TITLE

Causal relationships between gut microbiome and hundreds of age-related traits: evidence of a replicable effect on ApoM protein levels.

Authors

Federica Grosso[1], Daniela Zanetti[1], Serena Sanna[1]

Affiliations

[1]Institute for Genetic and Biomedical Research (IRGB) of the National Research Council (CNR), Monserrato (CA), Italy

Contact Information:

Serena Sanna

Institute for Genetic and Biomedical Research (IRGB) of the National Research Council (CNR), Cittadella Universitaria di Monserrato, SS555 Km 4500, Monserrato (CA), Italy

Email: serena.sanna@cnr.it

Tel: +39 0706754593

Keywords

causal inference, aging, gut microbiome, inflammatory proteins, age-related macular degeneration

Abstract

In the past 20 years, the involvement of gut microbiome in human health has received particular attention, but its contribution to age-related diseases remains unclear. To address this, we performed a comprehensive two-sample Mendelian randomization investigation, testing 55130 potential causal relationships between 37 traits representing gut microbiome composition and function and age-related phenotypes, including 1472 inflammatory and cardiometabolic circulating plasma proteins from UK Biobank Pharma Proteomic Project and 18 complex traits. A total of 91 causal relationships remained significant after multiple testing correction (false discovery rate p-value<0.05) and sensitivity analyses, notably two with the risk of developing age-related macular degeneration and 89 with plasma proteins. The link between purine nucleotides degradation II aerobic pathway and apolipoprotein M was further replicated using independent genome-wide association study data. Finally, by taking advantage of previously reported biological function of *Faecalibacterium prausnitzii* we found evidence of regulation of six proteins by its function as mucosal-A antigen utilization. These results support the role of gut microbiome as modulator of the inflammatory and cardiometabolic circuits, that may contribute to the onset of age-related diseases, albeit future studies are needed to investigate the underlying biological mechanisms.

Introduction

Many human disorders, such as cancer, diabetes, neurodegeneration and cardiovascular diseases, are closely linked to the intricate and multifactorial process of aging, which involves physiological changes. To identify possible therapeutic targets for age-related disorders and to understand the mechanism of aging, it is essential to study changes within different body compartments during lifespan, such as those in the gut microbiome [1]. The human gastrointestinal tract (gut) hosts a vast

microbial community of approximately 100 trillion microorganisms, which changes significantly throughout life. Furthermore, it is composed of a large number of immune cells that constantly communicate with the gut microbiota, an essential process for maintaining immune homeostasis. Disruption of this interaction can lead to dysbiosis, which can contribute to the development of autoimmune diseases, inflammatory diseases, cardiovascular diseases, susceptibility to infections, and other health problems [2], [3]. Therefore, changes in the gut microbiota influence various aspects of both gut and systemic immune and inflammatory responses, suggesting a link to the age-related decline in cardiovascular and immune function, often referred to as immune aging, immunosenescence or inflammaging [3].

During aging, significant changes occur in the gut microbiome. For example, opportunistic pathogens such as *Enterobacteria*, which can induce intestinal inflammation, increase, while by contrast beneficial commensals such as *Bacteroides*, *Bifidobacteria* and *Lactobacilli* decrease [3]. These changes in microbial composition underline the importance of studying the microbiota throughout life to better understand its role in aging and associated diseases. However, it remains uncertain whether changes in the microbiota drive these conditions or whether these or the related therapies impact the microbiota. Given this ambiguity, we investigated the causal relationships between the gut microbiome and age-related complex traits, along with associated immune-related and cardiometabolic circulating proteins.

To investigate potential causal links between the gut microbiome and age-related traits and proteins, the two samples Mendelian randomization (MR) technique was employed, a causal inference approach that uses genome-wide association study (GWAS) data to infer causal relationships while controlling for confounding factors and reverse causality. Other studies [4], [5], [6], [7], [8], [9], [10] have used MR to assess the causal relationships between gut microbiome and specific age-related traits. For example, Mao et al. [6] investigated its causality with age-related macular degeneration; Chen et al. [7], [10] explored links with longevity traits; and Bo et al. [8] focused specifically on

70 frailty. In our work, we opted for a comprehensive analysis with an extensive number of age-related
71 outcomes. Furthermore, we emphasized the importance of a rigorous approach to MR analyses. This
72 involves i) employing GWAS carried out on large data set to ensure sufficient statistical power, ii) the
73 use of stringent parameters to ensure validity of MR assumptions such as the independence and strong
74 association of the instrumental variables (IVs), iii) performing multiple testing correction and
75 sensitivity analysis, and iv) exclusion of reverse causality bias via bidirectional MR. Moreover, our
76 study followed the STROBE-MR (Strengthening the Reporting of Observational Studies in
77 Epidemiology using Mendelian Randomization) checklist [11] and guidelines from Burgess et al.
78 [12].

79 Unlike previous studies, we performed replication analyses for the significant results using
80 independent GWAS datasets, a fundamental step that has often been overlooked. Replication [13] is
81 one way to prevent false positive results from being spread [14], [15]. This approach strengthens the
82 robustness and reliability of our findings, making them a more solid reference point for understanding
83 the causal relationships between the gut microbiome and aging-related phenotypes.

84 Results

85 A two-sample MR was performed to investigate the causal relationships between the gut microbiome
86 (exposure) and selected age-related phenotypes (outcome). As exposure we considered 37
87 microbiome features, with at least one variant associated at genome-wide significant level with the
88 trait (**Supplementary Table 1a**). All the exposures showed an F-statistic > 10, indicating sufficient
89 instrument strength and reducing the risk of weak instrument bias (**Supplementary Table 2a**). As
90 outcomes we selected 1490 age-related phenotypes including diseases and quantitative traits **(Table
91 1)**. We searched for publicly available GWASs related to the most common age-related diseases,
92 excluding cancer and neurological disorders, and for traits related to overall aging (like lifespan and
93 longevity). We identified 18 age-related trait GWASs that included at least 20,000 European ancestry
94 individuals and for which the necessary information to perform MR analysis (SNP chromosome and

position, effect allele, other allele, beta, standard error, p-value) was available. We also included 1472 inflammatory and cardiometabolic proteins, selected from the study of Sun et al., 2023 [16].

Out of the 55130 MR tests performed, 91 causal relationships remained significant after multiple-testing correction (FDR p-value<0.05). The results highlighted a causal link between the gut microbiome and age-related macular degeneration (AMD) and levels of 62 distinct circulating protein levels (**Supplementary Table 2**).

Age-related macular degeneration

AMD GWAS was collected from the study of Jiang et al. [17], which analyzed data of 1,295 cases and 455,053 controls from UK Biobank [18]. We found that a genetic predisposition to higher levels of bacteria of the order of *Coriobacteriales* or of family *Coriobacteriaceae* increases the risk of developing AMD disease ($\beta_c = 0.61$, corresponding to an OR=1.84, $p_{BH} = 0.047$). All the sensitivity analyses **(Table 2)** and graphical inspections **(Supplementary Figures 2 and 3**) confirmed this result, further supported by the absence of reverse causality (all $p > 0.87$).

We tested the replicability of these causal relationships using three more GWAS data of AMD **(Supplementary Table 1c)**. Albeit the definition of AMD was different, their large sample size provided sufficient power for replication **(Supplementary Figure 5)**, i.e. p-value<0.05 and same direction of the effect as in the main MR analysis. Despite the sufficient power, none of these three GWASs led to significant results **(Supplementary Figure 4)**.

Cardiometabolic and inflammatory proteins

The GWAS summary statistics for circulating inflammatory and cardiometabolic proteins refer to the study of Sun et al., 2023 [16], which analyzed a total of 2923 proteins in 54219 UK Biobank participants **(Supplementary Table 1b)**. Among all these proteins, we analyzed as outcomes a total of 1472 proteins from the inflammatory and cardiometabolic panels, markers of interest for our study. Out of all the 54464 MR analyses carried out with the 37 microbiome features, 89 showed significant MR results after FDR correction (adjusted p<0.05) and all sensitivity analyses.

Interestingly, out of these 89 total causal relationships detected, 57 were with proteins listed in the inflammatory panel (36 distinct proteins) and 32 with proteins from the cardiometabolic panel (26 distinct proteins) (**Figure 2**).

We replicated significant causal relationships for proteins for which GWASs were available in other studies. Specifically, we used three key datasets: Zhao et al. [19], Sun et al., 2018 [20], and Folkersen et al. [21] (**Supplementary Table 1c**). A relationship was considered replicated if the IVW p-value was less than 0.05 and the direction of the effect was consistent with our main analysis. With these criteria, we found that the causal relationship between the purine nucleotides degradation II aerobic pathway and apolipoprotein M (ApoM) was replicated using the ApoM GWAS from Sun et al., 2018 [20] (GCST90240318) (**Figure 3**). In particular, in both analyses an increase in pathways abundance was causally linked with a decrease in ApoM circulating levels (main analysis: $\beta_c = -0.11, p_{BH} = 0.02$; replication analysis: $\beta_c = -0.22,\ p = 0.01$). All other results are detailed in the **Supplementary Table 2k**.

Of note, unlike the UK Biobank Plasma Proteomic Project (UKB-PPP) proteins used in Sun et al., 2023 [16], measured using the OLINK platform, the INTERVAL study (Sun et al., 2018 [20]) employed the SOMAscan assay for protein quantification. Additionally, the INTERVAL study has a much smaller sample size compared to the UK Biobank (3622 vs. 54219 participants). In contrast, the other two studies – Folkersen et al. [21], and Zhao et al. [19] – also used OLINK for protein measurements albeit with a slightly different methodology, and have sample sizes of 21758 and 14824 participants, respectively.

Further investigation of GalNAc-linked utilization

We further investigated the underlying biological meaning of the significant causal relationships between the lactose-galactose degradation I pathway (LACTOSECAT) pathway, and plasma protein levels, being this pathway strongly associated with a genetic variant in the *ABO* gene [22]. We therefore aimed to investigate whether the causal relationships observed in our study might be

connected to the utilization of N-acetylgalactosamine (GalNAc) in blood type A individuals, since a recent study from our collaborators [23] showed that strains of *ABO*-associated species, such as strains of *Faecalibacterium prausnitzii*, can utilize this sugar from individuals who secrete the mucosal A-antigen.

