World Journal of Surgery and Surgical Research

Research Article
Published: 19 Aug, 2024

The Causal Link Between Colon Adenocarcinoma and Gut Microbiota: A Bi-Directional Mendelian Randomization Study

Wu ZY[1#], Huang SK[2#], Ye YY[3#], Zheng WT[4], Chen HL[2], Hou XT[2], Shen RJ[5], Vithayaphalert V[6], Juengpanich S[5]* and Zheng HM[1]**

[1]Department of Surgical Oncology, Sir Run Run Shaw Hospital, School of Medicine, Zhejiang University, China

[2]Department of Clinical Medicine, The Third Clinical School of Guangzhou Medical University, China

[3]Department of Immunology and Pathogen Biology, School of Basic Medical Sciences, Hangzhou Normal University, China

[4]Department of Clinical Medicine, The First Clinical School of Guangzhou Medical University, China

[5]School of Medicine, Zhejiang University, Hangzhou, China

[6]Department of Internal Medicine, King memorial Chulalongkorn Hospital, Thailand

[#]These authors contributed equally to this work

Abstract

Purpose: This study aims to explore the causal relationships between gut microbiota and colon cancer using a bidirectional Mendelian Randomization (MR) approach.

Methods: We employed a bidirectional MR research, utilizing genetic variants as Instrumental Variables (IVs) to infer causality. Genetic data for gut microbiota were obtained from the MiBioGen Consortium, encompassing 211 gut microbiota taxa, while data for colon cancer were sourced from the FinnGen Consortium, including 290,221 participants. We conducted MR analyses to evaluate the influence of gut microbiota on colon cancer and *vice versa*. IVs were selected based on stringent criteria to ensure the robustness of our findings. Multiple statistical methods, including Inverse Variance Weighted (IVW) and MR-Egger regression, were used to assess causality.

Results: Forward MR analysis identified eight gut microbiota taxa associated with colon cancer risk. Notably, *Enterobacteriales* (order), *Peptococcaceae* (family), and genera like *Dialister* and *Roseburia* were protective factors, whereas *Enterobacteriaceae*, *Porphyromonadaceae*, and *Escherichia-Shigella* were risk factors. Reverse MR analysis revealed that colon cancer was associated with changes in the abundance of eight gut microbiota taxa, including an increase in *Lactobacillales* (order) and *Streptococcaceae* (family), with genera such as *Eubacterium_eligens_group* and *Roseburia* being protective.

Conclusion: Our bidirectional MR study underscores a significant causal relationship between gut microbiota and colon cancer. Specific gut microbiota taxa were identified as either risk or protective factors for colon cancer, suggesting potential avenues for microbiota-targeted interventions in colon cancer prevention and treatment. Further research is warranted to elucidate the underlying mechanisms and validate these findings in diverse populations.

Keyword: Gut microbiota; Colon adenocarcinoma; Mendelian randomization

OPEN ACCESS

Correspondence:
He-Ming Zheng, Department of Surgical Oncology, Sir Run Run Shaw Hospital, School of Medicine, Zhejiang University, Hangzhou 310000, China, Sarun Juengpanich, School of Medicine, Zhejiang University, Hangzhou 310058, China, Vasinth Vithayaphalert, Department of Internal Medicine, King memorial Chulalongkorn Hospital, Bangkok 10330, Thailand,

Received Date: *22 Jul 2024*
Accepted Date: *13 Aug 2024*
Published Date: *19 Aug 2024*

Citation:
Wu ZY, Huang SK, Ye YY, Zheng WT, Chen HL, Hou XT, et al. The Causal Link Between Colon Adenocarcinoma and Gut Microbiota: A Bi-Directional Mendelian Randomization Study. World J Surg Surgical Res. 2024; 7: 1566.

Introduction

Colorectal cancer is a major malignant tumor posing a significant threat to human health. It ranks third globally and second in China in terms of incidence [1,2]. Intestinal microbiota, especially gut flora, plays a crucial role in the human body by promoting digestion and absorption, eliminating pathogenic bacteria, and improving the intestinal microenvironment [3]. However, research has also indicated that gut flora significantly contributes to diseases such as cancer, inflammatory bowel disease, and metabolic disorders like diabetes [4].

There is substantial evidence suggesting that various types of gut microbiota are associated with the development of colorectal cancer [5]. However, the precise nature of this relationship remains

unclear.

Mendelian Randomization (MR) is an analytical approach employing genetic variants as Instrumental Variables (IVs) for exposure [6]. These variants are randomly distributed during conception, thereby avoiding influences from reverse causation and confounding factors. This method provides estimates of disease risk with minimal interference from external variables. MR surpasses limitations of traditional epidemiological studies, yielding robust results, contingent upon the absence of pleiotropy (where genetic variants affect disease outcomes through multiple pathways).

In this study, we conducted bidirectional MR analyses to investigate causal relationships between gut microbiota and colon cancer. Through reverse MR methods, we also explored whether SNPs associated with colon cancer causally impact specific microbial communities, examining potential influences of tumor predisposition and/or presence on gut microbiota.

Method

Study design

We performed a bidirectional Mendelian Randomization (MR) to investigate the relationship between gut microbiota and colon adenocarcinoma. MR is a causal inference method based on genetic variation by finding SNPs which are associated with biological factors, *i.e.*, Instrumental Variables (IVs), and using these IVs to explore the effects of exposure on outcome. In this study, the influence of gut microbiota on colon adenocarcinoma was explored, and the causal direction of gut microbiota and colon adenocarcinoma was investigated using bidirectional MR. Due to the existence of gene variability, the influence of the research of causal relationship can be avoided, and the reliability of this study is improved. In our research, we adhere to three assumptions of MR [6].

Data source

To avoid overlap bias caused by the same source of exposure and outcome, we collected data from different databases, and most of them are from European populations. GWAS data for 211 gut microbiotas were obtained from the MiBioGen Consortium (mibiogen.gcc.rug.nl). The MiBioGen Consortium is the largest genomic meta-analysis of gut microbiota data published to date [7]. This research included 18,340 participants from 24 cohorts, 13,266 of whom were European. MiBioGen Consortium used microbiota Quantitative Trait Loci (mbQTL) mapping analysis to identify host genetic variation associated with genetic loci associated with the abundance of bacterial classes in the gut microbiota. In this study, genus (131, of which 12 are unknown) was the lowest classification level, and other classification levels such as phylum (9), class (16), order (20), and family (35, of which 3 are unknown) were brought into study [7]. Finally, a total of 196 taxa containing 9 phyla, 16 classes, 20 orders, 32 families and 119 genera were included in the study. GWAS data for colon adenocarcinoma were obtained from the FinnGen Consortium (www.finngen.fi) R9 data, which included 290221 participants, including 3,084 patients and 287,137 controls [8]. Other additional details were available in Supplementary Table 1.

