

Paired plasma cfRNA and fecal metagenomics jointly resolve the two-compartment bacterial landscape of SCZ

Yudan Zhang

First Affiliated Hospital of Xi'an Jiaotong University

Chun'e Li

First Affiliated Hospital of Xi'an Jiaotong University

Yiyang Wang

First Affiliated Hospital of Xi'an Jiaotong University

Wen Lu

First Affiliated Hospital of Xi'an Jiaotong University

Huiyao Zhao

Shaan Probiomicros Co.,Ltd

Ruina Liu

First Affiliated Hospital of Xi'an Jiaotong University

Cailing Wei

First Affiliated Hospital of Xi'an Jiaotong University

Pu Lei

First Affiliated Hospital of Xi'an Jiaotong University

Yuan Gao

First Affiliated Hospital of Xi'an Jiaotong University

Yanmei Ju

Shaan Probiomicros Co.,Ltd

Ruijin Guo

First Affiliated Hospital of Xi'an Jiaotong University

Xueqin Song

First Affiliated Hospital of Zhengzhou University

Wei Wang

First Affiliated Hospital of Xi'an Jiaotong University

Xiancang Ma

First Affiliated Hospital of Xi'an Jiaotong University

Feng Zhu

zhufeng1982@xjtu.edu.cn

Article

Keywords: Schizophrenia, cell-free RNA, gut microbiome, metagenomics, biomarker, liquid-biopsy

Posted Date: October 13th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10719072/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

Blood-based microbial signals are increasingly investigated as liquid-biopsy biomarkers, yet their relevance to psychiatric disorders remains unestablished. Here, we profiled plasma cell-free RNA (cfRNA) and paired fecal microbiome from 89 individuals with schizophrenia (SCZ) and 89 healthy controls (HCs) recruited from the same parent cohort and constructed a two-compartment bacterial landscape. Plasma microbial alpha diversity remained unchanged; beta-diversity exhibited modest yet statistically significant variation between groups, and twelve bacterial species exhibited differential abundance. In contrast, fecal communities showed extensive restructuring. Although group-level plasma-fecal taxonomic concordance was significantly stronger in SCZ than in HC, within-individual correlations did not differ by diagnosis. At the functional level, plasma cfRNA-derived enzyme abundances were significantly associated with symptom severity. Integrating plasma cfRNA features with fecal metagenomic features modestly improved internal exploratory classification performance (AUC = 0.976). Collectively, our results demonstrate that plasma bacterial cfRNA captures a non-redundant, symptom-associated microbial signal in SCZ and supports systemic microecological reorganization in the disorder.

Introduction

Schizophrenia (SCZ) is a heterogeneous psychiatric disorder in which genetic susceptibility interacts with environmental, immune, and metabolic factors¹. While classic psychiatric genetics, from landmark twin studies to large-scale genome-wide association studies, has robustly established SCZ's high heritability (60–80%) and polygenic architecture, its etiological “dark matter” resides in the dynamic interplay between genetic vulnerability and environmental exposures. Crucially, although host genetics confers substantial disease risk, the human microbiome, increasingly recognized as a “second genome”, plays an indispensable role in maintaining systemic homeostasis and driving disease pathogenesis². The microbiota-gut-brain axis offers a mechanistic bridge in this context, translating environmental signals into systemic pathophysiological cascades and thus connecting genetic predisposition with phenotypic manifestation in SCZ³. Moreover, the microbiota-gut-brain axis links peripheral physiology to central nervous system function via immune, endocrine, metabolic, and neural pathways^{1,3}. Over the past decade, gut microbiome studies in SCZ have reported alterations in community structure, species abundance, and functional potential^{4–7}. More importantly, experimental transfer of SCZ-associated microbial communities into germ-free or antibiotic-treated rodents has recapitulated neurochemical and behavioral phenotypes relevant to the disorder^{8–10}. These findings collectively establish the gut microbiome not merely as a disease correlate but as a functionally active modulator of SCZ pathophysiology.

