Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Cohort dependent associations between gut Bacteroides and circulating sphingolipid balance

Hangyul Song
Gwangju Institute of Science and Technology
Dong Jin Yu
Gwangju Institute of Science and Technology
Younghoon Go
Korea Institute of Oriental Medicine
Chang-Myung Oh
cmoh@gist.ac.kr

Gwangju Institute of Science and Technology

Research Article

Keywords: gut microbiome, Bacteroides, serine palmitoyltransferase, sphingolipids, heart failure, replication

Posted Date: November 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11318398/v1

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

Additional Declarations: No competing interests reported.

Abstract

Taxonomic and gene-level measurements offer complementary views of microbial contributions to host lipid metabolism. We examined gut *Bacteroides* abundance, bacterial serine palmitoyltransferase (SPT) gene read abundance and circulating sphingolipids in heart failure and MetaCardis. In 58 patients, the *Bacteroides* association with the lactosylceramide(16:0)-to-ceramide(16:0) ratio persisted across 85 observations (β = 0.296, P = 0.007). *Bacteroides* abundance correlated with SPT read abundance (ρ = 0.389, P = 0.0026), but SPT did not improve the adjusted lipid model (likelihood-ratio P = 0.828). The primary MetaCardis analysis did not reproduce the ratio association (n = 544; permutation P = 0.127). An exploratory analysis identified an inverse *Bacteroides*–lactosylceramide association (n = 972; P = 0.0029). These findings establish a framework for distinguishing cohort-specific taxonomic associations from measured microbial gene abundance and prioritise direct functional validation of microbiome–sphingolipid relationships.

Introduction

Bacterial sphingolipids provide a plausible link between gut microbes and host lipid metabolism. In controlled colonisation experiments, sphingolipids produced by *Bacteroides thetaiotaomicron* enter host metabolic pathways and alter hepatic ceramide pools.[1] *Bacteroides*-derived sphingolipids also contribute to intestinal homeostasis.[2] These findings motivate testing microbial sphingolipid-related exposures against circulating lipids in human cohorts.

Human studies have linked gut microbiome composition to circulating sphingolipids across several clinical settings. The HELIUS case-cohort study related microbiome composition and plasma sphingolipid patterns to type 2 diabetes risk.[3] A microscopic colitis study described associated microbial and metabolic signatures,[4] while a pregnancy cohort examined microbiome–lipid associations with gestational diabetes and included experimental follow-up.[5] The taxa, lipid species, populations and endpoints differ across these studies; their results cannot be treated as direct replications of a single taxon–lipid association.

Taxonomic abundance, microbial gene abundance and enzyme activity represent different levels of biological measurement. A taxonomic association may identify a microbial community feature without establishing the gene or biochemical reaction responsible. Our literature search did not identify a directly comparable human cohort analysis pairing bacterial SPT read abundance with this circulating lipid endpoint. This search result describes the records retrieved and does not establish that no such study exists.

In an earlier analysis of a publicly deposited heart-failure dataset,[6,7] we identified an association between a gut *Bacteroides* abundance summary and the plasma lactosylceramide(16:0)-to-ceramide(16:0) ratio. The ratio was selected post hoc after a 756-pair taxon–lipid screen in which no

individual pair survived false-discovery correction. The small, cross-sectional analysis therefore required independent assessment and evaluation of repeated observations.

Here, we combine three approaches: read-level quantification of bacterial SPT in the discovery samples, repeated-measures analysis of the discovery cohort, and an external test of the taxonomic association in MetaCardis.[8] We use the supplied analysis plans to distinguish planned comparisons from exploratory component analyses. The study asks whether a taxonomic signal is accompanied by a measurable SPT gene signal and whether the taxonomic association extends across populations.

Results

The discovery association is reproduced from the frozen analysis set

Refitting the original baseline specification reproduced $\beta = 0.4201$, SE = 0.1142, $P = 5.67 \times 10^{-4}$ and $R^2 = 0.229$ in 58 participants. The unadjusted Spearman correlation was $\rho = 0.4613$ ($P = 2.68 \times 10^{-4}$). Refitting from the source deposit gave the same regression estimates. Covariate-adjusted Freedman–Lane permutation inference in the SPT comparison returned P = 0.00040. This verified baseline was used for the subsequent exposure comparisons (Supplementary Data 1).

