European Journal of Clinical Nutrition www.nature.com/ejcn

ARTICLE

Check for updates

Nutrition and Health (including climate and ecological aspects)

# Causal relationship between gut microbiota and serum vitamin D: evidence from genetic correlation and Mendelian randomization study

Xiaolin Yang[1,2], Qingtian Zhu[1,2], Lei Zhang[3], Yufang Pei[4], Xiaojian Xu[5], Xinnong Liu[1,2], Guotao Lu[1,2], Jiajia Pan[6]✉ and Ying Wang[7]✉



**BACKGROUND:** The gastrointestinal microbiota is emerging as an important mediator in intestinal metabolism, such as vitamin D absorption.
**METHODS:** To elucidate the causality of microbiota and vitamin D, we used linkage disequilibrium score (LDSC) regression and two-sample Mendelian randomization (MR) methods with largest genome-wide association study (GWAS) summary statistics to identify specific taxa that are linked to serum 25-hydroxyvitamin D (25(OH)D).
**RESULTS:** We found that *Ruminiclostridium9* was significantly genetically correlated with 25(OH)D at nominal significance ($r_g = 0.43$, $P = 0.04$). Applying the inverse variance weighted (IVW) method, we identified that doubling the genetic liability of abundance of *Erysipelotrichia, Erysipelotrichaceae* and *Erysipelotrichales* reduced the concentration of 25(OH)D by 0.06 standard deviation (SD) ($\beta_{IVW} = -0.06$, s.e. = 0.01, $P = 1.48 \times 10^{-6}$, $P_{FDR} = 1.93 \times 10^{-4}$) and, in turn, one SD increment in genetically determined serum 25 (OH)D caused a 0.16 SD decrease in the relative abundance of *Phascolarctobacterium* ($\beta_{IVW} = -0.16$, s.e. = 0.04, $P = 2.48 \times 10^{-4}$, $P_{FDR} = 0.02$) after removing pleiotropic instruments and outliers. Moreover, four MR methods were also used to evaluate causality, the results of which supported these findings. Leave-one-out analyses showed that the results were robust with regard to alterations in the single nucleotide polymorphisms (SNPs) we selected.
**CONCLUSIONS:** In conclusion, our results support the hypothesis that the gut microbiota mediates the absorption of serum vitamin D supplementation and interacts with it closely. These microbiota are potential therapeutic targets for promoting serum vitamin D homeostasis.

*European Journal of Clinical Nutrition*; https://doi.org/10.1038/s41430-021-01065-3

## INTRODUCTION

More than 100 trillion bacteria and archaea in the human gastrointestinal tract make up a complete gut microbiota environment, which has been recognized to be involved in the pathology of such diseases as inflammatory bowel disease and obesity [1]. To date, the gut microbiota has been implicated as an endocrine organ, having an effect on disease progression by an increasing number of studies [2, 3]. Animal studies have demonstrated a causal role for gut microbiota in the development of type 2 diabetes [4, 5] and faecal microbiota transplantation underlines the applicable value of gut microbiota in treating obesity [6–9]. In addition, an increasing number of microbiome-wide association studies have also revealed a broad range of associations between the gut microbiome and other complex traits, including vitamin D [10]. Causality, however, remains largely undetermined.

Vitamin D, a significant substance in bone metabolism, has also been reported to be associated with gut microbiota in patients with various diseases, including multiple sclerosis [11], depression [12] and cancers [13, 14]. Nevertheless, epidemiological evidence on the association of circulating vitamin D and gut microbiota in the general population remains inconclusive. A cross-sectional study identified a negative association with *Prevotella* and a strongly positive association with *Bacteroides* in healthy participants [15], while another similar study observed contrasting results [16]. Additionally, other epidemiological studies have been conducted as well and have obtained inconsistent results [17–19]. On the other hand, the vitamin D receptor (VDR) has been

[1]Institute of digestive diseases, The Affiliated Hospital of Yangzhou University, Yangzhou University, Yangzhou, Jiangsu, PR China. [2]Pancreatic Center, Department of Gastroenterology, The Affiliated Hospital of Yangzhou University, Yangzhou University, Yangzhou, Jiangsu, PR China. [3]Center for Genetic Epidemiology and Genomics, School of Public Health, Medical College of Soochow University, Suzhou, Jiangsu, PR China. [4]Department of Epidemiology and Health Statistics, School of Public and Health, Medical College of Soochow University, Suzhou, Jiangsu, PR China. [5]Jiangsu Provincial Yangzhou Environmental Monitoring Center, Yangzhou, Jiangsu, PR China. [6]Intensive Care Unit, The Affiliated Hospital of Yangzhou University, Yangzhou University, Yangzhou, Jiangsu, PR China. [7]Department of Endocrinology, The Affiliated Hospital of Yangzhou University, Yangzhou University, Yangzhou, Jiangsu, PR China. ✉email: jjpan@yzu.edu.cn; wangying1972@yzu.edu.cn

Received: 19 November 2020 Revised: 10 December 2021 Accepted: 30 December 2021
Published online: 19 January 2022