Despite these advances, the field faces a persistent reproducibility challenge. Reported taxonomic shifts vary substantially across cohorts, with inconsistent directionality for key genera, including *Bifidobacterium*, *Lactobacillus*, and *Roseburia*, and no consensus on alpha diversity changes^{10–12}. This

heterogeneity likely arises from differences in geography, diet, antipsychotic medication, metabolic comorbidities, and bioinformatic pipelines^{12,13}. Although such challenges cannot be fully resolved by a single-site study, a standardized design can improve internal comparability and enhance the robustness of results when comparing plasma and fecal compartments. This yields two underappreciated implications. First, given modest effect sizes and pervasive confounding, single-cohort fecal metagenomic studies remain limited in their capacity to distinguish disease-related microbial signals from environmentally driven variation. Second, and more fundamentally, fecal metagenomics characterizes the intestinal microbial reservoir but offers only an indirect proxy for the molecular signals that actually engage host systems. What host immune and metabolic systems encounter is not the fecal community itself, but the subset of microbial molecules that may be released from or interact with the intestinal barrier and become systemically detectable¹⁴. This conceptual distinction, between the resident gut microbiota and the host-exposed circulating microbial signal, remains largely unexplored in psychiatric microbiome research.

Evidence of gastrointestinal inflammation, increased intestinal permeability, and elevated serological markers of bacterial translocation in SCZ^{16–20} motivates direct interrogation of circulating microbial molecules. Soluble CD14 and lipopolysaccharide-binding protein are elevated in subsets of patients¹⁸, and plasma IgM/IgA reactivity against commensal gut bacteria correlates with negative symptoms and the deficit subtype^{16,19}. While these immunological surrogates indicate systemic exposure to microbial components, they lack taxonomic and functional resolution: they cannot identify which organisms contributed the translocated material or what molecular functions were carried. A direct, sequence-based assay of circulating microbial nucleic acids is therefore essential.

Plasma cell-free RNA (cfRNA) is a dynamic liquid-biopsy analyte composed of short extracellular RNA fragments released from diverse tissues and cell types^{20,21}. Its abundance reflects integrated biological processes, including transcription, release, extracellular stability, and clearance, and tissue-of-origin deconvolution of human cfRNA has enabled non-invasive monitoring of organ health²². Microbial cfRNA can be recovered from plasma^{23,24} and may capture a circulating layer of host-exposed microbial signals that is mechanistically distinct from fecal microbial DNA. Unlike fecal DNA which reflects standing biomass and functional potential, plasma microbial cfRNA captures circulating microbial RNA fragments that may reflect microbial release, translocation, persistence, and clearance. This distinction is biologically consequential: it implies that plasma and fecal measurements are complementary, not redundant, sources of information about host-microbe interactions. A prior whole-blood RNA-seq study reported altered microbial read profiles in SCZ²⁵; however, whole blood contains abundant cellular RNA and was not designed to isolate plasma cfRNA or link it to paired fecal metagenomes. Plasma microbial profiling poses substantial technical challenges: microbial cfRNA reads are exceedingly sparse, and reagent- or environment-derived contaminants can dominate low-biomass specimens. Rigorous contamination-control procedures, including extraction blanks, dual-indexed library preparation, orthogonal taxonomic classification, and blank-control-based contaminant filtering, should be implemented before biological interpretation.^{19–24}.

Here, we integrated plasma cfRNA sequencing with paired fecal shotgun metagenomic sequencing from the same individuals to reconstruct the two-compartment bacterial landscape in SCZ. We sought to: (i) identify plasma bacterial cfRNA features differentially abundant in SCZ versus healthy controls (HCs); (ii) characterize the relationship between plasma and fecal microbial profiles—including cross-compartment taxon sharing and group-level concordance; (iii) evaluate associations of plasma taxonomic and enzyme-level features with Positive and Negative Syndrome Scale (PANSS) subscale scores; and (iv) assess whether plasma cfRNA features improve diagnostic classification performance when combined with fecal metagenomic features. By integrating these two biologically distinct compartments within the same individuals, we move beyond single-modality description toward a systems-level understanding of host-exposed microbial signals in SCZ. By jointly profiling paired fecal and plasma samples from the same individuals, we can distinguish the intestinal microbial reservoir from circulating microbial signals, thereby addressing a key limitation of fecal-only studies.