Taxonomic and SPT read abundance show distinct associations with the lipid ratio

We quantified reads assigned to bacterial SPT in the same 58 baseline samples to compare a gene-related exposure with the *Bacteroides* taxonomic summary. The assay measures sequence read abundance; it does not measure SPT protein abundance or catalytic activity. An amino-acid seed screen recruited 120,818 reads against a 1,990-sequence bacterial SPT reference set. Of these, 13,253 had an SPT alignment and underwent comparison with 33,291 decoy sequences representing related α-oxoamine synthases and eukaryotic SPT subunits. The supplied classification results contain 2,757 SPT-assigned reads, 4,017 decoy assignments and 6,479 score ties. The median SPT count was 48.5 per sample (range 7–125); the median abundance was 60.4 hits per million parsed reads (range 9.7–145.9). These are rates among parsed reads without dedicated adapter trimming, quality filtering or host-read subtraction. The large proportion of ties (48.9%) illustrates the difficulty of assigning short reads to related enzymes (Supplementary Data 2–10). SPT read abundance varied across samples, but its association with the lipid ratio was not statistically significant in the tested specifications (Table 1). The wide range of observed counts does not by itself establish assay sensitivity or adequate power for the biological effect of interest.

Table 1
Separate taxonomic and SPT read exposure models in the discovery cohort

Exposure	β	SE	P (model)	P (permutation)
Bacteroides CLR	+ 0.4201	0.1142	5.67×10^{-4}	0.00040
SPT hits per million	+ 0.0035	0.0032	**0.281**	**0.281**
SPT per Bacteroidetes relative abundance	−0.1565	0.1923	0.420	0.424
SPT per *Bacteroides* relative abundance	−0.1343	0.0838	0.115	0.114

All models use n = 58 and adjust for age, sex, BMI, NYHA class and statin use. β is the unstandardised coefficient and SE its standard error. Model P values are two-sided coefficient tests; permutation P values use 10,000 Freedman–Lane residual permutations. Taxonomy-normalised SPT rates are not per-genome copy numbers.

We next fitted the *Bacteroides* summary and SPT read abundance jointly with the discovery covariates (Table 2). The *Bacteroides* coefficient remained positive in the joint model (β = 0.4297, P = 0.0012), whereas the SPT coefficient was close to zero (β = −0.00063, P = 0.841). Adding SPT did not detectably improve fit over the *Bacteroides*-only model (likelihood-ratio $\chi^2_1 = 0.047$, P = 0.828; Fig. 1). This is a model comparison rather than a test of causal mediation or proof that bacterial SPT has no biological effect. *Bacteroides* and SPT read abundance were positively correlated (Spearman ρ = 0.389, P = 0.0026). Nevertheless, their adjusted associations with the lipid endpoint differed. Two additional exposures normalised SPT hits per million by the summed relative abundance of Bacteroidetes or *Bacteroides* features; neither reached nominal significance (Table 1). These taxonomy-based normalisations do not estimate copies per bacterial genome. Interpretation is limited by panel-dependent read recruitment, short-read ambiguity and library preparation. The source run metadata label the libraries as whole-genome amplified, and amplification can introduce composition-dependent bias.[9] Applying the same pipeline to all samples does not guarantee equal bias across samples. These limitations reduce the strength of conclusions drawn from the nonsignificant SPT association.

Table 2
Joint model of *Bacteroides* and SPT read abundance

Term in joint model	β	SE	P
Bacteroides CLR	**+ 0.4297**	0.1247	**0.0012**
SPT hits per million	−0.0006	0.0031	0.841

The model includes both exposures and the discovery covariates (n = 58). P values are two-sided coefficient tests. Adding SPT to the *Bacteroides*-only model gave likelihood-ratio $\chi^2_1 = 0.047$ and P = 0.828.

The discovery association persists when repeated observations are included

The source deposit contains 135 records from 108 participants across three visits. Paired metagenomics and lipidomics were available for 85 observations from 58 participants, including 26 participants with at least two paired visits. The repeated-measures analysis therefore included 27 observations beyond baseline.A model with a participant-specific random intercept retained a positive association between *Bacteroides* abundance and the lipid ratio (β = 0.2964, SE = 0.1099, 95% CI 0.081–0.512, P = 0.0070; Supplementary Data 14). The coefficient was smaller than the baseline estimate.