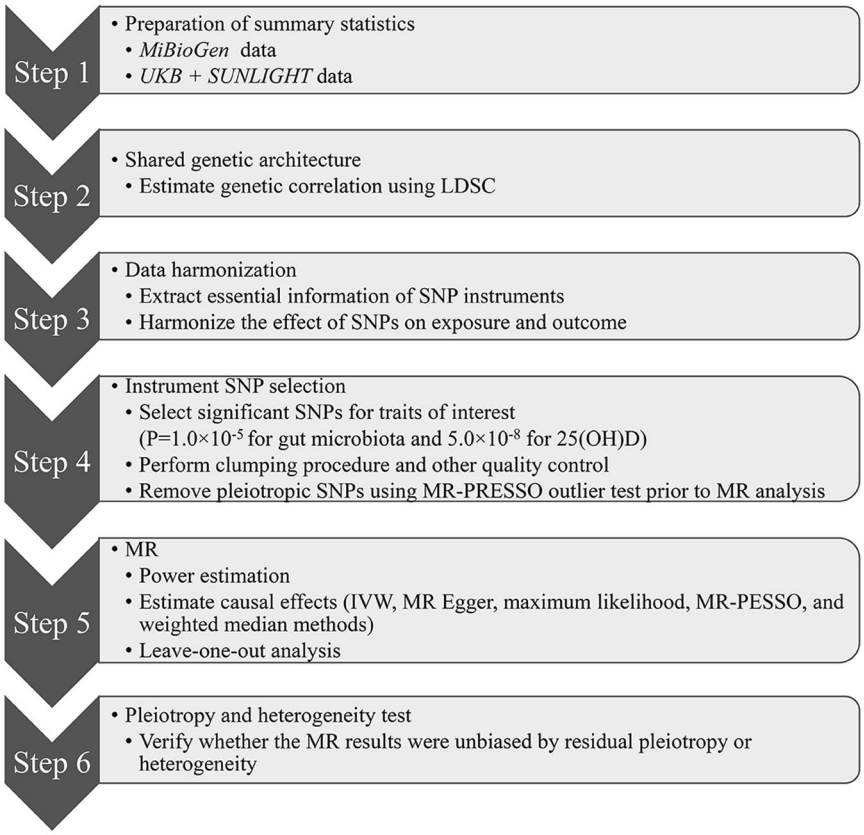


**Fig. 1 Diagrammatic description of MR analysis in this study.** MR Mendelian randomization, LDSC linkage disequilibrium score regression.

reported to be associated with gut microbiota [20, 21], adding evidence of the close relationship between vitamin D and gut microbiota. Given the presence of reverse causality and underlying confounding, such findings from observational studies are difficult to infer causality. Only two RCTs have investigated the relationship between circulating vitamin D and gut microbiota, and their sample sizes were modest ($n = 26$ and $n = 127$) [22, 23]. Therefore, a comprehensive investigation of their causal role in the general population is warranted.

Linkage disequilibrium score (LDSC) regression [24] and Mendelian randomization (MR) [25] allow researchers to dissect associations by leveraging genetic data generated from cross-sectional studies. LDSC allows assessment of genetic correlation between two traits using summary data and is robust with regard to overlapping samples. MR enables an assessment of potential causal associations between two traits based on Mendelian law, which is equivalent to a natural RCT. Recently, published large summary data on gut microbiota and 25-hydroxyvitamin D (25(OH)D) GWAS have provided an opportunity to deploy LDSC and MR approaches to infer shared and causal relationships. In the present work, we adopted a combination of LDSC and MR to evaluate the shared genetic architecture and detect promising causality, utilizing two large-scale GWAS datasets of gut microbiota and 25(OH)D mainly from two consortia in the community.

## METHODS

### Study design

We used LDSC to assess the correlation in allele effects across the entire genome and to find a global relationship between gut microbiota and 25(OH)D. The direction of causal association was assessed by MR and the power of the causality estimate was also evaluated. Before formal MR analyses, we used MR-PRESSO global and outlier tests [26] to identify and remove pleiotropic single nucleotide polymorphisms (SNPs). Causal effects were estimated using an inverse variance weighted (IVW) method [27]. Four additional MR methods, including MR-Egger [28], maximum likelihood [29], MR-PRESSO [26], and weighted median [30] models, were adopted to consider situations slightly violating MR assumptions. Significant signals were tested again to verify whether the results were unbiased by residual pleiotropic or heterogeneous effects. Additionally, leave-one-out analysis was used to examine the robustness of the results (Fig. 1).

### Data sources

All contributing studies in this work were approved by institutional review boards, and written informed consent was available from primary studies, as described elsewhere [10, 31]. Both of the GWAS data are publicly available.

*UKB + SUNLIGHT data.* The GWAS of 25(OH)D we adopted is a meta-analysis of a population-based cohort from two large consortia, consisting of 496,946 valid individuals of European ancestry [31]. In the UKB cohort, 25(OH)D concentrations from 417,580 UKB participants were measured in blood assays using chemiluminescent immunoassays and converted by natural-log transformation. The SUNLIGHT consortium comprised 79,366 subjects from 31 cohorts, mainly in Europe and the USA. Every constituent cohort adopted different measurement approaches, mainly radioimmunoassays, mass spectrometry and immunoassays, to evaluate serum 25(OH)D concentrations. To ensure consistent quality control in these two studies, the 25(OH)D concentration was natural-log transformed; the covariates in the additive model were mainly similar. After data re-management, a total of 6,912,294 common SNPs were extracted to perform a meta-analysis weighted by sample size. GWAS summary statistics are publicly available at https://cnsgenomics.com/data/revez_20/Revezetal2020_25OHD_SUNLIGHTmeta.gz.

