

## Identifying metabolic features of colorectal cancer liability using Mendelian randomization

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

<sup>1</sup> MRC Integrative Epidemiology Unit at the University of Bristol, Bristol, UK

<sup>2</sup> Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

<sup>3</sup> Translational Health Sciences, Bristol Medical School, University of Bristol, UK

<sup>4</sup> Public Health Sciences Division, Fred Hutchinson Cancer Center, Seattle, Washington, USA

<sup>5</sup> Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany

<sup>6</sup> Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany

<sup>7</sup> German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany

<sup>8</sup> 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

<sup>9</sup> Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA

<sup>10</sup> Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA

<sup>11</sup> Clinical and Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA

<sup>12</sup> Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA

<sup>13</sup> Department of Epidemiology, Harvard T.H. Chan School of Public Health, Harvard University, Boston, Massachusetts, USA

<sup>14</sup> Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Harvard University, Boston, Massachusetts, USA

<sup>15</sup> Department of Preventive Medicine, Chonnam National University Medical School, Gwangju, Korea

<sup>16</sup> 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.

<sup>17</sup> University of Hawaii Cancer Center, Honolulu, Hawaii, USA

<sup>18</sup> Department of Family Medicine, University of Virginia, Charlottesville, Virginia, USA

<sup>19</sup> Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, California, USA

<sup>20</sup> University of California, San Francisco Helen Diller Family Comprehensive Cancer Center, San Francisco, San Francisco, California, USA

<sup>21</sup> Department of Pathology and Laboratory Medicine, Mayo Clinic, Arizona, Scottsdale, Arizona, USA

<sup>22</sup> Department of Medicine, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA

<sup>23</sup> Nutrition and Metabolism Branch, International Agency for Research on Cancer, Lyon, France

<sup>24</sup> 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](mailto: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)

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

1 **Introduction**

2

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

14

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

30

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

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

47  
48 Here, we applied a reverse MR framework to identify circulating metabolites which are associated  
49 with CRC liability across different stages of the early life course (spanning childhood to young  
50 adulthood, when use of medications and CRC are both rare) using data from a birth cohort study.  
51 We then attempted to replicate these results using reverse two-sample MR in an independent  
52 cohort of middle-aged adults (UK Biobank). We then performed conventional ‘forward’ MR of  
53 metabolites onto CRC risk using large-scale cancer consortia data to identify metabolites which  
54 may have a causal role in CRC development.

55

56

57

## 58 **Methods**

59

### 60 *Study populations*

61

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

68

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

87  
88 UK Biobank is a population-based cohort study based in 22 centres across the UK.<sup>36</sup> The cohort  
89 is made up of around 500,000 adults aged 40-80 years old, who were enrolled between 2006 and  
90 2010. Genotyping data is available for 488,377 participants.<sup>37</sup> Participants were genotyped using  
91 one of two arrays – either the Applied Biosystems UK BiLEVE Axiom Array by Affymetrix (now  
92 part of Thermo Fisher Scientific), or the closely related Applied Biosystems UK Biobank Axiom  
93 Array. Approaches based on Principal Component Analysis (PCA) were used to account for  
94 population structure. Individuals were excluded: if reported sex differed from inferred sex based  
95 on genotyping data; if they had sex chromosome karyotypes which were not XX or XY; if they  
96 were outliers in terms of heterozygosity and missing rates; or if they had high relatedness to  
97 another participant. Multiallelic SNPs or those with a minor allele frequency of below 1% were  
98 removed. Imputation was performed using the UK10K haplotype and HRC reference panels.

99  
100 The GWAS meta-analysis for CRC included up to 52,775 cases and 45,940 controls.<sup>38,39</sup> This  
101 sample excluded cases and controls from UK Biobank to avoid potential bias due to sample

102 overlap which may be problematic in MR analyses.<sup>40</sup> Cases were diagnosed by a physician and  
103 recorded overall and by site (colon, 28,736 cases; proximal colon, 14,416 cases; distal colon,  
104 12,879 cases; and rectal, 14,150 cases). Colon cancer included proximal colon (any primary  
105 tumour arising in the cecum, ascending colon, hepatic flexure, or transverse colon), distal colon  
106 (any primary tumour arising in the pleenic flexure, descending colon or sigmoid colon), and colon  
107 cases with unspecified site. Rectal cancer included any primary tumour arising in the rectum or  
108 rectosigmoid junction.<sup>38</sup> Approximately 92% of participants in the overall CRC GWAS were white-  
109 European (~8% were East Asian). All participants included in site-specific CRC analyses were of  
110 European ancestry. Imputation was performed using the Michigan imputation server and HRC  
111 r1.0 reference panel. Regression models were further adjusted for age, sex, genotyping platform,  
112 and genomic principal components as described previously.<sup>38</sup>

113

#### 114 *Assessment of CRC genetic liability*

115

116 Genetic liability to CRC was based on single nucleotide polymorphisms (SNPs) associated with  
117 CRC case status at genome-wide significance ( $P < 5 \times 10^{-8}$ ). 108 independent SNPs reported by  
118 two major GWAS meta-analyses were eligible for inclusion in a CRC genetic risk score (GRS).<sup>38,41</sup>  
119 The set of SNPs was filtered, excluding 36 SNPs that were in linkage disequilibrium based on  
120  $R^2 > 0.001$  using the TwoSampleMR package (SNPs with the lowest P-values were retained).<sup>42</sup>  
121 This left 72 SNPs independently associated with CRC (**Supplementary File 1a**), 65 of which  
122 were available in imputed ALSPAC genotype data post quality control. As GWAS of site-specific  
123 CRC have identified marked heterogeneity,<sup>43</sup> GRS describing site-specific CRCs were  
124 constructed for sensitivity analyses using the same process outlined above. The GRS for colon  
125 cancer, rectal cancer, proximal colon cancer and distal colon cancer were comprised of 38, 25,  
126 20 and 24 variants, respectively (**Supplementary File 1a**). For overall CRC and site-specific CRC  
127 analyses, sensitivity analyses excluding any SNPs in the FADS cluster (i.e. within the gene  
128 regions of *FADS1*, *FADS2*, or *FADS3*) (**Supplementary File 1a**) were performed given a likely  
129 role for these SNPs in influencing circulating metabolite levels directly, in particular via lipid  
130 metabolism (i.e., not primarily due to CRC).<sup>44-50</sup>

131

#### 132 *Assessment of circulating metabolites*

133

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

136 during a clinic visit while aged approximately 8y, and fasting blood samples from clinic visits while  
137 aged approximately 16y, 18y, and 25y. Proton nuclear magnetic resonance (<sup>1</sup>H-NMR)  
138 spectroscopy was performed on Ethylenediaminetetraacetic acid (EDTA) plasma (stored at or  
139 below -70 degrees Celsius pre-processing) to quantify a maximum of 231 metabolites.<sup>51</sup>  
140 Quantified metabolites included the cholesterol and triglyceride content of lipoprotein particles;  
141 the concentrations and diameter/size of these particles; apolipoprotein B and apolipoprotein A-1  
142 concentrations; as well as fatty acids and their ratios to total fatty acid concentration, branched  
143 chain and aromatic amino acids, glucose and pre-glycaemic factors including lactate and citrate,  
144 fluid balance factors including albumin and creatinine, and the inflammatory marker glycoprotein  
145 acetyls (GlycA). This metabolomics platform has limited coverage of fatty acids. In UK Biobank,  
146 EDTA plasma samples from 117,121 participants, a random subset of the original ~500,000 who  
147 provided samples at assessment centres between 2006 and 2013, were analysed between 2019  
148 and 2020 for levels of 249 metabolic traits (168 concentrations plus 81 ratios) using the same  
149 high-throughput <sup>1</sup>H-NMR platform. Data pre-processing and QC steps are described  
150 previously.<sup>51-53</sup> To allow comparability between MR and GRS estimates all metabolite measures  
151 were standardised and normalised using rank-based inverse normal transformation. For  
152 descriptive purposes in ALSPAC, body mass index (BMI) was calculated at each time point as  
153 weight (kg) divided by squared height (m<sup>2</sup>) based on clinic measures of weight to the nearest  
154 0.1 kg using a Tanita scale and height measured in light clothing without shoes to the nearest 0.1  
155 cm using a Harpenden stadiometer.

156

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

162

### 163 *Statistical approach*

164

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



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

173

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

175

176 Separate linear regression models with robust standard errors were used to estimate coefficients  
177 and 95% confidence intervals for associations of GRSs with each metabolite as a dependent  
178 variable measured on the same individuals at age 8y, 16y, 18y, and 25y, adjusted for sex and  
179 age at the time of metabolite assessment. To aid interpretations, estimates were multiplied by  
180 0.693 ( $\log_e 2$ ) to reflect SD-unit differences in metabolites per doubling of genetic liability to CRC.<sup>55</sup>  
181 The Benjamini-Hochberg method was used to adjust P-values for multiple testing and an adjusted  
182 P-value of  $<0.05$  was used as a heuristic for evidence for association given current sample sizes.<sup>56</sup>

183

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

185

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

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

207

### 208 *3. Forward MR of the effects of metabolites on CRC*

209

210 Forward MR analyses were conducted using summary statistics from UK Biobank for the same  
211 NMR-measured metabolites (SNP-exposure) and from GECCO/CORECT/CCFR as outlined  
212 above (SNP-outcome). We identified SNPs that were independently associated ( $R^2 < 0.001$  and  
213  $P < 5 \times 10^{-8}$ ) with metabolites from a GWAS of 249 metabolites in UK Biobank described above. As  
214 before, we used up to 3 statistical methods to generate MR estimates of the effect of circulating  
215 metabolites on CRC risk (overall and site-specific): random-effects IVW, weighted-median, and  
216 weighted-mode. The Benjamini-Hochberg method was used to adjust P-values for multiple testing  
217 and an adjusted P-value of  $< 0.05$  was used as a heuristic for nominal evidence for a causal  
218 effect.<sup>56</sup> MR outputs are beta coefficients representing the logOR for CRC per SD higher  
219 metabolite, exponentiated to reflect the OR for CRC per SD metabolite.

