

RESEARCH ARTICLE

Open Access



Causal relationship between gut microbiota and cancers: a two-sample Mendelian randomisation study

Yiwen Long^{1,2†}, Lanhua Tang^{1,2†}, Yangying Zhou^{1,2}, Shushan Zhao^{2,3*} and Hong Zhu^{1,2*}

Abstract

Background Evidence from observational studies and clinical trials suggests that the gut microbiota is associated with cancer. However, the causal association between gut microbiota and cancer remains to be determined.

Methods We first identified two sets of gut microbiota based on phylum, class, order, family, and genus level information, and cancer data were obtained from the IEU Open GWAS project. We then performed two-sample Mendelian randomisation (MR) to determine whether the gut microbiota is causally associated with eight cancer types. Furthermore, we performed a bi-directional MR analysis to examine the direction of the causal relations.

Results We identified 11 causal relationships between genetic liability in the gut microbiome and cancer, including those involving the genus *Bifidobacterium*. We found 17 strong associations between genetic liability in the gut microbiome and cancer. Moreover, we found 24 associations between genetic liability in the gut microbiome and cancer using multiple datasets.

Conclusions Our MR analysis revealed that the gut microbiota was causally associated with cancers and may be useful in providing new insights for further mechanistic and clinical studies of microbiota-mediated cancer.

Keywords Gut microbiota, Cancer, Mendelian randomisation, Genetics, SNPs

Background

Gut microbiota residing in the gastrointestinal tract can be considered a potential environmental factor influencing human life. Currently, the gut microbiota has been

implicated as a risk or preventive factor for a variety of diseases, including cancers, and is closely associated with the onset of colorectal cancer (CRC) [1]. Conversely, it has been pointed out that cancer also affects the gut microbiota in mice, which could induce gut microbiota disorders and cancer growth [2].

In contrast to observational studies, randomised controlled trials of gut microbiota could potentially help establish a causal relationship. Unfortunately, owing to the influence of objective factors, such as technology and research methods, the screening of strains involved in early diagnosis and prognosis still has great limitations. Consequently, most of the current research conclusions are based on observation of the composition and changes in the gut microbiota in patients' faeces and on the results of trials in which gut

[†]Yiwen Long and Lanhua Tang contributed equally to this work.

*Correspondence:

Shushan Zhao

zhaoshuiquan@126.com

Hong Zhu

zhuHong0719@126.com

¹ Department of Oncology, Xiangya Hospital, Central South University, Changsha, Hunan 410008, People's Republic of China

² National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, Hunan 410008, People's Republic of China

³ Department of Orthopedics, Xiangya Hospital, Central South University, Changsha, Hunan 410008, People's Republic of China



microbiota were transplanted into gnotobiotic mice, which are influenced by a variety of factors, such as diet and antibiotic use [3–6]. In summary, whether the associations between the gut microbiota and cancers are causal and the direction of the causal associations are still unknown. It is thus important to explore the causal relationship between the gut microbiota and cancers.

Genome-wide association studies (GWASs) have tested millions of genetic variants across the genomes of many individuals to identify genotype–phenotype associations and have revolutionised the field of complex disease genetics over the past decade [7]. GWASs provide an agnostic approach for investigating the genetic basis of complex diseases. As of October 2022, the GWAS Catalog contained 6041 publications and 427,870 associations. In oncology, over 450 genetic variants associated with increased risks of common cancers have been identified. The clinical application of GWAS data has been providing opportunities for cancer prevention [8].

Mendelian randomisation (MR) analysis exploits the inherent properties of common genetic variations for a modifiable environmental exposure of interest and has become a widely used approach to explore the potential causal relationships between environmental exposures and diseases [9–11]. Two-sample MR analysis can utilise single-nucleotide polymorphism (SNP)-exposure and SNP-outcome associations from independent GWASs and combine them into a single causal estimate. As the number of GWASs on gut microbiota and diseases has increased rapidly [12, 13], large-scale summary statistics have become more widely available, allowing for two-sample MR analysis with significantly improved statistical power.

In the present study, we investigated the causal relationship between gut microbiota and a broad range of cancers by conducting a comprehensive two-sample MR analysis of eight cancers derived from the IEU Open GWAS project, including breast cancer, colorectal cancer, ovarian cancer, head and neck cancer, lung cancer, endometrial cancer, and prostate cancer. By applying a bi-directional MR approach, we can explore whether gut microbiota casually affects cancer risk and we can also examine whether the genetic predisposition to cancer risk causally influences the gut microbiota. Based on these, we tried to clarify the role of the gut microbiota in cancer development to eventually help to develop new treatment strategies, such as probiotic therapy, dietary modulations, and faecal microbiota transplantation (FMT) [14].

Methods

Exposure data

SNPs related to the human gut microbiome composition were selected as instrumental variables (IVs) from a GWAS dataset of the international consortium MiBioGen [13]. This was a multi-ethnic large-scale GWAS that coordinated 16S ribosomal RNA gene sequencing profiles and genotyping data from 18,340 participants from 24 cohorts from the USA, Canada, Israel, South Korea, Germany, Denmark, the Netherlands, Belgium, Sweden, Finland, and the UK to explore the association between autosomal human genetic variants and the gut microbiome. A total of 211 taxa (131 genera, 35 families, 20 orders, 16 classes, and 9 phyla) were included.

Outcome data

We downloaded all traits reported in the IEU Open GWAS project <https://gwas.mrcieu.ac.uk/> (updated to 2022.04.06, $N = 40,427$) and derived all cancer-related GWAS summary-level data. After screening the dataset and excluding duplicate studies, non-malignant tumours, and non-European ancestry, the GWAS summary-level data for the associations between genetic variants and cancers included those from the UK Biobank [15], the International Lung Cancer Consortium (ILCCO) [16, 17], the Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome (PRACTICA-L) consortium [18], the Medical Research Council-Integrative Epidemiology Unit (MRC-IEU) [19], the Ovarian Cancer Association Consortium (OCAC) [20], the Oncoarray oral cavity and oropharyngeal cancer [21], the Breast Cancer Association Consortium (BCAC) [22], FINNGEN [23], and Neale Lab (<http://www.nealelab.is-uk-biobank/>). Detailed information is provided in Additional file 1: Table S1.

Instrumental variable selection

The flowchart of the study is presented in Fig. 1. Briefly, the gut microbiota served as the exposure, whereas cancer served as the outcome.

Bacterial taxa were analysed at five levels (phylum, class, order, family, and genus), and a distinct taxon was defined as a feature. To ensure the authenticity and accuracy of the conclusions on the causal link between the gut microbiome and cancer risk, the following quality control steps were used to select the optimal IVs. First, SNPs that were significantly related to the gut microbiome were selected as the IVs. Two thresholds were used to select the IVs. The first threshold selected SNPs less than the genome-wide statistical significance threshold (5×10^{-8}) to serve as IVs. Unfortunately, after we selected SNPs, only a small number of gut microbiota were selected as

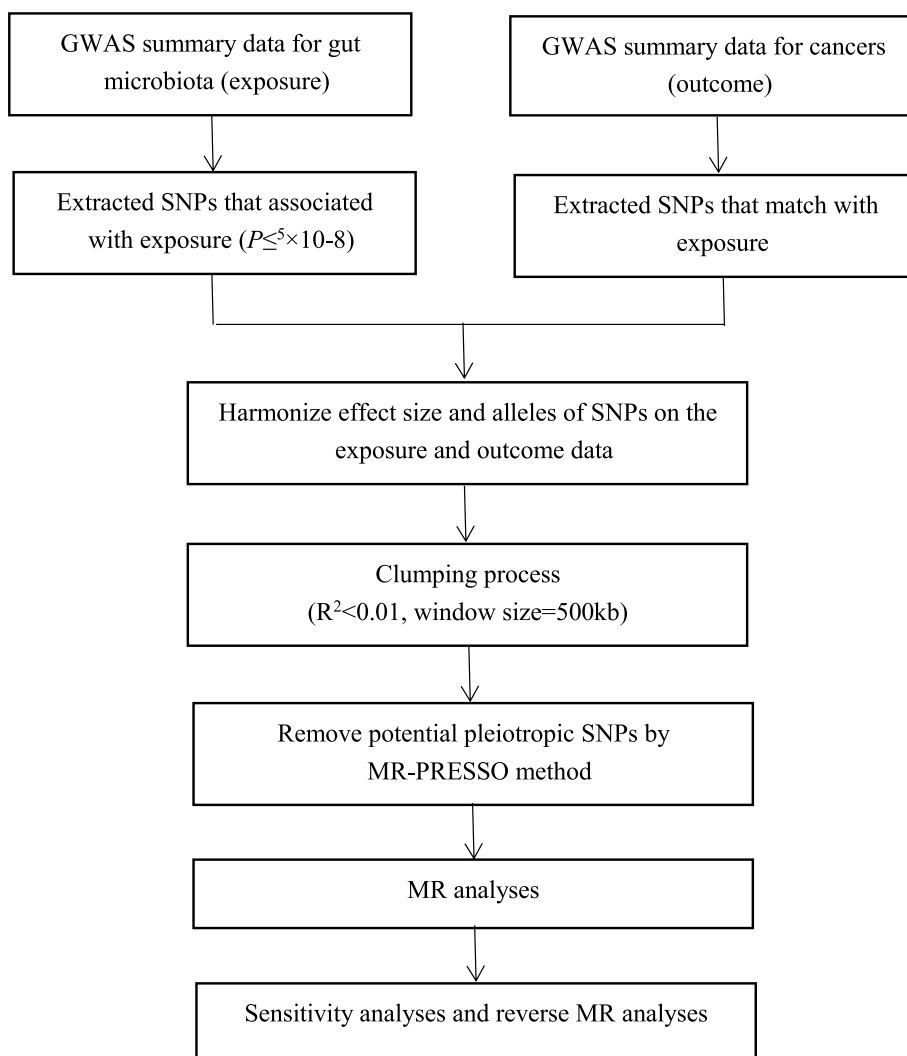


Fig. 1 Study design and workflow

IVs, and to explore more relations between cancers and gut microbiota to obtain more comprehensive results, we used the second threshold that identified SNPs that were smaller than the locus-wide significance level (1×10^{-6}) and selected them as the second IVs set to find more potential causal associations. Second, the minor allele frequency (MAF) threshold of the variants of interest was 0.01. Third, one of the principles of the MR approach is that there is no linkage disequilibrium (LD) among the included IVs, as the presence of strong LD might result in biased results. In the current study, the clumping process ($R^2 < 0.01$ and clumping distance = 10,000 kb) was conducted to assess the LD between the included SNPs. Fourth, an important step in MR is to ensure that the effects of the SNPs on the exposure correspond to the same allele as the effects on the outcome. To