Methods

Study design and participants

This cross-sectional study enrolled 178 participants drawn from an ongoing parent cohort: 89 individuals diagnosed with SCZ and 89 HCs. All participants provided paired fecal and plasma samples for shotgun metagenomic sequencing and cfRNA sequencing, respectively. SCZ diagnoses were established by board-certified psychiatrists using DSM-5 criteria. HCs were recruited from the same geographic region and underwent structured clinical interviews to confirm absence of current or lifetime psychiatric disorders. Exclusion criteria included: (1) acute infection or systemic inflammatory disease; (2) autoimmune disease; (3) malignancy; (4) severe or unstable neurological or medical conditions; (5) substance use disorder (including alcohol dependence); (6) pregnancy or lactation; and (7) any condition likely to impair reliability of clinical or cognitive assessment.

The study protocol was approved by the Ethics Committee of Xi'an Jiaotong University First Affiliated Hospital (No. XJTU1AF2025LSYY-368) and conducted in full compliance with the Declaration of Helsinki. Written informed consent was obtained from all participants; for those lacking decision-making capacity, consent was obtained from their legally authorized representatives.

Plasma collection, cfRNA extraction and sequencing

Peripheral venous blood was collected between 7:00–9:00 a.m. following an overnight fast (8–10 h). Approximately 5 mL was drawn into K₂EDTA-coated tubes and processed within 60 min at 4°C. Plasma was isolated via single centrifugation at 800 × g for 10 min at 4°C; supernatant was carefully aspirated without disturbing the buffy coat, aliquoted into RNase-free tubes, and stored at -80°C until processing.

cfRNA was extracted from 200 μ L plasma using the RNeasy Circulating Reagent Kit (Qiagen, 12321), per manufacturer instructions. RNA concentration and integrity were quantified using the

Qubit RNA High Sensitivity Assay Kit and assessed on a Q-seq 100 Analyzer (Qsep100 system). Library preparation was performed with the Fast RNA-seq Library Preparation Kit V2 (ABclonal, RK20306), incorporating human rRNA depletion. Sequencing was conducted on the MGIEQ-2000RS platform generating 150-bp paired-end reads.

To determine optimal sequencing depth, ten representative libraries (five SCZ, five HC) were sequenced to target depths of ~ 10, ~20, and ~ 30 Gb (Fig. 1a). Gene and transcript detection increased sharply below 10 Gb and plateaued thereafter, with diminishing returns beyond this threshold. Accordingly, 10 Gb per sample was adopted for the primary cfRNA profiling analysis.

Human-read depletion and plasma microbial profiling

Quality-trimmed reads were aligned to the GRCh38 human reference genome using HISAT2³² v2.2.1. Non-human-mapping read pairs were retained for downstream microbial analysis. Species-level taxonomic profiling was performed using MetaPhlAn v4.2.4³³ with the mpa_vJan25_CHOCOPhIAnSGB_202503 database. As an orthogonal validation method, Kraken2 v2.1.2³⁴ was applied using the PlusPF database under stringent parameters: --confidence 0.20, --minimum-hit-groups 3, and --minimum-base-quality 20. Kraken2 species counts were normalized to counts per million (CPM) of post-quality-control read pairs. MetaPhlAn relative abundances constituted the primary taxonomic table; Kraken2 outputs were used exclusively for contamination assessment, bacterial burden estimation, and independent taxonomic verification.

Process controls and contaminant identification

A total of 27 process controls were co-processed alongside the 178 biological samples: nine extraction blanks, nine skin-associated controls (swabs from operator hands and collection surfaces), and nine pure-culture controls (three replicates each of *Phocaeicola vulgatus*, *Bacteroides fragilis*, and *Megamonas funiformis*). Extraction blanks served as the primary basis for formal contaminant filtering. Skin-associated controls enabled identification of ambient skin microbiota background. Pure-culture controls verified expected-species recovery, assessed cross-classification fidelity, and monitored potential carryover, none were applied as automatic exclusion rules.

