Identifying the causal relationship between gut microbiota and lymphoma disease: a Mendelian randomization study and meta-analysis

Ruizhe Huang[1,2], Boyi Liu[1,2], Wenchao Zhang[1,2], Jie Tian[1,2], Kaiwei Yu[1,2], Danna Chen[1,3,4]*

[1]Changsha Medical University, Changsha, China
[2]The First Clinical College, Changsha Medical University, Changsha, China
[3]Department of Basic Medical Sciences, Changsha Medical University, Changsha, Hunan, China
[4]Hunan Provincial University Key Laboratory of the Fundamental and Clinical Research on Functional Nucleic Acid

*Corresponding author:
Danna Chen
Department of Basic
Medical Sciences
Changsha Medical University
1501 Leifeng St
Changsha, Hunan
410219, China
Phone: +86-731-84805339
Fax: +86–731-84 478 152
E-mail: chendanna007@
gmail.com

Abstract

Introduction: Increasing evidence from observational studies and clinical trials suggests that gut microbiota (GM) are associated with lymphoma. However, it is unclear whether there is a causal relationship between GM and lymphoma.
Material and methods: To evaluate the causal relationship between GM and lymphoma, we conducted a two-sample Mendelian randomization analysis based on the genome-wide association study (GWAS). GWAS summary statistics of the GM were obtained from the MiBioGen study that including 211 taxa. GWAS summary statistics of four lymphomas were obtained from the IEU Open GWAS study. Then, we systematically performed sensitivity analyses and heterogeneity analysis to verify the reliability of our findings. Finally, we used an external cohort for validation and performed a meta-analysis of the positive results.
Results: We identified 37 causal relationships between GM and lymphoma. The results of the sensitivity analyses demonstrated the reliability of the MR analysis. Furthermore, the combined meta-analysis yielded nine significant results, with the most notable being *Terrisporobacter* (MetaOR = 2.39, 95% CI = 1.07–5.32, p = 0.03), which was identified as a pathogenic factor for diffuse large B-cell lymphoma (DLBCL). *Methanobrevibacter* (MetaOR = 0.49, 95% CI = 0.27–0.92, p = 0.03) was identified as a protective factor against DLBCL. Cyanobacteria (MetaOR = 1.94, 95% CI = 1.24–3.03, p = 0.004) were identified as a pathogenic factor for FL.
Conclusions: Our study identified a causal relationship between gut microbes and four lymphoma diseases. To further confirm the causal relationship, external validation was performed, thereby providing new insights into the subsequent mechanisms by which gut microbes mediate lymphoma development.

Key words: lymphoma, gut microbiota, Mendelian randomization, causal effect, meta-analysis.

Introduction

Lymphoma is a malignant neoplasm originating from lymphatic tissue, capable of metastasizing to any region of the body, and presenting

AMS

Ruizhe Huang, Boyi Liu, Wenchao Zhang, Jie Tian, Kaiwei Yu, Danna Chen

a range of clinical manifestations [1]. The most prevalent hematologic malignancy, non-Hodgkin's lymphoma (NHL), which accounts for 90% of all cases, and Hodgkin's lymphoma, which accounts for 10% of all cases, are the two traditional categories for lymphomas [2]. NHL has both aggressive and inert subtypes with a 5-year overall survival rate that ranges from 25% to 75%, and the prognosis is variable [3]. Genetic factors play a 10% role in the onset of lymphoma, and there is no significant difference in heritability between the sexes [4]. Many subtypes of lymphoma remain incurable with current management strategies, and more clinical trials are needed to determine novel therapies with promising activity in this disease [5].

The human intestinal flora is a collection of various microorganisms, including bacteria, fungi, and viruses, that live on the surface of the epithelial barrier of the gastrointestinal tract [6, 7]. With the advancement of molecular tools and technologies, such as 16S ribosomal RNA sequencing and metabolomics, complex host-microbiota interactions are gradually being discovered [8, 9]. Numerous studies have shown that gut microbiota (GM) play key roles in immune, metabolic and inflammatory responses [10, 11].

Studies have shown that the combination of gut microbiota and vitamin D is promising for the treatment and prevention of autoimmune and allergic diseases and that Th17 cells also play a key role [12, 13].

In addition, it has been suggested that dysregulation of gut microbial ecology jeopardizes the integrity of the immune network and may lead to acute lymphoblastic leukemia [14]. Among them, a group of potential tumor-promoting or anti-tumor microbial species have been identified [15], which lays the foundation for modulating the gut flora in cancer therapy.

An observational study found significant enrichment of *Faecalibacterium*, *Bifidobacterium*, and *Ruminococcus* abundance in patients with NHL treated with CD19 CAR-T cells [16].

Genetic variation is used as an instrumental variable (IV) in Mendelian randomization (MR), a technique that is widely used to explore potential causal relationships between environmental exposures and disease and also a widely accepted way to control for potential confounders and avoid reverse causality bias [17–19]. Combining single nucleotide polymorphism (SNP) exposure and SNP outcome associations from a separate genome-wide association study (GWAS) with MR studies of two samples allows for the creation of a single causal estimate. MR analysis must be predicated on three fundamental tenets in order to guarantee the validity of our findings: (1) genetic variants must exhibit a strong correlation with the exposure factor; (2) genetic variants must not have a direct impact on the outcome; and (3) genetic variants must not be causally linked to any potential confounders [20].

Randomized controlled trials of the gut microbiota, as opposed to observational studies, may aid in the establishment of causation. Unfortunately, due to objective elements such as technology and study technique, strain screening involving early diagnosis and prognosis still has considerable limitations. These experiments were impacted by a number of variables, including antibiotic use and diet [21, 22]. In conclusion, it is still unclear whether GM and lymphomas (Hodgkin's and non-Hodgkin's lymphomas) are causally related.

Material and methods

Exposure data

Summary statistics of GM from the global consortium MiBioGen's GWAS dataset served as exposure IVs [23, 24]. This was a multi-ethnic large-scale GWAS that coordinated genome-wide genotype and 16S fecal microbiome data from 18,340 participants from 24 cohorts in the United States, Canada, Israel, Korea, Germany, Denmark, the Netherlands, Belgium, Sweden, Finland, and the United Kingdom to explore the relationship between human autosomal genetic variation and the GM. A total of 211 taxa were included, belonging to 35 families, 20 orders, 16 phyla, 9 phyla and 131 genera. The externally validated genetic data on the gut microbiota were obtained from the GWAS summary data available on the IEU Open GWAS Project website (https://gwas.mrcieu.ac.uk/).

Outcome data

GWAS summary statistics for Hodgkin's and follicular lymphoma (HL and FL), mature T/NK-cell lymphoma (MT/NKCL), and diffuse large B-cell lymphoma (DLBCL) (all with exclusion of other cancers) were obtained from the FinnGen Consortium (updated to 2021, total Ncase = 1,250, Ncontrol = 716,274). To facilitate analysis and data extraction, we accessed and analyzed outcome data directly through the OPEN GWAS website (https://∨/GWAS.mrcieu.ac.uk/). Detailed information is provided in Table I.

Instrumental variable selection

GWAS summary statistics that are available to the public were used in this study. There was no need for additional ethical approvals. The study flow chart is displayed in Figure 1.

