scientific reports

OPEN

Application of Mendelian randomization to explore the causal role of the human gut microbiome in colorectal cancer

Charlie Hatcher[1,2,28], George Richenberg[1,2,28], Samuel Waterson[2,3], Long H. Nguyen[4], Amit D. Joshi[4], Robert Carreras-Torres[5,6], Victor Moreno[5,7,8,9], Andrew T. Chan[4], Marc Gunter[10], Yi Lin[11], Conghui Qu[12], Mingyang Song[13,14,15], Graham Casey[16], Jane C. Figueiredo[17], Stephen B. Gruber[18], Jochen Hampe[19], Heather Hampel[20], Mark A. Jenkins[21], Temitope O. Keku[22], Ulrike Peters[12], Catherine M. Tangen[23], Anna H. Wu[24], David A. Hughes[1,2], Malte C. Rühlemann[25], Jeroen Raes[26,27], Nicholas J. Timpson[1,2] & Kaitlin H. Wade[1,2✉]

The role of the human gut microbiome in colorectal cancer (CRC) is unclear as most studies on the topic are unable to discern correlation from causation. We apply two-sample Mendelian randomization (MR) to estimate the causal relationship between the gut microbiome and CRC. We used summary-level data from independent genome-wide association studies to estimate the causal effect of 14 microbial traits (n = 3890 individuals) on overall CRC (55,168 cases, 65,160 controls) and site-specific CRC risk, conducting several sensitivity analyses to understand the nature of results. Initial MR analysis suggested that a higher abundance of *Bifidobacterium* and presence of an unclassified

[1]MRC Integrative Epidemiology Unit, University of Bristol, Oakfield House, Oakfield Grove, Bristol BS8 2BN, UK. [2]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol BS8 2BN, UK. [3]North Bristol NHS Trust, Bristol BS10 5NB, UK. [4]Massachusetts General Hospital, Boston, MA 02114, USA. [5]Colorectal Cancer Group, ONCOBELL Program, Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, 08908 Barcelona, Spain. [6]Digestive Diseases and Microbiota Group, Girona Biomedical Research Institute (IDIBGI), 17190 Salt, Girona, Spain. [7]Oncology Data Analytics Program, Catalan Institute of Oncology (ICO), Hospitalet de Llobregat, Barcelona, Spain. [8]Biomedical Research Centre Network for Epidemiology and Public Health (CIBERESP), Madrid, Spain. [9]Department of Clinical Sciences, Universitat de Barcelona Institute of Complex Systems (UBICS), Faculty of Medicine and Health Sciences, University of Barcelona, Barcelona, Spain. [10]International Agency for Research on Cancer, 150 Cours Albert Thomas, 69372 Lyon, CEDEX 08, France. [11]Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. N., Seattle, WA 98109, USA. [12]Public Health Sciences Division, Fred Hutchinson Cancer Center, Seattle, WA, USA. [13]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA. [14]Division of Gastroenterology, Clinical and Translational Epidemiology Unit, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02115, USA. [15]Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA. [16]Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA. [17]Department of Medicine, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA. [18]Department of Preventive Medicine, USC Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA. [19]Department of Medicine I, University Hospital Dresden, Technische Universität Dresden (TU Dresden), Dresden, Germany. [20]Division of Human Genetics, Department of Internal Medicine, The Ohio State University Comprehensive Cancer Center, Columbus, OH, USA. [21]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, VIC, Australia. [22]Center for Gastrointestinal Biology and Disease, University of North Carolina, Chapel Hill, NC, USA. [23]SWOG Statistical Center, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [24]Preventative Medicine, University of Southern California, Los Angeles, CA, USA. [25]Institute of Clinical Molecular Biology, Christian Albrechts University of Kiel, Kiel, Germany. [26]Department of Microbiology and Immunology, Rega Instituut, KU Leuven, University of Leuven, Leuven, Belgium. [27]Center for Microbiology, VIB, Leuven, Belgium. [28]These authors contributed equally: Charlie Hatcher and George Richenberg. ✉email: Kaitlin.Wade@bristol.ac.uk

group of bacteria within the *Bacteroidales* order in the gut increased overall and site-specific CRC risk. However, sensitivity analyses suggested that instruments used to estimate relationships were likely complex and involved in many potential horizontal pleiotropic pathways, demonstrating that caution is needed when interpreting MR analyses with gut microbiome exposures. In assessing reverse causality, we did not find strong evidence that CRC causally affected these microbial traits. Whilst our study initially identified potential causal roles for two microbial traits in CRC, importantly, further exploration of these relationships highlighted that these were unlikely to reflect causality.

Colorectal cancer (CRC) is the third most common cause of cancer death worldwide, with over 16,000 deaths due to CRC every year in the UK alone[1,2]. Whilst research suggests that over half of CRC cases are likely explained by preventable causes[1,3], CRC remains an ever-increasing burden, particularly among young people. Therefore, it is important to identify novel modifiable risk factors to help reduce the global incidence of, and deaths from, CRC.

There is increasing evidence that the human gut microbiome—the naturally occurring complex community of microorganisms housed within the gastrointestinal tract—plays a role in human health, including influencing the risk and prognosis of CRC. As the gut microbiome has a substantial impact on host metabolism[4,5], inflammation[6,7], and host immune response to both commensal and pathogenic microbes[7,8], there are many plausible biological mechanisms by which the gut microbiome could influence cancer development[4,9,10]. However, findings within this context have been inconsistent and unclear, making it difficult to draw conclusive evidence. For example, whilst there have been some compelling results from in vivo and in vitro studies showing a reduced incidence of CRC through modification of single or a small consortia of microbial constituents (e.g., with pre- or pro-biotics)[11], few of these findings have been successfully translated between model organisms and into humans harbouring more complex communities.

Despite this emerging evidence, studies generating this body of work have been unconvincing in their ability to offer causal evidence given their observational nature and lack of temporality. For example, evidence has suggested an overall lower diversity in microbiota in cases of CRC compared to controls in addition to lower levels of *Bifidobacterium* and *Roseburia* spp. and higher levels of *Fusobacterium* and *Porphyromonas* bacteria[12–14]. However, these existing epidemiological studies are usually cross-sectional or case–control designs and have a limited ability to discern correlation from causation. In addition, they often have small sample sizes, limiting power to detect associations. Furthermore, human randomized controlled trials (RCTs) have been unable to conclusively show strong evidence for a reduction in the incidence of CRC using treatments designed to alter the gut microbiome[15]. Furthermore, discrepancies in the literature are likely due to the challenges in multi-omic technologies (e.g., sequencing and metagenomics), sensitive experimental models and important limitations of conventional epidemiological studies such as confounding, reverse causation and bias. Evidence that has not translated between model organisms impedes opportunity for harnessing the gut microbiome for improving population health[16].

In the absence of large-scale RCTs, Mendelian randomization (MR) is a method that offers improved causal inference by utilizing human germline genetic variation (usually, single nucleotide polymorphisms [SNPs]) as instruments for a clinically relevant trait (here, the gut microbiome) in a manner that is analogous to RCTs to mitigate some of the biases present within conventional epidemiological studies[17–20]. As germline genetic variation is randomly inherited and fixed at conception, results of MR analyses should be largely independent from confounding and traditional reverse causation (i.e., the outcome leading to variation in the exposure). There is an increasing literature utilising MR to assess the causal implications of variation in the gut microbiome (using host genetic variation associated with gut microbial composition) on various health outcomes. However, there is an unmet requirement for careful examination and interpretation of these estimates given the potential complexity of host (i.e., human) genetic effects on the gut microbiome, despite their ability to yield apparently causal estimates. In this study, we utilised and explored the properties of two-sample MR analyses based on host microbiome genetic effects to interrogate causal relationships between variation in the human gut microbiome (specifically, 14 microbial traits representing either relative abundances or the likelihood of presence versus absence of certain bacteria) on overall and site-specific CRC risk.

Results

Effects of gut microbial traits on CRC.   We used two-sample MR using genome-wide association study (GWAS) summary statistics to examine the causal relationship between features of the human gut microbiome and the risk of developing overall CRC and site-specific CRC (distal, proximal, colon, and rectal) (see Methods, Supplementary Table 1 and Fig. 1). Genetic variants associated with 14 microbial traits were obtained from a microbiome GWAS (mGWAS) meta-analyses[21] conducted in 3,890 individuals. Specifically, 13 SNPs exceeding a genome-wide meta-analysis threshold ($P < 2.5 \times 10^{-08}$) were selected as genetic instruments for 13 microbial traits (with each microbial trait instrumented by a single SNP), in addition to one SNP associated with bacteria of the *Bifidobacterium* genus that has previously consistently been reported in the literature. F-statistics were greater than 12 (range: 12.73–38.46) (Supplementary Table 2), suggesting that the likelihood of weak instrument bias was relatively low. The variance explained (R^2) by each associated SNP in each microbial trait ranged between 0.8% (the genus, *Veillonella*) and 1.6% (the unclassified group of bacteria in the *Porphyromonadaceae* family).

