement occasion in ALSPAC are shown in **Supplementary File 1b**.

In the GRS analysis, there was no strong evidence of association of CRC liability with metabolites at age 8y (**Supplementary File 1c**). At age 16y, there was evidence for association with several lipid traits including higher cholesteryl esters to total lipids ratio in large low-density lipoprotein (LDL) (SD change per doubling CRC liability = 0.06, 95% CI = 0.02 to 0.10) and higher cholesterol in very small very low-density lipoprotein (VLDL) (SD change per doubling CRC liability = 0.06, 95% CI = 0.03 to 0.10). There was strong evidence for association with several traits at age 18y including higher non-high-density lipoprotein (non-HDL) lipids, e.g., a 1 doubling CRC liability was associated with higher levels of total cholesterol (SD change = 0.05 95% CI = 0.01 to 0.09), VLDL-cholesterol (SD change = 0.05, 95% CI = 0.01 to 0.09), LDL-cholesterol (SD change = 0.06, 95% CI = 0.02 to 0.09)), apolipoproteins (apolipoprotein B (SD change = 0.06, 95% CI = 0.02 to 0.09)), and fatty acids (omega-3 (SD change = 0.08, 95% CI = 0.04 to 0.11), docosahexaenoic acid (DHA) (SD change = 0.05, 95% CI = 0.02 to 0.09)) (**Supplementary File 1c**). **Figure 2(-figure supplements 1-6)** shows results for all clinically validated metabolites. At age 25y, there was no strong evidence of association of CRC liability with metabolites. In anatomical site-specific analyses, there was strong evidence for association of liability to colon cancer with omega-3 (SD change = 0.07, 95% CI = 0.03 to 0.11) and DHA (SD change = 0.07, 95% CI = 0.03 to 0.10) at age 18y. There was little evidence for any associations at any other CRC site or age (**Supplementary File 1c**). When SNPs in the FADS cluster gene regions were excluded due to possible horizontal pleiotropy given the role of FADS in lipid metabolism, there was a reduction in strength of evidence for an association of liability to CRC with any metabolite measured, although estimates were in a largely consistent direction with the prior analysis (**Supplementary File 1d**).

Reverse MR of the effects of CRC liability on circulating metabolites in middle adulthood

All instrument sets from the reverse MR analysis had an F-statistic greater than 10 (minimum F-statistic = 36, median = 40), suggesting our analyses did not suffer from weak instrument bias (**Supplementary File 1e**). There was little evidence of an association of CRC liability (overall or by anatomical site) on any of the circulating metabolites investigated, including when the SNP in the *FADS* gene region was excluded, based on our pre-determined cut-off of FDR-P < 0.05; however, the direction of effect estimates was largely consistent with those seen in ALSPAC GRS analyses, with higher CRC liability weakly associated with higher non-HDLs, lipoproteins and fatty acid levels (**Supplementary File 1f-g)**. **Figure 3(-figure supplements 1-3)** shows the results for clinically validated metabolites. In subsite stratified analyses, there was strong evidence for a causal effect of genetic liability to proximal colon cancer on several traits, including total fatty acids (SD change per doubling of liability = 0.02, 95% CI = 0.01 to 0.04) and omega-6 fatty acids (SD change per doubling of liability = 0.03, 95% CI = 0.01 to 0.05).

Forward MR for the effects of metabolites on CRC risk

All instrument sets from the forward MR analysis had an F-statistic greater than 10 (minimum F-statistic = 54, median = 141), suggesting that our analyses were unlikely to suffer from weak instrument bias (**Supplementary File 1h-i**). There was strong evidence for an effect of several fatty acid traits on overall CRC risk, including of omega-3 fatty acids (CRC OR = 1.13, 95% CI = 1.06 to 1.21), DHA (OR CRC = 1.76, 95% CI = 1.08 to 1.28), ratio of omega-3 fatty acids to total fatty acids (OR CRC = 1.18, 95% CI = 1.11 to 1.25), ratio of DHA to total fatty acids (CRC OR = 1.20, 95% CI = 1.10 to 1.31), and ratio of omega-6 fatty acids to omega-3 fatty acids (CRC OR = 0.86, 95% CI = 0.80 to 9.13) (**Supplementary File 1j, Figure 4-figure supplements 1-3**). These estimates were overlapping with variable precision in MR sensitivity models. When SNPs in the *FADS* gene region were excluded, there was little evidence for a causal effect of any metabolite investigated on CRC risk based on the predetermined FDR-P cut of off < 0.05, although the directions of effect estimates were consistent with previous analyses (**Supplementary File 1k**).

In anatomical subtype stratified analyses evidence was strongest for an effect of fatty acid traits on higher CRC risk, and this appeared specific to the distal colon, e.g., omega-3 (distal CRC OR = 1.20, 95% CI = 1.09 to 1.32), and ratio of DHA to total fatty acids (distal colon OR = 1.29, 95%

CI = 1.16 to 1.43). There was also evidence of a negative effect of ratio of omega-6 to omega-3 fatty acids (distal CRC OR = 0.80, 95% CI = 0.74 to 0.88) and a positive effect of ratio of omega-3 fatty acids to total fatty acids (distal CRC = 1.24, 95% CI = 1.15 to 1.35; seen also for proximal CRC OR = 1.15, 95% CI = 1.07 to 1.23) (**Supplementary File 1j**). These estimates were also directionally consistent in MR sensitivity models.

Discussion

Here, we used a reverse MR framework to identify circulating metabolites which are associated with genetic CRC liability across different stages of the early life course and attempted to replicate results in an independent cohort of middle-aged adults. We then performed forward MR to characterise the causal direction of the relationship between metabolites and CRC. Our GRS analysis provided evidence for an association of genetic liability to CRC with higher circulating levels of lipoprotein lipids (including total cholesterol, VLDL-cholesterol, and LDL-cholesterol), apolipoproteins (including apolipoprotein B), and fatty acids (including omega-3 and DHA) in young adults. These results were largely consistent in direction (though smaller in magnitude and weaker in strength of evidence) in a two-sample MR analysis in an independent cohort of middle-aged adults. Results were attenuated, but consistent in direction, when potentially pleiotropic SNPs in the *FADS* gene regions were excluded. However, it should be noted that use of a narrow window for exclusion based on being within one of the three *FADS* genes may mean that some pleiotropic SNPs remain. Our subsequent forward MR analysis highlighted polyunsaturated fatty acids as potentially having a causal role in the development of CRC.

Our analyses highlight a potentially important role of polyunsaturated fatty acids in colorectal cancer liability. However, these analyses may be biased by substantial genetic pleiotropy among fatty acid traits. SNPs which are associated with levels of one fatty acid are generally associated with levels of many more fatty acid (and non-fatty acid) traits.[70,71] For instance, genetic instruments within the FADS cluster of genes will likely affect both omega-3 and omega-6 fatty acids, given *FADS1* and *FADS2* encode enzymes which catalyse the conversion of both from shorter chain into longer chain fatty acids.[71] In addition, the NMR metabolomics platform utilised in the analyses outlined here has limited coverage of fatty acids, meaning many putative causal metabolites for CRC, for example arachidonic acid, could not be investigated. Therefore, although our results indicate that polyunsaturated fatty acids may be important in colorectal cancer risk, given the

pleiotropic nature of the fatty acid genetic instruments and the limited coverage of the NMR platform, we are unable to determine with any certainty which specific classes of fatty acids may be driving these associations.

Our analyses featured evaluating the effect of genetic liability to CRC on circulating metabolites across repeated measures in the ALSPAC cohort. The mean ages at the time of the repeated measures were 8y, 16y, 18y, and 25y, representing childhood, early adolescence, late adolescence, and young adulthood respectively, and therefore individuals in this cohort are unlikely to be taking metabolite-altering medication such as statins, and unlikely to have CRC. The strongest evidence for an effect of liability to CRC on metabolite levels was seen in late adolescence. The reason for this remains unclear. It is possible that this represents a true biological phenomenon if late adolescence is a critical window in CRC development or metabolite variability, which may be likely given the limited variance in metabolite levels at the later age of 25y (**Supplementary File 1b**). The lack of an effect at the younger ages could be explained by the fact that the CRC GRS may capture many key life events or experiences which could impact the metabolome (e.g., initiation of smoking, higher category of BMI reached, educational attainment level set, etc) but may not have yet happened at younger ages, thus obscuring an effect of genetic liability to CRC on the metabolome. Our results suggest that puberty could be important, with an effect seen seemingly particularly at the end of puberty. Repeating our analysis with sex-stratified data may aid in determining whether this is likely to be the case; sex-stratified GWAS for metabolites are not currently available to replicate such analyses. An alternative explanation is selection bias due to loss of follow-up, leading to a change in sample characteristics over time.

Another key finding in the reverse MR analysis was that genetic liability to CRC was associated with increased levels of total cholesterol, VLDL-cholesterol, LDL-cholesterol, and apolipoprotein B, though we find little evidence for a causal effect of these traits on risk of CRC in the forward MR, replicating previous forward MR analyses for total and LDL-cholesterol.[9,72–74] This suggests that these traits may either be only predictive of (i.e., non-causal for) later CRC development, or may be influenced by the development of CRC and could have diagnostic or predictive potential. Given that the participants in the ALSPAC cohort are many decades younger than the average age of diagnosis for CRC (mean age 25 years in the latest repeated measure analysed in ALSPAC; whereas the median age at diagnosis of CRC is 64 years),[75] the former seems the most likely scenario. Previous conventional observational studies have presented conflicting results

when investigating the association between measures of cholesterol and CRC risk with some finding an inverse association and others a positive association, possibly reflecting residual confounding in conventional observational analyses.[76–82] Previous MR studies have had similar findings to our forward MR analysis, in that there seems to be little evidence for a causal effect of cholesterol on CRC development.[72–74] One possible explanation for how circulating levels of total cholesterol, VLDL-cholesterol, LDL-cholesterol and apolipoprotein B could predict (without necessarily causing) future CRC development could be linked to diet. A previous MR analysis suggested an effect of increased BMI on several measures of circulating cholesterol.[9] Consuming a diet which is high in fat may increase CRC risk both through and possibly independently of adiposity, alongside increasing levels of circulating cholesterol.[83–88] The potential for lipoprotein or apolipoprotein lipid measures in future CRC risk prediction should be further investigated.