All 205 profiles (178 biological + 27 controls) were analyzed jointly with MetaPhlAn and Kraken2. Species-level classifications from the two pipelines were harmonized using the corresponding NCBI species TaxID before assessing concordant detection across classifiers. A species was designated a high-confidence blank contaminant if it met all four criteria: (i) detected by both classifiers in ≥ 2 extraction blanks; (ii) concordantly detected by both classifiers in ≥ 2 of the same blanks; (iii) mean Kraken2 CPM in blanks ≥ 2 -fold higher than in biological samples; and (iv) significantly enriched in blanks versus biological samples (one-sided Wilcoxon rank-sum test, Benjamini-Hochberg-adjusted $q < 0.05$). Of these, 47 corresponded to MetaPhlAn species-level features present in the biological-sample matrix and

were removed, reducing the matrix from 610 to 563 features. The remaining relative abundances were then re-closed to 100% within each sample. The complete 563-feature matrix was used for alpha-diversity, beta-diversity and taxonomic-composition analyses. For differential-abundance and downstream taxon-level analyses, 109 species meeting the prespecified criterion of relative abundance $\geq 0.01\%$ in at least 18 of the 178 plasma samples were retained.

Microbial functional profiling and contaminant adjustment

Functional profiles were inferred from human-unmapped plasma cfRNA reads using HUMAnN v4.0.0.alpha.1³⁵. Gene-family abundances were aggregated to Enzyme Commission (EC) numbers and Kyoto Encyclopedia of Genes and Genomes Orthology (KO) identifiers and expressed as reads per kilobase per million mapped reads (RPKM). Rows taxon-stratified to the 54 high-confidence blank contaminants were excluded. For each functional feature and sample, contaminant-adjusted unstratified abundance was computed by subtracting the summed abundance assigned to contaminant-stratified rows from the original unstratified abundance; negative values were set to zero. This procedure eliminated 1,343 contaminant-stratified KO rows and 2,665 EC rows, resulting in adjusted unstratified matrices comprising 4,495 KO and 2,996 EC features across all 178 biological samples.

Fecal collection and shotgun metagenomic profiling

Fecal samples were immediately frozen at -80°C upon collection and stored until DNA extraction. Genomic DNA was isolated using the QIAamp DNA Stool Mini Kit (Qiagen), with modifications to maximize yield and minimize bias. Shotgun metagenomic sequencing (150-bp paired-end) was performed on the MGISEQ-T7 platform. Taxonomic profiling was conducted using MetaPhlAn v4.2.4 with the identical mpa_vJan25_CHOCOPhIAnSGB_202503 database employed for plasma analysis. Species with relative abundance $\geq 0.01\%$ in ≥ 18 participants were retained, yielding 294 fecal species.

Microbial diversity and differential abundance

All primary downstream analyses were adjusted for age, sex, and BMI. These covariates were included in the PERMANOVA, differential-abundance, taxon–PANSS association, and relevant correlation analyses. Alpha diversity was quantified using Shannon and Simpson indices and compared between SCZ and HC groups using Wilcoxon rank-sum tests. beta diversity was assessed via Bray-Curtis dissimilarity and visualized using principal-coordinate analysis (PCoA). Group-level differences in overall community composition were tested using `vegan::adonis2` (9,999 permutations) with the formula `distance ~ Group + age + sex + BMI` and marginal tests (`by = "margin"`). The effect of Group was evaluated while accounting for age, sex, and BMI. Differential species abundance was evaluated using `MaAsLin2`³⁶, with age, sex, and BMI included as covariates. Benjamini–Hochberg correction was applied separately within the plasma and fecal taxonomic analyses, across the tested species within each compartment. An

exploratory threshold of $q < 0.10$ was applied to plasma cfRNA results, whereas $q < 0.05$ was used for fecal metagenomic results.

Cross-compartment taxonomic correspondence

Species passing compartment-specific abundance filters ($\geq 0.01\%$ in ≥ 18 samples per compartment) were matched by exact MetaPhlAn taxon name, identifying 86 taxa jointly detected in both plasma and fecal profiles. For each participant, Spearman's rank correlation was computed between paired plasma and fecal relative-abundance vectors across these 86 shared species; group-wise correlation distributions were compared using Wilcoxon rank-sum tests. To assess whether observed plasma-fecal correlations exceeded chance expectation, sample pairing labels were randomly permuted 10,000 times within each diagnostic group. Paired co-detection rates and compartment-specific median abundances were summarized for the three differentially abundant species identified in both compartments.

For taxon-to-taxon co-variation analysis, the 12 plasma and 81 fecal differentially abundant species were subjected to centered-log-ratio (CLR) transformation and residualized against age, sex, and BMI. Spearman correlations were then calculated for all 972 possible plasma-fecal species pairs. Multiple testing correction was applied across the entire set using the Benjamini-Hochberg procedure.