In main MR analyses, there was evidence for a causal role of a higher abundance of bacteria within the *Bifidobacterium* genus (*G. Bifidobacterium*) increasing the risk of overall CRC by approximately 41% (odds ratio [OR] per standard deviation [SD] higher relative abundance: 1.41; 95% confidence interval [CI]: 1.20, 1.65, $P = 1.83 \times 10^{-05}$; Fig. 2). The presence (versus absence) of an unclassified group of bacteria within the *Bacteroidales*
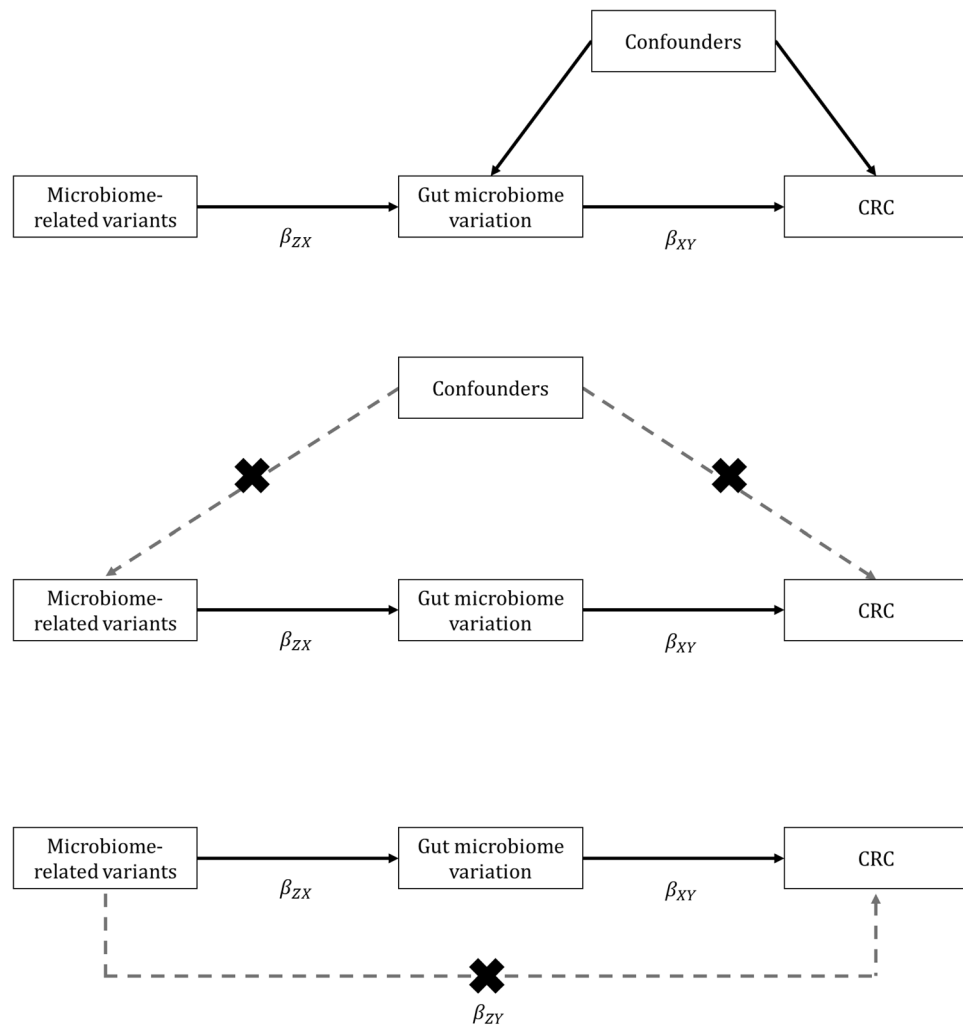
Figure 1. Mendelian randomization framework applied to assess the causal effect of the human gut microbiome on CRC risk. CRC = colorectal cancer; MR = Mendelian randomization; SNP = single nucleotide polymorphism. MR relies on three key assumptions: (top panel) the SNPs are associated with the exposure (here, the gut microbiome); (middle panel) there are no common causes of the SNPs and the outcome (here, CRC), meaning any confounders driven by population substructure, dynastic effects or assortative mating (these may not and are unlikely to be the same confounders of the association between the exposure and outcome); and (bottom panel) the SNPs are not independently associated with the outcome (here, CRC) other than pathways through the exposure (here, the gut microbiome). Given these key assumptions, microbiome-related variants can be used to assess the causal effect of the human gut microbiome on CRC, overcoming limitations of observational epidemiological studies. In two-sample MR analyses, the causal effect of the exposure on outcome (β_{XY}) is generated by a ratio of the SNP-outcome (β_{ZY}) and the SNP-exposure (β_{ZX}) effect estimates derived from two independent samples.

order (*G. unclassified, O. Bacteroidales*) was also found to increase the risk of overall CRC (OR per approximate doubling of the genetic liability to presence versus absence: 1.08; 95% CI: 1.02, 1.16; $P = 0.01$). Similar relationships were observed between these two microbial traits and the four CRC sites studied (Table 1). Specifically, a higher abundance of *Bifidobacterium* increased the risk of all four CRC sites: distal colon cancer (OR: 1.61; 95% CI: 1.26, 2.07; $P = 1.40 \times 10^{-04}$), proximal colon cancer (OR: 1.31; 95% CI: 1.03, 1.67; $P = 0.03$), colon cancer (OR: 1.40; 95% CI: 1.16, 1.69; $P = 3.75 \times 10^{-04}$) and rectal cancer (OR: 1.61; 95% CI: 1.26, 2.05; $P = 1.16 \times 10^{-04}$). Results for the unclassified group of bacteria within the *Bacteroidales* order also showed consistently positive directions of effect for all CRC sites: distal colon cancer (OR: 1.04; 95% CI: 0.93, 1.15; $P = 0.51$), proximal colon cancer (OR: 1.12; 95% CI: 1.02, 1.24; $P = 0.02$), colon cancer (OR: 1.09; 95% CI: 1.01, 1.18; $P = 0.03$) and rectal cancer (OR: 1.11; 95% CI: 1.00, 1.23; $P = 0.05$). Whilst the effect estimate for distal colon cancer was positive, the CIs were wide, likely due to the smaller sample size; therefore, the strength of evidence for a causal effect of bacteria within the *Bacteroidales* order and distal colon cancer was weaker than for other CRC sites. MR effect estimates for the remaining 12 microbial traits on overall and site-specific CRC risk were much smaller in size and had 95% CIs that crossed the null (Table 1).
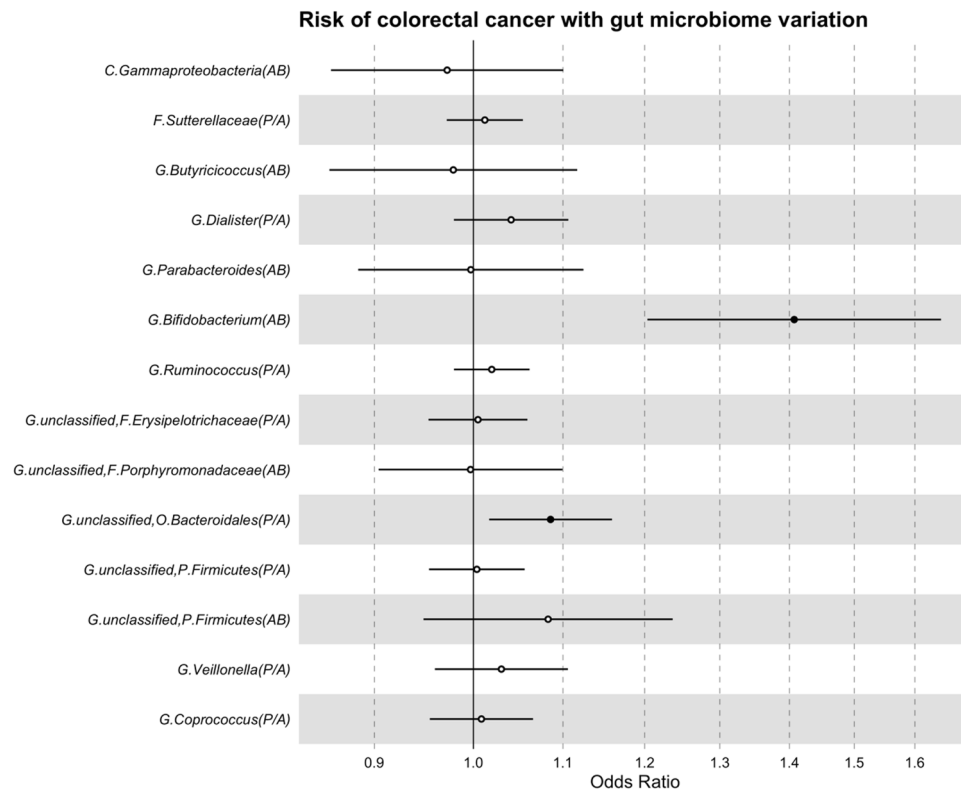
Figure 2. MR estimates of the effect of each microbial trait on overall CRC risk. AB = abundance; CI = confidence interval; CRC = colorectal cancer; MR = Mendelian randomization; OR = odds ratio; P/A = presence versus absence; SD = standard deviation. Letters in the microbial trait name represent the taxon classification level from which that microbial trait was observed, with "C", "F", "G", "O" and "P" representing "class", "family", "genus", "order" and "phylum", respectively. All microbial traits that were not confidently classified at the genus level were organised into unclassified groups within higher classification levels (represented by "unclassified"). MR estimates represent the OR for CRC risk and 95% CI per SD unit change for continuous microbial traits (labelled as "AB" in brackets) or per approximate doubling of the genetic liability to presence (versus absence) of each binary microbial trait (labelled as "P/A" in brackets).

Sensitivity analyses.   *Colocalisation.*   We performed several sensitivity analyses to test the robustness of our findings to violations of MR assumptions and explore possible explanations of the observed causal effects of *G. Bifidobacterium* and *G. unclassified, O. Bacteroidales* on CRC and site-specific CRC risk. Firstly, we performed genetic colocalisation to determine whether the single genetic variant being used as an instrument for each microbial trait was associated with both variation in that microbial trait and CRC, which is a necessary (but not a sufficient) criterion for causality. Specifically, we performed colocalisation on the two microbial traits which showed strong evidence of a causal effect in the MR analyses (*G. Bifidobacterium* and *G. unclassified, O. Bacteroidales*) with overall and site-specific CRC (distal CRC was not considered in the colocalisation analyses with *G. unclassified, O. Bacteroidales* since evidence for a causal effect was weaker in the main MR analyses). Genome-wide data for microbial traits were only available from the Flemish Gut Flora Project (FGFP) GWAS limiting analyses to 2223 participants. Additionally, unlike in the mGWAS meta-analysis, neither lead SNP (rs4988235 and rs116135844) reached genome-wide significance in the FGFP dataset. Therefore, focusing on the difference between the tested hypotheses (see Methods), colocalisation results provided little evidence that *G. unclassified, O. Bacteroidales* was likely to share a causal variant with overall or site-specific CRC (Table 2).

