Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Identification of Host Gene-Microbiome Associations in colorectal cancer patients Using Mendelian Randomization

yaoxian Xiang (hsdxyxky@163.com)
Beijing Luhe Hospital https://orcid.org/0009-0000-2393-3389
Chan Zhang
Beijing Luhe Hospital
Jing Wang
Beijing Luhe Hospital
Yurong Cheng
Beijing Luhe Hospital
Li Wang
Beijing Luhe Hospital
Yingying Tong
Beijing Luhe Hospital
Dong Yan
Beijing Luhe Hospital

Research Article

Keywords: Mendelian randomization (MR), gut microbiota, gene, colorectal cancer (CRC), causal relationship

Posted Date: March 24th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2683275/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Version of Record: A version of this preprint was published at Journal of Translational Medicine on August 10th, 2023. See the published version at https://doi.org/10.1186/s12967-023-04335-9.

Abstract

Background: There are many studies indicating that alterations in the abundance of certain gut microbiota are associated with colorectal cancer (CRC). However, a causal relationship has not been identified due to confounding factors such as lifestyle, environmental, and possible reverse causal associations between the two. Furthermore, certain host gene mutations can also contribute to the development of CRC. However, the association between genes and gut microbes in patients with CRC has not been extensively studied.

Methods: We conducted a two-sample Mendelian randomization (MR) study to reveal the causal relationship between gut microbiota and CRC. We obtained SNPs associated with gut microbiome abundance as instrumental variables (IVs) from a large-scale, multi-ethnic GWAS study, and extracted CRC-related datasets from an East Asian Population genetic consortia GWAS(AGWAS) study and FinnGen consortium, respectively. We analyzed a total of 166 bacterial features at four taxonomic levels, including order, family, genus, and species. The inverse-variance-weighted (IVW), weighted median, MR-Egger, and simple median methods were applied to the MR analysis, and the robustness of the results were tested using a series of sensitivity analyses. We extracted IVs of gut microbiota with direct causal association with CRC for SNP annotation to identify the genes in which these genetic variants were located to reveal the possible host gene-microbiome associations in CRC patients.

Results: The findings from our MR analysis based on CRC-associated GWAS datasets from AGWAS revealed causal relationships between 6 bacterial taxa and CRC at a locus-wide significance level ($P < 1 \times 10^{-5}$). The IVW method found that family *Porphyromonadaceae*, genera *Anaerotruncus*, *Intestinibacter*, *Slackia*, and *Ruminococcaceae* UCG004, and species *Eubacterium coprostanoligenes* group were positively associated with CRC risk, which was generally consistent with the results of other complementary analyses. The results of a meta-analysis of the MR estimates from the AGWAS and the FinnGen datasets showed that family *Porphyromonadaceae* and genera *Slackia*, *Anaerotruncus*, and *Intestinibacter* replicated the same causal association. Sensitivity analysis of all causal associations did not indicate significant heterogeneity, horizontal pleiotropy, or reverse causal associations. We annotated the SNPs at a locus-wide significance level of the above intestinal flora and identified 24 host genes that may be related to pathogenic intestinal microflora in CRC patients.

Conclusion: This study supported the causal relationship of gut microbiota on CRC and revealed a possible correlation between genes and pathogenic microbiota in CRC. These findings suggested that the study of the gut microbiome and its further multi-omics analysis was important for the prevention and treatment of CRC.

1. Introduction

Colorectal cancer (CRC) is a common malignancy of the digestive system that mainly originates from epithelial cells. It currently ranks third in incidence among common malignancies worldwide and is the second leading cause of tumor-related deaths.[1, 2] In recent years, the incidence of CRC has increased in many Asian countries including China.[3] It has become imperative to identify as many risk factors associated with CRC as possible for the prevention and treatment of CRC.

The human gastrointestinal tract hosts a large population of microorganisms that can interact with each other as well as with the intestinal microenvironment and other species in the environment. The relative abundance of certain gut microbiota may change under the influence of gene, drugs, and various metabolic and environmental factors, which can lead to a decrease in beneficial commensal flora and an increase in conditionally pathogenic and disease-causing bacteria,[4] causing further changes in flora metabolism that can lead to disease in the intestine or in other target organs through a series of complex mechanisms. Several animal models have found an association between intestinal flora and CRC. In a study by Wong et al, feces from CRC patients and non-CRC patients were fed to healthy mice by gavage, and the results showed that the ratio of Th1 to Th17 cells, level of inflammatory markers, number of polyps, and proliferation levels of intestinal mucosal cells were significantly higher in mice fed feces from CRC patients compared to controls.[5] The association between intestinal flora and CRC has also been found in CRC patients with familial adenomatous polyposis (FAP), a precancerous condition of hereditary CRC. Dejea et al. found *E. coli* that formed biofilms as the predominant flora in surgically resected tissue from the colon of FAP patients, demonstrating that intestinal flora can form biofilms that induce upregulation of colonic epithelial interleukin 17 expression, causing abnormal alteration of colonic epithelial DNA, heterogeneous proliferation of epithelial cells, and subsequent progression to malignant tumor.[6] However, it is difficult to prove the causal association between gut microbiota and CRC by randomized controlled trials due to confounding factors such as diet, lifestyle, and the underdeveloped technology used in fecal transplantation experiments. In addition, recent studies have found a correlation between abnormal expression of genes related to CRC occurrence and the abundance of pathogenic bacteria[7, 8]. However, most studies have focused only on the association between a limited number of genes and gut microbes or specific bacteria[9, 10]. Therefore, the association of host genes with the gut microbiome in CRC needs to be further discovered and studied.

Mendelian randomization (MR) uses genetic variants in non-experimental data to infer the causal effect of an exposure on an outcome. The idea of MR is to use genome-wide association studies (GWAS) to obtain single-nucleotide polymorphisms (SNPs) that exhibit strong correlations with specific outcomes that can serve as a tool to infer causal associations between exposure factors and outcomes. These SNPs can be used to test for causal associations between exposure factors and outcomes while avoiding the effects of confounding factors because they are based on random Mendelian genetic variation. Biological genotypes are formed by random assignment during meiosis, a process that is generally not influenced by external factors. We therefore conducted an MR study to evaluate the causal association of gut microbiota on CRC. Annotation of the SNPs of the intestinal flora validated by MR analysis can find associated genes.

2. Methods

2.1 Data sources

We obtained SNPs associated with gut microbial abundance from the MiBioGen consortium's GWAS study, which included 25 cohorts of 18,340 subjects from countries including the United States, Italy, and South Korea, and which focused on identifying genetic loci that influence the relative abundance of gut microbes by analyzing the 16SrRNA sequencing profiles of their subjects.[11] We obtained a dataset of genetic variants associated with CRC from a large GWAS study of East Asian populations, which included three cohorts with a total of 6692 CRC patients and 27278 controls.[12] In addition, we obtained the CRC risk-related dataset from the FinnGen consortium for validation, which included 7427 CRC patients and 25600 controls (**Table 1**).[13] The GWAS studies selected for this MR analysis were ethically approved, and materials such as informed consent forms were available in the supplemental materials of the respective original publications.