Clinical assessment and microbial feature-PANSS associations

Psychopathology severity in SCZ participants was assessed using the 30-item Positive and Negative Syndrome Scale (PANSS³⁷). Each item was rated on a 1–7 Likert scale; Positive, Negative, General Psychopathology, and Total subscale scores were computed per standard conventions. Subscale scores were only calculated when all constituent items were present and scored within the valid range (1–7). Final complete-case analytic datasets comprised 51 participants for the Positive subscale, 50 for the Negative subscale, and 48 each for the General Psychopathology and Total subscales.

For taxonomic analyses, relative abundances were converted to proportions, arcsine-square-root transformed, and z-standardized. Effect estimates presented in Fig. 4 derive from linear regression models with HC3 heteroscedasticity-robust standard errors, adjusted for age, sex, and BMI. Multiple testing correction was applied separately across plasma (48 tests) and fecal (324 tests) taxon-PANSS associations; $q < 0.05$ was considered statistically significant and $q < 0.10$ exploratory.

EC features detected in $\geq 10\%$ of samples underwent total-sum scaling followed by $\log_2$ transformation. MaAsLin3³⁸ models included age, sex, and BMI as covariates. Benjamini-Hochberg correction was performed independently within each PANSS outcome (Positive, Negative, General Psychopathology, Total) across all tested EC features; $q < 0.10$ denoted exploratory associations. KO-level analyses followed the identical analytical workflow.

SCZ-HC classification analysis

Three complementary feature modalities were evaluated: (1) plasma bacterial cfRNA species abundances; (2) fecal metagenomic species abundances; and (3) a combined matrix formed by participant-wise concatenation of the two. For the combined analysis, plasma and fecal features were concatenated for each participant. Logistic Lasso regression with fivefold cross-validation was then applied to the combined matrix, and features with non-zero coefficients at $\lambda_{\min}$ were retained. Sixteen machine learning classifiers were benchmarked using the R caret framework: ridge regression, partial least-squares discriminant analysis (PLS-DA), elastic net, multilayer perceptron neural network, radial and linear support vector machines, linear discriminant analysis, Lasso, random forest, quadratic discriminant analysis, gradient boosting machine, naïve Bayes, k-nearest neighbors, C5.0 decision tree, logistic regression, and classification tree. Model performance was evaluated using repeated fivefold cross-validation (three repetitions), with metrics including area under the receiver-operating-characteristic curve (AUC), accuracy, sensitivity, and specificity. Plasma or fecal species consistently ranked among the top five most important features in ≥ 2 classifiers across the combined-view analysis were defined as cross-model consensus features.

Results

A contamination-aware plasma bacterial cfRNA landscape in SCZ

We enrolled 178 participants, 89 individuals with SCZ and 89 HCs, from whom paired plasma and fecal samples were collected. Demographic characteristics of all participants and disease duration in the SCZ group are summarized in Table 1; no statistically significant differences were observed between groups in age, sex, or BMI. In the pilot phase, saturation analysis of ten representative plasma cfRNA libraries revealed that gene and transcript detection reached a plateau at approximately 10 Gb of sequencing data (Fig. 1a). Accordingly, for all participants, we generated an average of ~ 11.38 Gb of high-quality (clean) total cfRNA reads per sample (median: 11.04 Gb; IQR: 9.31–13.01 Gb), including ~ 2.50 Gb of bacterial-derived cfRNA reads (median: 2.34 Gb; IQR: 1.74–3.09 Gb) (Supplementary Table 1). Neither total cfRNA nor bacterial-derived cfRNA sequencing depth showed significant variation by diagnosis (SCZ vs. HC), sex, age, or BMI (all $P > 0.05$; Supplementary Table 2).

Table 1

Characteristic	Total	HC	SCZ	P value
N	178	89	89	-
Age, years	35.43 ± 10.63	36.62 ± 10.34	34.25 ± 10.85	0.09
Range	18–64	19–58	18–64	-
Gender, Male, N (%)	126 (70.8%)	65 (73.0%)	61 (68.5%)	0.621
Female, N (%)	52 (29.2%)	24 (27.0%)	28 (31.5%)	-
BMI, kg/m ²	24.21 ± 3.86	24.45 ± 3.87	23.97 ± 3.86	0.437
Range	15.0-33.3	18.3–32.9	15.0-33.3	-
Disease duration, years	-	-	5.56 ± 2.23	-
Range	-	-	0–9	-
Data are presented as mean ± SD for continuous variables and N (%) for categorical variables. <i>P</i> values were calculated using independent t-test or Mann-Whitney U test for continuous variables, and Chi-square test or Fisher's exact test for categorical variables. Normality was assessed by the Shapiro-Wilk test.				