Regional association plots confirmed findings from genetic colocalisation for *G. unclassified, O. Bacteroidales* (Fig. 3; Supplementary Fig. 1). Whilst there was evidence for genetic colocalisation for *G. Bifidobacterium* with overall (with a posterior probability of 0.87) CRC risk and, to a lower extent, with site-specific CRC risk (posterior probabilities ranged from 0.05–0.72), regional association plots suggest the wider genomic region surrounding the rs4988235 SNP is important in both abundance of *G. Bifidobacterium* and CRC risk (Fig. 3; Supplementary Fig. 2).

Manual exploration of pleiotropy.   In the absence of being able to apply pleiotropy-robust methods, which require multiple genetic instruments associated with each microbial trait, we performed a manual exploration of horizontal pleiotropy to identify any associations between the microbiome-related SNPs and CRC (either directly or indirectly through alternative phenotypes), which would invalidate one of the core MR assumptions

Microbial trait	MR estimate using the Wald ratio (95% CI); *p*-value				
	Overall CRC risk	Distal CRC	Proximal CRC	Colon cancer	Rectal cancer
C. Gammaproteobacteria (AB)	0.97 (0.86, 1.00); 0.66	0.94 (0.77, 1.14); 0.52	0.96 (0.80, 1.17); 0.71	0.92 (0.79, 1.07); 0.28	1.04 (0.86, 1.26); 0.66
F. Sutterellaceae (P/A)	1.01 (0.97, 1.05); 0.56	1.02 (0.96, 1.09); 0.50	0.99 (0.93, 1.05); 0.75	1.02 (0.97, 1.07); 0.49	0.99 (0.93, 1.05); 0.65
G. Bifidobacterium (AB)	1.41 (1.20, 1.65); 1.83×10^{-05}	1.61 (1.26, 2.07); 1.40×10^{-04}	1.31 (1.03, 1.67); 0.03	1.40 (1.16, 1.69); 3.75×10^{-04}	1.61 (1.26, 2.05); 1.16×10^{-04}
G. Butyricicoccus (AB)	0.98 (0.86, 1.11); 0.75	0.97 (0.78, 1.19); 0.75	1.02 (0.83, 1.25); 0.83	0.99 (0.85, 1.16); 0.92	0.99 (0.81, 1.22); 0.94
G. Coprococcus (P/A)	1.00 (0.95, 1.07); 0.76	0.97 (0.89, 1.06); 0.54	1.05 (0.96, 1.14); 0.28	1.01 (0.95, 1.08); 0.70	1.04 (0.95, 1.13); 0.40
G. Dialister (P/A)	1.04 (0.98, 1.10); 0.20	1.05 (0.95, 1.16); 0.31	1.00 (0.91, 1.10); 0.95	1.04 (0.97, 1.12); 0.30	1.02 (0.93, 1.13); 0.59
G. Parabacteroides (AB)	1.00 (0.88, 1.12); 0.96	0.92 (0.76, 1.11); 0.39	0.95 (0.79, 1.14); 0.57	0.97 (0.84, 1.12); 0.69	1.03 (0.86, 1.15); 0.71
G. Ruminococcus (P/A)	1.02 (0.98, 1.06); 0.34	1.00 (0.93, 1.07); 0.97	1.00 (0.93, 1.07); 0.96	1.00 (0.95, 1.06); 0.93	1.04 (0.97, 1.11); 0.32
G. unclassified, F. Erysipelotrichaceae (P/A)	1.00 (0.95, 1.06); 0.86	0.99 (0.91, 1.07); 0.76	0.93 (0.86, 1.01); 0.11	0.97 (0.91, 1.04); 0.41	1.00 (0.92, 1.08); 0.95
G. unclassified, F. Porphyromonadaceae (AB)	1.00 (0.90, 1.10); 0.95	1.04 (0.89, 1.21); 0.66	1.00 (0.85, 1.16); 0.96	1.05 (0.90, 1.14); 0.79	0.93 (0.80, 1.09); 0.38
G. unclassified, O. Bacteroidales (P/A)	1.08 (1.02, 1.16); 0.01	1.04 (0.93, 1.15); 0.51	1.12 (1.02, 1.24); 0.02	1.09 (1.01, 1.18); 0.03	1.11 (1.00, 1.23); 0.05
G. unclassified, P. Firmicutes (P/A)	1.00 (0.96, 1.05); 0.89	0.98 (0.91, 1.07); 0.67	1.03 (0.95, 1.11); 0.54	1.01 (0.95, 1.08); 0.64	1.03 (0.95, 1.18); 0.45
G. unclassified, P. Firmicutes (AB)	1.08 (0.95, 1.24); 0.24	1.00 (0.81, 1.23); 0.97	1.09 (0.89, 1.34); 0.42	1.06 (0.88, 1.22); 0.65	1.21 (0.98, 1.49); 0.07
G. Veillonella (P/A)	1.03 (0.96, 1.10); 0.41	0.97 (0.86, 1.08); 0.57	1.08 (0.97, 1.20); 0.14	1.07 (0.95, 1.12); 0.50	1.08 (0.98, 1.21); 0.13

Table 1. Two-sample MR estimates of the effect of each microbial trait on overall and site-specific CRC risk. AB = abundance; CI = confidence interval; CRC = colorectal cancer; MR = Mendelian randomization; OR = odds ratio; P/A = presence versus absence. Letters in the microbial trait name represent the taxon classification level from which that microbial trait was observed, with "*C*", "*F*", "*G*", "*O*" and "*P*" representing "class", "family", "genus", "order" and "phylum", respectively. All microbial traits that were not confidently classified at the genus level were organised into unclassified groups within higher classification levels (represented by "unclassified"). Effect estimates represent the OR for CRC risk for a standard deviation higher AB of each continuous microbial trait and the OR for CRC for an approximate doubling of the genetic liability to P/A of each binary microbial trait.

Microbial trait and CRC risk	H0	H1	H2	H3	H4
G. Bifidobacterium (AB) and overall CRC	0.01	0.03	0.02	0.07	0.87
G. Bifidobacterium (AB) and distal colon cancer risk	0.03	0.12	0.03	0.11	0.71
G. Bifidobacterium (AB) and proximal colon cancer risk	0.17	0.67	0.02	0.09	0.05
G. Bifidobacterium (AB) and colon cancer risk	0.08	0.31	0.03	0.11	0.47
G. Bifidobacterium (AB) and rectal cancer risk	0.02	0.08	0.04	0.14	0.72
G. unclassified, O. Bacteroidales (P/A) and overall CRC	0.13	0.63	0.02	0.09	0.13
G. unclassified, O. Bacteroidales (P/A) and proximal colon cancer risk	0.12	0.60	0.03	0.13	0.12
G. unclassified, O. Bacteroidales (P/A) and colon cancer risk	0.14	0.66	0.02	0.10	0.08
G. unclassified, O. Bacteroidales (P/A) and rectal cancer risk	0.13	0.63	0.03	0.14	0.07

Table 2. Posterior probabilities relating to associations between microbial traits and both overall and site-specific colorectal cancer. AB = abundance; CRC = colorectal cancer; P/A = presence versus absence. Letters in the microbial trait name represent the taxon classification level from which that microbial trait was observed, with "*G*" and "*O*" representing "genus" and "order", respectively. All microbial traits that were not confidently classified at the genus level were organised into unclassified groups within higher classification levels (represented by "unclassified"). H0: Neither trait has a genetic association in the region, H1: Only the first trait (i.e., the microbial trait) has a genetic association in the region, H2; Only the second trait (i.e., CRC risk) has a genetic association in the region, H3: Both traits are associated but have different causal variants and H4: Both traits are associated and have the same causal variant.