Table I. Overview of the source of lymphoma data

GWAS ID	Year	Trait	Consortium	Case	Control	Number of SNPs	Population
DLBCL	2021	Diffuse large B-cell lymphoma (all cancers excluded)	FinnGen	209	174,006	16,380,303	European
FL	2021	Follicular lymphoma (all cancers excluded)	FinnGen	522	180,756	16,380,337	European
HL	2021	Hodgkin lymphoma (all cancers excluded)	FinnGen	369	180,756	16,380,338	European
MNKTCL	2021	Mature T/NK-cell lymphomas (all cancers excluded)	FinnGen	150	180,756	16,380,337	European

Exposure

Gut microbiota GWAS
(1) MiBioGen consortium
(2) 18340 samples with 24 cohorts
(3) 211 taxa, 5 level groups, including 119 known genera

SNPs

Confounder factors

Selection of instrumental variables (SNPs)
(1) Genome-wide significance: $p < 1 \times 10^{-5}$
(2) LD tendency: $R^2 < 0.001$ and clumping window = 10 Mb
(3) Removal of weak instrumental variables (IVs) ($F < 10$)

Harmonization of data
Deletion of palindromes SNPs

Outcome

Lymphoma diseases
(1) Mature T/NK cell lymphoma
(2) Diffuse large β-cell lymphoma
(3) Hodgkin's lymphoma
(4) Follicular lymphoma

MR analysis
IVW; MR-Egger weighted median estimate

Sensitivity analysis
(1) Cochrane's Q test
(2) MR-Egger intercept test
(3) MR-PRESSO test

Meta-analysis
External validation to avoid false positives

Figure 1. Graphical abstract

We implemented rigorous quality control procedures to select the most robust IVs and ensure the validity and accuracy of our results. Initially, we selected SNPs strongly associated with GM, applying a significance threshold of $p = 1 \times 10^{-5}$. Secondly, an important step in MR analysis is to ensure that the effect of the SNP on exposure corresponds to the same allele as the effect on outcome. After matching the results, we removed the palindromic SNP. (Palindromic SNPs are SNPs with A/T or G/C alleles.) Thirdly, only independent

SNPs were kept after clustering the SNPs in each bacterial classification unit. The linkage disequilibrium (LD) threshold was set to $r^2 < 0.01$ and the clustering window size to 10,000 kb [25]. Furthermore, the influence of pleiotropy was eliminated by applying the MR pleiotropy residuals sum and outlier (MR-PRESSO) test, which was used to identify possible horizontal pleiotropy by removing outliers [26]. In addition, F statistics were computed in order to assess weak instrument bias $F = (R^2(n - 1 - k))/((1 - R^2)k) < 10$ was considered

a weak IV, R^2 is the proportion of genetic variation explaining exposure, N is the sample size of exposure, and k is the number of SNPs used for MR analysis. The weak instrumental variable was excluded from the subsequent analysis [12].

Statistical analysis

MR analysis

To investigate possible causal relationships between lymphoma and GM we performed MR analysis. The gut microbiome with multiple IVs was characterized using three prominent MR methods: the weighted median estimator (WME) [27], MR-Egger regression [28], and the inverse variance weighted (IVW) method [29]. If there was only a single IV, only the Wald ratio method was used to perform the analysis. The results for multiple IVs are mainly based on the IVW method, supplemented by the other two methods. This is because it has been observed that in some cases, the IVW method is more effective than other methods [30].

Sensitivity analysis

We tested for heterogeneity using Cochrane's Q test. $P < 0.05$ was considered heterogeneous. The intercept of MR-Egger regression was used to assess the presence of horizontal pleiotropy in IVS, and MR-PRESSO was used to propose SNPs with pleiotropy. Finally, FDR correction was applied to reduce the possibility of false positives; $P_{FDR} < 0.05$ was considered a significant result.

Validation based on external cohorts and meta-analysis

To further validate the reliability of the NMR analysis results, two large-scale external cohort studies were used to validate the NMR-positive results: the FINRISK study in Finland and the Lifeline study in the northern Netherlands. It is noteworthy that to enhance the robustness of the results, datasets that were at disparate levels of bacterial classification were excluded. In the absence of relevant cohorts in the datasets, a meta-analysis was not performed. The external cohort validation process was conducted using the identical parameters and procedures as those employed for the MR analysis described above. A meta-analysis was performed on the analyzed results to incorporate effect sizes. The I^2 statistic was employed to assess heterogeneity in the meta-analysis. When $I^2 < 50\%$ and $p > 0.05$, a fixed-effects model was utilized. Conversely, when $I^2 \geq 50\%$ and $p \leq 0.05$, a random-effects model was employed.

R (version 4.3.2) packages that use TwoSampleMR, data.table, dplyr, doParallel, ggthemes, magrittr, p.adjust and readr were used for all statistical studies.

Results

Two-sample MR analysis of lymphoma

The F-statistics for all IVs were greater than 10, indicating that there is no weak IV bias in our analysis (Supplementary Table SI). MR analyses revealed 37 causal associations between genetically predicted GM and four lymphoma diseases. In Hodgkin's lymphoma, we found 15 causal relationships. In follicular lymphoma disease, we found 11 causal relationships. In diffuse large B-cell lymphoma, we found 5 causal relationships. In mature T/NK-cell lymphomas, we found 6 causal relationships.

MR analysis showed that 6 GM were associated with an increase in the incidence of Hodgkin's lymphoma and 9 were associated with a decrease in the incidence of Hodgkin's lymphoma, with the more significant ones being genera. *Butyricimonas* (OR = 2.125, 95% CI = 1.173–3.849, $p = 0.013$) and order Bifidobacteriales (OR = 2.154, 95% CI = 1.117–3.962, $p = 0.014$) were significantly associated with increased risk of Hodgkin's lymphoma. The genus *Peptococcus* (OR = 0.573, 95% CI = 0.390–0.840, $p = 0.004$) was associated with significantly decreased risk of Hodgkin's lymphoma.

Three GM were associated with increased incidence of diffuse large B-cell lymphoma, and two GM were associated with decreased incidence of diffuse large B-cell lymphoma, the more significant of which were the *Eubacterium* genus group (OR = 0.236, 95% CI = 0.079–0.712, $p = 0.01$), which was negatively correlated, and the genus *Terrisporobacter* (OR = 3.607, 95% CI = 1.189–10.938, $p = 0.023$), which was positively correlated.

Four GM were associated with increased incidence of follicular lymphoma and seven were associated with decreased incidence of follicular lymphoma, the more significant of which were *Eubacterium* (OR = 0.450, 95% CI = 0.224–0.900, $p = 0.024$), which was associated with a reduced risk, and the phylum *Proteobacteria* (OR = 1.984, 95% CI = 1.063–3.704, $p = 0.031$), associated with elevated risk.