We conducted MR analysis using as exposures host genetic variants associated with the deletion SV (dSV) region (577-579 kb) present in strains of *F. prausnitzii* known to be active in degradation of secreted mucosal A-antigens (**Supplementary Table 1d**), and proteins and LACTOSECAT pathway as outcomes. First, we corroborated the existing biological evidence with MR, and showed that predisposition to higher load of *F. prausnitzii* SVs (thus higher GalNAc utilization activity) is causally linked with higher levels of the lactose-galactose degradation I pathway, supporting our hypothesis that the activity of this microbial pathway is related to mucosal-A antigen utilization (IVW $p < 0.05$ for all SVs) (**Supplementary Table 2r**). We also detected a causal link between *F. prausnitzii* SV and 2 cardiometabolic and 4 inflammatory protein levels in blood. Specifically, we found that higher load of *F. prausnitzii* SVs and consequently higher mucosal-A antigen utilization, was causally linked with an increase in circulating levels of Pulmonary surfactant-associated protein D (SFTPD), Sialic acid-binding Ig-like lectin 8 (SIGLEC8), Triggering receptor expressed on myeloid cells 2 (TREM2), Alpha-1-antichymotrypsin (SERPINA3) and Complement component C7 (C7), and with decreasing levels of T-cell surface glycoprotein CD3 gamma chain (CD3G) (all IVW $p < 0.05$). Furthermore, the direction of the effect of the MR analysis between *F. prausnitzii* SV and proteins is consistent with the direction of the effect of the MR analysis detected between LACTOSECAT pathway and proteins (**Figure 4**).

Discussion

In this study, we investigated the causal effect of variation in gut microbiome composition and function on age-related traits, using the Mendelian randomization approach. Among all the

170 investigated outcomes, the analyses highlighted a potential causal effect on AMD, and on 36
171 inflammatory proteins and 26 cardiometabolic circulating plasma protein levels.

172 A strong and replicable causal relationship highlighted by our study is the one between purine
173 nucleotides degradation II (aerobic) pathway and apolipoprotein M protein in plasma. This pathway
174 leads to the stepwise breakdown of purine nucleotides into urate, which in humans is the final product
175 due to the absence of the enzyme uricase. Urate is mostly reabsorbed by the kidneys, with only ~10%
176 excreted [24]. Its accumulation has been linked to cardiovascular risk - hyperuricemia is associated
177 with higher incidence of cardiovascular events [25], [26]. Our findings suggest that increased purine
178 degradation may contribute to cardiovascular risk through two converging mechanisms: elevated
179 urate levels and reduced ApoM concentrations. ApoM, mainly produced in the liver and kidneys, is
180 essential for HDL formation and function, promoting cholesterol efflux and exerting anti-
181 inflammatory and atheroprotective effects. Its location in the Major Histocompatibility Complex
182 (MHC) class III region also links it to immune and inflammatory processes [27]. An increase in urate
183 levels may be a risk factor for endothelial dysfunction [28], a process that, particularly in older
184 individuals who are more vulnerable to chronic inflammation, can lead to downregulation of ApoM
185 expression. This may contribute to the development of atherosclerosis and subsequent vascular events
186 [29].

187 Another interesting result is the identification of causal relationships between the LACTOSECAT
188 pathway and six circulating protein levels, which we showed to be attributable to GalNAc utilization
189 of secreted mucosal A antigen in blood type A individuals. Using MR and the existing knowledge of
190 host-interaction function of *F. prausnitzii* strains, we confirmed that higher abundance of
191 LACTOSECAT pathway implicates higher GalNAc utilization in blood type A individuals, and this
192 in turn increases or decreases the levels of these proteins. Specifically, GalNAc utilization in blood
193 type A individuals leads to higher activity of the pathway, and in turn to increased levels of SIGLEC8,
194 SFTPD, TREM2, SERPINA3, and C7. Whereas CD3G levels decrease with both higher *F. prausnitzii*

and increased LACTOSECAT pathway activity. Based on these results, we speculate that differences in disease risk among blood type groups for age-related conditions on which these proteins are potentially involved may be different due to a modulation of gut microbiome on host protein levels. In particular some studies demonstrated that TREM2 is protective for atherosclerosis [30], while the increase of the presence of some other proteins, like SERPINA3 and C7 led to an increase of cardiac mortality and coronary artery disease respectively [31], [32].

We acknowledge that there is an extensive growth in literature on the associations between gut microbiome and the onset of many chronic and age-related diseases. However, reliability and reproducibility of most of these findings are limited by several pitfalls in the statistical protocol used, such as a lenient selection of IVs, being limited to one single MR test and not performing sensitivity analyses, and the lack of correction for multiple testing. Furthermore, very few studies have aimed to replicate their findings in independent datasets, leaving their results inconclusive.

Chen et al. [7], [10], for example, investigated the presence of potential causal relationships between gut microbiome and five longevity traits – frailty, health span, lifespan, longevity, parental longevity – using the same GWAS data that we have used in our study. Among their exposures there were traits derived from the GWAS carried out on the Dutch Microbiome Project, which we also used in our study. They identified several causal relationships that we were unable to reproduce. The primary differences between their study and ours lie in the methodology, particularly in the threshold used for selecting IVs. Indeed, to avoid overestimating the causal effect, we only focused on microbiome traits with at least one genome-wide significant hit and opted for a more stringent association p-value of $5 \cdot 10^{-6}$, instead of $1 \cdot 10^{-5}$. Furthermore, we did not base our significance on nominal p-values but rather employed multiple testing correction. This approach reduces the number of significant findings but strengthens the robustness and reliability of our results. Out of the 168 exposure-outcome significant pairs reported by Chen et al. [7], [10], 12 would survive our pipeline for selection of exposure and only 3 were nominal significant for the IVW test. However, none were significant after

220 adjusting for multiple testing (all p_{BH}>0.16). Additionally, none would pass the sensitivity analyses
221 therefore they would not be considered significant even without applying FDR (**Supplementary**
222 **Table 3**).

223 Of note, another study (Mao et al. [6]) has specifically investigated the causal link between gut
224 microbiome and AMD, but the two GWAS used for analyses were different from those used by us.
225 As exposure (microbiome GWASs) they used data from the MiBioGen consortium [23], and as
226 outcome (AMD GWAS) data were derived from FinnGen biobank analysis (round 5), which includes
227 3763 cases and 205359 controls. Using a $p < 1 \cdot 10^{-5}$ as significant threshold to select IV, 6
228 microbiome features (out of 211 tested) were found to be causally linked to AMD at p < 0.05 (smallest
229 p=0.005). None of these features include species of the order of *Coriobacteriales*. These results,
230 however, were not supported by a correction for multiple testing, therefore they could not be
231 considered statistically significant. In addition, the other MR methods they implemented were not
232 taken into account; for example, the p-values from MR-PRESSO were never significant (all p>0.05).
233 In contrast, the causal relationship we detected with AMD was robust to FDR correction and to all
234 sensitivity analysis we opted for. It must be noted however that despite we have used a stringent and
235 rigorous pipeline, we were unable to replicate the causal relationship using different datasets.

236 Our findings highlight the importance of a rigorous methodological approach in the application of
237 MR to study the causal relationships between the gut microbiome and age-related phenotypes. A strict
238 observance to MR guidelines [12], [33] and STROBE-MR [11] is essential to ensure the integrity and
239 reliability of the results and a proper evaluation of finding, concept that extends beyond the specific
240 exposure and outcome we selected. Several findings from previous studies [6], [7], [10] that did not
241 follow the guidelines were in fact not supported by our analyses. Our study also focus attention on
242 the need for replication of the results in independent GWAS datasets, rather than assuming that an
243 initial statistical significance can be interpreted as definitive proof of causality. This concept is
244 essential to avoid misinterpretation and to progress towards the discovery of effective therapeutic

245 interventions. The range of applications of MR and related methods for understanding causal
246 mechanisms has in fact expanded rapidly over the past 20 years, coupled with the increasing number
247 of public GWAS studies, and therefore the adherence to a rigorous pipeline is mandatory.

248 We also acknowledge the study limitations. First, the genetic of microbiome traits has a limited impact
249 on explaining the total variation, therefore only one IV could be selected at $5 \cdot 10^{-8}$ for each trait,
250 significantly impacting power of MR. We mitigated this limitation by using a lower threshold of $5 \cdot$
251 10^{-6} to detect more IVs and rigorously assessed the coherence of causal estimates from each IVs
252 using several sensitivity analysis methods. Secondly, we acknowledge that genetics of gut
253 microbiome traits can be influenced by other factors, including diet and lifestyle factors which may
254 confound MR analyses by deviating from the independence assumption. While we have no data to
255 explore this confounding, we attempted to reduce its impact by considering only those microbiome
256 traits with strong genetic link (at least one genome-wide significant hit), most of which are known to
257 consistently replicate in other cohorts regardless of diet [34]. It must be noted that some of these
258 effects may be even stronger depending on diet. For example, the association between ABO genetic
259 variants with bacterial species and pathways, many of which show significant causal relationships in
260 our analyses, is enhanced by fiber intake [35]. Third, we recognize the lack of replication of some
261 results in independent cohorts. Furthermore, we are aware that differences in data collection methods
262 and characteristics of the cohorts analyzed may have introduced sources of bias, which may have
263 prevented the detection of weaker but still relevant relationships. For instance, in the case of AMD
264 we identified three independent GWASs for the disease, each employing different case definitions.
265 The primary AMD dataset considered AMD as a binary trait coded as 362.29 in the ICD-9-CM
266 (**Supplementary Table 1b**), which differs from the definitions used in the other three datasets
267 employed for replication (**Supplementary Table 1c**). Thus, despite the higher statistical power in the
268 replication datasets (**Supplementary Figure 5**), inconsistent phenotypic definitions may have
269 confounded replication results. Similarly, in the case of circulating inflammatory and cardiometabolic

proteins, all replication datasets had lower sample sizes and thus lower power compared with the main analyses.

While our analyses are robust and the results compelling, we were unable to elucidate the underlying biological mechanisms driving the observed causal relationships. Therefore, any clinical application aimed at microbiome modulation remains premature.

In conclusion, we investigated the causal relationships between the gut microbiome and age-related phenotypes using the MR approach and ensured the integrity and reliability of the results by a strict adherence to MR guidelines. Our results support a causal role of gut microbiome in age-related macular degeneration and in both upregulating and downregulating the expression of 36 inflammatory and 26 cardiometabolic protein levels, some of which occur via mucosal-A antigen utilization. Particularly robust was the causal link between a microbial purine nucleotides degradation II aerobic pathway and and levels of the protein ApoM, which was successfully replicated in an independent cohort. While these links are particularly intriguing, we acknowledge that future studies are needed to investigate the underlying biological mechanisms and to further confirm our evidence.

Materials and Methods

Exposure and outcome selection

Two-sample MR was performed to investigate the causal relationship between the gut microbiome (exposure) and selected age-related phenotypes (outcome) (**Figure 1**). As exposure data we used genome-wide association study (GWAS) summary statistics collected from a previous study by Lopera-Maya et al. [22]. Lopera-Maya et al. used shotgun genome sequencing on fecal samples from 7738 individuals participating in the Dutch Microbiome Project [36] to derive quantitative information on 207 taxa and 205 pathways. By analyzing these 412 microbiome variables with GWAS method, they identified significant genetic associations ($p < 5 \cdot 10^{-8}$) for 37 of these: 18 taxa and 19 pathways; these 37 were considered as exposure for our study. The relevant GWAS summary statistics

can be accessed through the GWAS Catalog (https://www.ebi.ac.uk/gwas/home) [37] **(Supplementary Table 1a)**.