(see Methods). We searched summary-level GWAS data collated by PhenoScanner[22] and the IEU OpenGWAS[23] to identify whether SNPs used to instrument *G. Bifidobacterium* and *G. unclassified, O. Bacteroidales* had previously been reported to be associated with CRC or any other trait that could be an independent cause of CRC. Searching GWAS summary statistics collated by PhenoScanner found evidence for potential horizontal pleiotropy between the SNP used as an instrument for *G. Bifidobacterium* (rs4988235) and CRC risk. A total of 51 associations with complex traits and diseases were observed at a defined multiple testing p-value threshold ($P < 1 \times 10^{-04}$, see Methods) including positive associations with numerous anthropometric measures (i.e., fat
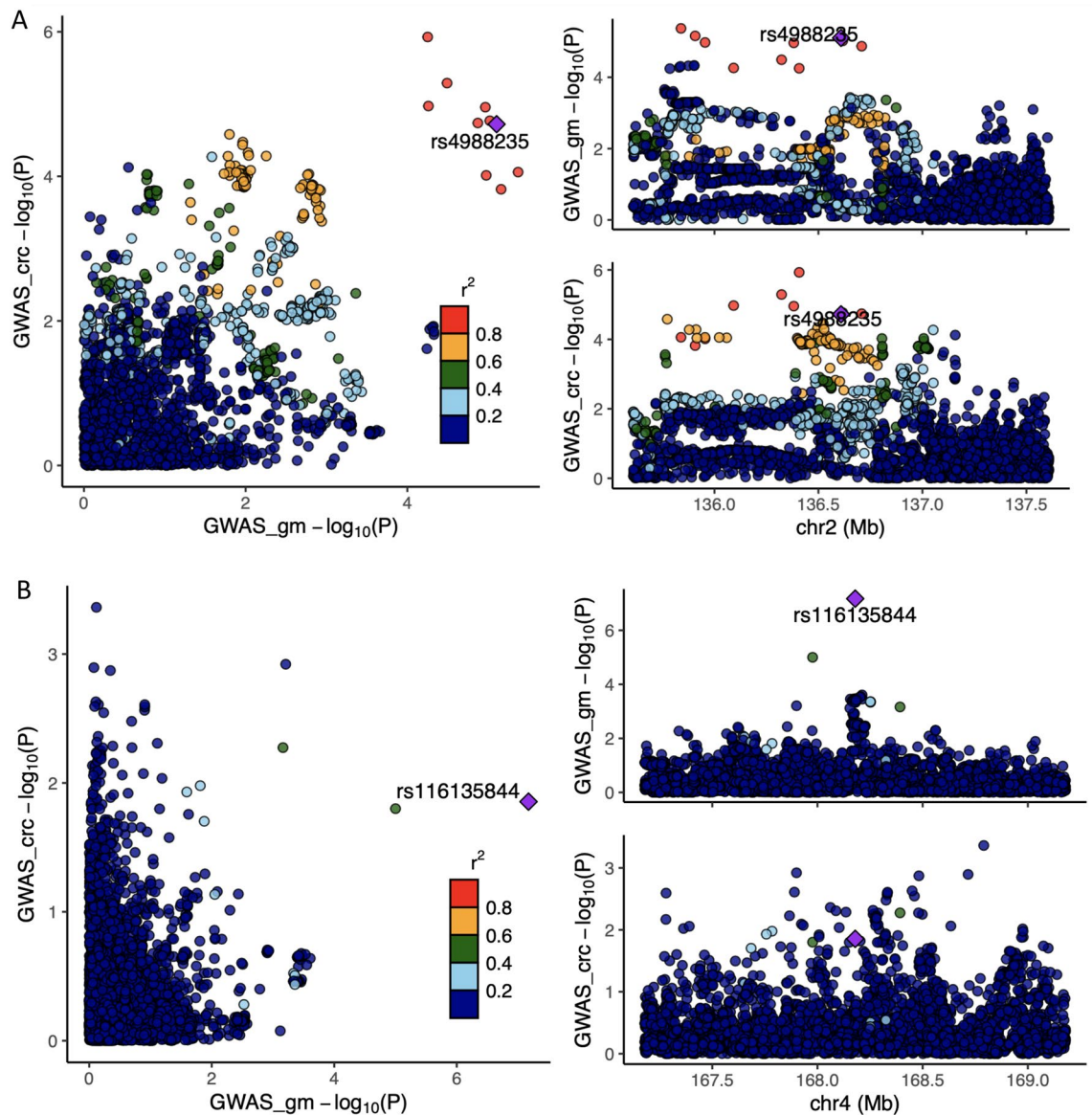
Figure 3. Colocalisation results for bacteria within the (**A**) *Bifidobacterium* AB microbial trait and (**B**) *P/A* of unclassified genera within the *Bacteroidales* order with overall CRC. AB = abundance; CRC = colorectal cancer; FGFP = Flemish Gut Flora Project; GECCO = Genetics and Epidemiology of Colorectal Cancer Consortium; GM = gut microbiome; GWAS = genome-wide association study; P/A = presence versus absence. Regional association plots, generated from LocusCompareR, showing the − log10(*P*-value) where each lead SNP is represented by a purple diamond (panel A: rs4988235 associated with the *Bifidobacterium* AB microbial trait and panel B: rs116135844 associated with the unclassified *Bacteroidales* P/A microbial trait) in relation to overall CRC. These plots were created using the FGFP and GECCO full summary-level data for microbial traits and CRC, respectively.

mass, body mass index and waist circumference). Additionally, the rs4988235 SNP has also previously been associated with 103 expression quantitative traits (eQTs), 6 metabolites, 59 methylation quantitative traits (mQTs) and one protein (Supplementary Tables 3–7). In the IEU OpenGWAS, over 230 traits were identified as being associated with rs4988235 at the same p-value threshold (Supplementary Table 8). Given the number of associations between this SNP and both complex and molecular traits, there are likely many plausible pathways between this SNP and CRC, which could be independent of the *G. Bifidobacterium*.

In contrast, we found very little evidence for horizontal pleiotropy between the SNP used as an instrument for the *G. unclassified, O. Bacteroidales* (rs116135844) and CRC risk, with only three associations observed at the pre-defined p-value threshold ($P < 1 \times 10^{-04}$; Supplementary Tables 9–12). These traits included disorders of the patella, which is less likely to be biologically linked to CRC and therefore unlikely to reflect horizontal pleiotropy between the instrument and outcome. The other traits implicated in relation to rs116135844 included gene expression of the *UBE2J2* gene and an uncharacterized genetic probe (2490351). The *UBE2J2* gene encodes

a member of the E2 ubiquitin-conjugating enzyme family that modifies abnormal proteins or short-lived proteins targeted for degradation. The GWAS Catalog showed various associations with *UBE2J2*, including many cardiometabolic measures (e.g., body mass index, systolic blood pressure and coronary heart disease), autoimmune diseases (e.g., inflammatory bowel diseases and ulcerative colitis), CRC and others, suggesting *UBE2J2* gene expression could have a downstream impact on CRC. Further to this, no associations of this SNP with any phenotypes in the IEU OpenGWAS met our p-value threshold (Supplementary Table 13).

Two-sample MR using a lenient p-value threshold for selection of genetic instruments. To allow the use of more formal pleiotropy-robust methods, a more lenient p-value threshold of 1×10^{-05} was used to select genetic instruments for *G. Bifidobacterium* and *G. unclassified, O. Bacteroidales*, as is commonly used in the literature[24,25]. In addition to using this lenient p-value threshold (which may increase the risk of weak instrument bias), associated SNPs were also then restricted to those which had directionally consistent effect estimates in each of the three studies in the mGWAS (FGFP, FoCus and PopGen). When using a more lenient *p*-value threshold ($P < 1 \times 10^{-05}$) and restricting to SNPs with directionally consistent effect estimates across the three studies comprising the gut microbiome GWAS (Supplementary Table 14), nine of the original 14 microbial traits had additional associated SNPs that could be used as theoretical instruments for further MR analyses. Focusing on those microbial traits for which there was evidence of a causal effect on CRC risk, *G. Bifidobacterium* had five associated SNPs and *G. unclassified, O. Bacteroidales* had three associated SNPs in total. Inverse variance weighted (IVW)-derived estimates for the causal effect of these microbial traits on overall CRC were consistent in direction to those obtained in the main analyses; however, the magnitude of all estimates was attenuated, and CIs spanned the null (Fig. 4, Supplementary Table 15). Specifically, a higher abundance of *G. Bifidobacterium* and presence of *G. unclassified, O. Bacteroidales* increased the risk of overall CRC by 16% (OR: 1.16; 95% CI: 0.97, 1.38; *P* = 0.10) and 5% (OR:
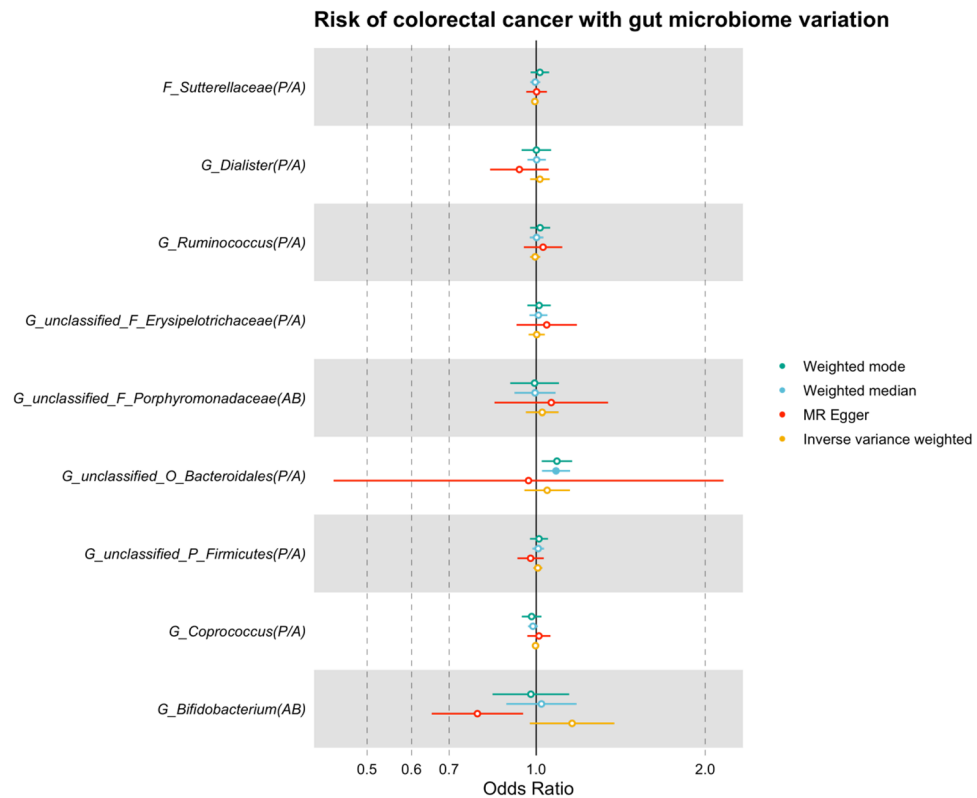
Figure 4. MR estimates of the effect of each microbial trait on overall CRC risk using a more lenient *p*-value threshold for selecting genetic instruments. AB = abundance; CI = confidence interval; CRC = colorectal cancer; mGWAS = microbiome genome-wide association study; MR = Mendelian randomization; OR = odds ratio; P/A = presence versus absence; SD = standard deviation. Letters in the microbial trait name represent the taxon classification level from which that microbial trait was observed, with "C", "F", "G", "O" and "P" representing "class", "family", "genus", "order" and "phylum", respectively. All microbial traits that were not confidently classified at the genus level were organised into unclassified groups within higher classification levels (represented by "unclassified"). MR estimates represent the OR for CRC risk and 95% CI per SD unit change for AB microbial traits or per approximate doubling of the genetic liability to *P/A* of each binary microbial trait. Results for the inverse variance weighted, weighted median, weighted mode and MR-Egger methods are presented when using the multi-SNP instrument for each microbial trait using a lenient p-value threshold ($P < 1 \times 10^{-05}$) and directional consistency across cohorts included in the mGWAS and were compared to the Wald ratio estimates obtained from the main analysis.