After removal of human-mapped reads and bioinformatic filtering of contaminants, Kraken-estimated bacterial read fractions were numerically higher in SCZ plasma than in HC plasma, but fell just short of conventional statistical significance (Wilcoxon rank-sum test, $P = 0.051$; Fig. 1b). MetaPhlAn identified 610 bacterial species across all plasma samples; of these, 47 species failed to meet our pre-specified high-confidence criteria for excluding blank-derived contamination and were therefore excluded, leaving 563 species for compositional analysis (Supplementary Fig. 1). Sex, age, and BMI exerted no statistically significant effect on overall microbial community structure (adonis, all $P > 0.05$; Supplementary Table 3). Alpha diversity, as measured by both Shannon and Simpson indices, did not differ significantly between SCZ patients and HCs (both $P > 0.05$; Fig. 1c). Likewise, the top 20 most abundant bacterial species exhibited highly similar relative abundance profiles across groups (Fig. 1d). Bray-Curtis-based PCoA revealed substantial inter-individual overlap; after adjusting for covariates, diagnosis (SCZ vs. HC) explained only a small proportion of community variation ($R^2 = 0.0141$, PERMANOVA $P = 0.0316$; Fig. 1e).

For differential taxon analysis, we prioritized bacterial taxa demonstrating both high detection confidence and sufficient prevalence. Specifically, we excluded species with mean relative abundance below 0.01% and detection in fewer than 10% of participants (i.e., < 18 individuals), resulting in a final set of 109 species for abundance-differential species (ADS) identification. Using MaAsLin2 with adjustment for age, sex, and BMI, we identified 12 bacterial species significantly associated with SCZ at an exploratory FDR threshold of $q < 0.10$: *Flavonifractor plautii*, *Aeribacillus pallidus*, *Ruminococcus bromii*, *Bacteroides intestinalis*, *Pilosibacter fragilis*, *Bifidobacterium longum*, *Blautia obeum*, *Clostridiaceae bacterium*, *Evtepia gabavorous*, *Sphingomonas sp. 3F27F9*, *Streptococcus salivarius*, and *Paeniclostridium*

hominis (Fig. 1f). Collectively, these results define a contamination-aware, species-resolved plasma bacterial cfRNA signature in SCZ, despite limited global separation between diagnostic groups.

Extensive compositional restructuring of the fecal microbiome in SCZ

To validate gut dysbiosis in SCZ and investigate its association with the plasma microbiome as inferred from cfRNA, we performed shotgun metagenomic profiling of fecal samples paired with plasma from 178 participants. Consistent with our findings in plasma cfRNA, no significant difference in alpha diversity was observed between SCZ patients and HCs ($P > 0.05$; Fig. 2a). In contrast, Bray-Curtis-based PCoA revealed a highly significant diagnostic-group effect on overall community structure (PCo1: 13.53%, PCo2: 10.06%; PERMANOVA $P = 0.0001$; Fig. 2b, Supplementary Table 4). The most abundant species including *Segatella copri*, *Phocaeicola vulgatus*, *Megamonas funiformis*, *Faecalibacterium prausnitzii*, *Ruminococcus bromii*, and multiple *Bacteroides spp.* were broadly shared across both groups (Fig. 2c).

Of the 294 species detected, 52 were significantly enriched and 29 significantly depleted in SCZ relative to HCs (FDR correction, $q < 0.05$), whereas 213 showed no differential abundance (Fig. 2d). SCZ-enriched taxa included *Coprococcus phoceensis*, *Fusobacterium mortiferum*, *Sarcina ventriculi*, *Eisenbergiella tayi*, and *Clostridium saudiense*; SCZ-depleted taxa included *Streptococcus salivarius*, *Blautia wexlerae*, *Faecalibacillus faecis*, and *Fusicatenibacter saccharivorans*. Collectively, these results indicate that fecal microbial communities undergo more pronounced, species-level restructuring in SCZ than the plasma cfRNA-associated microbial signal.