220

221 MR analyses were performed in R version 4.0.3.<sup>67</sup> and GRS analyses in Stata 16.1 (StataCorp,  
222 College Station, Texas, USA). The ggforestplot R package was used to generate results  
223 visualisations.<sup>68</sup>

224

### 225 *Ethics*

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

233

### 234 **Results**

235

236

237 *Associations of CRC liability with circulating metabolites in early life*

238

239 At the time the ALSPAC blood samples were taken, the mean age of participants was 7.5y  
240 (N=4,767), 15.5y (N=2,930), 17.8y (N=2,613), and 24.5y (N=2,559) for the childhood, early  
241 adolescence, late adolescence and young adulthood time points respectively. The proportion of  
242 participants which were male were 50.5%, 47.4%, 44.5%, and 39.1% and mean BMI was 16.2,  
243 21.4, 22.7, and 24.8 kg/m<sup>2</sup> for each time point respectively. The socio-demographic profile of  
244 ALSPAC offspring participants has been reported previously.<sup>69</sup> Mean and standard deviation (SD)  
245 values for metabolites on each measurement occasion in ALSPAC are shown in **Supplementary**  
246 **File 1b**.

247

248 In the GRS analysis, there was no strong evidence of association of CRC liability with metabolites  
249 at age 8y (**Supplementary File 1c**). At age 16y, there was evidence for association with several  
250 lipid traits including higher cholesteryl esters to total lipids ratio in large low-density lipoprotein  
251 (LDL) (SD change per doubling CRC liability = 0.06, 95% CI = 0.02 to 0.10) and higher cholesterol  
252 in very small very low-density lipoprotein (VLDL) (SD change per doubling CRC liability = 0.06,  
253 95% CI = 0.03 to 0.10). There was strong evidence for association with several traits at age 18y  
254 including higher non-high-density lipoprotein (non-HDL) lipids, e.g., a 1 doubling CRC liability was  
255 associated with higher levels of total cholesterol (SD change = 0.05 95% CI = 0.01 to 0.09), VLDL-  
256 cholesterol (SD change = 0.05, 95% CI = 0.01 to 0.09), LDL-cholesterol (SD change = 0.06, 95%  
257 CI = 0.02 to 0.09)), apolipoproteins (apolipoprotein B (SD change = 0.06, 95% CI = 0.02 to 0.09)),  
258 and fatty acids (omega-3 (SD change = 0.08, 95% CI = 0.04 to 0.11), docosahexaenoic acid  
259 (DHA) (SD change = 0.05, 95% CI = 0.02 to 0.09)) (**Supplementary File 1c**). **Figure 2(-figure**  
260 **supplements 1-6)** shows results for all clinically validated metabolites. At age 25y, there was no  
261 strong evidence of association of CRC liability with metabolites. In anatomical site-specific  
262 analyses, there was strong evidence for association of liability to colon cancer with omega-3 (SD  
263 change = 0.07, 95% CI = 0.03 to 0.11) and DHA (SD change = 0.07, 95% CI = 0.03 to 0.10) at  
264 age 18y. There was little evidence for any associations at any other CRC site or age  
265 (**Supplementary File 1c**). When SNPs in the FADS cluster gene regions were excluded due to  
266 possible horizontal pleiotropy given the role of FADS in lipid metabolism, there was a reduction in  
267 strength of evidence for an association of liability to CRC with any metabolite measured, although  
268 estimates were in a largely consistent direction with the prior analysis (**Supplementary File 1d**).

269

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

271

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

284

285 *Forward MR for the effects of metabolites on CRC risk*

286

287 All instrument sets from the forward MR analysis had an F-statistic greater than 10 (minimum F-  
288 statistic = 54, median = 141), suggesting that our analyses were unlikely to suffer from weak  
289 instrument bias (**Supplementary File 1h-i**). There was strong evidence for an effect of several  
290 fatty acid traits on overall CRC risk, including of omega-3 fatty acids (CRC OR = 1.13, 95% CI =  
291 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  
292 fatty acids (OR CRC = 1.18, 95% CI = 1.11 to 1.25), ratio of DHA to total fatty acids (CRC OR =  
293 1.20, 95% CI = 1.10 to 1.31), and ratio of omega-6 fatty acids to omega-3 fatty acids (CRC OR =  
294 0.86, 95% CI = 0.80 to 9.13) (**Supplementary File 1j, Figure 4-figure supplements 1-3**). These  
295 estimates were overlapping with variable precision in MR sensitivity models. When SNPs in the  
296 *FADS* gene region were excluded, there was little evidence for a causal effect of any metabolite  
297 investigated on CRC risk based on the predetermined FDR-P cut of off < 0.05, although the  
298 directions of effect estimates were consistent with previous analyses (**Supplementary File 1k**).

299

300 In anatomical subtype stratified analyses evidence was strongest for an effect of fatty acid traits  
301 on higher CRC risk, and this appeared specific to the distal colon, e.g., omega-3 (distal CRC OR  
302 = 1.20, 95% CI = 1.09 to 1.32), and ratio of DHA to total fatty acids (distal colon OR = 1.29, 95%

303 CI = 1.16 to 1.43). There was also evidence of a negative effect of ratio of omega-6 to omega-3  
304 fatty acids (distal CRC OR = 0.80, 95% CI = 0.74 to 0.88) and a positive effect of ratio of omega-  
305 3 fatty acids to total fatty acids (distal CRC = 1.24, 95% CI = 1.15 to 1.35; seen also for proximal  
306 CRC OR = 1.15, 95% CI = 1.07 to 1.23) (**Supplementary File 1j**). These estimates were also  
307 directionally consistent in MR sensitivity models.

308

309

## 310 **Discussion**

311

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

326

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

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

340  
341 Our analyses featured evaluating the effect of genetic liability to CRC on circulating metabolites  
342 across repeated measures in the ALSPAC cohort. The mean ages at the time of the repeated  
343 measures were 8y, 16y, 18y, and 25y, representing childhood, early adolescence, late  
344 adolescence, and young adulthood respectively, and therefore individuals in this cohort are  
345 unlikely to be taking metabolite-altering medication such as statins, and unlikely to have CRC.  
346 The strongest evidence for an effect of liability to CRC on metabolite levels was seen in late  
347 adolescence. The reason for this remains unclear. It is possible that this represents a true  
348 biological phenomenon if late adolescence is a critical window in CRC development or metabolite  
349 variability, which may be likely given the limited variance in metabolite levels at the later age of  
350 25y (**Supplementary File 1b**). The lack of an effect at the younger ages could be explained by  
351 the fact that the CRC GRS may capture many key life events or experiences which could impact  
352 the metabolome (e.g., initiation of smoking, higher category of BMI reached, educational  
353 attainment level set, etc) but may not have yet happened at younger ages, thus obscuring an  
354 effect of genetic liability to CRC on the metabolome. Our results suggest that puberty could be  
355 important, with an effect seen seemingly particularly at the end of puberty. Repeating our analysis  
356 with sex-stratified data may aid in determining whether this is likely to be the case; sex-stratified  
357 GWAS for metabolites are not currently available to replicate such analyses. An alternative  
358 explanation is selection bias due to loss of follow-up, leading to a change in sample characteristics  
359 over time.

360  
361 Another key finding in the reverse MR analysis was that genetic liability to CRC was associated  
362 with increased levels of total cholesterol, VLDL-cholesterol, LDL-cholesterol, and apolipoprotein  
363 B, though we find little evidence for a causal effect of these traits on risk of CRC in the forward  
364 MR, replicating previous forward MR analyses for total and LDL-cholesterol.<sup>9,72–74</sup> This suggests  
365 that these traits may either be only predictive of (i.e., non-causal for) later CRC development, or  
366 may be influenced by the development of CRC and could have diagnostic or predictive potential.  
367 Given that the participants in the ALSPAC cohort are many decades younger than the average  
368 age of diagnosis for CRC (mean age 25 years in the latest repeated measure analysed in  
369 ALSPAC; whereas the median age at diagnosis of CRC is 64 years),<sup>75</sup> the former seems the most  
370 likely scenario. Previous conventional observational studies have presented conflicting results

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

382  
383 Our analyses stratified by anatomical subsite highlighted fatty acids as being affected by genetic  
384 liability to colon and proximal colon cancer, with the forward MR confirming that fatty acid traits  
385 may be particularly important in the development of these subsites of CRC as well as distal colon  
386 cancer.