1.05; 95% CI: 0.95, 1.15; *P* = 0.34), respectively. Results for other microbial traits were consistent with the main analyses and provided little evidence for a causal effect on CRC risk.

Estimates of the causal effect of *G. Bifidobacterium* on overall CRC risk using the pleiotropy-robust methods were inconsistent with those derived using the Wald ratio and IVW methods, where the weighted median estimate was positive but both the MR-Egger and weighted mode-derived estimates were in the inverse direction (MR-Egger OR: 0.79; 95% CI: 0.65, 0.95; *P* = 0.09 and weighted mode OR: 0.99; 95% CI: 0.82, 1.14; *P* = 0.80). The inconsistent results obtained by these approaches suggest that horizontal pleiotropy may in part explain the relationship between *G. Bifidobacterium* and CRC risk, whereby the SNP associated with *G. Bifidobacterium* (rs4988325) may impact the risk of CRC independent of the microbial trait. For *G. unclassified, O. Bacteroidales*, estimates using the weighted median and weighted mode methods were consistent in direction suggesting that presence of this group of bacteria increased the risk of overall CRC; however, the effect estimate was in the reverse direction using the MR-Egger method and CIs were predictably wide (Fig. 4, Supplementary Table 15).

For site-specific analyses, IVW-derived effect estimates for the effect of *G. Bifidobacterium* were directionally consistent with the main analyses using the Wald ratio, though some estimates had wide CIs. The estimates for the causal effect of *G. Bifidobacterium* on distal, proximal and colon cancer were also consistent; however, results using the MR-Egger method were inverse for all sites of CRC with wide CIs. The estimates for the effect of *G. unclassified, O. Bacteroidales* on distal, proximal and colon cancer risk using the IVW and pleiotropy-robust methods were all directionally consistent (though with wide CIs) with main analyses, suggesting presence of this bacteria also increased the risk of site-specific CRC. However, estimates for the effect of both microbial traits on rectal cancer were inconsistent across all models, with the IVW and weighted median-based estimates being positive (and thus consistent with the main analyses) but either/both of the MR-Egger and weighted mode-based estimates were inverse. Overall, these results may indicate that the originally observed causal effects in main analyses may be driven by horizontal pleiotropy (and downstream effects), which is also supported by our manual exploration of possible horizontal pleiotropic pathways.

Reverse MR. To examine whether the causal effects identified between microbial traits and CRC risk were either bi-directional in nature or driven by reverse causation of the microbial trait-associated SNPs (i.e., the "microbial trait-related" SNP actually having a direct or indirect effect on CRC, which, in turn, actually impacts microbial trait variation), we performed reverse two-sample MR analyses. These analyses aimed to assess whether overall and site-specific CRC risk had a causal effect on *G. Bifidobacterium* and *G. unclassified, O. Bacteroidales* microbial traits. All of the 57 associated with overall CRC risk and all but one SNP associated with all site-specific CRC risk were available in the FGFP GWAS summary-level data for these two microbial traits (i.e., leaving 30, 16, 39 and 30 independent SNPs associated with distal colon, proximal colon, colon and rectal CRC sites for reverse MR analyses). Results gave little evidence that overall CRC or CRC sites had a causal effect on either microbial trait, with CIs spanning the null for all IVW-derived estimates (Supplementary Table 16).

Discussion

In this study, we performed analyses to explore the causal role played by the gut microbiome and CRC. Whilst our initial MR analyses provided evidence that bacteria within the *Bifidobacterium* genera and an unclassified group of bacteria in the *Bacteroidales* order may increase the risk of overall and site-specific CRC, our sensitivity analyses highlighted that these relationships, and the tools used to assess them, are complex and, certainly in the case of *Bifidobacterium*, are unlikely to reflect causality. Our study therefore highlights the need for performing sensitivity analyses exploring the nature of the instruments and derived estimates when attempting to establish the causal role of the gut microbiome in health and disease.

Firstly, there was strong evidence for pleiotropy in the relationship between *G. Bifidobacterium* and both overall and site-specific CRC risk, where there were comfortably over 200 other traits that had previously been associated with the rs4988235 SNP that could be independently associated with CRC. Furthermore, when using multiple SNPs at a more lenient p-value threshold, the direction of effect was inconsistent across pleiotropy-robust methods. In reverse MR analyses, there was also no evidence to suggest a reverse effect in that a greater genetic liability to CRC was unlikely to drive variation in *Bifidobacterium* abundance. Therefore, the initially observed finding suggesting that a higher abundance of bacteria with the *Bifidobacterium* genus was likely not to reflect causality and, instead, reflect horizontal pleiotropy or a complex (potentially gene-environment) interaction between the genomic region surrounding the rs4988235 SNP (i.e., the *LCT/MCM6* gene) with both *Bifidobacterium* abundance and CRC risk, as indicated by our colocalisation analyses.

Conversely, our main results suggesting that the presence of an unclassified group of bacteria within the *Bacteroidales* order increased risk of CRC seemed not to be driven entirely by horizontal pleiotropy, as there were only a small number of plausible traits associated with the rs116135844 SNP that could be independently associated with CRC. In addition, results when using a more lenient *p*-value threshold were consistent with those from our main analyses. There was also no evidence to suggest a reverse effect in that a genetic liability to CRC was unlikely to change the greater genetic liability to presence of this unclassified group of *Bacteroidales* bacteria. Indeed, these findings support previous observational evidence suggesting that bacteria within the *Bacteroidales* order are more present in CRC cases compared to controls[26–28]. There are also many existing studies showing a relationship between various *Bacteroides* species including *Bacteroides fragilis* and other species including *B. dorei, B. vulgatus* and *B. massilensis* with CRC[29]. Whilst we were unable to classify the specific bacteria within this unclassified *Bacteroidales* microbial trait, these bacteria could plausibly contribute to the effect observed in the current analyses. Therefore, research using more granular measures of the gut microbiome (e.g., those capturing species- and strain-level bacteria or those able to characterize functionality such as metagenomics)

will help provide clarity on these results specifically. Nevertheless, these results provide further evidence that the previously reported association between *Bacteroidales* genera and CRC may reflect causality.

It is worth noting that, whilst limited by power, our colocalisation analyses provided some evidence to suggest that *G. Bifidobacterium* (but not *G. unclassified, O. Bacteroidales*) shared causal variants with CRC, which is a necessary (but not a sufficient) criterion for causality. Therefore, further analyses with a much larger quantity of microbiome-related SNPs, larger GWASs from which these SNPs are discovered and importantly replicated, plus a better understanding of the relationships between host genetic variation and the gut microbiome and comprehensive sensitivity analyses that we present here are required to provide conclusive evidence in this context.

Recently, a similar two-sample MR study assessing the causal relationship between gut microbiota and CRC reported evidence for a causal role of bacteria within two families (*Verrucomicrobiaceae* and *Enterobacteriaceae*) and three genera (*Akkermansia, Blautia, and Ruminococcus*) in CRC, which we did not replicate here[30]. Whilst the estimates for the effect of bacteria within the *Ruminococcus* genus on CRC risk found in our study were directionally consistent with those presented by Ni et al., the magnitude of these estimates was smaller in our study, and CIs spanned the null. There are several likely reasons for the difference in findings reported across these two papers. Firstly, with the exception of *Ruminococcus,* we did not consider the other microbial traits examined by Ni et al. as exposures in our analyses due to our selection of genetic instruments at the genome-wide p-value threshold defined by the mGWAS ($P < 2.5 \times 10^{-08}$), which is best practice in MR analyses. Ni et al., selected genetic instruments at a lenient p-value threshold within their study ($P < 1 \times 10^{-05}$), without consideration for heterogeneity across cohorts, which could induce bias through weak instrumentation, horizontal pleiotropy, genetic confounding and reverse causation. The sample size of the mGWAS used by Ni et al. was also much smaller than that which was used in our study, and was conducted within a single sample.

Generally, there is poor replication of microbiome-related SNPs reaching genome-wide significance across different mGWASs (likely due to differences in sampling, data processing and GWAS analysis methodology)[16]. Unlike the majority of the current studies applying MR to understand the causal implications of the gut microbiome and various health outcomes, including that published by Ni et al., the main strength of this study is that the microbiome-related SNPs used as instruments for the microbial traits are some of the first persistent signals across multiple cohorts[21]. Whilst acknowledging that instrumentation of the gut microbiome is complex, we used genetic variants that were robustly related to each microbial trait (i.e., those reaching a genome-wide p-value threshold in the mGWAS). Compared to existing studies of this kind (which tend to use a lenient p-value threshold of 1×10^{-05} as a main analysis), this reduces the likelihood of including invalid SNPs within MR instruments. We did, however, opt to use more lenient p-value threshold for selection of genetic instruments ($P < 1 \times 10^{-05}$) as a sensitivity analysis but importantly restricted SNPs to those which had a consistent direction of effect across multiple studies under the assumption that, in a larger GWAS, these SNPs would have reached a traditional genome-wide p-value threshold. Together, this provides more confidence that these are biologically relevant SNPs or at least those that were consistently associated with the microbial traits in this analysis. Additionally, GECCO is the largest GWAS of CRC to date, meaning that we were very well powered to detect a modest association in this current MR analysis (i.e., assuming an alpha level of 0.05 and an R^2 of 0.01, we were 89% powered to detect an odds ratio of 1.2 in CRC). Generally, our study has shown the importance of performing several sensitivity analyses to assess the robustness of MR findings.