Within-participant estimates remain imprecise

A within–between decomposition[10] used 53 observations from the 26 participants with repeated paired measurements (Table 3). Neither the within-participant coefficient (β = 0.1614, P = 0.498) nor the between-participant coefficient in this restricted subset (β = 0.2387, P = 0.199) reached nominal significance. The within-participant interval included zero (95% CI − 0.305 to 0.628). These data do not establish that changes in *Bacteroides* abundance track changes in the lipid ratio within an individual. They also do not establish that the association is exclusively between participants, because both components were imprecisely estimated in the repeated-measures subset.

Table 3
Within and between participant estimates in the repeat observation subset

Component	β	95% CI	P
Within-participant	+ 0.1614	−0.305 to + 0.628	0.498
Between-participant	+ 0.2387	−0.125 to + 0.603	0.199

The decomposition includes 53 observations from 26 participants. Intervals are Wald 95% confidence intervals calculated as β ± 1.96 × SE. P values are model-based. Neither estimate distinguishes the corresponding component from zero.

The ratio association is not reproduced in MetaCardis

In MetaCardis, 994 unique participants had the taxonomic exposure and both endpoint components. The primary complete-case model included 544 participants after adjustment for age, sex, BMI, statin use, left ventricular ejection fraction (LVEF) and nationality. LVEF substituted for the unavailable NYHA class variable. Models and uncertainty estimates are reported in Table 4 and Supplementary Data 11. The primary estimate was β = −0.0321 (SE = 0.0203; model-based P = 0.114), with participant-permutation P = 0.127 and bootstrap 95% CI − 0.0704 to 0.0051. The point estimate did not support the predicted positive direction. Raw coefficients were not treated as directly exchangeable across the two lipidomic platforms.

Table 4
External cohort estimates for the *Bacteroides* and lipid ratio association

Model	n	β	95% CI	P
Primary (core covariates incl. LVEF)	544	**−0.0321**	−0.0718 to + 0.0077	0.114
Participant permutation, 10,000	544	—	—	**0.127**
Sensitivity (LVEF omitted)	972	−0.0306	−0.0607 to − 0.0005	0.046
Heart-failure stratum (core covariates)	71	−0.0645	−0.1635 to + 0.0345	—
Heart-failure stratum (LVEF omitted)	83	−0.0749	−0.1607 to + 0.0110	—

All models adjust for age, sex, BMI, statin use and nationality; the primary model also includes LVEF. Displayed coefficient intervals are Wald 95% intervals. The primary participant-permutation P value is reported separately; its bootstrap interval is − 0.0704 to 0.0051. Heart-failure strata were designated for effect-size estimation; inferential P values are omitted.

The association did not meet the documented replication criterion requiring a positive coefficient and permutation $P < 0.05$. This finding does not establish a negative association on the primary test.

The sensitivity model excluding LVEF included 972 participants and yielded a small negative coefficient ($\beta = -0.0306$, model-based $P = 0.046$). Because the primary test was nonsignificant and this sensitivity estimate was opposite to the prediction, it was not interpreted as successful replication. Heart-failure stratum estimates were also negative, but their intervals included zero; these small subgroup analyses were designated for effect-size estimation. Discovery, repeated-measures and external-cohort estimates are presented together in Fig. 2. Cohort, assay and covariate differences remain relevant when interpreting their disagreement.

Auxiliary checks provide context for the replication result

We assessed statin–sterol associations and calculated approximate detectable-effect thresholds after the primary result was available. These checks provide context but cannot validate every aspect of the primary analysis. Statin use was inversely associated with cholesterol ($\beta = -0.2851$, $P = 7.0 \times 10^{-46}$; n = 1,018) and cholesterol sulfate in auxiliary models adjusted for age, BMI, sex and nationality (Table 5; Supplementary Data 12). The control analyses used the same source tables and composite-key join strategy, but not the same complete-case participants or LVEF-adjusted covariate set as the primary model.