2.2 Study Design

Our overall study design was shown in Fig. 1. We screened eligible SNPs from the GWAS dataset of the MiBioGen consortium using specific criteria as instrumental variables (shown in 2.3) for the gut microbiota. As shown in the Fig. 2, our MR study design satisfied the three necessary assumptions,[14] and also followed the requirements of STROBE-MR.[15] (**Supplementary Table 1**)

2.3 Instrument selection

First, we screened for SNPs associated with bacterial abundance from the GWAS study at the locus-wide significance level ($P < 1 \times 10^{-5}$) for each bacterial taxa at four taxonomic levels: order, family, genus, and species. Second, we screened and removed SNPs located on chromosome 23 and also removed SNPs containing multiple alleles (> 2) to avoid unwanted effects on our MR analysis results. Third, we removed SNPs with a minor allele frequency (MAF) of less than 0.01. Fourth, we used samples from the 1000 Genomes European Project as a reference to examine the linkage disequilibrium (LD) between instrumental variables (IVs), following the criteria of $r^2 < 0.01$ and window size > 10,000 kb, thus avoiding the effect of LD between IVs. Fifth, some IVs may be strongly correlated ($P < 5 \times 10^{-8}$) with confounders or outcome events, referred to as horizontal pleiotropy, and the reliability of the results would be affected if these SNPs were included as instrumental variables for MR analysis.[16] Therefore, we obtained SNPs significantly associated with confounding characteristics (such as BMI and age) using PhenoScanner to preliminarily exclude the effect of horizontal pleiotropy. As a result, we did not detect SNPs with strong correlations with other confounding factors. Finally, we used SNPs that met all the above criteria as IVs for downstream MR analysis. We also screened for SNPs associated with gut microbial abundance from the GWAS study at a genome-wide significance level ($P < 5 \times 10^{-8}$) to include as IVs to make the analysis more comprehensive. The screening process for instrumental variables was shown in Fig. 3.

2.4 Efficacy estimation of instrumental variables

The regression R^2 value is often used in MR studies as a measure of how much the variance in the exposure outcome can be explained by the IVs. It is calculated as $R^2 = 2\times EAF\times(1-EAF)\times beta^2/(2\times EAF\times\ 1-EAF\ \times beta^2) + 2\times EAF\times\ 1-EAF\ \times se\times N\times beta^2$.[17, 18] Weak IVs in MR studies can cause bias in the causal association between exposure factors and outcome events. The F-statistic, derived from the regression of exposure outcomes on instrumental variables, can respond to the degree of correlation between exposure factors and outcomes and detect weak IVs. It is used to represent the degree of bias when estimating causal associations and is calculated using the formula $F = R^2\times\ N-2\ /(1-R^2)$, where N represents the sample size of the exposed data.[19] An F-statistic less than 10 indicates the presence of weakly predictive instruments. This is derived from the observation that when $F < 10$, the bias of the IV estimate is more than 10% of the bias in the observational association estimate (relative bias > 1/10).

2.5 Statistical Analysis

We first obtained eligible SNPs as IVs using the process outlined above. For bacterial taxa containing only one IV, we used the Wald ratio for MR analysis. For bacterial taxa containing multiple IVs, we used the inverse-variance-weighted (IVW) approach as the main analysis method to examine the correlation between bacterial taxa and CRC. The IVW method is commonly used for obtaining variant-specific causal estimates, and can combine the effect values of multiple IVs into one estimate and provide a more accurate analysis of the causal relationships among variables. We also used the weighted median method, MR-Egger, simple median method, and MR-PRESSO as complementary analysis methods. The weighted median method is characterized by consistent results even when the weight of invalid IVs reach 50% (or < 50%).[20] The MR-Egger method has relatively low statistical power,[21] similar to the IVW method, except that the regression model contains an intercept term $\theta 0$ and the p-value of this intercept term can help identify horizontal pleiotropy.[22] We also applied the MR-PRESSO global test to detect horizontal pleiotropy, which is implemented using a weighted regression of all the genetic variants and then computing a residual sum of squares (RSS). Each IV would be removed in turn and the corresponding RSS value would be calculated. If the RSS value decreased significantly from the previous iteration and reached statistical significance ($p < 0.05$), it would suggest that the SNP exhibited horizontal pleiotropy. We tested for outlier SNPs using the MR-PRESSO outlier test and recalculated the estimates after removing any outliers, thus avoiding pleiotropic effects on our MR analysis.[23]

We detected potential reverse causal associations between SNPs associated with the gut microbiota and CRC using the MR Steiger Filtering Test.[24] We used a series of sensitivity analyses to test the robustness of the results. We quantified heterogeneity by calculating Cochran's Q statistic, which considers a result to be heterogeneous if the p-value is less than 0.05.[25] The I^2 statistic can also be used to quantify the degree of heterogeneity, and is calculated as $I^2 = (Q-Q_df)/Q$. It can be assumed that there is heterogeneity if I^2 is greater than 25%.[25, 26] The results of the analysis, based on the random effects model of the IVW method, may be more reliable if there is a high degree of heterogeneity among SNPs.[27] We assessed the heterogeneity between variant-specific causal estimates using meta-analysis techniques and identified outliers using scatter and funnel plots. In addition, we performed Leave-one-out analysis on IVs, in which all IVs of bacterial taxa were removed one by one, and recalculated MR estimates using all remaining SNPs to examine the correlation between the gut microbiota and CRC.

We performed MR analysis with the FinnGen consortium dataset to verify the accuracy of our results and meta-analyzed the MR estimates from the FinnGen and MiBioGen datasets. We used the mRnd online tool to calculate statistical power,[28] which represents the ability to detect a particular magnitude of causal effect in a given sample size and should generally be greater than 80% to have confidence in the results. All statistical analyses were performed using the TwoSampleMR[29] and MR-PRESSO packages[23] in R4.2.0.[30]

2.6 SNP annotation

The online network tool was used for SNP annotation[31]. **g:SNPense** maps a list of human SNP rs-codes to gene names, receives chromosomal coordinates and predicted variant effects. Mapping is enabled only for variants that overlap with at least one protein coding Ensembl gene. All underlying data are retrieved from the Ensembl Variation data.

3. Results

3.1 Instrumental Variables Selection

11,237 SNPs at the locus-wide significance level ($P < 1 \times 10^{-5}$) and 1,035 SNPs at the genome-wide significance level ($P < 5 \times 10^{-8}$) were selected based on 166 bacterial features in the MiBioGen consortium. After identifying and removing SNPs in LD, the remaining 2,271 SNPs at the locus-wide significance level and 12 SNPs at the genome-wide significance level were used as IVs. We extracted the effect allele, other allele, beta, SE, and p-value of these SNPs for MR analysis.

3.2 Mendelian Randomization Analysis

3.2.1 Locus-wide significance level

The results of the IVW analysis showed that the family *Porphyromonadaceae* (OR = 1.26, 95% CI, 1.03–1.55, P = 0.0267), genera *Anaerotruncus* (OR = 1.17, 95% CI, 1.01–1.36, P = 0.0390), *Intestinibacter* (OR = 1.31, 95% CI, 1.09–1.57, P = 0.0038), *Slackia* (OR = 1.24, 95% CI, 1.06–1.45, P = 0.0071), and *Ruminococcaceae* UCG004 (OR = 1.27, 95% CI, 1.03–1.57, P = 0.0232), and species *Eubacterium coprostanoligenes* group (OR = 1.25, 95% CI, 1.00–1.56, P = 0.0467) exhibited significant causal associations with CRC risk. The results of weighted median method showed that the genus *Intestinibacter* (OR = 1.28, 95% CI, 1.00–1.64, P = 0.0520) significantly increased the risk of CRC. According to the results of the simple median method, genus *Intestinibacter* (OR = 1.39, 95% CI, 1.08–1.78, P = 0.0093) and species *Eubacterium coprostanoligenes* group (OR = 1.62, 95% CI, 1.14–2.30, P = 0.0073) were positively associated with CRC risk, which was consistent with the results of the IVW analysis. The MR estimates from supplementary analysis all supported their negative effect on CRC (**Table 2**). Details on the SNPs used as bacterial features are shown in **Supplementary Table 2**. The F-statistics of the SNPs were all greater than 10, indicating no weak IVs were included. MR analysis based on the FinnGen database showed that family *Porphyromonadaceae* (OR = 1.50, 95% CI, 1.11–2.03, P = 0.0079) and genus *Slackia* (OR = 1.17, 95% CI, 1.02–1.36, P = 0.0298) were risk factors for CRC (**Table 2**). We combined MR estimates from both the AGWAS and FinnGen databases by meta-analysis and found that genus *Anaerotruncus* (OR = 1.16, 95% CI, 1.01–1.33, P = 0.0303) and genus *Intestinibacter* (OR = 1.31, 95% CI, 1.12–1.52, P = 0.0005) were positively associated with CRC. However, we found no associations between genus *Ruminococcaceae* UCG004 (OR = 1.13, 95% CI, 0.96–1.32, P = 0.1560) and species *Eubacterium coprostanoligenes* group (OR = 1.09, 95% CI, 0.94–1.28, P = 0.2656) with CRC. In summary, we found that family *Porphyromonadaceae*, genus *Slackia*, genus *Anaerotruncus*, and genus *Intestinibacter* all exhibited a significant causal association with CRC risk (Fig. 4).