387  
388 In our forward MR analyses we were unable to replicate the findings of three previous MR studies  
389 which found evidence for a causal effect of circulating linoleic acid levels on CRC development in  
390 terms of strength of evidence, though the direction of the effect estimate was similar to previous  
391 studies.<sup>89–91</sup> This is surprising as all three previous analyses had a much smaller sample size than  
392 that included in our analysis (the largest had sample size of 24,748 for exposure vs 118,466  
393 presently; and 11,016 cases and 13,732 controls for outcome vs 52,775 cases and 45,940  
394 controls presently). Our analysis using updated genetic instruments to proxy fatty acids may be  
395 more successful in accurately instrumenting heterogenous phenotypes such as metabolite levels  
396 compared with previous analyses. All other findings in our forward MR analysis are consistent  
397 with previous MR studies where they exist.<sup>72–74</sup>

398  
399 *Limitations*

400  
401 The limitations of this study include firstly the relatively small sample size included in the ALSPAC  
402 analysis, which may have implications for power and precision. Secondly, mostly due to the  
403 longitudinal nature of the ASLAPC study, our sample at each time point is composed of slightly  
404 different individuals. This could be influencing our results, and should be taken into account when

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

417

## 418 *Conclusions*

419

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

424

## 425 **Figure legends**

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

434

435 **Figure 2:** Associations of genetic liability to adult colorectal cancer (based on a 72 SNP genetic  
436 risk score) with clinically validated metabolic traits at different early life stages among ALSPAC  
437 offspring (age 8y (N=4,767), 16y (N=2,930), 18y (N=2,613), and 25y (N=2,559)). Estimates  
438 shown are beta coefficients representing the SD difference in metabolic trait per doubling of  
439 genetic liability to colorectal cancer (purple, 8y; turquoise, 16y; red, 18y; black, 25y). Filled point  
440 estimates are those that pass a Benjamini–Hochberg FDR multiple-testing correction (FDR <  
441 0.05).

442

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



445 8y, 16y, 18y, and 25y). Estimates shown are beta coefficients representing the SD difference in  
446 metabolic trait per doubling of genetic liability to colon cancer (purple, 8y; turquoise, 16y; red,  
447 18y; black, 25y). Filled point estimates are those that pass a Benjamini–Hochberg FDR multiple-  
448 testing correction (FDR < 0.05).

449

450 **Figure 2-figure supplement 2:** Associations of genetic liability to proximal colon cancer with  
451 clinically validated metabolic traits at different early life stages among ALSPAC offspring (age  
452 8y, 16y, 18y, and 25y). Estimates shown are beta coefficients representing the SD difference in  
453 metabolic trait per doubling of genetic liability to proximal colon cancer (purple, 8y; turquoise,  
454 16y; red, 18y; black, 25y). Filled point estimates are those that pass a Benjamini–Hochberg  
455 FDR multiple-testing correction (FDR < 0.05).

456

457 **Figure 2-figure supplement 3:** Associations of genetic liability to distal colon cancer with  
458 clinically validated metabolic traits at different early life stages among ALSPAC offspring (age  
459 8y, 16y, 18y, and 25y). Estimates shown are beta coefficients representing the SD difference in  
460 metabolic trait per doubling of genetic liability to distal colon cancer (purple, 8y; turquoise, 16y;  
461 red, 18y; black, 25y). Filled point estimates are those that pass a Benjamini–Hochberg FDR  
462 multiple-testing correction (FDR < 0.05).

463

464 **Figure 2-figure supplement 4:** Associations of genetic liability to rectal cancer with clinically  
465 validated metabolic traits at different early life stages among ALSPAC offspring (age 8y, 16y,  
466 18y, and 25y). Estimates shown are beta coefficients representing the SD difference in  
467 metabolic trait per doubling of genetic liability to rectal cancer (purple, 8y; turquoise, 16y; red,  
468 18y; black, 25y). Filled point estimates are those that pass a Benjamini–Hochberg FDR multiple-  
469 testing correction (FDR < 0.05).

470

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

477

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

484

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

491

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

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

498  
499 **Figure 3-figure supplement 2:** Associations of genetic liability to colorectal cancer (excluding  
500 genetic variants in the FADS gene region) with clinically validated metabolic traits in an  
501 independent sample of adults based on reverse two sample Mendelian randomization analyses.  
502 Estimates shown are beta coefficients representing the SD-unit difference in metabolic trait per  
503 doubling of liability to colorectal cancer. Filled point estimates are those that pass a Benjamini–  
504 Hochberg FDR multiple-testing correction ( $FDR < 0.05$ ).

505  
506 **Figure 3-figure supplement 3:** Associations of genetic liability to colorectal and colon cancer  
507 with clinically validated metabolic traits in an independent sample of adults based on reverse  
508 two sample Mendelian randomization analyses with FADS variants excluded from colorectal  
509 cancer instruments. Estimates shown are beta coefficients representing the SD-unit difference  
510 in metabolic trait per doubling of liability to colorectal cancer by site (colorectal, colon). Filled  
511 point estimates are those that pass a Benjamini–Hochberg FDR multiple-testing correction  
512 ( $FDR < 0.05$ ).

513  
514 **Figure 4:** Associations of clinically validated metabolites with colorectal cancer based on  
515 conventional (forward) two sample Mendelian randomization analyses in individuals from UK  
516 Biobank ( $N=118,466$ , median age 58y) . Estimates shown are beta coefficients representing the  
517 logOR for colorectal cancer per SD metabolite. Filled point estimates are those that pass a  
518 Benjamini–Hochberg FDR multiple-testing correction ( $FDR < 0.05$ ).

519  
520 **Figure 4-figure supplement 1:** Associations of clinically validated metabolites with colorectal  
521 cancer by site (colorectal, colon, distal colon, proximal colon and rectal cancer) based on  
522 conventional (forward) two sample Mendelian randomization analyses. Estimates shown are  
523 ORs for colorectal cancer per SD metabolite. Filled point estimates are those that pass a  
524 Benjamini–Hochberg FDR multiple-testing correction ( $FDR < 0.05$ ).

525  
526 **Supplementary file legends**

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

529  
530 **Supplementary File 1b:** Mean and SD values for raw metabolic traits at different life stages  
531 among ALSPAC offspring

532  
533 **Supplementary File 1c:** Associations of genetic liability to colorectal cancer with metabolic  
534 traits at different early life stages among ALSPAC offspring

535  
536 **Supplementary File 1d:** Associations of genetic liability to colorectal cancer (excluding SNPs in  
537 the FADS gene region) with metabolic traits at different early life stages among ALSPAC  
538 offspring

539  
540 **Supplementary File 1e:** Assessment of instrument strength for MR analyses

541  
542 **Supplementary File 1f:** Associations of genetic liability to colorectal cancer with metabolic traits  
543 based on two-sample MR

544

545 **Supplementary File 1g:** Associations of genetic liability to colorectal cancer with metabolic  
546 traits based on two-sample MR, excluding variants in the FADS gene region  
547  
548 **Supplementary File 1h:** Genetic variants used to instrument circulating metabolites  
549  
550 **Supplementary File 1i:** Assessment of instrument strength for MR analyses  
551  
552 **Supplementary File 1j:** Estimated effects of circulating metabolites on colorectal cancer risk  
553 based on two-sample MR  
554  
555 **Supplementary File 1k:** Estimated effects of circulating metabolites on colorectal cancer risk  
556 based on two-sample MR, excluding variants in the FADS gene region