Two GM were associated with increased incidence of mature T/NK-cell lymphoma, and four GM were associated with decreased incidence of mature T/NK-cell lymphoma, the more significant of which were the genus *Ruminococcaceae UCG003* (OR = 0.289, 95% CI = 0.102–0.821, $p = 0.020$), the genus *LachnospiraceaeUCG001* (OR = 0.381, 95% CI = 0.161–0.901, $p = 0.028$), which were associated with a reduced risk of morbidity, and the order Lactobacillales (OR = 4.751, 95% CI
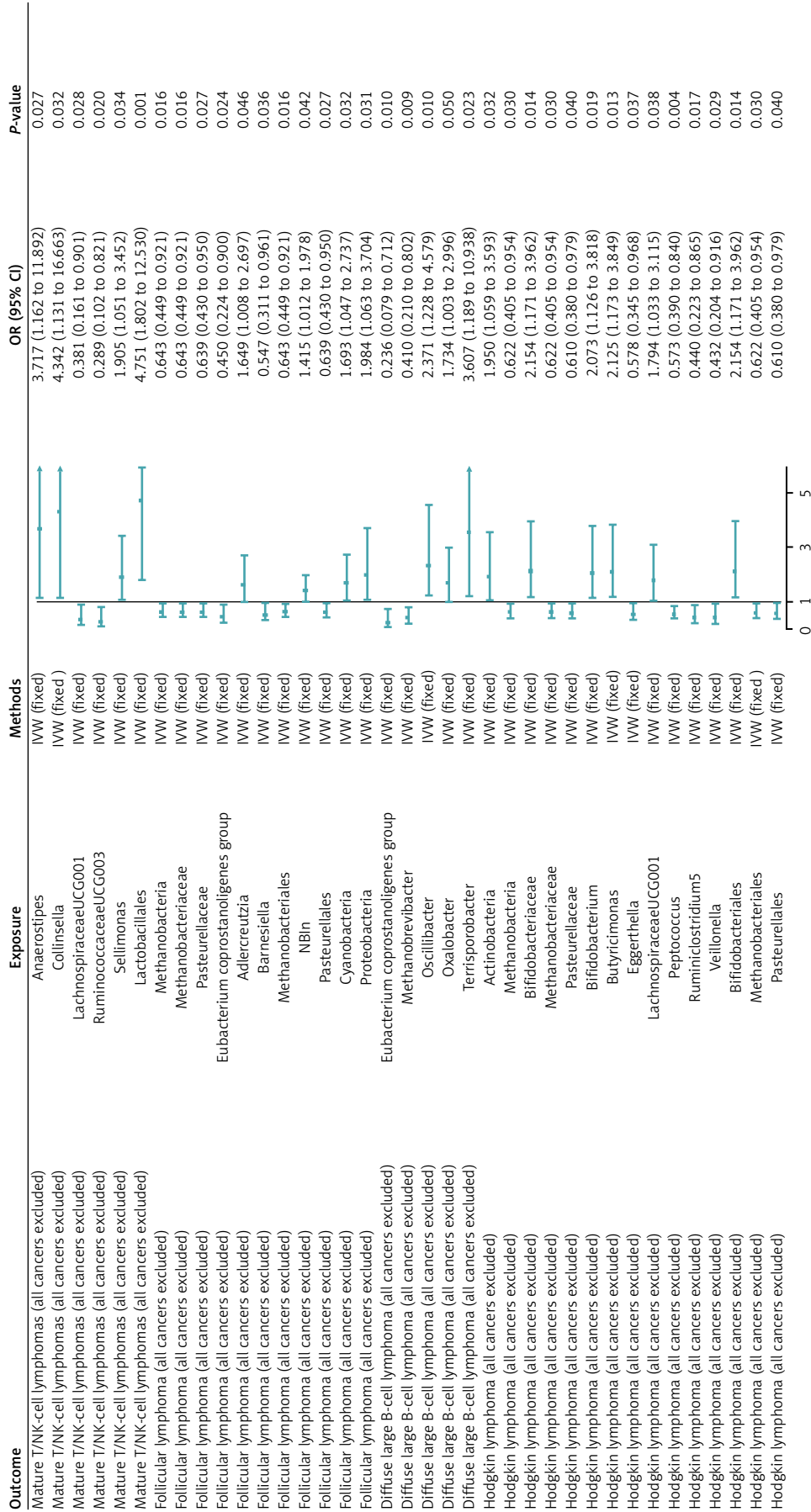
Outcome	Exposure	Methods		OR (95% CI)	P-value
Mature T/NK-cell lymphomas (all cancers excluded)	Anaerostipes	IVW (fixed)		3.717 (1.162 to 11.892)	0.027
Mature T/NK-cell lymphomas (all cancers excluded)	Collinsella	IVW (fixed)		4.342 (1.131 to 16.663)	0.032
Mature T/NK-cell lymphomas (all cancers excluded)	LachnospiraceaeUCG001	IVW (fixed)		0.381 (0.161 to 0.901)	0.028
Mature T/NK-cell lymphomas (all cancers excluded)	RuminococcaceaeUCG003	IVW (fixed)		0.289 (0.102 to 0.821)	0.020
Mature T/NK-cell lymphomas (all cancers excluded)	Sellimonas	IVW (fixed)		1.905 (1.051 to 3.452)	0.034
Mature T/NK-cell lymphomas (all cancers excluded)	Lactobacillales	IVW (fixed)		4.751 (1.802 to 12.530)	0.001
Follicular lymphoma (all cancers excluded)	Methanobacteria	IVW (fixed)		0.643 (0.449 to 0.921)	0.016
Follicular lymphoma (all cancers excluded)	Methanobacteriaceae	IVW (fixed)		0.643 (0.449 to 0.921)	0.016
Follicular lymphoma (all cancers excluded)	Pasteurellaceae	IVW (fixed)		0.639 (0.430 to 0.950)	0.027
Follicular lymphoma (all cancers excluded)	Eubacterium coprostanoligenes group	IVW (fixed)		0.450 (0.224 to 0.900)	0.024
Follicular lymphoma (all cancers excluded)	Adlercreutzia	IVW (fixed)		1.649 (1.008 to 2.697)	0.046
Follicular lymphoma (all cancers excluded)	Barnesiella	IVW (fixed)		0.547 (0.311 to 0.961)	0.036
Follicular lymphoma (all cancers excluded)	Methanobacteriales	IVW (fixed)		0.643 (0.449 to 0.921)	0.016
Follicular lymphoma (all cancers excluded)	NBln	IVW (fixed)		1.415 (1.012 to 1.978)	0.042
Follicular lymphoma (all cancers excluded)	Pasteurellales	IVW (fixed)		0.639 (0.430 to 0.950)	0.027
Follicular lymphoma (all cancers excluded)	Cyanobacteria	IVW (fixed)		1.693 (1.047 to 2.737)	0.032
Follicular lymphoma (all cancers excluded)	Proteobacteria	IVW (fixed)		1.984 (1.063 to 3.704)	0.031
Diffuse large B-cell lymphoma (all cancers excluded)	Eubacterium coprostanoligenes group	IVW (fixed)		0.236 (0.079 to 0.712)	0.010
Diffuse large B-cell lymphoma (all cancers excluded)	Methanobrevibacter	IVW (fixed)		0.410 (0.210 to 0.802)	0.009
Diffuse large B-cell lymphoma (all cancers excluded)	Oscillibacter	IVW (fixed)		2.371 (1.228 to 4.579)	0.010
Diffuse large B-cell lymphoma (all cancers excluded)	Oxalobacter	IVW (fixed)		1.734 (1.003 to 2.996)	0.050
Diffuse large B-cell lymphoma (all cancers excluded)	Terrisporobacter	IVW (fixed)		3.607 (1.189 to 10.938)	0.023
Hodgkin lymphoma (all cancers excluded)	Actinobacteria	IVW (fixed)		1.950 (1.059 to 3.593)	0.032
Hodgkin lymphoma (all cancers excluded)	Methanobacteria	IVW (fixed)		0.622 (0.405 to 0.954)	0.030
Hodgkin lymphoma (all cancers excluded)	Bifidobacteriaceae	IVW (fixed)		2.154 (1.171 to 3.962)	0.014
Hodgkin lymphoma (all cancers excluded)	Methanobacteriaceae	IVW (fixed)		0.622 (0.405 to 0.954)	0.030
Hodgkin lymphoma (all cancers excluded)	Pasteurellaceae	IVW (fixed)		0.610 (0.380 to 0.979)	0.040
Hodgkin lymphoma (all cancers excluded)	Bifidobacterium	IVW (fixed)		2.073 (1.126 to 3.818)	0.019
Hodgkin lymphoma (all cancers excluded)	Butyricimonas	IVW (fixed)		2.125 (1.173 to 3.849)	0.013
Hodgkin lymphoma (all cancers excluded)	Eggerthella	IVW (fixed)		0.578 (0.345 to 0.968)	0.037
Hodgkin lymphoma (all cancers excluded)	LachnospiraceaeUCG001	IVW (fixed)		1.794 (1.033 to 3.115)	0.038
Hodgkin lymphoma (all cancers excluded)	Peptococcus	IVW (fixed)		0.573 (0.390 to 0.840)	0.004
Hodgkin lymphoma (all cancers excluded)	Ruminiclostridium5	IVW (fixed)		0.440 (0.223 to 0.865)	0.017
Hodgkin lymphoma (all cancers excluded)	Veillonella	IVW (fixed)		0.432 (0.204 to 0.916)	0.029
Hodgkin lymphoma (all cancers excluded)	Bifidobacteriales	IVW (fixed)		2.154 (1.171 to 3.962)	0.014
Hodgkin lymphoma (all cancers excluded)	Methanobacteriales	IVW (fixed)		0.622 (0.405 to 0.954)	0.030
Hodgkin lymphoma (all cancers excluded)	Pasteurellales	IVW (fixed)		0.610 (0.380 to 0.979)	0.040

0 1   3   5

Figure 2. MR results and its forest plot. Causal effects for gut microbiota on T/NK cell lymphoma; causal effects for gut microbiota on diffuse large B-cell lymphoma; causal effects for gut microbiota on follicular lymphoma; causal effects for gut microbiota on Hodgkin's lymphoma

OR – odds radio, 95% CI – 95% confidence interval, IVW – inverse-variance weighted. P < 0.05 was considered statistically significant.