Ethics statement

All of the GWAS data used in this study were from original studies which had received ethical approval. This study did not require additional ethical approval.

Selection of instrumental variables

In our bidirectional MR study, we first explored the influence of gut microbiota on colon adenocarcinoma, and then explored the influence of colon adenocarcinoma on gut microbiota using reserve MR. The following are the main analysis steps completed in our study. For the filtrating of IVs of gut microbiota as exposure, as the number of SNPs obtained in most data was very small or even zero under the threshold of P<5e-8, we selected SNPs related to gut microbiota with the threshold of P<1e-5. In order to reduce weak instrumental variable bias, IVs with F-statistic <10 were excluded. To avoid the linkage disequilibrium, all selected IVs were clumped with r2<0.001 and LD<10000 kb. For the filtrating of colon adenocarcinoma as outcome, we selected the corresponding SNPs according to the selected IVs and ensured that the P-value of the outcome variable was ≥ 5e-8. When colon adenocarcinoma was used as exposure variable for reverse MR analysis, SNPs associated with it were selected with a threshold of P<5e-6 to ensure that a sufficient number of SNPs were obtained. Similarly, the IVs were excluded when the F-statistic was <10, and the selected IVs were clumped with r^2<0.001 and LD<10000 kb. Finally, effective IVs were obtained after harmonizing the effects of exposure and outcome and removing SNPs that cannot be harmonized (Supplementary Table 2, 3). Finally, Inverse Variance Weighted (IVW) [9,10], MR-egger regression [11,12], Simple Mode (SM) [13], weighted mode [13] and weighted median method [14] were used to evaluate the effective IVs.

Statistical analysis

All analyses in this study were performed using R v4.2.2, TwoSampleMR [15], MR-PRESSO [16] and Forestploter software packages. The formula for calculating the F statistic was: $\beta^2_{exposure}/se^2_{exposure}$ (β, effect; se, standard error), referring to the MR study by Kjaerggaard et al. and Chen et al. [17,18]. Prior to the MR analysis, we performed a Mendelian Randomization Pleiotropy Residual Sum and Outlier (MR-PRESSO) analysis to ensure the confidence of the MR results by removing any outliers with potential pleiotropy (Supplementary Table 4a, 4b, 5a, 5b). Then five methods, IVW, MR-Egger, SM, weighted median method and weighted mode, were used for MR analysis. In our study, IVW was used as the main analysis method, which integrated Wald Ratio estimates of each IVs to evaluate the causal effect between exposure and outcome. The remaining four analysis methods were as supplementary methods. The MR-Egger method assumes that IVs is independent of the direct effect and adds the intercept term compared with the IVW analysis method. The closer the intercept is to 0, the less likely horizontal pleiotropy is (P>0.05). The weighted median method is the median of the distribution function obtained after all the individual SNP effect values are sorted by weight. When the effective IVs reach at least 50%, the weighted median method can obtain a robust estimate. The SM classifies SNPs according to causal effect and groups SNPs with similar values, using the group with the largest number of SNPs to estimate the causal effect. The weighted mode weights the causal effect value of SNPs in each cluster with the number of SNPs and returns a provisional estimate of the number of SNPs with the greatest weight.

Sensitivity analyses were performed after MR Analysis. In our study, Cochran's Q test was used to estimate the heterogeneity of IVW method and MR-Egger method. When there was heterogeneity, IVW method of random effects model was used to estimate causal effect (Supplementary Table 6a, 6b, 7a, 7b). Secondly, we used MR-Egger regression analysis to judge whether there was pleiotropy

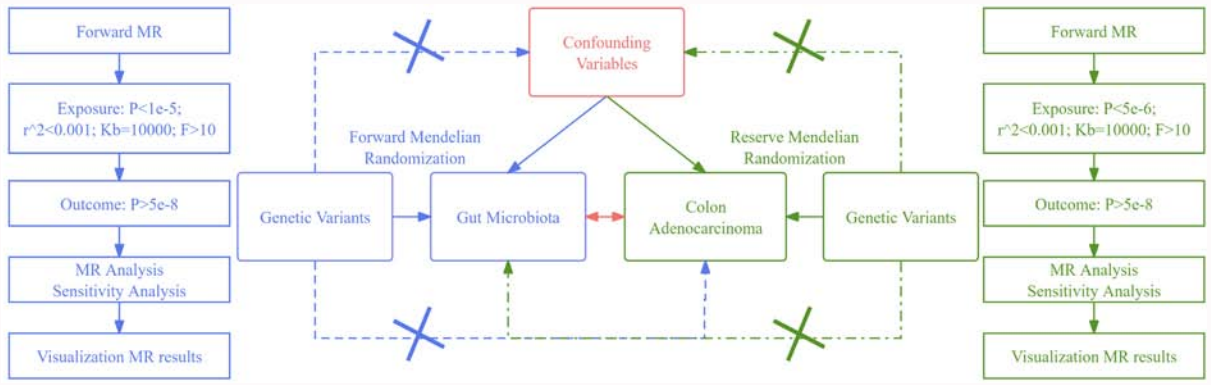

Figure 1: Flowchart of this study.

(Supplementary Table 8a, 8b). Finally, a Leave-One-Out (LOO) test was performed for the significant results to assess whether the total IVW estimate was affected by any single SNP and to detect bias due to heterogeneity by a method that excluded IV one by one (Supplementary Table 9a, 9b).

Results

MR of gut microbiota on colon adenocarcinoma

In this study, a total of 8 gut microbiota were associated with colon adenocarcinoma, including 1 bacterial order: *Enterobacteriales*, 1 bacterial family: *Peptococcaceae*, *Enterobacteriaceae* and *Porphyromonadaceae*, and 4 bacterial genera: *Roseburia*, *Family_XIII_AD3011_group*, *Dialister* and *Escherichia-Shigella*. In the taxon at the order level, *Enterobacteriales* was a protective factor in colon adenocarcinoma (IVW: OR (95% CI): 1.481 (1.017~2.157), P=0.04). In the family taxon, *Peptococcaceae* (IVW: OR (95% CI): 0.726 (0.56~0.941), P=0.015) was a protective factor for colon adenocarcinoma, while *Enterobacteriaceae* (IVW: OR (95% CI): 1.481 (1.017~2.157), P=0.04) and *Porphyromonadaceae* (IVW: OR (95% CI): 1.521 (1.036~2.232), P=0.032) were risk factors for colon adenocarcinoma. In the generic taxon, *Dialister* (IVW: OR (95% CI): 0.751 (0.582~0.968), P=0.027), *Family_XIII_AD3011_group* (IVW: OR (95% CI): 0.73 (0.551~0.966), P=0.028) and *Roseburia* (IVW: OR (95% CI): 0.707 (0.527~0.949), P=0.021) were protective factors for colon adenocarcinoma, while *Escherichia-Shigella* (IVW: OR (95% CI): 1.368 (1.033~1.811), P=0.029) was a risk factor for colon adenocarcinoma (Figure 1, 2, Supplementary Table 10a).