The outcomes of interest included 18 age-related phenotypes, both diseases and quantitative traits **(Table 1)**, of which summarized GWAS results were also collected on GWAS Catalog [37] and, for "lifespan", from the Edinburgh DataShare repository (https://datashare.ed.ac.uk/handle/10283/3209). Additionally, we included 1472 circulating proteins from inflammatory and cardiometabolic panels of UK Biobank Plasma Proteomic Project (UKB-PPP), for which the GWAS summary statistics were reported in the study by Sun et al., 2023 [16]. These outcomes were selected for their relevance to aging and cardiometabolic health, as they comprise key traits related to longevity, frailty, cardiovascular and metabolic disorders, as well as systemic inflammation. Furthermore, these GWASs were selected because they were carried out in very large cohorts of European ancestry (28967<N<1012240), (**Table 1**), so they provided sufficient statistical power for analyses.

GWAS quality control and selection of instrumental variables

We performed a quality control (QC) screening for each of the downloaded GWAS files (37 for the microbiome and 1490 for the outcomes). In particular, we confirmed that all studies were aligned to build 37 (genome assembly GRCh37/hg19) or otherwise to a compatible assembly such as hg38 (cardiovascular aging and circulating protein levels) and assigned missing rsIDs to genetic variants when needed. Furthermore, we removed variants in the exposure if their effect size was very large and likely unreliable (+/- 4 standard deviation units).

After QC, we selected genetic variants to be used as IVs for MR analyses. Since MR assumptions require independent (not in linkage disequilibrium) and highly associated IVs, we selected genome-wide significant ($p < 5 \cdot 10^{-6}$) independent hits, after linkage disequilibrium clumping using a window of 10 Mb and R^2 cutoff = 0.001.

To avoid the weak instrument bias, we computed the F statistics [38] as:

$$F = \frac{R^2(n-k-1)}{k(1-R^2)}$$

where k is the number of SNPs in the instrument and n the sample size of the exposure GWAS. R^2 is the total proportion of explained variance for each exposure, calculated by summing all R^2_{SNP} values of all variants included in the instrument. It can be computed, using the following formula [39]:

$$R^2_{SNP} = \frac{2\beta^2 MAF(1-MAF)}{2\beta^2 MAF(1-MAF) + 2n\big(se(\beta)\big)^2 MAF(1-MAF)}$$

where MAF is the minor allele frequency, β is the effect size for a given variant and n the sample size of the exposure GWAS. Instruments with F >10 were considered sufficiently strong and retained for analysis.

To reduce computational time in the clumping process, we used a local version of the European population (503 samples) genotypes from the 1000 Genomes Project (phase 3) [40], [41], that excluded variants with minor allele frequency (MAF) less than 1%.

Methods for causal inference

After the general QC of GWAS, two-sample MR analysis was performed, using "TwoSampleMR" version 0.6.8 package [42], [43] in R environment version 4.4.1. The main MR method we employed to evaluate causality is the "Inverse Variance Weighted" (IVW) method [44], [45], as it is the strongest method with the greatest discovery power. To avoid the risk of overestimating the significance of our findings, we applied a correction for multiple testing within each of the 1490 outcomes – the Benjamini-Hochberg (BH) false discovery rate (FDR) correction [46] – considering a significance level of 5% after correction.

The MR method relies on three important assumptions: (1) IVs are associated with exposure (relevant exposure); (2) IVs are not associated with outcome due to confounding pathways (independence); (3) IVs do not impact the outcome directly, except potentially via exposure (exclusion restriction) [44].

While the first assumption is met by our selection of IV with the clumping method, the other two assumptions cannot be investigated *a priori*. Therefore, in addition of the IVW method, we employed other MR methods as sensitivity analyses: weighted median method [47] and MR-PRESSO (using "MRPRESSO" version 1.0 package on R) [48], [49]. Given the limited power, we considered significant those tests with nominal p-value less than 0.05 (thus not applying multiple testing correction). Of note, while we derivate causal estimates also from the MR Egger test and these are provided on Supplementary Tables, they were not used to define a causal relationship significant, as these estimates have inflated Type 1 error rates [50]. Nevertheless, even if not significant in some cases, the direction of the causal effect was concordant with other MR methods. In addition, we specifically investigated the presence of constant horizontal pleiotropy, heterogeneity, and the influence of a single variant on causal estimates with several statistical tests [50]. In particular, these included the Egger intercept term which assumes all IVs display the same amount of pleiotropy, the Cochran's Q statistics which assesses heterogeneity across single IV causal estimates [51], and the leave-one-out test – where IVW is repeated removing in turn one IV from the set of IVs. The results were also visualized graphically using scatter plots and leave-one-out plots.

Significant results after IVW were also subjected to bidirectional MR analysis, it means using the outcome as exposure and the exposure as outcome, to confirm the lack of significant reverse causal relationships. This approach ensures that observed associations are not merely due to reverse causation or confounding, thereby strengthening the validity of causal inferences drawn from the study.

We then run replication analyses for all significant findings (those passing IVW and sensitivity analyses), using independent GWAS' datasets as outcome and exposure, if available (**Supplementary Table 1c**).

Post-hoc power analysis

364 For all our significant results we calculated the post-hoc power to compare the power of the main and
365 replication analyses and have further information about the strength of our findings. We used for
366 continuous outcomes the formula [52]:

367 $$Power_C = \Phi\left(\sqrt{(R^2\beta_c^2 N)} - z_{1-\alpha/2}\right)$$

368 and for binary outcomes:

369 $$Power_B = \Phi\left(\sqrt{NR^2\beta_c^2\frac{ratio}{1+ratio}} - z_{1-\alpha/2}\right)$$

370 where β_c is the causal effect obtained from main MR analysis, N is the sample size of the outcome's
371 GWAS, R^2 is the exposure's explained variance by the selected IVs and *ratio* is calculated by dividing
372 the number of cases by the number of controls of the outcome GWAS. The function Φ is the
373 cumulative distribution function of the standard normal distribution and $z_{1-\alpha/2}$ is the quantile of the
374 standard normal distribution corresponding $1-\alpha/2$, for a significance level $\alpha = 0.05$.

375 To compute the explained variance R^2, required to calculate the statistical power, we used the formula
376 described in the QC section [39].

377 **Further investigation of GalNAc-linked utilization**

378 Since in the study by Lopera-Maya et al. [22] the LACTOSECAT pathway was highly associated to
379 a genetic variant in the *ABO* gene, the gene responsible for determining blood group types
380 (**Supplementary Table 1**), we further investigated the significant relationships related to this
381 pathway. A recent study by our collaborators [23] has shown that *ABO*-associated species can also
382 utilize N-acetylgalactosamine (GalNAc) sugar from blood type A individuals who can secret the
383 mucosal A-antigen. We therefore aimed to investigate whether the causal relationships observed in
384 our study might be connected to utilization of GalNAc in blood type A individuals.

385 Specifically, we assessed the causal relationships between the presence of structural variants (SVs)
386 of *Faecalibacterium prausnitzii*, which contain genes that perform GalNAc degradation activity [53],
387 and the LACTOSECAT pathway and significant outcome results. Of note, this bacteria species was
388 not detected in Lopera-Maya et al. [22], and thus no GWAS was available for our MR analyses.
389 However, the study of Zhernakova et al. [37] identified human single nucleotide polymorphisms
390 (SNPs) that alter the abundance of the GalNAc utilization gene region, and thus these SNPs could be
391 used to investigate the causal effect of *F. prausnitzii*.

392 We conducted MR analysis using the abundance of the deletion SV (dSV) region (577-579 kb) related
393 to GalNAc activity in *F. prausnitzii* as exposures (**Supplementary Table 1d**), with proteins and
394 lactose-galactose degradation I pathway and significant related phenotypes as outcomes.

395 We performed clumping of GWAS related to abundance of the dSV with the same criteria used in the
396 main analysis and then we tested the causal relationships with MR method, considering as significant
397 the results with IVW p-value < 0.05.

398 # Abbreviations

399 AMD: Age-related macular degeneration

400 BH: Benjamini-Hochberg

401 C7: Complement component C7

402 CD3G: T-cell surface glycoprotein CD3 gamma chain

403 dSV: deletion structural variant

404 FDR: False Discovery Rate

405 GalNAc: N-acetylgalactosamine

406 GV: Genetic variant

407 GWAS: Genome-wide association study

408 HDL: high-density lipoproteins

409 IV: Instrumental variable

410 IVW: Inverse Variance Weighted

411 LD: Linkage disequilibrium

412 MAF: Minor allele frequency

413 MR: Mendelian Randomization

414 MR-PRESSO: MR-Pleiotropy Residual Sum and Outlier

415 QC: Quality control

416 SERPINA3: Alpha-1-antichymotrypsin

417 SFTPD: Pulmonary surfactant-associated protein D

418 SIGLEC8: Sialic acid-binding Ig-like lectin 8

419 SNP: Single-nucleotide polymorphism

420 STROBE: Strengthening the reporting of observational studies in epidemiology

421 TREM2: Triggering receptor expressed on myeloid cells 2

422 UKB-PPP: UK Biobank Pharma Proteomic Project

423 WM: Weighted Median

424 Authors contributions

425 Conceptualization: SS

426 Project Supervision: DZ, SS

Statistical analyses: FG

Manuscript Draft Preparation: FG, DZ, SS

Funding: SS

Acknowledgements

We thank Daria Zhernakova and Valeria Lo Faro for critical revisions on the manuscript, and Davide Murrau for support and management of our IT infrastructure.

Conflicts of Interest

None of the authors has conflict of interest to declare.

Ethical Statement

For this study only summarized non-individual data was used.

Funding

This study was supported by project DSB AD006.371 "InvAt" FOE 2022 and project DBA.AD005.225 "NutrAGE" FOE 2021. In addition, we acknowledge Marie-Curie Fellowship to D.Z., acronym 'Sex Dimorphism' (GA n. 101066678) and grants ERC Stg n. 101075624, Next Generation EU -PNRR Investment PE8 – Project Age-It: "Ageing Well in an Ageing Society" PE00000015 to S.S. that indirectly made this study possible.

Availability of data and materials

No data was generated for this study. We used public data [16], [17], [19], [20], [21], [22], [53], [54], [55], [56], [57], [58], [59], [60], [61], [62], [63], [64], [65], [66], download links are available in **Supplementary Table 1**. The scripts used for all the analyses are publicly available in the GitHub repository: https://github.com/Sanna-s-LAB/Mendelian-randomization-Project.git

Use of AI statement

We acknowledge the use of artificial intelligence tools, namely ChatGPT 4.0, to improve code efficiency and parallelization.