*MiBioGen data.* *MiBioGen* is an international consortium devoted to better understanding the genetic architecture of gut microbiota and has assembled 24 population-based cohorts incorporating approximately 18,340 participants [10]. Each cohort surveyed the gut microbiota via 16S rRNA sequencing and genotyped their participants with full-genome

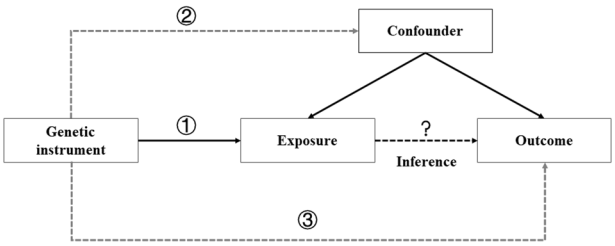


**Fig. 2 Mendelian randomization study rationale.** Mendelian randomization relies on three key assumptions: ① genetic instruments (*G*) are associated with exposure (*X*); ② genetic instruments are not associated with confounders; ③ genetic instruments influence outcome (*Y*) only through exposure. The establishment of association between genetic instruments and outcomes is supposed to be causality of exposure to outcome.

SNP arrays. Only genera and taxa at higher taxonomic levels were analyzed. All 16S data from individual cohorts were processed with a uniform procedure; transformation procedures, corrections for covariates and the thresholds set for the bacterial taxa to be included in the analysis in the original research [10]. After imputation with the Haplotype Reference Consortium [32] (HRC 1.0 or 1.1) as a reference panel, each cohort yielded approximately 39.1 million SNPs, of which 4 to 6 million variants passed standard quality control (https://mibiogen.gcc.rug.nl/).

### Estimate of genetic correlations

Cross-trait linkage disequilibrium score (LDSC) regression is an extended methodology of single-trait LDSC that allows researchers to estimate genetic correlations between two traits [24]. In cross-trait LDSC analysis, sample overlap only affects the intercept of the regression model but does not affect the slope, which means that the genetic correlation estimate is robust with regard to sample overlap. We used a standalone script developed by Brendan et al. [24] to calculate the genetic correlation with default parameters, in which the reference sample from the 1000 Genomes Project with European ancestry used in this analysis was in accordance with the participants we studied.

### Instrument selection

Considering the small number of loci identified for gut microbiota, we set the significance level at $P = 1.0 \times 10^{-5}$ to extract sufficient candidate instruments. Unlike the above inclusion criterion, conventional whole genome-wide association significance ($P = 5.0 \times 10^{-8}$) can ensure sufficient instruments for 25(OH)D. The clumping algorithm was implemented in PLINK 1.9 with --window-size 250 KB and --r2 0.01 to select independent SNPs ($r^2 < 0.01$), in which European subjects from the 1000 Genomes Project Phase III European were used as the linkage disequilibrium reference panel. SNPs with discordant (A/C vs. T/C) or ambiguous (A/T or C/G) alleles were removed. Other quality controls were applied to satisfy assumptions of MR: 1) first-stage F-statistic of SNPs ≥10; 2) minor allele frequency (MAF) ≥ 0.01; 3) imputation score ≥0.8, if available; 4) an acceptable allele frequency difference (≤ 0.02) comparing with populations from the 1000G Phase III European.

### Pleiotropy and heterogeneity test

Pleiotropy, defined as a gene property that affects multiple functions, is widespread in genetic studies. The association can be overestimated arising from overall unbalanced horizontal pleiotropy. We conducted pleiotropy and heterogeneity tests before and after MR analysis to ensure that the results were free of horizontal pleiotropy. Prior to MR analysis, we used MR-PRESSO global and outlier analyses to identify and remove pleiotropic SNPs [26]. Furthermore, the influence of pleiotropic SNPs on MR analyses was evaluated by the MR-PRESSO distortion test. We adopted a conservative strategy in which we removed pleiotropic SNPs regardless of whether they had significant influences on estimation. A subset of SNPs after excluding pleiotropic SNPs was used for subsequent analyses. After MR analysis, significant signals were tested again to determine whether the results were biased by residual pleiotropic or heterogeneous effects using modified Cochran's Q statistic, MR-Egger intercept and funnel plot for corresponding MR methods.

### Power evaluation

The power to detect truly causal effects was calculated using an online tool mRnd (http://cnsgenomics.com/shiny/mRnd/). This method uses a non-centrality parameter to calculate the statistical power of a continuous outcome inferred with a two-sample MR approach [33]. In the current work, this parameter was used to estimate power based on the explainable variance, effect size and sample size. All the GWAS summary results were based on standardized phenotypes (i.e., with mean 0 and standard deviation 1). Therefore, the individual SNP effect size was estimated as the explainable variance with the formula $2f(1-f)\beta^2$, where $f$ is the allele frequency and $\beta$ is the regression coefficient. The sum of single SNP effect sizes was used as the explainable variance explained by all instrument variants to calculate power.