Our analyses stratified by anatomical subsite highlighted fatty acids as being affected by genetic liability to colon and proximal colon cancer, with the forward MR confirming that fatty acid traits may be particularly important in the development of these subsites of CRC as well as distal colon cancer.

In our forward MR analyses we were unable to replicate the findings of three previous MR studies which found evidence for a causal effect of circulating linoleic acid levels on CRC development in terms of strength of evidence, though the direction of the effect estimate was similar to previous studies.[89–91] This is surprising as all three previous analyses had a much smaller sample size than that included in our analysis (the largest had sample size of 24,748 for exposure vs 118,466 presently; and 11,016 cases and 13,732 controls for outcome vs 52,775 cases and 45,940 controls presently). Our analysis using updated genetic instruments to proxy fatty acids may be more successful in accurately instrumenting heterogenous phenotypes such as metabolite levels compared with previous analyses. All other findings in our forward MR analysis are consistent with previous MR studies where they exist.[72–74]

Limitations

The limitations of this study include firstly the relatively small sample size included in the ALSPAC analysis, which may have implications for power and precision. Secondly, mostly due to the longitudinal nature of the ASLAPC study, our sample at each time point is composed of slightly different individuals. This could be influencing our results, and should be taken into account when

comparing across time points. Thirdly, our analyses involving genetic instruments for CRC liability may have suffered from horizontal pleiotropy, even after excluding genetic variants in or near the *FADS* gene. Fourthly, our analyses were mostly restricted to white Europeans, which limits the generalisability of our findings to other populations. Fifthly, our analysis would benefit from being repeated with sex-stratified data. Although such GWAS results for metabolites are not currently available, the data to perform such GWAS are available in UK Biobank for future analyses. Sixthly, for our forward MR analysis, we used the UK Biobank for our exposure data. The UK Biobank has a median age of 58 at the time these measurements were taken, meaning statin use may be widespread in this population, which could be attenuating our effect estimates. Future work could attempt to replicate our analysis in a population with lower prevalence of statins intake. Finally, we included only metabolites measured using NMR. Confirming whether our results replicate using metabolite data measured with an alternative method would strengthen our findings.

Conclusions

Our analysis provides evidence that genetic liability to CRC is associated with altered levels of metabolites at certain ages, some of which may have a causal role in CRC development. Further investigating the role of polyunsaturated fatty acids in CRC risk and circulating cholesterol in CRC prediction may be promising avenues for future research.

Figure legends

Figure 1: Study design. First, linear regression models were used to examine the relationship between genetic susceptibility to adult colorectal cancer and circulating metabolites measured in ALSPAC participants at age 8, 16, 18 and 25 years. Next, we performed a reverse Mendelian randomization analysis to identify metabolites influenced by CRC susceptibility in an independent population of adults. Finally, we performed a conventional (forward) Mendelian randomization analysis of circulating metabolites on CRC to identify metabolites causally associated with CRC risk. Consistent evidence across all three methodological approaches was interpreted to indicate a causal role for a given metabolite in CRC aetiology.

Figure 2: Associations of genetic liability to adult colorectal cancer (based on a 72 SNP genetic risk score) with clinically validated metabolic traits at different early life stages among ALSPAC offspring (age 8y (N=4,767), 16y (N=2,930), 18y (N=2,613), and 25y (N=2,559)). Estimates shown are beta coefficients representing the SD difference in metabolic trait per doubling of genetic liability to colorectal cancer (purple, 8y; turquoise, 16y; red, 18y; black, 25y). Filled point estimates are those that pass a Benjamini–Hochberg FDR multiple-testing correction (FDR < 0.05).

Figure 2-figure supplement 1: Associations of genetic liability to adult colon cancer with clinically validated metabolic traits at different early life stages among ALSPAC offspring (age

8y, 16y, 18y, and 25y). Estimates shown are beta coefficients representing the SD difference in metabolic trait per doubling of genetic liability to colon cancer (purple, 8y; turquoise, 16y; red, 18y; black, 25y). Filled point estimates are those that pass a Benjamini–Hochberg FDR multiple-testing correction (FDR < 0.05).

Figure 2-figure supplement 2: Associations of genetic liability to proximal colon cancer with clinically validated metabolic traits at different early life stages among ALSPAC offspring (age 8y, 16y, 18y, and 25y). Estimates shown are beta coefficients representing the SD difference in metabolic trait per doubling of genetic liability to proximal colon cancer (purple, 8y; turquoise, 16y; red, 18y; black, 25y). Filled point estimates are those that pass a Benjamini–Hochberg FDR multiple-testing correction (FDR < 0.05).

Figure 2-figure supplement 3: Associations of genetic liability to distal colon cancer with clinically validated metabolic traits at different early life stages among ALSPAC offspring (age 8y, 16y, 18y, and 25y). Estimates shown are beta coefficients representing the SD difference in metabolic trait per doubling of genetic liability to distal colon cancer (purple, 8y; turquoise, 16y; red, 18y; black, 25y). Filled point estimates are those that pass a Benjamini–Hochberg FDR multiple-testing correction (FDR < 0.05).

Figure 2-figure supplement 4: Associations of genetic liability to rectal cancer with clinically validated metabolic traits at different early life stages among ALSPAC offspring (age 8y, 16y, 18y, and 25y). Estimates shown are beta coefficients representing the SD difference in metabolic trait per doubling of genetic liability to rectal cancer (purple, 8y; turquoise, 16y; red, 18y; black, 25y). Filled point estimates are those that pass a Benjamini–Hochberg FDR multiple-testing correction (FDR < 0.05).

Figure 2-figure supplement 5: Associations of genetic liability to adult colorectal cancer (excluding rs174533) with clinically validated metabolic traits at different early life stages among ALSPAC offspring (age 8y, 16y, 18y, and 25y). Estimates shown are beta coefficients representing the SD difference in metabolic trait per doubling of genetic liability to colorectal cancer (purple, 8y; turquoise, 16y; red, 18y; black, 25y). Filled point estimates are those that pass a Benjamini–Hochberg FDR multiple-testing correction (FDR < 0.05).

Figure 2-figure supplement 6: Associations of genetic liability to adult colon cancer (excluding rs174535) with clinically validated metabolic traits at different early life stages among ALSPAC offspring (age 8y, 16y, 18y, and 25y). Estimates shown are beta coefficients representing the SD difference in metabolic trait per doubling of genetic liability to colorectal cancer cancer (purple, 8y; turquoise, 16y; red, 18y; black, 25y). Filled point estimates are those that pass a Benjamini–Hochberg FDR multiple-testing correction (FDR < 0.05).

Figure 3: Associations of genetic liability to colorectal cancer with clinically validated metabolic traits in an independent sample of adults (UK Biobank, N=118,466, median age 58y) based on reverse two sample Mendelian randomization analyses. Estimates shown are beta coefficients representing the SD-unit difference in metabolic trait per doubling of liability to colorectal cancer. Filled point estimates are those that pass a Benjamini–Hochberg FDR multiple-testing correction (FDR < 0.05).

Figure 3-figure supplement 1: Associations of genetic liability to colorectal cancer with clinically validated metabolic traits in an independent sample of adults based on reverse two sample Mendelian randomization analyses. Estimates shown are beta coefficients representing the SD-unit difference in metabolic trait per doubling of liability to colorectal cancer by site

(colorectal, colon, distal colon, proximal colon and rectal cancer). Filled point estimates are those that pass a Benjamini–Hochberg FDR multiple-testing correction (FDR < 0.05).

Figure 3-figure supplement 2: Associations of genetic liability to colorectal cancer (excluding genetic variants in the FADS gene region) with clinically validated metabolic traits in an independent sample of adults based on reverse two sample Mendelian randomization analyses. Estimates shown are beta coefficients representing the SD-unit difference in metabolic trait per doubling of liability to colorectal cancer. Filled point estimates are those that pass a Benjamini–Hochberg FDR multiple-testing correction (FDR < 0.05).

Figure 3-figure supplement 3: Associations of genetic liability to colorectal and colon cancer with clinically validated metabolic traits in an independent sample of adults based on reverse two sample Mendelian randomization analyses with FADS variants excluded from colorectal cancer instruments. Estimates shown are beta coefficients representing the SD-unit difference in metabolic trait per doubling of liability to colorectal cancer by site (colorectal, colon). Filled point estimates are those that pass a Benjamini–Hochberg FDR multiple-testing correction (FDR < 0.05).

Figure 4: Associations of clinically validated metabolites with colorectal cancer based on conventional (forward) two sample Mendelian randomization analyses in individuals from UK Biobank (N=118,466, median age 58y) . Estimates shown are beta coefficients representing the logOR for colorectal cancer per SD metabolite. Filled point estimates are those that pass a Benjamini–Hochberg FDR multiple-testing correction (FDR < 0.05).

Figure 4-figure supplement 1: Associations of clinically validated metabolites with colorectal cancer by site (colorectal, colon, distal colon, proximal colon and rectal cancer) based on conventional (forward) two sample Mendelian randomization analyses. Estimates shown are ORs for colorectal cancer per SD metabolite. Filled point estimates are those that pass a Benjamini–Hochberg FDR multiple-testing correction (FDR < 0.05).