557 **References**

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

- 603 13. Ward PS, Thompson CB. Metabolic reprogramming: a cancer hallmark even warburg did  
604 not anticipate. *Cancer Cell* [Internet]. Cancer Cell; 2012 Mar 20 [cited 2022 Nov  
605 17];**21**(3):297–308. Available from: <https://pubmed.ncbi.nlm.nih.gov/22439925/>
- 606 14. Rattray NJW, Charkoftaki G, Rattray Z, Hansen JE, Vasiliou V, Johnson CH.  
607 Environmental Influences in the Etiology of Colorectal Cancer: the Premise of  
608 Metabolomics. *Current Pharmacology Reports* 2017 3:3 [Internet]. Springer; 2017 Apr 7  
609 [cited 2022 Nov 22];**3**(3):114–125. Available from:  
610 <https://link.springer.com/article/10.1007/s40495-017-0088-z>
- 611 15. Singla P, Bardoloi A, Parkash AA. Metabolic effects of obesity: A review. *World J*  
612 *Diabetes* [Internet]. Baishideng Publishing Group Inc; 2010 Jul 7 [cited 2022 Nov  
613 17];**1**(3):76. Available from: </pmc/articles/PMC3083889/>
- 614 16. Papandreou C, García-Gavilán J, Camacho-Barcia L, et al. Circulating Metabolites  
615 Associated with Body Fat and Lean Mass in Adults with Overweight/Obesity. *Metabolites*  
616 *2021, Vol 11, Page 317* [Internet]. Multidisciplinary Digital Publishing Institute; 2021  
617 May 13 [cited 2022 Nov 17];**11**(5):317. Available from: [https://www.mdpi.com/2218-](https://www.mdpi.com/2218-1989/11/5/317/htm)  
618 [1989/11/5/317/htm](https://www.mdpi.com/2218-1989/11/5/317/htm)
- 619 17. Ahmad S, Hammar U, Kennedy B, et al. Effect of General Adiposity and Central Body  
620 Fat Distribution on the Circulating Metabolome: A Multicohort Nontargeted  
621 Metabolomics Observational and Mendelian Randomization Study. *Diabetes* [Internet].  
622 American Diabetes Association; 2022 Feb 1 [cited 2022 Nov 17];**71**(2):329–339.  
623 Available from: [https://diabetesjournals.org/diabetes/article/71/2/329/138937/Effect-of-](https://diabetesjournals.org/diabetes/article/71/2/329/138937/Effect-of-General-Adiposity-and-Central-Body-Fat)  
624 [General-Adiposity-and-Central-Body-Fat](https://diabetesjournals.org/diabetes/article/71/2/329/138937/Effect-of-General-Adiposity-and-Central-Body-Fat)
- 625 18. Haycock PC, Carolina Borges M, Burrows K, et al. The association between genetically  
626 elevated polyunsaturated fatty acids and risk of cancer. *www.thelancet.com* [Internet].  
627 2023 [cited 2023 Jul 31];**91**. Available from:  
628 [http://practical.icr.ac.uk/.ajhttps://epi.grants.cancer.gov/interlymph/.akhttps://dceg.cancer.](http://practical.icr.ac.uk/.ajhttps://epi.grants.cancer.gov/interlymph/.akhttps://dceg.cancer.gov/research/cancer-types/pancreas/panscan)  
629 [gov/research/cancer-types/pancreas/panscan](http://practical.icr.ac.uk/.ajhttps://epi.grants.cancer.gov/interlymph/.akhttps://dceg.cancer.gov/research/cancer-types/pancreas/panscan).
- 630 19. Yarmolinsky J, Wade KH, Richmond RC, et al. Causal Inference in Cancer  
631 Epidemiology: What Is the Role of Mendelian Randomization? [cited 2022 Nov 17];  
632 Available from: <http://aacrjournals.org/cebp/article-pdf/27/9/995/2285730/995.pdf>
- 633 20. GD S, S E. ‘Mendelian randomization’: can genetic epidemiology contribute to  
634 understanding environmental determinants of disease? *Int J Epidemiol* [Internet]. Int J  
635 Epidemiol; 2003 [cited 2022 Nov 17];**32**(1). Available from:  
636 <https://pubmed.ncbi.nlm.nih.gov/12689998/>
- 637 21. Holmes M v., Smith GD. Can Mendelian Randomization Shift into Reverse Gear? *Clin*  
638 *Chem* [Internet]. Oxford Academic; 2019 Mar 1 [cited 2022 Nov 17];**65**(3):363–366.  
639 Available from: <https://academic.oup.com/clinchem/article/65/3/363/5608004>
- 640 22. Gao R, Wu C, Zhu Y, et al. Integrated Analysis of Colorectal Cancer Reveals Cross-  
641 Cohort Gut Microbial Signatures and Associated Serum Metabolites. *Gastroenterology*.  
642 W.B. Saunders; 2022 Oct 1;**163**(4):1024-1037.e9.
- 643 23. Rattner JI, Kopciuk KA, Vogel HJ, et al. Early detection of treatment futility in patients  
644 with metastatic colorectal cancer. *Oncotarget* [Internet]. Impact Journals, LLC; 2022  
645 [cited 2022 Nov 22];**13**:61. Available from: </pmc/articles/PMC8746015/>
- 646 24. Leichtle AB, Nuoffer JM, Ceglarek U, et al. Serum amino acid profiles and their  
647 alterations in colorectal cancer. *Metabolomics* [Internet]. Springer; 2012 Aug 16 [cited

- 648 2022 Nov 22];**8**(4):643–653. Available from:  
649 <https://link.springer.com/article/10.1007/s11306-011-0357-5>
- 650 25. Ritchie SA, Ahiahonu PWK, Jayasinghe D, et al. Reduced levels of hydroxylated,  
651 polyunsaturated ultra long-chain fatty acids in the serum of colorectal cancer patients:  
652 Implications for early screening and detection. *BMC Med* [Internet]. BioMed Central;  
653 2010 Feb 15 [cited 2022 Nov 22];**8**(1):1–20. Available from:  
654 <https://bmcmedicine.biomedcentral.com/articles/10.1186/1741-7015-8-13>
- 655 26. Nishiumi S, Kobayashi T, Ikeda A, et al. A Novel Serum Metabolomics-Based Diagnostic  
656 Approach for Colorectal Cancer. *PLoS One* [Internet]. Public Library of Science; 2012 Jul  
657 11 [cited 2022 Nov 22];**7**(7):e40459. Available from:  
658 <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0040459>
- 659 27. Ma Y, Zhang P, Wang F, Liu W, Yang J, Qin H. An integrated proteomics and  
660 metabolomics approach for defining oncofetal biomarkers in the colorectal cancer. *Ann*  
661 *Surg* [Internet]. 2012 Apr [cited 2022 Nov 22];**255**(4):720–730. Available from:  
662 [https://journals.lww.com/annalsofsurgery/Fulltext/2012/04000/An\\_Integrated\\_Proteomics](https://journals.lww.com/annalsofsurgery/Fulltext/2012/04000/An_Integrated_Proteomics_and_Metabolomics_Approach.17.aspx)  
663 [\\_and\\_Metabolomics\\_Approach.17.aspx](https://journals.lww.com/annalsofsurgery/Fulltext/2012/04000/An_Integrated_Proteomics_and_Metabolomics_Approach.17.aspx)
- 664 28. Bertini I, Cacciatore S, Jensen B v., et al. Metabolomic NMR fingerprinting to identify  
665 and predict survival of patients with metastatic colorectal cancer. *Cancer Res* [Internet].  
666 American Association for Cancer Research; 2012 Jan 1 [cited 2022 Nov 22];**72**(1):356–  
667 364. Available from:  
668 [https://aacrjournals.org/cancerres/article/72/1/356/575915/Metabolomic-NMR-](https://aacrjournals.org/cancerres/article/72/1/356/575915/Metabolomic-NMR-Fingerprinting-to-Identify-and)  
669 [Fingerprinting-to-Identify-and](https://aacrjournals.org/cancerres/article/72/1/356/575915/Metabolomic-NMR-Fingerprinting-to-Identify-and)
- 670 29. Zhang A, Sun H, Yan G, Wang P, Han Y, Wang X. Metabolomics in diagnosis and  
671 biomarker discovery of colorectal cancer. *Cancer Lett*. Elsevier; 2014 Apr 1;**345**(1):17–  
672 20.
- 673 30. Farshidfar F, Weljie AM, Kopciuk KA, et al. A validated metabolomic signature for  
674 colorectal cancer: exploration of the clinical value of metabolomics. *British Journal of*  
675 *Cancer* 2016 *115*:7 [Internet]. Nature Publishing Group; 2016 Aug 25 [cited 2022 Nov  
676 22];**115**(7):848–857. Available from: <https://www.nature.com/articles/bjc2016243>
- 677 31. Farshidfar F, Weljie AM, Kopciuk K, et al. Serum metabolomic profile as a means to  
678 distinguish stage of colorectal cancer. *Genome Med* [Internet]. BioMed Central; 2012 May  
679 14 [cited 2022 Nov 22];**4**(5):1–13. Available from:  
680 <https://genomemedicine.biomedcentral.com/articles/10.1186/gm341>
- 681 32. Boyd A, Golding J, Macleod J, et al. Cohort Profile: The ‘Children of the 90s’—the index  
682 offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol*. 2013  
683 Feb;**42**(1):111–127.
- 684 33. Fraser A, Macdonald-Wallis C, Tilling K, et al. Cohort Profile: The Avon Longitudinal  
685 Study of Parents and Children: ALSPAC mothers cohort. *Int J Epidemiol*. 2013 Feb  
686 1;**42**(1):97–110.
- 687 34. Northstone K, Lewcock M, Groom A, et al. The Avon Longitudinal Study of Parents and  
688 Children (ALSPAC): an update on the enrolled sample of index children in 2019.  
689 *Wellcome Open Res* [Internet]. The Wellcome Trust; 2019 [cited 2023 Feb 28];**4**.  
690 Available from: [/pmc/articles/PMC6464058/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6464058/)
- 691 35. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research Electronic  
692 Data Capture (REDCap) - A metadata-driven methodology and workflow process for  
693 providing translational research informatics support. *J Biomed Inform* [Internet]. NIH

