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.

Gut microbiome-disease associations revealed by large-scale Mendelian randomization analyses

Xiaohui Che
Chongqing Medical University
Xu Zhang
Chongqing Medical University
Zishun Zhang
Chongqing Medical University
Jianbo Pan
panjianbo@cqmu.edu.cn
Chongqing Medical University

Research Article

Keywords: Gut microbiome, Mendelian randomization, colocalization, mediation analysis, circulating metabolites, cardiovascular disease, phenome-wide association

Posted Date: October 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11314330/v1

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

Additional Declarations: No competing interests reported.

Abstract

Background

The gut microbiome has been associated with diverse human diseases, but systematic genetic evidence for potential causal relevance, disease-system clustering, and circulating metabolic pathways remains limited.

Methods

We integrated four gut microbiome genome-wide association study (GWAS) datasets with 4,934 disease GWAS outcomes. Two-sample Mendelian randomization (MR), using the inverse-variance weighted (IVW) method as the primary analysis, was performed to evaluate potential causal associations between gut microbial traits and diseases. Multiple sensitivity analyses and reverse MR were conducted to assess robustness and potential reverse causality. Hypergeometric tests were used to examine disease-system enrichment of microbial lineages with significant MR associations. For significant MR associations, Bayesian colocalization analysis was performed to assess shared local genetic signals. Circulating metabolite GWAS data were further incorporated into two-step MR mediation analysis to explore candidate metabolic pathways.

Results

We identified 1,943 significant associations between 630 gut microbial traits and 925 disease outcomes across multiple disease systems. Enrichment analysis revealed system-specific clustering patterns, including enrichment of Burkholderiales-related taxa in circulatory diseases, Bacteroidota-related taxa in mental, behavioural, or neurodevelopmental disorders, *Eisenbergiella*-related taxa in musculoskeletal diseases, and *Faecalicatena*-related taxa in endocrine and metabolic diseases. Colocalization analysis identified 58 colocalization signals with supportive evidence of shared local genetic signals, involving 34 microbial traits, 45 disease outcomes, and 33distinct genomic windows. Mediation analysis highlighted triglycerides in intermediate-density lipoprotein (IDL-TG) as a frequent candidate mediator, particularly linking Burkholderiales-related taxa to ischaemic cardiovascular diseases.

Conclusions

This large-scale genetic study provides a systematic map of potential causal associations between the gut microbiome and multi-system diseases. By integrating MR, disease-system enrichment, colocalization, and mediation analyses, our findings suggest that specific gut microbial lineages may contribute to disease susceptibility through shared genetic architecture and circulating metabolic pathways. In particular, the Burkholderiales-IDL-TG-ischaemic cardiovascular disease axis may represent a candidate microbiome-metabolism-disease pathway that warrants further validation in experimental and clinical studies.

Background

The gut microbiome is a highly complex and dynamic internal ecosystem that plays indispensable roles in host nutrient absorption, immune defense, and metabolic homeostasis[1]. With advances in high-throughput sequencing and multi-omics cohorts, increasing evidence indicates that changes in gut microbiome composition and function

are not limited to local gastrointestinal diseases, but are also closely related to the development of cardiovascular diseases, endocrine and metabolic disorders, neuropsychiatric and neurodevelopmental abnormalities, and multiple malignancies[2]. For example, the concepts of the gut-brain axis and gut-liver axis have greatly expanded our understanding of the pathogenesis of complex systemic diseases[3, 4].

Although observational epidemiological studies have documented hundreds to thousands of microbiome-disease associations, these findings face important methodological challenges. First, unmeasured confounders such as diet, antibiotic exposure, and geographic environment may distort true association estimates[5]. Second, in many pathological states, systemic inflammation or physiological changes caused by the disease itself can reshape the gut microbiome, giving rise to reverse causality[6]. In addition, large-scale population-level genetic evidence remains limited for the molecular mechanisms by which specific gut taxa may influence distant organs by releasing or regulating circulating metabolites, such as triglycerides in intermediate-density lipoprotein (IDL-TG)[7–9].

Mendelian randomization (MR) uses genetic variants as instrumental variables based on the random allocation of alleles during meiosis, providing a framework to evaluate potential causal relationships between exposures and outcomes while reducing confounding and reverse causation in conventional observational studies[10–12]. However, MR alone may not distinguish causal associations from false-positive findings driven by linkage disequilibrium (LD)[13]. Therefore, combining MR with high-resolution regional genetic colocalization and incorporating genome-wide association study (GWAS) data for circulating metabolites in two-step mediation analysis have become important strategies for dissecting complex causal pathways[14, 15].

Accordingly, this study integrated data from four independent gut microbiome GWAS resources and 4,934 disease summary datasets covering 23 disease systems to conduct a large-scale bidirectional MR analysis. To ensure robustness, we performed multiple sensitivity analyses and applied the Benjamini–Hochberg procedure for multiple-testing correction to control the false discovery rate (FDR). We further incorporated system-level enrichment analysis to identify disease-system-specific microbial clustering patterns. In addition, we combined colocalization analysis with two-step metabolic mediation models to explore candidate mechanistic pathways through which representative taxa, such as Burkholderiales, may influence cardiometabolic outcomes via circulating lipid-related metabolites. This study aims to provide population-level genetic insights into the etiology of cross-system complex diseases and to identify potential targets for future microbiome-oriented translational research.

Methods

Study design

Figure 1 summarizes the data sources and overall analytical framework of this study. We used two-sample Mendelian randomization (MR) to systematically evaluate potential causal relationships between gut microbial traits and multiple diseases. This approach uses single nucleotide polymorphisms (SNPs) that are strongly associated with the exposure as instrumental variables (IVs) to infer causal effects of the exposure (gut microbial traits) on the outcome (disease). Because genetic variants are randomly allocated at conception, MR can reduce confounding and reverse causation that commonly affect conventional observational studies.

MR relies on three core assumptions for unbiased causal inference: (1) the instrumental variables are strongly associated with the exposure (relevance); (2) the instrumental variables are independent of confounders

(independence); and (3) the instrumental variables influence the outcome only through the exposure (exclusion restriction)[12, 16]. The GWAS data sources for gut microbiota, metabolites, and diseases are summarized in Additional file 1.

Gut microbiome genome-wide association study data

Summary-level gut microbiome GWAS data were obtained from four independent large-scale datasets. First, we used 16S rRNA sequencing data from the MiBioGen consortium, including 18,340 individuals (mainly of European ancestry, N = 13,266)[17]. After removing 15 taxa with unknown taxonomic levels, 196 bacterial taxa were retained, including 9 phyla, 16 classes, 20 orders, 32 families, and 119 genera. Second, we included metagenomic shotgun sequencing data from 7,738 participants in the Dutch Microbiome Project (DMP)[18], covering 207 taxa (including 105 species) and 205 metabolic pathways. Third, we included GWAS data from Rühlemann et al.[19], involving 8,956 German individuals and 430 gut microbial taxa. Finally, we used GWAS data from the Finnish FINRISK 2002 (FR02) cohort, including 5,959 individuals[20], covering 473 taxa resolved to the species level.

Disease genome-wide association study data

GWAS summary statistics for disease outcomes were obtained from the DMRdb database[21]. This database integrates public resources from the MRC IEU OpenGWAS platform[22], the GWAS Catalog[23], and the FinnGen project[24]. The initial dataset contained 6,640 disease-related phenotypes. Because FinnGen R9 and R10 contained overlapping data, duplicate phenotypes were removed, resulting in 4,934 disease phenotypes from 23 disease systems for subsequent causal inference.

Metabolite genome-wide association study data

To explore potential mechanisms by which the gut microbiome may influence disease risk, plasma metabolites were introduced as mediators. The first dataset was from a GWAS study by Chen et al. (2023) based on the Metabolon platform, covering 1,400 plasma metabolites (including 1,091 independent metabolites and 309 metabolite ratios), obtained from the GWAS Catalog (GCST90199621-GCST90201020)[25]. The second dataset focused on high-resolution quantitative lipidomics and used nuclear magnetic resonance (NMR)-based meta-analysis data from Karjalainen et al. (2024), involving 33 cohorts and 136,016 participants. This dataset covered 233 circulating metabolic traits, mainly lipoprotein subclasses and small molecules, and was obtained from the GWAS Catalog (GCST90301941–GCST90302173)[26].