Figures legends

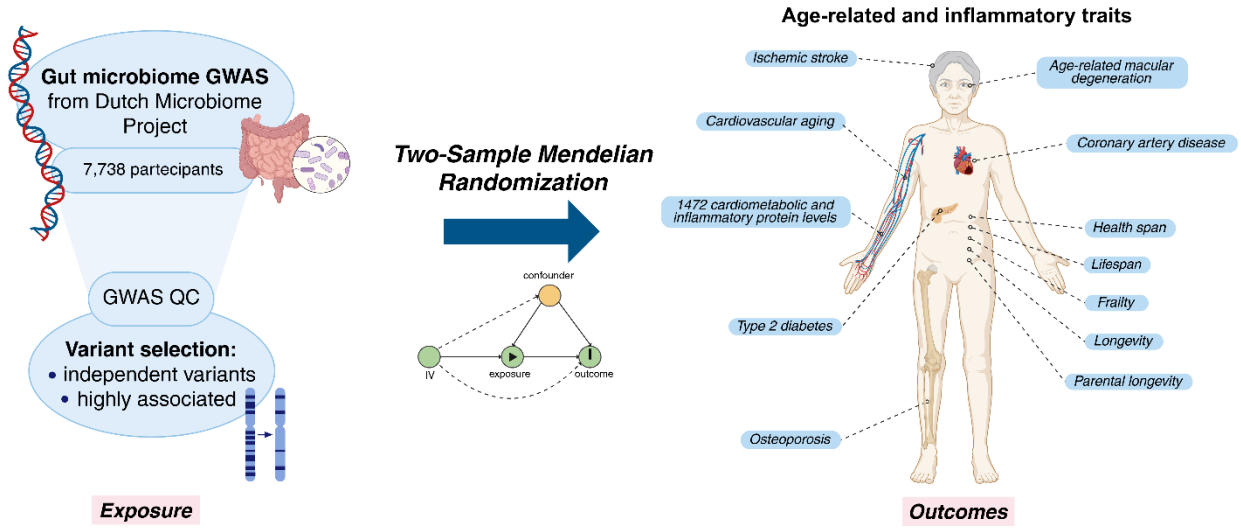


Figure 1. Graphical representation of the study

Created in BioRender. Sanna, S. (2025) https://BioRender.com/a45o861

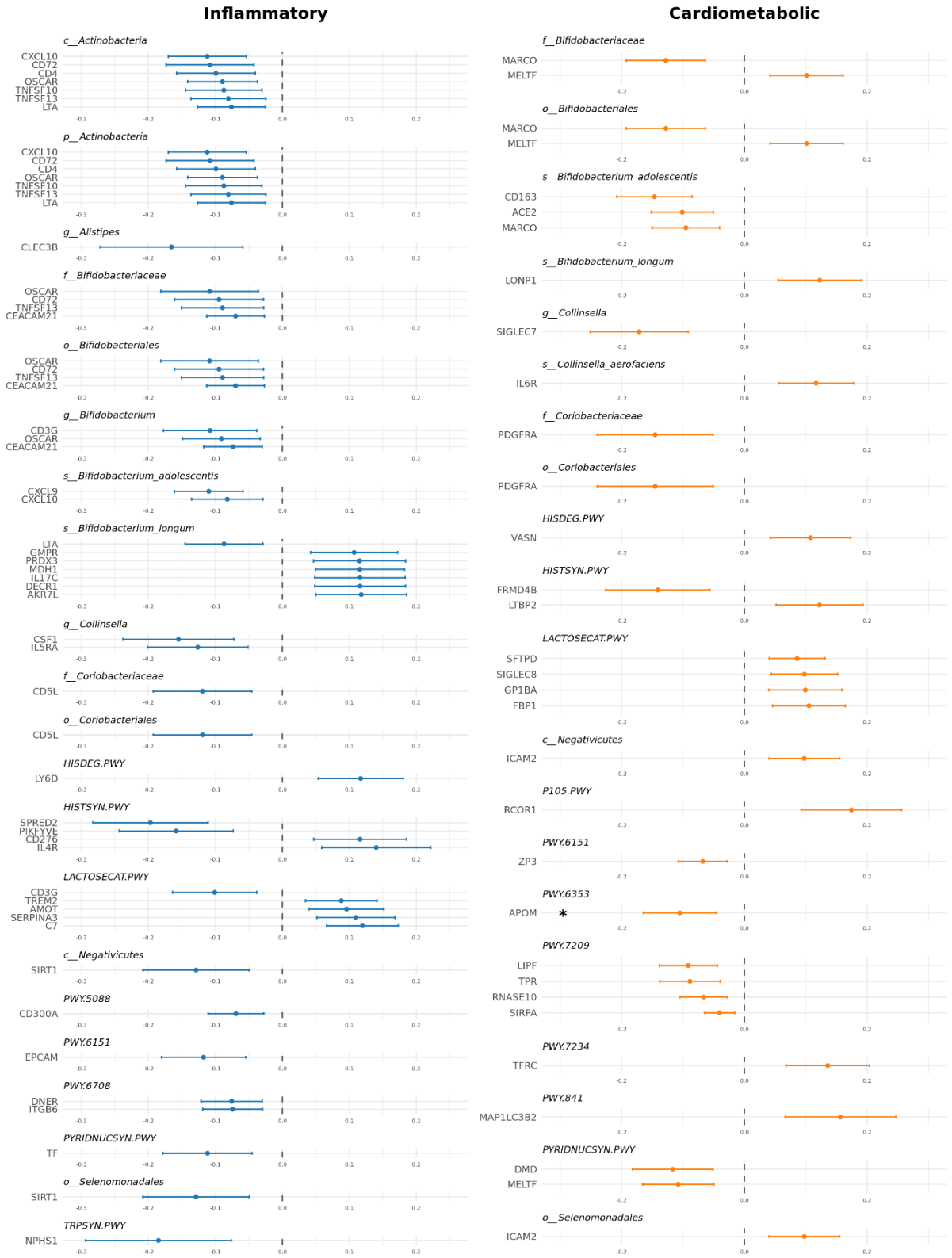


Figure 2. Significant causal relationships between gut microbiome and inflammatory and cardiometabolic proteins

Forest plot showing causal estimates for the 89 significant causal relationships identified.

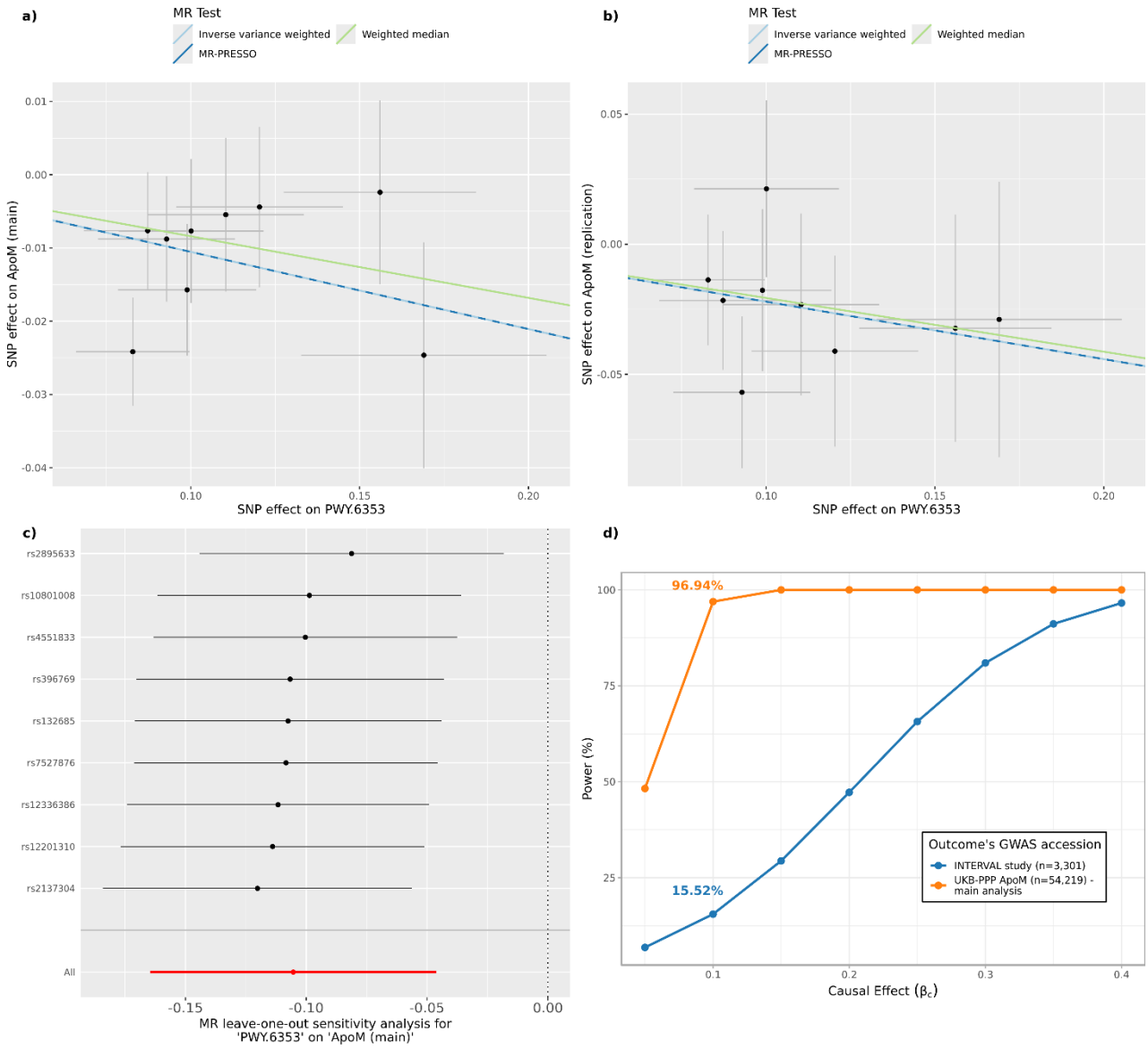


Figure 3. The causal relationship between purine nucleotides degradation pathway and ApoM levels.

The **panel a)** shows the results of the main MR analysis between purine nucleotides degradation II aerobic pathway and ApoM protein. In the scatter plot, each dot is an IV and the x and y-axis represents the association coefficients with the exposure and outcome, respectively. The three lines represent the results of the three different MR tests, with the slope of the lines being equal to causal estimates of each test. The **panel b)** follows the same definition of panel a) but refers to results obtained using the GWAS data of the ApoM protein from Sun et al., 2018 (GCST90240318) as outcome. In the **panel c)** the causal estimates from the leave-one-out analyses are shown and compared with the causal estimate from main analysis (red line). In the **panel d)** we show the post-hoc power estimates at varying causal effect sizes for different studies of ApoM as outcome.

The causal effect identified in the main analysis (β_c = -0.11) and corresponding power are highlighted in the curves.