### Statistical analysis

MR must conform to three key assumptions to properly infer causality (Fig. 2) [34]: i) The genetic instruments (IVs) used for analysis should be associated with exposure (i.e., a risk factor); ii) the genetic instruments should not be associated with any confounders of exposure or outcome; iii) and the association of genetic instruments with the outcome should be only through exposure (i.e., no horizontal pleiotropic effect). We utilized the IVW method to identify potential causality of 25(OH)D and gut microbiota. The estimator of IVW was a meta-analysis of Wald ratios derived from a 2-step, least square (2SLS) approach using multiple SNPs. The combined exposure-outcome (*X-Y*) causality estimator was defined as

$$\beta_{IVW} = \frac{\sum_{j=1}^{m} \hat{\beta}_{Yj}\hat{\beta}_{Xj} se\left(\hat{\beta}_{Yj}\right)^{-2}}{\sum_{j=1}^{m} \hat{\beta}_{Xj}^2 se\left(\hat{\beta}_{Yj}\right)^{-2}} \quad var(\beta_{IVW}) = \frac{1}{\sum_{j=1}^{m} \hat{\beta}_{Xj}^2 se\left(\hat{\beta}_{Yj}\right)^{-2}},$$

where $m$ is the number of SNPs used for estimation; $\hat{\beta}_{Xj}$ is the coefficient of exposure-genotype (*X-G*) regression of the *i*-th SNP; similarly, $\hat{\beta}_{Yj}$ is the coefficient of outcome-genotype (*Y-G*) regression of the *i*-th SNP; and *se* is the corresponding standard error. MR Egger, maximum likelihood, MR-PRESSO, and weighted median methods can evaluate causal effects in more complex circumstances that may slightly violate three MR assumptions. These four methods were also performed, although these methods have less statistical power than does IVW. Briefly, we adopted all these methods to comprehensively study causality with IVW as the main result in the present study. The threshold of statistical significance was adjusted using the Benjamini and Hochberg FDR correction. All *P* values were two-sided. All analyses were mainly based on MR-base [35] in R version 3.6.0 (R Project for Statistical Computing) and Python 3.7. The intact pipelines were deposited onto the GitHub website (https://github.com/kylin734984/conductGMMR).

## RESULTS

In the present study, approximately 18,340 and 496,946 subjects of European descent (or mainly European) were analyzed for gut microbiota and 25(OH)D, respectively. The basic characteristics of the study population have been described elsewhere [10, 31]. We estimated the genetic correlation between microbial taxa and 25(OH)D. Of these trait pairs, half were assessable, but most were not significant (Supplementary Table 1), showing a few genetic overlaps between them. Only one significant association showed that 25(OH)D and *Ruminiclostridium9* were highly positively correlated ($r_g = 0.43$, $P = 0.04$) at nominal significance.

In total, 211 taxa were included to comprehensively evaluate causality. Since some biologically similar taxa used the same sequences to estimate relative abundance or presence, the SNP effect sizes derived from GWAS summary statistics were identical among these traits. As shown in Supplementary Table 2, pleiotropic SNPs varying from 1 to 5 were identified for gut microbiota and 25(OH)D. After instrument SNP selection, a number of independent and nonpleiotropic SNPs (2-189, median = 129) clumped by PLINK 1.9 were extracted for every microbial taxon. We evaluated the causality of microbial taxa to 25(OH)D (Supplementary Table 3) and identified three taxa (*Erysipelotrichia*, *Erysipelotrichaceae*, and *Erysipelotrichales*) that had identical significant effects on 25(OH)D using ten SNPs, as these three taxa

**Table 1.** Mendelian randomization estimate using inverse variance weighted model.

| Exposure | Outcome | NSNP | $R^2$ | Beta | s.e. | *P* | $P_{FDR}$ | Q | $P_{het}$ |
|---|---|---|---|---|---|---|---|---|---|
| Erysipelotrichia, Erysipelotrichaceae, Erysipelotrichales | 25(OH)D | 10 | 0.025 | −0.06 | 0.01 | $1.48 \times 10^{-6}$ | $1.93 \times 10^{-4}$ | 7.59 | 0.58 |
| 25(OH)D | *Phascolarctobacterium* | 215 | 0.04 | −0.16 | 0.04 | $2.48 \times 10^{-4}$ | 0.02 | 183.23 | 0.94 |

*NSNP* Number of SNPs used in MR estimation, $R^2$ proportion of outcome variance explained by SNPs we used as instruments, *P* *P* value of the MR test, $P_{FDR}$ corrected *P* value after Benjamini and Hochberg FDR correction, *Q* Cochran's Q statistic derived from heterogeneity test, $P_{het}$ *P* value of Cochran's Q test.

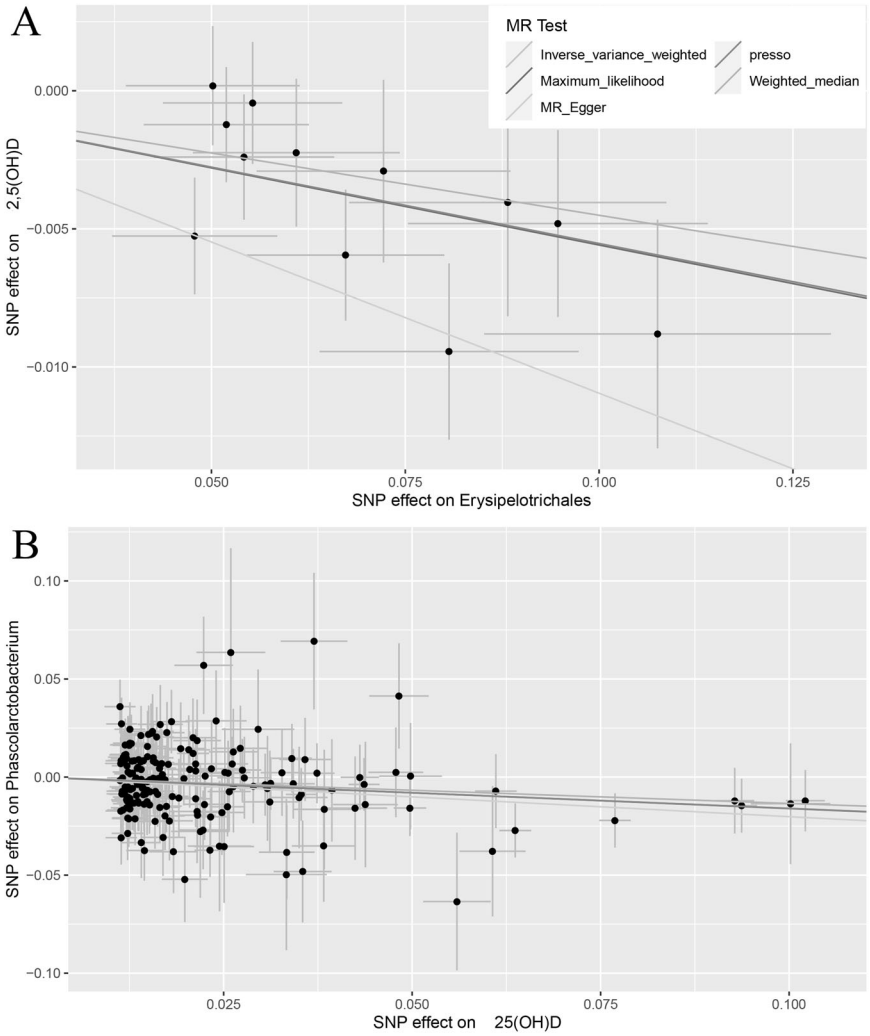


**Fig. 3 Mendelian randomization estimate using multiple single nucleotide polymorphisms.** Lines depict associations between exposure and outcome estimated by MR, where line colors refer to different methods. Vertical and horizontal black lines around points show 95% confidence interval for each polymorphism-exposure association and polymorphism-outcome association respectively. **A** represents the causal effect of Erysipelotrichales on 25(OH)D, and **B** represents the causal effect of 25(OH)D on Phascolarctobacterium.

were quantified by the same sequences categorized by the SILVA 128 database and RDP classifier (Table 1, Supplementary Table 4) [10]. Hence, we grouped the three into one category (Fig. 3A). One SNP was removed from MR analysis due to potential pleiotropy, although it had insignificant bias on MR estimation (Supplementary Table 2). *Erysipelotrichia*, *Erysipelotrichaceae* and *Erysipelotrichales* were found to have higher bacterial abundance in populations with lower 25(OH)D concentrations by IVW ($\beta_{IVW} = -0.06$, s.e. = 0.01, $P_{FDR} = 1.93 \times 10^{-4}$) (Table 1), in line with the results of the maximum likelihood method ($\beta = -0.06$, s.e. = 0.01, $P = 3.31 \times 10^{-6}$); the weighted median method ($\beta = -0.05$, s.e. = 0.02, $P = 8.44 \times 10^{-3}$); and the MR-PRESSO method ($\beta = -0.06$, s.e. = 0.01, $P = 5.35 \times 10^{-4}$). However, the association was not replicated in the MR-Egger method ($\beta = -0.11$, s.e. = 0.05, $P = 0.09$) (Table 2). With a type I error rate of 0.05, power assessment showed that we had 82% power to identify true causal effects ($\beta = -0.06$) in MR analysis with a large sample size (18,473) and multiple instrument SNPs ($R^2 = 0.03$). Modified Cochran's Q test and MR-Egger intercept, in conjunction with funnel plot, proved that there was no residual pleiotropy or heterogeneity (Supplementary Fig. 1).

For reverse MR, 25(OH)D incorporated more SNPs to perform MR. In total, 215 SNPs (Supplementary Table 5) were eligible and passed the pleiotropy test to conduct MR. 25(OH)D was found to

**Table 2.** Mendelian randomization estimate using other four methods.

| Exposure | Outcome | Method | NSNP | Beta | s.e. | *P* | Q | $P_{het}$ |
|---|---|---|---|---|---|---|---|---|
| Erysipelotrichia, Erysipelotrichaceae, Erysipelotrichales | 25(OH)D | Maximum likelihood | 10 | −0.06 | 0.01 | **$3.31 \times 10^{-6}$** | 6.88 | 0.65 |
| | | MR-Egger | 10 | −0.11 | 0.05 | 0.09 | 6.93 | 0.54 |
| | | Weighted median | 10 | −0.05 | 0.02 | **$8.44 \times 10^{-3}$** | – | – |
| | | MR-PRESSO | 10 | −0.06 | 0.01 | **$5.35 \times 10^{-4}$** | – | – |
| 25(OH)D | *Phascolarctobacterium* | Maximum likelihood | 215 | −0.16 | 0.04 | **$2.38 \times 10^{-4}$** | 183.17 | 0.94 |
| | | MR-Egger | 215 | −0.19 | 0.07 | **$8.77 \times 10^{-3}$** | 182.97 | 0.93 |
| | | Weighted median | 215 | −0.16 | 0.07 | **0.03** | – | – |
| | | MR-PRESSO | 215 | −0.16 | 0.04 | **$1.02 \times 10^{-4}$** | – | – |

– Cochran's Q test is not available for weighted median and MR-PRESSO methods. Statistically significant values ($P < 0.05$) are marked in bold.

have a negative effect on *Phascolarctobacterium* ($\beta_{IVW} = -0.16$, s.e. = 0.04, $P_{FDR} = 0.02$) based on IVW (Table 2, Fig. 3B), consistent with the results of maximum likelihood ($\beta = -0.16$, s.e. = 0.04, $P = 2.38 \times 10^{-4}$), the MR-Egger method ($\beta = -0.19$, s.e. = 0.07, $P = 8.77 \times 10^{-3}$); weighted median method ($\beta = -0.16$, s.e. = 0.07, $P = 0.03$) and MR-PRESSO ($\beta = -0.16$, s.e. = 0.04, $P = 1.02 \times 10^{-4}$). With the current sample size, the combination of 215 instrumental SNPs had excellent power (≥90%) to detect associations with *Phascolarctobacterium* at $\beta = -0.10$. Secondary pleiotropy and heterogeneity tests suggest that genetic pleiotropy did not drive the MR results. The results were robust with regard to the alterations of SNPs in the leave-one-out analysis (Supplementary Fig. 2).

## DISCUSSION

To the best of our knowledge, our present study represents the first to investigate SNV-based genetic correlations and potential MR associations between gut microbiota and 25(OH)D using large-scale GWAS summary data. In summary, our findings suggest that *Ruminiclostridium9* might be genetically associated with 25(OH)D. Furthermore, we found evidence for causal effects of specific microbiota on 25(OH)D as well as for reverse causality (i.e., *Erysipelotrichia*, *Erysipelotrichaceae* and *Erysipelotrichales* →□25(OH)D, and 25(OH)D → *Phascolarctobacterium*). These results were robust with regard to various MR methods and demonstrate the existence of interactions between these two traits.

Few studies highlight the impact of *Ruminiclostridium9* on vitamin homeostasis [36, 37]. Cross-trait LDSC regression showed a possible genetic correlation between *Ruminiclostridium9* and 25(OH)D. There was no causal relationship between them in MR analysis, although Zou et al. have suggested that the contribution of gut microbiota dysbiosis, including that of *Ruminiclostridium9*, to hypertension is partially mediated by influencing the metabolism of vitamin D3 [36]. Another animal study supported the link between *Ruminiclostridium9* and hypertension, but the role of vitamin D was not investigated [38]. Future studies need to explore whether these findings can translate into effective clinical intervention.

Clinical trials studying the benefits of vitamin D supplementation have shown the salutary effect of vitamin D on maintaining health [39], whereas absorption of exogenous vitamin D is largely determined by a healthy commensal gut microbiota environment [40]. Our results provide new evidence to support the hypothesis that *Erysipelotrichaceae* mediate the pathway of exogenous vitamin D absorption (exogenous vitamin D → *Erysipelotrichaceae* → serum vitamin D) [41]. A randomized controlled clinical trial suggested that *Veillonella* and *Erysipelotrichaceae* are significantly enriched in cystic fibrosis patients treated with placebo compared with those treated with vitamin D3 supplementation [42]. This means that vitamin D supplementation diminishes the abundance of *Veillonella* and *Erysipelotrichaceae*. Interestingly, significantly higher *Erysipelotrichaceae* and *Veillonellaceae* abundance was found in women with vitamin D deficiency than in the vitamin D nondeficient group [41]. Taken together, *Erysipelotrichaceae* is considered a vital mediator of vitamin D absorption. To what extent the taxon influences the pathway merits further investigation.

Logically, altered vitamin D could influence the composition and metabolism of gut microbiota in the body, such as *Bacteroidetes* and *Proteobacteria* phyla [43]. Our study newly identified a bacterial species, *Phascolarctobacterium*, that was negatively influenced by serum vitamin D status. *Phascolarctobacterium* abundantly colonizes the human gastrointestinal gut and increases with age [44]. Prior studies have reported the associations of *Phascolarctobacterium* with both vitamin D and major depressive disorder (MDD) [45, 46]. *Phascolarctobacterium* can produce short-chain fatty acids, including acetate and propionate, and can be associated with the metabolic state and mood of the host [44]. Specifically, *Phascolarctobacterium* was found to be abundant in MDD patients in multiple cross-sectional studies [46]. Intriguingly, Lin et al. found that propionate promotes vitamin D receptor expression in the intestine via activity in vitro and in vivo [47]. Moreover, serum vitamin D has been shown to be negatively associated with MDD in both cross-sectional studies and randomized clinical trials [45]. The causal inhibitory effect of vitamin D on *Phascolarctobacterium* established in the present work leads to a better understanding that vitamin D protects the population from MDD via the microbiota-gut-brain axis pathway.