Instrumental variable selection

To improve the validity and accuracy of causal inference between the gut microbiome and diseases, we applied the following quality-control procedures for instrumental variable (IV) selection. First, $P < 1 \times 10^{-5}$ was used as the selection threshold for gut microbiome instruments. Second, LD clumping was performed using PLINK (v1.90b6.21) based on the European-ancestry reference panel from the 1000 Genomes Project, with a 10,000-kb window and an $r^2 < 0.001$ threshold to ensure the relative independence of the instruments. Third, F statistics were calculated to assess instrument strength, and SNPs with $F > 10$ were retained as strong instruments. We then harmonized exposure and outcome GWAS data and removed palindromic SNPs with ambiguous orientation and

SNPs with inconsistent alleles. Finally, to reduce bias due to instruments being directly and strongly associated with outcomes, SNPs reaching $P < 1 \times 10^{-5}$ in the outcome GWAS were removed[27]. The same instrument-selection criteria and quality-control procedures were applied in the reverse MR analyses.

Mendelian randomization analysis

The inverse-variance weighted (IVW) method was used as the primary approach for estimating the causal effects of gut microbial traits on disease outcomes. For exposures with only one valid instrumental variable, the Wald ratio method was used. For each disease outcome, P values from the primary MR analyses were adjusted separately within each gut microbiome GWAS data source using the Benjamini–Hochberg (BH) procedure. Associations with FDR < 0.05 were considered statistically significant.

To assess the robustness of the identified associations, complementary MR methods, including MR-RAPS, MR-Egger, simple mode, and simple median, were applied where appropriate. Cochran's Q test was used to assess heterogeneity among instrumental variables[28], while MR-PRESSO was used to detect potential outlier variants and horizontal pleiotropy[29]. Leave-one-out analysis was performed to evaluate whether individual instrumental variants disproportionately influenced the causal estimates. Funnel plots were additionally examined for representative multi-instrument associations.

Potential reverse causal effects were further assessed using reverse MR analysis, with disease phenotypes treated as exposures and gut microbial traits as outcomes. Instrumental-variable selection and quality-control procedures followed those used in the forward MR analysis. IVW was used as the primary method for exposures with multiple instruments, whereas the Wald ratio was used for single-instrument exposures.

These complementary analyses were used to assess the robustness and consistency of the primary MR findings rather than as criteria for defining statistical significance. All MR analyses and sensitivity analyses were performed in R software (version 4.3.1), mainly using the TwoSampleMR, mr.raps, and MR-PRESSO packages.

Disease-system enrichment analysis

To characterize the distribution of significant MR associations across disease systems, gut microbial traits were first taxonomically harmonized and annotated using the Genome Taxonomy Database (GTDB)[30] and the NCBI Taxonomy database[31]. All disease phenotypes were then classified according to their corresponding disease systems.

Hypergeometric tests were subsequently used to evaluate whether significant gut microbial trait–disease GWAS outcome association pairs involving a given taxon were overrepresented in a particular disease system relative to chance expectation[32]. The probability was calculated as follows:

$$P(X \geq k) = \sum_{i=k}^{min(K,n)} \frac{\binom{K}{i}\binom{N-K}{n-i}}{\binom{N}{n}}$$

Here, *X* denotes the number of significant MR association pairs involving the specified disease system. For each taxon, *N* denotes the total number of possible gut microbial trait–disease GWAS outcome pairs; *K* denotes the number of possible pairs involving disease GWAS outcomes within the specified disease system; *n* denotes the

total number of significant MR association pairs involving the taxon; and *k* denotes the observed number of these significantly associated pairs involving that disease system.

Hypergeometric P values were calculated using the phyper function in the stats package in R[33]. To control the false discovery rate, P values from all enrichment tests were adjusted using the Benjamini–Hochberg procedure separately within each taxonomic rank, and associations with FDR < 0.05 were considered statistically significant.

Colocalization analysis

To further assess whether significant microbiota-disease associations were supported by a shared causal variant, approximate Bayes factor (ABF) colocalization analysis was performed for significant MR associations. Colocalization analysis was implemented using the coloc R package. Summary statistics from microbiome GWAS and disease GWAS were extracted within ± 500 kb of each MR instrumental variant. For reporting purposes, genomic regions were defined according to distinct MR lead SNPs, with each region centered on a lead SNP and extending ± 500 kb.

The posterior probability of hypothesis 4 (PP.H4.abf) was the focus of analysis, where H4 indicates that the microbial trait and disease outcome share the same causal variant within a genomic region. PP.H4.abf > 0.6 was considered supportive evidence of colocalization, whereas PP.H4.abf > 0.8 was considered strong evidence. Associations meeting these criteria were considered to have supportive evidence for a shared causal variant within the local genomic region. All colocalization analyses used default prior parameters from the coloc package.

To visually present colocalization results, the locuszoomr R package[34] was used to plot regions with supportive colocalization evidence.

Mediation analysis

Circulating metabolites were evaluated as potential mediators using a two-step MR framework, and indirect effects were estimated using the product-of-coefficients method. The analytical pathway was defined as gut microbial trait → circulating metabolite → disease outcome.

MR analyses were performed to estimate the effects of gut microbial traits on circulating metabolites, circulating metabolites on disease outcomes, and gut microbial traits on disease outcomes. For exploratory screening, only gut microbiome–metabolite–disease combinations for which all three associations reached nominal statistical significance (P < 0.01) were retained. This threshold was selected to provide a more stringent screening criterion than P < 0.05 while retaining sufficient candidate pathways for exploratory analysis.

The MR effect estimate for the gut microbial trait on the metabolite was denoted as β_a, the estimate for the metabolite on the disease outcome as β_b, and the MR estimate for the total effect of the gut microbial trait on the disease outcome as *βtotal*.

The indirect effect was calculated using the product-of-coefficients method:

$$\beta_{indirect} = \beta_a * \beta_b$$

The standard error of the indirect effect was calculated using the Sobel approximation based on the first-order delta method:

$$SE_{indirect} = \sqrt{\beta_b^2 SE_a^2 + \beta_a^2 SE_b^2}$$

The corresponding *z* statistic was calculated as:

$$z = \frac{\beta_{indirect}}{SE_{indirect}}$$

A two-sided P value for the indirect effect was calculated from the standard normal distribution.

The direct effect was approximated as:

$$\beta_{direct} = \beta_{total} - \beta_{indirect}$$

The mediation proportion was calculated as:

$$Mediation\ proportion = \frac{\beta_{indirect}}{\beta_{total}} \times 100\%$$

Candidate mediation paths were retained when the indirect effect had a two-sided P < 0.05, the indirect and total effects had the same direction, and the mediation proportion was greater than 0% and no greater than 100%. Given the use of nominal significance thresholds for exploratory screening, the identified paths were interpreted as candidate mediation signals rather than definitive evidence of biological mediation.

Results

Bidirectional Mendelian randomization analysis and sensitivity analysis

We performed two-sample MR analyses of four gut microbiome trait datasets and 4,934 disease trait datasets to systematically evaluate potential causal relationships between the gut microbiome and multiple diseases. Using the prespecified primary MR analyses, with IVW applied to exposures with multiple instruments and the Wald ratio applied to single-instrument exposures, we identified 1,943 significant potential causal associations between 630 gut microbial traits and 925 disease GWAS outcomes after Benjamini–Hochberg correction(Additional file 2).

The pie chart shows the proportional distribution of disease outcomes with significant MR associations across disease systems. Each sector represents a disease system, and the sector area reflects the relative proportion of disease outcomes in that system among the 925 disease outcomes with significant MR associations.

Significant MR associations spanned 23 disease systems (Fig. 2), with prominent representation in neoplasms, digestive diseases, endocrine, nutritional, and metabolic diseases, circulatory diseases, and mental, behavioural, or neurodevelopmental disorders. Representative associations in neoplasms, endocrine and metabolic diseases, musculoskeletal system and connective tissue diseases, and circulatory diseases are highlighted below.

Neoplasms

Among neoplastic diseases, 201 significant MR associations involving 111 tumor-related outcomes were identified (Fig. 3A), spanning malignancies of the digestive tract, lung and bronchus, breast, genitourinary system, skin, hematologic system, and central nervous system. Representative associations included an inverse association of

Prevotella-associated ASV (TestASV_19) with esophageal cancer (OR = 0.679, FDR < 0.001) and a positive association of Bifidobacteriaceae- and Bifidobacteriales-related traits with colorectal adenocarcinoma (OR ≈ 1.40, FDR = 0.009).

Endocrine, nutritional, and metabolic diseases

A total of 152 significant MR associations involving 66 endocrine and metabolic disease outcomes were identified (Fig. 3B), including type 2 diabetes, hypercholesterolemia, diabetes mellitus, and thyrotoxicosis. Bacteroidota-, Bacteroidia-, and Bacteroidales-related taxa showed consistent inverse associations with type 2 diabetes (OR ≈ 0.905, FDR = 0.026). In contrast, *Faecalicatena*-related traits were associated with an increased risk of diabetes mellitus (OR = 1.197, FDR = 0.015), whereas *Phascolarctobacterium* was positively associated with thyrotoxicosis (OR = 1.372, FDR = 0.009).