Ruizhe Huang, Boyi Liu, Wenchao Zhang, Jie Tian, Kaiwei Yu, Danna Chen

= 1.802–12.530, p = 0.001), which was associated with an increased risk of morbidity. See Figure 2 for details.

After FDR correction, Lactobacillales showed a significant causal association with Mature T/NK-cell lymphomas (OR = 4.751, 95% CI = 1.802–12.530, p_{FDR} = 0.030). Detailed information is provided in Supplementary Table SII.

Sensitivity analyses of the 37 significant causal associations showed that neither MR-Egger nor MR-PRESSO analyses detected the presence of horizontal pleiotropy. In addition, there was no significant heterogeneity among the selected SNPs according to the Cochrane Q test (p > 0.05). Detailed information is provided in Table II.

Validation based on external cohorts and meta-analysis

The dataset of 16 intestinal flora used to validate the results of the MRI analysis was obtained from the IEU Open GWAS project and is consistent with the present study at the level of intestinal flora classification to ensure good homogeneity. No meta-analysis was performed for the other flora as there was no corresponding dataset for them. For DLBCL, the heterogeneity test showed that no dataset with I^2 > 50% and p < 0.05 existed, so all were meta-analyzed using a fixed-effects model. Meta-analysis results showed that *Oxalobacter*

(MetaOR = 1.58, 95% CI = 1.06–2.36, p = 0.02), *Oscillibacter* (MetaOR = 1.94, 95% CI = 1.13–3.35, p = 0.02) and *Terrisporobacter* (MetaOR = 2.39, 95% CI = 1.07–5.32, p = 0.03) had a pathogenic effect on DLBCL. *Methanobrevibacter* (MetaOR = 0.49, 95% CI = 0.27–0.92, p = 0.03) was protective against DLBCL. See Figure 3 for details.

For HL, the heterogeneity test showed that there was no dataset with I^2 > 50% and p < 0.05, so all were meta-analyzed using a fixed-effects model. The meta-analysis showed that *Actinobacteria* (MetaOR = 2.02, 95% CI = 1.34–3.03, p < 0.001), *Bifidobacteriaceae* (MetaOR = 1.74, 95% CI = 1.19–2.53, p = 0.004), *Bifidobacterium* (MetaOR = 1.56, 95% CI = 1.05–2.31, p = 0.02) had a pathogenic effect on HL and the other two bacterial groups showed no significance. See Figure 4 for details.

For FL, the heterogeneity test showed the presence of one colony with I^2 > 50% and p < 0.05 in the dataset, so the meta-analysis was performed using the random effects model, and the other meta-analysis was performed using the fixed effects model, and the results of the meta-analysis showed the pathogenic effect of *Adlercreutzia* (MetaOR = 1.70, 95% CI = 1.14–2.52, p = 0.008), *Cyanobacteria* (MetaOR = 1.94, 95% CI = 1.24–3.03, p = 0.004) on FL, and the other results were not significant. See Figure 5 for details.

For NKTCL, the meta-analysis result was non-significant (Figure 6).

Table II. Sensitivity analysis of the causal association between gut microbiota and lymphoma

Outcome	Exposure	Method	Q p-val	MR-PRESSO p-val	Egger intercept	Egger p-val
Diffuse large B-cell lymphoma	Eubacterium coprostanoligenes group	Inverse variance weighted	0.30	0.735	0.127	0.37
Diffuse large B-cell lymphoma	Methanobrevibacter	Inverse variance weighted	0.86	0.898	−0.119	0.53
Diffuse large B-cell lymphoma	Oscillibacter	Inverse variance weighted	0.11	0.140	0.0122	0.93
Diffuse large B-cell lymphoma	Oxalobacter	Inverse variance weighted	0.88	0.890	−0.154	0.40
Diffuse large B-cell lymphoma	Terrisporobacter	Inverse variance weighted	0.29	0.378	0.008	0.97
Follicular lymphoma	Methanobacteria	Inverse variance weighted	0.98	0.986	0.019	0.86
Follicular lymphoma	Methanobacteriaceae	Inverse variance weighted	0.98	0.990	0.019	0.86
Follicular lymphoma	Pasteurellaceae	Inverse variance weighted	0.81	0.811	0.023	0.67
Follicular lymphoma	Eubacterium coprostanoligenes group	Inverse variance weighted	0.84	0.847	0.015	0.87
Follicular lymphoma	Adlercreutzia	Inverse variance weighted	0.67	0.709	0.087	0.46
Follicular lymphoma	Barnesiella	Inverse variance weighted	0.49	0.515	0.118	0.24

Table II. Cont.