Supplementary file legends

Supplementary File 1a: Genetic variants used to construct genetic risk scores reflecting colorectal cancer liability

Supplementary File 1b: Mean and SD values for raw metabolic traits at different life stages among ALSPAC offspring

Supplementary File 1c: Associations of genetic liability to colorectal cancer with metabolic traits at different early life stages among ALSPAC offspring

Supplementary File 1d: Associations of genetic liability to colorectal cancer (excluding SNPs in the FADS gene region) with metabolic traits at different early life stages among ALSPAC offspring

Supplementary File 1e: Assesment of instrument strength for MR analyses

Supplementary File 1f: Associations of genetic liability to colorectal cancer with metabolic traits based on two-sample MR

Supplementary File 1g: Associations of genetic liability to colorectal cancer with metabolic traits based on two-sample MR, excluding variants in the FADS gene region

Supplementary File 1h: Genetic variants used to instrument circulating metabolites

Supplementary File 1i: Assesment of instrument strength for MR analyses

Supplementary File 1j: Estimated effects of circulating metabolites on colorectal cancer risk based on two-sample MR

Supplementary File 1k: Estimated effects of circulating metabolites on colorectal cancer risk based on two-sample MR, excluding variants in the FADS gene region

References

1. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* [Internet]. Wiley-Liss Inc.; 2015 Mar 1 [cited 2022 Nov 17];**136**(5):E359–E386. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/ijc.29210
2. World Cancer Research Fund/American Institute for Cancer Research. Continuous Update Project Expert Report. Diet, nutrition, physical activity and colorectal cancer. 2018;
3. Huyghe JR, Bien SA, Harrison TA, et al. Discovery of common and rare genetic risk variants for colorectal cancer. *Nature Genetics 2018 51:1* [Internet]. Nature Publishing Group; 2018 Dec 3 [cited 2023 Mar 9];**51**(1):76–87. Available from: https://www.nature.com/articles/s41588-018-0286-6
4. Czene K, Lichtenstein P, Hemminki K. Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish family-cancer database. *Int J Cancer* [Internet]. John Wiley & Sons, Ltd; 2002 May 10 [cited 2023 Mar 9];**99**(2):260–266. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/ijc.10332
5. Aul P, Ichtenstein L, Olm IVH, et al. Environmental and Heritable Factors in the Causation of Cancer — Analyses of Cohorts of Twins from Sweden, Denmark, and Finland. *https://doi.org/101056/NEJM200007133430201* [Internet]. Massachusetts Medical Society ; 2000 Jul 13 [cited 2023 Mar 9];**343**(2):78–85. Available from: https://www.nejm.org/doi/full/10.1056/NEJM200007133430201
6. Händel MN, Rohde JF, Jacobsen R, et al. Processed meat intake and incidence of colorectal cancer: a systematic review and meta-analysis of prospective observational studies. *European Journal of Clinical Nutrition 2020 74:8* [Internet]. Nature Publishing Group; 2020 Feb 6 [cited 2022 Nov 17];**74**(8):1132–1148. Available from: https://www.nature.com/articles/s41430-020-0576-9
7. McNabb S, Harrison TA, Albanes D, et al. Meta-analysis of 16 studies of the association of alcohol with colorectal cancer. *Int J Cancer* [Internet]. John Wiley & Sons, Ltd; 2020 Feb 1 [cited 2022 Nov 17];**146**(3):861–873. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/ijc.32377
8. Lauby-Secretan B, Scoccianti C, Loomis D, Grosse Y, Bianchini F, Straif K. Body Fatness and Cancer--Viewpoint of the IARC Working Group. *N Engl J Med* [Internet]. N Engl J Med; 2016 Aug 25 [cited 2022 Nov 17];**375**(8):794–798. Available from: https://pubmed.ncbi.nlm.nih.gov/27557308/
9. Bull CJ, Bell JA, Murphy N, et al. Adiposity, metabolites, and colorectal cancer risk: Mendelian randomization study. *BMC Med* [Internet]. BioMed Central Ltd; 2020 Dec 1 [cited 2023 Jan 10];**18**(1):1–16. Available from: https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-020-01855-9
10. Ladabaum U, Dominitz JA, Kahi C, Schoen RE. Strategies for Colorectal Cancer Screening. *Gastroenterology*. W.B. Saunders; 2020 Jan 1;**158**(2):418–432.
11. Cardoso R, Guo F, Heisser T, et al. Colorectal cancer incidence, mortality, and stage distribution in European countries in the colorectal cancer screening era: an international population-based study. *Lancet Oncol*. Elsevier; 2021 Jul 1;**22**(7):1002–1013.
12. Qiu Y, Cai G, Su M, et al. Serum Metabolite Profiling of Human Colorectal Cancer Using GC−TOFMS and UPLC−QTOFMS. *J Proteome Res* [Internet]. American Chemical Society; 2009 [cited 2022 Nov 17];**8**(10):4844–4850. Available from: https://pubs.acs.org/doi/full/10.1021/pr9004162

13. Ward PS, Thompson CB. Metabolic reprogramming: a cancer hallmark even warburg did not anticipate. *Cancer Cell* [Internet]. Cancer Cell; 2012 Mar 20 [cited 2022 Nov 17];**21**(3):297–308. Available from: https://pubmed.ncbi.nlm.nih.gov/22439925/
14. Rattray NJW, Charkoftaki G, Rattray Z, Hansen JE, Vasiliou V, Johnson CH. Environmental Influences in the Etiology of Colorectal Cancer: the Premise of Metabolomics. *Current Pharmacology Reports 2017 3:3* [Internet]. Springer; 2017 Apr 7 [cited 2022 Nov 22];**3**(3):114–125. Available from: https://link.springer.com/article/10.1007/s40495-017-0088-z
15. Singla P, Bardoloi A, Parkash AA. Metabolic effects of obesity: A review. *World J Diabetes* [Internet]. Baishideng Publishing Group Inc; 2010 Jul 7 [cited 2022 Nov 17];**1**(3):76. Available from: /pmc/articles/PMC3083889/
16. Papandreou C, García-Gavilán J, Camacho-Barcia L, et al. Circulating Metabolites Associated with Body Fat and Lean Mass in Adults with Overweight/Obesity. *Metabolites 2021, Vol 11, Page 317* [Internet]. Multidisciplinary Digital Publishing Institute; 2021 May 13 [cited 2022 Nov 17];**11**(5):317. Available from: https://www.mdpi.com/2218-1989/11/5/317/htm
17. Ahmad S, Hammar U, Kennedy B, et al. Effect of General Adiposity and Central Body Fat Distribution on the Circulating Metabolome: A Multicohort Nontargeted Metabolomics Observational and Mendelian Randomization Study. *Diabetes* [Internet]. American Diabetes Association; 2022 Feb 1 [cited 2022 Nov 17];**71**(2):329–339. Available from: https://diabetesjournals.org/diabetes/article/71/2/329/138937/Effect-of-General-Adiposity-and-Central-Body-Fat
18. Haycock PC, Carolina Borges M, Burrows K, et al. The association between genetically elevated polyunsaturated fatty acids and risk of cancer. *www.thelancet.com* [Internet]. 2023 [cited 2023 Jul 31];**91**. Available from: http://practical.icr.ac.uk/.ajhttps://epi.grants.cancer.gov/interlymph/.akhttps://dceg.cancer.gov/research/cancer-types/pancreas/panscan.
19. Yarmolinsky J, Wade KH, Richmond RC, et al. Causal Inference in Cancer Epidemiology: What Is the Role of Mendelian Randomization? [cited 2022 Nov 17]; Available from: http://aacrjournals.org/cebp/article-pdf/27/9/995/2285730/995.pdf
20. GD S, S E. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* [Internet]. Int J Epidemiol; 2003 [cited 2022 Nov 17];**32**(1). Available from: https://pubmed.ncbi.nlm.nih.gov/12689998/
21. Holmes M v., Smith GD. Can Mendelian Randomization Shift into Reverse Gear? *Clin Chem* [Internet]. Oxford Academic; 2019 Mar 1 [cited 2022 Nov 17];**65**(3):363–366. Available from: https://academic.oup.com/clinchem/article/65/3/363/5608004
22. Gao R, Wu C, Zhu Y, et al. Integrated Analysis of Colorectal Cancer Reveals Cross-Cohort Gut Microbial Signatures and Associated Serum Metabolites. *Gastroenterology*. W.B. Saunders; 2022 Oct 1;**163**(4):1024-1037.e9.
23. Rattner JI, Kopciuk KA, Vogel HJ, et al. Early detection of treatment futility in patients with metastatic colorectal cancer. *Oncotarget* [Internet]. Impact Journals, LLC; 2022 [cited 2022 Nov 22];**13**:61. Available from: /pmc/articles/PMC8746015/
24. Leichtle AB, Nuoffer JM, Ceglarek U, et al. Serum amino acid profiles and their alterations in colorectal cancer. *Metabolomics* [Internet]. Springer; 2012 Aug 16 [cited

2022 Nov 22];**8**(4):643–653. Available from: https://link.springer.com/article/10.1007/s11306-011-0357-5

25. Ritchie SA, Ahiahonu PWK, Jayasinghe D, et al. Reduced levels of hydroxylated, polyunsaturated ultra long-chain fatty acids in the serum of colorectal cancer patients: Implications for early screening and detection. *BMC Med* [Internet]. BioMed Central; 2010 Feb 15 [cited 2022 Nov 22];**8**(1):1–20. Available from: https://bmcmedicine.biomedcentral.com/articles/10.1186/1741-7015-8-13

26. Nishiumi S, Kobayashi T, Ikeda A, et al. A Novel Serum Metabolomics-Based Diagnostic Approach for Colorectal Cancer. *PLoS One* [Internet]. Public Library of Science; 2012 Jul 11 [cited 2022 Nov 22];**7**(7):e40459. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0040459

27. Ma Y, Zhang P, Wang F, Liu W, Yang J, Qin H. An integrated proteomics and metabolomics approach for defining oncofetal biomarkers in the colorectal cancer. *Ann Surg* [Internet]. 2012 Apr [cited 2022 Nov 22];**255**(4):720–730. Available from: https://journals.lww.com/annalsofsurgery/Fulltext/2012/04000/An_Integrated_Proteomics_and_Metabolomics_Approach.17.aspx

28. Bertini I, Cacciatore S, Jensen B v., et al. Metabolomic NMR fingerprinting to identify and predict survival of patients with metastatic colorectal cancer. *Cancer Res* [Internet]. American Association for Cancer Research; 2012 Jan 1 [cited 2022 Nov 22];**72**(1):356–364. Available from: https://aacrjournals.org/cancerres/article/72/1/356/575915/Metabolomic-NMR-Fingerprinting-to-Identify-and

29. Zhang A, Sun H, Yan G, Wang P, Han Y, Wang X. Metabolomics in diagnosis and biomarker discovery of colorectal cancer. *Cancer Lett*. Elsevier; 2014 Apr 1;**345**(1):17–20.