Reverse MR of gut microbiota on colon adenocarcinoma

Reverse Mendelian Randomization have shown that colon adenocarcinoma was associated with the abundance of 8 gut microbiota. There was 1 from the order taxon: *Lactobacillales*; 1 from the family taxon: *Streptococcaceae*, and 6 genera: *Eubacterium_eligens_*Group, *Howardella*, *Roseburia*, *Prevotella* 7, *Ruminococcus* 1 and *Slackia*. In the taxon at the order level, colon adenocarcinoma increased *Lactobacillales* abundance (IVW: OR (95% CI): 1.046 (1.002~1.092), P=0.042). In a family taxon, colon adenocarcinoma increased the abundance of *Streptococcaceae* (IVW: OR (95% CI): 1.045 (1.0002-1.093), P=0.048). In the generic taxon, colon adenocarcinoma was harmful to *Eubacterium_eligens_group* (IVW: OR (95% CI): 0.953 (0.91~0.999), P=0.047), *Roseburia* (IVW: OR (95% CI): 0.941 (0.9~0.985), P=0.009), *Prevotella* 7 (IVW: OR (95% CI): 0.88 (0.783~0.989), P=0.032) and *Ruminococcus* 1 (IVW: OR (95% CI): 0.937 (0.893~0.984), P=0.009). While colon adenocarcinoma was beneficial to *Howardella* (IVW: OR (95% CI): 1.119 (1.019~1.229), P=0.019) and *Slackia* (IVW: OR (95% CI): 1.08 (1.004~1.161), P=0.04) (Figure 3, Supplementary Table 10b).

F-statistics and sensitivity analysis

Each SNP included in the analysis had an F-statistic greater than 10 (Supplementary Table 2, 3). For the heterogeneity test, MR-Egger regression method and IVW method were used to calculate Cochran's Q statistics. And IVW was as a main method. When the P value of Cochran's IVW Q statistics were <0.05, there was heterogeneity, and results for IVW method of random effects model were used. In the forward Mendelian Randomization study of gut microbiota on colon adenocarcinoma, the relations of *Bifidobacteriales* from the order taxon, *Bifidobacteriaceae* and *Defluviitaleaceae* from the family taxon, and *Eubacterium_rectale_group*, *Ruminococcus* 1 and *Coprococcus* 1 from the genus taxon, *Ruminococcaceae_UCG_005* and *Eubacterium_coprostanoligenes_group* with colon adenocarcinoma existed heterogeneity (p<0.05), and the rest microbiota did not exist heterogeneity (Supplementary Table 6a, 7a). For the pleiotropy test, MR-Egger was used for gene pleiotropy test. With the exception of *Allisonella, Intestinibacter, and Olsenella* from genus taxon with P-value <0.05 of pleiotropy test, gene pleiotropy was not found in any other group (Supplementary Table 8a). For the reverse Mendelian Randomization study of gut microbiota on colon adenocarcinoma, in addition to *Alphaproteobacteria* and *Clostridia* from class taxon *Rhodospirillales* and *Clostridiales* from order taxon, *Rhodospirillaceae* from family taxon and *Oxalobacter* from genus taxon, there was no heterogeneity in the relationship between colon adenocarcinoma and other microbiota (Supplementary Table 6b, 7b). In the pleiotropy test, except for *Coprococcus* 2, *Sutterella*, *Eisenbergiella* from genus level and *Burkholderiales* from order level, no gene pleiotropy was found between colon adenocarcinoma and other microbiota (Supplementary Table 8b).

Visualization of MR results

We visualized the groups with significant MR results in the bidirectional Mendelian Randomization study described above. The scatter plot shows correlations in the same direction as the forest plot (Figure 4, 5). According to the data of LOO test, LOO diagrams were shown, which were consistent with the results of the above LOO test (Figure 6, 7). Forest plots of bidirectional IVW results show a clear causal relationship, which were consistent with the results of the MR analysis above (Supplementary Figure 1, 2). Finally, the funnel plots show the distribution of effective IVs in the bidirectional MR analysis, indicating that our data keep to a random distribution