- 694 Public Access; 2009 Apr [cited 2023 Feb 28];**42**(2):377. Available from:  
695 /pmc/articles/PMC2700030/  
696 36. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: An Open Access Resource for  
697 Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age.  
698 *PLoS Med* [Internet]. Public Library of Science; 2015 Mar 1 [cited 2022 Nov  
699 22];**12**(3):e1001779. Available from:  
700 <https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001779>  
701 37. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping  
702 and genomic data. *Nature* 2018 562:7726 [Internet]. Nature Publishing Group; 2018 Oct  
703 10 [cited 2022 Nov 22];**562**(7726):203–209. Available from:  
704 <https://www.nature.com/articles/s41586-018-0579-z>  
705 38. Huyghe JR, Bien SA, Harrison TA, et al. Discovery of common and rare genetic risk  
706 variants for colorectal cancer. *Nat Genet.* 2019;**51**(1):76–87.  
707 39. Huyghe JR, Harrison TA, Bien SA, et al. Genetic architectures of proximal and distal  
708 colorectal cancer are partly distinct. *Gut* [Internet]. Gut; 2021 Jul 1 [cited 2023 Jan  
709 27];**70**(7):1325–1334. Available from: <https://pubmed.ncbi.nlm.nih.gov/33632709/>  
710 40. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample  
711 Mendelian randomization. *Genet Epidemiol.* 2016;**40**(7):597–608.  
712 41. Law PJ, Timofeeva M, Fernandez-Rozadilla C. Association analyses identify 31 new risk  
713 loci for colorectal cancer susceptibility.  
714 42. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal  
715 inference across the human phenome. *Elife.* 2018 May 30;**7**.  
716 43. Huyghe JR, Harrison TA, Bien SA, et al. Genetic architectures of proximal and distal  
717 colorectal cancer are partly distinct. *Gut.* Gut; 2021 Jul 1;**70**(7):1325–1334.  
718 44. Lu T, Sun L, Wang Z, Zhang Y, He Z, Xu C. Fatty acid synthase enhances colorectal  
719 cancer cell proliferation and metastasis via regulating AMPK/mTOR pathway. *Onco*  
720 *Targets Ther* [Internet]. Dove Press; 2019 [cited 2022 Nov 22];**12**:3339. Available from:  
721 /pmc/articles/PMC6504633/  
722 45. Zaytseva YY, Rychahou PG, Le AT, et al. Preclinical evaluation of novel fatty acid  
723 synthase inhibitors in primary colorectal cancer cells and a patient-derived xenograft  
724 model of colorectal cancer. *Oncotarget* [Internet]. Impact Journals, LLC; 2018 May 5  
725 [cited 2022 Nov 22];**9**(37):24787. Available from: /pmc/articles/PMC5973868/  
726 46. TVB 2640 for Resectable Colon Cancer Other Resectable Cancers; a Window Trial. - Full  
727 Text View - ClinicalTrials.gov [Internet]. [cited 2022 Nov 22]. Available from:  
728 <https://clinicaltrials.gov/ct2/show/NCT02980029>  
729 47. Fhu CW, Ali A. Fatty Acid Synthase: An Emerging Target in Cancer. *Molecules*  
730 [Internet]. Multidisciplinary Digital Publishing Institute (MDPI); 2020 Sep 1 [cited 2022  
731 Nov 22];**25**(17). Available from: /pmc/articles/PMC7504791/  
732 48. Chen HY, Cairns BJ, Small AM, et al. Association of FADS1/2 Locus Variants and  
733 Polyunsaturated Fatty Acids With Aortic Stenosis. *JAMA Cardiol* [Internet]. American  
734 Medical Association; 2020 Jun 1 [cited 2022 Nov 22];**5**(6):694–702. Available from:  
735 <https://jamanetwork.com/journals/jamacardiology/fullarticle/2762845>  
736 49. Kathiresan S, Willer CJ, Peloso GM, et al. Common variants at 30 loci contribute to  
737 polygenic dyslipidemia. *Nature Genetics* 2008 41:1 [Internet]. Nature Publishing Group;  
738 2008 Dec 7 [cited 2022 Nov 22];**41**(1):56–65. Available from:  
739 <https://www.nature.com/articles/ng.291>

- 740 50. Tanaka T, Shen J, Abecasis GR, et al. Genome-Wide Association Study of Plasma  
741 Polyunsaturated Fatty Acids in the InCHIANTI Study. *PLoS Genet* [Internet]. Public  
742 Library of Science; 2009 Jan [cited 2022 Nov 22];**5**(1):e1000338. Available from:  
743 <https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1000338>
- 744 51. Würtz P, Kangas AJ, Soininen P, Lawlor DA, Davey Smith G, Ala-Korpela M.  
745 Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Large-Scale  
746 Epidemiology: A Primer on -Omic Technologies. *Am J Epidemiol*. Oxford Academic;  
747 2017 Nov 1;**186**(9):1084–1096.
- 748 52. Julkunen H, Cichońska A, Slagboom PE, Würtz P. Metabolic biomarker profiling for  
749 identification of susceptibility to severe pneumonia and COVID-19 in the general  
750 population. *Elife*. eLife Sciences Publications Ltd; 2021 May 1;**10**.
- 751 53. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping  
752 and genomic data. *Nature* 2018 562:7726. Nature Publishing Group; 2018 Oct  
753 10;**562**(7726):203–209.
- 754 54. Holmes M V., Smith GD. Can Mendelian randomization shift into reverse gear? *Clin*  
755 *Chem* American Association for Clinical Chemistry Inc.; 2019. p. 363–366.
- 756 55. Burgess S, Labrecque JA. Mendelian randomization with a binary exposure variable:  
757 interpretation and presentation of causal estimates. *Eur J Epidemiol* [Internet]. Eur J  
758 Epidemiol; 2018 Oct 1 [cited 2023 Jan 9];**33**(10):947–952. Available from:  
759 <https://pubmed.ncbi.nlm.nih.gov/30039250/>
- 760 56. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and  
761 Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B*  
762 *(Methodological)* [Internet]. John Wiley & Sons, Ltd; 1995 Jan 1 [cited 2022 Nov  
763 22];**57**(1):289–300. Available from:  
764 <https://onlinelibrary.wiley.com/doi/full/10.1111/j.2517-6161.1995.tb02031.x>
- 765 57. Clayton GL, Borges MC, Lawlor DA. From menarche to menopause: the impact of  
766 reproductive factors on the metabolic profile of over 65,000 women. 2022 Apr 18 [cited  
767 2023 Jan 27]; Available from: <https://europepmc.org/article/PPR/PPR484113>
- 768 58. Borges MC, Haycock PC, Zheng J, et al. Role of circulating polyunsaturated fatty acids on  
769 cardiovascular diseases risk: analysis using Mendelian randomization and fatty acid  
770 genetic association data from over 114,000 UK Biobank participants. *BMC Med*  
771 [Internet]. BMC Med; 2022 Dec 1 [cited 2023 Jan 27];**20**(1). Available from:  
772 <https://pubmed.ncbi.nlm.nih.gov/35692035/>
- 773 59. MRC IEU UK Biobank GWAS pipeline version 2 - Datasets - data.bris [Internet]. [cited  
774 2022 Oct 14]. Available from:  
775 <https://data.bris.ac.uk/data/dataset/pnoat8cxo0u52p6ynfaekeigi>
- 776 60. Elsworth B, Lyon M, Alexander T, et al. The MRC IEU OpenGWAS data infrastructure.
- 777 61. Gibran Hemani Kaitlin H Wade, Charles Laurin, Benjamin Elsworth, Stephen Burgess,  
778 Jack Bowden, Ryan Langdon, Vanessa Tan, James Yarmolinsky, Hashem A. Shihab,  
779 Nicholas Timpson, David M Evans, Caroline Relton, Richard M Martin, George Davey  
780 Smith, Tom R G JZ. MR-Base: a platform for systematic causal inference across the  
781 phenome using billions of genetic associations. Biorxiv; 2016.
- 782 62. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian  
783 Randomization with Some Invalid Instruments Using a Weighted Median Estimator.  
784 *Genet Epidemiol*. 2016;**40**(4):304–314.



- 785 63. Hartwig FP, Smith GD, Bowden J. Robust inference in summary data Mendelian  
786 randomization via the zero modal pleiotropy assumption. *Int J Epidemiol*. Oxford  
787 Academic; 2017 Dec 1;**46**(6):1985–1998.
- 788 64. Hartwig FP, Smith GD, Bowden J. Robust inference in summary data Mendelian  
789 randomization via the zero modal pleiotropy assumption. *Int J Epidemiol* [Internet]. Int J  
790 Epidemiol; 2017 Dec 1 [cited 2023 Jan 9];**46**(6):1985–1998. Available from:  
791 <https://pubmed.ncbi.nlm.nih.gov/29040600/>
- 792 65. Hartwig FP, Smith GD, Bowden J. Robust inference in summary data Mendelian  
793 randomization via the zero modal pleiotropy assumption. [cited 2023 Aug 1]; Available  
794 from: <https://academic.oup.com/ije/article/46/6/1985/3957932>
- 795 66. Burgess S, Labrecque JA. Mendelian randomization with a binary exposure variable:  
796 interpretation and presentation of causal estimates.
- 797 67. R Core Team. R: A language and environment for statistical computing. R Foundation for  
798 Statistical Computing, Vienna, Austria.; 2020.
- 799 68. Forestplots of Measures of Effects and Their Confidence Intervals • ggforestplot  
800 [Internet]. [cited 2022 Oct 14]. Available from:  
801 <https://nightingalehealth.github.io/ggforestplot/index.html>
- 802 69. Boyd A, Golding J, Macleod J, et al. Cohort Profile: The ‘Children of the 90s’—the index  
803 offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol*  
804 [Internet]. Oxford Academic; 2013 Feb 1 [cited 2023 Jan 27];**42**(1):111–127. Available  
805 from: <https://academic.oup.com/ije/article/42/1/111/694290>
- 806 70. Borges MC, Haycock PC, Zheng J, et al. Role of circulating polyunsaturated fatty acids on  
807 cardiovascular diseases risk: analysis using Mendelian randomization and fatty acid  
808 genetic association data from over 114,000 UK Biobank participants. *BMC Med*  
809 [Internet]. BioMed Central Ltd; 2022 Dec 1 [cited 2023 Jan 30];**20**(1):1–14. Available  
810 from: <https://bmcmmedicine.biomedcentral.com/articles/10.1186/s12916-022-02399-w>
- 811 71. Borges MC, Haycock P, Zheng J, et al. The impact of fatty acids biosynthesis on the risk  
812 of cardiovascular diseases in Europeans and East Asians: a Mendelian randomization  
813 study. *Hum Mol Genet* [Internet]. Oxford Academic; 2022 Nov 28 [cited 2023 Feb  
814 28];**31**(23):4034–4054. Available from:  
815 <https://academic.oup.com/hmg/article/31/23/4034/6633188>
- 816 72. Rodriguez-Broadbent H, Law PJ, Sud A, et al. Mendelian randomisation implicates  
817 hyperlipidaemia as a risk factor for colorectal cancer. *Int J Cancer* [Internet]. Int J Cancer;  
818 2017 Jun 15 [cited 2022 Nov 28];**140**(12):2701–2708. Available from:  
819 <https://pubmed.ncbi.nlm.nih.gov/28340513/>
- 820 73. Cornish AJ, Law PJ, Timofeeva M, et al. Modifiable pathways for colorectal cancer: a  
821 mendelian randomisation analysis. *Lancet Gastroenterol Hepatol* [Internet]. Lancet  
822 Gastroenterol Hepatol; 2020 Jan 1 [cited 2022 Nov 28];**5**(1):55–62. Available from:  
823 <https://pubmed.ncbi.nlm.nih.gov/31668584/>
- 824 74. Luo X, Tu Z, Chen H, Ding J. Blood lipids and risk of colon or rectal cancer: a Mendelian  
825 randomization study. *J Cancer Res Clin Oncol* [Internet]. J Cancer Res Clin Oncol; 2021  
826 Dec 1 [cited 2022 Nov 28];**147**(12):3591–3599. Available from:  
827 <https://pubmed.ncbi.nlm.nih.gov/34490583/>
- 828 75. Granados-Romero JJ, Valderrama-Treviño A, Hazzel E, Flores C, Herrera M. Colorectal  
829 cancer: a review. *Article in International Journal of Research in Medical Sciences*