Musculoskeletal system and connective tissue diseases

We identified 162 significant MR associations involving 87 microbial traits and 71 musculoskeletal or connective tissue disease outcomes (Fig. 3C). Among representative findings, *Collinsella aerofaciens* was associated with an increased risk of arthropathies in other diseases classified elsewhere (OR = 2.576, FDR = 0.002), while *Eisenbergiella* was positively associated with hallux rigidus and related disorders (OR = 1.693, FDR = 0.012).

Circulatory system diseases

A total of 171 significant MR associations involving 62 circulatory disease outcomes were identified (Fig. 3D). Burkholderiales-related taxa showed consistent inverse associations with ischaemic heart disease (OR ≈ 0.912, FDR < 0.001). By contrast, Acidaminococcaceae was associated with an increased risk of ischaemic stroke due to large artery atherosclerosis (OR = 1.328, FDR = 0.034), while Betaproteobacteria- and Burkholderiales-related taxa were positively associated with mitral valve disease (OR = 3.016, FDR = 0.009 and OR = 2.861, FDR = 0.019, respectively). These findings highlight disease-specific differences in the direction and magnitude of associations between gut microbial traits and cardiovascular phenotypes.

Representative significant MR associations are shown for (A) neoplasms, (B) endocrine, nutritional, and metabolic diseases, (C) musculoskeletal system and connective tissue diseases, and (D) circulatory system diseases. Points indicate odds ratios (ORs), and error bars indicate 95% confidence intervals. OR > 1 indicates an association with increased disease risk, whereas OR < 1 indicates an association with reduced disease risk.

Sensitivity analysis

To assess the robustness of the main findings, we applied several additional MR methods, including MR-RAPS, MR-Egger, simple median, and the simple mode methods, together with sensitivity analyses comprising single-SNP analysis, leave-one-out analysis, and funnel plot assessment. As summarized in Fig. 4, the selected significant MR associations generally showed consistent effect directions across complementary MR methods. Detailed sensitivity plots for representative associations are provided in Additional file 6. These analyses suggested that the main findings were not strongly driven by method selection or by a single instrumental variant. Among associations for which the MR-Egger intercept test was applicable, no evidence of directional horizontal pleiotropy was observed.

The outer ring of the circular heatmap represents disease system categories, and the inner rings correspond to different MR methods, including IVW, MR-RAPS, Wald ratio, simple mode, and simple median. Colors indicate

causal effect estimates (Beta/LogOR), with orange representing increased disease risk (positive causal effect) and purple representing decreased risk (protective effect).

To further assess potential reverse causality, we used 4,934 disease phenotypes as exposures and gut microbial traits as outcomes in reverse MR analyses. No significant evidence supporting reverse causal effects of diseases on gut microbial traits was identified.

Disease-system enrichment analysis

To explore whether significant gut microbial trait–disease GWAS outcome association pairs were overrepresented within particular disease systems, we performed hypergeometric enrichment analyses across different taxonomic ranks. Figure 5 displays enrichment signals that remained significant after multiple-testing correction (FDR < 0.05) and involved at least three distinct disease phenotypes (Additional file 3).

Taxa belonging to Bacteroidota were significantly enriched for associations with mental, behavioural, or neurodevelopmental disorders, indicating an overrepresentation of significant MR associations involving these outcomes. Burkholderiales-related taxa were enriched for diseases of the circulatory system. The corresponding significant MR associations included cardiovascular outcomes such as coronary atherosclerosis and broadly defined ischaemic heart disease.

Actinomycetota- and *Eisenbergiella*-related taxa, together with Bacillota, were enriched for diseases of the musculoskeletal system or connective tissue, whereas *Faecalicatena* was enriched for endocrine, nutritional, or metabolic diseases. Together with the MR associations involving metabolic outcomes such as T2D, these findings indicate that significant gut microbial trait–disease associations were unevenly distributed across disease systems and identify microbial lineages that warrant further biological investigation.

The bubble plot shows the enrichment of significant MR association pairs involving microbial taxa across disease systems. The x-axis represents disease systems, and the y-axis represents microbial taxa. Bubble color represents enrichment significance (− log10[FDR]), and bubble size represents fold enrichment. Only enrichment results with FDR < 0.05 that were supported by at least three distinct disease phenotypes are shown.

Colocalization analysis

To further validate MR results and reduce potential confounding by linkage disequilibrium (LD) from nearby independent variants, we performed Bayesian colocalization analyses for all significant gut microbiota-disease MR association pairs. Using posterior probability of colocalization (PP.H4.abf) > 0.6 as the threshold for supportive colocalization evidence, we obtained 58 colocalization signals involving 34 microbial traits, 45 disease outcomes, and 33 distinct genomic windows centered on distinct lead SNPs

(Additional file 4).

Overall, these colocalization loci were distributed across multiple systems, including cardiovascular, neoplastic, musculoskeletal, digestive, and endocrine/metabolic systems. This indicates that some MR associations between the gut microbiome and multi-system diseases may not be purely statistical signals, but may reflect shared local genetic architecture or effects of the same or highly correlated genetic variants.

Among neoplasm-related phenotypes, *Bifidobacterium*-related taxonomic units showed consistent colocalization evidence in the chr2:136.12–137.12 Mb region. Bifidobacteriaceae and Bifidobacteriales were colocalized with

both benign neoplasm of the transverse colon (PP.H4.abf = 0.902 for both) and colorectal adenocarcinoma (PP.H4.abf ≈ 0.951 for both). These findings suggest that the MR associations between *Bifidobacterium*-related taxonomic levels and colorectal tumor-related phenotypes may be supported by shared local genetic signals.

In musculoskeletal and connective tissue disorders, Bifidobacterium breve showed strong colocalization evidence with other joint disorders (PP.H4.abf = 0.864), internal derangement of knee (PP.H4.abf = 0.879), and polyarthrosis (PP.H4.abf = 0.950). All three signals were located within the chr2:135.61–136.61 Mb region. Eisenbergiella sp900066775 also showed strong colocalization evidence with gonarthrosis (PP.H4.abf = 0.840).

In circulatory diseases, the lactose and galactose degradation I pathway showed the strongest colocalization signal with atrial fibrillation and flutter at chr9:135.65–136.65 Mb (PP.H4.abf = 0.978). Burkholderiales-related taxonomic units also showed consistent colocalization with ischaemic heart disease, wide definition, at chr11:1–714,169 bp, with rs1533824 as the representative lead SNP and PP.H4.abf values ranging from 0.752 to 0.756. In addition, Acidaminococcaceae was colocalized with large artery atherosclerotic ischaemic stroke (PP.H4.abf = 0.744).

For endocrine and metabolic outcomes, *Bacteroides nordii* showed supportive colocalization with thyrotoxicosis with toxic multinodular goitre (PP.H4.abf = 0.815), while *Dorea formicigenerans* was colocalized with type 2 diabetes (PP.H4.abf = 0.688). Among neuropsychiatric phenotypes, Parabacteroides showed strong colocalization evidence with childhood-onset emotional disorders and disorders of social functioning (PP.H4.abf = 0.889).

Mediation analysis

Mediation MR analysis identified 1,652 candidate mediation paths meeting the predefined exploratory criteria involving 465 gut microbial traits, 533 metabolites, and 596 disease outcomes. To characterize the overall distribution of these mediation paths, we selected the top 50 circulating metabolites that mediated the largest number of disease outcomes and constructed a metabolite–disease matrix plot (Fig. 6A). The complete mediation results are provided in Additional file 5. The results showed that candidate mediation paths were not uniformly distributed across all metabolites and disease outcomes, but instead were concentrated in specific circulating metabolites and disease systems. Lipid-related metabolites repeatedly appeared across multiple disease outcomes, including cholesterol, cholesteryl esters, triglycerides, phospholipids, apolipoproteins, and lipoprotein particle-related traits. The left bar plot further showed that several lipid metabolites mediated a relatively large number of disease outcomes, whereas the top bar plot indicated that some disease outcomes involved multiple mediating metabolites.

Triglycerides in intermediate-density lipoprotein (IDL-TG) appeared in 29 nominally significant and directionally consistent candidate mediation paths, including 24 paths involving three Burkholderiales-related taxonomic traits and eight disease outcomes. These IDL-TG-related candidate paths are illustrated in the Sankey plot (Fig. 6B).