Exposure	Outcome	method	nsnp	OR	pvalue
Dialister	Colon Adenocarcinoma	Inverse variance weighted	11	0.751	0.027
Dialister	Colon Adenocarcinoma	MR Egger	11	0.429	0.146
Dialister	Colon Adenocarcinoma	Simple mode	11	0.775	0.331
Dialister	Colon Adenocarcinoma	Weighted median	11	0.790	0.178
Dialister	Colon Adenocarcinoma	Weighted mode	11	0.771	0.327
Escherichia–Shigella	Colon Adenocarcinoma	Inverse variance weighted	10	1.368	0.029
Escherichia–Shigella	Colon Adenocarcinoma	MR Egger	10	1.458	0.419
Escherichia–Shigella	Colon Adenocarcinoma	Simple mode	10	1.483	0.205
Escherichia–Shigella	Colon Adenocarcinoma	Weighted median	10	1.409	0.069
Escherichia–Shigella	Colon Adenocarcinoma	Weighted mode	10	1.495	0.210
FamilyXIIIAD3011group	Colon Adenocarcinoma	Inverse variance weighted	13	0.730	0.028
FamilyXIIIAD3011group	Colon Adenocarcinoma	MR Egger	13	1.047	0.949
FamilyXIIIAD3011group	Colon Adenocarcinoma	Simple mode	13	0.788	0.431
FamilyXIIIAD3011group	Colon Adenocarcinoma	Weighted median	13	0.754	0.123
FamilyXIIIAD3011group	Colon Adenocarcinoma	Weighted mode	13	0.749	0.311
Roseburia	Colon Adenocarcinoma	Inverse variance weighted	14	0.707	0.021
Roseburia	Colon Adenocarcinoma	MR Egger	14	1.045	0.926
Roseburia	Colon Adenocarcinoma	Simple mode	14	0.789	0.465
Roseburia	Colon Adenocarcinoma	Weighted median	14	0.798	0.239
Roseburia	Colon Adenocarcinoma	Weighted mode	14	0.809	0.463
Enterobacteriaceae	Colon Adenocarcinoma	Inverse variance weighted	6	1.481	0.040
Enterobacteriaceae	Colon Adenocarcinoma	MR Egger	6	1.382	0.794
Enterobacteriaceae	Colon Adenocarcinoma	Simple mode	6	1.886	0.170
Enterobacteriaceae	Colon Adenocarcinoma	Weighted median	6	1.736	0.028
Enterobacteriaceae	Colon Adenocarcinoma	Weighted mode	6	1.828	0.169
Peptococcaceae	Colon Adenocarcinoma	Inverse variance weighted	9	0.726	0.015
Peptococcaceae	Colon Adenocarcinoma	MR Egger	9	0.745	0.474
Peptococcaceae	Colon Adenocarcinoma	Simple mode	9	0.640	0.186
Peptococcaceae	Colon Adenocarcinoma	Weighted median	9	0.723	0.072
Peptococcaceae	Colon Adenocarcinoma	Weighted mode	9	0.998	0.995
Porphyromonadaceae	Colon Adenocarcinoma	Inverse variance weighted	8	1.521	0.032
Porphyromonadaceae	Colon Adenocarcinoma	MR Egger	8	0.570	0.498
Porphyromonadaceae	Colon Adenocarcinoma	Simple mode	8	1.152	0.738
Porphyromonadaceae	Colon Adenocarcinoma	Weighted median	8	1.256	0.371
Porphyromonadaceae	Colon Adenocarcinoma	Weighted mode	8	1.124	0.762
Enterobacteriales	Colon Adenocarcinoma	Inverse variance weighted	6	1.481	0.040
Enterobacteriales	Colon Adenocarcinoma	MR Egger	6	1.382	0.794
Enterobacteriales	Colon Adenocarcinoma	Simple mode	6	1.886	0.170
Enterobacteriales	Colon Adenocarcinoma	Weighted median	6	1.736	0.028
Enterobacteriales	Colon Adenocarcinoma	Weighted mode	6	1.828	0.169

Figure 2: Forest plot of forward MR results.

(Supplementary Figure 3, 4).

Discussion

In this study, we utilized bidirectional Mendelian Randomization (MR) to investigate the relationships between gut microbiota and colon cancer. Forward MR analysis identified several microbial groups associated with colon cancer risk. Ample evidence suggests a link between gut microbiota and the incidence of colon cancer. Specifically, our forward MR analysis found that members of the *Firmicutes phylum*, such as *Peptococcaceae*, *Dialister*, *Family_XIII_AD3011_group*, and *Roseburia*, served as protective factors against colon cancer, whereas *Escherichia-Shigella* was identified as a risk factor. These findings align with numerous other studies.

Several studies investigating the relationship between specific gut microbiota and colorectal cancer consistently reveal significant differences in overall microbiota composition between patients with colorectal cancer and healthy individuals. Moore et al. [19] employed fecal bacterial isolation and culture methods, identifying 15 bacterial species strongly associated with an increased risk of colon cancer. Among these, *Bacteroides fragilis*, *Streptococcus* species, and *Fusobacterium* were notably more abundant in individuals at

Exposure	Outcome	method	nsnp	OR	pvalue
Colon Adenocarcinoma	Eubacteriumeligensgroup	Inverse variance weighted	14	0.953	0.047
Colon Adenocarcinoma	Eubacteriumeligensgroup	MR Egger	14	0.952	0.559
Colon Adenocarcinoma	Eubacteriumeligensgroup	Simple mode	14	0.963	0.522
Colon Adenocarcinoma	Eubacteriumeligensgroup	Weighted median	14	0.960	0.241
Colon Adenocarcinoma	Eubacteriumeligensgroup	Weighted mode	14	0.966	0.473
Colon Adenocarcinoma	Howardella	Inverse variance weighted	13	1.119	0.019
Colon Adenocarcinoma	Howardella	MR Egger	13	1.105	0.544
Colon Adenocarcinoma	Howardella	Simple mode	13	1.155	0.208
Colon Adenocarcinoma	Howardella	Weighted median	13	1.144	0.038
Colon Adenocarcinoma	Howardella	Weighted mode	13	1.149	0.205
Colon Adenocarcinoma	Prevotella7	Inverse variance weighted	12	0.880	0.032
Colon Adenocarcinoma	Prevotella7	MR Egger	12	0.650	0.031
Colon Adenocarcinoma	Prevotella7	Simple mode	12	0.866	0.289
Colon Adenocarcinoma	Prevotella7	Weighted median	12	0.851	0.030
Colon Adenocarcinoma	Prevotella7	Weighted mode	12	0.851	0.160
Colon Adenocarcinoma	Roseburia	Inverse variance weighted	14	0.941	0.009
Colon Adenocarcinoma	Roseburia	MR Egger	14	0.926	0.350
Colon Adenocarcinoma	Roseburia	Simple mode	14	0.982	0.756
Colon Adenocarcinoma	Roseburia	Weighted median	14	0.958	0.200
Colon Adenocarcinoma	Roseburia	Weighted mode	14	0.981	0.699
Colon Adenocarcinoma	Ruminococcus1	Inverse variance weighted	14	0.937	0.009
Colon Adenocarcinoma	Ruminococcus1	MR Egger	14	1.015	0.854
Colon Adenocarcinoma	Ruminococcus1	Simple mode	14	0.900	0.115
Colon Adenocarcinoma	Ruminococcus1	Weighted median	14	0.926	0.022
Colon Adenocarcinoma	Ruminococcus1	Weighted mode	14	0.894	0.098
Colon Adenocarcinoma	Slackia	Inverse variance weighted	13	1.080	0.040
Colon Adenocarcinoma	Slackia	MR Egger	13	1.158	0.251
Colon Adenocarcinoma	Slackia	Simple mode	13	1.157	0.086
Colon Adenocarcinoma	Slackia	Weighted median	13	1.096	0.060
Colon Adenocarcinoma	Slackia	Weighted mode	13	1.132	0.104
Colon Adenocarcinoma	Streptococcaceae	Inverse variance weighted	14	1.045	0.048
Colon Adenocarcinoma	Streptococcaceae	MR Egger	14	1.102	0.214
Colon Adenocarcinoma	Streptococcaceae	Simple mode	14	1.035	0.537
Colon Adenocarcinoma	Streptococcaceae	Weighted median	14	1.035	0.318
Colon Adenocarcinoma	Streptococcaceae	Weighted mode	14	1.032	0.561
Colon Adenocarcinoma	Lactobacillales	Inverse variance weighted	14	1.046	0.042
Colon Adenocarcinoma	Lactobacillales	MR Egger	14	1.124	0.136
Colon Adenocarcinoma	Lactobacillales	Simple mode	14	0.970	0.565
Colon Adenocarcinoma	Lactobacillales	Weighted median	14	1.043	0.196
Colon Adenocarcinoma	Lactobacillales	Weighted mode	14	1.049	0.375