Table 5
Auxiliary statin and sterol associations in MetaCardis

Control endpoint	n	β	SE	P	Control interpretation
Cholesterol	1,018	**−0.2851**	0.0191	**7.0×10^{-46}**	Inverse association
Cholesterol sulfate	1,018	−0.2415	0.0207	1.2×10^{-29}	Inverse association
Campesterol	1,018	+ 0.2921	0.1271	0.022	Positive association
β-sitosterol	872	+ 0.0026	0.0119	0.824	No detected association

Exposure is statin use; models adjust for age, BMI, sex and nationality and omit LVEF. Sample sizes differ from the primary microbiome–lipid analysis. β and SE use the deposited metabolite scale. Two-sided P values are nominal. These controls do not independently validate the microbiome-to-lipid join.

Recovery of these associations supports the plausibility of the auxiliary data processing. It does not prove that the exposure–lipid model is correctly specified, exclude residual confounding, or establish the quality of the microbiome-to-lipid join. Using the supplied model standard errors, an approximate 80%-power threshold at two-sided α = 0.05 was calculated as 2.802 × SE. On the reported standardised scale, the thresholds were 0.118 for the primary model and 0.090 for the sensitivity model, compared with a discovery coefficient of 0.469. The supplied script calculated the scaling factors from the available source tables rather than the matched complete-case samples; these thresholds are therefore descriptive approximations, not a prospective power guarantee or a formal cross-cohort equivalence test.

Exploratory component analysis identifies an inverse lactosylceramide association

We separately examined the two endpoint components and three neighbouring sphingolipid species using the LVEF-free model (n = 972). These five analyses were performed after the primary result and were exploratory (Table 6; Supplementary Data 13). Lactosylceramide(16:0) was inversely associated with *Bacteroides* abundance (β = −0.0314, P = 0.0029), meeting the Bonferroni threshold of 0.010 for these five tests. Ceramide(16:0) was not detectably associated (β = 0.0023, P = 0.855). The component estimates suggest that lactosylceramide contributes to the negative ratio estimate in the sensitivity analysis. This is a candidate association for independent testing; it is not a replicated primary finding, and a difference in significance between components is not itself a test of their difference.

Table 6
Exploratory associations with individual sphingolipid features in MetaCardis

Endpoint	n	β	SE	P	standardised β
Lactosylceramide(16:0)	972	**−0.0314**	0.0105	**0.0029**	−0.095
Glycosyl-N-palmitoyl-sphingosine (d18:1/16:0)	972	−0.0153	0.0121	0.206	−0.041
Ceramide(18:0)	972	+ 0.0124	0.0135	0.357	+ 0.029
Ceramide(16:0)	972	**+ 0.0023**	0.0124	**0.855**	+ 0.006
Dihydroceramide(16:0)	972	−0.0007	0.0167	0.965	−0.001

All models use 972 participants and adjust for age, sex, BMI, statin use and nationality. P values are two-sided and nominal; the Bonferroni threshold for these five tests is 0.010. Standardised β uses the supplied scaling. These are post-primary exploratory analyses, not replication of a prespecified inverse association.

Supporting analyses show context-dependent myocardial expression patterns

Additional class-level and myocardial analyses were used to assess the scope of the proposed microbiome–sphingolipid relationship. The glycosphingolipid class contrast was D = 0.166, with participant-permutation P = 0.249 and bootstrap 95% CI − 0.077 to 0.260. Per-taxon and covariate-adjusted analyses likewise did not reach nominal significance (P = 0.292 and P = 0.307, respectively). These estimates do not establish equivalence between lipid classes. The cardiac sphingolipid gene-set comparison did not distinguish the selected set from fatty-acid oxidation (P = 0.627) or cholesterol synthesis (P = 0.884). These exploratory comparisons do not demonstrate pathway equivalence or cardiac specificity.

In GSE322803,[11] cardiac *B4galt5* expression was higher in conventionally raised than germ-free mice ($\log_2$ expression difference 0.637, 95% CI 0.215–1.059, nominal P = 0.018; n = 3 per group). The faecal-transfer contrast was smaller (0.059, 95% CI − 0.136 to 0.254), as was the estimate in an independent wild-type gnotobiotic myocardial series (GSE14929;[12] −0.090, 95% CI − 0.329 to 0.148). These contrasts use different experimental designs and are not all independent. Same-direction changes in several B4galt paralogues and a larger *B4galt5* kidney estimate (1.070) do not establish a cardiac-specific response or prove a technical batch effect. *B4galt6*, which was lower in the supplied human failing-myocardium analyses including GSE116250,[13] had small cardiac colonisation estimates (Fig. 3; Supplementary Data 15–17).