avoid distortion of strand orientation or allele coding, we deleted palindromic SNPs (e.g. with A/T or G/C alleles). During the harmonisation process, we aligned the alleles to the human genome reference sequence (build 37) and removed ambiguous and duplicated SNPs.

We applied MR-PRESSO and MR-Egger regression tests to monitor the potential horizontal pleiotropy effect. For each SNP, the MR-PRESSO outlier test calculated a p -value for its pleiotropy significance, whereas the MR-PRESSO global test calculated a p -value for overall horizontal pleiotropy. SNPs were sorted in ascending order in terms of their MR-PRESSO outlier test p -values and were then removed one by one. The MR-PRESSO global test was performed on the remaining SNPs each time an SNP was removed from the list. Recursion was repeated until the p -value for the global test was insignificant ($p > 0.05$).

The list of SNPs remaining after removing pleiotropic SNPs was used for the subsequent MR analysis.

MR analysis

We performed an MR analysis to investigate the causal relationship between microbiome features and common cancers. For features containing only one IV, the Wald ratio test was used to estimate the association between the identified IV and each cancer [24]. Five popular MR methods were used for features containing multiple IVs: inverse-variance weighted (IVW) test [25], weighted mode [26], MR-Egger regression [27], weighted median estimator (WME) [28], and MR-PRESSO [29]. The IVW method is reported to be slightly more powerful than the others under certain conditions [28]; therefore, the results with more than one IV were mainly based on the IVW method, with the other four methods serving as complements.

Additionally, we established a multiple testing significance threshold at each feature level (phylum, class, order, family, and genus), defined as $p < 0.05/n$ (where n is the effective number of independent bacterial taxa at the corresponding taxonomic level). To assess the robustness of the results, several sensitivity analyses were performed. Leave-one-out analysis was performed to determine whether the causal signal was driven by a single SNP. This approach compares the variance explained by the IVs for both the exposure and outcome. If the IVs explain a greater variance in the exposure than that in the outcome, then the identified causal association could be considered directionally credible [30]. Furthermore, we calculated F statistics to evaluate weak instrument bias [31]. An F -value less than 10 indicated a weak instrument and was excluded.

All statistical analyses were performed using the R packages: two-sample MR [17] and MR-PRESSO [29].

Heterogeneity

We performed a test for heterogeneity using Cochran's Q statistics and the two-sample MR package between instruments. A Q larger than the number of instruments minus one provides evidence for heterogeneity and invalid instruments, or Q statistics significant at a p -value < 0.05 can imply the presence of heterogeneity [32, 33].

Reverse MR analysis

To explore whether cancers have any causal impact on the identified significant bacterial genera, we also performed a reverse MR analysis (i.e. cancers as the exposure and the identified causal bacterial genus as the outcome) using SNPs that are associated with cancers as IVs. We used the MR Steiger directionality test [30] to

examine whether exposure was directionally causal for the outcome.

Results

SNP selection

First, we identified 91, 307, 289, 310, and 397 SNPs associated with gut microbiota at the phylum, class, order, family, and genus levels, respectively, at a significance level of $p < 5 \times 10^{-8}$. We identified 228, 499,488, 811, and 1374 SNPs at the phylum, class, order, family, and genus levels, respectively, at a significance level of $p < 1 \times 10^{-6}$. After a series of quality control steps, 277 (genome-wide statistical significance threshold, $p < 1 \times 10^{-6}$) and 23 (genome-wide statistical significance threshold, $p < 5 \times 10^{-8}$) SNPs were selected as IVs.

The F statistics of the IVs were all largely > 10 (Additional file 2: Table. S2), indicating no evidence of weak instrument bias, and no evidence of pleiotropic effects was detected by the MR-PRESSO global test ($p > 0.05$). Eventually, after removing pleiotropic SNPs identified by the MR-PRESSO outlier test and the MR-Egger regression, there was no evidence of horizontal pleiotropy of the IVs (both MR-PRESSO global test $p > 0.05$ and MR-Egger regression $p > 0.05$).

Bacterial genera containing multiple SNPs were tested using the four MR methods to consider multiple testing corrections. In the set of SNPs less than the genome-wide statistical significance threshold (5×10^{-8}) that served as IVs, the significance threshold for various taxa levels was set to the following: phylum $p = 5 \times 10^{-2}$ (0.05/1), class $p = 5 \times 10^{-2}$ (0.05/1), order $p = 2.5 \times 10^{-2}$ (0.05/2), family $p = 1.25 \times 10^{-2}$ (0.05/4), and genus $p = 4.54 \times 10^{-3}$ (0.05/11). In the set of SNPs less than the genome-wide statistical significance threshold (1×10^{-6}) that served as IVs, the significance threshold for various taxa levels was set to the following: phylum $p = 5.55 \times 10^{-3}$ (0.05/9), class $p = 3.33 \times 10^{-3}$ (0.05/15), order $p = 2.5 \times 10^{-2}$ (0.05/15), family $p = 2.08 \times 10^{-3}$ (0.05/24), and genus $p = 6.25 \times 10^{-4}$ (0.05/80).

Causal effects of gut microbiota on the development of eight cancer types

Breast cancer

In the set of IVs ($p < 5 \times 10^{-8}$), we found that the phylum *Actinobacteria* (odds ratio (OR) = 1.011, 95% CI = 1.001–1.020, $p = 1.75 \times 10^{-2}$, Wald ratio) was causally associated with breast cancer, and class *Actinobacteria* was causally associated with patients with breast cancer; the causal association between class *Actinobacteria* and breast cancer was identified in the Neale Lab (OR = 1.010, 95% CI = 1.003–1.018, $p = 5.62 \times 10^{-3}$, Wald ratio), UK Biobank (OR = 1.012, 95% CI = 1.001–1.022, $p = 2.58 \times 10^{-2}$, IVW), and MRC-IEU (OR = 1.006, 95%

CI = 1.000–1.012, $p = 3.32 \times 10^{-2}$, Wald ratio). In addition, the genus *Ruminococcaceae* UCG013 was also causally associated with breast cancer (OR = 0.983, 95% CI = 0.972–0.994, $p = 4.35 \times 10^{-3}$, Wald ratio). Surprisingly, the family *Bifidobacteriaceae* and order *Bifidobacteriales* were also causally associated with breast cancer (OR = 1.010, 95% CI = 1.002–1.017, $p = 5.62 \times 10^{-3}$, Wald ratio); therefore, we performed an MR analysis in the UK Biobank database, which showed a similar result (OR = 1.009, 95% CI = 1.000–1.018, $p = 3.57 \times 10^{-2}$, Wald ratio) (Table 1, Fig. 2).

We identified more gut microbiota related to breast cancer in the other set of IVs ($p < 1 \times 10^{-6}$), and we also found that the genus *Ruminococcus gnavus* was causally associated with breast cancer (OR = 1.466, 95% CI = 1.251–1.718, $p = 2.15 \times 10^{-6}$, IVW), especially ER⁺ breast cancer (OR = 1.549, 95% CI = 1.285–1.866, $p = 4.27 \times 10^{-6}$, IVW); the genus *Oscillibacter* was causally associated with ER[–] breast cancer (OR = 2.045, 95% CI = 1.393–3.002, $p = 2.58 \times 10^{-4}$, IVW) (Table 2, Fig. 3).

Lung cancer

In the set of IVs ($p < 5 \times 10^{-8}$), we found that the phylum *Actinobacteria* (OR = 1.005, 95% CI = 1.000–1.010, $p = 2.42 \times 10^{-2}$, Wald ratio) and class *Actinobacteria* (OR = 1.004, 95% CI = 1.000–1.008, $p = 3.12 \times 10^{-2}$, Wald ratio) were causally associated with lung cancer. The genus *Tyzzereella3* was causally associated with lung adenocarcinoma (OR = 4.486, 95% CI = 1.641–12.263, $p = 3.43 \times 10^{-3}$, Wald ratio) (Table 1, Fig. 2).

We identified more gut microbiota related to lung cancer in the other set of IVs ($p < 1 \times 10^{-6}$), and we found that the order *Lactobacillales* was causally associated with squamous cell lung cancer (OR = 3.181, 95% CI = 1.517–6.666, $p = 2.17 \times 10^{-3}$, IVW) and small cell lung cancer (OR = 68.83, 95% CI = 4.182–1132.79, $p = 3.06 \times 10^{-3}$, IVW), while the order *Burkholderiales* (OR = 0.553, 95% CI = 0.375–0.815, $p = 2.75 \times 10^{-3}$, IVW) was causally associated with lung cancer (Table 2, Fig. 4).