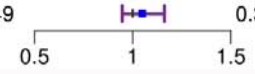

Figure 3: Forest plot of reverse MR results.

higher risk for colon cancer. Burkhardt et al. [20] utilized 16S rRNA gene sequencing to study fecal and mucosal bacteria diversity in 59 colorectal cancer patients, 21 colorectal polyp patients, and 56 healthy controls. Their analysis using principal component analysis revealed distinct differences in gut microbiota composition among these groups. In colorectal cancer patients, there was a significant increase in the abundance of *Bacteroides* species, *Helicobacter* species, and certain pathogenic bacteria originating from the oral cavity, such as *Streptococcus*, *Prevotella*, and *Clostridium* species.

In reverse Mendelian Randomization (MR), colon cancer might cause specific changes in gut microbiota composition. Among patients with colon cancer, there is a notable increase in the abundance of *Lactobacillales* and *Streptococcaceae*. These findings align consistently with previous research findings. Sobhani et al. [21] indicated that Colorectal Cancer (CRC) patients exhibit significant imbalances in their gut microbiota, marked by changes in both dominant and less prevalent bacterial species present in their fecal samples. Specifically, there is a notable increase in genera such as *Bacteroides* and *Prevotella* compared to healthy individuals. Wang et al. research further supports these findings, revealing elevated levels of frail

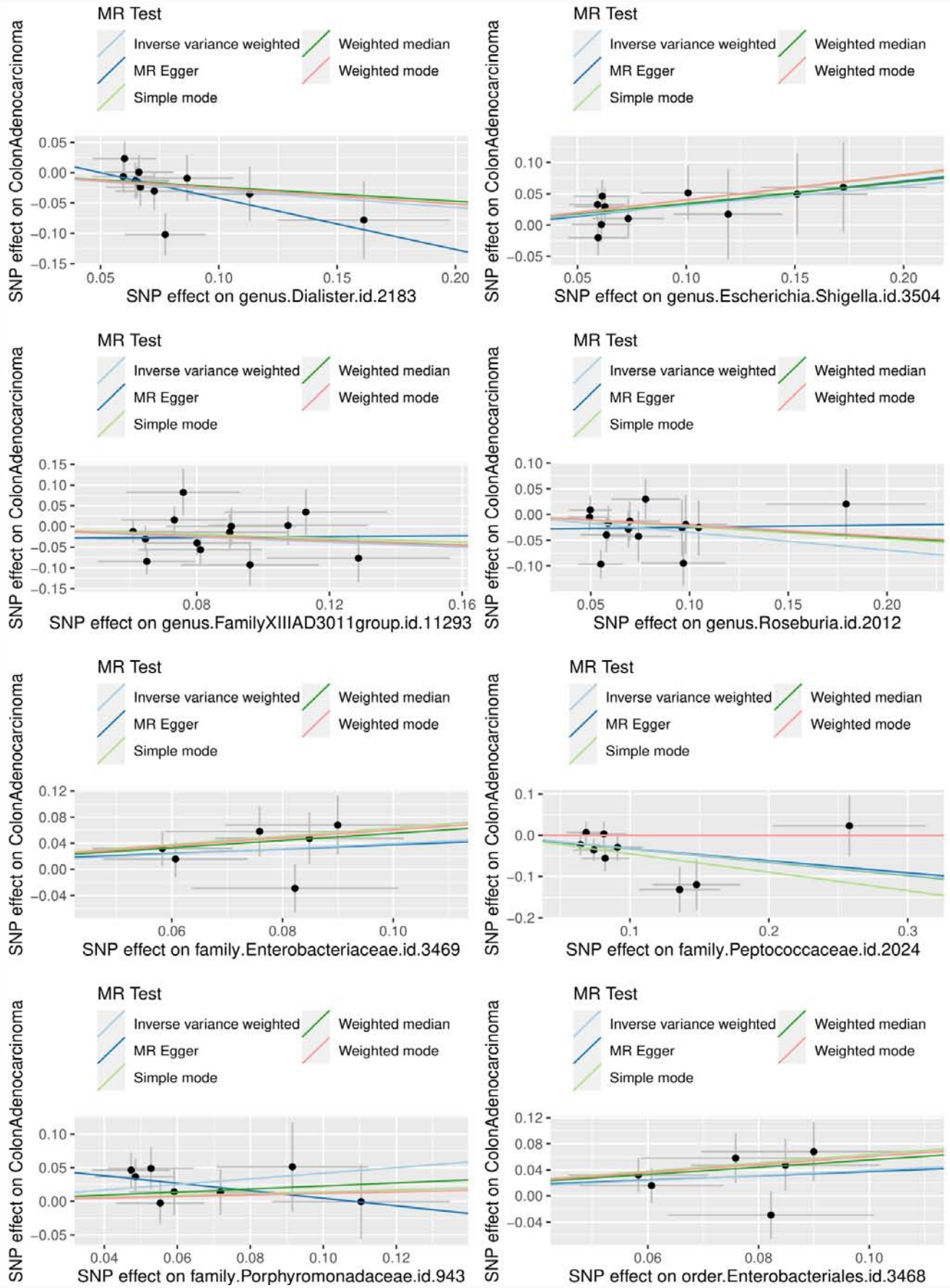

Figure 4: Scatter plots of forward MR results.