- 830 [Internet]. 2017 [cited 2022 Nov 17]; Available from: [http://dx.doi.org/10.18203/2320-](http://dx.doi.org/10.18203/2320-6012.ijrms20174914)  
831 6012.ijrms20174914
- 832 76. Park SK, Joo JS, Kim DH, Kim YE, Kang D, Yoo KY. Association of serum lipids and  
833 glucose with the risk of colorectal adenomatous polyp in men: a case-control study in  
834 Korea. *J Korean Med Sci* [Internet]. Korean Academy of Medical Science; 2000 Dec 1  
835 [cited 2022 Nov 17];**15**(6):690–695. Available from:  
836 <https://doi.org/10.3346/jkms.2000.15.6.690>
- 837 77. Yang Z, Tang H, Lu S, Sun X, Rao B. Relationship between serum lipid level and  
838 colorectal cancer: a systemic review and meta-analysis. *BMJ Open* [Internet]. British  
839 Medical Journal Publishing Group; 2022 Jun 1 [cited 2022 Nov 17];**12**(6):e052373.  
840 Available from: <https://bmjopen.bmj.com/content/12/6/e052373>
- 841 78. Tian Y, Wang K, Li J, et al. The association between serum lipids and colorectal  
842 neoplasm: a systemic review and meta-analysis. *Public Health Nutr* [Internet]. Cambridge  
843 University Press; 2015 Mar 16 [cited 2022 Nov 17];**18**(18):3355–3370. Available from:  
844 [https://www.cambridge.org/core/journals/public-health-nutrition/article/association-](https://www.cambridge.org/core/journals/public-health-nutrition/article/association-between-serum-lipids-and-colorectal-neoplasm-a-systemic-review-and-metaanalysis/4401DD8E09A7A257C95F1F3189E3B845)  
845 [between-serum-lipids-and-colorectal-neoplasm-a-systemic-review-and-](https://www.cambridge.org/core/journals/public-health-nutrition/article/association-between-serum-lipids-and-colorectal-neoplasm-a-systemic-review-and-metaanalysis/4401DD8E09A7A257C95F1F3189E3B845)  
846 [metaanalysis/4401DD8E09A7A257C95F1F3189E3B845](https://www.cambridge.org/core/journals/public-health-nutrition/article/association-between-serum-lipids-and-colorectal-neoplasm-a-systemic-review-and-metaanalysis/4401DD8E09A7A257C95F1F3189E3B845)
- 847 79. Yao X, Tian Z. Dyslipidemia and colorectal cancer risk: a meta-analysis of prospective  
848 studies. *Cancer Causes and Control* [Internet]. Kluwer Academic Publishers; 2015 Feb 1  
849 [cited 2022 Nov 17];**26**(2):257–268. Available from:  
850 <https://link.springer.com/article/10.1007/s10552-014-0507-y>
- 851 80. Mayengbam SS, Singh A, Pillai AD, Bhat MK. Influence of cholesterol on cancer  
852 progression and therapy. *Transl Oncol*. Elsevier; 2021 Jun 1;**14**(6):101043.
- 853 81. Mannes GA, Maier A, Thieme C, Wiebecke B, Paumgartner G. Relation between the  
854 Frequency of Colorectal Adenoma and the Serum Cholesterol Level.  
855 <http://dx.doi.org/101056/NEJM198612253152602> [Internet]. Massachusetts Medical  
856 Society ; 2009 Nov 16 [cited 2022 Nov 17];**315**(26):1634–1638. Available from:  
857 <https://www.nejm.org/doi/full/10.1056/NEJM198612253152602>
- 858 82. Mamtani R, Lewis JD, Scott FI, et al. Disentangling the Association between Statins,  
859 Cholesterol, and Colorectal Cancer: A Nested Case-Control Study. *PLoS Med* [Internet].  
860 Public Library of Science; 2016 Apr 1 [cited 2022 Nov 17];**13**(4):e1002007. Available  
861 from: <https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1002007>
- 862 83. Ocvirk S, Wilson AS, Appolonia CN, Thomas TK, O’Keefe SJD. Fiber, Fat, and  
863 Colorectal Cancer: New Insight into Modifiable Dietary Risk Factors. *Curr Gastroenterol*  
864 *Rep* [Internet]. Springer; 2019 Nov 1 [cited 2022 Nov 17];**21**(11):1–7. Available from:  
865 <https://link.springer.com/article/10.1007/s11894-019-0725-2>
- 866 84. Tang FY, Pai MH, Chiang EPI. Consumption of high-fat diet induces tumor progression  
867 and epithelial–mesenchymal transition of colorectal cancer in a mouse xenograft model. *J*  
868 *Nutr Biochem*. Elsevier; 2012 Oct 1;**23**(10):1302–1313.
- 869 85. Bell AE, Culp PA. Reduction in Saturated Fat Intake for Cardiovascular Disease. *Am Fam*  
870 *Physician* [Internet]. NLM (Medline); 2022 Jan 1 [cited 2022 Nov 17];**105**(1). Available  
871 from: <https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD011737.pub2/full>
- 872 86. Chiu S, Williams PT, Krauss RM. Effects of a very high saturated fat diet on LDL  
873 particles in adults with atherogenic dyslipidemia: A randomized controlled trial. *PLoS*  
874 *One* [Internet]. Public Library of Science; 2017 Feb 1 [cited 2022 Nov

- 875 17];**12**(2):e0170664. Available from:  
876 <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0170664>  
877 87. Han Q, Yeung SC, Ip MSM, Mak JCW. Dysregulation of cardiac lipid parameters in high-  
878 fat high-cholesterol diet-induced rat model. [cited 2022 Nov 17]; Available from:  
879 <https://doi.org/10.1186/s12944-018-0905-3>  
880 88. Vincent MJ, Allen B, Palacios OM, Haber LT, Maki KC. Meta-regression analysis of the  
881 effects of dietary cholesterol intake on LDL and HDL cholesterol. *Am J Clin Nutr*  
882 [Internet]. Oxford Academic; 2019 Jan 1 [cited 2022 Nov 17];**109**(1):7–16. Available  
883 from: <https://academic.oup.com/ajcn/article/109/1/7/5266898>  
884 89. May-Wilson S, Sud A, Law PJ, et al. Pro-inflammatory fatty acid profile and colorectal  
885 cancer risk: A Mendelian randomisation analysis. *Eur J Cancer* [Internet]. Eur J Cancer;  
886 2017 Oct 1 [cited 2022 Nov 28];**84**:228–238. Available from:  
887 <https://pubmed.ncbi.nlm.nih.gov/28829991/>  
888 90. Khankari NK, Banbury BL, Borges MC, et al. Mendelian Randomization of Circulating  
889 Polyunsaturated Fatty Acids and Colorectal Cancer Risk. *Cancer Epidemiol Biomarkers*  
890 *Prev* [Internet]. Cancer Epidemiol Biomarkers Prev; 2020 [cited 2022 Nov 23];**29**(4):860–  
891 870. Available from: <https://pubmed.ncbi.nlm.nih.gov/32051193/>  
892 91. Liyanage UE, Ong JS, An J, Gharahkhani P, Law MH, MacGregor S. Mendelian  
893 Randomization Study for Genetically Predicted Polyunsaturated Fatty Acids Levels on  
894 Overall Cancer Risk and Mortality. *Cancer Epidemiol Biomarkers Prev* [Internet]. Cancer  
895 Epidemiol Biomarkers Prev; 2019 Jun 1 [cited 2022 Nov 23];**28**(6):1015–1023. Available  
896 from: <https://pubmed.ncbi.nlm.nih.gov/30948449/>  
897

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

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

906  
907

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

915 available in the manuscript and supporting material. R scripts used in this study have been made  
916 publicly available on GitHub at: [https://github.com/cb12104/adiposity\\_metabolites\\_crc](https://github.com/cb12104/adiposity_metabolites_crc).