(ApoM: Apolipoprotein M; GWAS: genome-wide association study, IV: instrumental variable, MR: Mendelian randomization, PWY.6353: purine nucleotides degradation II aerobic pathway, SNP: single nucleotide polymorphism)

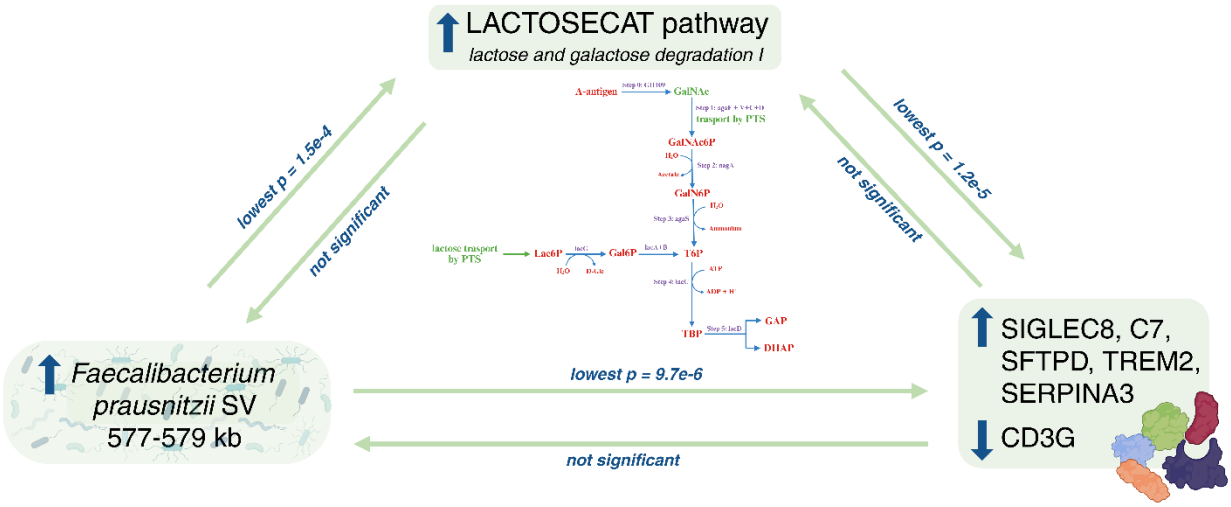


Figure 4. Directed causal relationships involving the abundance of *Faecalibacterium prausnitzii structural variant 577_579*, the LACTOSECAT pathway, and plasma proteins

The figure illustrates the causal relationship between the abundance of *Faecalibacterium prausnitzii* structural variant 577_579 and LACTOSECAT pathway, between LACTOSECAT pathway and six plasma proteins and the one between *Faecalibacterium prausnitzii* and the same six plasma proteins. It also shows the lactose and galactose degradation I pathway (horizontal) along with the GalNAc degradation pathway (vertical), with annotated steps detailing the processes involved in GalNAc degradation, as described in Zhernakova et al. (2024).

Created in BioRender. Sanna, S. (2025) https://BioRender.com/o9k1q5m

(ApoM: Apolipoprotein M; DHAP: glycerone phosphate; GalNAc: N-acetylgalactosamine; Gal6P: D-galactopyranose 6-phosphate; GalNAc6P: N-acetyl-D-glucosamine-6-phosphate; GalN6P: N-Acetylgalactosamine-6-phosphate; GAP: D-glyceraldehyde 3-phosphate; LACTOSECAT pathway: lactose and galactose degradation pathway; Lac6P: lactose 6'-phosphate; T6P: D-tagatofuranose-6-phosphate; TBP: D-tagatofuranose 1,6-biphosphate)

491 # Tables

OUTCOME NAME	DATA ARTICLE	SAMPLE SIZE	TYPE OF TRAIT	CHOICE OF THE TRAIT
Age-related macular degeneration	Jiang et al., 2021 (PMID: 34737426)	456348: 1295 cases and 455053 controls	Binary (ICD-9-CM: 362.29, based on hospital record)	European ancestry with higher sample size
Cardiovascular ageing	Shah M. et al., 2023 (PMID: 37604819)	29506	Continuous	European ancestry with higher sample size
Coronary artery disease	van der Harst P et al., 2017 (PMID: 29212778)	296525: 34541 cases and 261984 controls	Binary	European ancestry with higher sample size
Frailty	Atkins JL et al., 2021 (PMID: 34431594)	175226	Continuous	European ancestry with higher sample size
Health span	Zenin A et al., 2019 (PMID: 30729179)	300447	Continuous	European ancestry with higher sample size
Ischemic stroke	Malik R. et al, 2018 (PMID: 29531354)	440,328: 34,217 cases and 406,111 controls	Binary	European ancestry with higher sample size
Lifespan	Timmers et al., 2019 (PMID: 30642433)	1012240	Continuous	European ancestry with higher sample size
Longevity 90th percentile	Deelen J et al., 2019 (PMID: 31413261)	36745: 11262 cases and 25483 controls	Binary	European ancestry with higher sample size
Longevity 99th percentile	Deelen J et al., 2019 (PMID: 31413261)	28967: 3484 cases and 25483 controls	Binary	European ancestry with higher sample size
Osteoporosis	Dönertaş HM et al., 2021 (PMID: 33959723)	484598: 7751 cases and 476847 controls	Binary	European ancestry with higher sample size
Parental longevity (combined parental attained age)	Pilling LC et al., 2017 (PMID: 29227965)	389166	Continuous	European ancestry with higher sample size
Parental longevity (both parents in top 10%)	Pilling LC et al., 2017 (PMID: 29227965)	86949	Continuous	European ancestry with higher sample size
Parental longevity (mother's age at death)	Pilling LC et al., 2017 (PMID: 29227965)	246941	Continuous	European ancestry with higher sample size
Parental longevity (mother's attained age)	Pilling LC et al., 2017 (PMID: 29227965)	412937	Continuous	European ancestry with higher sample size
Parental longevity (father's age at death)	Pilling LC et al., 2017 (PMID: 29227965)	317652	Continuous	European ancestry with higher sample size
Parental longevity (father's attained age)	Pilling LC et al., 2017 (PMID: 29227965)	415311	Continuous	European ancestry with higher sample size
Parental longevity (combined parental age at death)	Pilling LC et al., 2017 (PMID: 29227965)	208118	Continuous	European ancestry with higher sample size
Type 2 diabetes	Loh PR et al., 2018 (PMID: 29892013)	468298	Continuous	European ancestry with higher sample size
Circulating inflammatory and cardiometabolic proteins	Sun et al., 2023 (PMID: 37794186)	54219	Continuous (OLINK)	Most extensive recent panel of circulating inflammatory and cardiovascular proteins

492 **Table 1. Age-related outcomes selected for the MR analysis**

493 This table shows all the GWAS summary statistics used as outcomes in the MR analysis and the reason for the
494 choice of these specific GWAS. GWAS: genome-wide association study, MR: mendelian randomization.

495

EXPOSURE	OUTCOME	β_c	IVW post-FDR correction pval	WM pval	MR-PRESSO pval	Pleiotropy pval	Heterogeneity pval	Reverse causality pval	Lowest replication pval
Coriobacteriales	AMD	0.61	0.047	0.007	0.009	0.44	0.75	0.98	0.234
Coriobacteriaceae	AMD	0.61	0.047	0.007	0.009	0.44	0.75	0.98	0.234

496

497 **Table 2. Significant causal relationships between gut microbiome features and age-related**
498 **macular degeneration.**

This table shows significant results of the MR analysis between gut microbiome and age-related macular degeneration. For each exposure-outcome pair, we show the causal estimate β_c (corresponding effect on disease risk liability for 1 standard deviation unit increase on the gut microbiome trait), the p-value obtained with IVW method, after FDR multiple testing correction, and the p-value from all sensitivity analyses (weighted median, MR-PRESSO as pleiotropy aware MR-tests, tests of pleiotropy (Egger intercept) and heterogeneity (Cochran's Q), and the reverse causality). The last column indicates the lowest p-value of the replicated causal relationship using independent datasets.

(AMD: age-related macular degeneration, FDR: false discovery rate, IVW: inverse variance weighted, MR: Mendelian randomization, WM: weighted median)

References

[1] X. Shen *et al.*, 'Nonlinear dynamics of multi-omics profiles during human aging', *Nat. Aging*, pp. 1–16, Aug. 2024, doi: 10.1038/s43587-024-00692-2.

[2] K. Hou *et al.*, 'Microbiota in health and diseases', *Signal Transduct. Target. Ther.*, vol. 7, no. 1, pp. 1–28, Apr. 2022, doi: 10.1038/s41392-022-00974-4.

[3] R. Nagpal *et al.*, 'Gut microbiome and aging: Physiological and mechanistic insights', *Nutr. Healthy Aging*, vol. 4, no. 4, pp. 267–285, Jun. 2018, doi: 10.3233/NHA-170030.

[4] H. Li and C. Li, 'Causal relationship between gut microbiota and type 2 diabetes: a two-sample Mendelian randomization study', *Front. Microbiol.*, vol. 14, Aug. 2023, doi: 10.3389/fmicb.2023.1184734.

[5] 'Gut microbiome and risk of ischaemic stroke: a comprehensive Mendelian randomization study | European Journal of Preventive Cardiology | Oxford Academic'. Accessed: Jan. 07, 2025. [Online]. Available: https://academic.oup.com/eurjpc/article/30/7/613/7044714

[6] D. Mao *et al.*, 'Causal Effects of Gut Microbiota on Age-Related Macular Degeneration: A Mendelian Randomization Study', *Invest. Ophthalmol. Vis. Sci.*, vol. 64, no. 12, p. 32, Sep. 2023, doi: 10.1167/iovs.64.12.32.

[7] S. Chen, W. Chen, X. Wang, and S. Liu, ‘Mendelian randomization analyses support causal relationships between gut microbiome and longevity’, *J. Transl. Med.*, vol. 22, no. 1, p. 1032, Nov. 2024, doi: 10.1186/s12967-024-05823-2.

[8] F. Bo *et al.*, ‘Exploring the causal relationship between gut microbiota and frailty: a two-sample mendelian randomization analysis’, *Front. Med.*, vol. 11, May 2024, doi: 10.3389/fmed.2024.1354037.

[9] A. D. Rosa, ‘Causal link between gut microbiota and osteoporosis analyzed via Mendelian randomization’, European Review. Accessed: Jan. 07, 2025. [Online]. Available: https://www.europeanreview.org/article/35052

[10] S. Chen *et al.*, ‘Consistent signatures in the human gut microbiome of longevous populations’, *Gut Microbes*, vol. 16, no. 1, p. 2393756, Dec. 2024, doi: 10.1080/19490976.2024.2393756.

[11] V. W. Skrivankova *et al.*, ‘Strengthening the Reporting of Observational Studies in Epidemiology Using Mendelian Randomization: The STROBE-MR Statement’, *JAMA*, vol. 326, no. 16, p. 1614, Oct. 2021, doi: 10.1001/jama.2021.18236.