Colorectal cancer

In the set of IVs ($p < 5 \times 10^{-8}$), we found that the genus *Tyzzereella3* (OR = 0.991, 95% CI = 0.987–0.996, $p = 5.43 \times 10^{-4}$, Wald ratio) and the genus *Ruminococcustorquesgroup* (OR = 1.014, 95% CI = 1.005–1.023, $p = 1.63 \times 10^{-3}$, Wald ratio) were causally associated with colorectal cancer (Table 1, Fig. 2).

In determining more gut microbiota related to colorectal cancer in the other set of IVs ($p < 1 \times 10^{-6}$), we found that the order *Verrucomicrobiales*, class *Verrucomicrobiae*, and family *Verrucomicrobiaceae* (OR = 1.013, 95% CI = 1.004–1.021, $p = 1.98 \times 10^{-3}$, Wald ratio) were causally associated with colorectal cancer. In addition,

the order *Desulfovibrionales* (OR = 1.015, 95% CI = 1.005–1.025, $p = 2.98 \times 10^{-3}$, Wald ratio) and class *Delataproteobacteria* (OR = 1.015, 95% CI = 1.005–1.026, $p = 2.98 \times 10^{-3}$, Wald ratio) were causally associated with colorectal cancer (Table 2, Fig. 4).

Prostate cancer

In the set of IVs ($p < 5 \times 10^{-8}$), we found that the genus *Ruminococcustorquesgroup* (OR = 0.350, 95% CI = 0.171–0.718, $p = 4.21 \times 10^{-3}$, Wald ratio) was causally associated with prostate cancer (Table 1, Fig. 5).

In identifying more gut microbiota related to prostate cancer in the other set of IVs ($p < 1 \times 10^{-6}$), we found the class *Verrucomicrobiae*, family *Verrucomicrobiaceae*, order *Verrucomicrobiales* (OR = 0.964, 95% CI = 0.943–0.985, $p = 8.72 \times 10^{-4}$, Wald ratio), genus *Terrisporobacter* (OR = 0.381, 95% CI = 0.237–0.612, $p = 6.78 \times 10^{-5}$, Wald ratio), genus *Roseburia* (OR = 1.727, 95% CI = 1.276–2.337, $p = 3.98 \times 10^{-4}$, Wald ratio), and class *Alphaproteobacteria* (OR = 1.289, 95% CI = 1.104–1.505, $p = 1.28 \times 10^{-3}$, Wald ratio) to be causally associated with prostate cancer (Table 2, Fig. 4).

Gastric cancer

In the set of IVs ($p < 5 \times 10^{-8}$), we found that the family *Peptostreptococcaceae* (OR = 12.516, 95% CI = 2.049–76.43, $p = 6.19 \times 10^{-3}$, Wald ratio) was causally associated with gastric cancer (Table 1, Fig. 5).

However, in identifying more gut microbiota related to gastric cancer in the other set of IVs ($p < 1 \times 10^{-6}$), we found no genetic liability to gut microbiota that was causally associated with gastric cancer after the Bonferroni test (Table 2, Fig. 4).

Head and neck cancer

In the set of IVs ($p < 5 \times 10^{-8}$), we found that the order *Gastranaerophilales* was causally associated with oropharyngeal cancer (OR = 0.277, 95% CI = 0.094–0.811, $p = 1.92 \times 10^{-2}$, Wald ratio), and the phylum *Actinobacteria* (OR = 0.041, 95% CI = 0.002–0.676, $p = 2.53 \times 10^{-2}$, Wald ratio) and class *Actinobacteria* (OR = 0.131, 95% CI = 0.019–0.910, $p = 3.99 \times 10^{-2}$, Wald ratio) were causally associated with oral cavity cancer (Table 1, Fig. 5).

In identifying more gut microbiota related to head and neck cancer in the other set of IVs ($p < 1 \times 10^{-6}$), we found that the genus *Ruminococcus1* was causally associated with head and neck cancer (OR = 1.008, 95% CI = 1.004–1.013, $p = 1.98 \times 10^{-4}$, Wald ratio), especially oral and oropharyngeal cancers (OR = 1.007, 95% CI = 1.003–1.011, $p = 3.71 \times 10^{-4}$, Wald ratio) (Table 2, Fig. 4).

Table 1 Mendelian randomisation (MR) results of causal effects between gut microbiome and cancer risk ($P < 5 \times 10^{-8}$)

Gut microbiota (exposure)	Cancer type (outcome)	Consortium	Method	Number of SNPs	β	SE	p-value	OR	95% CI	Correct causal direction	Steiger p-value
Phylum Actinobacteria	Breast cancer	Neale Lab	Wald ratio	1	1.11×10^{-2}	4.66×10^{-3}	1.75×10^{-2}	1.01	1.001–1.020	True	6.36×10^{-24}
Class Actinobacteria	Breast cancer	Neale Lab	Wald ratio	1	1.05×10^{-2}	3.78×10^{-3}	5.62×10^{-3}	1.01	1.003–1.018	True	1.64×10^{-33}
Class Actinobacteria	Breast cancer	UK Biobank	IWV	2	1.21×10^{-2}	5.41×10^{-3}	2.58×10^{-2}	1.01	1.001–1.022	True	2.09×10^{-17}
Class Actinobacteria	Breast cancer	MRC-IEU	Wald ratio	1	6.55×10^{-3}	3.08×10^{-3}	3.32×10^{-2}	1.01	1.000–1.012	True	3.04×10^{-17}
Family Bifidobacteriaceae	Breast cancer	Neale Lab	Wald ratio	1	9.97×10^{-3}	3.60×10^{-3}	5.62×10^{-3}	1.01	1.002–1.017	True	7.79×10^{-46}
Order Bifidobacteriales	Breast cancer	Neale Lab	Wald ratio	1	9.97×10^{-3}	3.60×10^{-3}	5.62×10^{-3}	1.01	1.002–1.017	True	7.79×10^{-46}
Genus <i>Ruminococcaceae</i>	Breast cancer	Neale Lab	Wald ratio	1	-1.63×10^{-2}	5.70×10^{-3}	4.35×10^{-3}	0.98	0.972–0.994	True	1.12×10^{-14}
UCG013											
Phylum Actinobacteria	Lung cancer	UK Biobank	Wald ratio	1	5.49×10^{-3}	2.44×10^{-3}	2.42×10^{-2}	1.01	1.000–1.010	True	1.06×10^{-12}
Class Actinobacteria	Lung cancer	UK Biobank	Wald ratio	1	4.23×10^{-3}	1.97×10^{-3}	3.12×10^{-2}	1.00	1.000–1.008	True	7.06×10^{-18}
Genus <i>Tyzzerella3</i>	Lung adenocarcinoma	FINNGEN	Wald ratio	1	1.50	0.51	3.43×10^{-3}	4.49	1.641–12.263	True	3.12×10^{-4}
Genus <i>Tyzzerella3</i>	Colorectal cancer	UK Biobank	Wald ratio	1	-8.05×10^{-3}	2.33×10^{-3}	5.43×10^{-4}	0.99	0.987–0.996	True	2.95×10^{-7}
Genus <i>Ruminococcus-torques</i>	Colorectal cancer	UK Biobank	Wald ratio	1	1.44×10^{-2}	4.58×10^{-3}	1.63×10^{-3}	1.01	1.005–1.023	True	1.01×10^{-6}
Genus <i>Ruminococcus-torques</i>	Prostate cancer	FINNGEN	Wald ratio	1	-1.05	0.37	4.21×10^{-3}	0.35	0.171–0.718	True	5.24×10^{-7}
Family Peptostreptococaceae	Gastric cancer	FINNGEN	Wald ratio	1	2.53	0.92	6.19×10^{-3}	12.52	2.049–76.43	True	4.21×10^{-6}
Order Gastranaerophilales	Oropharyngeal cancer	Oncoarray oral cavity and oropharyngeal cancer	Wald ratio	1	-1.28	0.55	1.92×10^{-2}	0.28	0.094–0.811	True	2.23×10^{-7}
Phylum Actinobacteria	Oral cavity cancer	Oncoarray oral cavity and oropharyngeal cancer	Wald ratio	1	-3.17	1.42	2.53×10^{-2}	0.04	0.002–0.676	True	1.52×10^{-7}
Class Actinobacteria	Oral cavity cancer	Oncoarray oral cavity and oropharyngeal cancer	Wald ratio	1	-2.03	0.99	3.99×10^{-2}	0.13	0.019–0.910	True	2.01×10^{-18}