The results of the MR steiger filtering test (**Supplementary Table 3**) did not reveal an inverse causal association between the bacterial taxa mentioned previously and CRC. There was no significant heterogeneity among SNPs for gut microbiome-CRC association, with low heterogeneity among all SNPs that served as IVs in all bacterial taxa ($I^2 < 25\%$, p Cochran's Q > 0.01) except genus *Slackia*($I^2 = 39\%$, p Cochran's Q = 0.11)and genus *Anaerotruncus* $I^2 = 45\%$, p Cochran's Q = 0.06 (**Table 3**). Visualized scatter and funnel plots are shown in **Supplemental Fig. 1–12**. Neither the Egger Intercept test nor the MR-PRESSO Global test detected significant horizontal pleiotropy. Similarly, the MR-PRESSO outlier test did not find any outlier SNPs that could lead to horizontal pleiotropy. The results of the Leave-one-out analyses showed no significant effect of individual SNPs on gut microbiome-CRC association. We had 97%, 99%, 72%, and 100% statistical power to detect ORs of 1.26, 1.24, 1.17, and 1.31 for associations of family *Porphyromonadaceae*, genus *Slackia*, genus *Anaerotruncus*, and genus *Intestinibacter* with CRC in the MiBioGen consortium, respectively. We had 100%, 99%, 60%, and 97% statistical power to detect the corresponding ORs of 1.41, 1.23, 1.07, and 1.24 in FinnGen.

3.2.2 Genome-Wide statistical significance level

We first performed MR analysis of the 12 eligible SNPs in aggregate using IVW (OR = 1.01, 95% CI, 0.88–1.15, P = 0.9062), the weighted median method (OR = 0.96, 95% CI, 0.79–1.16, P = 0.6493), MR Egger (OR = 0.79, 95% CI, 0.46–1.35, P = 0.4124), and the simple median method (OR = 1.12, 95% CI, 0.93–1.35, P = 0.2284), none of which suggested that gut microbes were associated with CRC risk. Heterogeneity among IVs was low (p Cochran's Q = 0.5720, $I^2 = 0$), and the Egger intercept test and the MR-PRESSO Global Test results showed no significant levels of pleiotropy (Egger intercept p = 0.3820, MR-PRESOO global test p = 0.604). We did not find any bacterial taxa associated with CRC risk (**Table 4.**), We could not perform further tests for heterogeneity and pleiotropy because the number of IVs in each bacterial feature was less than 2.

3.3 SNP annotation

We annotated the SNPs at a locus-wide significance level of the four intestinal flora and identified 24 host genes that may be related to pathogenic intestinal microflora in CRC patients. (**Table 5**)

4. Discussion

The human intestine is a diverse and nutrient-rich micro-ecological system, consisting of 100 trillion microbes mixed with digestive secretions, epithelial cells, and food-borne abiotic components. The intestinal flora regulates itself in healthy individuals to maintain the balance among the intestinal micro-ecological system while providing energy for the body through the digestion and absorption of food. The results from studies on intestinal flora in recent years have shown that changes in the structure, abundance, and function of intestinal flora are closely associated with many diseases including CRC.[32] There are significant differences in the number and species of intestinal flora between CRC patients and healthy people.[33] The degree of intestinal flora imbalance is positively correlated with the progression rate of CRC.[34] Several observational studies have found significant differences in gut flora composition between healthy patients and CRC patients at different stages of the disease from proliferative polyps and early cancer to metastatic malignancies, supporting the role of gut flora in the development of CRC.[35] However, other risk factors for CRC such as obesity, diet, lifestyle, and geography can also influence the composition of the gut microbiome. We thus do not know whether the alterations in the gut microbiome in CRC patients is secondary to the tumor or an active process that contributes to tumorigenesis. This potential reverse causal association prevents us from determining the direction of effect of the gut microbiome on CRC. In addition, previous studies have shown that microbiota can influence gene expression and that gene expression correlates with the abundance of gut microbiota, but studies on the association between broad gut microbiota and genes in CRC are limited [36, 37].We conducted this study to explore the causal association of the gut microbiome on CRC and identify possible associations between pathogenic bacteria and host genes in CRC. The results of the meta-analysis based on combining the MR estimates from the AGWAS and FinnGen datasets showed that the family *Porphyromonadaceae* and genera *Slackia*, *Anaerotruncus*, and *Intestinibacter* have a direct causal association on CRC.

The family *Porphyromonadaceae* contains a variety of genera such as *Parabacteroides*, *Odoribacter* and *Porphyromonas* that are rarely seen in healthy populations.[38] Zackular et al constructed a mouse model that replicated the progression of CRC from chronic inflammation to heterogeneous hyperplasia to adenocarcinoma.[39] Their analysis of the gut microbiome composition of the mouse model showed a significantly elevated abundance of genus *Odoribacter* (belonging to family *Porphyromonadaceae*).[40] Baxter et al analyzed the gut microbial composition of the feces of several CRC patients (serving as the experimental group) and that of healthy individuals (serving as the control group), and then transplanted the feces into healthy mice to observe the differences in the number of tumors in the mice. The results showed a positive correlation between the genus *Parabacteroides* (belonging to family *Porphyromonadaceae*) and the incidence of CRC in the experimental group in contrast to the control group.[41] These studies suggest a pathogenic role of family *Porphyromonadaceae* in CRC, on the basis of which our study further revealed its causal association to CRC. However, because the family *Porphyromonadaceae* is relatively rare, research on its pathogenic mechanisms is limited and further studies on its role in the development of CRC are needed in the future.

For the genus *Anaerotruncus*, Loke et al compared intestinal microbial composition and metabolomic differences between paired tumor tissue and normal tissue in 17 Asian CRC patients and found that the relative abundance of genus *Anaerotruncus* could influence steroid and terpene biosynthesis as well as bile metabolism, resulting in increased tumor-associated metabolites such as S-Adenosylmethionine (SAM) and S-Adenosyl-Homocysteine (SAH)[42]. Similarly, Satoh et al identified significantly higher levels of SAM in tumor tissues of CRC patients compared to normal tissues.[43] Loke et al revealed that gut microbiota dysbiosis caused local metabolic abnormalities at the primary tumor site, leading to significant upregulation of SAH levels.[42] Sibani et al. found that SAM and SAH levels were positively correlated with tumor number in animal models and could be used as a measure of abnormal cell transformation.[44] In addition, *Anaerotruncus* stimulates an increase in lipopolysaccharides (LPS) in humans which can disrupt the integrity of gastrointestinal epithelial cells and lead to impaired intestinal mucosal barrier function. Upregulated LPS promotes the release of pro-inflammatory cytokines and inhibits tight junction proteins, increasing oxidative stress and abnormal differentiation of colorectal epithelial cells.[45, 46] Enterotoxigenic *Bacteroides fragilis* (ETBF) is a Gram-negative anaerobic bacterium and Liu et al [47] found that increased abundance of ETBF was closely associated with colorectal cancer. ETBF can produce *B. fragilis* toxin (BFT), which when bound to intestinal mucosal epithelial receptors, can promote the activation of Wnt and NF-KB signaling pathways, facilitating cell proliferation and DNA damage, leading to abnormal cell transformation.[48–51] ETBF can also cause the release of reactive oxygen species from inflammatory cells and promote the expression of cytokines and chemokines, leading to DNA damage which in turn promotes the development of CRC. These findings suggest that the genus Anaerotruncus plays an important role in the pathogenesis of CRC and can influence host gene expression, which is consistent with our results. Therefore, we speculate that the altered relative abundance of the genus *Anaerotruncus* affects local metabolism, leading to increased levels of metabolites such as SAM and SAH, which in turn cause host gene damage and results in the transformation of normal cells to tumors.