917

## 918 **Funding:**

### 919 **Author funding**

920 JAB is supported by the Elizabeth Blackwell Institute for Health Research, University of Bristol  
921 and the Wellcome Trust Institutional Strategic Support Fund (204813/Z/16/Z) and works in a Unit  
922 funded by the Medical Research Council (MC\_UU\_00011/1) and the University of Bristol. EEV,  
923 DNL and CB are supported by Diabetes UK (17/0005587). NJT is a Wellcome Trust Investigator  
924 (202802/Z/16/Z), is the PI of the Avon Longitudinal Study of Parents and Children (MRC & WT  
925 102215/2/13/2), is supported by the University of Bristol NIHR Biomedical Research Centre (BRC-  
926 1215-20011), the MRC Integrative Epidemiology Unit (MC\_UU\_00011/1) and works within the  
927 CRUK Integrative Cancer Epidemiology Programme (C18281/A19169). EH is supported by a  
928 Cancer Research UK Population Research Committee Studentship (C18281/A30905), is  
929 supported by the CRUK Integrative Cancer Epidemiology Programme (C18281/A29019) and is  
930 part of the Medical Research Council Integrative Epidemiology Unit at the University of Bristol  
931 which is supported by the Medical Research Council (MC\_UU\_00011/4) and the University of  
932 Bristol. AC acknowledges funding from grant MR/N013794/1 for the GW4 BIOMED MRC DTP,  
933 awarded to the Universities of Bath, Bristol, Cardiff and Exeter from the Medical Research Council  
934 (MRC)/UKRI. MCB was supported by the UK Medical Research Council (MRC) Skills  
935 Development Fellowship (MR/P014054/1), University of Bristol Vice-Chancellor's Fellowship and  
936 MRC Integrative Epidemiology Unit (MC\_UU\_00011/6). The funders had no role in study design,  
937 data collection and analysis, decision to publish, or preparation of the manuscript. This work used  
938 the computational facilities of the Advanced Computing Research Centre, University of Bristol -  
939 <http://www.bristol.ac.uk/acrc/>.

940

### 941 **Study funding**

942 ALSPAC: The UK Medical Research Council and Wellcome (Grant ref: 217065/Z/19/Z) and the  
943 University of Bristol provide core support for ALSPAC. This publication is the work of the authors  
944 who will serve as guarantors for the contents of this paper. A comprehensive list of grants funding  
945 is available on the ALSPAC website ([http://www.bristol.ac.uk/alspac/external/documents/grant-  
946 acknowledgements.pdf](http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf)); This research was specifically funded by The UK Medical Research  
947 Council (Grant ref: MC\_UU\_12013/1). GWAS data was generated by Sample Logistics and

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

950

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

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

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

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

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

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

982 The Colon Cancer Family Registry (CCFR, [www.coloncfr.org](http://www.coloncfr.org)) is supported in part by funding  
983 from the National Cancer Institute (NCI), National Institutes of Health (NIH) (award U01  
984 CA167551). Support for case ascertainment was provided in part from the Surveillance,  
985 Epidemiology, and End Results (SEER) Program and the following U.S. state cancer registries:  
986 AZ, CO, MN, NC, NH; and by the Victoria Cancer Registry (Australia) and Ontario Cancer  
987 Registry (Canada). The CCFR Set-1 (Illumina 1M/1M-Duo) and Set-2 (Illumina Omni1-Quad)  
988 scans were supported by NIH awards U01 CA122839 and R01 CA143237 (to GC). The CCFR  
989 Set-3 (Affymetrix Axiom CORECT Set array) was supported by NIH award U19 CA148107 and  
990 R01 CA81488 (to SBG). The CCFR Set-4 (Illumina OncoArray 600K SNP array) was supported  
991 by NIH award U19 CA148107 (to SBG) and by the Center for Inherited Disease Research  
992 (CIDR), which is funded by the NIH to the Johns Hopkins University, contract number  
993 HHSN268201200008I. Additional funding for the OFCCR/ARCTIC was through award GL201-  
994 043 from the Ontario Research Fund (to BWZ), award 112746 from the Canadian Institutes of  
995 Health Research (to TJH), through a Cancer Risk Evaluation (CaRE) Program grant from the  
996 Canadian Cancer Society (to SG), and through generous support from the Ontario Ministry of  
997 Research and Innovation. The SFCCR Illumina HumanCytoSNP array was supported in part  
998 through NCI/NIH awards U01/U24 CA074794 and R01 CA076366 (to PAN). The content of this  
999 manuscript does not necessarily reflect the views or policies of the NCI, NIH or any of the  
1000 collaborating centers in the Colon Cancer Family Registry (CCFR), nor does mention of trade  
1001 names, commercial products, or organizations imply endorsement by the US Government, any  
1002 cancer registry, or the CCFR.

1003 COLON: The COLON study is sponsored by Wereld Kanker Onderzoek Fonds, including funds  
1004 from grant 2014/1179 as part of the World Cancer Research Fund International Regular Grant  
1005 Programme, by Alpe d'Huzes and the Dutch Cancer Society (UM 2012–5653, UW 2013-5927,  
1006 UW2015-7946), and by TRANSCAN (JTC2012-MetaboCCC, JTC2013-FOCUS). The Nqplus  
1007 study is sponsored by a ZonMW investment grant (98-10030); by PREVIEW, the project  
1008 PREvention of diabetes through lifestyle intervention and population studies in Europe and  
1009 around the World (PREVIEW) project which received funding from the European Union Seventh  
1010 Framework Programme (FP7/2007–2013) under grant no. 312057; by funds from TI Food and

1011 Nutrition (cardiovascular health theme), a public–private partnership on precompetitive research  
1012 in food and nutrition; and by FOODBALL, the Food Biomarker Alliance, a project from JPI  
1013 Healthy Diet for a Healthy Life.

1014 COLO2&3: National Institutes of Health (R01 CA060987)

1015 Colorectal Cancer Transdisciplinary (CORECT) Study: The CORECT Study was supported by  
1016 the National Cancer Institute, National Institutes of Health (NCI/NIH), U.S. Department of Health  
1017 and Human Services (grant numbers U19 CA148107, R01 CA081488, P30 CA014089, R01  
1018 CA197350; P01 CA196569; R01 CA201407; R01 CA242218), National Institutes of  
1019 Environmental Health Sciences, National Institutes of Health (grant number T32 ES013678) and  
1020 a generous gift from Daniel and Maryann Fong.

1021 CORSA: The CORSA study was funded by Austrian Research Funding Agency (FFG) BRIDGE  
1022 (grant 829675, to Andrea Gsur), the “Herzfelder’sche Familienstiftung” (grant to Andrea Gsur)  
1023 and was supported by COST Action BM1206.

1024 CPS-II: The American Cancer Society funds the creation, maintenance, and updating of the  
1025 Cancer Prevention Study-II (CPS-II) cohort. The study protocol was approved by the institutional  
1026 review boards of Emory University, and those of participating registries as required.

1027 CRCGEN: Colorectal Cancer Genetics & Genomics, Spanish study was supported by Instituto  
1028 de Salud Carlos III, co-funded by FEDER funds –a way to build Europe– (grants PI14-613 and  
1029 PI09-1286), Agency for Management of University and Research Grants (AGAUR) of the  
1030 Catalan Government (grant 2017SGR723), Junta de Castilla y León (grant LE22A10-2), the  
1031 Spanish Association Against Cancer (AECC) Scientific Foundation grant GCTRA18022MORE  
1032 and the Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP),  
1033 action Genrisk. Sample collection of this work was supported by the Xarxa de Bancs de Tumors  
1034 de Catalunya sponsored by Pla Director d’Oncologia de Catalunya (XBTC), Plataforma  
1035 Biobancos PT13/0010/0013 and ICOBIOBANC, sponsored by the Catalan Institute of Oncology.  
1036 We thank CERCA Programme, Generalitat de Catalunya for institutional support.

1037 Czech Republic CCS: This work was supported by the Grant Agency of the Czech Republic (21-  
1038 04607X, 20-03997S), by the Grant Agency of the Ministry of Health of the Czech Republic  
1039 (grants AZV NU21-07-00247 and AZV NU21-03-00506), and Charles University Research Fund  
1040 (Cooperation 43-Surgical disciplines)

1041 DACHS: This work was supported by the German Research Council (BR 1704/6-1, BR 1704/6-  
1042 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  
1043 1704/17-1), the Interdisciplinary Research Program of the National Center for Tumor Diseases  
1044 (NCT), Germany, and the German Federal Ministry of Education and Research (01KH0404,  
1045 01ER0814, 01ER0815, 01ER1505A and 01ER1505B).