OR Odds ratio, SE Standard error, CI Confidence interval

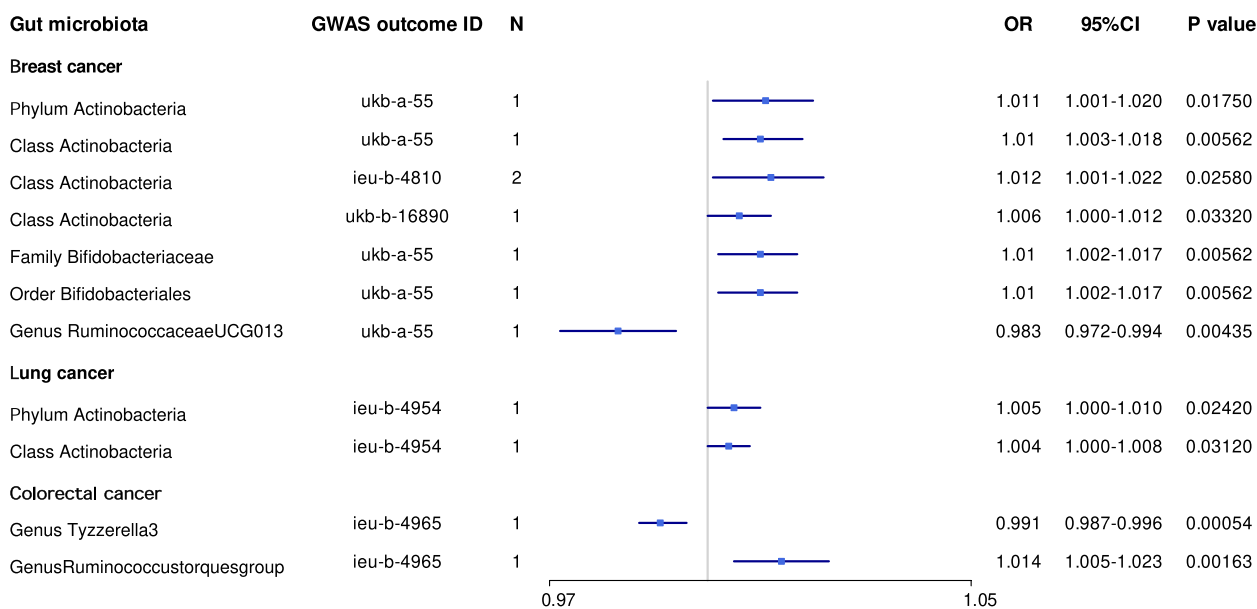


Fig. 2 Mendelian randomisation results of causal effects between gut microbiome and cancer risk ($p < 5 \times 10^{-8}$)

Endometrial cancer

In the set of IVs ($p < 5 \times 10^{-8}$), we found that genetic liability to the gut microbiota was not causally associated with endometrial cancer, as per the Bonferroni test.

When determining more gut microbiota related to endometrial cancer in the other set of IVs ($p < 1 \times 10^{-6}$), we found that the class *Gammaproteobacteria* was causally associated with endometrial cancer (OR = 0.9936, 95% CI = 0.989–0.997, $p = 8.43 \times 10^{-4}$, Wald ratio) (Table 2, Fig. 4).

Ovarian cancer

In the set of IVs ($p < 5 \times 10^{-8}$), we did not find any genetic liability to the gut microbiota that was causally associated with ovarian cancer after the Bonferroni test.

When identifying more gut microbiota related to ovarian cancer in the other set of IVs ($p < 1 \times 10^{-6}$), we found that the genus *Ruminiclostridium 6* was causally associated with a low malignant potential in serous ovarian cancer (OR = 11.869, 95% CI = 3.178–44.327, $p = 2.33 \times 10^{-4}$, Wald ratio) (Table 2, Fig. 4).

Potential causal associations between the gut microbiota and cancers

Moreover, we found some potential causal associations between the gut microbiota and cancers. Those results were found in at least two different datasets in the set

of IVs ($p < 1 \times 10^{-6}$), with $p < 0.05$, but did not pass the Bonferroni test. Detailed information is provided in Additional file 3: Table S3.

The genus *Ruminococcaceae* UCG013, genus *Ruminococcaceae* NK4A214 group, genus *Oxalobacter*, genus *Holdemanella*, genus *Eubacterium eligens* group, and class *Alphaproteobacteria* were highly associated with breast cancer. The order *Selenomonadales*, genus *Turicibacter*, genus *Ruminococcus1*, genus *Ruminococcaceae* UCG014, genus *Odoribacter*, genus *Dorea*, and class *Negativicutes* were highly associated with head and neck cancer. The genus *Eubacterium coprostanoligenes* group, genus *Parasutterella*, genus *Ruminococcaceae* UCG003, and order *Lactobacillales* were highly associated with lung cancer. The family *Verrucomicrobiaceae*, class *Verrucomicrobiae*, order *Verrucomicrobiales*, genus *Terriporobacter*, genus *Roseburia*, and genus *Akkermansia* were highly associated with prostate cancer. The genus *Adlercreutzia* was highly associated with prostate and endometrial cancer.

Sensitivity analyses

The MR-Egger, weighted mode, simple mode, and weighted median methods yielded similar causal estimates for magnitude and direction. We found no evidence of horizontal pleiotropy for gut microbiota in cancers with $p > 0.05$ when using the MR-Egger regression intercept approach. MR-PRESSO analysis revealed no outliers in the results. In addition, the results of the Cochrane Q statistics showed no significant heterogeneity ($p > 0.05$).

Table 2 MR results of causal effects between gut microbiome and cancer risk ($p < 1 \times 10^{-6}$)

Gut microbiota (exposure)	Cancer type (outcome)	Consortium	Method	Number of SNPs	β	SE	p-value	OR	95% CI	Correct causal direction	Steiger p-value
Genus <i>Ruminococcusgnavus</i> -group	Breast cancer	BCAC	IWW	2	0.38	8.08 × 10 ⁻²	2.16 × 10 ⁻⁶	1.47	1.25–1.73	True	4.13 × 10 ⁻⁷
Genus <i>Ruminococcusgnavus</i> -group	ER+ breast cancer	BCAC	IWW	2	0.44	9.52 × 10 ⁻²	4.27 × 10 ⁻⁶	1.55	1.29–1.87	True	1.57 × 10 ⁻⁶
Genus <i>Oscillibacter</i>	ER– breast cancer	BCAC	IWW	2	0.72	0.20	2.59 × 10 ⁻⁴	2.05	1.39–3.00	True	1.17 × 10 ⁻⁵
Class Gammaproteobacteria	Endometrial cancer	MRC-IEU	Wald ratio	1	-6.38 × 10 ⁻³	1.91 × 10 ⁻³	8.44 × 10 ⁻⁴	0.99	0.99–1.00	True	2.22 × 10 ⁻⁵
Genus <i>Ruminococcus1</i>	Head and neck cancer	UK Biobank	Wald ratio	1	8.69 × 10 ⁻³	2.34 × 10 ⁻³	1.98 × 10 ⁻⁴	1.01	1.00–1.01	True	9.17 × 10 ⁻¹²
Genus <i>Ruminococcus1</i>	Oral and oropharyngeal cancer	UK Biobank	Wald ratio	1	7.25 × 10 ⁻³	2.04 × 10 ⁻³	3.71 × 10 ⁻⁴	1.01	1.00–1.01	True	8.76 × 10 ⁻¹²
Genus <i>Ruminiclostridium6</i>	Serous ovarian cancer: low malignant potential	OCAC	Wald ratio	1	2.47	0.67	2.33 × 10 ⁻⁴	11.87	3.18–44.33	True	2.83 × 10 ⁻⁹
Order Lactobacillales	Squamous cell lung cancer	ILCCO	IWW	2	1.16	0.38	2.17 × 10 ⁻³	3.18	1.52–6.67	True	8.74 × 10 ⁻⁶
Order Burkholderiales	Lung cancer	ILCCO	IWW	3	-0.59	0.20	2.75 × 10 ⁻³	0.55	0.38–0.82	True	7.49 × 10 ⁻⁹
Order Lactobacillales	Small cell Lung cancer	FINNGEN	IWW	2	4.23	1.43	3.06 × 10 ⁻³	68.83	4.18–1132.80	True	6.08 × 10 ⁻⁶
Genus <i>Terrisporobacter</i>	Prostate cancer	FINNGEN	Wald ratio	1	-0.96	0.24	6.79 × 10 ⁻⁵	0.38	0.24–0.61	True	4.56 × 10 ⁻⁹
Genus <i>Roseburia</i>	Prostate cancer	PRACTICA-L	Wald ratio	1	0.55	0.15	3.89 × 10 ⁻⁴	1.73	1.28–2.34	True	3.08 × 10 ⁻¹³
Class Verrucomicrobiae	Prostate cancer	UK Biobank	Wald ratio	1	-3.66 × 10 ⁻²	1.10 × 10 ⁻²	8.73 × 10 ⁻⁴	0.96	0.94–0.99	True	1.01 × 10 ⁻⁴
Family Verrucomicrobiaceae	Prostate cancer	UK Biobank	Wald ratio	1	-3.66 × 10 ⁻²	1.10 × 10 ⁻²	8.73 × 10 ⁻⁴	0.96	0.94–0.99	True	1.01 × 10 ⁻⁴
Order Verrucomicrobiales	Prostate cancer	UK Biobank	Wald ratio	1	-3.66 × 10 ⁻²	1.10 × 10 ⁻²	8.73 × 10 ⁻⁴	0.96	0.94–0.99	True	1.01 × 10 ⁻⁴
Class Alphaproteobacteria	Prostate cancer	PRACTICA-L	IWW	2	0.25	7.90 × 10 ⁻²	1.29 × 10 ⁻³	1.29	1.10–1.51	True	7.72 × 10 ⁻¹¹
Order Verrucomicrobiales	Colorectal cancer	UK Biobank	Wald ratio	1	1.32 × 10 ⁻²	4.27 × 10 ⁻³	1.98 × 10 ⁻³	1.01	1.00–1.02	True	1.80 × 10 ⁻⁵
Order Desulfotribionales	Colorectal cancer	UK Biobank	Wald ratio	1	1.53 × 10 ⁻²	5.15 × 10 ⁻³	2.98 × 10 ⁻³	1.02	1.00–1.03	True	6.97 × 10 ⁻⁶
Family Verrucomicrobiaceae	Colorectal cancer	UK Biobank	Wald ratio	1	1.32 × 10 ⁻²	4.27 × 10 ⁻³	1.98 × 10 ⁻³	1.01	1.00–1.02	True	1.80 × 10 ⁻⁵
Class Verrucomicrobiae	Colorectal cancer	UK Biobank	Wald ratio	1	1.32 × 10 ⁻²	4.27 × 10 ⁻³	1.98 × 10 ⁻³	1.01	1.00–1.02	True	1.80 × 10 ⁻⁵
Class Deltaproteobacteria	Colorectal cancer	UK Biobank	Wald ratio	1	1.55 × 10 ⁻²	5.15 × 10 ⁻³	2.98 × 10 ⁻³	1.02	1.02–1.03	True	9.00 × 10 ⁻⁶