30. Farshidfar F, Weljie AM, Kopciuk KA, et al. A validated metabolomic signature for colorectal cancer: exploration of the clinical value of metabolomics. *British Journal of Cancer 2016 115:7* [Internet]. Nature Publishing Group; 2016 Aug 25 [cited 2022 Nov 22];**115**(7):848–857. Available from: https://www.nature.com/articles/bjc2016243

31. Farshidfar F, Weljie AM, Kopciuk K, et al. Serum metabolomic profile as a means to distinguish stage of colorectal cancer. *Genome Med* [Internet]. BioMed Central; 2012 May 14 [cited 2022 Nov 22];**4**(5):1–13. Available from: https://genomemedicine.biomedcentral.com/articles/10.1186/gm341

32. Boyd A, Golding J, Macleod J, et al. Cohort Profile: The ‘Children of the 90s’—the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol*. 2013 Feb;**42**(1):111–127.

33. Fraser A, Macdonald-Wallis C, Tilling K, et al. Cohort Profile: The Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int J Epidemiol*. 2013 Feb 1;**42**(1):97–110.

34. Northstone K, Lewcock M, Groom A, et al. The Avon Longitudinal Study of Parents and Children (ALSPAC): an update on the enrolled sample of index children in 2019. *Wellcome Open Res* [Internet]. The Wellcome Trust; 2019 [cited 2023 Feb 28];**4**. Available from: /pmc/articles/PMC6464058/

35. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research Electronic Data Capture (REDCap) - A metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* [Internet]. NIH

Public Access; 2009 Apr [cited 2023 Feb 28];**42**(2):377. Available from: /pmc/articles/PMC2700030/

36. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Med* [Internet]. Public Library of Science; 2015 Mar 1 [cited 2022 Nov 22];**12**(3):e1001779. Available from: https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001779
37. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature 2018 562:7726* [Internet]. Nature Publishing Group; 2018 Oct 10 [cited 2022 Nov 22];**562**(7726):203–209. Available from: https://www.nature.com/articles/s41586-018-0579-z
38. Huyghe JR, Bien SA, Harrison TA, et al. Discovery of common and rare genetic risk variants for colorectal cancer. *Nat Genet*. 2019;**51**(1):76–87.
39. Huyghe JR, Harrison TA, Bien SA, et al. Genetic architectures of proximal and distal colorectal cancer are partly distinct. *Gut* [Internet]. Gut; 2021 Jul 1 [cited 2023 Jan 27];**70**(7):1325–1334. Available from: https://pubmed.ncbi.nlm.nih.gov/33632709/
40. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. *Genet Epidemiol*. 2016;**40**(7):597–608.
41. Law PJ, Timofeeva M, Fernandez-Rozadilla C. Association analyses identify 31 new risk loci for colorectal cancer susceptibility.
42. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife*. 2018 May 30;**7**.
43. Huyghe JR, Harrison TA, Bien SA, et al. Genetic architectures of proximal and distal colorectal cancer are partly distinct. *Gut*. Gut; 2021 Jul 1;**70**(7):1325–1334.
44. Lu T, Sun L, Wang Z, Zhang Y, He Z, Xu C. Fatty acid synthase enhances colorectal cancer cell proliferation and metastasis via regulating AMPK/mTOR pathway. *Onco Targets Ther* [Internet]. Dove Press; 2019 [cited 2022 Nov 22];**12**:3339. Available from: /pmc/articles/PMC6504633/
45. Zaytseva YY, Rychahou PG, Le AT, et al. Preclinical evaluation of novel fatty acid synthase inhibitors in primary colorectal cancer cells and a patient-derived xenograft model of colorectal cancer. *Oncotarget* [Internet]. Impact Journals, LLC; 2018 May 5 [cited 2022 Nov 22];**9**(37):24787. Available from: /pmc/articles/PMC5973868/
46. TVB 2640 for Resectable Colon Cancer Other Resectable Cancers; a Window Trial. - Full Text View - ClinicalTrials.gov [Internet]. [cited 2022 Nov 22]. Available from: https://clinicaltrials.gov/ct2/show/NCT02980029
47. Fhu CW, Ali A. Fatty Acid Synthase: An Emerging Target in Cancer. *Molecules* [Internet]. Multidisciplinary Digital Publishing Institute (MDPI); 2020 Sep 1 [cited 2022 Nov 22];**25**(17). Available from: /pmc/articles/PMC7504791/
48. Chen HY, Cairns BJ, Small AM, et al. Association of FADS1/2 Locus Variants and Polyunsaturated Fatty Acids With Aortic Stenosis. *JAMA Cardiol* [Internet]. American Medical Association; 2020 Jun 1 [cited 2022 Nov 22];**5**(6):694–702. Available from: https://jamanetwork.com/journals/jamacardiology/fullarticle/2762845
49. Kathiresan S, Willer CJ, Peloso GM, et al. Common variants at 30 loci contribute to polygenic dyslipidemia. *Nature Genetics 2008 41:1* [Internet]. Nature Publishing Group; 2008 Dec 7 [cited 2022 Nov 22];**41**(1):56–65. Available from: https://www.nature.com/articles/ng.291

50. Tanaka T, Shen J, Abecasis GR, et al. Genome-Wide Association Study of Plasma Polyunsaturated Fatty Acids in the InCHIANTI Study. *PLoS Genet* [Internet]. Public Library of Science; 2009 Jan [cited 2022 Nov 22];**5**(1):e1000338. Available from: https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1000338
51. Würtz P, Kangas AJ, Soininen P, Lawlor DA, Davey Smith G, Ala-Korpela M. Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Large-Scale Epidemiology: A Primer on -Omic Technologies. *Am J Epidemiol*. Oxford Academic; 2017 Nov 1;**186**(9):1084–1096.
52. Julkunen H, Cichońska A, Slagboom PE, Würtz P. Metabolic biomarker profiling for identification of susceptibility to severe pneumonia and COVID-19 in the general population. *Elife*. eLife Sciences Publications Ltd; 2021 May 1;**10**.
53. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature 2018 562:7726*. Nature Publishing Group; 2018 Oct 10;**562**(7726):203–209.
54. Holmes M V., Smith GD. Can Mendelian randomization shift into reverse gear? *Clin Chem* American Association for Clinical Chemistry Inc.; 2019. p. 363–366.
55. Burgess S, Labrecque JA. Mendelian randomization with a binary exposure variable: interpretation and presentation of causal estimates. *Eur J Epidemiol* [Internet]. Eur J Epidemiol; 2018 Oct 1 [cited 2023 Jan 9];**33**(10):947–952. Available from: https://pubmed.ncbi.nlm.nih.gov/30039250/
56. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* [Internet]. John Wiley & Sons, Ltd; 1995 Jan 1 [cited 2022 Nov 22];**57**(1):289–300. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/j.2517-6161.1995.tb02031.x
57. Clayton GL, Borges MC, Lawlor DA. From menarche to menopause: the impact of reproductive factors on the metabolic profile of over 65,000 women. 2022 Apr 18 [cited 2023 Jan 27]; Available from: https://europepmc.org/article/PPR/PPR484113
58. Borges MC, Haycock PC, Zheng J, et al. Role of circulating polyunsaturated fatty acids on cardiovascular diseases risk: analysis using Mendelian randomization and fatty acid genetic association data from over 114,000 UK Biobank participants. *BMC Med* [Internet]. BMC Med; 2022 Dec 1 [cited 2023 Jan 27];**20**(1). Available from: https://pubmed.ncbi.nlm.nih.gov/35692035/
59. MRC IEU UK Biobank GWAS pipeline version 2 - Datasets - data.bris [Internet]. [cited 2022 Oct 14]. Available from: https://data.bris.ac.uk/data/dataset/pnoat8cxo0u52p6ynfaekeigi
60. Elsworth B, Lyon M, Alexander T, et al. The MRC IEU OpenGWAS data infrastructure.
61. Gibran Hemani Kaitlin H Wade, Charles Laurin, Benjamin Elsworth, Stephen Burgess, Jack Bowden, Ryan Langdon, Vanessa Tan, James Yarmolinsky, Hashem A. Shihab, Nicholas Timpson, David M Evans, Caroline Relton, Richard M Martin, George Davey Smith, Tom R G JZ. MR-Base: a platform for systematic causal inference across the phenome using billions of genetic associations. Biorxiv; 2016.
62. 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):304–314.

63. Hartwig FP, Smith GD, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol*. Oxford Academic; 2017 Dec 1;**46**(6):1985–1998.
64. Hartwig FP, Smith GD, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol* [Internet]. Int J Epidemiol; 2017 Dec 1 [cited 2023 Jan 9];**46**(6):1985–1998. Available from: https://pubmed.ncbi.nlm.nih.gov/29040600/
65. Hartwig FP, Smith GD, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. [cited 2023 Aug 1]; Available from: https://academic.oup.com/ije/article/46/6/1985/3957932
66. Burgess S, Labrecque JA. Mendelian randomization with a binary exposure variable: interpretation and presentation of causal estimates.
67. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.; 2020.
68. Forestplots of Measures of Effects and Their Confidence Intervals • ggforestplot [Internet]. [cited 2022 Oct 14]. Available from: https://nightingalehealth.github.io/ggforestplot/index.html
69. Boyd A, Golding J, Macleod J, et al. Cohort Profile: The 'Children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol* [Internet]. Oxford Academic; 2013 Feb 1 [cited 2023 Jan 27];**42**(1):111–127. Available from: https://academic.oup.com/ije/article/42/1/111/694290
70. Borges MC, Haycock PC, Zheng J, et al. Role of circulating polyunsaturated fatty acids on cardiovascular diseases risk: analysis using Mendelian randomization and fatty acid genetic association data from over 114,000 UK Biobank participants. *BMC Med* [Internet]. BioMed Central Ltd; 2022 Dec 1 [cited 2023 Jan 30];**20**(1):1–14. Available from: https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-022-02399-w
71. Borges MC, Haycock P, Zheng J, et al. The impact of fatty acids biosynthesis on the risk of cardiovascular diseases in Europeans and East Asians: a Mendelian randomization study. *Hum Mol Genet* [Internet]. Oxford Academic; 2022 Nov 28 [cited 2023 Feb 28];**31**(23):4034–4054. Available from: https://academic.oup.com/hmg/article/31/23/4034/6633188
72. Rodriguez-Broadbent H, Law PJ, Sud A, et al. Mendelian randomisation implicates hyperlipidaemia as a risk factor for colorectal cancer. *Int J Cancer* [Internet]. Int J Cancer; 2017 Jun 15 [cited 2022 Nov 28];**140**(12):2701–2708. Available from: https://pubmed.ncbi.nlm.nih.gov/28340513/
73. Cornish AJ, Law PJ, Timofeeva M, et al. Modifiable pathways for colorectal cancer: a mendelian randomisation analysis. *Lancet Gastroenterol Hepatol* [Internet]. Lancet Gastroenterol Hepatol; 2020 Jan 1 [cited 2022 Nov 28];**5**(1):55–62. Available from: https://pubmed.ncbi.nlm.nih.gov/31668584/
74. Luo X, Tu Z, Chen H, Ding J. Blood lipids and risk of colon or rectal cancer: a Mendelian randomization study. *J Cancer Res Clin Oncol* [Internet]. J Cancer Res Clin Oncol; 2021 Dec 1 [cited 2022 Nov 28];**147**(12):3591–3599. Available from: https://pubmed.ncbi.nlm.nih.gov/34490583/
75. Granados-Romero JJ, Valderrama-Treviño A, Hazzel E, Flores C, Herrera M. Colorectal cancer: a review. *Article in International Journal of Research in Medical Sciences*