1046 DAL5: National Institutes of Health (R01 CA048998 to M. L. Slattery).

1047 EDRN: This work is funded and supported by the NCI, EDRN Grant (U01-CA152753).

1048 EPIC: The coordination of EPIC is financially supported by International Agency for Research  
1049 on Cancer (IARC) and also by the Department of Epidemiology and Biostatistics, School of  
1050 Public Health, Imperial College London which has additional infrastructure support provided by  
1051 the NIHR Imperial Biomedical Research Centre (BRC). The national cohorts are supported by:  
1052 Danish Cancer Society (Denmark); Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle  
1053 Générale de l'Education Nationale, Institut National de la Santé et de la Recherche Médicale  
1054 (INSERM) (France); German Cancer Aid, German Cancer Research Center (DKFZ), German  
1055 Institute of Human Nutrition Potsdam- Rehbruecke (DIfE), Federal Ministry of Education and  
1056 Research (BMBF) (Germany); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy,  
1057 Compagnia di SanPaolo and National Research Council (Italy); Dutch Ministry of Public Health,  
1058 Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch  
1059 Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund  
1060 (WCRF), Statistics Netherlands (The Netherlands); Health Research Fund (FIS) - Instituto de  
1061 Salud Carlos III (ISCIII), Regional Governments of Andalucía, Asturias, Basque Country, Murcia  
1062 and Navarra, and the Catalan Institute of Oncology - ICO (Spain); Swedish Cancer Society,  
1063 Swedish Research Council and and Region Skåne and Region Västerbotten (Sweden); Cancer  
1064 Research UK (14136 to EPIC-Norfolk; C8221/A29017 to EPIC-Oxford), Medical Research  
1065 Council (1000143 to EPIC-Norfolk; MR/M012190/1 to EPIC-Oxford). (United Kingdom).

1066 EPICOLON: This work was supported by grants from Fondo de Investigación Sanitaria/FEDER  
1067 (PI08/0024, PI08/1276, PS09/02368, PI11/00219, PI11/00681, PI14/00173, PI14/00230,  
1068 PI17/00509, 17/00878, PI20/00113, PI20/00226, Acción Transversal de Cáncer), Xunta de  
1069 Galicia (PGIDIT07PXIB9101209PR), Ministerio de Economía y Competitividad (SAF07-64873,  
1070 SAF 2010-19273, SAF2014-54453R), Fundación Científica de la Asociación Española contra el  
1071 Cáncer (GCB13131592CAST, PRYGN211085CAST), Beca Grupo de Trabajo "Oncología" AEG  
1072 (Asociación Española de Gastroenterología), Fundación Privada Olga Torres, FP7 CHIBCHA



1073 Consortium, Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR, Generalitat de  
1074 Catalunya, 2014SGR135, 2014SGR255, 2017SGR21, 2017SGR653, 2021SGR00716,  
1075 2021SGR01185), Catalan Tumour Bank Network (Pla Director d'Oncologia, Generalitat de  
1076 Catalunya), PERIS (SLT002/16/00398, Generalitat de Catalunya), Marató TV3 (202008-10),  
1077 CERCA Programme (Generalitat de Catalunya) and COST Actions BM1206 and CA17118.  
1078 CIBERehd is funded by the Instituto de Salud Carlos III. ESTHER/VERDI. This work was  
1079 supported by grants from the Baden-Württemberg Ministry of Science, Research and Arts and  
1080 the German Cancer Aid.

1081 Harvard cohorts: HPFS is supported by the National Institutes of Health (P01 CA055075, UM1  
1082 CA167552, U01 CA167552, R01 CA137178, R01 CA151993, and R35 CA197735), NHS by the  
1083 National Institutes of Health (P01 CA087969, UM1 CA186107, R01 CA137178, R01 CA151993,  
1084 and R35 CA197735), and PHS by the National Institutes of Health (R01 CA042182).

1085 Hawaii Adenoma Study: NCI grants R01 CA072520.

1086 HCES-CRC: the Hwasun Cancer Epidemiology Study–Colon and Rectum Cancer (HCES-CRC;  
1087 grants from Chonnam National University Hwasun Hospital, HCRI15011-1).

1088 Kentucky: This work was supported by the following grant support: Clinical Investigator Award  
1089 from Damon Runyon Cancer Research Foundation (CI-8); NCI R01CA136726.

1090 LCCS: The Leeds Colorectal Cancer Study was funded by the Food Standards Agency and  
1091 Cancer Research UK Programme Award (C588/A19167).

1092 MCCS cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS  
1093 was further supported by Australian NHMRC grants 509348, 209057, 251553 and 504711 and  
1094 by infrastructure provided by Cancer Council Victoria. Cases and their vital status were  
1095 ascertained through the Victorian Cancer Registry (VCR) and the Australian Institute of Health  
1096 and Welfare (AIHW), including the National Death Index and the Australian Cancer Database.  
1097 BMLynch was supported by MCRF18005 from the Victorian Cancer Agency.

1098 MEC: National Institutes of Health (R37 CA054281, P01 CA033619, and R01 CA063464).

1099 MECC: This work was supported by the National Institutes of Health, U.S. Department of Health  
1100 and Human Services (R01 CA081488, R01 CA197350, U19 CA148107, R01 CA242218, and a  
1101 generous gift from Daniel and Maryann Fong.

1102 MSKCC: The work at Sloan Kettering in New York was supported by the Robert and Kate  
1103 Niehaus Center for Inherited Cancer Genomics and the Romeo Milio Foundation. Moffitt: This  
1104 work was supported by funding from the National Institutes of Health (grant numbers R01  
1105 CA189184, P30 CA076292), Florida Department of Health Bankhead-Coley Grant 09BN-13,  
1106 and the University of South Florida Oehler Foundation. Moffitt contributions were supported in  
1107 part by the Total Cancer Care Initiative, Collaborative Data Services Core, and Tissue Core at  
1108 the H. Lee Moffitt Cancer Center & Research Institute, a National Cancer Institute-designated  
1109 Comprehensive Cancer Center (grant number P30 CA076292).

1110 NCCCS I & II: We acknowledge funding support for this project from the National Institutes of  
1111 Health, R01 CA066635 and P30 DK034987.

1112 NFCCR: This work was supported by an Interdisciplinary Health Research Team award from  
1113 the Canadian Institutes of Health Research (CRT 43821); the National Institutes of Health, U.S.  
1114 Department of Health and Human Services (U01 CA074783); and National Cancer Institute of  
1115 Canada grants (18223 and 18226). The authors wish to acknowledge the contribution of  
1116 Alexandre Belisle and the genotyping team of the McGill University and Génome Québec  
1117 Innovation Centre, Montréal, Canada, for genotyping the Sequenom panel in the NFCCR  
1118 samples. Funding was provided to Michael O. Woods by the Canadian Cancer Society  
1119 Research Institute.

1120 NSHDS: The research was supported by Biobank Sweden through funding from the Swedish  
1121 Research Council (VR 2017-00650, VR 2017-01737), the Swedish Cancer Society (CAN  
1122 2017/581), Region Västerbotten (VLL-841671, VLL-833291), Knut and Alice Wallenberg  
1123 Foundation (VLL-765961), and the Lion's Cancer Research Foundation (several grants) and  
1124 Insamlingsstiftelsen, both at Umeå University.

1125 OSUMC: OCCPI funding was provided by Pelotonia and HNPCC funding was provided by the  
1126 NCI (CA016058 and CA067941).

1127 PLCO: Intramural Research Program of the Division of Cancer Epidemiology and Genetics and  
1128 supported by contracts from the Division of Cancer Prevention, National Cancer Institute, NIH,  
1129 DHHS. Funding was provided by National Institutes of Health (NIH), Genes, Environment and  
1130 Health Initiative (GEI) Z01 CP 010200, NIH U01 HG004446, and NIH GEI U01 HG 004438.

1131 SEARCH: The University of Cambridge has received salary support in respect of PDPP from  
1132 the NHS in the East of England through the Clinical Academic Reserve. Cancer Research UK

1133 (C490/A16561); the UK National Institute for Health Research Biomedical Research Centres at  
1134 the University of Cambridge.

1135 SELECT: Research reported in this publication was supported in part by the National Cancer  
1136 Institute of the National Institutes of Health under Award Numbers U10 CA037429 (CD Blanke),  
1137 and UM1 CA182883 (CM Tangen/IM Thompson). The content is solely the responsibility of the  
1138 authors and does not necessarily represent the official views of the National Institutes of Health.

1139 SMS and REACHS: This work was supported by the National Cancer Institute (grant P01  
1140 CA074184 to J.D.P. and P.A.N., grants R01 CA097325, R03 CA153323, and K05 CA152715 to  
1141 P.A.N., and the National Center for Advancing Translational Sciences at the National Institutes  
1142 of Health (grant KL2 TR000421 to A.N.B.-H.)

1143 The Swedish Low-risk Colorectal Cancer Study: The study was supported by grants from the  
1144 Swedish research council; K2015-55X-22674-01-4, K2008-55X-20157-03-3, K2006-72X-20157-  
1145 01-2 and the Stockholm County Council (ALF project).

1146 Swedish Mammography Cohort and Cohort of Swedish Men: This work is supported by the  
1147 Swedish Research Council /Infrastructure grant, the Swedish Cancer Foundation, and the  
1148 Karolinska Institute's Distinguished Professor Award to Alicja Wolk.

1149 UK Biobank: This research has been conducted using the UK Biobank Resource under  
1150 Application Number 8614

1151 VITAL: National Institutes of Health (K05 CA154337).

1152 The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes  
1153 of Health, U.S. Department of Health and Human Services through contracts 75N92021D00001,  
1154 75N92021D00002, 75N92021D00003, 75N92021D00004, 75N92021D00005