[12] S. Burgess *et al.*, ‘Guidelines for performing Mendelian randomization investigations: update for summer 2023’, *Wellcome Open Res.*, vol. 4, p. 186, 2019, doi: 10.12688/wellcomeopenres.15555.3.

[13] Directorate-General for Research and Innovation (European Commission) *et al.*, *Reproducibility of scientific results in the EU: scoping report*. Publications Office of the European Union, 2020. Accessed: Sep. 03, 2024. [Online]. Available: https://data.europa.eu/doi/10.2777/341654

[14] M. R. Munafò *et al.*, ‘A manifesto for reproducible science’, *Nat. Hum. Behav.*, vol. 1, no. 1, pp. 1–9, Jan. 2017, doi: 10.1038/s41562-016-0021.

[15] M. Baker, ‘1,500 scientists lift the lid on reproducibility’, *Nature*, vol. 533, no. 7604, pp. 452–454, May 2016, doi: 10.1038/533452a.

[16] B. B. Sun *et al.*, ‘Plasma proteomic associations with genetics and health in the UK Biobank’, *Nature*, vol. 622, no. 7982, pp. 329–338, Oct. 2023, doi: 10.1038/s41586-023-06592-6.

[17] L. Jiang, Z. Zheng, H. Fang, and J. Yang, ‘A generalized linear mixed model association tool for biobank-scale data’, *Nat. Genet.*, vol. 53, no. 11, pp. 1616–1621, Nov. 2021, doi: 10.1038/s41588-021-00954-4.

[18] ‘UKBiobank PheWeb’. Accessed: Jul. 04, 2024. [Online]. Available: https://pheweb.org/UKB-Neale/

[19] J. H. Zhao *et al.*, ‘Genetics of circulating inflammatory proteins identifies drivers of immune-mediated disease risk and therapeutic targets’, *Nat. Immunol.*, vol. 24, no. 9, pp. 1540–1551, Sep. 2023, doi: 10.1038/s41590-023-01588-w.

[20] B. B. Sun *et al.*, ‘Genomic atlas of the human plasma proteome’, *Nature*, vol. 558, no. 7708, pp. 73–79, Jun. 2018, doi: 10.1038/s41586-018-0175-2.

[21] L. Folkersen *et al.*, ‘Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals’, *Nat. Metab.*, vol. 2, no. 10, pp. 1135–1148, Oct. 2020, doi: 10.1038/s42255-020-00287-2.

[22] E. A. Lopera-Maya *et al.*, ‘Effect of host genetics on the gut microbiome in 7,738 participants of the Dutch Microbiome Project’, *Nat. Genet.*, vol. 54, no. 2, pp. 143–151, Feb. 2022, doi: 10.1038/s41588-021-00992-y.

[23] A. Kurilshikov *et al.*, ‘Large-scale association analyses identify host factors influencing human gut microbiome composition’, *Nat. Genet.*, vol. 53, no. 2, pp. 156–165, Feb. 2021, doi: 10.1038/s41588-020-00763-1.

[24] ‘MetaCyc purine nucleotides degradation II (aerobic)’. Accessed: Jun. 03, 2025. [Online]. Available: https://metacyc.org/pathway?orgid=META&id=PWY-6353

[25] M. L. Muiesan, C. Agabiti-Rosei, A. Paini, and M. Salvetti, ‘Uric Acid and Cardiovascular Disease: An Update’, *Eur. Cardiol. Rev.*, vol. 11, no. 1, pp. 54–59, Aug. 2016, doi: 10.15420/ecr.2016:4:2.

[26] G.-Z. Liao *et al.*, ‘Serum uric acid fluctuations and cardiovascular outcomes in hospitalized acute heart failure patients’, *Nutr. Metab. Cardiovasc. Dis. NMCD*, vol. 35, no. 6, p. 104025, Jun. 2025, doi: 10.1016/j.numecd.2025.104025.

[27] R. Yadav, H. Obinata, and K. Venkataraman, ‘Delineating the intricacies of polymorphisms, structures, functions and therapeutic applications of biological high-density lipoprotein-apolipoprotein M: A review’, *Int. J. Biol. Macromol.*, vol. 310, no. Pt 2, p. 143187, May 2025, doi: 10.1016/j.ijbiomac.2025.143187.

[28] W. Cai *et al.*, ‘Uric Acid Induces Endothelial Dysfunction by Activating the HMGB1/RAGE Signaling Pathway’, *BioMed Res. Int.*, vol. 2017, p. 4391920, 2017, doi: 10.1155/2017/4391920.

[29] K. R. Feingold, J. K. Shigenaga, L. G. Chui, A. Moser, W. Khovidhunkit, and C. Grunfeld, 'Infection and inflammation decrease apolipoprotein M expression', *Atherosclerosis*, vol. 199, no. 1, pp. 19–26, Jul. 2008, doi: 10.1016/j.atherosclerosis.2007.10.007.

[30] M. Piollet *et al.*, 'TREM2 protects from atherosclerosis by limiting necrotic core formation', *Nat. Cardiovasc. Res.*, vol. 3, no. 3, pp. 269–282, Mar. 2024, doi: 10.1038/s44161-024-00429-9.

[31] R. Aarsetøy *et al.*, 'Complement component 7 is associated with total- and cardiac death in chest-pain patients with suspected acute coronary syndrome', *BMC Cardiovasc. Disord.*, vol. 21, no. 1, p. 496, Oct. 2021, doi: 10.1186/s12872-021-02306-w.

[32] B. Li *et al.*, 'The Association and Pathogenesis of SERPINA3 in Coronary Artery Disease', *Front. Cardiovasc. Med.*, vol. 8, p. 756889, Dec. 2021, doi: 10.3389/fcvm.2021.756889.

[33] S. Burgess, B. Woolf, A. M. Mason, M. Ala-Korpela, and D. Gill, 'Addressing the credibility crisis in Mendelian randomization', *BMC Med.*, vol. 22, no. 1, p. 374, Sep. 2024, doi: 10.1186/s12916-024-03607-5.

[34] S. Sanna, A. Kurilshikov, A. van der Graaf, J. Fu, and A. Zhernakova, 'Challenges and future directions for studying effects of host genetics on the gut microbiome', *Nat. Genet.*, vol. 54, no. 2, pp. 100–106, Feb. 2022, doi: 10.1038/s41588-021-00983-z.

[35] Y. Qin *et al.*, 'Combined effects of host genetics and diet on human gut microbiota and incident disease in a single population cohort', *Nat. Genet.*, vol. 54, no. 2, pp. 134–142, Feb. 2022, doi: 10.1038/s41588-021-00991-z.

[36] R. Gacesa *et al.*, 'The Dutch Microbiome Project defines factors that shape the healthy gut microbiome', Nov. 27, 2020, *bioRxiv*. doi: 10.1101/2020.11.27.401125.

[37] E. Sollis *et al.*, 'The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource', *Nucleic Acids Res.*, vol. 51, no. D1, pp. D977–D985, Jan. 2023, doi: 10.1093/nar/gkac1010.

[38] S. Burgess, S. G. Thompson, and CRP CHD Genetics Collaboration, 'Avoiding bias from weak instruments in Mendelian randomization studies', *Int. J. Epidemiol.*, vol. 40, no. 3, pp. 755–764, Jun. 2011, doi: 10.1093/ije/dyr036.

[39] J. Yarmolinsky *et al.*, ‘Circulating Selenium and Prostate Cancer Risk: A Mendelian Randomization Analysis’, *JNCI J. Natl. Cancer Inst.*, vol. 110, no. 9, pp. 1035–1038, May 2018, doi: 10.1093/jnci/djy081.

[40] 1000 Genomes, ‘1000 Genomes | A Deep Catalog of Human Genetic Variation’. Accessed: Jul. 09, 2024. [Online]. Available: https://www.internationalgenome.org/

[41] K. Rogers, ‘1000 Genomes Project | Genome Mapping & Sequencing’, *Encycl. Br.*, Jul. 2024, Accessed: Aug. 26, 2024. [Online]. Available: https://www.britannica.com/event/1000-Genomes-Project

[42] G. Hemani *et al.*, ‘The MR-Base platform supports systematic causal inference across the human phenome’, *eLife*, vol. 7, p. e34408, May 2018, doi: 10.7554/eLife.34408.

[43] G. Hemani, K. Tilling, and G. D. Smith, ‘Orienting the causal relationship between imprecisely measured traits using GWAS summary data’, *PLOS Genet.*, vol. 13, no. 11, p. e1007081, Nov. 2017, doi: 10.1371/journal.pgen.1007081.

[44] S. Burgess and S. G. Thompson, *Mendelian Randomization: Methods for Causal Inference Using Genetic Variants*, 2nd ed. Chapman and Hall/CRC, 2021. Accessed: Apr. 11, 2024. [Online]. Available: https://www.mendelianrandomization.com/index.php/book-chapters

[45] S. Burgess, A. Butterworth, and S. G. Thompson, ‘Mendelian Randomization Analysis With Multiple Genetic Variants Using Summarized Data’, *Genet. Epidemiol.*, vol. 37, no. 7, pp. 658–665, Nov. 2013, doi: 10.1002/gepi.21758.

[46] Benjamini & Hochberg, ‘Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing - Benjamini - 1995 - Journal of the Royal Statistical Society: Series B (Methodological) - Wiley Online Library’, Journal of the Royal Statistical Society, 1995. Accessed: Jun. 28, 2024. [Online]. Available: https://rss.onlinelibrary.wiley.com/doi/10.1111/j.2517-6161.1995.tb02031.x

[47] J. Bowden, G. Davey Smith, P. C. Haycock, and S. Burgess, ‘Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator’, *Genet. Epidemiol.*, vol. 40, no. 4, pp. 304–314, May 2016, doi: 10.1002/gepi.21965.

[48] M. Verbanck, C.-Y. Chen, B. Neale, and R. Do, 'Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases', *Nat. Genet.*, vol. 50, no. 5, pp. 693–698, May 2018, doi: 10.1038/s41588-018-0099-7.

[49] R. Do, *Performs the Mendelian Randomization Pleiotropy RESidual Sum and Outlier (MR-PRESSO) method | https://github.com/rondolab/MR-PRESSO*. (Jun. 15, 2024). R. Accessed: Jul. 01, 2024. [Online]. Available: https://github.com/rondolab/MR-PRESSO

[50] S. Burgess and S. G. Thompson, 'Interpreting findings from Mendelian randomization using the MR-Egger method', *Eur. J. Epidemiol.*, vol. 32, no. 5, pp. 377–389, 2017, doi: 10.1007/s10654-017-0255-x.

[51] N. A. Sheehan and V. Didelez, 'Epidemiology, genetic epidemiology and Mendelian randomisation: more need than ever to attend to detail', *Hum. Genet.*, vol. 139, no. 1, pp. 121–136, Jan. 2020, doi: 10.1007/s00439-019-02027-3.

[52] S. Burgess, 'Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome', *Int. J. Epidemiol.*, vol. 43, no. 3, pp. 922–929, Jun. 2014, doi: 10.1093/ije/dyu005.