[Internet]. 2017 [cited 2022 Nov 17]; Available from: http://dx.doi.org/10.18203/2320-6012.ijrms20174914

76. Park SK, Joo JS, Kim DH, Kim YE, Kang D, Yoo KY. Association of serum lipids and glucose with the risk of colorectal adenomatous polyp in men: a case-control study in Korea. *J Korean Med Sci* [Internet]. Korean Academy of Medical Science; 2000 Dec 1 [cited 2022 Nov 17];**15**(6):690–695. Available from: https://doi.org/10.3346/jkms.2000.15.6.690

77. Yang Z, Tang H, Lu S, Sun X, Rao B. Relationship between serum lipid level and colorectal cancer: a systemic review and meta-analysis. *BMJ Open* [Internet]. British Medical Journal Publishing Group; 2022 Jun 1 [cited 2022 Nov 17];**12**(6):e052373. Available from: https://bmjopen.bmj.com/content/12/6/e052373

78. Tian Y, Wang K, Li J, et al. The association between serum lipids and colorectal neoplasm: a systemic review and meta-analysis. *Public Health Nutr* [Internet]. Cambridge University Press; 2015 Mar 16 [cited 2022 Nov 17];**18**(18):3355–3370. Available from: https://www.cambridge.org/core/journals/public-health-nutrition/article/association-between-serum-lipids-and-colorectal-neoplasm-a-systemic-review-and-metaanalysis/4401DD8E09A7A257C95F1F3189E3B845

79. Yao X, Tian Z. Dyslipidemia and colorectal cancer risk: a meta-analysis of prospective studies. *Cancer Causes and Control* [Internet]. Kluwer Academic Publishers; 2015 Feb 1 [cited 2022 Nov 17];**26**(2):257–268. Available from: https://link.springer.com/article/10.1007/s10552-014-0507-y

80. Mayengbam SS, Singh A, Pillai AD, Bhat MK. Influence of cholesterol on cancer progression and therapy. *Transl Oncol*. Elsevier; 2021 Jun 1;**14**(6):101043.

81. Mannes GA, Maier A, Thieme C, Wiebecke B, Paumgartner G. Relation between the Frequency of Colorectal Adenoma and the Serum Cholesterol Level. *http://dx.doi.org/101056/NEJM198612253152602* [Internet]. Massachusetts Medical Society ; 2009 Nov 16 [cited 2022 Nov 17];**315**(26):1634–1638. Available from: https://www.nejm.org/doi/full/10.1056/NEJM198612253152602

82. Mamtani R, Lewis JD, Scott FI, et al. Disentangling the Association between Statins, Cholesterol, and Colorectal Cancer: A Nested Case-Control Study. *PLoS Med* [Internet]. Public Library of Science; 2016 Apr 1 [cited 2022 Nov 17];**13**(4):e1002007. Available from: https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1002007

83. Ocvirk S, Wilson AS, Appolonia CN, Thomas TK, O'Keefe SJD. Fiber, Fat, and Colorectal Cancer: New Insight into Modifiable Dietary Risk Factors. *Curr Gastroenterol Rep* [Internet]. Springer; 2019 Nov 1 [cited 2022 Nov 17];**21**(11):1–7. Available from: https://link.springer.com/article/10.1007/s11894-019-0725-2

84. Tang FY, Pai MH, Chiang EPI. Consumption of high-fat diet induces tumor progression and epithelial–mesenchymal transition of colorectal cancer in a mouse xenograft model. *J Nutr Biochem*. Elsevier; 2012 Oct 1;**23**(10):1302–1313.

85. Bell AE, Culp PA. Reduction in Saturated Fat Intake for Cardiovascular Disease. *Am Fam Physician* [Internet]. NLM (Medline); 2022 Jan 1 [cited 2022 Nov 17];**105**(1). Available from: https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD011737.pub2/full

86. Chiu S, Williams PT, Krauss RM. Effects of a very high saturated fat diet on LDL particles in adults with atherogenic dyslipidemia: A randomized controlled trial. *PLoS One* [Internet]. Public Library of Science; 2017 Feb 1 [cited 2022 Nov

17];**12**(2):e0170664. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0170664

87. Han Q, Yeung SC, Ip MSM, Mak JCW. Dysregulation of cardiac lipid parameters in high-fat high-cholesterol diet-induced rat model. [cited 2022 Nov 17]; Available from: https://doi.org/10.1186/s12944-018-0905-3

88. Vincent MJ, Allen B, Palacios OM, Haber LT, Maki KC. Meta-regression analysis of the effects of dietary cholesterol intake on LDL and HDL cholesterol. *Am J Clin Nutr* [Internet]. Oxford Academic; 2019 Jan 1 [cited 2022 Nov 17];**109**(1):7–16. Available from: https://academic.oup.com/ajcn/article/109/1/7/5266898

89. May-Wilson S, Sud A, Law PJ, et al. Pro-inflammatory fatty acid profile and colorectal cancer risk: A Mendelian randomisation analysis. *Eur J Cancer* [Internet]. Eur J Cancer; 2017 Oct 1 [cited 2022 Nov 28];**84**:228–238. Available from: https://pubmed.ncbi.nlm.nih.gov/28829991/

90. Khankari NK, Banbury BL, Borges MC, et al. Mendelian Randomization of Circulating Polyunsaturated Fatty Acids and Colorectal Cancer Risk. *Cancer Epidemiol Biomarkers Prev* [Internet]. Cancer Epidemiol Biomarkers Prev; 2020 [cited 2022 Nov 23];**29**(4):860–870. Available from: https://pubmed.ncbi.nlm.nih.gov/32051193/

91. Liyanage UE, Ong JS, An J, Gharahkhani P, Law MH, MacGregor S. Mendelian Randomization Study for Genetically Predicted Polyunsaturated Fatty Acids Levels on Overall Cancer Risk and Mortality. *Cancer Epidemiol Biomarkers Prev* [Internet]. Cancer Epidemiol Biomarkers Prev; 2019 Jun 1 [cited 2022 Nov 23];**28**(6):1015–1023. Available from: https://pubmed.ncbi.nlm.nih.gov/30948449/

Acknowledgements: ALSPAC: We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses.

EPICOLON: We are sincerely grateful to all patients participating in this study who were recruited as part of the EPICOLON project. We acknowledge the Spanish National DNA Bank, Biobank of Hospital Clínic–IDIBAPS and Biobanco Vasco for the availability of the samples. The work was carried out (in part) at the Esther Koplowitz Centre, Barcelona.

Data availability: Individual-level ALSPAC data are available following an application. This process of managed access is detailed at www.bristol.ac.uk/alspac/researchers/access. Cohort details and data descriptions for ALSPAC are publicly available at the same web address. Summary-level GWAS data used in this study from UK Biobank are publicly available without the need for application through the MR-Base platform, which is accessible at http://www.mrbase.org/. The summary-level GWAS data for CRC used in this study are available following an application to GECCO (managed access). All data generated by this study are

available in the manuscript and supporting material. R scripts used in this study have been made publicly available on GitHub at: https://github.com/cb12104/adiposity_metabolites_crc.

Funding:

Author funding

JAB is supported by the Elizabeth Blackwell Institute for Health Research, University of Bristol and the Wellcome Trust Institutional Strategic Support Fund (204813/Z/16/Z) and works in a Unit funded by the Medical Research Council (MC_UU_00011/1) and the University of Bristol. EEV, DNL and CB are supported by Diabetes UK (17/0005587). NJT is a Wellcome Trust Investigator (202802/Z/16/Z), is the PI of the Avon Longitudinal Study of Parents and Children (MRC & WT 102215/2/13/2), is supported by the University of Bristol NIHR Biomedical Research Centre (BRC-1215-20011), the MRC Integrative Epidemiology Unit (MC_UU_00011/1) and works within the CRUK Integrative Cancer Epidemiology Programme (C18281/A19169). EH is supported by a Cancer Research UK Population Research Committee Studentship (C18281/A30905), is supported by the CRUK Integrative Cancer Epidemiology Programme (C18281/A29019) and is part of the Medical Research Council Integrative Epidemiology Unit at the University of Bristol which is supported by the Medical Research Council (MC_UU_00011/4) and the University of Bristol. AC acknowledges funding from grant MR/N0137941/1 for the GW4 BIOMED MRC DTP, awarded to the Universities of Bath, Bristol, Cardiff and Exeter from the Medical Research Council (MRC)/UKRI. MCB was supported by the UK Medical Research Council (MRC) Skills Development Fellowship (MR/P014054/1), University of Bristol Vice-Chancellor's Fellowship and MRC Integrative Epidemiology Unit (MC_UU_00011/6). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. This work used the computational facilities of the Advanced Computing Research Centre, University of Bristol - http://www.bristol.ac.uk/acrc/.