IDL-TG-related candidate mediation paths predominantly linked Burkholderiales-related taxa to cardiometabolic outcomes, with directionally consistent indirect effects across family-, genus-, and species-level traits. The corresponding disease outcomes included coronary atherosclerosis, angina pectoris, strictly defined myocardial infarction, ischaemic heart disease (wide definition), and coronary revascularization. These paths had negative indirect effects, with mediation proportions ranging from 2.04% to 4.78%, suggesting a potential mediating role for IDL-TG in the associations between Burkholderiales-related taxa and ischaemic cardiovascular diseases.

Additional candidate mediation paths involving IDL-TG linked Ruminococcus bromii to familial combined hyperlipidemia and myocardial infarction, and Bifidobacterium to familial hypercholesterolemia-related outcomes. Notably, the mediation proportion of IDL-TG in the effect of *Ruminococcus bromii* on familial combined hyperlipidemia was relatively high (approximately 11.98%). Overall, the mediation analysis highlighted IDL-TG as a recurrent candidate mediator linking selected gut microbial traits to cardiometabolic outcomes.

Discussion

In this study, we integrated four independent gut microbiome GWAS datasets and 4,934 disease-related phenotypes to systematically investigate the potential causal relationships between gut microbiota and human diseases. Our MR analyses identified widespread associations across multiple disease systems, with distinct clustering patterns in circulatory, mental, behavioural or neurodevelopmental, endocrine and metabolic, and musculoskeletal diseases[35–37]. We further used colocalization and two-step MR mediation analyses to investigate the genetic and metabolic basis of selected associations. These analyses provided additional evidence for shared local genetic signals and suggested several candidate circulating metabolic pathways linking gut microbial traits to disease risk.

Previous studies have shown that the gut microbiome can influence extraintestinal organs through microbially derived metabolites, modulation of gut barrier integrity, and host immune and inflammatory responses[7]. Our findings further suggest that these effects may vary substantially across microbial lineages and disease contexts. For example, Bacteroidota-related taxa were inversely associated with type 2 diabetes, consistent with previous evidence linking Bacteroidota-related taxa to glucose and lipid metabolism[38], whereas related microbial traits showed heterogeneous associations across musculoskeletal phenotypes. Such differences may reflect taxonomic resolution, species- or strain-specific functions, host disease context, and distinct microbial functional pathways.

At the endocrine and metabolic level, animal studies have shown that *Bacteroides uniformis* can improve glucose and lipid metabolism through modulation of bile acid metabolism and the gut–liver axis[39]. B. uniformis CECT 7771 has also been reported to ameliorate high-fat diet-induced obesity, insulin resistance, and dyslipidaemia[40, 41], while *Bacteroides acidifaciens* may ameliorate obesity and improve insulin sensitivity in mice through GLP-1-related metabolic regulation[42, 43]. Although these experimental observations involve specific species rather than the exact taxonomic levels identified in our MR analyses, they provide biological context for the inverse associations between Bacteroidota-related taxa and type 2 diabetes observed in this study. The enrichment of *Faecalicatena*-related signals in endocrine and metabolic phenotypes further suggests that multiple microbial lineages may contribute to metabolic disease susceptibility.

In musculoskeletal and connective tissue diseases, several microbial traits, including *Collinsella aerofaciens*, *Eisenbergiella*, and *Bifidobacterium* breve, were implicated across MR, enrichment, or colocalization analyses. These findings are consistent with the proposed gut–joint axis, through which intestinal barrier disruption, microbial antigen translocation, molecular mimicry, and Th17/Treg imbalance may contribute to joint inflammation and connective tissue disorders[44]. Notably, C. aerofaciens has been experimentally shown to impair intestinal epithelial tight-junction integrity, increase intestinal permeability, and promote IL-17-related inflammatory responses[45], providing a plausible biological context for its association with arthropathy-related phenotypes. B. breve has also been shown to modulate dendritic-cell and T-cell responses through TLR2/MyD88-dependent pathways and extracellular polysaccharides[46, 47]. These immunomodulatory properties may partly explain the genetic associations observed here, although their relevance to specific musculoskeletal phenotypes remains to be experimentally established.

For neoplastic diseases, *Bifidobacterium*-related taxonomic levels showed both MR associations and supportive colocalization evidence with colorectal tumour-related phenotypes. Experimental studies suggest that selected *Bifidobacterium* strains may influence colorectal tumorigenesis through fatty-acid metabolism, microbially derived indole metabolites, macrophage differentiation, and antitumour immune responses[48–51]. However, genetically predicted *Bifidobacterium*-related traits in our study were not uniformly protective and, in some analyses, were associated with increased colorectal tumour risk. These apparently divergent findings suggest that the effects of *Bifidobacterium* should not be interpreted uniformly across taxonomic levels and disease settings, but may instead depend on strain-specific function, disease stage, tumour microenvironment, and host genetic background.

The strongest convergence across analytical layers was observed for circulatory diseases. Burkholderiales-related taxa were associated with a lower risk of coronary atherosclerosis and ischaemic heart disease in MR analyses, showed system-specific enrichment in circulatory diseases, and shared local genetic signals with ischaemic heart disease in the chr11 region. Mediation analysis further identified triglycerides in intermediate-density lipoprotein (IDL-TG) as a candidate intermediate linking Burkholderiales-related taxa to ischaemic cardiovascular outcomes. Previous studies have implicated bile acid metabolism, short-chain fatty acids, trimethylamine N-oxide, lipopolysaccharide-mediated low-grade inflammation, and cholesterol metabolism in microbiome–cardiovascular interactions[52, 53], whereas triglyceride-rich remnant lipoproteins and remnant cholesterol are closely associated with coronary heart disease risk[54, 55]. Although direct experimental evidence linking Burkholderiales to IDL-TG regulation is currently lacking, these findings collectively support a candidate Burkholderiales–IDL-TG–ischaemic cardiovascular disease axis that warrants further mechanistic investigation.

Several limitations should be acknowledged. First, residual pleiotropy and violations of instrumental-variable assumptions cannot be completely excluded despite multiple sensitivity analyses. Second, differences across microbiome GWAS datasets in population characteristics, taxonomic resolution, and profiling methods may introduce heterogeneity and limit generalizability. Some mediation paths involved nested taxonomic traits from the same microbial lineage and should therefore not be interpreted as independent biological pathways. Third, colocalization and mediation analyses provide complementary evidence but do not establish biological mechanisms, and the mediation analysis was exploratory. Further validation in diverse populations and experimental studies is therefore warranted.

Overall, our findings demonstrate that gut microbiome–disease associations are distributed non-randomly across disease systems and can be further prioritized by integrating genetic colocalization and circulating metabolic information. The convergence observed for Burkholderiales-related taxa and ischaemic cardiovascular disease provides a particularly informative example of how multi-layer genetic evidence can be used to nominate candidate microbiome–metabolism–disease pathways for future investigation.

Conclusions

Based on large-scale GWAS summary data, this study systematically mapped potential causal associations between gut microbial traits and diseases across multiple organ systems. Significant associations showed non-random disease-system-specific clustering, indicating that different microbial lineages may be preferentially associated with distinct groups of human diseases.

Integration of MR, disease-system enrichment, colocalization, and metabolic mediation analyses further prioritized selected associations with convergent evidence. In particular, Burkholderiales-related taxa showed consistent links with ischaemic cardiovascular phenotypes, with additional evidence implicating shared local genetic signals and

IDL-TG as a candidate metabolic intermediate. These findings provide a framework for prioritizing microbiome–disease associations for future mechanistic and experimental validation.

Abbreviations

ABF
approximate Bayes factor
BH
Benjamini-Hochberg
DMP
Dutch Microbiome Project
FDR
false discovery rate
GWAS
genome-wide association study
IDL-TG
triglycerides in intermediate-density lipoprotein
IV
instrumental variable
IVW
inverse variance weighted
LD
linkage disequilibrium
MR
Mendelian randomization
MR-PRESSO
Mendelian Randomization Pleiotropy RESidual Sum and Outlier
OR
odds ratio
PP.H4
posterior probability of hypothesis 4
SNP
single nucleotide polymorphism
T2D
type 2 diabetes.

Declarations

Ethics approval and consent to participate

This study was based entirely on publicly available summary-level GWAS datasets and did not involve the collection of new human samples or the use of individual-level identifiable data. Ethical approval and informed consent had been obtained in the original studies. Therefore, additional ethics approval and consent to participate were not required for this secondary analysis.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China (32470699).

Author Contribution

XC performed data collection, data processing, statistical analyses, visualization, and drafted the manuscript. XZ and ZZ contributed to interpretation of the results and revision of the manuscript. JP supervised the project, provided methodological and academic guidance, and critically reviewed and revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The computing work in this paper was partly supported by Supercomputing Center of Chongqing Medical University.

Availability of data and materials

The datasets analysed during the current study are available from the original public GWAS resources described in the Methods section and summarized in Additional file 1. Processed summary results are included within this article and its additional files.