| Outcome	Exposure	Method	Q p-val	MR-PRESSO p-val	Egger intercept	Egger p-val
Follicular lymphoma	Methanobacteriales	Inverse variance weighted	0.98	0.985	0.019	0.86
Follicular lymphoma	NB1n	Inverse variance weighted	0.36	0.391	−0.132	0.15
Follicular lymphoma	Pasteurellales	Inverse variance weighted	0.81	0.824	0.023	0.67
Follicular lymphoma	Cyanobacteria	Inverse variance weighted	0.56	0.566	0.008	0.93
Follicular lymphoma	Proteobacteria	Inverse variance weighted	0.80	0.814	−0.065	0.30
Hodgkin lymphoma	Actinobacteria	Inverse variance weighted	0.20	0.233	−0.046	0.57
Hodgkin lymphoma	Methanobacteria	Inverse variance weighted	0.72	0.757	−0.120	0.37
Hodgkin lymphoma	Bifidobacteriaceae	Inverse variance weighted	0.84	0.854	0.088	0.29
Hodgkin lymphoma	Methanobacteriaceae	Inverse variance weighted	0.72	0.770	−0.120	0.37
Hodgkin lymphoma	Pasteurellaceae	Inverse variance weighted	0.56	0.519	0.017	0.80
Hodgkin lymphoma	Bifidobacterium	Inverse variance weighted	0.51	0.537	0.058	0.39
Hodgkin lymphoma	Butyricimonas	Inverse variance weighted	0.35	0.401	0.128	0.22
Hodgkin lymphoma	Eggerthella	Inverse variance weighted	0.37	0.374	−0.178	0.22
Hodgkin lymphoma	LachnospiraceaeUCG001	Inverse variance weighted	0.28	0.291	0.202	0.08
Hodgkin lymphoma	Peptococcus	Inverse variance weighted	0.57	0.611	0.138	0.18
Hodgkin lymphoma	Ruminiclostridium5	Inverse variance weighted	0.58	0.616	−0.082	0.21
Hodgkin lymphoma	Veillonella	Inverse variance weighted	0.92	0.928	0.149	0.60
Hodgkin lymphoma	Bifidobacteriales	Inverse variance weighted	0.84	0.845	0.088	0.29
Hodgkin lymphoma	Methanobacteriales	Inverse variance weighted	0.72	0.751	−0.120	0.37
Hodgkin lymphoma	Pasteurellales	Inverse variance weighted	0.56	0.530	0.017	0.80
Mature T/NK-cell lymphomas	Anaerostipes	Inverse variance weighted	0.25	0.299	0.128	0.34
Mature T/NK-cell lymphomas	Collinsella	Inverse variance weighted	0.63	0.689	0.152	0.44
Mature T/NK-cell lymphomas	LachnospiraceaeUCG001	Inverse variance weighted	0.61	0.620	−0.117	0.50
Mature T/NK-cell lymphomas	RuminococcaceaeUCG003	Inverse variance weighted	0.74	0.781	−0.098	0.46
Mature T/NK-cell lymphomas	Sellimonas	Inverse variance weighted	0.45	0.488	0.510	0.09
Mature T/NK-cell lymphomas	Lactobacillales	Inverse variance weighted	0.13	0.155	0.014	0.91
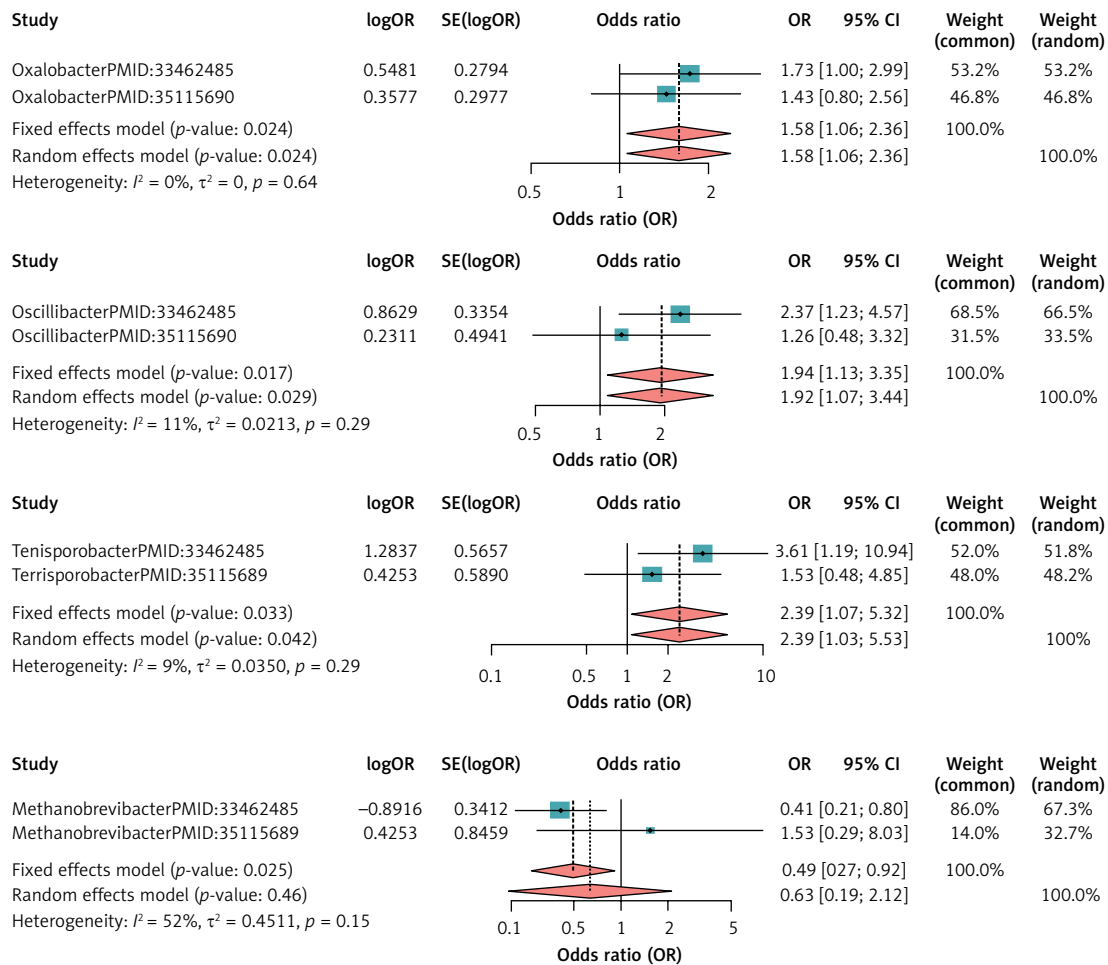
Figure 3. Meta-analysis of four gut microbiota against DLBCL

Discussion

In this study, MR analysis showed an association between gut microbiota and lymphoma. We used the extensive GWAS meta-analysis data on GM obtained from the MiBioGen consortium and lymphoma statistics published by the IEU to examine potential causal relationships. We identified 37 cases of GM with a significant causal association with lymphoma and performed a meta-analysis to obtain 9 cases of GM that remained significant.

Numerous studies have identified a possible link between the GM selected in our study and lymphoma. Our findings suggest that Eubacterium may act as a protective agent for diffuse large B-cell lymphoma and follicular lymphoma, which is consistent with previous reports [31]. A class of strictly anaerobic, Gram-positive, non-spore-forming bacteria known as the Eubacterium genus are frequently found in the human gut and are distinguished by their ability to produce butyrate [32]. Butyrate has strong anti-inflammatory properties, and early studies have shown that butyrate has anti-inflammatory properties through the inter-

action of G protein-coupled receptors GPR41 and GPR43 and strongly inhibits the release of the pro-inflammatory cytokine TNF from lamina propria mononuclear cells [33–35] in an inflammatory environment, as the pro-inflammatory cytokine TNF upregulates TLR4 in intestinal B cells. This makes B cells more sensitive to LPS [31]. TNF in combination with LPS may lead to lymphoma, while Eubacterium inhibits it, suggesting a correlation between Eubacterium and reduced risk of lymphoma [31].

The present study proposes that Terrisporobacter is a high-risk GM for the development of DLBCL and has the potential to be a specific marker or therapeutic target. Terrisporobacter is an anaerobic fungus that is frequently detected in postoperative patients suffering from comorbidities, such as abscesses and bloodstream infections [36], and is positively correlated with the risk of sepsis [37]. Furthermore, invasive fungal disease (IFD) represents a significant cause of morbidity and mortality in patients with hematologic malignancies.

Our study is the first to propose cyanobacteria as a high-risk GM for FL, and a study has suggested a possible carcinogenic pathway of cyanobacteria
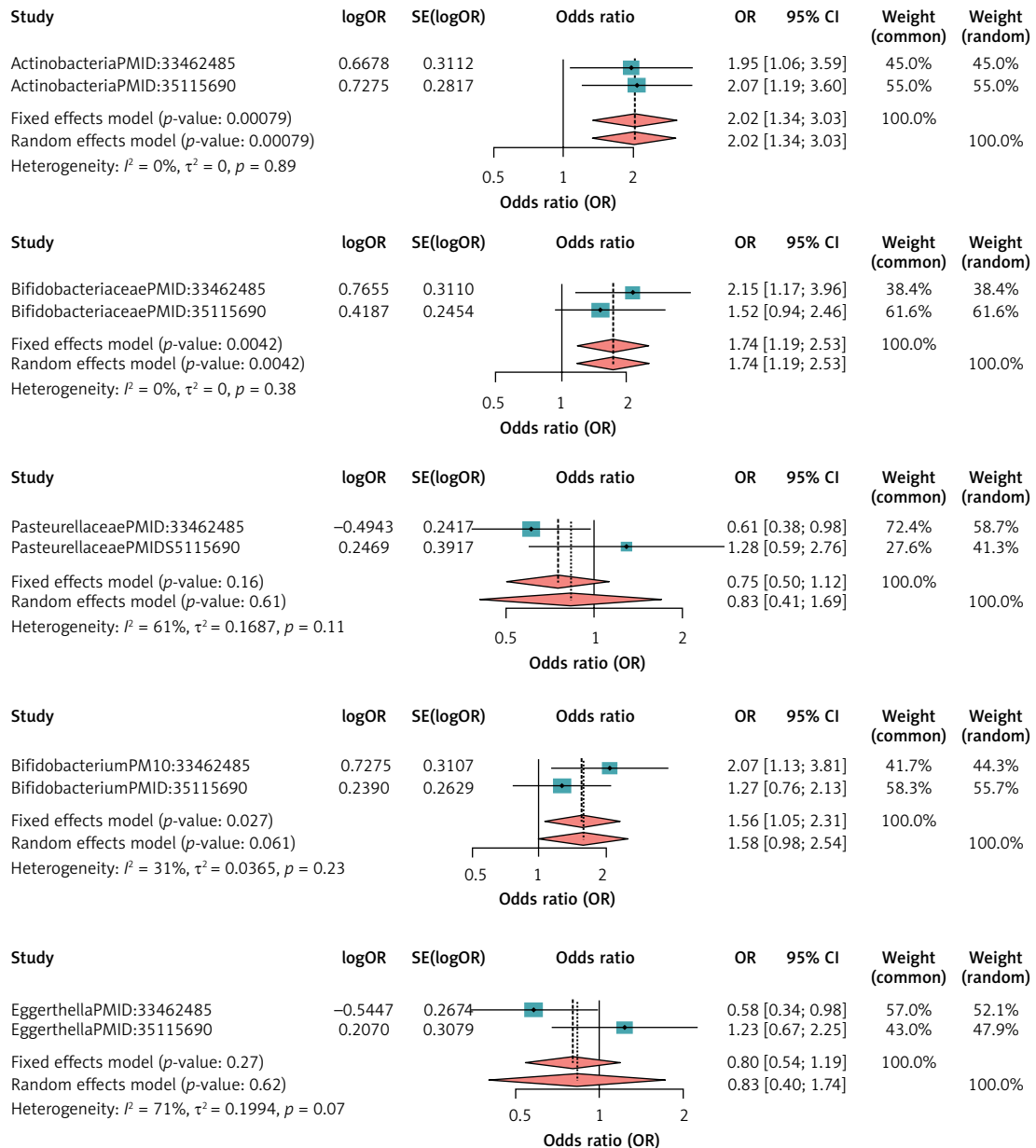
Study	logOR	SE(logOR)	Odds ratio	OR	95% CI	Weight (common)	Weight (random)
ActinobacteriaPMID:33462485	0.6678	0.3112		1.95	[1.06; 3.59]	45.0%	45.0%
ActinobacteriaPMID:35115690	0.7275	0.2817		2.07	[1.19; 3.60]	55.0%	55.0%
Fixed effects model (*p*-value: 0.00079)				2.02	[1.34; 3.03]	100.0%	
Random effects model (*p*-value: 0.00079)				2.02	[1.34; 3.03]		100.0%
Heterogeneity: $I^2 = 0\%$, $\tau^2 = 0$, $p = 0.89$							