Study funding

ALSPAC: The UK Medical Research Council and Wellcome (Grant ref: 217065/Z/19/Z) and the University of Bristol provide core support for ALSPAC. This publication is the work of the authors who will serve as guarantors for the contents of this paper. A comprehensive list of grants funding is available on the ALSPAC website (http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf); This research was specifically funded by The UK Medical Research Council (Grant ref: MC_UU_12013/1). GWAS data was generated by Sample Logistics and

Genotyping Facilities at Wellcome Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe.

Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO): National Cancer Institute, National Institutes of Health, U.S. Department of Health and Human Services (U01 CA137088, R01 CA059045, R01 201407). Genotyping/Sequencing services were provided by the Center for Inherited Disease Research (CIDR) contract number HHSN268201700006I and HHSN268201200008I . This research was funded in part through the NIH/NCI Cancer Center Support Grant P30 CA015704. Scientific Computing Infrastructure at Fred Hutch funded by ORIP grant S10OD028685.

ASTERISK: a Hospital Clinical Research Program (PHRC-BRD09/C) from the University Hospital Center of Nantes (CHU de Nantes) and supported by the Regional Council of Pays de la Loire, the Groupement des Entreprises Françaises dans la Lutte contre le Cancer (GEFLUC), the Association Anne de Bretagne Génétique and the Ligue Régionale Contre le Cancer (LRCC).

The ATBC Study is supported by the Intramural Research Program of the U.S. National Cancer Institute, National Institutes of Health, Department of Health and Human Services.

CLUE II funding was from the National Cancer Institute (U01 CA086308, Early Detection Research Network; P30 CA006973), National Institute on Aging (U01 AG018033), and the American Institute for Cancer Research. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US government.

Maryland Cancer Registry (MCR): Cancer data was provided by the Maryland Cancer Registry, Center for Cancer Prevention and Control, Maryland Department of Health, with funding from the State of Maryland and the Maryland Cigarette Restitution Fund. The collection and availability of cancer registry data is also supported by the Cooperative Agreement NU58DP006333, funded by the Centers for Disease Control and Prevention. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention or the Department of Health and Human Services.

ColoCare: This work was supported by the National Institutes of Health (grant numbers R01 CA189184 (Li/Ulrich), U01 CA206110 (Ulrich/Li/Siegel/Figueiredo/Colditz, 2P30CA015704- 40 (Gilliland), R01 CA207371 (Ulrich/Li)), the Matthias Lackas-Foundation, the German Consortium for Translational Cancer Research, and the EU TRANSCAN initiative.

The Colon Cancer Family Registry (CCFR, www.coloncfr.org) is supported in part by funding from the National Cancer Institute (NCI), National Institutes of Health (NIH) (award U01 CA167551). Support for case ascertainment was provided in part from the Surveillance, Epidemiology, and End Results (SEER) Program and the following U.S. state cancer registries: AZ, CO, MN, NC, NH; and by the Victoria Cancer Registry (Australia) and Ontario Cancer Registry (Canada). The CCFR Set-1 (Illumina 1M/1M-Duo) and Set-2 (Illumina Omni1-Quad) scans were supported by NIH awards U01 CA122839 and R01 CA143237 (to GC). The CCFR Set-3 (Affymetrix Axiom CORECT Set array) was supported by NIH award U19 CA148107 and R01 CA81488 (to SBG). The CCFR Set-4 (Illumina OncoArray 600K SNP array) was supported by NIH award U19 CA148107 (to SBG) and by the Center for Inherited Disease Research (CIDR), which is funded by the NIH to the Johns Hopkins University, contract number HHSN268201200008I. Additional funding for the OFCCR/ARCTIC was through award GL201-043 from the Ontario Research Fund (to BWZ), award 112746 from the Canadian Institutes of Health Research (to TJH), through a Cancer Risk Evaluation (CaRE) Program grant from the Canadian Cancer Society (to SG), and through generous support from the Ontario Ministry of Research and Innovation. The SFCCR Illumina HumanCytoSNP array was supported in part through NCI/NIH awards U01/U24 CA074794 and R01 CA076366 (to PAN). The content of this manuscript does not necessarily reflect the views or policies of the NCI, NIH or any of the collaborating centers in the Colon Cancer Family Registry (CCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government, any cancer registry, or the CCFR.

COLON: The COLON study is sponsored by Wereld Kanker Onderzoek Fonds, including funds from grant 2014/1179 as part of the World Cancer Research Fund International Regular Grant Programme, by Alpe d'Huzes and the Dutch Cancer Society (UM 2012–5653, UW 2013-5927, UW2015-7946), and by TRANSCAN (JTC2012-MetaboCCC, JTC2013-FOCUS). The Nqplus study is sponsored by a ZonMW investment grant (98-10030); by PREVIEW, the project PREVention of diabetes through lifestyle intervention and population studies in Europe and around the World (PREVIEW) project which received funding from the European Union Seventh Framework Programme (FP7/2007–2013) under grant no. 312057; by funds from TI Food and

Nutrition (cardiovascular health theme), a public–private partnership on precompetitive research in food and nutrition; and by FOODBALL, the Food Biomarker Alliance, a project from JPI Healthy Diet for a Healthy Life.

COLO2&3: National Institutes of Health (R01 CA060987)

Colorectal Cancer Transdisciplinary (CORECT) Study: The CORECT Study was supported by the National Cancer Institute, National Institutes of Health (NCI/NIH), U.S. Department of Health and Human Services (grant numbers U19 CA148107, R01 CA081488, P30 CA014089, R01 CA197350; P01 CA196569; R01 CA201407; R01 CA242218), National Institutes of Environmental Health Sciences, National Institutes of Health (grant number T32 ES013678) and a generous gift from Daniel and Maryann Fong.

CORSA: The CORSA study was funded by Austrian Research Funding Agency (FFG) BRIDGE (grant 829675, to Andrea Gsur), the “Herzfelder’sche Familienstiftung” (grant to Andrea Gsur) and was supported by COST Action BM1206.

CPS-II: The American Cancer Society funds the creation, maintenance, and updating of the Cancer Prevention Study-II (CPS-II) cohort. The study protocol was approved by the institutional review boards of Emory University, and those of participating registries as required.

CRCGEN: Colorectal Cancer Genetics & Genomics, Spanish study was supported by Instituto de Salud Carlos III, co-funded by FEDER funds –a way to build Europe– (grants PI14-613 and PI09-1286), Agency for Management of University and Research Grants (AGAUR) of the Catalan Government (grant 2017SGR723), Junta de Castilla y León (grant LE22A10-2), the Spanish Association Against Cancer (AECC) Scientific Foundation grant GCTRA18022MORE and the Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), action Genrisk. Sample collection of this work was supported by the Xarxa de Bancs de Tumors de Catalunya sponsored by Pla Director d’Oncología de Catalunya (XBTC), Plataforma Biobancos PT13/0010/0013 and ICOBIOBANC, sponsored by the Catalan Institute of Oncology. We thank CERCA Programme, Generalitat de Catalunya for institutional support.

Czech Republic CCS: This work was supported by the Grant Agency of the Czech Republic (21-04607X, 20-03997S), by the Grant Agency of the Ministry of Health of the Czech Republic (grants AZV NU21-07-00247 and AZV NU21-03-00506), and Charles University Research Fund (Cooperation 43-Surgical disciplines)

DACHS: This work was supported by the German Research Council (BR 1704/6-1, BR 1704/6-3, BR 1704/6-4, CH 117/1-1, HO 5117/2-1, HE 5998/2-1, KL 2354/3-1, RO 2270/8-1 and BR 1704/17-1), the Interdisciplinary Research Program of the National Center for Tumor Diseases (NCT), Germany, and the German Federal Ministry of Education and Research (01KH0404, 01ER0814, 01ER0815, 01ER1505A and 01ER1505B).

DALS: National Institutes of Health (R01 CA048998 to M. L. Slattery).

EDRN: This work is funded and supported by the NCI, EDRN Grant (U01-CA152753).

EPIC: The coordination of EPIC is financially supported by International Agency for Research on Cancer (IARC) and also by the Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London which has additional infrastructure support provided by the NIHR Imperial Biomedical Research Centre (BRC). The national cohorts are supported by: Danish Cancer Society (Denmark); Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Education Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research Center (DKFZ), German Institute of Human Nutrition Potsdam- Rehbruecke (DIfE), Federal Ministry of Education and Research (BMBF) (Germany); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy, Compagnia di SanPaolo and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands); Health Research Fund (FIS) - Instituto de Salud Carlos III (ISCIII), Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, and the Catalan Institute of Oncology - ICO (Spain); Swedish Cancer Society, Swedish Research Council and and Region Skåne and Region Västerbotten (Sweden); Cancer Research UK (14136 to EPIC-Norfolk; C8221/A29017 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk; MR/M012190/1 to EPIC-Oxford). (United Kingdom).

EPICOLON: This work was supported by grants from Fondo de Investigación Sanitaria/FEDER (PI08/0024, PI08/1276, PS09/02368, PI11/00219, PI11/00681, PI14/00173, PI14/00230, PI17/00509, 17/00878, PI20/00113, PI20/00226, Acción Transversal de Cáncer), Xunta de Galicia (PGIDIT07PXIB9101209PR), Ministerio de Economia y Competitividad (SAF07-64873, SAF 2010-19273, SAF2014-54453R), Fundación Científica de la Asociación Española contra el Cáncer (GCB13131592CAST, PRYGN211085CAST), Beca Grupo de Trabajo "Oncología" AEG (Asociación Española de Gastroenterología), Fundación Privada Olga Torres, FP7 CHIBCHA

Consortium, Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR, Generalitat de Catalunya, 2014SGR135, 2014SGR255, 2017SGR21, 2017SGR653, 2021SGR00716, 2021SGR01185), Catalan Tumour Bank Network (Pla Director d'Oncologia, Generalitat de Catalunya), PERIS (SLT002/16/00398, Generalitat de Catalunya), Marató TV3 (202008-10), CERCA Programme (Generalitat de Catalunya) and COST Actions BM1206 and CA17118. CIBERehd is funded by the Instituto de Salud Carlos III.ESTHER/VERDI. This work was supported by grants from the Baden-Württemberg Ministry of Science, Research and Arts and the German Cancer Aid.