Discussion

Combining taxonomic, gene-read and lipid measurements revealed both concordance and divergence. In the discovery cohort, *Bacteroides* abundance correlated with bacterial SPT read abundance, and its association with the circulating lactosylceramide-to-ceramide ratio persisted after repeated observations were included. However, adding SPT read abundance did not detectably improve the taxonomic model. These results distinguish a taxonomic association from the performance of one sequence-based functional proxy in the same participants.

A taxon–metabolite association does not by itself identify the biochemical pathway responsible. Diet, transit time, community composition, medication and host physiology are plausible alternative explanations, while imperfect gene measurement can attenuate a true functional association. Our comparison therefore supports measuring taxonomic and gene-level exposures separately; it does not establish that the taxonomic association is independent of SPT biology or that gene-level exposures are generally less informative. The external MetaCardis analysis did not reproduce the positive ratio association. Its primary estimate was close to zero and opposite in direction, with a permutation P value of 0.127. The auxiliary statin–sterol associations and descriptive precision calculations provide useful context, but neither turns the primary result into proof of biological absence.

Several explanations for the cross-cohort difference remain possible. Post hoc selection and sampling variability may have inflated the discovery estimate. The ratio was selected after a screen with no false-discovery-significant pairs, and subsequent analyses in the same participants are not independent replication. The lipidomic platforms and measurement scales also differ. Standardising coefficients removes units but does not remove assay-specific measurement error, annotation uncertainty or differences in biological matrices. In addition, the *Bacteroides* summaries average different species-feature sets across cohorts. Population and covariate differences remain relevant. MetaCardis represents a broader cardiometabolic spectrum, and its small heart-failure subgroup is not a sufficiently precise disease-matched replication to exclude a heart-failure-specific effect. The repeated-measures analysis strengthens the observation that the discovery association is not restricted to a single baseline fit. Nevertheless, the within–between decomposition remains inconclusive. Neither component was statistically resolved among the 26 participants with repeat paired measurements, so the data do not support an exclusively between-person or a longitudinal causal interpretation.

Published human cohorts provide biological context rather than direct directional confirmation. HELIUS, microscopic colitis and pregnancy studies differ in taxa, lipid definitions, populations and analytical models.[3,4,5] The exploratory inverse lactosylceramide association in MetaCardis should therefore be assessed on its own merits and tested in a new cohort, rather than described as established agreement with that literature. Experimental colonisation studies establish that bacterial sphingolipids can interact with host intestinal and hepatic metabolism.[1,2] Our circulating-lipid analyses address a different compartment and a different exposure. Neither a nonsignificant plasma association nor a small myocardial transcript change excludes microbial effects on tissue lipids, protein activity or metabolic flux. The value of the combined analysis is to define which observations can motivate mechanistic work. The positive taxon–gene correlation supports a measurable link between these two microbial

summaries, whereas the exploratory inverse lactosylceramide result identifies a specific circulating lipid for follow-up. Both require confirmation with more direct functional measurements and harmonised external data.

Several limitations affect interpretation. The discovery cohort was small and sex-imbalanced (46 male and 12 female participants); sex-specific effects were not assessed. Only 26 participants contributed repeat paired measurements. The primary MetaCardis analysis required complete covariates, including LVEF, and its heart-failure subgroup contained 71–83 participants. SPT recruitment depended on reference-panel coverage and short-read discrimination, nearly half of candidate reads were ties, and the libraries were labelled as whole-genome amplified. No dedicated quality trimming or host-read subtraction was performed. Identical processing does not imply identical measurement bias across samples. The SPT comparison was confined to the discovery cohort; the planned secondary gene-level analysis in MetaCardis was not completed. Host DNA fraction was not estimated despite its contribution to the parsed-read denominator. Taxonomy-normalised SPT rates are not genome-equivalent copy estimates. The five lipid-component tests were exploratory, and their multiplicity correction does not account for all earlier outcome-selection decisions. The supplied analysis plans document intended decisions, but a publicly timestamped registration record was not provided. Supplementary material also contains historical interim outputs; the final tables and data files used here are identified explicitly.