Odds ratio (OR)

Study	logOR	SE(logOR)	Odds ratio	OR	95% CI	Weight (common)	Weight (random)
BifidobacteriaceaePMID:33462485	0.7655	0.3110		2.15	[1.17; 3.96]	38.4%	38.4%
BifidobacteriaceaePMID:35115690	0.4187	0.2454		1.52	[0.94; 2.46]	61.6%	61.6%
Fixed effects model (*p*-value: 0.0042)				1.74	[1.19; 2.53]	100.0%	
Random effects model (*p*-value: 0.0042)				1.74	[1.19; 2.53]		100.0%
Heterogeneity: $I^2 = 0\%$, $\tau^2 = 0$, $p = 0.38$							

Odds ratio (OR)

Study	logOR	SE(logOR)	Odds ratio	OR	95% CI	Weight (common)	Weight (random)
PasteurellaceaePMID:33462485	−0.4943	0.2417		0.61	[0.38; 0.98]	72.4%	58.7%
PasteurellaceaePMIDS5115690	0.2469	0.3917		1.28	[0.59; 2.76]	27.6%	41.3%
Fixed effects model (*p*-value: 0.16)				0.75	[0.50; 1.12]	100.0%	
Random effects model (*p*-value: 0.61)				0.83	[0.41; 1.69]		100.0%
Heterogeneity: $I^2 = 61\%$, $\tau^2 = 0.1687$, $p = 0.11$							

Odds ratio (OR)

Study	logOR	SE(logOR)	Odds ratio	OR	95% CI	Weight (common)	Weight (random)
BifidobacteriumPM10:33462485	0.7275	0.3107		2.07	[1.13; 3.81]	41.7%	44.3%
BifidobacteriumPMID:35115690	0.2390	0.2629		1.27	[0.76; 2.13]	58.3%	55.7%
Fixed effects model (*p*-value: 0.027)				1.56	[1.05; 2.31]	100.0%	
Random effects model (*p*-value: 0.061)				1.58	[0.98; 2.54]		100.0%
Heterogeneity: $I^2 = 31\%$, $\tau^2 = 0.0365$, $p = 0.23$							

Odds ratio (OR)

Study	logOR	SE(logOR)	Odds ratio	OR	95% CI	Weight (common)	Weight (random)
EggerthellaPMID:33462485	−0.5447	0.2674		0.58	[0.34; 0.98]	57.0%	52.1%
EggerthellaPMID:35115690	0.2070	0.3079		1.23	[0.67; 2.25]	43.0%	47.9%
Fixed effects model (*p*-value: 0.27)				0.80	[0.54; 1.19]	100.0%	
Random effects model (*p*-value: 0.62)				0.83	[0.40; 1.74]		100.0%
Heterogeneity: $I^2 = 71\%$, $\tau^2 = 0.1994$, $p = 0.07$							

Odds ratio (OR)

Figure 4. Meta-analysis of four gut microbiota against HL

associated with colorectal cancer. Cyanobacteria produce a harmful secondary metabolite, microcystin-LR (MC-LR). MC-LR activates the PI3-K/AKT signaling pathway, leading to epithelial-mesenchymal transition, and regulates the expression of miR-221/PTEN and STAT3 signaling pathways, which promotes invasion and metastasis of CRC cells [38]. In addition, some epidemiologic studies have shown that long-term exposure to MC-LR increases the incidence of several cancers, including hepatocellular carcinoma [39] and CRC [40], and may lead to cancer progression [41]. Therefore, studies are needed to explore the mechanisms.

GM has also been studied as a non-invasive diagnostic and prognostic biomarker for natural killer/T-cell lymphoma [42]. There are few studies

on Hodgkin's lymphoma [15], and our study may provide some ideas for the impact of GM on Hodgkin's lymphoma.

This study has several notable strengths and innovations. Firstly, we employed a novel approach combining MR analysis with meta-analysis of the combined results, which allowed us to investigate the causal relationship between GM and lymphoma in a way that has not been previously attempted. Secondly, our study is less susceptible to confounding factors and reverse causation than traditional observational studies, which allows us to infer the etiology of complex diseases with greater precision.