Harvard cohorts: HPFS is supported by the National Institutes of Health (P01 CA055075, UM1 CA167552, U01 CA167552, R01 CA137178, R01 CA151993, and R35 CA197735), NHS by the National Institutes of Health (P01 CA087969, UM1 CA186107, R01 CA137178, R01 CA151993, and R35 CA197735), and PHS by the National Institutes of Health (R01 CA042182).

Hawaii Adenoma Study: NCI grants R01 CA072520.

HCES-CRC: the Hwasun Cancer Epidemiology Study–Colon and Rectum Cancer (HCES-CRC; grants from Chonnam National University Hwasun Hospital, HCRI15011-1).

Kentucky: This work was supported by the following grant support: Clinical Investigator Award from Damon Runyon Cancer Research Foundation (CI-8); NCI R01CA136726.

LCCS: The Leeds Colorectal Cancer Study was funded by the Food Standards Agency and Cancer Research UK Programme Award (C588/A19167).

MCCS cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further supported by Australian NHMRC grants 509348, 209057, 251553 and 504711 and by infrastructure provided by Cancer Council Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry (VCR) and the Australian Institute of Health and Welfare (AIHW), including the National Death Index and the Australian Cancer Database. BMLynch was supported by MCRF18005 from the Victorian Cancer Agency.

MEC: National Institutes of Health (R37 CA054281, P01 CA033619, and R01 CA063464).

MECC: This work was supported by the National Institutes of Health, U.S. Department of Health and Human Services (R01 CA081488, R01 CA197350, U19 CA148107, R01 CA242218, and a generous gift from Daniel and Maryann Fong.

MSKCC: The work at Sloan Kettering in New York was supported by the Robert and Kate Niehaus Center for Inherited Cancer Genomics and the Romeo Milio Foundation. Moffitt: This work was supported by funding from the National Institutes of Health (grant numbers R01 CA189184, P30 CA076292), Florida Department of Health Bankhead-Coley Grant 09BN-13, and the University of South Florida Oehler Foundation. Moffitt contributions were supported in part by the Total Cancer Care Initiative, Collaborative Data Services Core, and Tissue Core at the H. Lee Moffitt Cancer Center & Research Institute, a National Cancer Institute-designated Comprehensive Cancer Center (grant number P30 CA076292).

NCCCS I & II: We acknowledge funding support for this project from the National Institutes of Health, R01 CA066635 and P30 DK034987.

NFCCR: This work was supported by an Interdisciplinary Health Research Team award from the Canadian Institutes of Health Research (CRT 43821); the National Institutes of Health, U.S. Department of Health and Human Serivces (U01 CA074783); and National Cancer Institute of Canada grants (18223 and 18226). The authors wish to acknowledge the contribution of Alexandre Belisle and the genotyping team of the McGill University and Génome Québec Innovation Centre, Montréal, Canada, for genotyping the Sequenom panel in the NFCCR samples. Funding was provided to Michael O. Woods by the Canadian Cancer Society Research Institute.

NSHDS: The research was supported by Biobank Sweden through funding from the Swedish Research Council (VR 2017-00650, VR 2017-01737), the Swedish Cancer Society (CAN 2017/581), Region Västerbotten (VLL-841671, VLL-833291), Knut and Alice Wallenberg Foundation (VLL-765961), and the Lion's Cancer Research Foundation (several grants) and Insamlingsstiftelsen, both at Umeå University.

OSUMC: OCCPI funding was provided by Pelotonia and HNPCC funding was provided by the NCI (CA016058 and CA067941).

PLCO: Intramural Research Program of the Division of Cancer Epidemiology and Genetics and supported by contracts from the Division of Cancer Prevention, National Cancer Institute, NIH, DHHS. Funding was provided by National Institutes of Health (NIH), Genes, Environment and Health Initiative (GEI) Z01 CP 010200, NIH U01 HG004446, and NIH GEI U01 HG 004438.

SEARCH: The University of Cambridge has received salary support in respect of PDPP from the NHS in the East of England through the Clinical Academic Reserve. Cancer Research UK

(C490/A16561); the UK National Institute for Health Research Biomedical Research Centres at the University of Cambridge.

SELECT: Research reported in this publication was supported in part by the National Cancer Institute of the National Institutes of Health under Award Numbers U10 CA037429 (CD Blanke), and UM1 CA182883 (CM Tangen/IM Thompson). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

SMS and REACHS: This work was supported by the National Cancer Institute (grant P01 CA074184 to J.D.P. and P.A.N., grants R01 CA097325, R03 CA153323, and K05 CA152715 to P.A.N., and the National Center for Advancing Translational Sciences at the National Institutes of Health (grant KL2 TR000421 to A.N.B.-H.)

The Swedish Low-risk Colorectal Cancer Study: The study was supported by grants from the Swedish research council; K2015-55X-22674-01-4, K2008-55X-20157-03-3, K2006-72X-20157-01-2 and the Stockholm County Council (ALF project).

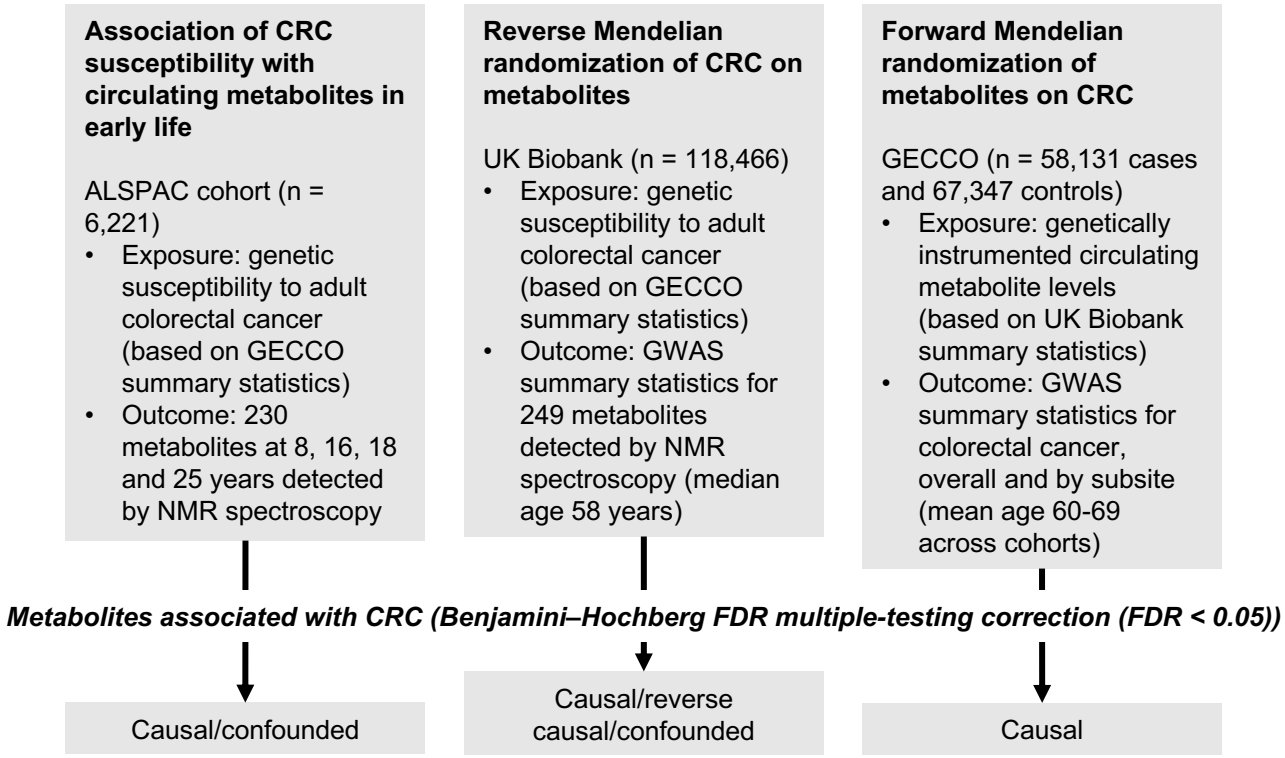
Swedish Mammography Cohort and Cohort of Swedish Men: This work is supported by the Swedish Research Council /Infrastructure grant, the Swedish Cancer Foundation, and the Karolinska Institute´s Distinguished Professor Award to Alicja Wolk.

UK Biobank: This research has been conducted using the UK Biobank Resource under Application Number 8614