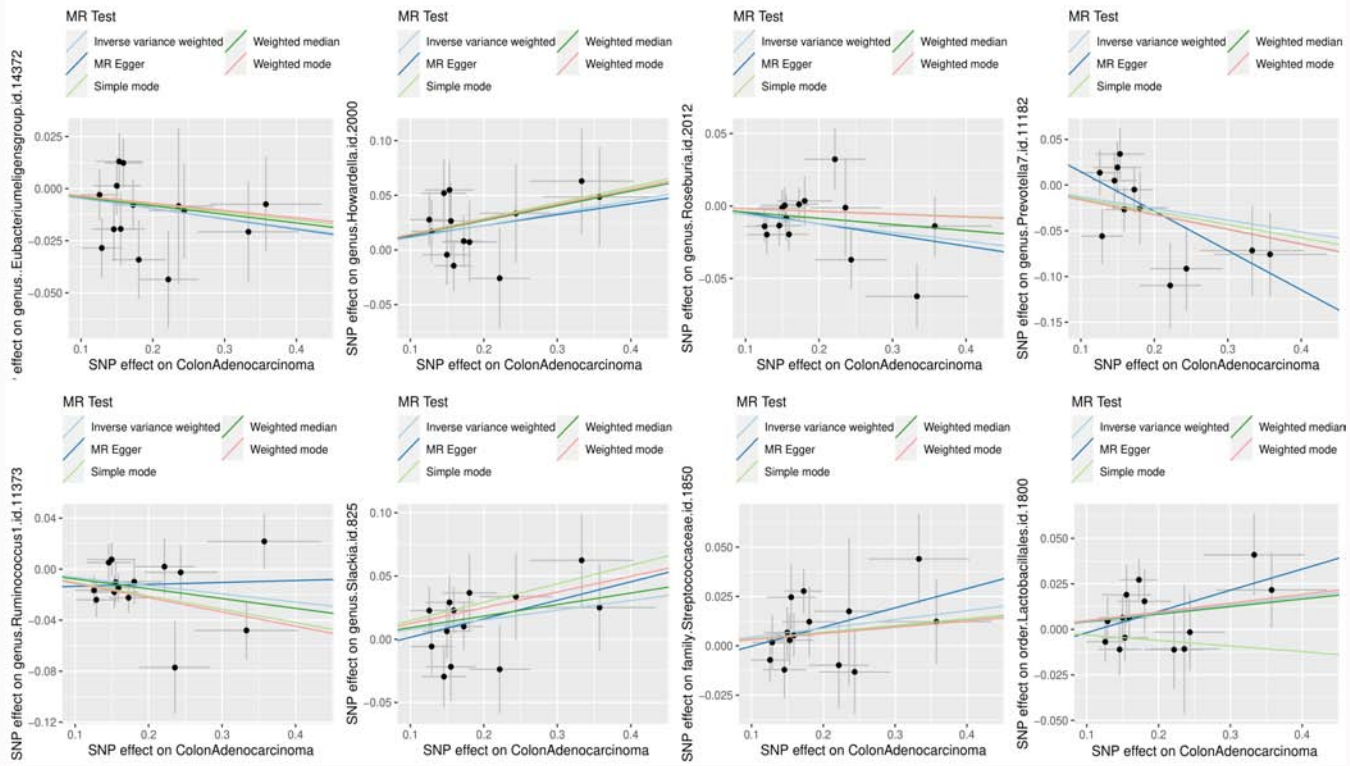

Figure 5: Scatter plots of reverse MR results.

Bacilli, *Enterococcus*, *Escherichia/Shigella*, *Klebsiella*, *Streptococcus*, and *Enterococcus* in colorectal cancer patients' feces. In contrast, *Roseburia* and *Faecalibacterium* are more abundant in fecal samples from healthy subjects. Moreover, Ahn et al. [22] observed reduced diversity in the gut microbiota of CRC patients, alongside decreased levels of *Clostridia* and increased presence of *Fusobacterium* and *Porphyromonas*. *Fusobacterium* was particularly enriched in rectal swabs from CRC patients compared to controls [23]. Additionally, Chen [24] found distinct differences in both fecal and tissue samples between CRC patients and healthy controls. They noted an enrichment of *Lactobacillales* in cancerous tissues and reduced levels of beneficial bacteria such as *Bifidobacterium* and *Faecalibacterium*. Analysis of mucosal-adherent bacteria from anal swabs showed decreases in *Bifidobacterium*, *Faecalibacterium*, and *Blautia*, while *Clostridium*, *Haemophilus*, *Enterococcus*, and *Mogibacterium* were more prevalent. These studies collectively highlight significant shifts in gut microbiota composition associated with CRC, suggesting potential implications for biomarker discovery and therapeutic interventions in the future. However, in our study, *Prevotella* is a risk factor for colon cancer. This may be attributed to differences in the sources of the sample data.

Roseburia is an important genus within the Firmicutes phylum, known for producing Short-Chain Fatty Acids (SCFAs) such as acetate, propionate, and butyrate. SCFAs play a crucial role in the metabolism of essential nutrients like carbohydrates and fats. Typical *Roseburia* strains are particularly adept at producing high levels of butyrate. These compounds can protect the gut from pathogens and diseases. Ohigashi et al. [25] found that the level of *Enterobacteriaceae* was significantly reduced in the feces of patients with colorectal cancer. Although the concentration of short-chain fatty acids was significantly reduced in the colorectal cancer group, gut pH increased. Therefore, it is speculated that the gut microbiota of patients with colorectal cancer changes. The decrease in short-chain fatty acids and the increase in pH may be related to the onset of colorectal cancer.

This study has several significant strengths. Firstly, we utilized bidirectional Mendelian Randomization (MR) analysis of GWAS data to comprehensively assess the bidirectional causal relationships between gut microbiota and colorectal cancer. These results are consistent with previous observational studies but provide stronger evidence through the MR approach, which minimizes bias from confounding factors and reverse causation. Secondly, sensitivity analyses conducted in our study demonstrated the robustness and reliability of our MR assumptions.

Despite its strengths, our study also faces several limitations. Firstly, for the filtrating of IVs of gut microbiota as exposure, as the number of SNPs obtained in most data was very small or even zero under the threshold of P<5e-8, we selected SNPs related to gut microbiota with the threshold of P<1e-5. However, using a lenient P-value threshold of 1e-5 to select IVs may introduce weak instrument bias and horizontal pleiotropy. Since many bacteria were represented by only one or two SNPs when applying a 5e-8 threshold, the genetic explanation (R^2) was low, potentially resulting in insufficient statistical power to detect modest or minor correlations. Therefore, the negative results of this study do not entirely rule out causal relationships between microbiota and colon cancer. Secondly, genetic factors influencing proximal and distal colorectal cancer differ [26], and similarly, the microbiota composition involved in left-sided versus right-sided colon cancer may vary substantially [27-29]. These distinctions were not addressed in our research. Thirdly, the composition of gut microbiota is influenced by numerous factors, including lifestyle variables such as dietary patterns, medication use, and overall health status [30]. These factors could obscure the interpretation of genetic tools, particularly among individuals