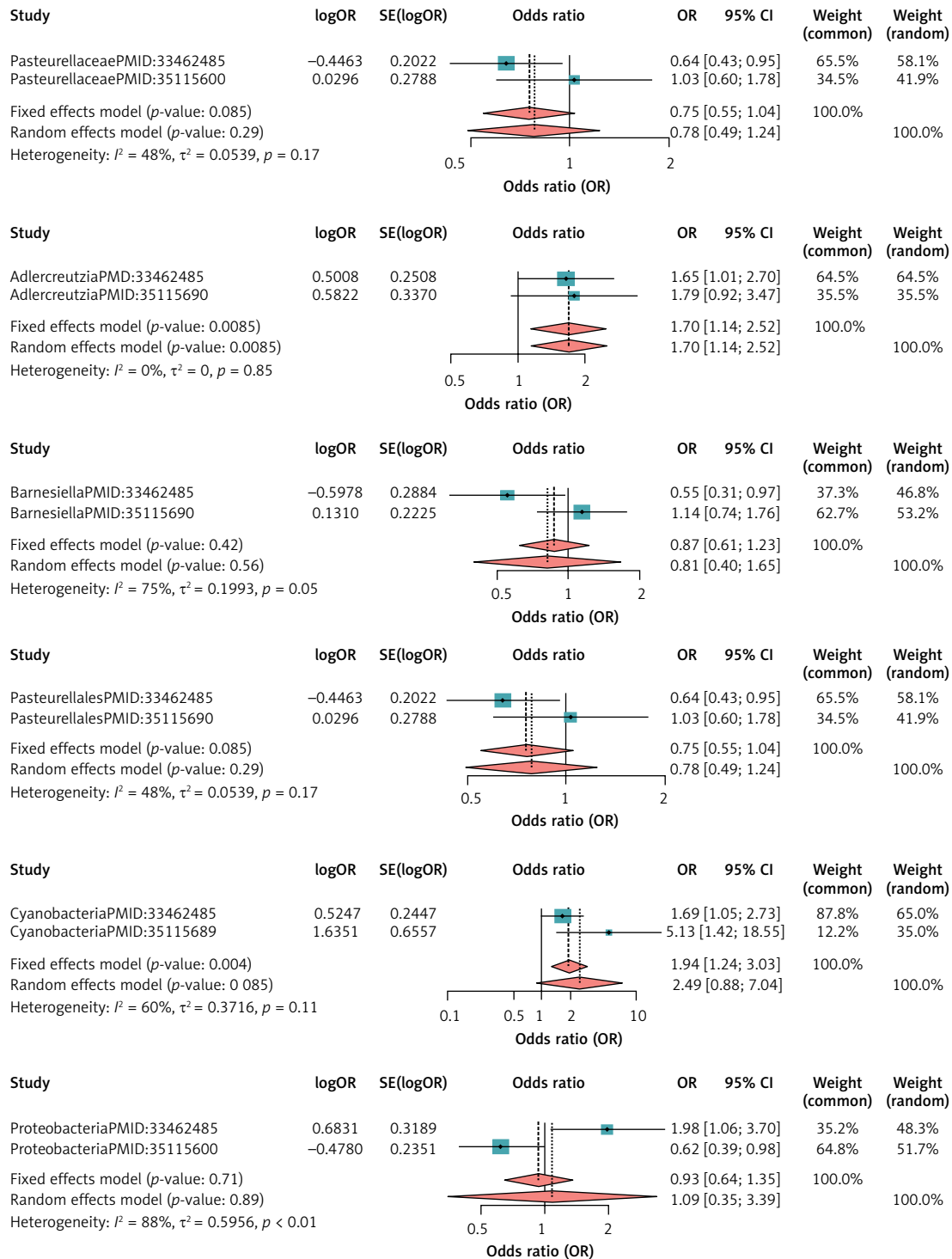However, it is important to recognize the limitations of this study. First, ethnicity affects the

Ruizhe Huang, Boyi Liu, Wenchao Zhang, Jie Tian, Kaiwei Yu, Danna Chen

Study	logOR	SE(logOR)	Odds ratio	OR	95% CI	Weight (common)	Weight (random)
PasteurellaceaePMID:33462485	−0.4463	0.2022		0.64 [0.43; 0.95]		65.5%	58.1%
PasteurellaceaePMID:35115600	0.0296	0.2788		1.03 [0.60; 1.78]		34.5%	41.9%
Fixed effects model (p-value: 0.085)				0.75 [0.55; 1.04]		100.0%	
Random effects model (p-value: 0.29)				0.78 [0.49; 1.24]			100.0%
Heterogeneity: I^2 = 48%, τ^2 = 0.0539, p = 0.17							

Study	logOR	SE(logOR)	Odds ratio	OR	95% CI	Weight (common)	Weight (random)
AdlercreutziaPMD:33462485	0.5008	0.2508		1.65 [1.01; 2.70]		64.5%	64.5%
AdlercreutziaPMID:35115690	0.5822	0.3370		1.79 [0.92; 3.47]		35.5%	35.5%
Fixed effects model (p-value: 0.0085)				1.70 [1.14; 2.52]		100.0%	
Random effects model (p-value: 0.0085)				1.70 [1.14; 2.52]			100.0%
Heterogeneity: I^2 = 0%, τ^2 = 0, p = 0.85							

Study	logOR	SE(logOR)	Odds ratio	OR	95% CI	Weight (common)	Weight (random)
BarnesiellaPMID:33462485	−0.5978	0.2884		0.55 [0.31; 0.97]		37.3%	46.8%
BarnesiellaPMID:35115690	0.1310	0.2225		1.14 [0.74; 1.76]		62.7%	53.2%
Fixed effects model (p-value: 0.42)				0.87 [0.61; 1.23]		100.0%	
Random effects model (p-value: 0.56)				0.81 [0.40; 1.65]			100.0%
Heterogeneity: I^2 = 75%, τ^2 = 0.1993, p = 0.05							

Study	logOR	SE(logOR)	Odds ratio	OR	95% CI	Weight (common)	Weight (random)
PasteurellalesPMID:33462485	−0.4463	0.2022		0.64 [0.43; 0.95]		65.5%	58.1%
PasteurellalesPMID:35115690	0.0296	0.2788		1.03 [0.60; 1.78]		34.5%	41.9%
Fixed effects model (p-value: 0.085)				0.75 [0.55; 1.04]		100.0%	
Random effects model (p-value: 0.29)				0.78 [0.49; 1.24]			100.0%
Heterogeneity: I^2 = 48%, τ^2 = 0.0539, p = 0.17							

Study	logOR	SE(logOR)	Odds ratio	OR	95% CI	Weight (common)	Weight (random)
CyanobacteriaPMID:33462485	0.5247	0.2447		1.69 [1.05; 2.73]		87.8%	65.0%
CyanobacteriaPMID:35115689	1.6351	0.6557		5.13 [1.42; 18.55]		12.2%	35.0%
Fixed effects model (p-value: 0.004)				1.94 [1.24; 3.03]		100.0%	
Random effects model (p-value: 0 085)				2.49 [0.88; 7.04]			100.0%
Heterogeneity: I^2 = 60%, τ^2 = 0.3716, p = 0.11							

Study	logOR	SE(logOR)	Odds ratio	OR	95% CI	Weight (common)	Weight (random)
ProteobacteriaPMID:33462485	0.6831	0.3189		1.98 [1.06; 3.70]		35.2%	48.3%
ProteobacteriaPMID:35115600	−0.4780	0.2351		0.62 [0.39; 0.98]		64.8%	51.7%
Fixed effects model (p-value: 0.71)				0.93 [0.64; 1.35]		100.0%	
Random effects model (p-value: 0.89)				1.09 [0.35; 3.39]			100.0%
Heterogeneity: I^2 = 88%, τ^2 = 0.5956, p < 0.01							

Figure 5. Meta-analysis of four gut microbiota against FL

results of gene-level studies. The genomics analysis data used in this study were primarily from European populations, which limits the possibility of generalizing the current findings to other ethnic groups. Larger studies are necessary to further corroborate these data. The control thresholds used in the selection of IVs were not stringent enough. Increasing the thresholds would help to identify more potentially valuable IVs. Second, this study did not externally validate all positive results, and there was a lack of data on some of the gut microbiota.

In conclusion, we performed MR analysis to determine the causal relationship between GM and four lymphoma diseases. In Hodgkin's lymphoma, we found 15 causal relationships. In follicular lymphoma, we found 11 causal relationships. In diffuse large B-cell lymphoma, we found 5 causal
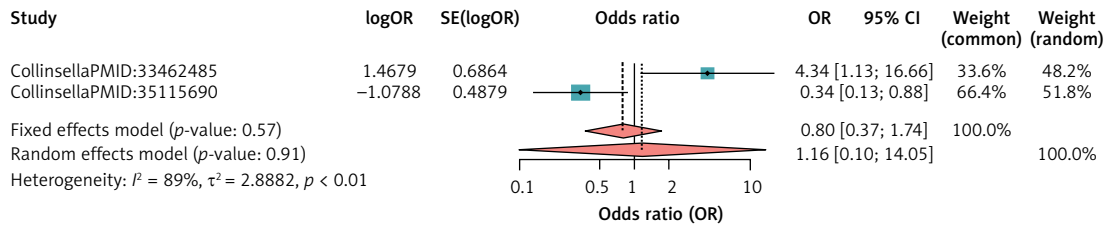
Figure 6. Meta-analysis of four gut microbiota against MT/NKCL

relationships. In mature T/NK-cell lymphomas, six causal relationships were identified and some of the positive results were externally validated. Furthermore, for the first time, it was proposed that cyanobacteria represent a high-risk factor for FL, thereby providing new insights into the mechanisms of intestinal microbe-mediated lymphoma progression and diagnostic and therapeutic follow-up.