References

1. Fan Y, Pedersen O. Gut microbiota in human metabolic health and disease. Nat Rev Microbiol. 2021;19(1):55–71.
2. Witkowski M, Weeks TL, Hazen SL. Gut Microbiota and Cardiovascular Disease. Circ Res. 2020;127(4):553–70.
3. Goralczyk-Binkowska A, Szmajda-Krygier D, Kozlowska E. The Microbiota-Gut-Brain Axis in Psychiatric Disorders. Int J Mol Sci. 2022;23(19):11245.
4. Pabst O, Hornef MW, Schaap FG, Cerovic V, Clavel T, Bruns T. Gut-liver axis: barriers and functional circuits. Nat Rev Gastroenterol Hepatol. 2023;20(7):447–61.
5. Vujkovic-Cvijin I, Sklar J, Jiang L, Natarajan L, Knight R, Belkaid Y. Host variables confound gut microbiota studies of human disease. Nature. 2020;587(7834):448–54.

6. Walter J, Armet AM, Finlay BB, Shanahan F. Establishing or Exaggerating Causality for the Gut Microbiome: Lessons from Human Microbiota-Associated Rodents. Cell. 2020;180(2):221–32.
7. de Vos WM, Tilg H, Van Hul M, Cani PD. Gut microbiome and health: mechanistic insights. Gut. 2022;71(5):1020–32.
8. Schroeder BO, Backhed F. Signals from the gut microbiota to distant organs in physiology and disease. Nat Med. 2016;22(10):1079–89.
9. Liu J, Tan Y, Cheng H, Zhang D, Feng W, Peng C. Functions of Gut Microbiota Metabolites, Current Status and Future Perspectives. Aging Dis. 2022;13(4):1106–26.
10. Sanna S, van Zuydam NR, Mahajan A, Kurilshikov A, Vich Vila A, Vosa U, et al. Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. Nat Genet. 2019;51(4):600–5.
11. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet. 2014;23(R1):R89–98.
12. Sanderson E, Glymour MM, Holmes MV, Kang H, Morrison J, Munafo MR, et al. Mendelian randomization. Nat Rev Methods Primers. 2022;2:6.
13. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 2014;10(5):e1004383.
14. Zuber V, Grinberg NF, Gill D, Manipur I, Slob EAW, Patel A, et al. Combining evidence from Mendelian randomization and colocalization: Review and comparison of approaches. Am J Hum Genet. 2022;109(5):767–82.
15. 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.
16. Skrivankova VW, Richmond RC, Woolf BAR, Davies NM, Swanson SA, Vanderweele TJ, et al. Strengthening the reporting of observational studies in epidemiology using mendelian randomisation (STROBE-MR): explanation and elaboration. BMJ. 2021;375:n2233.
17. 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.
18. Lopera-Maya EA, Kurilshikov A, van der Graaf A, Hu S, Andreu-Sanchez S, Chen L, et al. Effect of host genetics on the gut microbiome in 7,738 participants of the Dutch Microbiome Project. Nat Genet. 2022;54(2):143–51.
19. Ruhlemann MC, Hermes BM, Bang C, Doms S, Moitinho-Silva L, Thingholm LB, et al. Genome-wide association study in 8,956 German individuals identifies influence of ABO histo-blood groups on gut microbiome. Nat Genet. 2021;53(2):147–55.
20. Qin Y, Havulinna AS, Liu Y, Jousilahti P, Ritchie SC, Tokolyi A, et al. Combined effects of host genetics and diet on human gut microbiota and incident disease in a single population cohort. Nat Genet. 2022;54(2):134–42.
21. Zheng X, Tian Z, Che X, Zhang X, Xiang Y, Ge Z, et al. DMRdb: a disease-centric Mendelian randomization database for systematically assessing causal relationships of diseases with genes, proteins, CpG sites, metabolites and other diseases. Nucleic Acids Res. 2025;53(D1):D1363–71.
22. Elsworth B, Lyon M, Alexander T, Liu Y, Matthews P, Hallett J et al. The MRC IEU OpenGWAS data infrastructure. bioRxiv. 2020.

23. Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. Nucleic Acids Res. 2023;51(D1):D977–85.
24. Kurki MI, Karjalainen J, Palta P, Sipila TP, Kristiansson K, Donner KM, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. Nature. 2023;613(7944):508–18.
25. Chen Y, Lu T, Pettersson-Kymmer U, Stewart ID, Butler-Laporte G, Nakanishi T, et al. Genomic atlas of the plasma metabolome prioritizes metabolites implicated in human diseases. Nat Genet. 2023;55(1):44–53.
26. Karjalainen MK, Karthikeyan S, Oliver-Williams C, Sliz E, Allara E, Fung WT, et al. Genome-wide characterization of circulating metabolic biomarkers. Nature. 2024;628(8006):130–8.
27. Chen S, Zhang Z, Liu S, Chen T, Lu Z, Zhao W, et al. Consistent signatures in the human gut microbiome of longevous populations. Gut Microbes. 2024;16(1):2393756.
28. 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.
29. Verbanck M, Chen C, 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.
30. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P, Hugenholtz P. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Res. 2022;50(D1):D785–94.
31. Schoch CL, Ciufo S, Domrachev M, Hotton CL, Kannan S, Khovanskaya R et al. NCBI Taxonomy: a comprehensive update on curation, resources and tools. Database (Oxford). 2020;2020.
32. Rivals I, Personnaz L, Taing L, Potier M. Enrichment or depletion of a GO category within a class of genes: which test? Bioinformatics. 2007;23(4):401–7.
33. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2023.
34. Lewis MJ, Wang S. locuszoomr: an R package for visualising publication-ready regional gene locus plots. Bioinf Adv. 2025;5(1):vbaf006.
35. Rooks MG, Garrett WS. Gut microbiota, metabolites and host immunity. Nat Rev Immunol. 2016;16(6):341–52.
36. Hou K, Wu Z, Chen X, Wang J, Zhang D, Xiao C, et al. Microbiota in health and diseases. Signal Transduct Target Ther. 2022;7(1):135.
37. Metwaly A, Kriaa A, Hassani Z, Carraturo F, Druart C, Arnauts K, et al. A Consensus Statement on establishing causality, therapeutic applications and the use of preclinical models in microbiome research. Nat Rev Gastroenterol Hepatol. 2025;22(5):343–56.
38. Fu L, Baranova A, Cao H, Zhang F. Gut microbiome links obesity to type 2 diabetes: insights from Mendelian randomization. BMC Microbiol. 2025;25(1):253.
39. Zhu X, Zhao C, Meng X, Yu X, Ma L, Chen T, et al. Bacteroides uniformis Ameliorates Carbohydrate and Lipid Metabolism Disorders in Diabetic Mice by Regulating Bile Acid Metabolism via the Gut-Liver Axis. Pharmaceuticals (Basel). 2024;17(8):1015.
40. Romani-Perez M, Lopez-Almela I, Bullich-Vilarrubias C, Evtoski Z, Benitez-Paez A, Sanz Y. Bacteroides uniformis CECT 7771 requires adaptive immunity to improve glucose tolerance but not to prevent body weight gain in diet-induced obese mice. Microbiome. 2024;12(1):103.