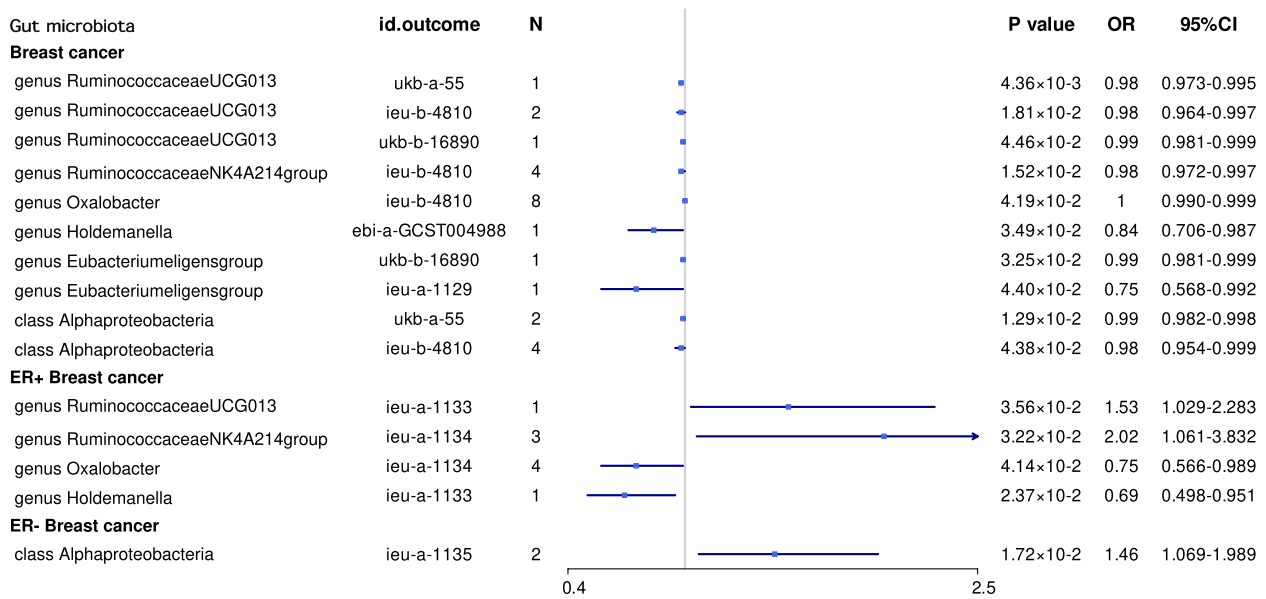


Fig. 3 Mendelian randomisation results of causal effects between gut microbiome and breast cancer risk ($p < 1 \times 10^{-6}$)

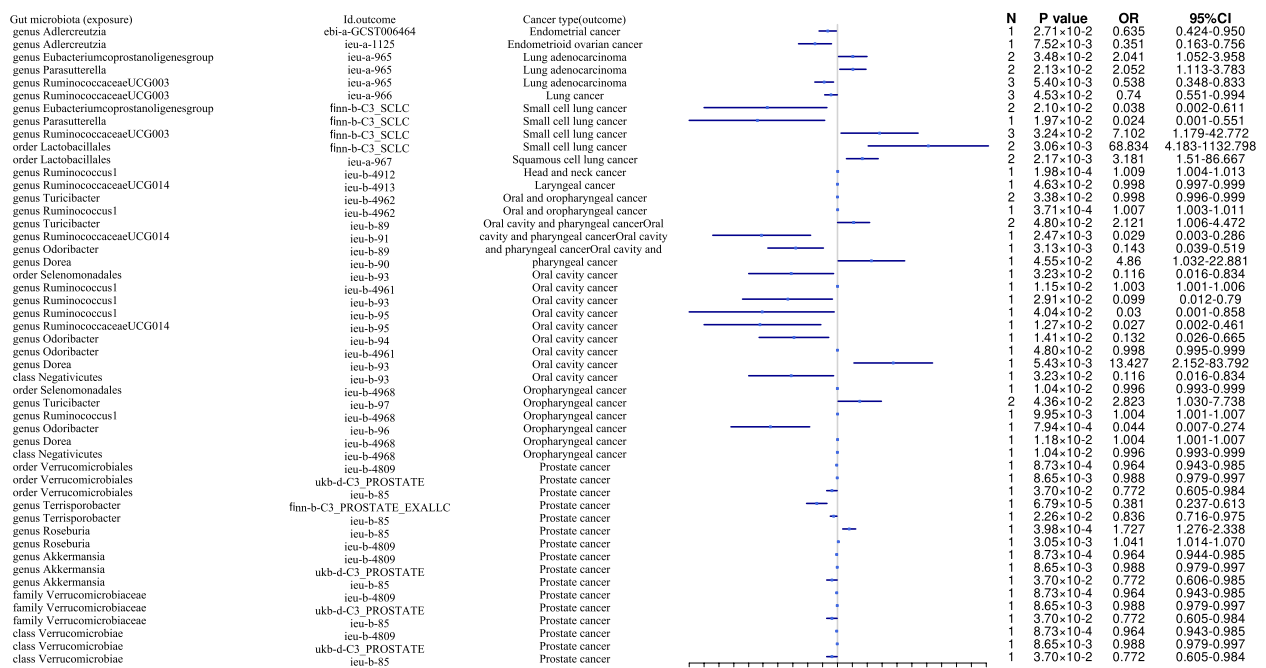


Fig. 4 Mendelian randomisation results of causal effects between gut microbiome and other cancer risks ($p < 1 \times 10^{-6}$)

Bi-directional causal effects between gut microbiota and cancer risk

To evaluate any reverse causation effects, we used cancer as exposure and gut microbiota as outcome, and 211 SNPs associated with cancers extracted from previous GWAS were used as IVs (Table 3). Based on the Bonferroni test, the significance threshold for various taxa levels

was set to $p = 6.25 \times 10^{-3}$ (0.05/8), and we found that lung adenocarcinoma was causally associated with the genus *Tyzzereella3* ($p = 1.02 \times 10^{-3}$, IVW), which indicates a bi-directional causal effect between lung adenocarcinoma and the genus *Tyzzereella3*. A summary network for a better understanding of the relationship between gut microbiota and cancer is presented in Fig. 6.

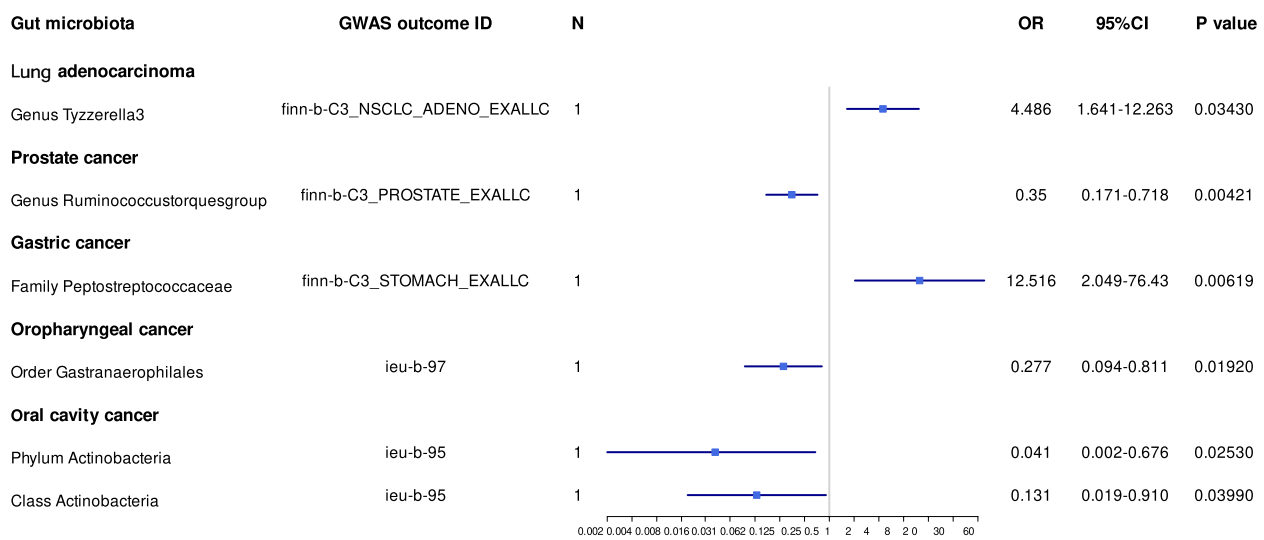


Fig. 5 Mendelian randomisation results of causal effects between gut microbiome and cancer risk (continue) ($p < 5 \times 10^{-8}$)

Table 3 Bi-directional MR results of the causal effects between gut microbiome and cancer risk ($p < 5 \times 10^{-8}$)

Cancer type (exposure)	Gut microbiota (outcome)	Method	Number of SNPs	β	SE	p -value	Correct causal direction	Steiger p -value
ProstateCancer	Genus <i>Roseburia</i>	IVW	23	-0.86	0.37	1.97×10^{-2}	True	1.49×10^{-12}
OropharyngealCancer	Genus <i>Turicibacter</i>	Wald ratio	1	79.16	33.11	1.68×10^{-2}	True	8.69×10^{-6}
Non-SmallCell Lung Cancer	Genus <i>Tyzzzeria3</i>	Wald ratio	1	-0.16	0.06	1.14×10^{-2}	True	4.48×10^{-5}
Non-SmallCell Lung Cancer	Genus <i>Parasutterella</i>	Wald ratio	1	0.11	0.04	8.04×10^{-3}	True	3.19×10^{-7}
Non-SmallCell Lung Cancer	Genus <i>Tyzzzeria3</i>	Wald ratio	1	-0.16	0.06	1.14×10^{-2}	True	4.48×10^{-5}
Non-SmallCell Lung Cancer	Genus <i>Parasutterella</i>	Wald ratio	1	0.12	0.04	8.03×10^{-3}	True	3.19×10^{-7}
LungCancer	Order <i>Lactobacillales</i>	IVW	2	-14.06	6.52	3.10×10^{-3}	False	0.46
LungCancer	Order <i>Lactobacillales</i>	IVW	2	-14.42	6.51	2.67×10^{-2}	False	0.48
LungAdenocarcinoma	Genus <i>Tyzzzeria3</i>	IVW	2	-0.17	0.05	1.02×10^{-3}	True	1.55×10^{-9}
ER+Breast Cancer	Genus <i>Eubacterium eli-Gens Group</i>	IVW	7	-0.04	0.02	4.80×10^{-2}	True	1.05×10^{-65}
BreastCancer	Genus <i>Oxalobacter</i>	IVW	47	0.08	0.04	2.51×10^{-3}	True	6.01×10^{-66}

Discussion

To the best of our knowledge, this is the first MR study to investigate whether gut microbiota is causally associated with cancers, which we think is a longitudinal microbiome study antecedent to human cancer. Robustly associated gene variants were identified in the largest GWAS of the gut microbiota. Based on comprehensive genetic data from over 450,000 European individuals, we found genetic liability to some gut microbiota causally associated with cancers. Surprisingly, the genetic liability to the family *Bifidobacteriaceae* and order *Bifidobacteriales* was causally associated with breast cancer. We also identified some gut microbiota that might be potential risk factors for cancer. These results could have implications for public health interventions aimed at reducing cancer risk.

A growing number of studies have found a possible link between the gut microbiota selected in our study and other cancers. For instance, *Ruminococcus* plays an important role in the digestion of resistant starch [34]. However, a previous study found that *Ruminococcus gnavus* is associated with Crohn’s disease [35] and *Ruminococcus gnavus* was identified as a signature taxon for patients with hepatocellular carcinoma infected with hepatitis B and/or hepatitis C viruses [36]. The abundance of members from *Peptostreptococcaceae* was increased in patients with intrahepatic cholangiocarcinoma (ICC) compared to that in patients with hepatocellular carcinoma or liver cirrhosis and healthy individuals. Patients with vascular invasion (VI) had a greater abundance of

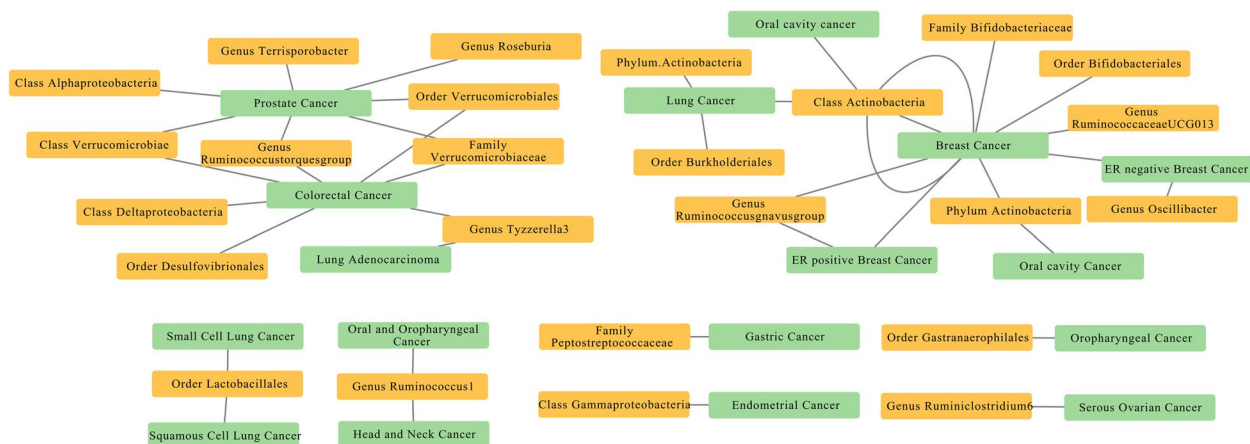


Fig. 6 The causal relationships between gut microbiota and cancers by Mendelian randomisation analysis

the family *Ruminococcaceae* than did patients with ICC without VI [37].

Actinobacteria, including 15 species of *Bifidobacterium*, decreased with age, and the overall richness or number of unique *Bifidobacterium* species present in an individual steadily declined throughout their life [38]. A previous study showed that Actinobacteria was present in a relatively high proportion of breast cancer tissue samples [39]. Meanwhile, some absolute numbers of *Bifidobacterium* were significantly different according to the clinical stages of cancer [40, 41], which suggests that the microbiome may be involved in the progression of breast cancer [41]. Recent studies have reported a higher abundance of *Bifidobacterium* in the tissues of patients with colorectal adenomas [42, 43] and advanced pancreatic cancer in mice [44], whereas other studies have reported that *Bifidobacterium* correlates with an increased anti-PD-L1 therapeutic response [45, 46]. One study showed that *Bifidobacteria* might be potential pathogens [47], which indicated that although *Bifidobacterium* is generally considered beneficial, specific species and strains of *Bifidobacterium* may have varying effects on human health [48]. As shown in our results, our MR results suggested that the phylum *Actinobacteria* and class *Actinobacteria* are risk factors for breast cancer and lung cancer but are protective factors for oral cavity cancer. The family *Bifidobacteriaceae* and order *Bifidobacteriales* are also risk factors for breast cancer, while the genus *Ruminococcaceae* UCG013 is a protective factor against breast cancer. The genus *Tyzzerella3* is a risk factor for lung adenocarcinoma, but a protective factor against colorectal cancer. The genus *Ruminococcustorquesgroup* is a risk factor for colorectal cancer, but a protective factor against prostate cancer. The family *Peptostreptococcaceae* is a risk factor for gastric cancer, and the order

Gastranaerophilales is a risk factor for oropharyngeal cancer. In summary, different species may have divergent effects on the tumour microenvironment [45].

Experimental models have suggested that gut microbiota can promote the induction and/or development of tumour formation through multiple mechanisms [49]. However, the exact mechanism by which the gut microbiota causes cancer has not been determined. Therefore, a mechanistic analysis of our results is required for further investigation.

Studies have determined whether the gut microbes are “beneficial” or “harmful” by comparing the relative abundance of gut microbiota between healthy people and patients. Engstrand and Graham suggested that the relative abundance of the dominant microbiota in the gut microbiota may not be a risk factor for cancers, but may represent a bystander effect [50] as well as a carcinogenic factor.

Many dietary components can influence cancer via targeting gut microbiota [51]. Nowadays, the prevalence of obesity is significantly increasing in developing countries such as China where people are adapting to high-fat diet [52, 53]. High-fat diet is dominated by carbohydrate and fat and lacks plant-based dietary fibre [54]. The consequence of this is that high-fat diet populations showed lower bacterial diversity compared with those of traditional rural population [55]. Obesity is correlated with excessive fat dietary intake. The positive association between obesity and cancers has been verified by several studies [56–58]. So apart from the genetic factors influencing on cancer risk, behaviour and lifestyle can also play an important role in cancer development. In the future, exploring the relationship of diet and cancer through gut microbiota may offer new insights to cancer treatment [59]. In consideration of the complex

relationships between diet, gut microbiota, and cancers [60, 61], more studies and mediation MR analysis are needed to discover the association and mechanism in detail [62].

A GWAS is unlikely to explain all the heritability of complex traits [63]. As linkage disequilibrium patterns vary across ethnic groups, it is not suitable for non-European populations in the past [63, 64]. With the development of a new generation of high-density arrays and the accumulation of more sequencing data from more diverse populations, this problem may have been improved [65]. Nowadays, clinical prediction by GWASs might also prove to be especially useful in small isolated populations where deleterious variants with strong effects have increased to a high frequency [66]. A previous study revealed that variants that are significantly correlated with each other tend to be in linkage disequilibrium or even form haplotypes [67]. Although we can find causal relations between variants and disease, it is difficult to identify causal variations from multiple variants located on the same haplotype [68]. Variable penetrance and variable expressivity are the common cause for the observation where individuals carrying the same variant display highly variable symptoms [69]. A case–control analysis of autism and cancer cohorts suggested that modified penetrance of coding variants by cis-regulatory variation contributes to disease risk [70]. Although knowledge of individual's genetic risk can improve readiness to adopt a healthier lifestyle, human behavior is complex [71, 72]. Both the environment and genes can influence disease symptoms. In addition to the separate effects of genotype and environmental factors, the effects of environmental factors on different individuals will be affected by genotype. Meanwhile, the role of genetic factors also depends on environmental influences [73, 74]. We used the MR approach to eliminate some confounders that are commonly observed in epidemiological studies. Moreover, our SNPs were strongly associated with gut microbiota and were compared with multiple cancer databases. Moreover, the reverse MR analysis and sensitivity analysis showed no pleiotropy or heterogeneity, which indicates that our results are statistically robust.

Nevertheless, our study had several limitations. First, while the majority of patients in the GWAS summary data used in our study were European, only a small number of the gut microbiota data were taken from other races, which may lead to bias estimates and affect universality. Second, the bacterial taxa were only analysed at the order or family level. If the GWASs had used more advanced shotgun metagenomic sequencing analyses, the results would be more specific and accurate. Third, due to our strict thresholds, many of the genetic liabilities

of the gut microbiota were excluded at the IV selection stage, which may result in some results being missed.

Recent research proposed that future research should take an integrative approach that uses multiple omics platforms to improve understanding of the pathogenesis of disease in the context of the complex interactions between genes and the environment over time [74].

Conclusions

In summary, we comprehensively assessed the causal association between the gut microbiota and a series of cancers. Our results suggest that there are four positive causal directions and one negative causal direction with breast cancer, three positive causal directions and one negative causal direction with lung cancer, two positive causal directions and four negative causal directions with prostate cancer, one positive causal direction with gastric cancer, one positive causal direction with ovarian cancer, one negative causal direction with endometrial cancer, six positive causal directions and one negative causal direction with colorectal cancer, and one positive causal direction and three negative causal directions with colorectal cancer. In addition, we found potential causal associations between the gut microbiota and cancer. This study may provide new insights into the mechanisms of gut microbiota-mediated cancer development.

Abbreviations

BCAC	Breast Cancer Association Consortium
CRC	Colorectal cancer
FMT	Faecal microbiota transplantation
GWAS	Genome-wide association study
ILCCO	International Lung Cancer Consortium
MR	Mendelian randomisation
MRC-IEU	Medical Research Council-Integrative Epidemiology Unit
PRACTICA-L	Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome
SNP	Single-nucleotide polymorphism

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12916-023-02761-6>.

Additional file 1: Table S1. Overview of the source of cancer data.

Additional file 2: Table S2. F statistics of instrumental variables (IVs).

Additional file 3: Table S3. The potential relationships between gut microbiota and cancer risk.

Acknowledgements

The authors thank the participants of all GWAS cohorts included in the present work and the investigators of the IEU Open GWAS project, MiBioGen, UK Biobank, BCAC, OCAC, MRC-IEU, Neale Lab, PRACTICA-L, ILCCO, FINNGEN and Oncoarray oral cavity and oropharyngeal cancer consortium for sharing the GWAS summary statistics.

Authors' contributions

Conceptualisation, YL; methodology, YL, SZ, HZ, and LT; verification of the underlying data, YL and SZ; writing—original draft, YL and YZ; writing—review

and editing, SZ and HZ. All the authors participated in planning, execution, and analysis and have read and approved the final submitted version.

Funding

This work was financially supported by the National Natural Science Foundation of China (Nos. 82203498), the Natural Science Foundation of Hunan Province (2022JJ40816), and the Youth Research Fund of XiangYa Hospital (2020Q05).

Availability of data and materials

The summary data of Neale Lab can be downloaded from the website <http://www.nealelab.is/uk-biobank/> (2022). The summary data of FINNGEN can be downloaded from the website https://www.finngen.fi/en/access_results (2022). The other datasets generated and/or analysed during the current study are publicly available and included in this published article and its supplementary information files.

Declarations

Ethics approval and consent to participate

Our analysis used publicly available genome-wide association study (GWAS) summary statistics. No new data were collected, and no new ethical approval was required.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 19 August 2022 Accepted: 30 January 2023

Published online: 21 February 2023

References

- Yachida S, Mizutani S, Shiroma H, Shiba S, Nakajima T, Sakamoto T, et al. Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nat Med*. 2019;25(6):968–76.
- Yonekura S, Terrisse S, Alves Costa Silva C, Lafarge A, Iebba V, Ferrere G, et al. Cancer induces a stress ileopathy depending on β -adrenergic receptors and promoting dysbiosis that contributes to carcinogenesis. *Cancer Discov*. 2022;12(4):1128–51.
- Malard F, Vekhoff A, Lapusan S, Isnard F, D'Incan-Corda E, Rey J, et al. Gut microbiota diversity after autologous fecal microbiota transfer in acute myeloid leukemia patients. *Nat Commun*. 2021;12(1):3084.
- Wan Y, Wang F, Yuan J, Li J, Jiang D, Zhang J, et al. Effects of dietary fat on gut microbiota and faecal metabolites, and their relationship with cardiometabolic risk factors: a 6-month randomised controlled-feeding trial. *Gut*. 2019;68(8):1417–29.
- Delannoy-Bruno O, Desai C, Raman AS, Chen RY, Hibberd MC, Cheng J, et al. Evaluating microbiome-directed fibre snacks in gnotobiotic mice and humans. *Nature*. 2021;595(7865):91–5.
- Fluhr L, Mor U, Kolodziejczyk AA, Dori-Bachash M, Leshem A, Itav S, et al. Gut microbiota modulates weight gain in mice after discontinued smoke exposure. *Nature*. 2021;600(7890):713–9.
- Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS discovery: biology, function, and translation. *Am J Hum Genet*. 2017;101(1):5–22.
- Sud A, Kinnersley B, Houlston RS. Genome-wide association studies of cancer: current insights and future perspectives. *Nat Rev Cancer*. 2017;17(11):692–704.
- Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol*. 2003;32(1):1–22.
- Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet*. 2014;23(R1):R89–98.
- Sekula P, Del Greco MF, Pattaro C, Köttgen A. Mendelian randomization as an approach to assess causality using observational data. *J Am Soc Nephrol*. 2016;27(11):3253–65.
- Wang J, Kurilshikov A, Radjabzadeh D, Turpin W, Croitoru K, Bonder MJ, et al. Meta-analysis of human genome-microbiome association studies: the MiBioGen consortium initiative. *Microbiome*. 2018;6(1):101. <https://doi.org/10.1186/s40168-018-0479-3>.
- Kurilshikov A, Medina-Gomez C, Bacigalupe R, Radjabzadeh D, Wang J, Demirkan A, et al. Large-scale association analyses identify host factors influencing human gut microbiome composition. *Nat Genet*. 2021;53(2):156–65. [https://mibiogen.gcc.rug.nl/\(2022\)](https://mibiogen.gcc.rug.nl/(2022)).
- Ting NL, Lau HC, Yu J. Cancer pharmacomicrobiomics: targeting microbiota to optimise cancer therapy outcomes. *Gut*. 2022;71(7):1412–25.
- Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12(3):e1001779. <https://doi.org/10.1371/journal.pmed.1001779>.
- Wang Y, McKay JD, Rafnar T, Wang Z, Timofeeva MN, Broderick P, et al. Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. *Nat Genet*. 2014;46(7):736–41. <https://doi.org/10.1038/ng.3002>.
- Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenotype. *Elife*. 2018;7. <https://doi.org/10.7554/elife.34408>.
- Schumacher FR, Al Olama AA, Berndt SI, Benlloch S, Ahmed M, Saunders EJ, et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat Genet*. 2018;50(7):928–36. <https://doi.org/10.1038/s41588-018-0142-8>.
- Batram T, Yousefi P, Crawford G, Prince C, Sheikhali Babaei M, Sharp G, et al. The EWAS Catalog: a database of epigenome-wide association studies. *Wellcome Open Res*. 2022;7:41. <https://doi.org/10.12688/wellcomeopenres.17598.2>.
- Phelan CM, Kuchenbaecker KB, Tyrer JP, Kar SP, Lawrenson K, Winham SJ, et al. Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. *Nat Genet*. 2017;49(5):680–91. <https://doi.org/10.1038/ng.3826>.
- Shete S, Liu H, Wang J, Yu R, Sturgis EM, Li G, et al. A genome-wide association study identifies two novel susceptible regions for squamous cell carcinoma of the head and neck. *Cancer Res*. 2020;80(12):2451–60. <https://doi.org/10.1158/0008-5472.can-19-2360>.
- Zhang H, Ahearn TU, Lecarpentier J, Barnes D, Beesley J, Qi G, et al. Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. *Nat Genet*. 2020;52(6):572–81. <https://doi.org/10.1038/s41588-020-0609-2>.
- Kurki MI, Karjalainen J, Palta P, Sipilä TP, Kristiansson K, Donner KM, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature*. 2023;613(7944):508–18. <https://doi.org/10.1038/s41586-022-05473-8>.
- Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization. *Stat Methods Med Res*. 2017;26(5):2333–55.
- Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol*. 2013;37(7):658–65.
- Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol*. 2017;46(6):1985–98.
- Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol*. 2015;44(2):512–25.
- Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol*. 2016;40(4):304–14.
- Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet*. 2018;50(5):693–8.
- Hemani G, Tilling K, Davey SG. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet*. 2017;13(11):e1007081.
- Burgess S, Thompson SG. Bias in causal estimates from Mendelian randomization studies with weak instruments. *Stat Med*. 2011;30(11):1312–23.