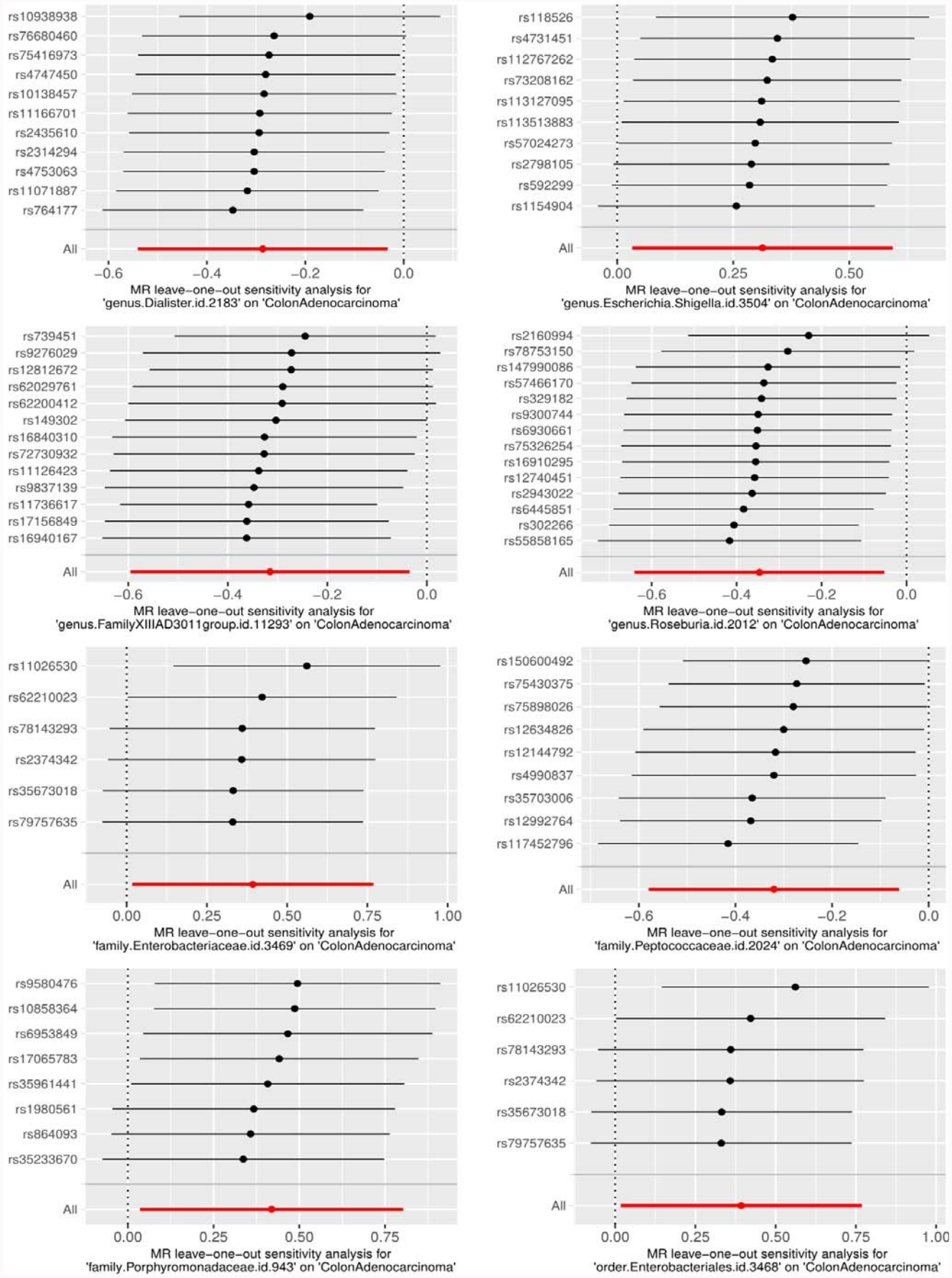

Figure 6: LOO plots of forward MR results.