41. Fabersani E, Portune K, Campillo I, Lopez-Almela I, la Paz SM, Romani-Perez M, et al. Bacteroides uniformis CECT 7771 alleviates inflammation within the gut-adipose tissue axis involving TLR5 signaling in obese mice. Sci Rep. 2021;11(1):11788.
42. Yang J, Lee Y, Kim Y, Lee S, Ryu S, Fukuda S, et al. Gut commensal Bacteroides acidifaciens prevents obesity and improves insulin sensitivity in mice. Mucosal Immunol. 2017;10(1):104–16.
43. Wang Q, Lin H, Shen C, Zhang M, Wang X, Yuan M, et al. Gut microbiota regulates postprandial GLP-1 response via ileal bile acid-TGR5 signaling. Gut Microbes. 2023;15(2):2274124.
44. Zaiss MM, Joyce Wu H, Mauro D, Schett G, Ciccia F. The gut-joint axis in rheumatoid arthritis. Nat Rev Rheumatol. 2021;17(4):224–37.
45. Chen J, Wright K, Davis JM, Jeraldo P, Marietta EV, Murray J, et al. An expansion of rare lineage intestinal microbes characterizes rheumatoid arthritis. Genome Med. 2016;8(1):43.
46. Jeon SG, Kayama H, Ueda Y, Takahashi T, Asahara T, Tsuji H, et al. Probiotic Bifidobacterium breve induces IL-10-producing Tr1 cells in the colon. PLoS Pathog. 2012;8(5):e1002714.
47. Hickey A, Stamou P, Udayan S, Ramon-Vazquez A, Esteban-Torres M, Bottacini F, et al. Bifidobacterium breve Exopolysaccharide Blocks Dendritic Cell Maturation and Activation of CD4(+) T Cells. Front Microbiol. 2021;12:653587.
48. Chen Y, Fang H, Chen H, Liu X, Zhao J, Stanton C, et al. Bifidobacterium inhibits the progression of colorectal tumorigenesis in mice through fatty acid isomerization and gut microbiota modulation. Gut Microbes. 2025;17(1):2464945.
49. Li Y, Li Q, Yuan R, Wang Y, Guo C, Wang L. Bifidobacterium breve-derived indole-3-lactic acid ameliorates colitis-associated tumorigenesis by directing the differentiation of immature colonic macrophages. Theranostics. 2024;14(7):2719–35.
50. 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.
51. Sun S, Luo L, Liang W, Yin Q, Guo J, Rush AM, et al. Bifidobacterium alters the gut microbiota and modulates the functional metabolism of T regulatory cells in the context of immune checkpoint blockade. Proc Natl Acad Sci U S A. 2020;117(44):27509–15.
52. Chakaroun RM, Olsson LM, Backhed F. The potential of tailoring the gut microbiome to prevent and treat cardiometabolic disease. Nat Rev Cardiol. 2023;20(4):217–35.
53. Li C, Strazar M, Mohamed AMT, Pacheco JA, Walker RL, Lebar T, et al. Gut microbiome and metabolome profiling in Framingham heart study reveals cholesterol-metabolizing bacteria. Cell. 2024;187(8):1834–e185219.
54. Bjornson E, Adiels M, Taskinen M, Burgess S, Rawshani A, Boren J, et al. Triglyceride-rich lipoprotein remnants, low-density lipoproteins, and risk of coronary heart disease: a UK Biobank study. Eur Heart J. 2023;44(39):4186–95.
55. Navarese EP, Vine D, Proctor S, Grzelakowska K, Berti S, Kubica J, et al. Independent Causal Effect of Remnant Cholesterol on Atherosclerotic Cardiovascular Outcomes: A Mendelian Randomization Study. Arterioscler Thromb Vasc Biol. 2023;43(9):e373–80.

Figures

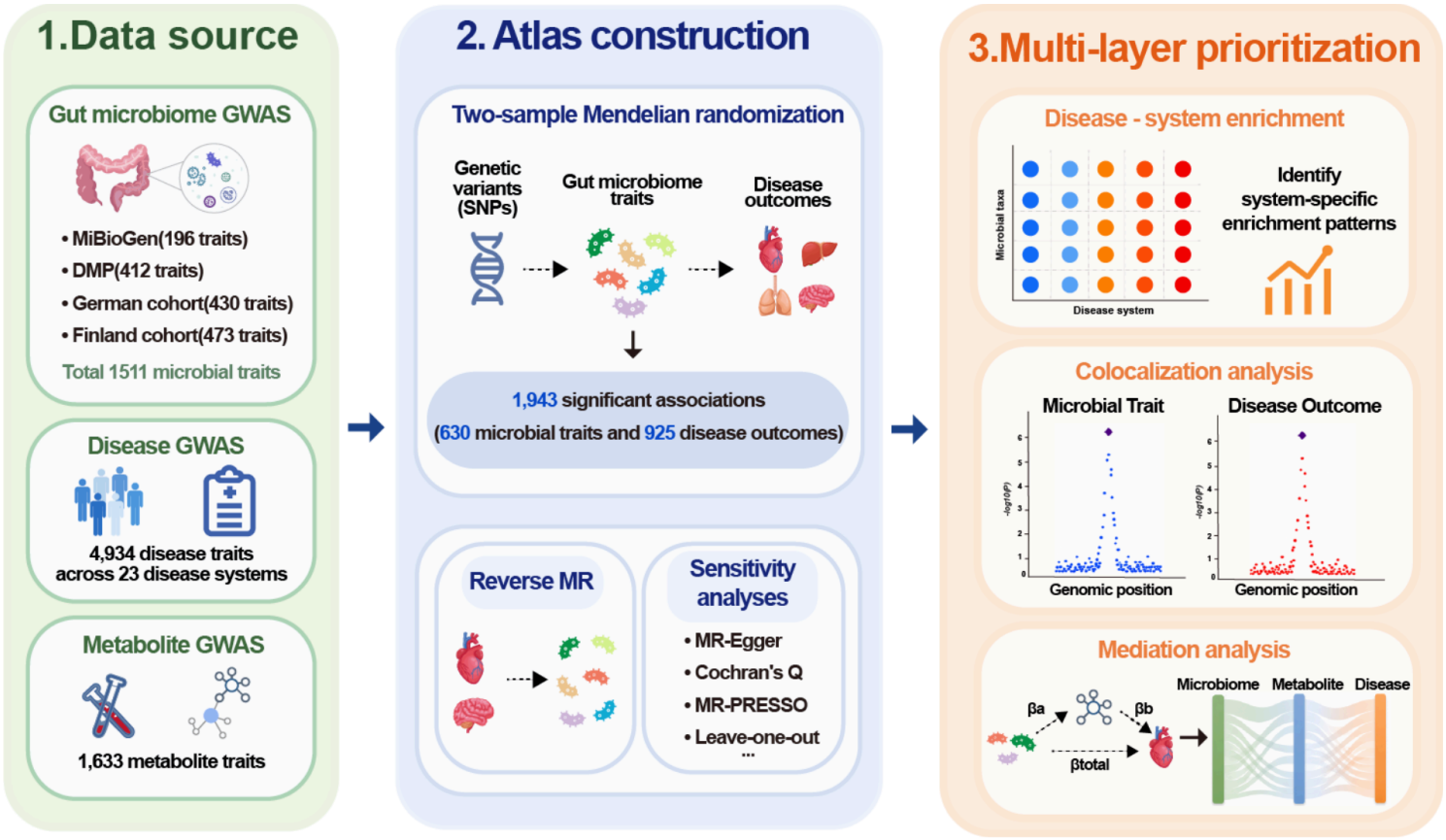


Figure 1

Study design workflow.

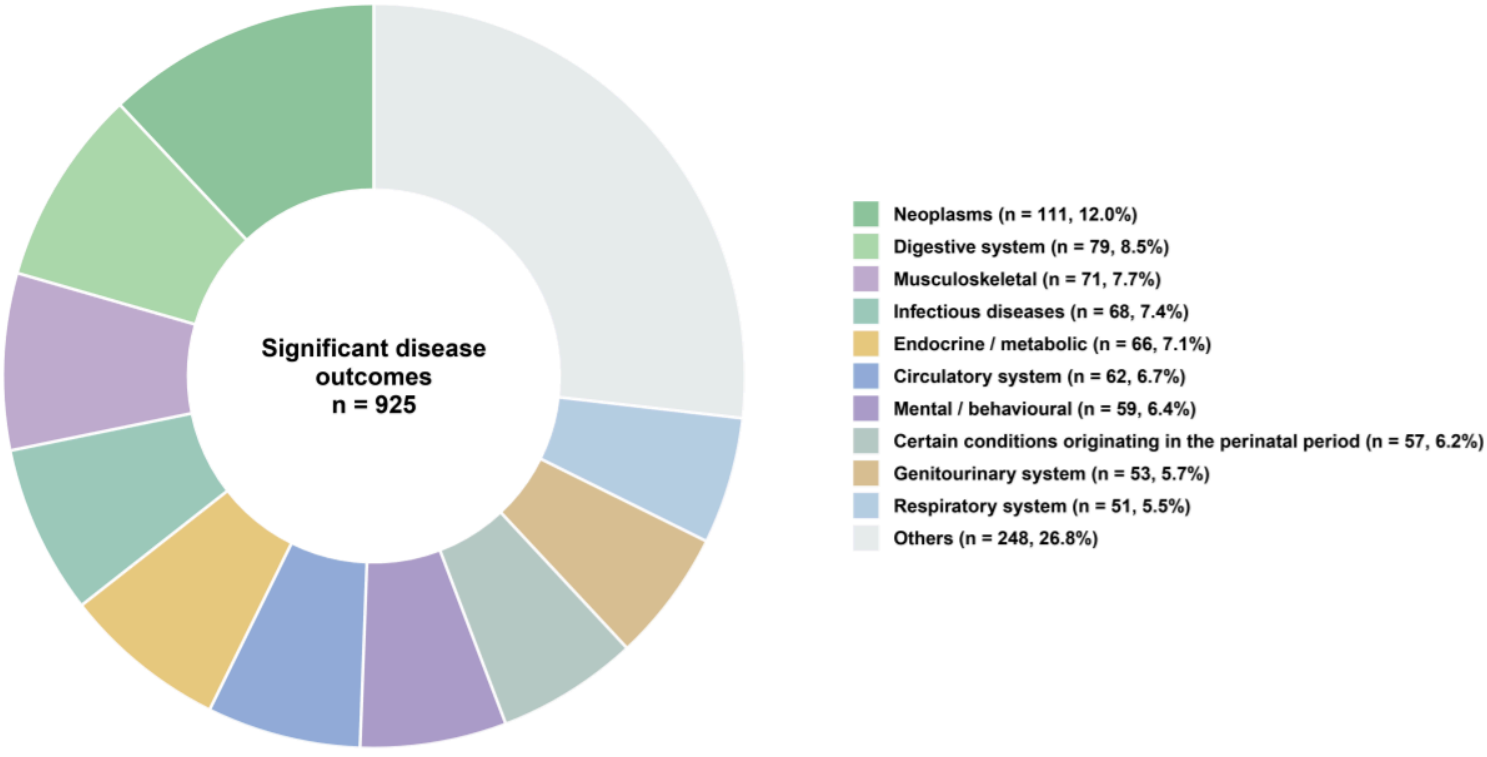


Figure 2

Distribution of disease outcomes with significant MR associations across disease systems.

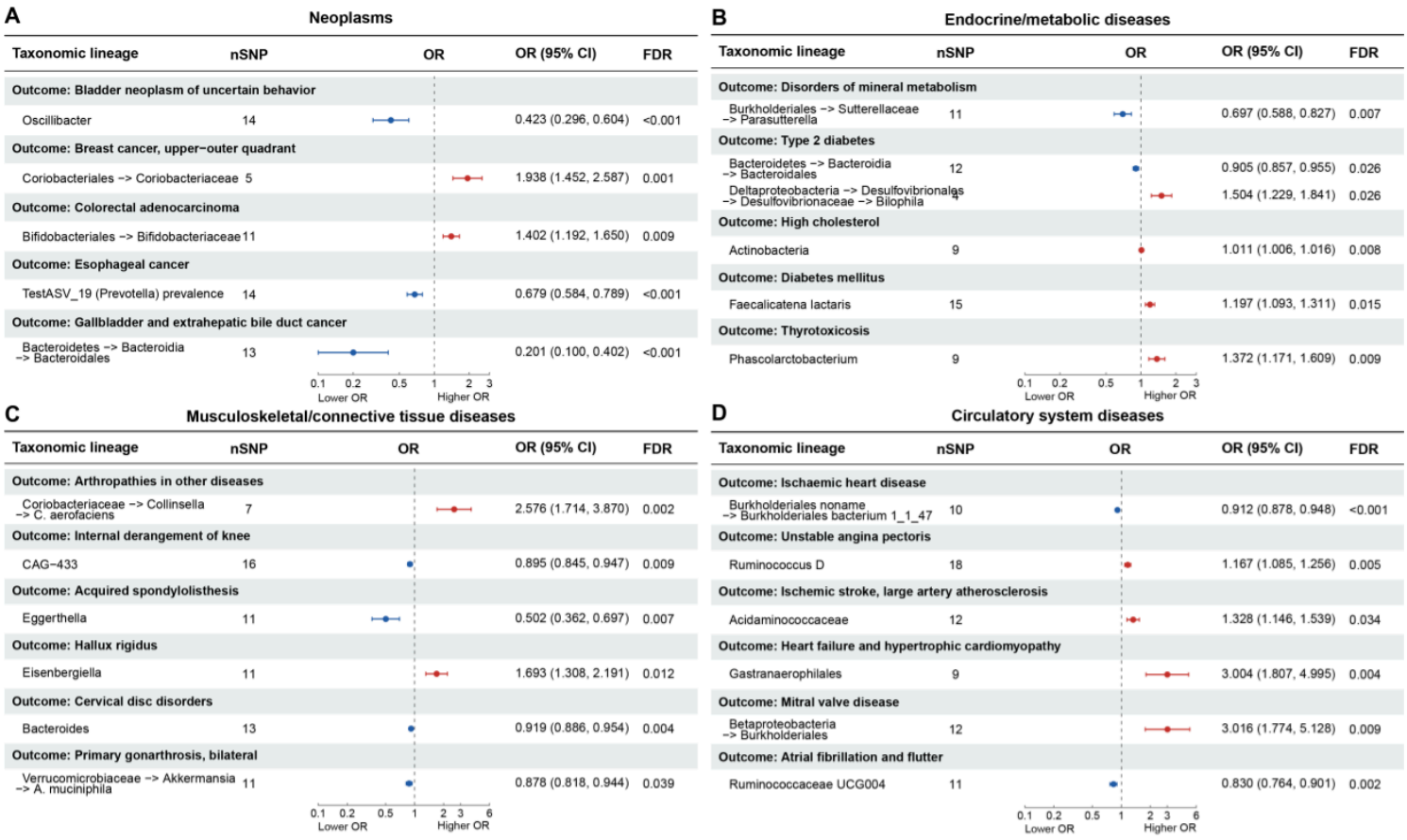


Figure 3

Representative significant MR associations across disease systems.

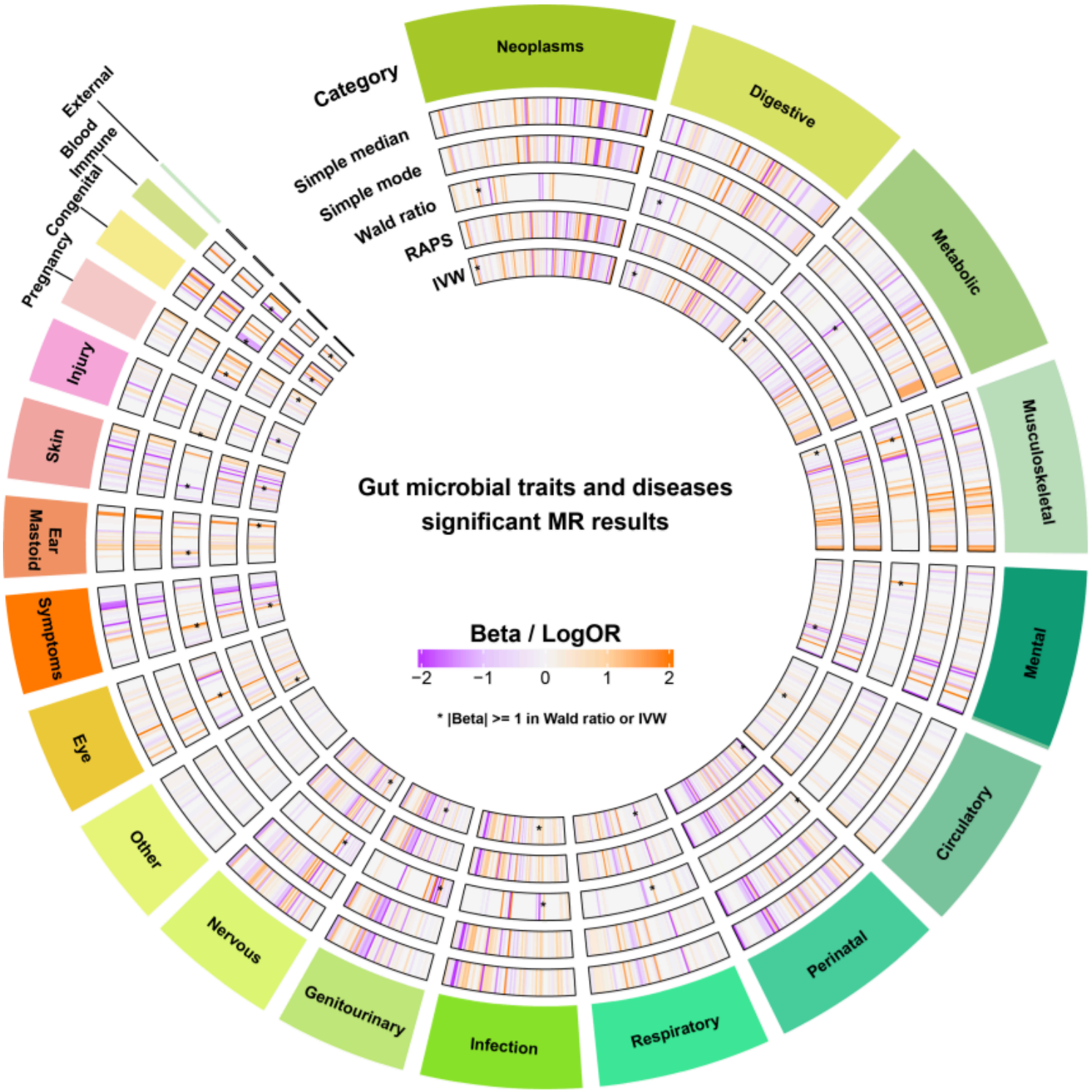


Figure 4

Multi-method sensitivity analysis of significant MR results.