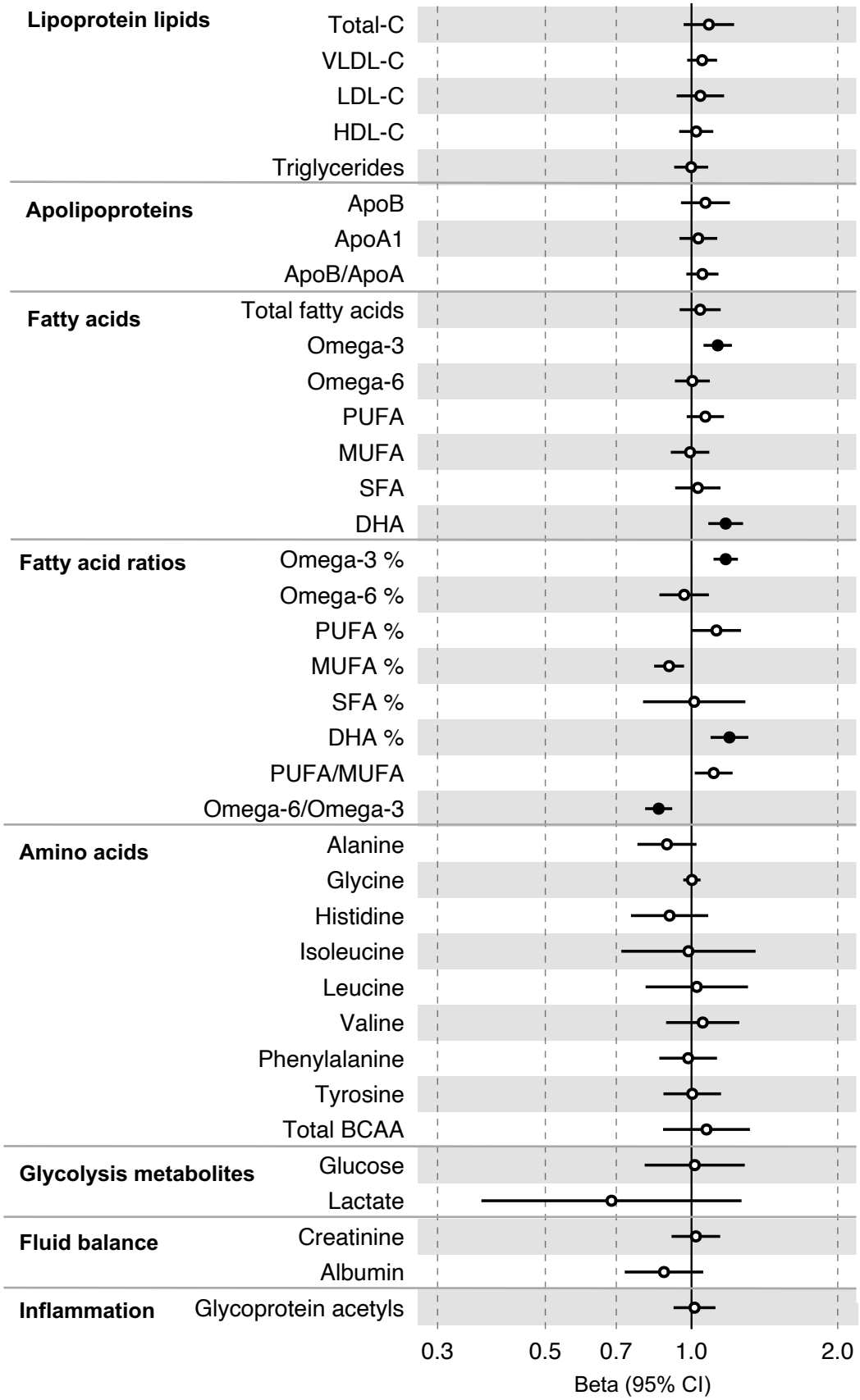
VITAL: National Institutes of Health (K05 CA154337).

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 75N92021D00001, 75N92021D00002, 75N92021D00003, 75N92021D00004, 75N92021D00005

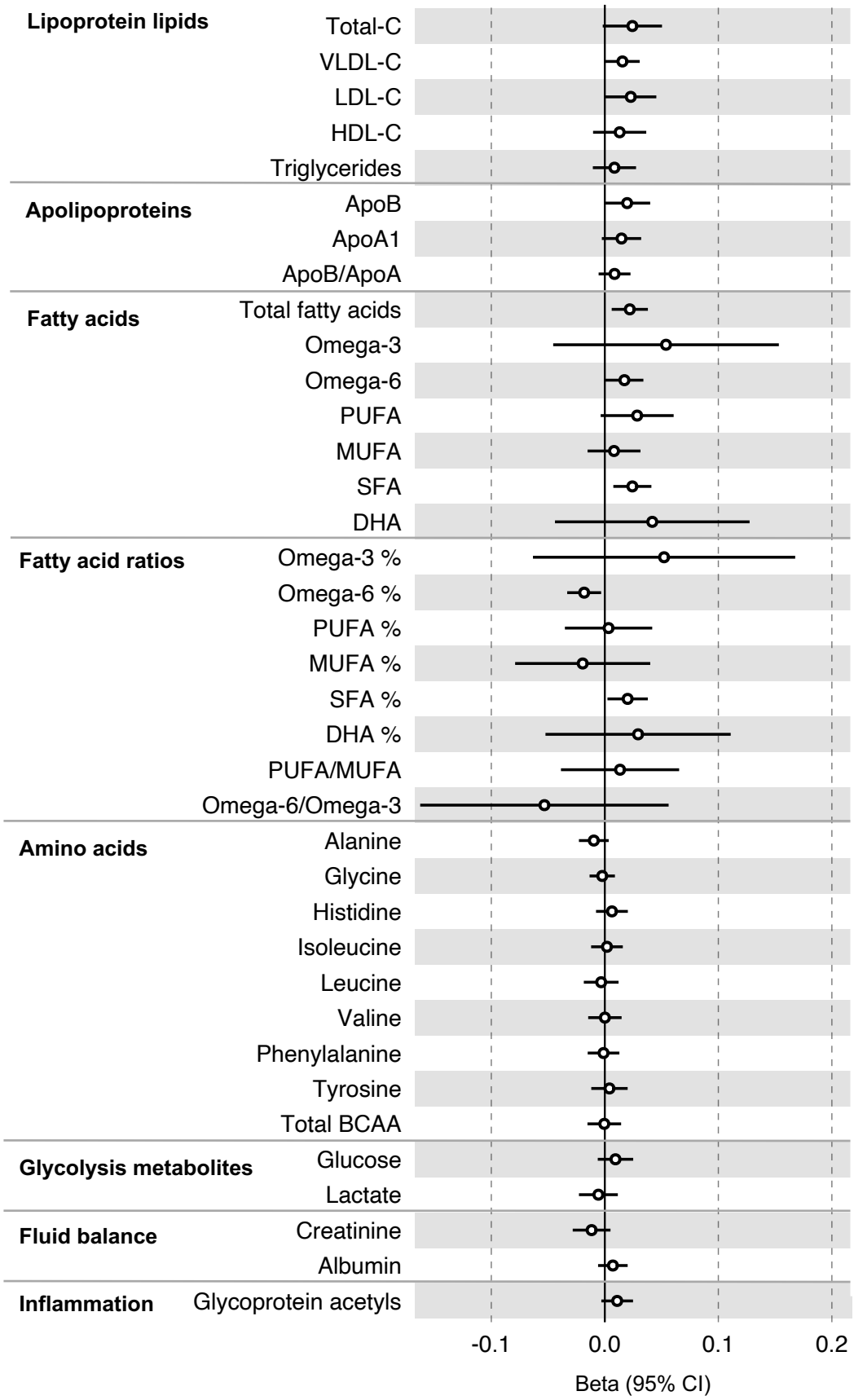
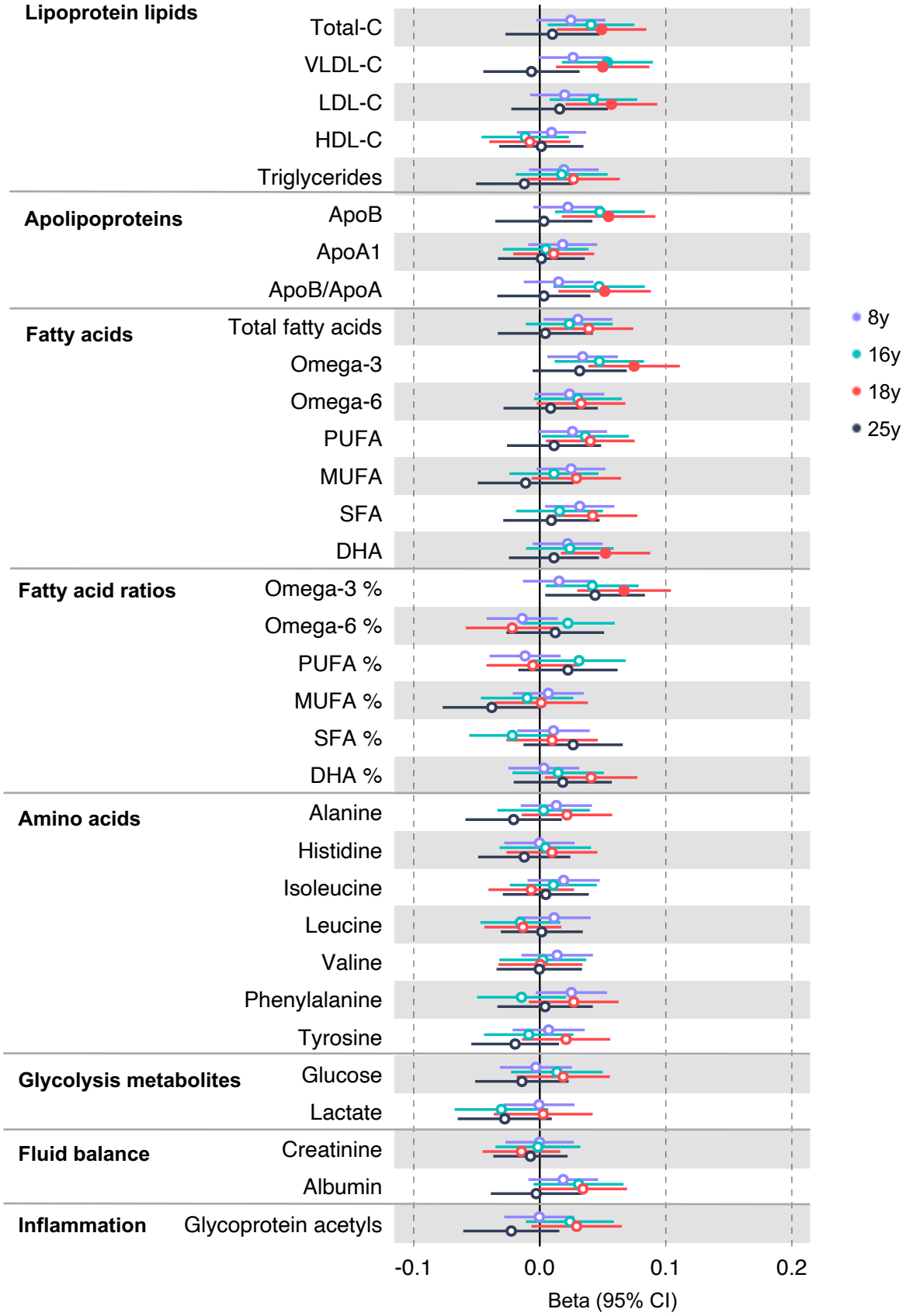
medRxiv preprint doi: https://doi.org/10.1101/2023.03.10.23287084; this version posted November 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .



medRxiv preprint doi: https://doi.org/10.1101/2023.03.10.23287084; this version posted November 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .



medRxiv preprint doi: https://doi.org/10.1101/2023.03.10.23287084; this version posted November 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .



medRxiv preprint doi: https://doi.org/10.1101/2023.03.10.23287084; this version posted November 9, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .