



OPEN ACCESS

EDITED BY

Matthew Cook,
University of Cambridge, United Kingdom

REVIEWED BY

Jiang-Shan Tan,
Chinese Academy of Medical Sciences and
Peking Union Medical College, China
Ana Maria Calderon De La Barca,
National Council of Science and
Technology (CONACYT), Mexico
Dimitri Poddighe,
Nazarbayev University, Kazakhstan

*CORRESPONDENCE

Iraia García-Santisteban
✉ irai.garcia@ehu.eus
Jose Ramon Bilbao
✉ joseramon.bilbao@ehu.eus

†These authors have contributed
equally to this work and share
senior authorship

RECEIVED 28 October 2022

ACCEPTED 07 June 2023

PUBLISHED 30 June 2023

CITATION

González-García BP, Marí S,
Cilleros-Portet A, Hernangomez-Laderas A,
Fernandez-Jimenez N,
García-Santisteban I and Bilbao JR (2023)
Two-Sample Mendelian Randomization
detects bidirectional causality between gut
microbiota and celiac disease in individuals
with high genetic risk.
Front. Immunol. 14:1082862.
doi: 10.3389/fimmu.2023.1082862

COPYRIGHT

© 2023 González-García, Marí,
Cilleros-Portet, Hernangomez-Laderas,
Fernandez-Jimenez, García-Santisteban and
Bilbao. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Two-Sample Mendelian Randomization detects bidirectional causality between gut microbiota and celiac disease in individuals with high genetic risk

Bárbara P. González-García¹, Sergi Marí¹,
Ariadna Cilleros-Portet¹, Alba Hernangomez-Laderas¹,
Nora Fernandez-Jimenez¹, Iraia García-Santisteban^{1*†}
and Jose Ramon Bilbao^{1,2*†}

¹Department of Genetics, Physical Anthropology and Animal Physiology, University of the Basque Country (UPV/EHU) and Biocruces Bizkaia Health Research Institute, Leioa, Spain, ²Spanish Biomedical Research Center in Diabetes and Associated Metabolic Disorders (CIBERDEM), Madrid, Spain

Background: Celiac Disease (CeD) is an autoimmune disorder triggered by gluten intake in genetically susceptible individuals. Highest risk individuals are homozygous for the Human Leucocyte Antigen (HLA) DQ2.5 haplotype or DQ2.5/DQ2.2 heterozygous. Both the HLA-DQ2-positive high genetic risk individuals and those that have developed the disease have altered intestinal microbiota, but it remains unclear whether these alterations are a cause or a consequence of CeD.

Objective: To investigate a potential bidirectional causality between gut microbiota (GM) and CeD in HLA-DQ2 high genetic risk individuals.

Materials and Methods: We performed a bidirectional Two-Sample Mendelian Randomization (2SMR) test using summary statistics from the largest publicly available Genome-Wide Association Study (GWAS) of GM and the summary statistics of the ImmunoChip CeD study of those individuals with the HLA-DQ2 high-risk haplotype. To test whether changes in GM composition were causally linked to CeD, GM data were used as exposure and CeD data as outcome; to test for reverse causation, the exposure and outcome datasets were inverted.

Results: We identified several bacteria from *Ruminococcaceae* and *Lachnospiraceae* families of the Firmicutes phylum as potentially causal in both directions. In addition, our results suggest that changes in the abundance of *Veillonellaceae* family might be causal in the development of CeD, while alterations in *Pasteurellaceae* family might be a consequence of the disease itself.

Conclusion: Our results suggest that the relationship between GM and HLA-DQ2 high risk individuals is highly complex and bidirectional.

KEYWORDS

Mendelian randomization, celiac disease, HLA-DQ2, gut microbiota, bidirectional causality

1 Introduction

Celiac Disease (CeD) is an autoimmune disorder caused by gluten intake that develops in genetically susceptible individuals (1). The genetic susceptibility to CeD arises principally from Human Leukocyte Antigen (HLA) genes (2). More than 90% of celiac patients present the HLA-DQ2 heterodimer, encoded in *cis* by the HLA-DQA1*05:01 and DQB1*02:01 alleles that form the DQ2.5 haplotype (3). Among them, those with two copies of the HLA-DQ2 heterodimer (homozygous for the HLA-DQ2.5 haplotype or heterozygous for haplotypes HLA-DQ2.5 and HLA-DQ2.2, formed by alleles HLA-DQA1*02:01 and DQB1*02:02) are at the highest genetic risk of developing the disease (4). However, not all individuals carrying the HLA-DQ risk alleles finally develop CeD, suggesting that additional factors are needed to trigger disease onset.

In recent years, several studies have suggested that an additional factor involved in the pathogenesis of CeD could be the gut microbiota (5). The gut microbiota (GM) is defined as the community of microorganisms that live in the human gastrointestinal tract. The core GM is shaped early in life and is strongly influenced by multiple environmental factors such as mode of delivery, type of milk feeding (6), diet, antibiotic consumption (7), and also, by host genetics (8). At approximately three years of age, a relative stability in GM composition and diversity is reached, with phyla Firmicutes and Bacteroidetes representing 90% of all bacteria (9); these, together with other less abundant phyla such as Proteobacteria, Actinobacteria and Verrucomicrobia shape the healthy GM (10).

Cross-sectional studies in children and adults have shown that celiac patients have an altered composition of intestinal microbiota compared to non-celiac individuals (11–13). However, it must be mentioned that a clear microbiome signature has not been evidenced yet, in part due to the heterogeneity of the experimental procedures and design of these kind of studies (14). Moreover, the control subjects enrolled in these kinds of observational studies might suffer from other pathologies that could cause microbiota alterations. Another important limitation is that the majority of celiac patients are on a gluten-free diet (GFD) at the time of the study, and gluten is a crucial factor that *per se* influences GM composition, and thus constitutes an inevitable confounder (15–17). Prospective studies that have followed up cohorts of infants at risk of developing CeD have reported that the HLA-DQ genotype itself influences intestinal microbiota composition (18–22). While these studies provide valuable information, they are often restricted to

very few individuals, since not all the HLA-DQ risk subjects finally develop CeD. Finally, it is unclear whether the alterations in the microbial community observed in celiac patients are a cause or a consequence of the disease (23).

Mendelian Randomization (MR) is an alternative approach that overcomes the limitations of observational studies, such as confounding factors or reverse causation. MR uses Single Nucleotide Polymorphisms (SNPs), that are randomly and independently allocated to each individual at conception, as instrumental variables to assess a potential causal relationship between an exposure and an outcome (24). Following this approach, our group identified a number of genetic variants associated with bacterial composition and function that might be driving CeD pathogenesis (25). More recently, using larger GM genomic datasets, other groups have performed bidirectional MR analyses to infer causal effects of gut microbial abundance on CeD risk, and *vice versa* (8, 26–28). However, no previous study has evaluated the potential bidirectional causality between GM composition and CeD in the context of the high-risk HLA (HR-HLA) haplotypes.

In the present work, we aim to explore whether the GM composition contributes to the development of CeD in HLA-DQ2 high genetic risk individuals, or if the disease itself is what modifies the composition of the GM in those genetically predisposed individuals (29).

2 Materials and methods

2.1 GWAS data sources

Two GWAS datasets were used for the Mendelian Randomization analysis. On the one hand, summary statistics from the MiBioGen microbiome consortium GWAS meta-analysis (8) were downloaded from www.mibiogen.org, and directly used for our analysis. This study represents the largest microbiome GWAS up to date including 18,340 individuals from 24 cohorts from different ancestries: 72.3% European, 6.0% American Hispanic/Latin, 4.4% East Asian, 2.6% from Middle Eastern countries, 0.7% African American and 14.0% from multiple ancestries. This large study analyzes the association of the host genotype with a total of 211 microbial taxa (9 phyla, 16 classes, 20 orders, 35 families, and 131 genera). On the other hand, individual-level genotype data from the Immunochip CeD study (29) were reanalyzed selecting only HR-HLA individuals, as described in the

following section. The Immunochip is a case-control study that aimed to identify non-HLA risk *loci* associated with CeD by genotyping 12,041 CeD cases and 12,228 healthy controls (97.5% of the individuals were of European ancestry, and 2.5% from India).

2.2 Quality control, genotype imputation and genome-wide association analysis

The genotypes from the Immunochip CeD dataset were subjected to SNP and sample quality controls. SNPs with a Minor Allele Frequency (MAF) below 1%, a call rate below 95%, or a Hardy-Weinberg equilibrium p -value under 1×10^{-6} were discarded. In addition, samples with a call rate below 97%, sex discrepancies and those with heterozygosity levels higher than four times the standard deviation of the mean were also excluded.

In order to select HR-HLA individuals, DQ haplotypes were imputed with the *HIBAG-HLA Genotype Imputation* R package (Version 1.32.0) (30) and 4,814 individuals (4,264 celiac patients and 550 controls), either homozygous for HLA-DQ2.5 (HLA-DQA1*05_01-DQB1*02:01) or heterozygous for HLA-DQ2.5/DQ2.2 (HLA-DQA1*05:01-DQB1*02:01/DQA1*02:01-DQB1*02:02) were selected. Genome-wide imputation of HR-HLA individuals was carried out in the Michigan Imputation Server (31) using the TOPMed Imputation Reference panel (32) with the GRCh38 genomic coordinates. Variants with an R^2 imputation accuracy value lower than 0.9 were filtered out.

The association analysis was performed using SNPTTEST (Version 2.5.6) (33), with score test and an additive model. A Manhattan plot was built using the *qqman* R package (Version 0.1.8) (34).

2.3 Bidirectional two-sample mendelian randomization

The reciprocal causality between GM and HR-HLA CeD was assessed following a bidirectional 2SMR approach (Figure 1) using the *TwoSampleMR* R package (Version 4.26) (35). To test if changes in GM composition were causally linked to CeD in HR-HLA individuals, public data from MiBioGen Consortium ($n=18,340$) were used as exposure data, and the association study results ($n=4,814$) from the HR-HLA Immunochip study were used as outcome data. To test for reverse causation, the exposure and outcome datasets were inverted.

Exposure data were prepared using the *format_data()* function. For Instrumental variables (IVs) associated with bacterial taxonomies (phylum, class, order, family, or genus) a nominal threshold p -value of 1×10^{-5} was chosen (as this cut-off value was previously used in analogous 2SMR studies by the MiBioGen consortium to extract sufficient candidate instruments) (8). In the case of the reanalyzed Immunochip dataset, the same nominal threshold p -value was used (Supplementary Table 1) given the reduction in the sample size from 24,000 individuals to less than 5,000. Independent IVs were selected using the strict default clumping parameters ($r^2 < 0.001$, window $> 10,000$ kb) of the

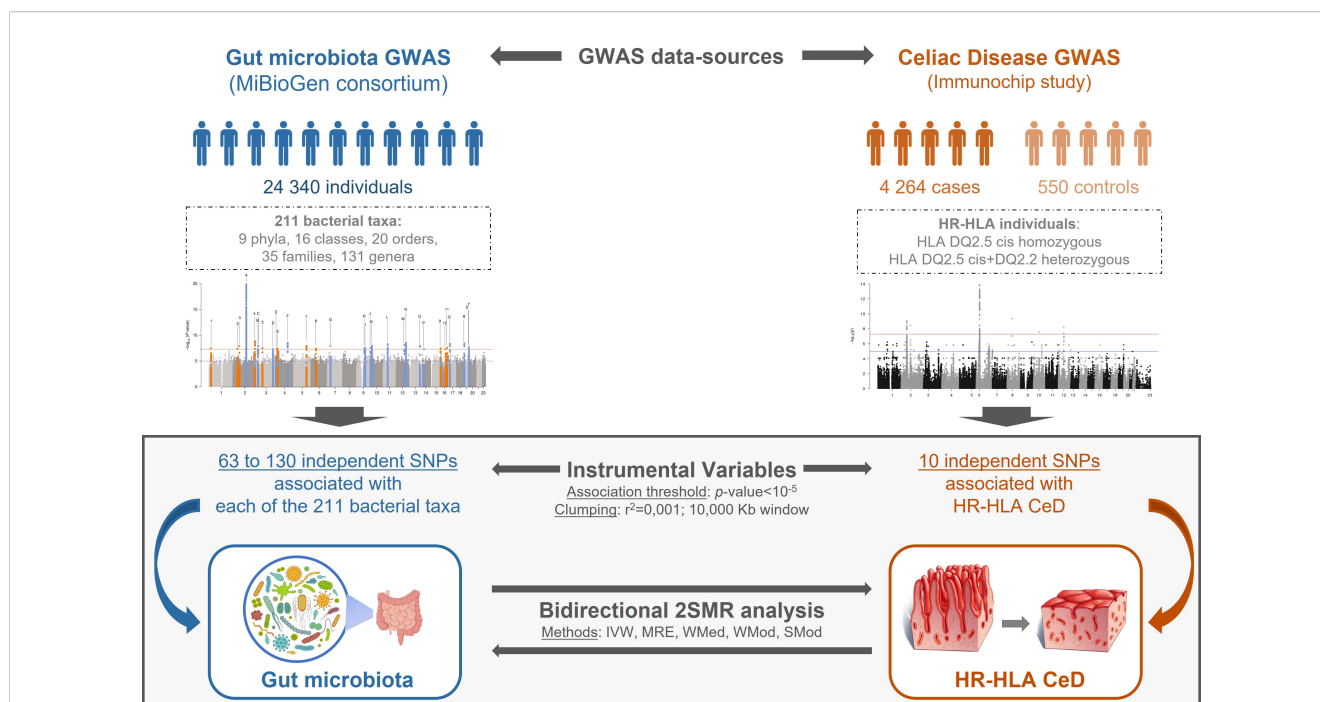


FIGURE 1

Schematic representation of the bidirectional 2SMR study. GM and CeD datasets were obtained from MiBioGen consortium and Immunochip study, respectively. GM dataset included genotype association of 211 taxa in 18,340 individuals; the CeD Immunochip was reanalyzed to only include genetic associations of HR-HLA individuals. Instrumental variables from exposure data were selected applying a p -value $< 10^{-5}$ association threshold, a pairwise LD $r^2 < 0.001$ and a 10,000 kb clumping window. The bidirectional 2SMR analysis: GM $>$ HR-HLA CeD or HR-HLA CeD $>$ GM was performed using five methods: Inverse Variance Weighted (IVW), MR Egger (MRE), Weighted Median (WMed), Weighted Mode (WMod), and Simple Mode (SMod).

clump_data() function. An *in-house* script was developed to extract the IVs from the GM and the HR-HLA CeD datasets to be used as outcome data (Supplementary Data 1).

Both exposure and outcome data were harmonized using the *harmonise_data()* function to guarantee that the effect that a given SNP has over the exposure and the outcome correspond to the same allele. After harmonization, the 2SMR analysis was completed using the different methods (Inverse Variance Weighted (IVW), MR Egger (MRE), Weighted Median (WMed), Weighted Mode (WMod), and Simple Mode (SMod)) available for the *mr()* function. Scatter plots were generated using the *mr_scatter_plot()* function, and forest plots with the *mr_forest_plot()* function.

3 Results

3.1 CeD susceptibility variants identified in individuals with high-risk HLA

In this study, we aimed to identify the reciprocal causality between the GM and CeD in HR-HLA individuals using a bidirectional 2SMR approach. Before performing the bidirectional 2SMR analysis, the Immunochip CeD genomic study (29) was reanalyzed considering only HR-HLA individuals (4) (Supplementary Table S1; Supplementary Figure S1). As shown in the Manhattan plot in Supplementary Figure S1, we obtained three association signals at genome-wide significance (p -value $< 5 \times 10^{-8}$): the top signal maps to the

MHC region on chromosome 6 (rs3128927, p -value = 1.35×10^{-14}), the second signal is on gene *PUS10* on chromosome 2 (rs10191951, p -value = 8.41×10^{-10}), and a novel association signal was located near *ATP23* gene on chromosome 12 (rs73344397, p -value = 5.77×10^{-9}). Along with these three regions, two additional signals corresponding to unique SNPs in those *loci* can be observed on chromosome 8 (rs138705229, p -value = 4.25×10^{-8}), and on chromosome 10 (rs375414915, p -value = 2.53×10^{-8}).

3.2 Bidirectional 2SMR identifies bacterial taxonomies causally related to high-risk HLA CeD

In order to explore whether the GM composition contributes to the development of CeD in HR-HLA individuals, or if on the contrary the disease modifies the composition of the GM in those individuals, we performed a 2SMR analysis in both directions: GM to HR-HLA CeD and HR-HLA CeD to GM (Table 1). As a first approach, the IVW method was used as an estimator of the potential causal effect that the exposure had on the outcome. In the GM to HR-HLA CeD direction (Table 1, upper panel), four microbial taxa from the phylum Firmicutes were identified, whereas in the HR-HLA CeD to GM direction (Table 1, lower panel), six microbial taxa from phyla Firmicutes, Bacteroidetes, and Proteobacteria resulted significant (IVW p -value < 0.05).

We next analyzed which of the bacterial taxa that resulted significant were common to both or specific to one direction. Some bacterial

TABLE 1 Summary of the potential causal associations between GM and HR HLA CeD.

GM to HR-HLA CeD								
TAXONOMIC LEVEL					2SMR RESULTS			
Phylum	Class	Order	Family	Genus	IVs	Beta	SE	p -value
Firmicutes	Clostridia	Clostridiales	<i>Ruminococcaceae</i>	<i>Ruminococcaceae</i> UCG010	3	1,690	0,828	0,041
Firmicutes	Clostridia	Clostridiales	<i>Ruminococcaceae</i>	<i>Ruminococcaceae</i> UCG011	2	-1,190	0,509	0,019
Firmicutes	Clostridia	Clostridiales	<i>Lachnospiraceae</i>	<i>Lachnospiraceae</i> UCG008	3	1,462	0,586	0,012
Firmicutes	Negativicutes	Veillonellales	<i>Veillonellaceae</i>	-	7	1,112	0,454	0,014
-	-	-	-	<i>FamilyXIII</i> UCG001	4	1,216	0,580	0,036
HR-HLA CeD to GM								
TAXONOMIC LEVEL					2SMR RESULTS			
Phylum	Class	Order	Family	Genus	IVs	Beta	SE	p -value
Firmicutes	Clostridia	Clostridiales	<i>Ruminococcaceae</i>	<i>Anaerotruncus</i>	3	-0,053	0,018	0,003
Firmicutes	Clostridia	Clostridiales	<i>Lachnospiraceae</i>	<i>Tyzzrella3</i>	3	-0,084	0,032	0,008
Bacteroidetes	Bacteroidia	Bacteroidales	<i>Rikenellaceae</i>	<i>Rikenellaceae</i> RC9 gutgroup	3	0,085	0,040	0,036
Proteobacteria	Gammaproteobacteria	-	-	-	3	0,047	0,018	0,011
Proteobacteria	Gammaproteobacteria	Pasteurellales	-	-	3	0,049	0,024	0,041
Proteobacteria	Gammaproteobacteria	Pasteurellales	<i>Pasteurellaceae</i>	-	3	0,049	0,024	0,041
-	-	-	-	<i>Unknowngenus</i>	3	-0,057	0,023	0,014

Significant threshold was set at p -value < 0.05 for the Inverse Variance Weighted method (IVW). IVs, Instrumental Variables; SE, standard error. In black are depicted the direct results from our bidirectional 2SMR analysis, while in grey are indicated the upper taxonomic level of each hit.

taxonomies observed, specifically the *Ruminococcaceae* and *Lachnospiraceae* families from the phylum Firmicutes (class Clostridia, order Clostridiales), are common to both directions, suggesting that they might play a role both as a cause and as a consequence of the disease. Other taxa only resulted significant in one of the two directions analyzed. In the GM to HR-HLA CeD direction, the class Negativicutes (p -value=0.014, $\beta=1.112 \pm 0.454$) resulted significant, suggesting a causal role of this taxon in the pathogenesis of HR-HLA CeD. Regarding the HR-HLA CeD to GM direction, classes Bacteroidia (p -value=0.036, $\beta=0.085 \pm 0.040$) or Gammaproteobacteria (p -value=0.041, $\beta=0.049 \pm 0.024$) gave a significant result, suggesting that their alteration could be a consequence rather than a cause of the disease itself. Of note, three of the seven hits in this direction are phylogenetically related to each other (class Gammaproteobacteria, order Pasteurellales, family *Pasteurellaceae*), highlighting their relevance in this direction. *FamilyXIII UCG001* and *unknowngenus* were excluded from subsequent analyses as there is no further information available.

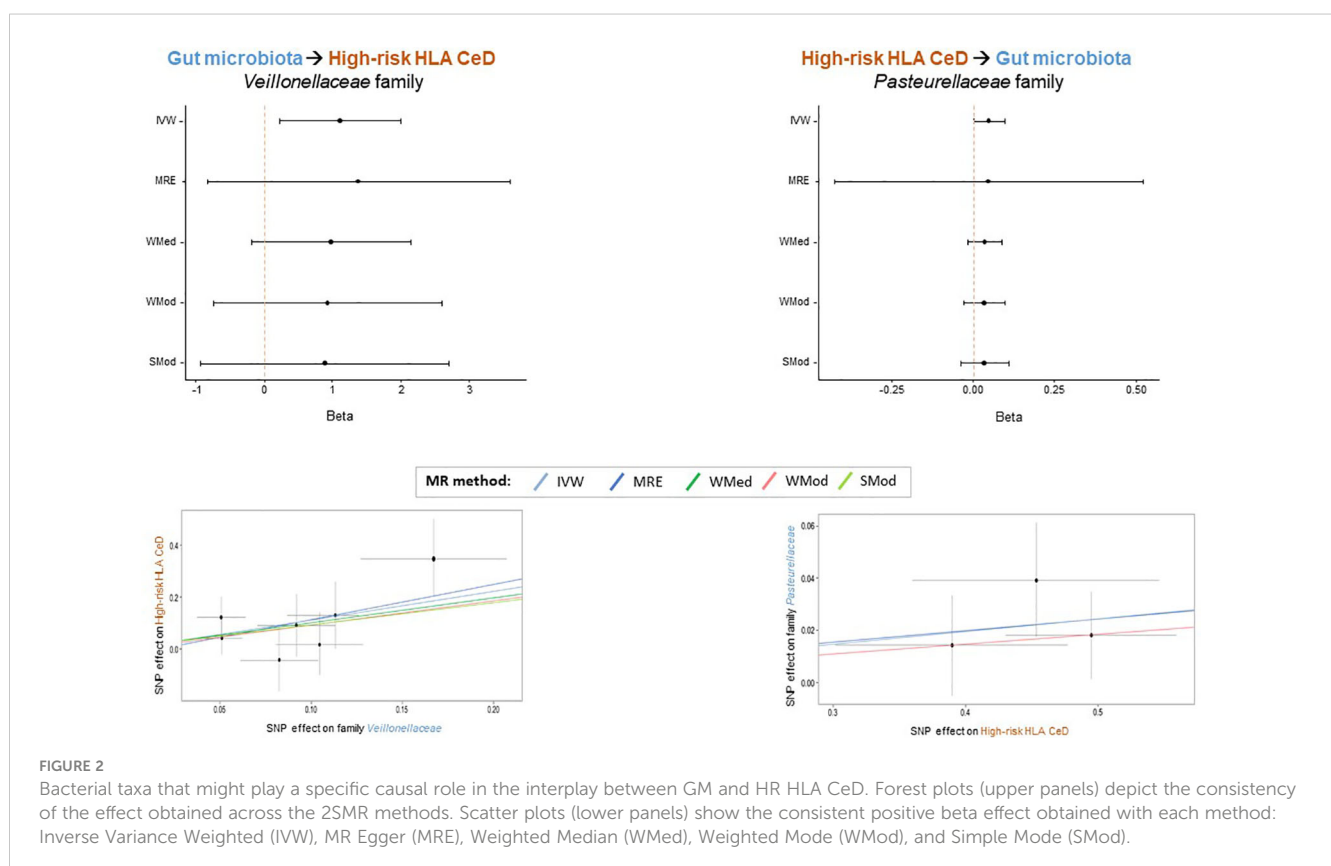
3.3 *Veillonellaceae* and *Pasteurellaceae* families might play specific causal roles in the interplay between the GM and high-risk HLA CeD

In order to prioritize the most relevant hits, in addition to the IVW method, four additional methods from the 2SMR package were applied (MRE, WMed, WMod and SMod). Even if the p -values were not significant for all methods (Supplementary Table

S2), forest plots illustrate a consistent direction in the effect for families *Veillonellaceae* (IVW $\beta=1.112$; MRE $\beta=1.362$; WMed $\beta=0.988$; WMod $\beta=0.930$; SMod $\beta=0.891$) and *Pasteurellaceae* (IVW $\beta=0.049$; MRE $\beta=0.045$; WMed $\beta=0.037$; WMod $\beta=0.037$; SMod $\beta=0.037$) in the GM to HR-HLA CeD and HR-HLA CeD to GM directions, respectively (Figure 2, upper panels). The absolute beta values were also very similar within each family, as evidenced in the slopes of the lines depicted in the scatter plots (Figure 2, lower panels). The consistency in the effect trends supports the causal role that each bacterial taxon might have in each direction, as will be discussed in the next section.

4 Discussion

An association between GM and CeD has been suggested in observational studies (11–13, 18–22). In this work, we evaluate for the first time the potential causality and directionality of this association in the context of the HR-HLA haplotypes by employing a bidirectional 2SMR approach. A previous study in our group also evaluated the interplay between host genetics, GM and CeD using a single-SNP 2SMR approach (25). This new analysis differs from the previous one in several aspects. Our current analysis follows a global MR approach that, using several SNPs as genetic instruments, allows the identification of causal associations between gut microbial taxa and HR-HLA CeD. In addition, this time we not only assess the potential causality that



microbial alterations might have over the development of CeD, but also look at the causal consequences that a celiac gut might exert over the GM. Other more recent reports also employ a bidirectional 2SMR approach to evaluate the reciprocal causality between gut microbial genera and several diseases, including type 1 diabetes and CeD (8, 26–28). However, the current study is specifically focused on individuals at high genetic risk of developing CeD.

Our analysis identified bacteria from *Ruminococcaceae* and *Lachnospiraceae* families of phylum Firmicutes as potentially causal in both directions. Firmicutes is one of the most frequently altered phyla not only in CeD patients (15), but also in those individuals at high genetic risk of developing the disease (21, 22). Specifically, family *Ruminococcaceae* was identified in a recent genomic study as potentially causal in the GM to CeD direction (26). Our results are in line with previous literature and go further by suggesting that GM alteration can also be a consequence of the disease. In line with this observation, a study characterizing the gut microbiota of children with CeD found alterations in *Ruminococcus* fiber-fermenters between treated and untreated individuals, suggesting that their modulation is a consequence of the treatment with a GFD (17). With regard to the *Lachnospiraceae*, a family that belongs to the core GM, the impact in host physiology is controversial. Some studies have identified associations between different taxa of *Lachnospiraceae* family with several intra- and extra intestinal diseases (36), including CeD. In this context, a recent study has reported that CeD patients in GFD had a decreased level of three taxa belonging to *Lachnospiraceae* family (37). A randomized controlled trial resulted in both decreases and increases in the abundances of different species of the *Lachnospiraceae* family during a low-gluten diet intervention (38). Thus, even if family *Lachnospiraceae* appears as potentially causal in both directions, it is more likely that the directionality is determined at a lower taxonomic level such as genus or species. Another possible interpretation of the potential dual role that *Ruminococcaceae* and *Lachnospiraceae* families have might come from a context-specific interaction, where depending on environmental factors, these bacteria might trigger the development of the disease, or be the consequence of a celiac gut microenvironment.

According to our data, other bacterial taxa might play a specific causal role in each of the directions analyzed. This is the case for the *Veillonellaceae* and *Pasteurellaceae* families, which yielded significant results with the IVW method in the GM to HR-HLA CeD and HR-HLA CeD to GM directions, respectively. Moreover, these significant findings were consistent across sensitivity analyses, and all the methods showed concordant effects, further supporting their specific causal role in each direction. On the one hand, our 2SMR analysis suggested that an increased abundance of family *Veillonellaceae* might have a causal effect on HR-HLA CeD. This potential causality might be extended to other gut diseases such as Crohn's disease, where an increased abundance of these type of pro-inflammatory bacteria has been consistently observed (39, 40). In addition, a decrease in *Veillonellaceae* abundance has been reported in healthy individuals following a GFD (16). In addition to the protective effect that gluten deprivation has over CeD, our results

could suggest that a GFD might slightly contribute through the reduction of *Veillonellaceae* bacteria. However, we must take into account that the reduction in *Veillonellaceae* abundance was observed in healthy volunteers, and not in individuals at high genetic risk of developing CeD. On the other hand, a higher genetic risk for CeD was predicted to increase the abundance in several taxa belonging to phylum Proteobacteria, specifically class Gammaproteobacteria, order Pasteurellales and family *Pasteurellaceae*. Interestingly, the phylum Proteobacteria has been reported to be more abundant not only in active celiac individuals compared to GFD-treated celiac patients and controls (41), but also in CeD individuals experiencing persistent symptoms, despite showing a normal histology and adhering to a GFD (15). It is conceivable to think that the higher abundance of Proteobacteria observed in these individuals could be a consequence rather than a cause of the disease itself.

Nevertheless, some limitations of our study should be considered. The first one is related to the small number of HLA-DQ2 non-celiac controls (n=550) compared to the celiac group (n=4,264) in the ImmunoChip study (29). However, this is an intrinsic limitation of the study design, as we intentionally selected those people with the highest genetic risk of developing CeD, which considerably increases the proportion of celiac individuals and restrains that of healthy controls. The second limitation relates to the ancestry of the participants in the ImmunoChip study. In our analysis, we used data of individuals mainly from European ancestry, where the HLA-DQ2 is the predominant risk haplotype. However, studies in Latin America indicate that HLA-DQ8 could confer a higher risk to CeD, and this might also modify the microbiota composition. It would be interesting to carry out a similar study using GWAS datasets from other ancestries to ascertain whether our findings can be generalized to other ethnic groups. The third limitation is related to the high inter-individual microbial variability observed across the 24 cohorts from the MiBioGen study (8), which reduces the statistical power of these kind of multi-cohort meta-analyses. Very recently, Qin et al. published a new microbiome GWAS in a single cohort of almost 6,000 European individuals, identifying genetic variants associated with certain microbes (29). The top genetic *loci* identified to be associated with bacterial taxa remain the same as in the MiBioGen study, suggesting that the strategy of using a single homogeneous cohort lies behind the constraint on the sample size. Fourth, the taxonomic resolution of the MiBioGen database reaches the genus level, leaving out more accurate genetic associations at the species or strain levels, and reports some uncharacterized hits. Thus, it would be interesting to repeat this 2SMR analysis when more specialized microbiota GWAS become available. To finalize, although in this study we have investigated the bidirectional causality between GM and CeD in high-risk individuals, microbiota from other niches such as the breast milk have been suggested to protect from autoimmune diseases, including CeD (6). Thus, it could be interesting to investigate the potential protective role of breastfeeding in a genetic setting predisposing to CeD. In summary, this is the first study that

evaluates the bidirectional causality between GM composition and CeD in the context of the HR-HLA haplotype using a 2SMR approach. We not only identify a set of microbial taxa that might be involved in the pathogenesis of CeD, but also identify other bacteria whose altered abundance might be a consequence of the disease itself. Our results, although preliminary and pending validation, could give clues to highlight potential focus on future research in the field of GM and CeD. Moreover, our computational approach could be applied to study other autoimmune diseases such as type 1 diabetes.

Data availability statement

The original contributions presented in the study are included in the article/[Supplementary Materials](#), further inquiries can be directed to the corresponding author/s.

Ethics statement

The data was based on published studies or databases and no additional ethical approval from an institutional review board was required.

Author contributions

JR-B designed the research. BG-G and SM collected the data. SM wrote the initial scripts, and BG-G performed the global analyses. AC-P and AH-L helped in the analyses. BG-G and IG-S performed the literature search. BG-G and IG-S drafted the article. NF-J revised the first draft and gave feedback. IG-S and JR-B supervised the study. All authors were involved in writing the paper. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Basque Dpt. of Health Projects GVSAN2018111086, GVSAN2019111085, and GVSAN2020111043 to JR-B, NF-J, and IG-S, respectively; by the Basque Government IT1739-22 grant to JR-B; by the Spanish Ministry of Equity Instituto de las Mujeres 12-4-ID22 to IG-S; by MCIN/AEI/10.13039/501100011033 PID2019-106382RB-I00 to JR-B; by the Instituto de Salud Carlos III grant PI21/0149 to NF-J, and by the Mexican National Council for Science and Technology grant 2021-000007-01EXTF-00209 to BG-G.

Acknowledgments

We thank the MiBioGen consortium for providing gut microbiota GWAS summary statistics data for our analyses. This study makes use of data from the CeD Immunochip study, generated by the Wellcome Trust Case-Control Consortium WTCC, data sets EGAD00010000246, EGAD00010000248 and EGAD00010000250 deposited in the European Genome-Phenome Archive (EGA). A full list of the investigators who contributed to the generation of the data is available from <http://www.wtccc.org.uk> Funding for the project was provided by the Wellcome Trust under awards 076113, 085475, and 099355.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2023.1082862/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

Manhattan plot of the Immunochip Association Study (GWAS) of HR-HLA individuals. Each dot represents a variant plotted as $-\log_{10}$ (p -value) on the y -axis against the corresponding variant position on the x -axis. The red line represents the genome-wide significance threshold value (p -value $< 5 \times 10^{-8}$), and the blue line represents the nominal significance threshold (p -value $< 1 \times 10^{-5}$).

SUPPLEMENTARY TABLE 1

Independent genetic variants below 1×10^{-5} nominal significance level in the HR-HLA CeD association analysis.

SUPPLEMENTARY TABLE 2

Results of the bidirectional 2SMR analysis with the five methods in both directions. HR-HLA CeD: High-risk HLA Celiac Disease; GM: Gut Microbiota; IVs: Instrumental Variables; SE: standard error; p -val: p -value.