Funding

Ethical approval

This study only used publicly available data. No original data were collected. Ethical approval for each of the studies included in the investigation can be found in the original publications.

Data availability statement

We thank the participants and researchers for providing the publicly available summary data used in this study. Gut microbiota data are from MiBioGen's (https://www.nature.com/articles/s41588-020-00763-1). Lymphoma data are available at the OPEN GWAS database (https://gwas.mrcieu.ac.uk/).

Conflict of interest

The authors declare no conflict of interest.

References

1. Yıldırım M, Kaya V, Demirpençe Ö, Paydaş S. The role of gender in patients with diffuse large b cell lymphoma treated with rituximab-containing regimens: a meta-analysis. Arch Med Sci 2015; 11: 708-14.
2. Armitage JO, Gascoyne RD, Lunning MA, Cavalli F. Non-Hodgkin lymphoma. Lancet 2017; 390: 298-310.
3. Mugnaini EN, Ghosh N. Lymphoma. Prim Care 2016; 43: 661-75.
4. Berndt SI, Vijai J, Benavente Y, et al. Distinct germline genetic susceptibility profiles identified for common non-Hodgkin lymphoma subtypes. Leukemia 2022; 36: 2835-44.
5. Ansell SM. Non-Hodgkin lymphoma: diagnosis and treatment. Mayo Clin Proc 2015; 90: 1152-63.
6. Aggarwal N, Kitano S, Puah GRY, Kittelmann S, Hwang IY, Chang MW. Microbiome and human health: current understanding, engineering, and enabling technologies. Chem Rev 2023; 123: 31-72.
7. Su YN, Wang MJ, Yang JP, et al. Effects of yulin tong bu formula on modulating gut microbiota and fecal metabolite interactions in mice with polycystic ovary syndrome. Front Endocrinol 2023; 14: 1122709.
8. de Vos WM, Tilg H, Van Hul M, Cani PD. Gut microbiome and health: mechanistic insights. Gut 2022; 71: 1020-32.
9. Liu Y, Li H, Wang X, et al. Anti-Alzheimers molecular mechanism of icariin: insights from gut microbiota, metabolomics, and network pharmacology. J Transl Med 2023; 21: 277.
10. Scott AJ, Alexander JL, Merrifield CA, et al. International cancer microbiome consortium consensus statement on the role of the human microbiome in carcinogenesis. Gut 2019; 68: 1624-32.
11. Wang K, Ma J, Li Y, et al. Corrigendum: Effects of essential oil extracted from artemisia argyi leaf on lipid metabolism and gut microbiota in high-fat diet-fed mice. Front Nutr 2024; 11: 1416210.
12. Murdaca G, Gerosa A, Paladin F, Petrocchi L, Banchero S, Gangemi S. Vitamin D and microbiota: is there a link with allergies? Int J Mol Sci 2021; 22: 4288.
13. Murdaca G, Colombo BM, Puppo F. The role of Th17 lymphocytes in the autoimmune and chronic inflammatory diseases. Intern Emerg Med 2011; 6: 487-95.
14. Peppas I, Ford AM, Furness CL, Greaves MF. Gut microbiome immaturity and childhood acute lymphoblastic leukaemia. Nat Rev Cancer 2023; 23: 565-76.
15. Bhatt AP, Redinbo MR, Bultman SJ. The role of the microbiome in cancer development and therapy. CA Cancer J Clin 2017; 67: 326-44..
16. Hu Y, Li J, Ni F, et al. Car-T cell therapy-related cytokine release syndrome and therapeutic response is modulated by the gut microbiome in hematologic malignancies. Nat Commun 2022; 13: 5313.
17. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med 2008; 27: 1133-63.

18. Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using gwas summary data. PLoS Genet 2017; 13: e1007081.

19. Burgess S, Thompson SG. Bias in causal estimates from mendelian randomization studies with weak instruments. Stat Med 2011; 30: 1312-23.

20. Sanderson E, Glymour MM, Holmes MV, et al. Mendelian randomization. Nat Rev Methods Primers 2022; 2: 6.

21. Wan Y, Wang F, Yuan J, et al. Effects of dietary fat on gut microbiota and faecal metabolites, and their relationship with cardiometabolic risk factors: a 6-month randomised controlled-feeding trial. Gut 2019; 68: 1417-29.

22. Delannoy-Bruno O, Desai C, Raman AS, et al. Evaluating microbiome-directed fibre snacks in gnotobiotic mice and humans. Nature 2021; 595: 91-5.

23. Kurilshikov A, Medina-Gomez C, Bacigalupe R, et al. Large-scale association analyses identify host factors influencing human gut microbiome composition. Nat Genet 2021; 53: 156-65.

24. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol 2003; 32: 1-22.

25. Myers TA, Chanock SJ, Machiela MJ. Ldlinkr: An R Package for rapidly calculating linkage disequilibrium statistics in diverse populations. Front Genet 2020; 11: 157.

26. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet 2018; 50: 693-8.

27. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. Genet Epidemiol 2016; 40: 304-14.

28. Burgess S, Thompson SG. Interpreting findings from mendelian randomization using the Mr-Egger method. Eur J Epidemiol 2017; 32: 377-89.

29. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol 2013; 37: 658-65..

30. Sadik A, Dardani C, Pagoni P, et al. Parental inflammatory bowel disease and autism in children. Nat Med 2022; 28: 1406-11..

31. Lu H, Xu X, Fu D, et al. Butyrate-producing eubacterium rectale suppresses lymphomagenesis by alleviating the Tnf-Induced Tlr4/Myd88/Nf- b Axis. Cell Host Microbe 2022; 30: 1139-50.e7.

32. Louis P, Hold GL, Flint HJ. The gut microbiota, bacterial metabolites and colorectal cancer. Nat Rev Microbiol 2014; 12: 661-72.

33. Ang Z, Ding JL. Gpr41 and Gpr43 in obesity and inflammation - protective or causative? Front Immunol 2016; 7: 28.

34. Buda A, Qualtrough D, Jepson MA, Martines D, Paraskeva C, Pignatelli M. Butyrate downregulates alpha2beta1 integrin: a possible role in the induction of apoptosis in colorectal cancer cell lines. Gut 2003; 52: 729-34.

35. Rooks MG, Garrett WS. Gut microbiota, metabolites and host immunity. Nat Rev Immunol 2016; 16: 341-52.

36. Cheng MP, Domingo MC, Lévesque S, Yansouni CP. A case report of a deep surgical site infection with Terrisporobacter glycolicus/T. mayombei and review of the literature. BMC Infect Dis 2016; 16: 529.

37. Chen JH, Zeng LY, Zhao YF, et al. Causal effects of gut microbiota on sepsis: a two-sample mendelian randomization study. Front Microbiol 2023; 14: 1167416.

38. Jiang X, Zhang H, Zhang H, et al. Microcystin-Lr-induced interaction between M2 tumor-associated macrophage and colorectal cancer cell promotes colorectal cancer cell migration through regulating the expression of Tgf- 1 and Cst3. Int J Mol Sci 2023; 24: 10527.

39. Liu W, Wang L, Yang X, et al. Environmental microcystin exposure increases liver injury risk induced by hepatitis B virus combined with aflatoxin: a cross-sectional study in Southwest China. Environ Sci Technol 2017; 51: 6367-78.

40. Martínez Hernández J, López-Rodas V, Costas E. Microcystins from tap water could be a risk factor for liver and colorectal cancer: a risk intensified by global change. Med Hypotheses 2009; 72: 539-40.

41. Xu P, Zhang XX, Miao C, et al. Promotion of melanoma cell invasion and tumor metastasis by microcystin-Lr via phosphatidylinositol 3-kinase/Akt pathway. Environ Sci Technol 2013; 47: 8801-8.

42. Shi Z, Hu G, Li MW, et al. Gut microbiota as non-invasive diagnostic and prognostic biomarkers for natural killer/T-cell lymphoma. Gut 2023; 72: 1999-2002.