Similarly, previous studies have found that genera *Slackia* and *Intestinibacter* are associated with CRC. Huo et al compared the gut microbial composition of tissue samples from patients with and without CRC recurrence and found that the relative abundance of genus *Slackia* was significantly higher in patients with CRC recurrence than in patients without recurrence, suggesting that it is a potential biomarker for prognosis in CRC patients.[52] For genus *Intestinibacter*, many studies have found a significant increase in the abundance of this bacterium both in animal models with CRC and in the fecal and mucosal tissues of CRC patients.[40, 41, 53] For example, *Fusobacterium nucleatum* (FN) (belonging to genus *Intestinibacter*) can be involved in the development and metastasis of CRC through multiple mechanisms. KOSTIC et al found that *Clostridium perfringens* suppressed anti-tumor immune responses by recruiting myeloid suppressor cells, tumor-associated macrophages, and regulatory T cells.[54]

Previous observational studies have found an association between the gut microbiota and CRC, but the results cannot be used as evidence to support a direct causal association due to the influence of certain confounding factors such as the environment, diet. The significant advantage of our MR study is the selection of genetic variants significantly associated with the composition of the gut microbiota as IVs, which do not directly contribute to CRC and are not influenced by other risk factors for CRC. This means that any association between IVs with CRC must arise via the variant's association with the gut microbiota, thus implying a causal effect of the gut microbiota on CRC. We identified host genes that may be associated with the abundance of gut microbes by SNP annotation. Sambhawa Priya et al [55]used multi-omics integration to identify human disease-specific host gene-microbiome associations and found

that gut microbes are associated with individual host genes. 755 significant host gene-microbial associations were identified in CRC, including the PCSK5 gene, which was consistent with our findings, and on this basis, we found that this gene may be associated with genus Anaerotruncus abundance

However, there are still unavoidable limitations of the present MR study. First, our MR analysis based on IVs at the genome-wide statistical significance level ($P < 5 \times 10^{-8}$) did not identify any causal association of the gut microbiome on CRC. All causal associations revealed by our MR study were obtained based on IVs at the locus-wide significance level ($P < 1 \times 10^{-5}$), which may have an impact on the accuracy of the results. Second, the causal association of genus *Anaerotruncus* on CRC did not reach the desired statistical power threshold of 80%, so the correlation needs to be further clarified. Third, since detailed baseline characteristics of study subjects (e.g., age, tumor markers, tumor stage, etc.) were not provided in the GWAS study of CRC, we could not further investigate the effect of gut microbiome on different subgroups of the population. Fourth, although we identified possible gene-gut microbiome associations through SNP annotation, future studies such as animal models are needed to further validate the causal association between genetic alterations and pathogenic microbiome.

In conclusion, this MR study demonstrates that several gut microbes are positively associated with CRC risk and can serve as potential biomarkers, on the basis of which this study also identified possible gene-gut microbiome associations in CRC. Further study of the pathogenic mechanisms of these intestinal flora and verification of the causal association between host genetic alterations and intestinal flora abundance will be important for optimizing the diagnosis and treatment of CRC in the future.

5. List Of Abbreviations

CRC: colorectal cancer

MR: Mendelian randomization

GWAS: Genome-wide association study

AGWAS: Asian Population Genome-wide association study

IVW: inverse-variance-weighted

FAP: familial adenomatous polyposis

SNP: single-nucleotide polymorphisms

MAF: minor allele frequency

LD: linkage disequilibrium

SE: standard error

EAF: effect allele frequency

RSS: residual sum of squares

SAM: S-Adenosylmethionine

SAH: S-Adenosyl-Homocysteine

LPS: lipopolysaccharide

ETBF: Enterotoxigenic Bacteroides Fragilis

BFT: B. fragilis toxin

6. Declarations

Ethics approval and consent to participate

All studies were previously approved by respective institutional review boards (IRBs). No new IRB approval was required. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Availability of data and materials

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author/s.

Consent for publication

Not applicable

Competing interests

All authors declare that they have no conflict of interest.

Funding

This work was supported by Beijing Municipal Education Commission Science and Technology Project (KM202010025005); The Capital Health Research and Development of Special Projects (2022-2-7083); Beijing Municipal Natural Science Foundation (7222100).

Authors' contributions

DY and YYT conceived the presented idea. YX, CZ performed the computations and manuscript writing. YX, CZ, YR, LW and JW were involved in acquisition of data. YX was involved in interpretation of data. All authors contributed to the article and approved the submitted version.

Acknowledgements

We appreciate all the volunteers and patients who participated in this study. We are grateful to the MiBioGen consortium study for releasing the gut microbiota GWAS summary statistics, East Asian Population genetic consortia study and FinnGen consortium, for releasing the CRC GWAS summary statistics.

Author details

1 Department of Oncology, Beijing Luhe Hospital Affiliated to Capital Medical University, Beijing, 101149 China.

Publisher's Note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. BRAY F, FERLAY J, SOERJOMATARAM I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries [J]. CA: a cancer journal for clinicians, 2018, 68(6): 394-424.
2. SIEGEL R L, MILLER K D. Cancer statistics, 2023 [J]. 2023, 73(1): 17-48.
3. CHEN W, ZHENG R, ZENG H, et al. Annual report on status of cancer in China, 2011 [J]. Chinese journal of cancer research = Chung-kuo yen cheng yen chiu, 2015, 27(1): 2-12.
4. GAINES S, SHAO C, HYMAN N, et al. Gut microbiome influences on anastomotic leak and recurrence rates following colorectal cancer surgery [J]. The British journal of surgery, 2018, 105(2): e131-e41.
5. WONG S H, ZHAO L, ZHANG X, et al. Gavage of Fecal Samples From Patients With Colorectal Cancer Promotes Intestinal Carcinogenesis in Germ-Free and Conventional Mice [J]. Gastroenterology, 2017, 153(6): 1621-33 e6.
6. DEJEA C M, FATHI P, CRAIG J M. Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria [J]. 2018, 359(6375): 592-7.
7. DAYAMA G, PRIYA S, NICCUM D E, et al. Interactions between the gut microbiome and host gene regulation in cystic fibrosis [J]. Genome medicine, 2020, 12(1): 12.
8. FLEMER B, LYNCH D B, BROWN J M, et al. Tumour-associated and non-tumour-associated microbiota in colorectal cancer [J]. Gut, 2017, 66(4): 633-43.
9. H SLER R, SHEIBANI-TEZERJI R, SINHA A, et al. Uncoupling of mucosal gene regulation, mRNA splicing and adherent microbiota signatures in inflammatory bowel disease [J]. Gut, 2017, 66(12): 2087-97.
10. LLOYD-PRICE J, ARZE C, ANANTHAKRISHNAN A N, et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases [J]. Nature, 2019, 569(7758): 655-62.
11. KURILSHIKOV A, MEDINA-GOMEZ C, BACIGALUPE R, et al. Large-scale association analyses identify host factors influencing human gut microbiome composition [J]. Nature genetics, 2021, 53(2): 156-65.
12. TANIKAWA C, KAMATANI Y, TAKAHASHI A, et al. GWAS identifies two novel colorectal cancer loci at 16q24.1 and 20q13.12 [J]. Carcinogenesis, 2018, 39(5): 652-60.
13. FinnGen_Consortium. FinnGen Data [M]. Available online:https://www.finngen.fi/.
14. BURGESS S, SCOTT R A, TIMPSON N J, et al. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors [J]. European journal of epidemiology, 2015, 30(7): 543-52.
15. SKRIVANKOVA V W, RICHMOND R C, WOOLF B A R, et al. Strengthening the Reporting of Observational Studies in Epidemiology Using Mendelian Randomization: The STROBE-MR Statement [J]. Jama, 2021, 326(16): 1614-21.
16. KAMAT M A, BLACKSHAW J A, YOUNG R, et al. PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations [J]. Bioinformatics (Oxford, England), 2019, 35(22): 4851-3.

17. PAPADIMITRIOU N, DIMOU N, TSILIDIS K K, et al. Physical activity and risks of breast and colorectal cancer: a Mendelian randomisation analysis [J]. Nature communications, 2020, 11(1): 597.
18. SHIM H, CHASMAN D I, SMITH J D, et al. A multivariate genome-wide association analysis of 10 LDL subfractions, and their response to statin treatment, in 1868 Caucasians [J]. PloS one, 2015, 10(4): e0120758.
19. BURGESS S, THOMPSON S G. Avoiding bias from weak instruments in Mendelian randomization studies [J]. International journal of epidemiology, 2011, 40(3): 755-64.
20. BOWDEN J, DAVEY SMITH G, HAYCOCK P C, et al. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator [J]. Genetic epidemiology, 2016, 40(4): 304-14.
21. BURGESS S, THOMPSON S G. Interpreting findings from Mendelian randomization using the MR-Egger method [J]. European journal of epidemiology, 2017, 32(5): 377-89.
22. BOWDEN J, DAVEY SMITH G, BURGESS S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression [J]. International journal of epidemiology, 2015, 44(2): 512-25.
23. VERBANCK M, CHEN C Y. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases [J]. 2018, 50(5): 693-8.
24. HEMANI G, TILLING K, DAVEY SMITH G. Correction: Orienting the causal relationship between imprecisely measured traits using GWAS summary data [J]. PLoS genetics, 2017, 13(12): e1007149.
25. GRECO M F, MINELLI C, SHEEHAN N A, et al. Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome [J]. Statistics in medicine, 2015, 34(21): 2926-40.
26. LU Y, XU Z, GEORGAKIS M K, et al. Smoking and heart failure: a Mendelian randomization and mediation analysis [J]. ESC heart failure, 2021, 8(3): 1954-65.
27. BOWDEN J, DEL GRECO M F, MINELLI C, et al. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization [J]. Statistics in medicine, 2017, 36(11): 1783-802.
28. BRION M J, SHAKHBAZOV K, VISSCHER P M. Calculating statistical power in Mendelian randomization studies [J]. International journal of epidemiology, 2013, 42(5): 1497-501.
29. HEMANI G, ZHENG J. The MR-Base platform supports systematic causal inference across the human phenome [J]. 2018, 7(
30. R_Core_Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna,Austria. [M]. Available online: https://www.R-project.org/
31. g:Profile :an ELIXIR Recommended Interoperability Resource. https://biit.cs.ut.ee/gprofiler/snpense. Accessed 12 May 2023 [J].
32. LYNCH S V, PEDERSEN O. The Human Intestinal Microbiome in Health and Disease [J]. The New England journal of medicine, 2016, 375(24): 2369-79.
33. LIU W, ZHANG R. Study of the Relationship between Microbiome and Colorectal Cancer Susceptibility Using 16SrRNA Sequencing [J]. 2020, 2020(7828392.
34. MIRA-PASCUAL L, CABRERA-RUBIO R, OCON S, et al. Microbial mucosal colonic shifts associated with the development of colorectal cancer reveal the presence of different bacterial and archaeal biomarkers [J]. Journal of gastroenterology, 2015, 50(2): 167-79.
35. AHN J, SINHA R, PEI Z, et al. Human gut microbiome and risk for colorectal cancer [J]. Journal of the National Cancer Institute, 2013, 105(24): 1907-11.
36. CAMP J G, FRANK C L, LICKWAR C R, et al. Microbiota modulate transcription in the intestinal epithelium without remodeling the accessible chromatin landscape [J]. Genome research, 2014, 24(9): 1504-16.
37. RICHARDS A L, MUEHLBAUER A L, ALAZIZI A, et al. Gut Microbiota Has a Widespread and Modifiable Effect on Host Gene Regulation [J]. 2019, 4(5):
38. WU N, YANG X, ZHANG R, et al. Dysbiosis signature of fecal microbiota in colorectal cancer patients [J]. Microbial ecology, 2013, 66(2): 462-70.
39. DE ROBERTIS M, MASSI E, POETA M L, et al. The AOM/DSS murine model for the study of colon carcinogenesis: From pathways to diagnosis and therapy studies [J]. Journal of carcinogenesis, 2011, 10(9.
40. ZACKULAR J P, BAXTER N T, IVERSON K D, et al. The gut microbiome modulates colon tumorigenesis [J]. mBio, 2013, 4(6): e00692-13.
41. BAXTER N T, ZACKULAR J P, CHEN G Y, et al. Structure of the gut microbiome following colonization with human feces determines colonic tumor burden [J]. Microbiome, 2014, 2(20.
42. LOKE M F, CHUA E G, GAN H M, et al. Metabolomics and 16S rRNA sequencing of human colorectal cancers and adjacent mucosa [J]. PloS one, 2018, 13(12): e0208584.
43. SATOH K, YACHIDA S, SUGIMOTO M, et al. Global metabolic reprogramming of colorectal cancer occurs at adenoma stage and is induced by MYC [J]. Proceedings of the National Academy of Sciences of the United States of America, 2017, 114(37): E7697-e706.
44. SIBANI S, MELNYK S, POGRIBNY I P, et al. Studies of methionine cycle intermediates (SAM, SAH), DNA methylation and the impact of folate deficiency on tumor numbers in Min mice [J]. Carcinogenesis, 2002, 23(1): 61-5.
45. BAIL N M, BRESSA C, MART NEZ-L PEZ S, et al. Microbiota Features Associated With a High-Fat/Low-Fiber Diet in Healthy Adults [J]. Frontiers in nutrition, 2020, 7(583608.
46. GAO Z, WU H, ZHANG K, et al. Protective effects of grape seed procyanidin extract on intestinal barrier dysfunction induced by a long-term high-fat diet [J]. Journal of Functional Foods, 2020, 64(103663.
47. LIU C J, ZHANG Y L, SHANG Y, et al. Intestinal bacteria detected in cancer and adjacent tissue from patients with colorectal cancer [J]. Oncology letters, 2019, 17(1): 1115-27.

48. GOODWIN A C, DESTEFANO SHIELDS C E, WU S, et al. Polyamine catabolism contributes to enterotoxigenic Bacteroides fragilis-induced colon tumorigenesis [J]. Proceedings of the National Academy of Sciences of the United States of America, 2011, 108(37): 15354-9.
49. WU S, LIM K C, HUANG J, et al. Bacteroides fragilis enterotoxin cleaves the zonula adherens protein, E-cadherin [J]. Proceedings of the National Academy of Sciences of the United States of America, 1998, 95(25): 14979-84.
50. WU S, POWELL J, MATHIOUDAKIS N, et al. Bacteroides fragilis enterotoxin induces intestinal epithelial cell secretion of interleukin-8 through mitogen-activated protein kinases and a tyrosine kinase-regulated nuclear factor-kappaB pathway [J]. Infection and immunity, 2004, 72(10): 5832-9.
51. WU S, SHIN J, ZHANG G, et al. The Bacteroides fragilis toxin binds to a specific intestinal epithelial cell receptor [J]. Infection and immunity, 2006, 74(9): 5382-90.
52. HUO R X, WANG Y J, HOU S B, et al. Gut mucosal microbiota profiles linked to colorectal cancer recurrence [J]. World journal of gastroenterology, 2022, 28(18): 1946-64.
53. ZHU Q, JIN Z, WU W, et al. Analysis of the intestinal lumen microbiota in an animal model of colorectal cancer [J]. PloS one, 2014, 9(6): e90849.
54. KOSTIC A D, GEVERS D, PEDAMALLU C S, et al. Genomic analysis identifies association of Fusobacterium with colorectal carcinoma [J]. Genome research, 2012, 22(2): 292-8.
55. PRIYA S, BURNS M B, WARD T, et al. Identification of shared and disease-specific host gene-microbiome associations across human diseases using multi-omic integration [J]. Nature microbiology, 2022, 7(6): 780-95.

Tables

Table 1. Detailed information of studies and datasets used for analyses.

Data Source	Phenotype	Sample Size	Cases	Population	Adjustment
MiBioGen consortium	gut microbial	18340	/	the United States, Canada, Israel, South Korea, Germany, Denmark, the Netherlands, Belgium, Sweden, Finland and the United Kingdom	age, sex, technical covariates, and genetic principal components
AGWAS	CRC	33970	6692	Asian	age, sex
FinnGen	CRC	33027	7427	European	age, sex

Table 2. MR results of causal links between gut microbiome and CRC risk ($P < 1 \times 10^{-5}$).

Data source	Classification		Nsnp	Methods	OR (95%CI)	p−value
MiBioGen consortium	family	Porphyromonadaceae.id.943	11	Inverse variance weighted (fixed effects)	1.26(1.03,1.55)	0.0267
				Inverse variance weighted (multiplicative random effects)	1.26(1.02,1.56)	0.0337
				Weighted median	1.25(0.93,1.67)	0.1337
				MR Egger	1.28(0.83,1.96)	0.2923
				Simple median	1.4(0.94,2.08)	0.1003
MiBioGen consortium	genus	Anaerotruncus.id.2054	10	Inverse variance weighted (fixed effects)	1.17(0.96,1.42)	0.1121
				Inverse variance weighted (multiplicative random effects)	1.17(1.01,1.36)	0.0390
				Weighted median	1.14(0.88,1.49)	0.3184
				MR Egger	1.08(0.63,1.85)	0.7807
				Simple median	1.15(0.87,1.51)	0.3265
MiBioGen consortium		Intestinibacter.id.11345	10	Inverse variance weighted (fixed effects)	1.31(1.09,1.57)	0.0038
				Inverse variance weighted (multiplicative random effects)	1.31(1.14,1.5)	0.0001
				Weighted median	1.28(1,1.64)	0.0520
				MR Egger	1.06(0.5,2.26)	0.8849
				Simple median	1.39(1.08,1.78)	0.0093
MiBioGen consortium		Slackia.id.825	9	Inverse variance weighted (fixed effects)	1.24(1.06,1.45)	0.0071
				Inverse variance weighted (multiplicative random effects)	1.24(1.01,1.51)	0.0357
				Weighted median	1.15(0.91,1.44)	0.2363
				MR Egger	0.62(0.24,1.64)	0.3692
				Simple median	1.14(0.88,1.48)	0.3161
MiBioGen consortium		RuminococcaceaeUCG004.id.11362	9	Inverse variance weighted (fixed effects)	1.27(1.03,1.57)	0.0232
				Inverse variance weighted (multiplicative random effects)	1.27(1.07,1.51)	0.0053
				Weighted median	1.30(0.99,1.71)	0.0580
				MR Egger	2.09(0.62,7.13)	0.2754
				Simple median	1.32(0.99,1.74)	0.0563
MiBioGen consortium		Eubacteriumcoprostanoligenesgroup.id.11375	12	Inverse variance weighted (fixed effects)	1.25(0.99,1.58)	0.0583
				Inverse variance weighted (multiplicative random effects)	1.25(1.00,1.56)	0.0467
				Weighted median	1.28(0.92,1.79)	0.1387
				MR Egger	0.86(0.31,2.38)	0.7746
				Simple median	1.62(1.14,2.30)	0.0073
FinnGen	family	Porphyromonadaceae.id.943	11	Inverse variance weighted (fixed effects)	1.50(1.11,2.03)	0.0079
				Inverse variance weighted (multiplicative random effects)	1.50(1.18,1.92)	0.0011
				Weighted median	1.44(0.95,2.2)	0.0892
				MR Egger	1.51(0.7,3.23)	0.3177
				Simple median	1.42(0.93,2.17)	0.1062
FinnGen	genus	Anaerotruncus.id.2054	10	Inverse variance weighted (fixed effects)	1.12(0.81,1.55)	0.4987

				Inverse variance weighted (multiplicative random effects)	1.12(0.72,1.73)	0.6149
				Weighted median	0.91(0.56,1.49)	0.7151
				MR Egger	1.64(0.38,7.03)	0.5247
				Simple median	0.89(0.55,1.45)	0.6506
FinnGen		Intestinibacter.id.11345	10	Inverse variance weighted (fixed effects)	1.30(0.99,1.71)	0.0610
				Inverse variance weighted (multiplicative random effects)	1.30(0.98,1.72)	0.0641
				Weighted median	1.27(0.86,1.88)	0.2207
				MR Egger	2.12(0.6,7.54)	0.2790
				Simple median	1.35(0.93,1.97)	0.1110
FinnGen		Slackia.id.825	9	Inverse variance weighted (fixed effects)	1.17(0.94,1.46)	0.1557
				Inverse variance weighted (multiplicative random effects)	1.17(1.02,1.36)	0.0298
				Weighted median	1.24(0.94,1.64)	0.1302
				MR Egger	0.56(0.16,1.98)	0.4003
				Simple median	1.24(0.92,1.67)	0.1514
FinnGen		RuminococcaceaeUCG004.id.11362	9	Inverse variance weighted (fixed effects)	0.94(0.73,1.22)	0.6549
				Inverse variance weighted (multiplicative random effects)	0.94(0.72,1.23)	0.6687
				Weighted median	0.79(0.56,1.12)	0.1839
				MR Egger	0.86(0.21,3.5)	0.8353
				Simple median	0.79(0.55,1.14)	0.2153
FinnGen	species	Eubacteriumcoprostanoligenesgroup.id.11375	12	Inverse variance weighted (fixed effects)	0.96(0.71,1.30)	0.7951
				Inverse variance weighted (multiplicative random effects)	0.96(0.77,1.19)	0.7138
				Weighted median	0.86(0.57,1.28)	0.4482
				MR Egger	1.05(0.25,4.35)	0.9509
				Simple median	0.90(0.61,1.34)	0.6130

Table 3. Evaluation of heterogeneity and directional pleiotropy using different methods.

Data source	Classification	Bacterial taxas	Heterogeneity			Horizontal pleiotropy			
			I^2 (%)	Cochran's Q	p-value	Egger intercept	SE	Pvalue	MR−PRESSO Global Test p
MiBioGen consortium	family	Porphyromonadaceae.id.943	8	10.89	0.37	0.00	0.02	0.94	0.5
MiBioGen consortium	genus	Anaerotruncus.id.2054	0	5.33	0.80	0.01	0.02	0.76	0.797
MiBioGen consortium	genus	Intestinibacter.id.11345	0	5.17	0.82	0.02	0.03	0.59	0.844
MiBioGen consortium	genus	RuminococcaceaeUCG004.id.11362	0	5.31	0.72	-0.04	0.05	0.44	0.779
MiBioGen consortium	genus	Slackia.id.825	39	13.13	0.11	0.07	0.05	0.20	0.134
MiBioGen consortium	species	Eubacteriumcoprostanoligenesgroup.id.11375	0	10.88	0.54	0.03	0.03	0.47	0.571
FinnGen	family	Porphyromonadaceae.id.943	0	6.64	0.76	0.00	0.03	0.99	0.80
FinnGen	genus	Anaerotruncus.id.2054	45	16.28	0.06	-0.03	0.05	0.60	0.07
FinnGen	genus	Intestinibacter.id.11345	2	9.22	0.42	-0.04	0.05	0.46	0.45
FinnGen	genus	RuminococcaceaeUCG004.id.11362	8	8.73	0.37	0.01	0.06	0.89	0.40
FinnGen	genus	Slackia.id.825	0	3.41	0.91	0.07	0.06	0.28	0.92
FinnGen	species	Eubacteriumcoprostanoligenesgroup.id.11375	0	5.51	0.90	-0.01	0.04	0.91	0.90

Table 4. MR results of causal links between gut microbiome and CRC risk ($P < 5 \times 10^{-8}$).

Data source	Classification		Nsnp	Methods	OR (95%CI)	p–value	Heterogeneity		
							I^2 (%)	Cochran's Q	p-value
MiBioGen consortium	total		12	Inverse variance weighted (fixed effects)	1.01(0.88,1.15)	0.906175	0	9.420471	0.571954
				Inverse variance weighted (multiplicative random effects)	1.01(0.89,1.14)	0.898653			
				Weighted median	0.96(0.79,1.16)	0.64935			
				MR Egger	0.79(0.46,1.35)	0.412427			
				Simple median	1.12(0.93,1.35)	0.228353			
MiBioGen consortium	family	Peptostreptococcaceae	1	Wald ratio	4.74(0.81,27.82)	0.085157	–	–	–
MiBioGen consortium	genus	Oxalobacter.id	1	Wald ratio	0.81(0.60,1.11)	0.198543	–	–	–
		Enterorhabdus.id	1	Wald ratio	1.11(0.74,1.66)	0.617075	–	–	–
		Erysipelatoclostridium	1	Wald ratio	1.20(0.73,1.99)	0.4751	–	–	–
		RuminococcaceaeUCG009	1	Wald ratio	1.36(0.87,2.12)	0.1755	–	–	–
		Bifidobacterium	1	Wald ratio	0.88(0.59,1.33)	0.5427	–	–	–
		RuminococcaceaeUCG013	1	Wald ratio	1.29(0.41,4.13)	0.6629	–	–	–
		Tyzzerella3	1	Wald ratio	0.92(0.63,1.34)	0.6599	–	–	–
		Allisonella	1	Wald ratio	0.9(0.54,1.5)	0.6729	–	–	–
MiBioGen consortium	order	Bifidobacteriales	1	Wald ratio	0.88(0.58,1.33)	0.5427	–	–	–
MiBioGen consortium	species	Eubacteriumcoprostanoli genesgroup	1	Wald ratio	0.8(0.44,1.46)	0.4634	–	–	–

Table 5. SNP annotation of intestinal flora IVs

		id	chr	start	end	strand	gene_ids	gene_names
family	Porphyromonadaceae	rs10119172		-1	-1		ENSG00000264615	RN7SL592P
		rs1029811		-1	-1		/	/
		rs10762312	10	69812107	69812107	+	ENSG00000197467 ENSG00000289193	COL13A1 /
		rs10858364		-1	-1		/	/
		rs12700163	7	2609042	2609042	+	ENSG00000106012	IQCE
		rs17065783	3	62049912	62049912	+	ENSG00000144724	PTPRG
		rs2066088	1	1.65E+08	1.65E+08	+	ENSG00000185630	PBX1
		rs2401072		-1	-1		/	/
		rs35233670	17	65754785	65754785	+	ENSG00000154240	CEP112
		rs35961441	1	2.41E+08	2.41E+08	+	ENSG00000226919 ENSG00000182901	/ RGS7
		rs7330827		-1	-1		/	/
genus	Slackia	rs1006200		-1	-1		/	/
		rs10409783	19	4555774	4555774	+	ENSG00000167680	SEMA6B
		rs11957560	5	31268861	31268861	+	ENSG00000113361 ENSG00000254138	CDH6 /
		rs12440440	15	33749695	33749695	+	ENSG00000198838	RYR3
		rs16894137		-1	-1		/	/
		rs35156985	7	99854092	99854092	+	ENSG00000021461	CYP3A43
		rs4492265	7	13484058	13484058	+	ENSG00000229618	/
		rs7710333	5	1.78E+08	1.78E+08	+	ENSG00000246596 ENSG00000290968	/ /
		rs8901	17	76270929	76270929	+	ENSG00000185262	UBALD2
	Anaerotruncus	rs10150232	14	29948802	29948802	+	ENSG00000184304, ENSG00000257904	PRKD1 /
		rs11018566		-1	-1		/	/
		rs12056802	8	73800865	73800865	+	ENSG00000104343 ENSG00000258677	UBE2W /
		rs1272208	9	76015978	76015978	+	ENSG00000099139	PCSK5
		rs1431492	3	1.51E+08	1.51E+08	+	ENSG00000144893	MED12L
		rs4669806	2	12060626	12060626	+	ENSG00000224184	MIR3681HG
		rs6563550	13	37484276	37484276	+	ENSG00000230390	LINC01048
		rs7675045	4	1.72E+08	1.72E+08	+	ENSG00000174473	GALNTL6
		rs7963258	12	1.13E+08	1.13E+08	+	ENSG00000089169	RPH3A
		rs9347879		-1	-1		/	/
	Intestinibacter	rs10805326		-1	-1		/	/
		rs11109097	12	97534659	97534659	+	ENSG00000255794	RMST
		rs112879476		-1	-1		/	/
		rs16938435	9	21502924	21502924	+	ENSG00000171889	MIR31HG
		rs2098844		-1	-1		/	/
		rs2702387		-1	-1		/	/
		rs4327025	15	91903453	91903453	+	ENSG00000176463	SLCO3A1
		rs447950		-1	-1		/	/
		rs6875660	5	1.6E+08	1.6E+08	+	ENSG00000135083	CCNJL
		rs9348442		-1	-1		/	/

Supplementary Materials

The Supplementary Materials, Tables S1-S3 and Figures S1-S12 are not available with this version.