Despite this, there are several limitations to this work that centre around the core assumptions of the MR framework. Firstly, there must be no pathway between genetic instruments and the outcome (here, CRC) independent of the microbial exposure. Given that very little is known about the biology of the host genetic effects on the gut microbiome, there is a greater likelihood that the observed GWAS "microbiome-related" signals are reflective of either host-driven effects upstream of the gut microbiome (i.e., reverse causation in an MR context) or independent associations between these host genetic variants and, in this case, CRC (i.e., horizontal pleiotropy). This will remain a substantial limitation within the field without further large-scale GWASs of the gut microbiome and inter-disciplinary analyses clarifying the mechanistic pathways explaining the relationships between host genetic variation and gut microbiome variation.

Secondly, there are many microbiota of lower taxonomic units within the group of unclassified genera in the *Bacteroidales* order, emphasising a need to classify the exact species or strain of bacteria that could be driving this relationship and understand the mechanisms by which this occurs. Even if these early results do suggest causality (with the case of *G. unclassified, O. Bacteroidales*), we do not currently know whether attempts to alter microbiota to reduce CRC risk would have other, unforeseen effects on other aspects of health. This is particularly important for the inclusion of the gut microbiome in the development of targeted therapeutics or preventative strategies for CRC. Lastly, these analyses were undertaken with only individuals of European ancestry; therefore, it is not clear how generalisable these results are to other populations, especially those with differing gut microbial compositions.

Our study initially provided evidence that individuals who have a higher abundance of bacteria within the *Bifidobacterium* genus and presence of an unclassified group of bacteria within the *Bacteroidales* order within the gut may be at an increased risk of CRC and its sites. However, through exploration of invalidation of MR assumptions, we performed numerous sensitivity analyses which showed that these relationships are complex and may not reflect direct or causal relationships. Therefore, caution is required when interpreting these results and MR estimates alone are insufficient to determine causality in these contexts. Inter-disciplinary collaboration and triangulation across multiple, complementary study designs are required to understand the causal mechanisms that link both host genetic variation with the gut microbiome and the relationships between the gut microbiome and human health.

Materials and methods

Study design. Two-sample MR was used to examine the causal relationship between features of the human gut microbiome and both overall and site-specific CRC risk. In brief, SNP-exposure and SNP-outcome associations were obtained from independent non-overlapping GWASs to generate causal effect estimates of 14 microbial traits on overall and site-specific CRC risk. This study and all methods have been conducted in line with the Strengthening the Reporting of Observational Studies in Epidemiology MR (STROBE-MR) reporting guidelines for MR studies[31]. All methods have been carried out in accordance with relevant guidelines and regulations.

Gut microbiome GWAS data and instrument selection. Genetic variants associated with microbial traits were obtained from one of the largest microbiome GWAS (mGWAS) meta-analyses of bacterial abundance, presence (versus absence), alpha- and beta-diversity and enterotype conducted within the FGFP (n = 2,223) and two independent validation cohorts (the Food-Chain Plus study [FoCus, n = 950] and the PopGen study [n = 717])[21]. Full details of the data sampling, preparation and analyses have been described previously[21,32,33]. Briefly, DNA was extracted from frozen fecal samples provided by participants and hypervariable regions of the 16S rRNA gene were amplified and sequenced (the V4 hypervariable region for FGFP and the V1-V2 hypervariable regions for both FoCus and PopGen). Resulting sequences were analysed per sample using the DADA2 pipeline to provide taxonomical classifications of all bacteria in each sample[34]. Bacterial classifications that were not confidently assigned at the genus level were organised into arbitrary unclassified groups within higher classification levels. Informed consent and appropriate ethical approval were obtained for each study.

In the FGFP cohort, which was used the discovery cohort in the GWAS analysis, the DADA2 pipeline yielded 499 taxonomical units across five levels of microbiota phylogeny from phylum to genus. Four metrics on alpha-diversity (within-individual diversity) and beta-diversity (inter-individual differences) were estimated using data for all 288 genus-level taxa. An enterotype phenotype, describing the different phylum-level community compositions, was also estimated.

After performing taxa-level quality control (QC), 92 taxa were available for the mGWAS (details previously described[21]). Given the ecological count nature of the 16S rRNA data, many microbial traits had zero-inflated distributions, which is problematic for linear modelling. Of the 92 taxa, 62 contained substantial zero-inflation; therefore, these taxa were modelled using a two-step hurdle binary analysis, which includes a binary presence (versus absence) analysis (denoted throughout as "P/A") and a zero-truncated rank normal transformed abundance analysis (denoted throughout as "AB"). All other taxa were relatively normally distributed; therefore, were treated as AB phenotypes and rank normal transformed accordingly. Thus, a total of 159 microbial traits (i.e., 62 P/A, 92 AB, 3 alpha-diversity, 1 beta-diversity and 1 enterotype phenotypes) were analysed in the mGWAS conducted in the FGFP cohort.

Individuals from the FGFP cohort were genotyped on the Human Core Exome v1.0 and the Human Core Exome v1.1. Imputation was conducted using the 1000 Genomes data (phase 3) as the reference panel. After variant-level and individual-level QC followed by imputation, 7,711,310 variants and 2259 individuals remained, 2223 of whom also had data on the gut microbiome and all covariates used in the mGWAS analysis. All individuals were of European ancestry.

All mGWAS analyses were adjusted for extraction type and year, aliquot year, person performing the aliquot, library preparation plate, the first 10 genetic principal components, sex and age. Assuming an additive genetic model and accounting for genotype uncertainty, all AB and alpha-diversity microbial traits were regressed on covariates and residuals were regressed on genotype probability data in univariate linear models, all P/A microbial traits were analysed using a multivariate logistic model, all beta-diversity microbial traits were analysed using a multivariate model and the enterotype microbial trait was analysed using a multinomial logistic regression for categorial traits. All SNPs that reached an inclusive association test p-value threshold of 1×10^{-05} in the FGFP data set (n = 23,735) were taken forward into a targeted meta-analysis including the independent FoCus and PopGen studies. Three genera were not present in these German cohorts; therefore, after excluding both P/A and AB microbial traits in these three instances (i.e., six microbial traits in total), the meta-analysis was limited to 153 microbial traits and 23,067 SNPs.

The FoCus and PopGen cohorts were genotyped using the Illumina Omni Express + Exome array and the Affymetrix Genome-Wide Human SNP Array 6.0, respectively. Imputation was conducted using the Sanger Imputation Service with the Human Reference Consortium (HRC) version 1.1 as the reference panel. The mGWAS meta-analyses were performed using the inverse-variance fixed effects method, with SNPs considered "meta-supported" if the p-value of association became smaller in the meta-analysis compared to that obtained in FGFP alone. To identify all independent SNPs associated with microbial traits, all meta-supported markers were clumped using plink and its default settings. The genome-wide meta-analysis p-value threshold was defined as 2.5×10^{-08} and the study-wide p-value threshold was defined as 1.57×10^{-10} using Bonferroni correction (assuming 2 million independent genetic association tests across 159 microbial traits).

Of the 153 microbial traits tested, two SNPs showed evidence of association that exceeded the study-wide meta-analysis p-value threshold and a further 11 exceeded the genome-wide meta-analysis threshold, where each SNP was associated with one microbial trait. These 13 SNPs were selected as genetic instruments for the 13 microbial traits within this two-sample MR analysis with overall and site-specific CRC, in addition to one SNP associated with bacteria of the *Bifidobacterium* genus that has previously been reported in the literature (Supplementary Table 2). Effect estimates from the mGWAS represent an increase in the log-transformed OR for P/A microbial traits and the standard deviation (SD) change for rank normalised AB microbial traits with each effect allele carried.

Colorectal cancer GWAS data. Data for CRC were obtained from the most comprehensive GWAS of overall CRC to date from a meta-analysis of 120,328 individuals (comprising 55,168 CRC cases and 65,160 controls). The GWAS meta-analysis included the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO), Colorectal Cancer Transdisciplinary Study (CORECT), OncoArray + Custom, OmniExpress + Exome Chip, COloRectal cancer Study of Austria (CORSA) and UK Biobank (Supplementary Table 1) and full information on genotyping, imputation and QC have been described previously[35,36]. Data for CRC sites were available across colon cancer (N = 32,002), which is a sum of distal CRC, proximal CRC and cancer cases with unspecified sites, distal CRC (N = 14,376), proximal CRC (N = 15,706) and rectal cancer (N = 16,212). CRC case status was diagnosed by a physician and approximately 92% of participants were white and of European ancestry. Data access was granted and the study was approved by the GECCO Coordinating Centre.

Statistical analyses. *Two-sample Mendelian randomization.* In our primary analyses, we performed two-sample MR using the *TwoSampleMR*[37] package (version 0.5.6) in R to examine the causal relationship between 14 microbial traits and both overall and site-specific CRC risk.