32. Greco MF, Minelli C, Sheehan NA, Thompson JR. Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. *Stat Med*. 2015;34(21):2926–40.
33. Bowden J, Del Greco MF, Minelli C, Zhao Q, Lawlor DA, Sheehan NA, et al. Improving the accuracy of two-sample summary-data Mendelian randomization: moving beyond the NOME assumption. *Int J Epidemiol*. 2019;48(3):728–42.
34. Ze X, Ben David Y, Laverde-Gomez JA, Dassa B, Sheridan PO, Duncan SH, et al. Unique organization of extracellular amylases into amylozymes in the resistant starch-utilizing human colonic Firmicutes Bacterium *Ruminococcus bromii*. *mBio*. 2015;6(5):e01058–15.
35. Henke MT, Kenny DJ, Cassilly CD, Vlamakis H, Xavier RJ, Clardy J. *Ruminococcus gnavus*, a member of the human gut microbiome associated with Crohn's disease, produces an inflammatory polysaccharide. *Proc Natl Acad Sci U S A*. 2019;116(26):12672–7.
36. Komiya S, Yamada T, Takemura N, Kokudo N, Hase K, Kawamura YI. Profiling of tumour-associated microbiota in human hepatocellular carcinoma. *Sci Rep*. 2021;11(1):10589.
37. Jia X, Lu S, Zeng Z, Liu Q, Dong Z, Chen Y, et al. Characterization of gut microbiota, bile acid metabolism, and cytokines in intrahepatic cholangiocarcinoma. *Hepatology*. 2020;71(3):893–906.
38. Byrd AL, Liu M, Fujimura KE, Lyalina S, Nagarkar DR, Charbit B, et al. Gut microbiome stability and dynamics in healthy donors and patients with non-gastrointestinal cancers. *J Exp Med*. 2021;218(1):e20200606.
39. Dohlmán AB, Argüjio Mendoza D, Ding S, Gao M, Dressman H, Iliev ID, et al. The cancer microbiome atlas: a pan-cancer comparative analysis to distinguish tissue-resident microbiota from contaminants. *Cell Host Microbe*. 2021;29(2):281–98.e5.
40. Bard J-M, Luu HT, Dravet F, Michel C, Moyon T, Pagniez A, Nazih H, Bobin-Dubigeon C. Relationship Between Intestinal Microbiota and Clinical Characteristics of Patients with Early Stage Breast Cancer. *The FASEB Journal*. 2015;29:914.2. https://doi.org/10.1096/fasebj.29.1_supplement.914.2.
41. Wu AH, Tseng C, Vigen C, Yu Y, Cozen W, Garcia AA, et al. Gut microbiome associations with breast cancer risk factors and tumor characteristics: a pilot study. *Breast Cancer Res Treat*. 2020;182(2):451–63.
42. Nugent JL, McCoy AN, Addamo CJ, Jia W, Sandler RS, Keku TO. Altered tissue metabolites correlate with microbial dysbiosis in colorectal adenomas. *J Proteome Res*. 2014;13(4):1921–9.
43. Sanapareddy N, Legge RM, Jovov B, McCoy A, Burcal L, Araujo-Perez F, et al. Increased rectal microbial richness is associated with the presence of colorectal adenomas in humans. *ISME J*. 2012;6(10):1858–68.
44. Pushalkar S, Hundeyin M, Daley D, Zambirinis CP, Kurz E, Mishra A, et al. The pancreatic cancer microbiome promotes oncogenesis by induction of innate and adaptive immune suppression. *Cancer Discov*. 2018;8(4):403–16.
45. Sivan A, Corrales L, Hubert N, Williams JB, Aquino-Michaels K, Earley ZM, et al. Commensal *Bifidobacterium* promotes antitumor immunity and facilitates anti-PD-L1 efficacy. *Science*. 2015;350(6264):1084–9.
46. Routy B, Le Chatelier E, Derosa L, Duong CPM, Alou MT, Daillère R, et al. Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. *Science*. 2018;359(6371):91–7.
47. Mahlen SD, Clarridge JE 3rd. Site and clinical significance of *Alloscardovia omnicolens* and *Bifidobacterium* species isolated in the clinical laboratory. *J Clin Microbiol*. 2009;47(10):3289–93.
48. Xiang K, Wang P, Xu Z, Hu YQ, He YS, Chen Y, et al. Causal effects of gut microbiome on systemic lupus erythematosus: a two-sample Mendelian randomization study. *Front Immunol*. 2021;12:667097.
49. Knippel RJ, Drewes JL, Sears CL. The cancer microbiome: recent highlights and knowledge gaps. *Cancer Discov*. 2021;11(10):2378–95.
50. Engstrand L, Graham DY. Microbiome and gastric cancer. *Dig Dis Sci*. 2020;65(3):865–73.
51. Tao J, Li S, Gan RY, Zhao CN, Meng X, Li HB. Targeting gut microbiota with dietary components on cancer: effects and potential mechanisms of action. *Crit Rev Food Sci Nutr*. 2020;60(6):1025–37.
52. Yang J, Yu J. The association of diet, gut microbiota and colorectal cancer: what we eat may imply what we get. *Protein Cell*. 2018;9(5):474–87.
53. Goss PE, Strasser-Weippl K, Lee-Bychkovsky BL, Fan L, Li J, Chavarri-Guerra Y, et al. Challenges to effective cancer control in China, India, and Russia. *Lancet Oncol*. 2014;15(5):489–538.
54. McGill CR, Fulgoni VL 3rd, Devareddy L. Ten-year trends in fiber and whole grain intakes and food sources for the United States population: National Health and Nutrition Examination Survey 2001–2010. *Nutrients*. 2015;7(2):1119–30.
55. Schnorr SL, Candela M, Rampelli S, Centanni M, Consolandi C, Basaglia G, et al. Gut microbiome of the Hadza hunter-gatherers. *Nat Commun*. 2014;5:3654.
56. Wang YC, McPherson K, Marsh T, Gortmaker SL, Brown M. Health and economic burden of the projected obesity trends in the USA and the UK. *Lancet*. 2011;378(9793):815–25.
57. Bhaskaran K, Douglas I, Forbes H, dos-Santos-Silva I, Leon DA, Smeeth L. Body-mass index and risk of 22 specific cancers: a population-based cohort study of 5.24 million UK adults. *Lancet*. 2014;384(9945):755–65.
58. Mariosa D, Smith-Byrne K, Richardson TG, Ferrari P, Gunter MJ, Papadimitriou N, et al. Body size at different ages and risk of 6 cancers: a Mendelian randomization and prospective cohort study. *J Natl Cancer Inst*. 2022;114(9):1296–300.
59. Weir TL, Trikha SRJ, Thompson HJ. Diet and cancer risk reduction: the role of diet-microbiota interactions and microbial metabolites. *Semin Cancer Biol*. 2021;70:53–60.
60. Schulz MD, Atay C, Heringer J, Romrig FK, Schwitalla S, Aydin B, et al. High-fat-diet-mediated dysbiosis promotes intestinal carcinogenesis independently of obesity. *Nature*. 2014;514(7523):508–12.
61. Song M, Chan AT, Sun J. Influence of the gut microbiome, diet, and environment on risk of colorectal cancer. *Gastroenterology*. 2020;158(2):322–40.
62. Carter AR, Sanderson E, Hammerton G, Richmond RC, Davey Smith G, Heron J, et al. Mendelian randomisation for mediation analysis: current methods and challenges for implementation. *Eur J Epidemiol*. 2021;36(5):465–78.
63. Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. *Science*. 2008;322(5903):881–8.
64. Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, Boehnke M. Genome-wide association studies in diverse populations. *Nat Rev Genet*. 2010;11(5):356–66.
65. Hoffmann TJ, Zhan Y, Kvale MN, Hesselson SE, Gollub J, Iribarren C, et al. Design and coverage of high throughput genotyping arrays optimized for individuals of East Asian, African American, and Latino race/ethnicity using imputation and a novel hybrid SNP selection algorithm. *Genomics*. 2011;98(6):422–30.
66. Manning A, Highland HM, Gasser J, Sim X, Tukiainen T, Fontanillas P, et al. A low-frequency inactivating *AKT2* variant enriched in the Finnish population is associated with fasting insulin levels and type 2 diabetes risk. *Diabetes*. 2017;66(7):2019–32.
67. Farh KK, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature*. 2015;518(7539):337–43.
68. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, et al. The structure of haplotype blocks in the human genome. *Science*. 2002;296(5576):2225–9.
69. Chen R, Shi L, Hakenberg J, Naughton B, Sklar P, Zhang J, et al. Analysis of 589,306 genomes identifies individuals resilient to severe Mendelian childhood diseases. *Nat Biotechnol*. 2016;34(5):531–8.
70. Castel SE, Cervera A, Mohammadi P, Aguet F, Reverter F, Wolman A, et al. Modified penetrance of coding variants by cis-regulatory variation contributes to disease risk. *Nat Genet*. 2018;50(9):1327–34.
71. Meisel SF, Beeken RJ, van Jaarsveld CH, Wardle J. Genetic susceptibility testing and readiness to control weight: results from a randomized controlled trial. *Obesity (Silver Spring)*. 2015;23(2):305–12.
72. Meisel SF, Walker C, Wardle J. Psychological responses to genetic testing for weight gain: a vignette study. *Obesity (Silver Spring)*. 2012;20(3):540–6.
73. Tucker-Drob EM, Bates TC. Large cross-national differences in gene x socioeconomic status interaction on intelligence. *Psychol Sci*. 2016;27(2):138–49.
74. Agustí A, Melén E, DeMeo DL, Breyer-Kohansal R, Faner R. Pathogenesis of chronic obstructive pulmonary disease: understanding the contributions of gene-environment interactions across the lifespan. *Lancet Respir Med*. 2022;10(5):512–24.

Publisher's Note

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