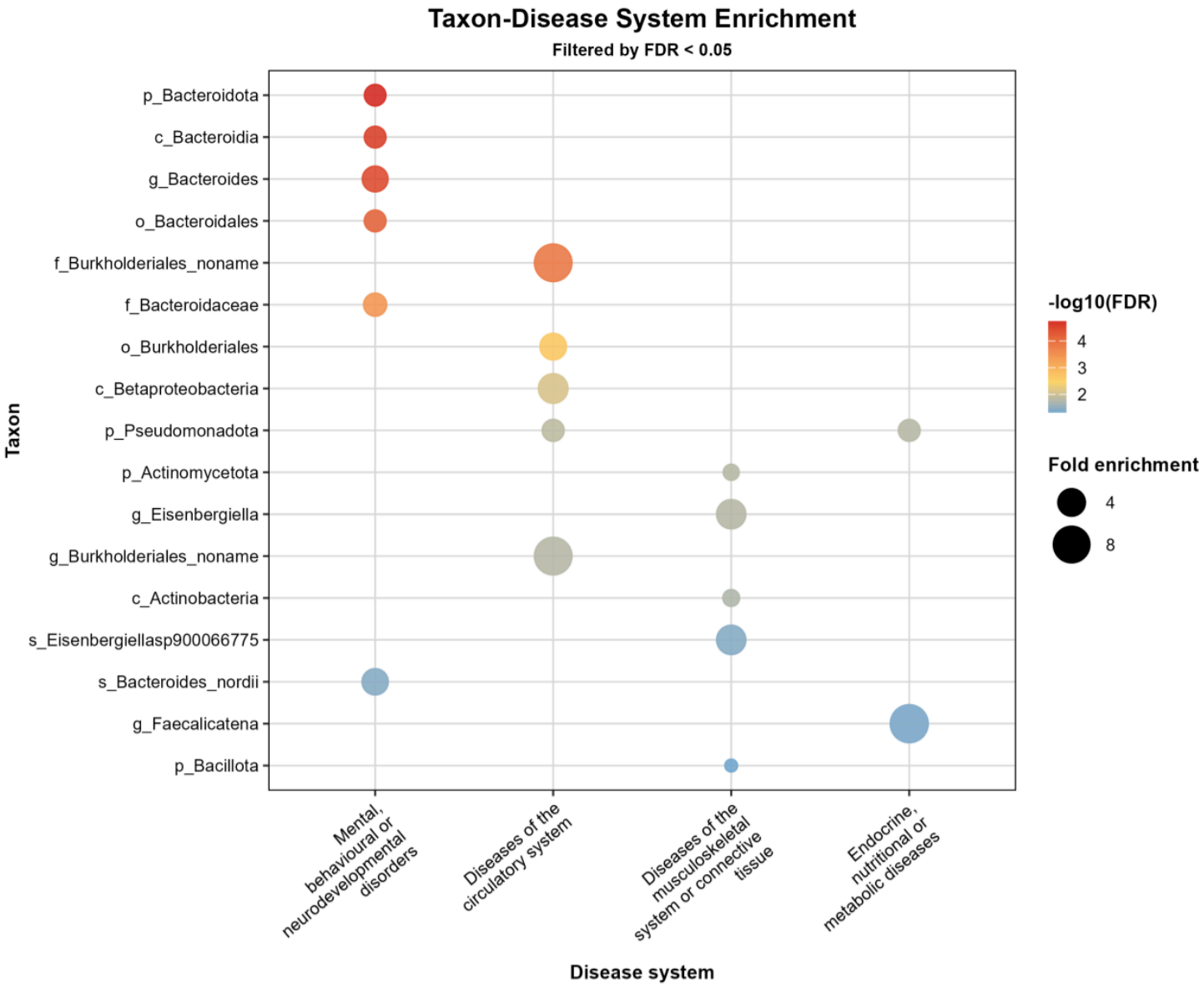


Figure 5

Enrichment of significant gut microbial trait–disease GWAS outcome association pairs across disease systems.

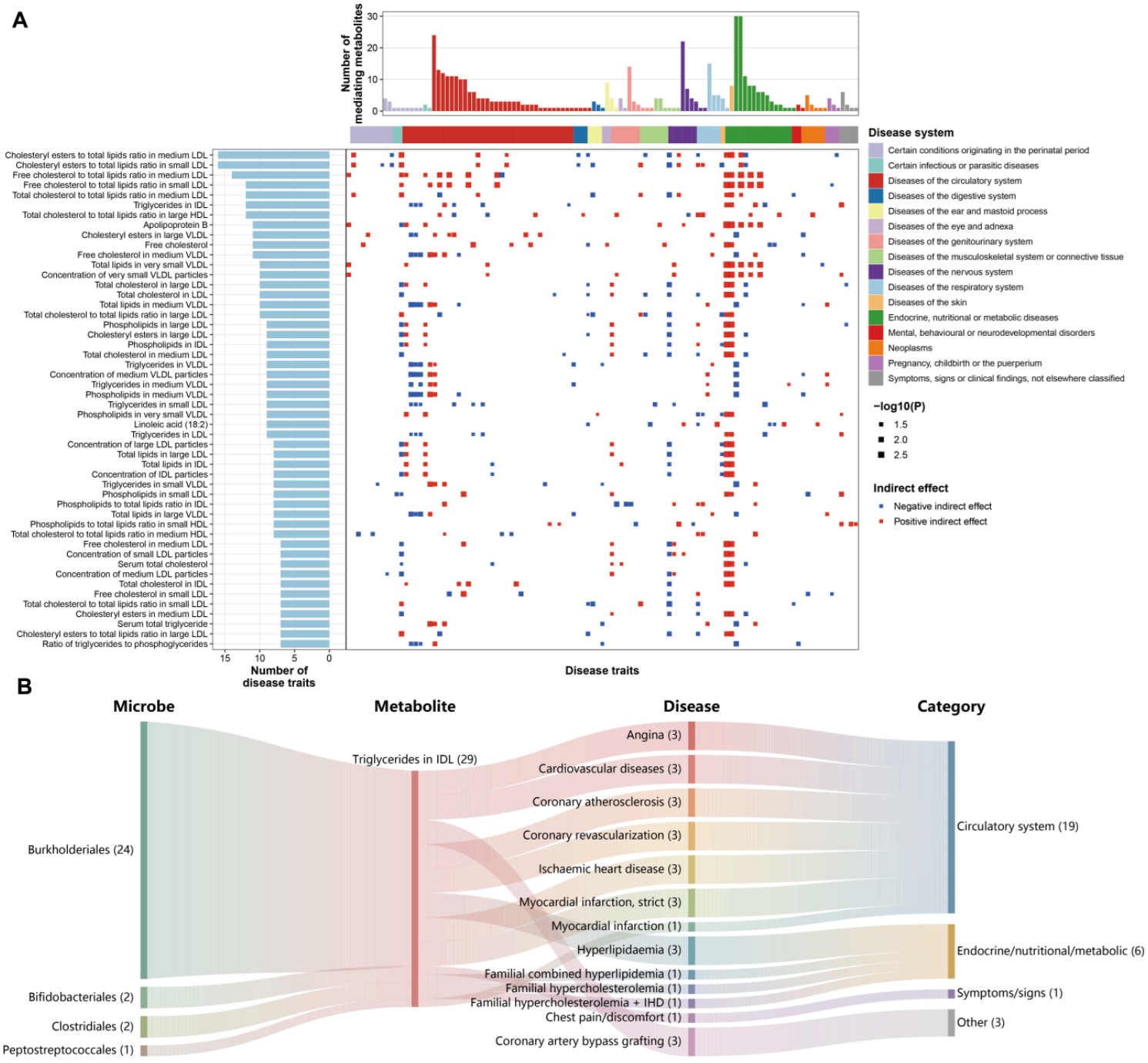


Figure 6

Candidate pathways through which gut microbiota mediate disease outcomes via metabolites.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable1.GWASdatasetsusedinthisstudy.xlsx
- SupplementaryTable2.SignificantgutmicrobiotadiseaseMRassociations.xlsx
- SupplementaryTable3.DiseasesystemenrichmentofsignificantgutmicrobialtraitdiseaseGWASoutcomepairs.xlsx
- SupplementaryTable4.Colocalizationresults.xlsx

- SupplementaryTable5.Candidatemediationpathsinvolvingcirculatingmetabolites.xlsx
- Additionalfile6SupplementaryFig.pdf