Summary-level data (i.e., the SNP name [rsid], effect allele, other allele, effect allele frequency, beta coefficient, standard error, p-value and sample size) for each of the 14 microbial trait-associated SNPs associated were extracted from both the mGWAS and CRC GWAS meta-analysis. The proportion of variance explained (R^2) in each microbial trait by each SNP and the strength of the instrument (assessed through the F-statistic) were calculated. For each binary microbial trait, the proportion of variance explained of the liability to presence (versus absence) was obtained using the "get_r_from_lor" function of the *TwoSampleMR* package[37] (the estimate of which was then squared to get the R^2). The required parameter for this function describing prevalence of each microbial trait was obtained from FGFP alone (as published by Hughes et al.[21]). For continuous microbial traits (i.e., abundances), the R^2 was calculated using the following formula[38]:

$$R^2 = \frac{2 \times \beta^2 \times MAF \times (1 - MAF)}{2 \times \beta^2 \times MAF \times (1 - MAF) + (SE(\beta))^2 \times (2 \times N) \times MAF \times (1 - MAF)}$$

where β is the effect size (beta-coefficient), *MAF* is the minor allele frequency and *SE(β)* is the standard error of the effect size for a given SNP, and *N* is the sample size of the mGWAS. The F-statistic was calculated as follows:

$$F = R^2 \times (N - 1 - k))/((1 - R^2) \times k)$$

where R^2 is the proportion of variance explained in the microbial trait by the SNP, *N* is the sample size of the GWAS, and *k* is the number of SNPs included in the instrument (i.e., for our main analysis, k = 1 for each microbial trait)[39].

The exposure and outcome datasets were harmonized such that the effect of each SNP on the exposure and outcome was relative to the same effect allele. For ambiguously coded SNPs (i.e., "palindromic" SNPs where the effect/other allele were either an A/T or G/C combination), we used the effect allele frequency to resolve strand ambiguity, where possible. Non-inferable SNPs (i.e., "palindromic" SNPs with a MAF > 0.42) were removed from the analysis.

Given that there was only one SNP associated with each microbial trait, the Wald ratio method was used as the main analyses, which estimates the effect of the exposure (here, each microbial traits) on the outcome (here, overall and site-specific CRC risk) by dividing the SNP-outcome association by the SNP-exposure association (Fig. 1)[19].

Sensitivity analyses. Three assumptions, namely (1) the relevance assumption (the genetic instruments used to proxy the exposure are strongly associated with that exposure), (2) the independence assumption (there is no confounding between the genetic instrument(s) used to proxy the exposure and the outcome) and (3) the exclusion restriction assumption (the genetic instruments used to proxy the exposure have no effect on the outcome other than through the exposure) must be met when assessing evidence for causal relationships between an exposure and outcome using MR[17–20]. We explored possible invalidations of each of these assumptions in our analyses by performing several sensitivity analyses that test the robustness of our findings and explored possible explanations of any observed causal effect via a process of elimination.

Firstly, genetic colocalisation was used to evaluate whether there was a shared causal variant at each locus responsible for variation in the gut microbiome and for conferring CRC risk, which is necessary but not sufficient for establishing causality. Secondly, in the absence of using formal statistical techniques that assess robustness of causal estimates owed to having only one SNP associated with each microbial trait, an alternative approach was used to test for any invalidation of the third MR assumption (i.e., the exclusion restriction assumption). We manually searched SNPs in two databases (PhenoScanner and the IEU OpenGWAS database) to identify traits they have previously been associated with. Thirdly, to formally test potential pleiotropy, we selected genetic variants to proxy microbial traits at a more lenient p-value threshold of 1×10^{-05} and repeated MR analyses with several complementary, pleiotropy-robust methods. Lastly, to assess whether any causal effect observed between microbial traits and CRC in our main analyses was indicative of reverse causation, we additionally performed two-sample MR in the reverse direction (i.e., with overall and site-specific CRC as the exposure and microbial traits as the outcome). Each of these sensitivity analyses are described in detail below.

Colocalisation. Colocalisation analyses were conducted using the 'coloc' R package with default parameters (i.e., with the prior probabilities of the SNP being associated with the exposure, the outcome or both traits being specified as 1×10^{-04}, and 1×10^{-05}, respectively[40]), additionally specifying that the exposures (i.e., microbial

traits) were continuous and that the outcome (i.e., CRC risk) was binary (with either the 'quant' or 'cc' option, respectively, specified as the type of data). Bayes factor computation was used to generate 5 posterior probabilities (H0-H4) characterised by the following outcomes: (H0) neither trait has a genetic association in the region; (H1) only the gut microbiome has a genetic association in the region; (H2) only CRC has a genetic association in the region; (H3) both traits are associated but have different causal variants and (H4) both traits are associated and have the same causal variant. We used a posterior probability threshold ≥ 0.80 to indicate evidence of a shared common causal variant between each microbial trait and CRC. Full summary statistics were only available for the FGFP cohort for gut microbiome variation; therefore, genetic variants ± 1 Mb of the lead SNP associated with any microbial trait for which there was evidence for a causal impact on CRC in our main analyses were extracted from FGFP and GECCO genome-wide datasets. Regional association plots were generated to visualise genetic colocalisation using the LocusCompareR package[41].

Manual exploration of pleiotropy. Where there was evidence for a causal effect of any microbial trait on CRC risk, summary-level GWAS data collated by PhenoScanner and the IEU OpenGWAS were searched to identify whether any SNP used as an instrument for those microbial traits had been reported to be associated with CRC or any other trait that could be an independent cause of CRC. A lenient p-value threshold of 1×10^{-04} was set as the multiple testing threshold at which we defined evidence for an association between the SNP and any disease or trait in PhenoScanner and the IEU OpenGWAS[22,42]. This multiple testing correction was based on the number of results returned by the online PhenoScanner interface (i.e., 1000, as of May 2022) and for consistency in the inclusive p-value threshold selected across the two platforms, where a 10% Bonferroni correction was applied due to the likelihood that traits in those databases were correlated.

Two-sample MR using a lenient p-value threshold to select genetic instruments. To allow the use of more formal pleiotropy-robust methods that require multiple genetic instruments, a more lenient p-value threshold of 1×10^{-05} was used to expand the number of SNPs associated with each of the microbial traits in the main analyses, reflecting the p-value threshold commonly used in the literature[24,25]. However, in addition to increasing this *p*-value threshold (which may increase the risk of weak instrument bias), SNPs reaching this lenient *p*-value threshold were then also restricted to those which had directionally consistent effect estimates in each of the three studies in the mGWAS[21] (FGFP, FoCus and PopGen; see list of all directionally consistent SNPs in Supplementary Table 14) to reduce heterogeneity in the effect of each SNP on the exposure. The IVW method was used and compared to the main analyses that utilised the Wald ratio estimator. The IVW method meta-analyses effect estimates across all SNPs weighted by the inverse variance of the SNP-outcome association using fixed effects.

The caveats of using multiple SNPs defined at a more lenient p-value threshold in MR analyses are increasing the likelihood of weak instrument bias and introducing horizontal pleiotropic pathways between the SNPs and the outcome. Therefore, to test for horizontal pleiotropy in these sensitivity analyses, the weighted median[43], weighted mode[44] and MR-Egger[45] regression methods were also applied and consistency of effect estimates were compared to those obtained from the IVW method.

The weighted median[43] requires that only half the SNPs are valid instruments (i.e., exhibiting no horizontal pleiotropy, no confounders of the instrument-outcome association and a robust association with the exposure) for the causal effect estimate to be unbiased. The mode-based estimator clusters the SNPs into groups based on similarity of causal effects and returns the causal effect estimate based on the cluster that has the largest number of SNPs. The weighted mode[44] introduces an extra element similar to the IVW and weighted median estimators, weighting the contribution of each SNP to the clustering by the inverse variance of its outcome effect.

The MR-Egger[45] method is similar to the IVW approach but relaxes the "no horizontal pleiotropy" assumption. MR-Egger regression allows a non-zero intercept in the relationship between multiple SNP-outcome and SNP-exposure associations, where the intercept provides a formal statistical test for the presence of directional (bias inducing) horizontal pleiotropy. The slope of the MR-Egger regression between multiple SNP-outcome and SNP-exposure associations can be considered as an unbiased causal effect between the exposure (here, microbial traits) and the outcome (here, CRC), assuming any horizontal pleiotropic effects are not correlated with the SNP-exposure effects (i.e., strength of the instrument). Violations of the third MR assumption were also assessed by visual inspection of funnel[19], forest, scatter and leave-one-out plots, and tests of heterogeneity[18] of effects between the SNPs using Cochan's Q statistic[46].

Effect estimates obtained from all two-sample MR analyses represent the OR for both overall and site-specific CRC risk for each SD higher abundance of each continuous microbial trait (those denoted with "AB" in tables) and the OR for CRC for an approximate doubling of the genetic liability to presence (versus absence) of each binary microbial trait (those denoted with "P/A" in tables)[47]. For all analyses, P-values were interpreted as continuous indicators of evidence strength and conclusions were drawn based on effect sizes and their precision. Given the high correlation between microbial traits, there was no correction for multiple testing. Analyses and data visualisation were conducted using the *TwoSampleMR* R package (version 0.5.5) and the ieugwasr R package (0.1.5) in with R (version 4.1.0) and PhenoScanner (version 2) online[22,37,42].

Reverse two-sample Mendelian randomization. For reverse MR analyses, 57 independent SNPs (i.e., 31 new and 26 that were previously known) that reached the conventional genome-wide p-value threshold ($P < 5 \times 10^{-08}$) were selected as instruments for overall CRC risk, obtained from a large GWAS meta-analysis of CRC comprising 34,627 CRC cases and 71,379 controls of European ancestry[48]. For CRC sites, 109 SNPs (i.e., 13 new and 96 that were previously known) associated with each of the four CRC sites ($P < 5 \times 10^{-08}$) were obtained from the largest and most recent GWAS meta-analysis of anatomical site-specific loci[35], where the number of SNPs associated with each were 102, 99, 108 and 103 for distal colon, proximal colon, colon and rectal, respectively.