Together, these analyses identify cohort-dependent relationships between gut *Bacteroides* and circulating sphingolipid balance. The association in the heart-failure cohort persisted when repeated observations were included, and *Bacteroides* abundance correlated positively with bacterial SPT read abundance. The lack of improved lipid-model fit after adding SPT, and the absence of replication of the ratio association in MetaCardis, delineate the limits of a taxonomic interpretation. The exploratory inverse lactosylceramide association provides a specific candidate for independent testing. Integrating taxonomy, gene read abundance and host lipid measurements therefore provides a practical framework for selecting microbiome–lipid hypotheses for functional validation, without equating gene abundance with enzyme activity or causality.

Methods

Discovery cohort

We reanalyzed publicly deposited paired faecal metagenomics and plasma lipidomics from chronic systolic heart failure (Figshare DOI 10.6084/m9.figshare.30422413; BioProject PRJNA1266263).[6,7] The frozen baseline set contains 58 participants, 221 species features and 36 sphingolipid species (NYHA I/II/III/IV = 20/25/11/2; 46 male and 12 female; LVEF 10–55%). Paired measurements across visits yielded 85 observations from these 58 participants. Controls without lipidomics could not contribute to the lipid analyses. The deposited frozen set is Supplementary Data 1.

Taxonomic exposure and lipid endpoint

The exposure was the mean centred log-ratio (CLR) across *Bacteroides*-labelled features: 17 of 221 species features in discovery and 30 of 729 in MetaCardis. CLR used natural logarithms across the full species table after replacement of zeros by half the minimum positive abundance. The discovery endpoint was the difference between the deposited $\log_2$ lactosylceramide(16:0) and ceramide(16:0) concentrations. The MetaCardis endpoint was the difference between the corresponding deposited log-transformed metabolite measurements, without an additional logarithm. Differences in platform, scaling and taxonomy were retained as limitations.

Discovery models and repeated observations

Baseline ordinary least squares models adjusted for age, sex, BMI, NYHA class and statin use. The SPT comparison used 10,000 Freedman–Lane permutations of reduced-model residuals with an absolute t-statistic and the (r + 1)/(10,001) correction.[14] The mixed model used a participant random intercept. The within–between model decomposed the exposure into participant means and within-participant deviations[10] among 26 participants with repeated paired observations. Model-based uncertainty was reported for these fits; it should not be described as permutation inference for every result. Random procedures used seed 20260924.

MetaCardis processing and models

The processed microbiome, metabolome, clinical and medication tables from the source publication[8] were joined by the composite key (ID, Status). Statin use was coded from Yes/No entries. Rows representing overlapping cohort labels were deduplicated by participant ID; they were not treated as repeated visits. The primary model adjusted for age, sex, BMI, statin use, LVEF and nationality. The sensitivity model omitted LVEF. Heart-failure strata were selected before deduplication to preserve group membership and were reported as effect-size analyses.

Replication inference and analysis plans

The primary replication used 10,000 participant-level permutations and 10,000 participant bootstrap samples. The supplied results report model-based P = 0.114, permutation P = 0.127 and a bootstrap interval separately. The planned positive association was considered reproduced only when the coefficient was positive and permutation $P < 0.05$. A documented amendment replaced the originally proposed random-intercept model with ordinary least squares after duplicate rows were identified. The public registration status and the chronology of the plan, amendment and data access require author verification.

Auxiliary control and precision analyses

Statin–sterol models used age, BMI, sex and nationality as covariates; they did not include LVEF or require the microbiome exposure and therefore used different complete-case samples from the primary model. Approximate detectable coefficients were calculated as $(z_{0.975} + z_{0.80}) \times SE$. Standardisation used $\beta \times SD(\text{exposure})/SD(\text{endpoint})$. The supplied script used source-table standard deviations rather than

model-specific matched-sample standard deviations; the resulting estimates are retained as descriptive precision summaries (Supplementary Data 12).

Exploratory lipid components

The two ratio components and three neighbouring sphingolipid features were analysed with the LVEF-free model (n = 972). The five tests were exploratory and interpreted against a Bonferroni threshold of 0.010 (Supplementary Data 13). A glycosylceramide feature is reported by its deposited annotation rather than assuming that the measurement resolves glucosylceramide from other isomers.