There were several limitations in the present study. Only taxa above the genus level were able to be evaluated, some of which were detected by the same sequences, which were limited by the classifier and sequencing depth. In addition, microbiota diversity indices were not included for analysis since they varied in distinct cohorts.

In summary, our study explored the associations between gut microbiota and vitamin D absorption by two-sample MR and LDSC analyses. Our findings provide support of the interactions between specific gut microbiota and vitamin D. Based on our results, we suggest that modulating the composition of gut microbiota could be a new avenue of research in vitamin D homeostasis.

## DATA AVAILABILITY

Data used in this study are publicly available at https://cnsgenomics.com/data/revez_20/ and https://mibiogen.gcc.rug.nl/.

## REFERENCES

1. Mondot S, de Wouters T, Dore J, Lepage P. The human gut microbiome and its dysfunctions. Dig Dis. 2013;31:278–85.
2. Busnelli M, Manzini S, Chiesa G. The gut microbiota affects host pathophysiology as an endocrine organ: a focus on cardiovascular disease. Nutrients. 2019;12:79.

3. Tang WH, Kitai T, Hazen SL. Gut microbiota in cardiovascular health and disease. Circ Res. 2017;120:1183–96.
4. Ley RE, Backhed F, Turnbaugh P, Lozupone CA, Knight RD, Gordon JI. Obesity alters gut microbial ecology. Proc Natl Acad Sci USA. 2005;102:11070–5.
5. Kreznar JH, Keller MP, Traeger LL, Rabaglia ME, Schueler KL, Stapleton DS, et al. Host genotype and gut microbiome modulate insulin secretion and diet-induced metabolic phenotypes. Cell Rep. 2017;18:1739–50.
6. Larsen N, Vogensen FK, van den Berg FW, Nielsen DS, Andreasen AS, Pedersen BK, et al. Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. PLoS One. 2010;5:e9085.
7. Karlsson FH, Tremaroli V, Nookaew I, Bergstrom G, Behre CJ, Fagerberg B, et al. Gut metagenome in European women with normal, impaired and diabetic glucose control. Nature. 2013;498:99–103.
8. Kootte RS, Levin E, Salojarvi J, Smits LP, Hartstra AV, Udayappan SD, et al. Improvement of insulin sensitivity after lean donor feces in metabolic syndrome is driven by baseline intestinal microbiota composition. Cell Metab. 2017;26:611–9. e6
9. Zhang X, Shen D, Fang Z, Jie Z, Qiu X, Zhang C, et al. Human gut microbiota changes reveal the progression of glucose intolerance. PLoS One. 2013;8:e71108.
10. 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;53:156–65.
11. Riccio P, Rossano R. Diet, gut microbiota, and vitamins D + A in multiple sclerosis. Neurotherapeutics. 2018;15:75–91.
12. Zheng P, Zeng B, Zhou C, Liu M, Fang Z, Xu X, et al. Gut microbiome remodeling induces depressive-like behaviors through a pathway mediated by the host's metabolism. Mol Psychiatry. 2016;21:786–96.
13. Zhang Y, Kang C, Wang XL, Zhou M, Chen MT, Zhu XH, et al. Dietary factors modulate colonic tumorigenesis through the interaction of gut microbiota and host chloride channels. Mol Nutr Food Res. 2018;62:1700554.
14. Huang R, Xiang J, Zhou P. Vitamin D, gut microbiota, and radiation-related resistance: a love-hate triangle. J Exp Clin Cancer Res. 2019;38:493.
15. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, et al. Linking long-term dietary patterns with gut microbial enterotypes. Science. 2011;334:105–8.
16. Luthold RV, Fernandes GR, Franco-de-Moraes AC, Folchetti LG, Ferreira SR. Gut microbiota interactions with the immunomodulatory role of vitamin D in normal individuals. Metabolism. 2017;69:76–86.
17. Thomas RL, Jiang L, Adams JS, Xu ZZ, Shen J, Janssen S, et al. Vitamin D metabolites and the gut microbiome in older men. Nat Commun. 2020;11:5997.
18. Seura T, Yoshino Y, Fukuwatari T. The relationship between habitual dietary intake and gut microbiota in young Japanese women. J Nutr Sci Vitaminol. 2017;63:396–404.
19. Waterhouse M, Hope B, Krause L, Morrison M, Protani MM, Zakrzewski M, et al. Vitamin D and the gut microbiome: a systematic review of in vivo studies. Eur J Nutr. 2019;58:2895–910.
20. Wang J, Thingholm LB, Skieceviciene J, Rausch P, Kummen M, Hov JR, et al. Genome-wide association analysis identifies variation in vitamin D receptor and other host factors influencing the gut microbiota. Nat Genet. 2016;48:1396–406.
21. Jin D, Wu S, Zhang YG, Lu R, Xia Y, Dong H, et al. Lack of vitamin D receptor causes dysbiosis and changes the functions of the murine intestinal microbiome. Clin Ther. 2015;37:996–1009. e7
22. Naderpoor N, Mousa A, Fernanda Gomez Arango L, Barrett HL, Dekker Nitert M, de Courten B. Effect of vitamin D supplementation on faecal microbiota: a randomised clinical trial. Nutrients. 2019;11:2888.
23. Jones ML, Martoni CJ, Prakash S. Oral supplementation with probiotic L. reuteri NCIMB 30242 increases mean circulating 25-hydroxyvitamin D: a post hoc analysis of a randomized controlled trial. J Clin Endocrinol Metab. 2013;98:2944–51.
24. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet. 2015;47:1236–41.
25. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol. 2003;32:1–22.
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. Dastani Z, Hivert MF, Timpson N, Perry JR, Yuan X, Scott RA, et al. Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multi-ethnic meta-analysis of 45,891 individuals. PLoS Genet. 2012;8:e1002607.
28. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. Eur J Epidemiol. 2017;32:377–89.
29. Pierce BL, Burgess S. Efficient design for Mendelian randomization studies: subsample and 2-sample instrumental variable estimators. Am J Epidemiol. 2013;178:1177–84.
30. 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.
31. Revez JA, Lin T, Qiao Z, Xue A, Holtz Y, Zhu Z, et al. Genome-wide association study identifies 143 loci associated with 25 hydroxyvitamin D concentration. Nat Commun. 2020;11:1647.
32. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet. 2016;48:1279–83.
33. Brion MJ, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. Int J Epidemiol. 2013;42:1497–501.
34. 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.
35. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife. 2018;7:e34408.
36. Zuo K, Li J, Xu Q, Hu C, Gao Y, Chen M, et al. Dysbiotic gut microbes may contribute to hypertension by limiting vitamin D production. Clin Cardiol. 2019;42:710–9.
37. Xiao L, Chen B, Feng D, Yang T, Li T, Chen J. TLR4 may be involved in the regulation of colonic mucosal microbiota by vitamin A. Front Microbiol. 2019;10:268.
38. Gutierrez-Calabres E, Ortega-Hernandez A, Modrego J, Gomez-Gordo R, Caro-Vadillo A, Rodriguez-Bobada C, et al. Gut microbiota profile identifies transition from compensated cardiac hypertrophy to heart failure in hypertensive rats. Hypertension. 2020;76:1545–54.
39. Yang Q, Liang Q, Balakrishnan B, Belobrajdic DP, Feng QJ, Zhang W. Role of dietary nutrients in the modulation of gut microbiota: a narrative review. Nutrients. 2020;12:381.
40. Sun J. Dietary vitamin D, vitamin D receptor, and microbiome. Curr Opin Clin Nutr Metab Care. 2018;21:471–4.
41. Palacios-Gonzalez B, Ramirez-Salazar EG, Rivera-Paredez B, Quiterio M, Flores YN, Macias-Kauffer L, et al. A multi-omic analysis for low bone mineral density in postmenopausal women suggests a RELATIONSHIP between diet, metabolites, and microbiota. Microorganisms. 2020;8:1630.
42. Kanhere M, He J, Chassaing B, Ziegler TR, Alvarez JA, Ivie EA, et al. Bolus weekly vitamin D3 supplementation impacts gut and airway microbiota in adults with cystic fibrosis: a double-blind, randomized, placebo-controlled clinical trial. J Clin Endocrinol Metab. 2018;103:564–74.
43. Wu S, Zhang YG, Lu R, Xia Y, Zhou D, Petrof EO, et al. Intestinal epithelial vitamin D receptor deletion leads to defective autophagy in colitis. Gut. 2015;64:1082–94.
44. Wu F, Guo X, Zhang J, Zhang M, Ou Z, Peng Y. Phascolarctobacterium faecium abundant colonization in human gastrointestinal tract. Exp Ther Med. 2017;14:3122–6.
45. Tsiglopoulos J, Pearson N, Mifsud N, Allott K, O'Donoghue B. The association between vitamin D and symptom domains in psychotic disorders: a systematic review. Schizophr Res. 2021;237:79–92.
46. Iglesias-Vazquez L, Van Ginkel Riba G, Arija V, Canals J. Composition of gut microbiota in children with autism spectrum disorder: a systematic review and meta-analysis. Nutrients. 2020;12:792.
47. Lin H, Huang Y, Tian T, Wang P, Li Y. Propionate promotes vitamin D receptor expression via yes-associated protein in rats with short bowel syndrome. Biochem Biophys Res Commun. 2020;523:645–50.

## ACKNOWLEDGEMENTS

We hereby express our thanks to consortia (UKB, SUNLIGHT, MiBioGen) for kindly releasing GWAS summary data. The study was funded by National Natural Science Foundation of China (81801970 to GTL, 81200596 to YW), Research Foundation from Jiangsu Commission of Health (Z2017018 to YW), and National Natural Science Foundation of Yangzhou (No. YZ2018091 to XNL).

## AUTHOR CONTRIBUTIONS

PJJ, WY, and YXL conceived the research questions and designed the analysis; YXL and ZQT analyzed data; ZL, PYF, LXN and LGT guided and provided feedback on the analysis and interpretation of results; YXL and XXJ wrote the paper. All authors critically reviewed and approved the final manuscript.

## COMPETING INTERESTS

The authors declare no competing interests.

## ETHICAL APPROVAL

This study was exempt from ethical approval because all data were anonymous and were based on results from previous publications.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41430-021-01065-3.

**Correspondence** and requests for materials should be addressed to Jiajia Pan or Ying Wang.