[53] D. V. Zhernakova *et al.*, 'Host genetic regulation of human gut microbial structural variation', *Nature*, vol. 625, no. 7996, pp. 813–821, Jan. 2024, doi: 10.1038/s41586-023-06893-w.

[54] M. Shah *et al.*, 'Environmental and genetic predictors of human cardiovascular ageing', *Nat. Commun.*, vol. 14, no. 1, p. 4941, Aug. 2023, doi: 10.1038/s41467-023-40566-6.

[55] P. van der Harst and N. Verweij, 'Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease', *Circ. Res.*, vol. 122, no. 3, pp. 433–443, Feb. 2018, doi: 10.1161/CIRCRESAHA.117.312086.

[56] J. L. Atkins *et al.*, 'A genome-wide association study of the frailty index highlights brain pathways in ageing', *Aging Cell*, vol. 20, no. 9, p. e13459, Sep. 2021, doi: 10.1111/acel.13459.

[57] A. Zenin *et al.*, 'Identification of 12 genetic loci associated with human healthspan', *Commun. Biol.*, vol. 2, p. 41, 2019, doi: 10.1038/s42003-019-0290-0.

[58] R. Malik *et al.*, 'Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes', *Nat. Genet.*, vol. 50, no. 4, pp. 524–537, Apr. 2018, doi: 10.1038/s41588-018-0058-3.

[59] P. R. Timmers *et al.*, ‘Genomics of 1 million parent lifespans implicates novel pathways and common diseases and distinguishes survival chances’, *eLife*, vol. 8, p. e39856, Jan. 2019, doi: 10.7554/eLife.39856.

[60] J. Deelen *et al.*, ‘A meta-analysis of genome-wide association studies identifies multiple longevity genes’, *Nat. Commun.*, vol. 10, no. 1, p. 3669, Aug. 2019, doi: 10.1038/s41467-019-11558-2.

[61] H. M. Dönertaş, D. K. Fabian, M. F. Valenzuela, L. Partridge, and J. M. Thornton, ‘Common genetic associations between age-related diseases’, *Nat. Aging*, vol. 1, no. 4, pp. 400–412, Apr. 2021, doi: 10.1038/s43587-021-00051-5.

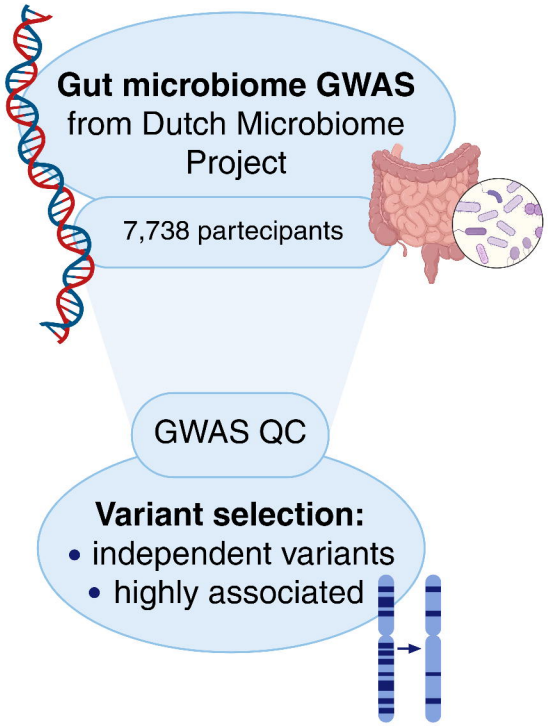
[62] L. C. Pilling *et al.*, ‘Human longevity: 25 genetic loci associated in 389,166 UK biobank participants’, *Aging*, vol. 9, no. 12, pp. 2504–2520, Dec. 2017, doi: 10.18632/aging.101334.

[63] P.-R. Loh, G. Kichaev, S. Gazal, A. P. Schoech, and A. L. Price, ‘Mixed-model association for biobank-scale datasets’, *Nat. Genet.*, vol. 50, no. 7, pp. 906–908, Jul. 2018, doi: 10.1038/s41588-018-0144-6.

[64] M. Guindo-Martínez *et al.*, ‘The impact of non-additive genetic associations on age-related complex diseases’, *Nat. Commun.*, vol. 12, no. 1, p. 2436, Apr. 2021, doi: 10.1038/s41467-021-21952-4.

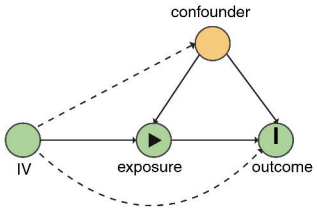
[65] T. W. Winkler *et al.*, ‘Genome-wide association meta-analysis for early age-related macular degeneration highlights novel loci and insights for advanced disease’, *BMC Med. Genomics*, vol. 13, no. 1, p. 120, Aug. 2020, doi: 10.1186/s12920-020-00760-7.

[66] M. I. Kurki *et al.*, ‘FinnGen: Unique genetic insights from combining isolated population and national health register data’, Mar. 06, 2022, *medRxiv*. doi: 10.1101/2022.03.03.22271360.