Literature search.

Europe PMC queries were used to assess comparable human microbiome–sphingolipid studies and gene-level designs. The supplied broad search contains 461 deduplicated records, with separate gene-level search outputs and cohort-assessment records (Supplementary Data 18–20). This was a targeted literature and novelty assessment, not a systematic review. Failure to identify a study in these searches was not treated as proof of an unprecedented design.

Analysis-plan availability

The submitted supplementary document contains the study analysis plans and historical amendments. [AUTHOR INPUT: provide a persistent public link and verifiable dates for the original plan and amendment; distinguish prospective specification from later revisions. Until verified, this manuscript uses “documented analysis plan” rather than claiming public preregistration.]

Bacterial SPT read quantification

The 58 baseline samples were matched to sequencing runs using record_id and ENA sample_alias (Supplementary Data 10). Only R1 reads were sampled, with a target compressed-byte budget of 60,000,000 bytes per sample; the number actually parsed was recorded. No adapter/quality trimming or host subtraction was performed. Reads longer than 1,000 nt were excluded as parsing artefacts. SPT hits per million were calculated using the number of parsed reads as the denominator, not the number of quality-filtered bacterial reads.

Reference sequences and read classification

The bacterial SPT reference set contained 1,990 UniProtKB sequences retrieved with EC 2.3.1.50 and bacterial taxonomy, including BT_0870 (UniProt Q8A9E5; https://www.uniprot.org/uniprotkb/Q8A9E5). The decoy set comprised 33,291 related BioF, ALAS, KBL and eukaryotic SPTLC sequences. Six-frame translation and exact amino-acid 8-mer matching recruited candidate reads. BLASTX confirmation used $E \leq 10^{-5}$, amino-acid identity ≥ 50%, alignment length ≥ 30 amino acids, query genetic code 11 and SEG masking.[15] SPT, decoy and tied assignments were compared using their best bitscores.

SPT normalisation and model comparison

The primary read exposure was SPT hits per million parsed reads. Additional exposures divided this rate by the summed relative abundance of Bacteroidetes or *Bacteroides* species features from the frozen table. No single-copy-marker or genome-equivalent normalisation was performed. Models used the discovery covariates above. A likelihood-ratio test compared the joint *Bacteroides*-plus-SPT model with the *Bacteroides*-only model, with one additional parameter. Spearman correlation described the relation between the two exposures (Supplementary Data 5–7).

Archived myocardial and tissue expression

GSE322803 supplied RNA-sequencing counts from six tissues, with three animals per microbial-status group.[11] The supplied analysis notes describe tissue-wise normalisation and $\log_2$ expression transformation. Conventional-versus-germ-free and faecal-transfer-versus-germ-free contrasts share the same germ-free group. GSE14929 microarray values were analysed on a $\log_2$ scale after per-array median centring; the wild-type model adjusted for physiological state, and the *Ppara*-null comparison used Welch tests.[12] These are heterogeneous experimental contrasts rather than four independent replications. Where several probes measured a gene, the highest-expressed representative probe was displayed. Reported intervals and P values were model-based or Welch estimates, as specified in Supplementary Data 15.

Supporting gene and lipid class analyses

The B4galt tissue panel was descriptive and unadjusted for multiple comparisons (Supplementary Data 16–17). Human myocardial estimates were carried forward from prior analyses, including GSE116250.[13] The lipid-class permutation unit was the participant. No formal equivalence test or causal mediation model was fitted.

Declarations

Data availability

Source data are available from Figshare (10.6084/m9.figshare.30422413; raw reads PRJNA1266263),[7] the MetaCardis publication and its supplementary tables,[8] and GEO series GSE322803, GSE14929 and GSE116250.[11,12,13] No new participants or experimental samples were collected. Derived tables, SPT panel sequences, counts, checksums, analysis summaries and search records are supplied as Supplementary Data 1–20

Ethics

This study reanalysed public human datasets and archived animal transcriptomes; no new participants or animals were recruited or experimentally studied.

Clinical trial number

Not applicable

Author contributions

Hangyul Song: Conceptualization; Formal analysis; Visualization; Writing – original draft. Dong Jin Yu: Data Curation; Formal Analysis; Investigation; Methodology. Younghoon Go: Conceptualization; Project Administration; Supervision. Chang-Myung Oh: Conceptualization; Formal analysis; Visualization; Writing – original draft.