References

- Lindfors K, Ciacci C, Kurppa K, Lundin KEA, Makharia GK, Mearin ML, et al. Coeliac disease. *Nat Rev Dis Primers* (2019) 5:1–18. doi: 10.1038/s41572018-0054-z
- de Haas EC, Kumar V, Wijmenga C. Immunogenetics of celiac disease. In: Rampertab SD, Mullin GE, editors. *Clinical gastroenterology*. New York, NY: Humana Press (2014). p. 53–66. doi: 10.1007/978-1-4614-8560-5
- Sollid LM, Thorsby E. HLA susceptibility genes in celiac disease: genetic mapping and role in pathogenesis. *Gastroenterology* (1993) 105:910–22. doi: 10.1016/0016-5085(93)90912-V
- Sollid LM, Markussen G, Ek J, Gjerde H, Vartdal F, Thorsby E. Evidence for a primary association of celiac disease to a particular HLA-DQ α/β heterodimer. *J Exp Med* (1989) 169:345–50. doi: 10.1084/jem.169.1.345
- Valitutti F, Cucchiara S, Fasano A. Celiac disease and the microbiome. *Nutrients* (2019) 11. doi: 10.3390/nu11102403
- Benítez-Páez A, Olivares M, Szajewska H, Pieścik-Lech M, Polanco I, Castillejo G, et al. Breast-Milk microbiota linked to celiac disease development in children: a pilot study from the PreventCD cohort. *Front Microbiol* (2020) 11:1335. doi: 10.3389/fmicb.2020.01335
- Tamburini S, Shen N, Wu HC, Clemente JC. The microbiome in early life: implications for health outcomes. *Nat Med* (2016) 22:713–22. doi: 10.1038/nm.4142
- 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. doi: 10.1038/s41588-020-00763-1
- Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, et al. Enterotypes of the human gut microbiome. *Nature* (2011) 473:174–80. doi: 10.1038/nature09944
- Rinninella E, Raoul P, Cintoni M, Franceschi F, Miggiano GAD, Gasbarrini A, et al. What is the healthy gut microbiota composition? a changing ecosystem across age, environment, diet, and diseases. *Microorganisms* (2019) 7:14. doi: 10.3390/microorganisms7010014
- Collado MC, Donat E, Ribes-Koninckx C, Calabuig M, Sanz Y. Specific duodenal and faecal bacterial groups associated with paediatric coeliac disease. *J Clin Pathol* (2009) 62:264–9. doi: 10.1136/jcp.2008.061366
- Schippa S, Iebba V, Barbato M, di Nardo G, Totino V, Checchi MP, et al. A distinctive “microbial signature” in celiac pediatric patients. *BMC Microbiol* (2010) 10:1–10. doi: 10.1186/1471-2180-10-175
- Nistal E, Caminero A, Herrán AR, Pérez-Andres J, Vivas S, Ruiz de Morales JM, et al. Study of duodenal bacterial communities by 16S rRNA gene analysis in adults with active celiac disease vs non-celiac disease controls. *J Appl Microbiol* (2016) 120:1691–700. doi: 10.1111/jam.13111
- Abdulkhakimova D, Dossybayeva K, Poddighe D. Fecal and duodenal microbiota in pediatric celiac disease. *Front Pediatr* (2021) 9:652208. doi: 10.3389/fped.2021.652208
- Wacklin P, Laurikka P, Lindfors K, Collin P, Salmi T, Lähdeaho ML, et al. Altered duodenal microbiota composition in celiac disease patients suffering from persistent symptoms on a long-term gluten-free diet. *Am J Gastroenterol* (2014) 109:1933–41. doi: 10.1038/ajg.2014.355
- Bonder MJ, Tigchelaar EF, Cai X, Trynka G, Cenit MC, Hrdlickova B, et al. The influence of a short-term gluten-free diet on the human gut microbiome. *Genome Med* (2016) 8:1–11. doi: 10.1186/s13073-016-0295-y
- Zafeiropoulou K, Nichols B, Mackinder M, Biskou O, Rizou E, Karanikolou A, et al. Alterations in intestinal microbiota of children with celiac disease at the time of diagnosis and on a gluten-free diet. *Gastroenterology* (2020) 159:2039–2051.e20. doi: 10.1053/j.gastro.2020.08.007
- Sellitto M, Bai G, Serena G, Fricke WF, Sturgeon C, Gajer P, et al. Proof of concept of microbiome-metabolome analysis and delayed gluten exposure on celiac disease autoimmunity in genetically at-risk infants. *PLoS One* (2012) 7:33387. doi: 10.1371/journal.pone.0033387
- de Palma G, Capilla A, Nova E, Castillejo G, Varea V, Pozo T, et al. Influence of milk-feeding type and genetic risk of developing coeliac disease on intestinal microbiota of infants: the PROFICEL study. *PLoS One* (2012) 7:e30791. doi: 10.1371/journal.pone.0030791
- de Palma G, Capilla A, Nadal I, Nova E, Pozo T, Varea V, et al. Interplay between human leukocyte antigen genes and the microbial colonization process of the newborn intestine. *Curr Issues Mol Biol* (2010) 12:1–10. doi: 10.21775/cimb.012.001
- Olivares M, Neef A, Castillejo G, de Palma G, Varea V, Capilla A, et al. The HLA-DQ2 genotype selects for early intestinal microbiota composition in infants at high risk of developing coeliac disease. *Gut* (2015) 64:406–17. doi: 10.1136/gutjnl-2014-306931
- Olivares M, Walker AW, Capilla A, Benítez-Páez A, Palau F, Parkhill J, et al. Gut microbiota trajectory in early life may predict development of celiac disease. *Microbiome* (2018) 6. doi: 10.1186/s40168-018-0415-6
- Cenit MC, Olivares M, Codoñer-Franch P, Sanz Y. Intestinal microbiota and celiac disease: cause, consequence or co-evolution? *Nutrients* (2015) 7:6900–23. doi: 10.3390/nu7085314
- Smith GD, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* (2014) 23:R89–98. doi: 10.1093/hmg/ddu328
- García-Santisteban I, Cilleros-Portet A, Moyua-Ormazabal E, Kurilshikov A, Zhernakova A, García-Etxebarria K, et al. A two-sample mendelian randomization analysis investigates associations between gut microbiota and celiac disease. *Nutrients* (2020) 12:1420. doi: 10.3390/nu12051420
- Rühlemann 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) 53:147–55. doi: 10.1038/s4158802000747-1
- 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) 54:134–42. doi: 10.1038/s41588-021-00991-z
- Xu Q, Ni JJ, Han BX, Yan SS, Wei XT, Feng GJ, et al. Causal relationship between gut microbiota and autoimmune diseases: a two-sample mendelian randomization study. *Front Immunol* (2022) 12:746998. doi: 10.3389/fimmu.2021.746998
- Trynka G, Hunt KA, Bockett NA, Romanos J, Mistry V, Szperl A, et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat Genet* (2011) 43:1193–201. doi: 10.1038/ng.998
- Zheng X, Shen J, Cox C, Wakefield JC, Ehm MG, Nelson MR, et al. HIBAG - HLA genotype imputation with attribute bagging. *Pharmacogenom J* (2014) 14:192–200. doi: 10.1038/tpj.2013.18
- Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. *Nat Genet* (2016) 48:1284–7. doi: 10.1038/ng.3656
- Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed program. *Nature* (2021) 590:290–9. doi: 10.1038/s4158602103205-y
- Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* (2007) 39:906–13. doi: 10.1038/ng2088
- Turner DS. Qqman: an R package for visualizing GWAS results using q-q and Manhattan plots. *J Open Source Softw* (2018) 3:731. doi: 10.21105/joss.00731
- 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. doi: 10.7554/eLife.34408
- Vacca M, Celano G, Calabrese FM, Portincasa P, Gobetti M, De Angelis M. The controversial role of human gut lachnospiraceae. *Microorganisms* (2020) 8(4):573. doi: 10.3390/microorganisms8040573
- Palmieri O, Castellana S, Bevilacqua A, Latiano A, Latiano T, Panza A, et al. Adherence to Gluten-Free diet restores alpha diversity in celiac people but the microbiome composition is different to healthy people. *Nutrients* (2022) 14(12):2452. doi: 10.3390/nu14122452
- Hansen LBS, Roager HM, Sønderstoft NB, Gøbel RJ, Kristensen M, Vallès-Colomer M, et al. A low-gluten diet induces changes in the intestinal microbiome of healthy Danish adults. *Nat Commun* (2018) 9(1):4630. doi: 10.1038/s41467-018-07019-x
- Haberman Y, Tickle TL, Dexheimer PJ, Kim MO, Tang D, Karns R, et al. Pediatric crohn disease patients exhibit specific ileal transcriptome and microbiome signature. *J Clin Invest* (2014) 124:3617–33. doi: 10.1172/JCI75436
- Gevers D, Kugathasan S, Denson LA, Vázquez-Baeza Y, van Treuren W, Ren B, et al. The treatment-naive microbiome in new-onset crohn's disease. *Cell Host Microbe* (2014) 15:382–92. doi: 10.1016/j.chom.2014.02.005
- Sánchez E, Donat E, Ribes-Koninckx C, Fernández-Murga ML, Sanz Y. Duodenal-mucosal bacteria associated with celiac disease in children. *Appl Environ Microbiol* (2013) 79:5472–9. doi: 10.1128/AEM.00869-13