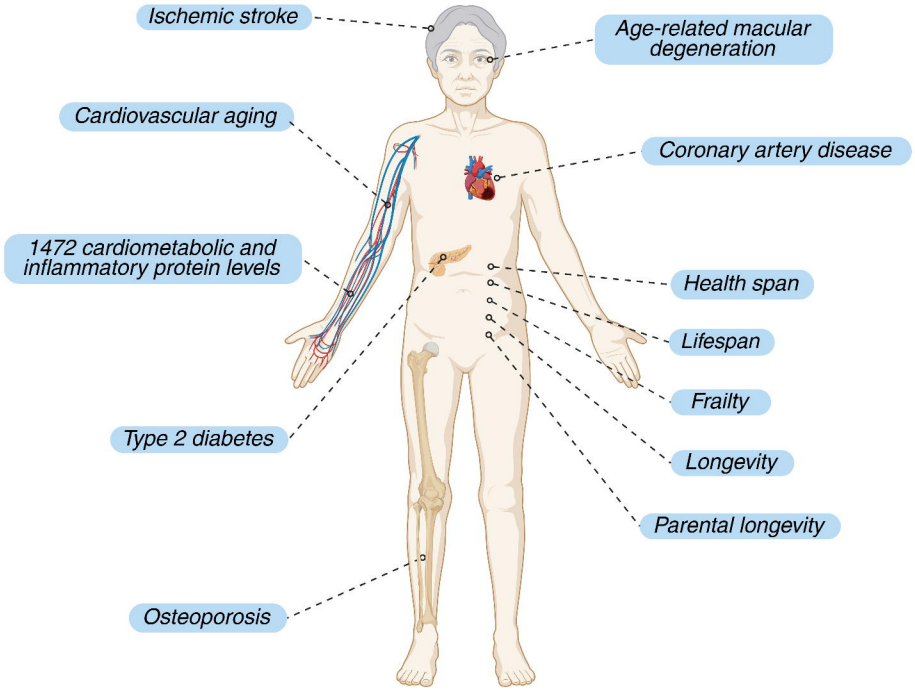


Exposure

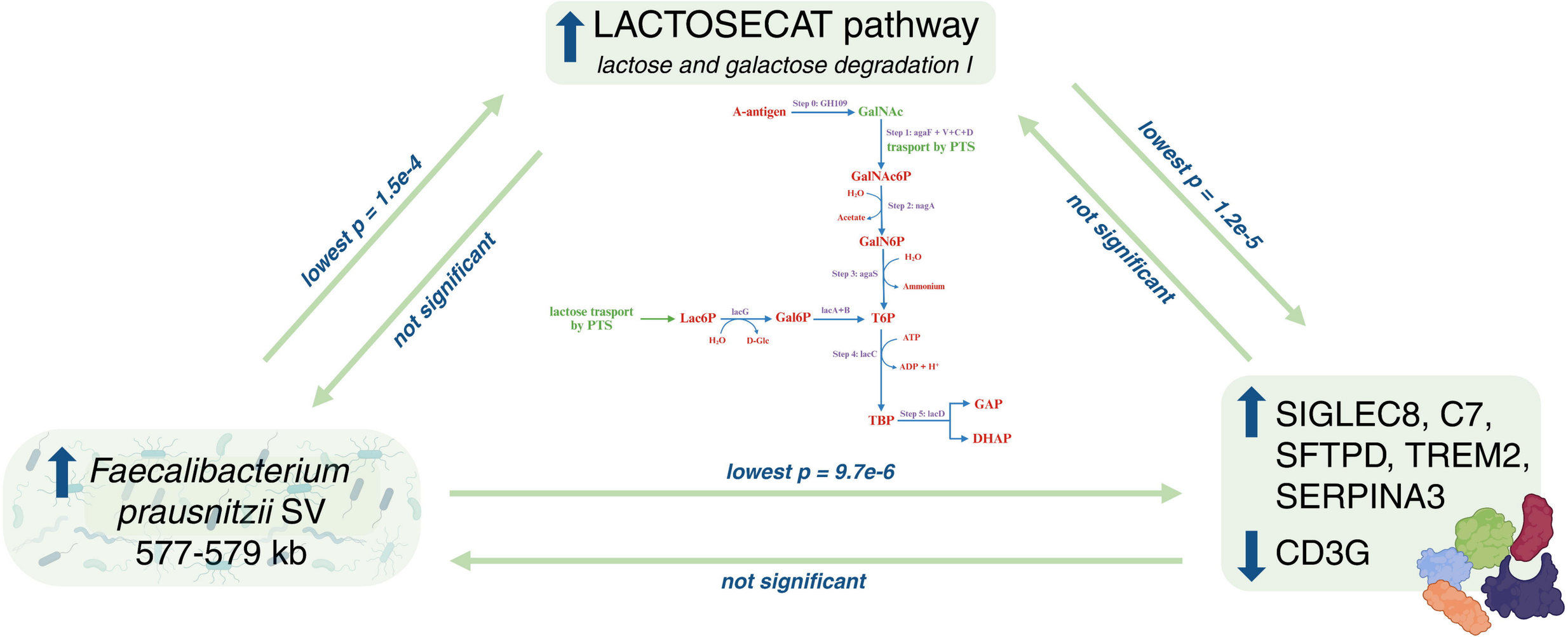
Two-Sample Mendelian Randomization

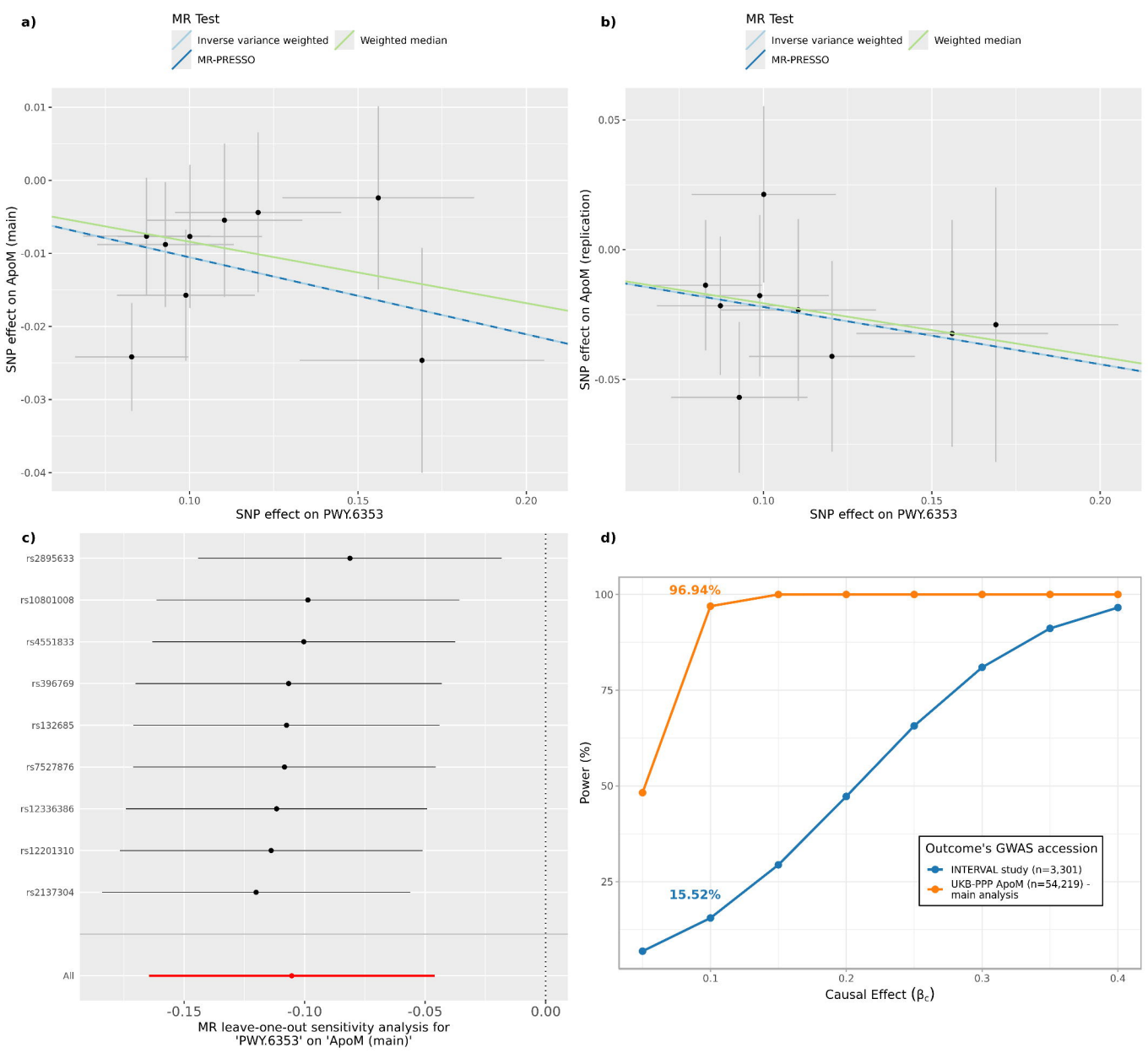


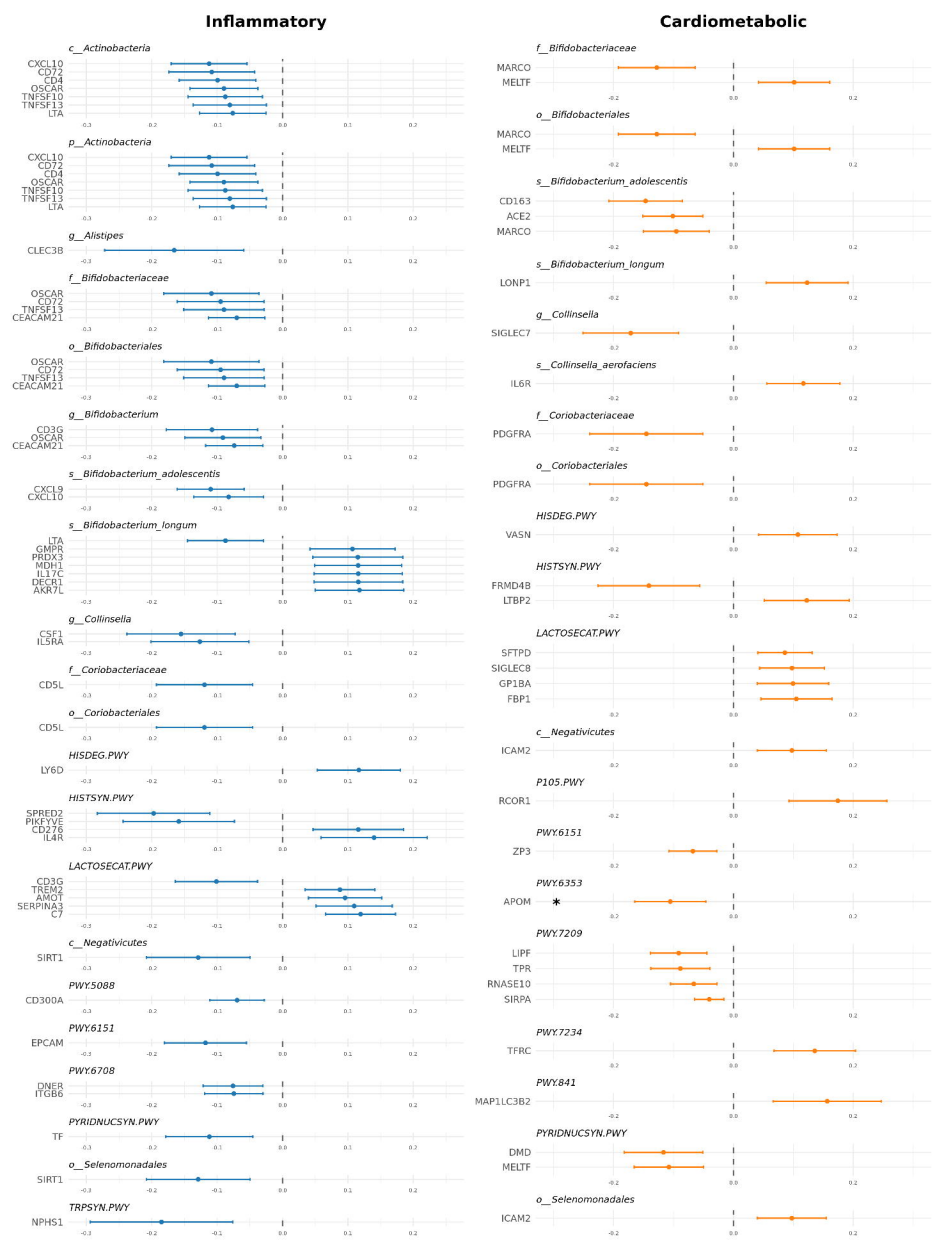
Age-related and inflammatory traits



Outcomes


LACTOSECAT pathway
lactose and galactose degradation I
lowest p = 1.5e-4
not significant
lowest p = 1.2e-5
not significant
A-antigen
Step 0: GH109
GalNAc
Step 1: agaF + V+C+D
trasport by PTS
GalNAc6P
H2O
Step 2: nagA
Acetate
GalN6P
H2O
Step 3: agaS
Ammonium
lactose trasport
by PTS
Lac6P
lacG
Gal6P
lacA+B
T6P
H2O
D-Glc
ATP
Step 4: lacC
ADP + H+
TBP
Step 5: lacD
GAP
DHAP
Faecalibacterium
prausnitzii SV
577-579 kb
lowest p = 9.7e-6
not significant
SIGLEC8, C7,
SFTPD, TREM2,
SERPINA3
CD3G


a)
MR Test
Inverse variance weighted
Weighted median
MR-PRESSO
SNP effect on ApoM (main)
SNP effect on PWY.6353
b)
MR Test
Inverse variance weighted
Weighted median
MR-PRESSO
SNP effect on ApoM (replication)
SNP effect on PWY.6353
c)
rs2895633
rs10801008
rs4551833
rs396769
rs132685
rs7527876
rs12336386
rs12201310
rs2137304
All
MR leave-one-out sensitivity analysis for 'PWY.6353' on 'ApoM (main)'
d)
96.94%
15.52%
Power (%)
Causal Effect (β_c)
Outcome's GWAS accession
INTERVAL study (n=3,301)
UKB-PPP ApoM (n=54,219) - main analysis


Inflammatory
c__Actinobacteria
CXCL10
CD72
CD4
OSCAR
TNFSF10
TNFSF13
LTA
p__Actinobacteria
CXCL10
CD72
CD4
OSCAR
TNFSF10
TNFSF13
LTA
g__Alistipes
CLEC3B
f__Bifidobacteriaceae
OSCAR
CD72
TNFSF13
CEACAM21
o__Bifidobacteriales
OSCAR
CD72
TNFSF13
CEACAM21
g__Bifidobacterium
CD3G
OSCAR
CEACAM21
s__Bifidobacterium_adolescentis
CXCL9
CXCL10
s__Bifidobacterium_longum
LTA
GMPR
PRDX3
MDH1
IL1C
DECR1
AKR7L
g__Collinsella
CSF1
IL5RA
f__Coriobacteriaceae
CD5L
o__Coriobacteriales
CD5L
HISDEG.PWY
LY6D
HISTSYN.PWY
SPRED2
PIKFYVE
CD276
IL4R
LACTOSECAT.PWY
CD3G
TREM2
AMOT
SERPINA3
C7
c__Negativicutes
SIRT1
PWY.5088
CD300A
PWY.6151
EPCAM
PWY.6708
DNER
ITGB6
PYRIDNUCSYN.PWY
TF
o__Selenomonadales
SIRT1
TRPSYN.PWY
NPHS1
Cardiometabolic
f__Bifidobacteriaceae
MARCO
MELTF
o__Bifidobacteriales
MARCO
MELTF
s__Bifidobacterium_adolescentis
CD163
ACE2
MARCO
s__Bifidobacterium_longum
LONP1
g__Collinsella
SIGLEC7
s__Collinsella_aerofaciens
IL6R
f__Coriobacteriaceae
PDGFRA
o__Coriobacteriales
PDGFRA
HISDEG.PWY
VASN
HISTSYN.PWY
FRMD4B
LTBP2
LACTOSECAT.PWY
SFTPD
SIGLEC8
GP1BA
FBP1
c__Negativicutes
ICAM2
P105.PWY
RCOR1
PWY.6151
ZP3
PWY.6353
APOM
*
PWY.7209
LIPF
TPR
RNASE10
SIRPA
PWY.7234
TFRC
PWY.841
MAP1LC3B2
PYRIDNUCSYN.PWY
DMD
MELTF
o__Selenomonadales
ICAM2