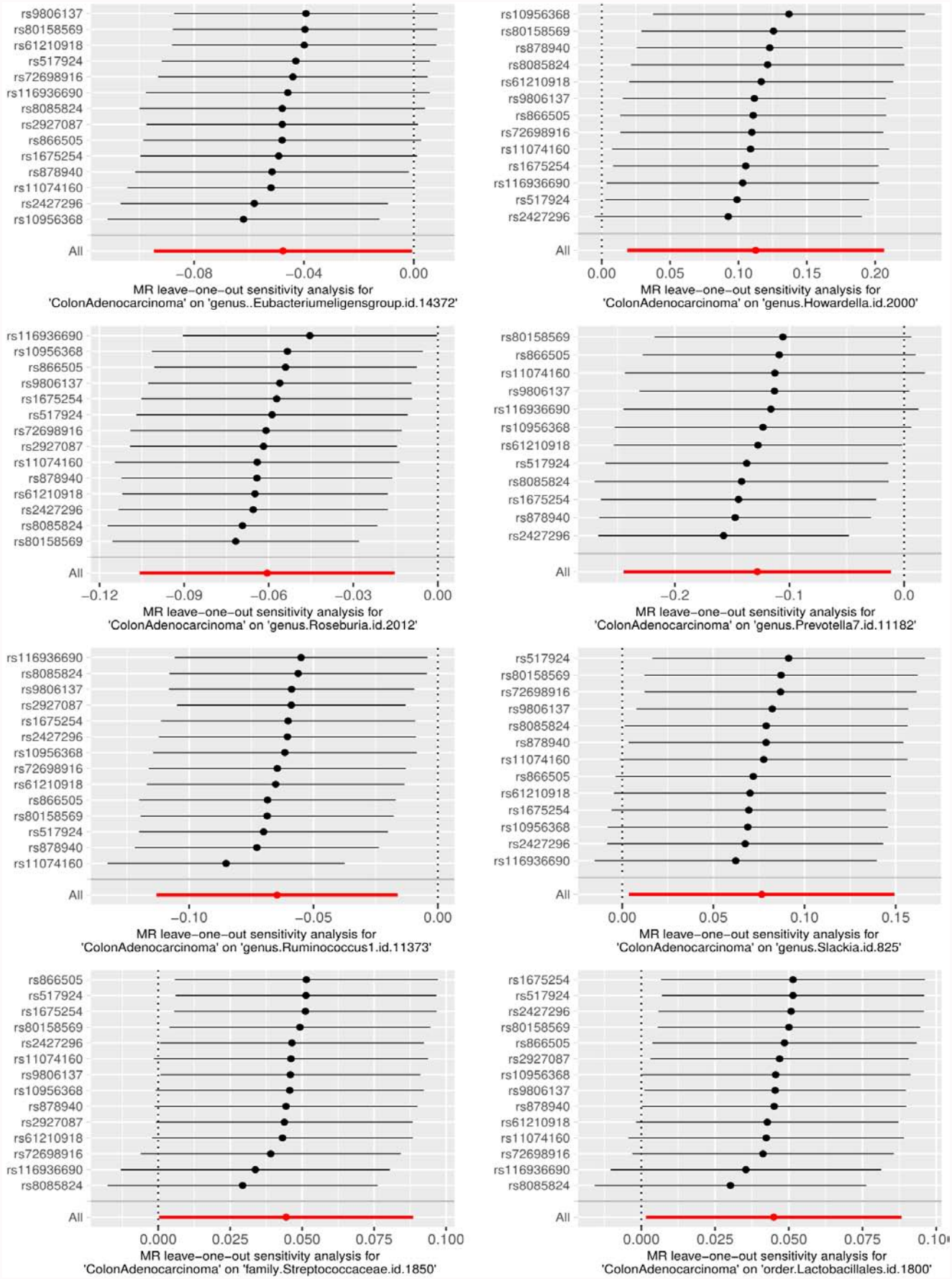

Figure 7: LOO plots of reverse MR results.

following a Westernized diet rich in saturated fats and red meat, and low in fiber. The potential interactions between diet and genetics, genetics and environmental factors, as well as non-genetic influences on outcomes, may also impact the observed results. Additionally, our study was limited to individuals of European descent, restricting the generalizability of our findings to broader populations. Our study design was predicated on the assumption of linear associations between gut microbiota composition, their metabolites, and colon cancer risk. This approach may overlook nonlinear effects that could be present in the data.

Our study demonstrates a close association between various gut microbiota and the incidence of colon cancer. Concurrently, the onset of colorectal cancer significantly alters the composition of gut microbiota. These findings suggest that detecting gut microbiota may help predict the risk of development trend of colon cancer, indicating its potential as a tumor biomarker. However, further experiments and clinical studies are necessary to validate this possibility.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71(3):209-49.
2. Zheng R, Zhang S, Zeng H, Wang S, Sun K, Chen R, et al. Cancer incidence and mortality in China, 2016. J Natl Cancer Cent. 2022;2(1):1-9.
3. De Vos WM, Tilg H, Van Hul M, Cani PD. Gut microbiome and health: Mechanistic insights. Gut. 2022;71(5):1020-32.
4. Sanna S, van Zuydam NR, Mahajan A, Kurilshikov A, Vila AV, Võsa U, et al. Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. Nature Genet. 2019;51(4):600-605.
5. Yu L, Zhao G, Wang L, Zhou X, Sun J, Li X, et al. A systematic review of microbial markers for risk prediction of colorectal neoplasia. Br J Cancer. 2022;126(9):1318-28.
6. Emdin CA, Khera AV, Kathiresan S. Mendelian randomization. JAMA. 2017;318(19):1925-6.
7. 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.
8. 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.
9. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol. 2013;37(7):658-65.
10. Burgess S, Bowden J, Fall T, Ingelsson E, Thompson SG. Sensitivity analyses for robust causal inference from Mendelian randomization analyses with multiple genetic variants. Epidemiology. 2017;28(1):30-42.
11. 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.
12. Slob EA, Groenen PJ, Thurik AR, Rietveld CA. A note on the use of Egger regression in Mendelian randomization studies. Int J Epidemiol. 2017;46(6):2094-7.
13. 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.
14. 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.
15. Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. PLoS Genet. 2017;13(11):e1007081.
16. Verbanck M, Chen C-Y, 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.
17. Kjaergaard AD, Marouli E, Papadopoulou A, Deloukas P, Kuś A, Sterenborg R, et al. Thyroid function, sex hormones and sexual function: A Mendelian randomization study. Eur J Epidemiol. 2021;36:335-44.
18. Chen L, Yang H, Li H, He C, Yang L, Lv G. Insights into modifiable risk factors of cholelithiasis: A Mendelian randomization study. Hepatology. 2022;75(4):785-96.
19. Moore W, Moore LH. Intestinal floras of populations that have a high risk of colon cancer. Appl Environ Microbiol. 1995;61(9):3202-7.
20. Flemer B, Lynch DB, Brown JM, Jeffery IB, Ryan FJ, Claesson MJ, et al. Tumor-associated and non-tumor-associated microbiota in colorectal cancer. Gut. 2017;66(4):633-43.
21. Sobhani I, Tap J, Roudot-Thoraval F, Roperch JP, Letulle S, Langella P, et al. Microbial dysbiosis in Colorectal Cancer (CRC) patients. PLoS One. 2011;6(1):e16393.
22. Ahn J, Sinha R, Pei Z, Dominianni C, Wu J, Shi J, et al. Human gut microbiome and risk for colorectal cancer. J Natl Cancer Inst. 2013;105(24):1907-11.
23. Mima K, Nishihara R, Qian ZR, Cao Y, Sukawa Y, Nowak JA, et al. Fusobacterium nucleatum in colorectal carcinoma tissue and patient prognosis. Gut. 2016;65(12):1973-80.
24. Chen W, Liu F, Ling Z, Tong X, Xiang C. Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer. PLoS One. 2012;7(6):e39743.
25. Ohigashi S, Sudo K, Kobayashi D, Takahashi O, Takahashi T, Asahara T, et al. Changes of the intestinal microbiota, short chain fatty acids, and fecal pH in patients with colorectal cancer. Dig Dis Sci. 2013;58(6):1717-26.
26. Huyghe JR, Harrison TA, Bien SA, Hampel H, Figueiredo JC, Schmit SL, et al. Genetic architectures of proximal and distal colorectal cancer are partly distinct. Gut. 2021;70(7):1325-34.
27. Xi Y, Yuefen P, Wei W, Quan Q, Jing Z, Jiamin X, et al. Analysis of prognosis, genome, microbiome, and microbial metabolome in different sites of colorectal cancer. J Transl Med. 2019;17(1):353.
28. Miyake T, Mori H, Yasukawa D, Hexun Z, Maehira H, Ueki T, et al. The comparison of fecal microbiota in left-side and right-side human colorectal cancer. Eur Surg Res. 2021;62(4):248-54.
29. Phipps O, Quraishi MN, Kumar A, Dickson EA, Oliver Ng, Acheson AG, et al. P317 Differences in normal mucosa and colorectal tumor microbiota between right and left colon. Gut. 2021;70(Suppl 1):A204.
30. Mehta RS, Song M, Nishihara R, Drew DA, Wu K, Qian ZR, et al. Dietary patterns and risk of colorectal cancer: analysis by tumor location and molecular subtypes. Gastroenterology. 2017;152(8):1944-53.e1.