These SNPs were then clumped (r-squared = 0.001) to obtain 31, 17, 40 and 31 independent SNPs associated with distal colon, proximal colon, colon and rectal, respectively. Summary statistics for CRC-associated SNPs were extracted for the 14 microbial traits from the FGFP cohort (n = 2223) from the Hughes et al., GWAS[21]. MR analyses were undertaken in the same was as described above, with the IVW method used as the main analyses, results of which were compared to those obtained from the MR-Egger, weighted median, and weighted mode-based estimators that test the assumptions of no pleiotropy among genetic instruments and outcomes.

Data availability

GWAS summary-level data used in this study were publicly available for microbiome GWAS (mGWAS) conducted by Hughes et al.[21]. Summary-level data for overall and site-specific CRC were obtained directly from GECCO (summary-level data are publicly available for the CRC GWASs used in the reverse MR analyses[35,48].

References

1. Brown, K. F. et al. The fraction of cancer attributable to modifiable risk factors in England, Wales, Scotland, Northern Ireland, and the United Kingdom in 2015. Br. J. Cancer 118(8), 1130–1141. https://doi.org/10.1038/s41416-018-0029-6 (2018).
2. Ferlay, J. et al. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. Int. J. Cancer 136(5), E359–E386 (2015).
3. Doll, R. & Peto, R. The causes of cancer: Quantitative estimates of avoidable risks of cancer in the United States today. J. Natl. Cancer Inst. 66(6), 1191–1308 (1981).
4. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. Cell 144(5), 646–674 (2011).
5. Visconti, A. et al. Interplay between the human gut microbiome and host metabolism. Nat. Commun. 10(1), 4505. https://doi.org/10.1038/s41467-019-12476-z (2019).
6. Francescone, R., Hou, V. & Grivennikov, S. I. Microbiome, inflammation, and cancer. Cancer J. 20(3), 181–189 (2014).
7. Blander, J. M., Longman, R. S., Iliev, I. D., Sonnenberg, G. F. & Artis, D. Regulation of inflammation by microbiota interactions with the host. Nat. Immunol. 18(8), 851–860 (2017).
8. Zheng, D., Liwinski, T. & Elinav, E. Interaction between microbiota and immunity in health and disease. Cell Res. 30(6), 492–506. https://doi.org/10.1038/s41422-020-0332-7 (2020).
9. Gopalakrishnan, V., Helmink, B. A., Spencer, C. N., Reuben, A. & Wargo, J. A. The influence of the gut microbiome on cancer, immunity, and cancer immunotherapy. Cancer Cell 33(4), 570–580 (2018).
10. Garrett, W. S. Cancer and the microbiota. Science 348(6230), 80–86 (2015).
11. Ambalam, P., Raman, M., Purama, R. K. & Doble, M. Probiotics, prebiotics and colorectal cancer prevention. Best Pract. Res. Clin. Gastroenterol. 30(1), 119–131 (2016).
12. Ahn, J. et al. Human gut microbiome and risk for colorectal cancer. J. Natl. Cancer Inst. 105(24), 1907–1911 (2013).
13. Zhang, Y. et al. Changes in gut microbiota and plasma inflammatory factors across the stages of colorectal tumorigenesis: A case–control study. BMC Microbiol. 18(1), 92 (2018).
14. Rezasoltani, S. et al. The association between fecal microbiota and different types of colorectal polyp as precursors of colorectal cancer. Microb. Pathog. 124, 244–249 (2018).
15. Ishikawa, H. et al. Randomized trial of dietary fiber and Lactobacillus casei administration for prevention of colorectal tumors. Int. J. Cancer 116(5), 762–767 (2005).
16. Wade, K. H. & Hall, L. J. Improving causality in microbiome research: Can human genetic epidemiology help?. Wellcome Open Res. 4, 199 (2019).
17. Davey Smith, G. & Hemani, G. Mendelian randomization: Genetic anchors for causal inference in epidemiological studies. Hum. Mol. Genet. 23(R1), R89-98. https://doi.org/10.1093/hmg/ddu328 (2014).
18. Lawlor, D. A., Wade, K., Borges, M. C., Palmer, T. M., Hartwig, F. P., Hemani, G. et al. A Mendelian Randomization Dictionary: Useful Definitions and Descriptions for Undertaking, Understanding and Interpreting Mendelian Randomization Studies. OSF Prepr. (2019).
19. Haycock, P. C. et al. Best (but oft-forgotten) practices: The design, analysis, and interpretation of Mendelian randomization studies. Am. J. Clin. Nutr. 103(4), 965–978 (2016).
20. Davies, N. M., Holmes, M. V. & Davey, S. G. Reading Mendelian randomisation studies: A guide, glossary, and checklist for clinicians. BMJ 362, 601 (2018).
21. Hughes, D. A. et al. Genome-wide associations of human gut microbiome variation and implications for causal inference analyses. Nat. Microbiol. 5(9), 1079–1087. https://doi.org/10.1038/s41564-020-0743-8 (2020).
22. Kamat, M. A. et al. PhenoScanner V2: An expanded tool for searching human genotype–phenotype associations. Bioinformatics 35(22), 4851–4853. https://doi.org/10.1093/bioinformatics/btz469 (2019).
23. Elsworth, B., Lyon, M., Alexander, T., Liu, Y., Matthews, P., Hallett, J, et al. The MRC IEU OpenGWAS data infrastructure. bioRxiv [Internet]. 2020 Jan 1;2020.08.10.244293. Available from: http://biorxiv.org/content/early/2020/08/10/2020.08.10.244293.abstract
24. Sanna, S. et al. Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. Nat. Genet. 51(4), 600–605. https://doi.org/10.1038/s41588-019-0350-x (2019).
25. Zhang, Z.-J. et al. Assessment of causal direction between gut microbiota and inflammatory bowel disease: A mendelian randomization analysis. Front Genet. 12, 631061 (2021).
26. Boleij, A. et al. The Bacteroides fragilis toxin gene is prevalent in the colon mucosa of colorectal cancer patients. Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am. 60(2), 208–215 (2015).
27. Montalban-Arques, A. & Scharl, M. Intestinal microbiota and colorectal carcinoma: Implications for pathogenesis, diagnosis, and therapy. EBioMedicine 48, 648–655 (2019).
28. Yang, J. et al. Development of a colorectal cancer diagnostic model and dietary risk assessment through gut microbiome analysis. Exp. Mol. Med. 51(10), 1–15. https://doi.org/10.1038/s12276-019-0313-4 (2019).
29. Tilg, H., Adolph, T. E., Gerner, R. R. & Moschen, A. R. The intestinal microbiota in colorectal cancer. Cancer Cell 33(6), 954–964 (2018).
30. Ni, J.-J. et al. Mendelian randomization study of causal link from gut microbiota to colorectal cancer. BMC Cancer 22(1), 1371. https://doi.org/10.1186/s12885-022-10483-w (2022).
31. Skrivankova, V. W. et al. Strengthening the reporting of observational studies in epidemiology using Mendelian randomization: The STROBE-MR statement. JAMA 326(16), 1614–1621 (2021).
32. Wang, J. et al. Genome-wide association analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota. Nat. Genet. 48(11), 1396–1406. https://doi.org/10.1038/ng.3695 (2016).

33. Krawczak, M. *et al.* PopGen: Population-based recruitment of patients and controls for the analysis of complex genotype–phenotype relationships. *Commun. Genet.* **9**(1), 55–61 (2006).
34. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**(7), 581–583. https://doi.org/10.1038/nmeth.3869 (2016).
35. Huyghe, J. R. *et al.* Genetic architectures of proximal and distal colorectal cancer are partly distinct. *Gut* **70**(7), 1325–1334. https://doi.org/10.1136/gutjnl-2020-321534 (2021).
36. Schmit, S. L. *et al.* Novel common genetic susceptibility loci for colorectal cancer. *J. Natl. Cancer Inst.* **111**(2), 146–157 (2019).
37. Hemani, G. *et al.* The MR-base platform supports systematic causal inference across the human phenome. *Elife* **7**, e34408. https://doi.org/10.7554/eLife.34408 (2018).
38. Yarmolinsky, J. *et al.* Circulating selenium and prostate cancer risk: A Mendelian randomization analysis. *J. Natl. Cancer Inst.* **110**(9), 1035–1038 (2018).
39. Palmer, T. M. *et al.* Using multiple genetic variants as instrumental variables for modifiable risk factors. *Stat. Methods Med. Res.* **21**(3), 223–242 (2012).
40. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**(5), e1004383. https://doi.org/10.1371/journal.pgen.1004383 (2014).
41. Liu, B., Gloudemans, M. J., Rao, A. S., Ingelsson, E. & Montgomery, S. B. Abundant associations with gene expression complicate GWAS follow-up. *Nat. Genet.* **51**, 768–769 (2019).
42. Staley, J. R. *et al.* Phenoscanner: A database of human genotype–phenotype associations. *Bioinformatics* **32**(20), 3207–3209. https://doi.org/10.1093/bioinformatics/btw373 (2016).
43. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* **40**(4), 304–314 (2016).
44. Hartwig, F. P., Davey Smith, G. & Bowden, J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int. J. Epidemiol.* **46**(6), 1985–1998 (2017).
45. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**(2), 512–525 (2015).
46. Davies, N. M., Holmesm M. V., Davey Smith, G. Reading Mendelian randomisation studies: A guide, glossary, and checklist for clinicians. *BMJ*. 362 (2018).
47. Burgess, S. & Labrecque, J. A. Mendelian randomization with a binary exposure variable: Interpretation and presentation of causal estimates. *Eur. J. Epidemiol.* **33**(10), 947–952 (2018).
48. Law, P. J. *et al.* Association analyses identify 31 new risk loci for colorectal cancer susceptibility. *Nat. Commun.* **10**(1), 2154. https://doi.org/10.1038/s41467-019-09775-w (2019).

Author contributions

Funding

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-31840-0.

Correspondence and requests for materials should be addressed to K.H.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.