Figures

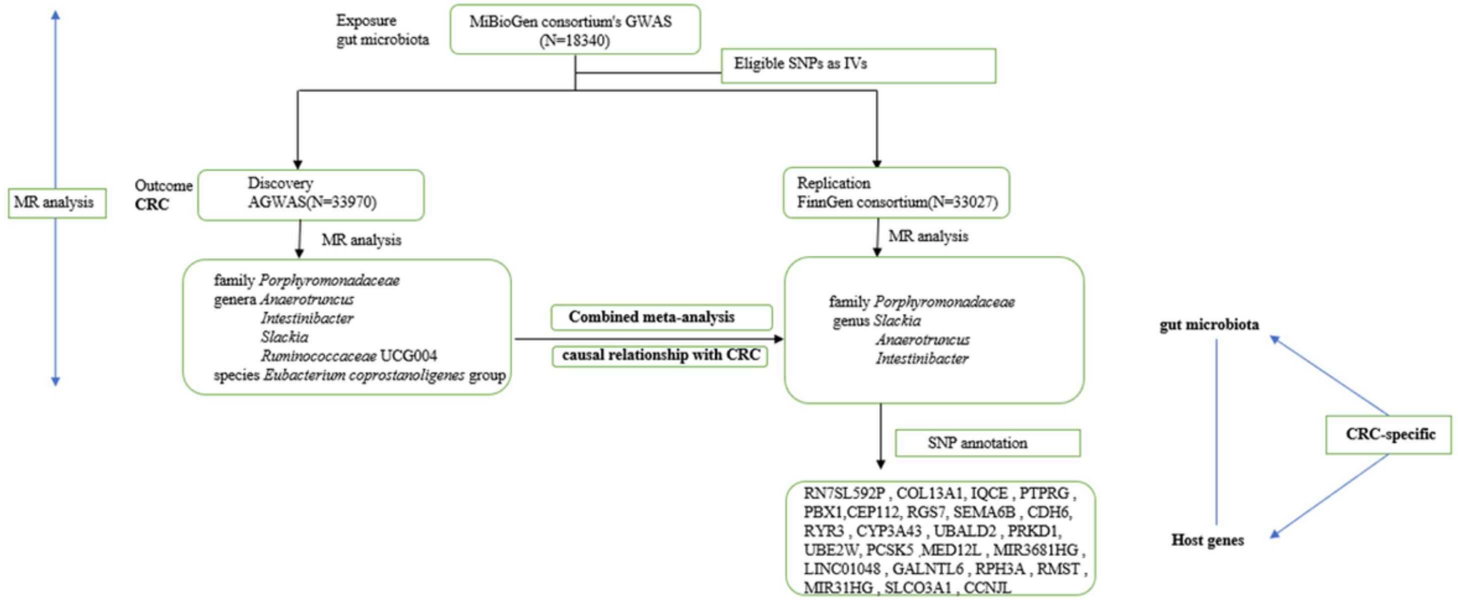

Figure 1

Overview of study design

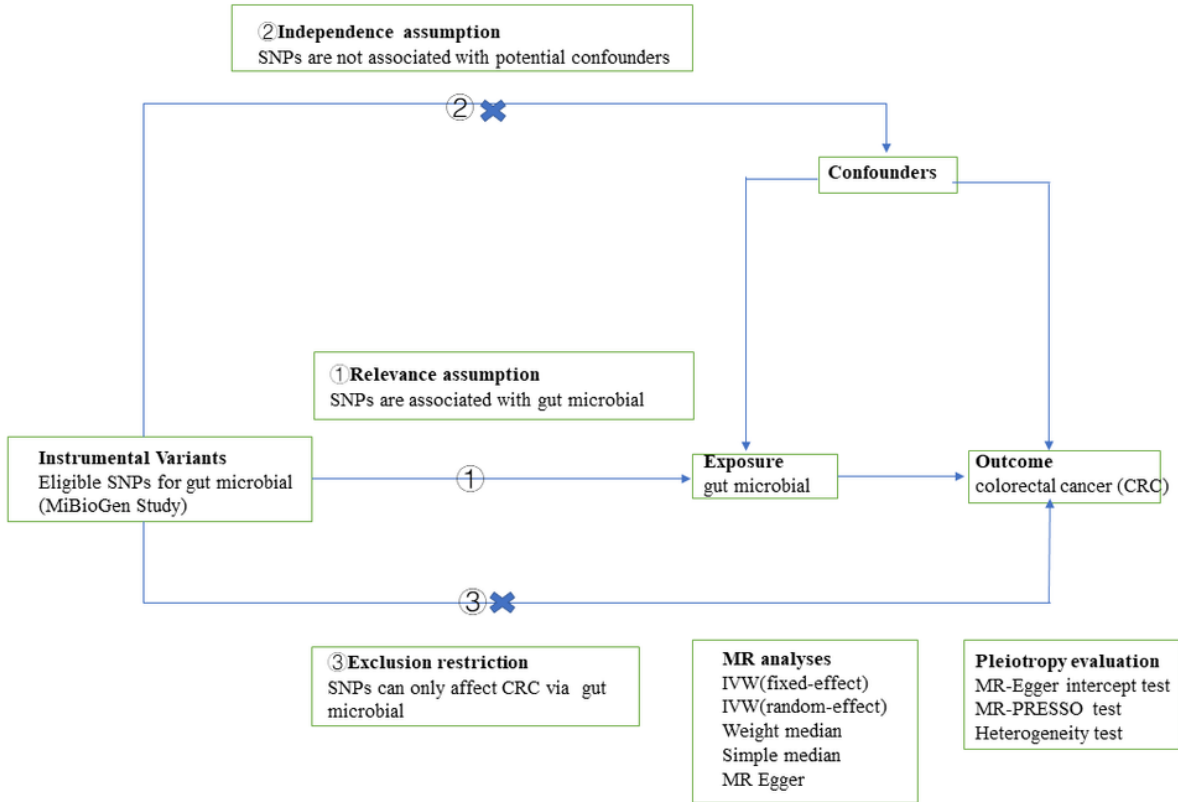

Figure 2

Schematic diagram of the present Mendelian randomization study.

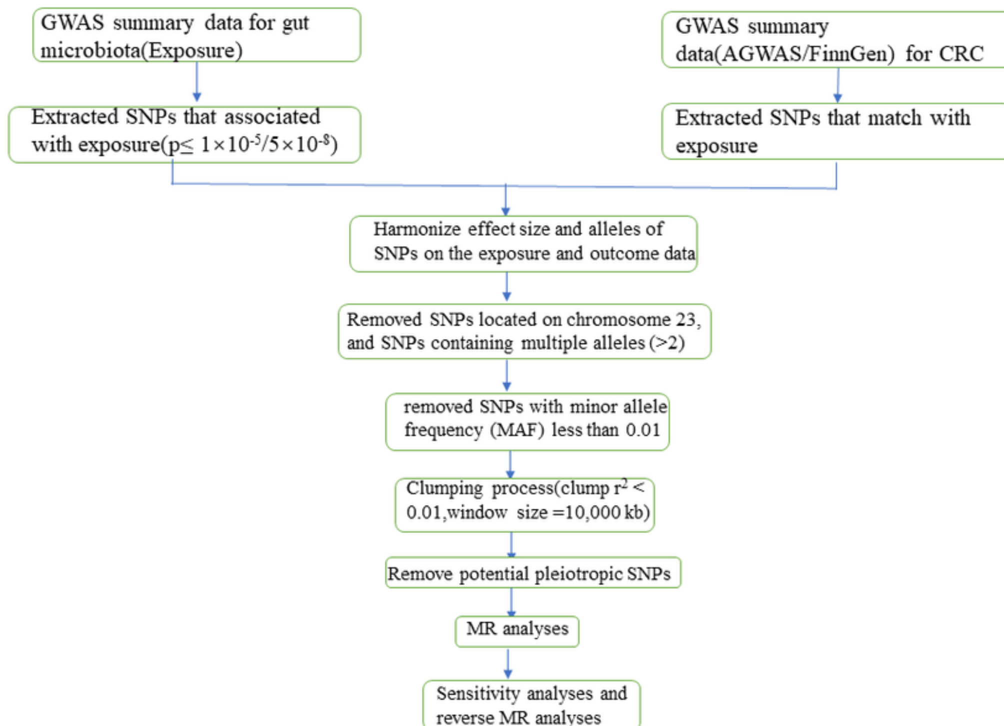

Figure 3

The whole workflow of MR analysis.

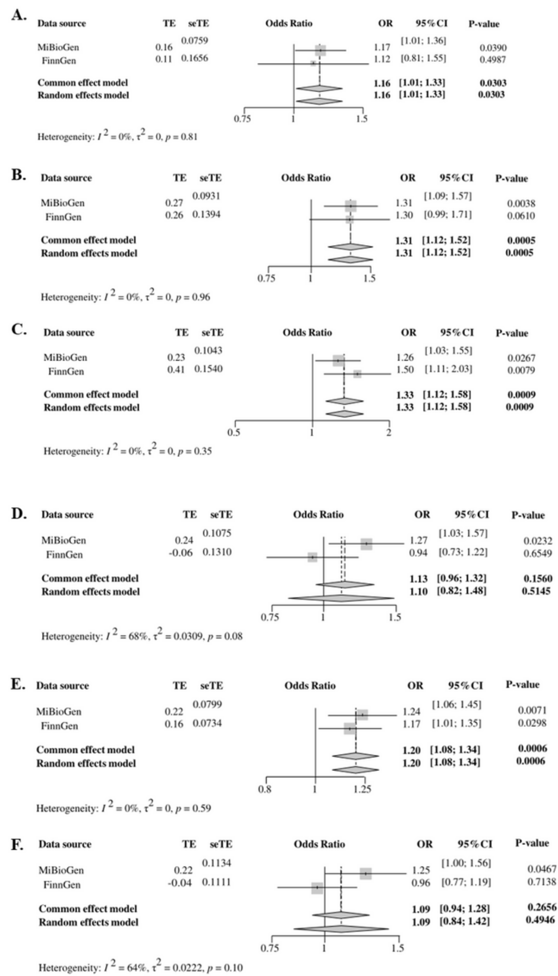

Figure 4

Association of genetically predicted Gut Microbiome with risk of CRC and combined MR estimates from both AGWAS and FinnGen databases by meta-analysis (A) genus Anaerotruncus (B) genus Intestinibacter (C) family Porphyromonadaceae (D) genus RuminococcaceaeUCG004 (E) genus Slackia (F) species Eubacteriumcoprostanoligenesgroup.