Competing interests

The author declares no competing interests.

Acknowledgements

This research was supported by the Bio & Medical Technology Development Program of the NRF funded by the Korean government (RS-2026-25477591), the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (Grant No. RS-2024-00439685), and the National Research Foundation of Korea (NRF) grant funded by the Korean government (Grant No. RS-2026-25505636), and by the National Research Council of Science & Technology(NST) grant by the Korea government (MSIT) (No. CAP23024-000).

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process

During the preparation of this work the author used GPT-6 for proofreading. BioRender AI assisted the initial graphical abstract generation. The resulting output was reviewed and edited by the author, who takes full responsibility for the content of the publication.

References

1. Johnson EL, et al. Sphingolipids produced by gut bacteria enter host metabolic pathways impacting ceramide levels. Nat Commun. 2020;11:2471. https://doi.org/10.1038/s41467-020-16274-w.
2. Brown EM, et al. Bacteroides-derived sphingolipids are critical for maintaining intestinal homeostasis and symbiosis. Cell Host Microbe. 2019;25:668–80. https://doi.org/10.1016/j.chom.2019.04.002. .e7.
3. Overbeek MF, et al. Plasma sphingolipids mediate the association between gut microbiome composition and type 2 diabetes risk in the HELIUS cohort: a case-cohort study. BMJ Open Diabetes Res Care. 2024;12:e004180. https://doi.org/10.1136/bmjdrc-2024-004180.
4. Chen AS-Y, et al. Association of distinct microbial and metabolic signatures with microscopic colitis. Nat Commun. 2025;16:4644. https://doi.org/10.1038/s41467-025-59566-9.
5. Sun Z et al. A gut microbiome-lipid axis in early pregnancy is associated with metabolic dysregulation and diabetes risk. *iMeta* 5, e70166 (2026). https://doi.org/10.1002/imt2.70166

6. Mamic P, et al. Gut microbiome shifts in chronic systolic heart failure are associated with disease severity and clinical improvement. Nat Cardiovasc Res. 2026. https://doi.org/10.1038/s44161-026-00854-y.
7. Mamic P, Shi H, Huang KC, Snyder MP. Systolic heart failure, gut microbiome, multi-omics dataset. Figshare. 2025. https://doi.org/10.6084/m9.figshare.30422413.v1.
8. Fromentin S, et al. *Microbiome* and metabolome features of the cardiometabolic disease spectrum. Nat Med. 2022;28:303–14. https://doi.org/10.1038/s41591-022-01688-4.
9. Direito SOL, Zaura E, Little M, Ehrenfreund P, Roling WF. M. Systematic evaluation of bias in microbial community profiles induced by whole genome amplification. Environ Microbiol. 2014;16:643–57. https://doi.org/10.1111/1462-2920.12365.
10. Mundlak Y. On the pooling of time series and cross section data. Econometrica. 1978;46:69–85. https://doi.org/10.2307/1913646.
11. Huang L, et al. Gut microbiota dysbiosis remodels the multi-tissue transcriptional landscape of the host. Microbiome. 2026. https://doi.org/10.1186/s40168-026-02495-6.
12. Crawford PA et al. Regulation of myocardial ketone body metabolism by the gut microbiota during nutrient deprivation. *Proc. Natl Acad. Sci. USA* 106, 11276–11281 (2009). https://doi.org/10.1073/pnas.0902366106
13. Sweet ME, et al. Transcriptome analysis of human heart failure reveals dysregulated cell adhesion in dilated cardiomyopathy and activated immune pathways in ischemic heart failure. BMC Genomics. 2018;19:812. https://doi.org/10.1186/s12864-018-5213-9.
14. Winkler AM, Ridgway GR, Webster MA, Smith SM, Nichols TE. Permutation inference for the general linear model. NeuroImage. 2014;92:381–97. https://doi.org/10.1016/j.neuroimage.2014.01.060.
15. Camacho C, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421. https://doi.org/10.1186/1471-2105-10-421.
16. von Elm E, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Bull World Health Organ. 2007;85:867–72. https://doi.org/10.2471/BLT.07.045120.