Taxonomic overlap and enhanced group-level concordance between plasma and fecal microbiota in SCZ

We next performed intersection analysis between plasma and fecal bacterial communities. Among the 294 species detected in fecal samples and the 109 species with high detection confidence and sufficient prevalence in plasma, 86 species were shared across both compartments. Of these shared taxa, *Blautia obeum*, *Streptococcus salivarius*, and *Evtepia gabavorous* showed statistically significant abundance differences between SCZ patients and HCs in both fecal and plasma compartments. *Blautia obeum* and *Streptococcus salivarius* were depleted in feces but increased in plasma cfRNA, whereas *Evtepia gabavorous* was increased in both compartments. (Fig. 3a; Supplementary Tables 5 and 6). Consistently, co-detection rates and relative abundances of these three taxa also differed significantly between SCZ and HCs (Fig. 3b).

Participant-level plasma-fecal correlations across the 86 shared species were significantly stronger in SCZ than in HCs (mean Spearman's $\rho = 0.352$ vs. 0.302 ; Wilcoxon $P = 0.0196$; Fig. 3c), suggesting enhanced cross-compartment coordination of microbial signals in SCZ. However, correctly paired plasma-fecal sample pairs did not yield higher correlation than randomly permuted within-group pairs (HC permutation $P = 0.796$; SCZ $P = 0.132$; Fig. 3d), indicating that the observed association reflects

group-level patterns rather than individual-level concordance. Collectively, these findings support enhanced group-level-rather than individual-level concordance-between plasma cfRNA and fecal microbiota in SCZ.

To further probe functional crosstalk between ADS in plasma and feces, we conducted correlation analysis between 12 plasma ADSs and 81 fecal ADSs. Among the 972 plasma-fecal taxon pairs tested, several associations reached nominal significance after covariate adjustment; however, none survived Benjamini-Hochberg FDR correction (minimum $q = 0.286$; Fig. 3e). Together, the taxonomic overlap and group-level correlation results indicate that plasma bacterial cfRNA reflects a biologically relevant, yet non-redundant signal relative to fecal microbial composition.

Microbial taxa and plasma cfRNA enzyme functions show distinct associations with PANSS dimensions

To assess the clinical relevance of microbial signals in plasma and feces, we conducted correlation analyses between PANSS subscale scores and (▢) plasma microbial taxon abundances, (▢) plasma-derived Enzyme Commission (EC) reaction abundances, and (▢) fecal microbial taxon abundances. Among plasma taxon-PANSS associations, *Blautia obeum* exhibited the strongest link: a 1-SD increase in its log-transformed plasma abundance was associated with a 3.68-point higher Positive subscale score (HC3-robust $P = 0.0021$; Fig. 4a, Supplementary Table 7). Its FDR-adjusted q value was 0.1009, slightly above the prespecified exploratory threshold of 0.10. Thus, no plasma taxon-PANSS association reached statistical significance. Several additional associations attained nominal significance ($P < 0.05$) but failed to survive family-wise error correction.

One FDR-significant association existed among 324 tests of fecal taxon-PANSS association: *Veillonella rogosae* abundance was inversely associated with General Psychopathology score ($P = 0.00011$; $q = 0.0358$; Fig. 4b, Supplementary Table 8). However, this species was detected in only 8 of 48 participants (16.7%), limiting the robustness and generalizability of this finding. *Anaerostipes hadrus* also showed a substantial inverse association with General Psychopathology ($\beta = -3.89$, $P = 0.00123$), though it did not survive FDR correction ($q = 0.194$). Overall, both plasma and fecal microbial taxa showed weak correlations with PANSS symptoms.

In contrast to microbial taxa, exploratory functional profiling on plasma microbial cfRNA identified 25 unique Enzyme Commission (EC) reactions that contributed to 40 significant abundance-component associations with PANSS subscale scores at $q < 0.10$: 14 with PANSS Positive, 6 with General Psychopathology, and 20 with Total score; no EC reaction reached this threshold for the Negative subscale (Supplementary Fig. 2). Of the 40 associations, 37 exhibited negative regression coefficients. Recurrently negative coefficients across Positive, General Psychopathology, and Total scores included acetaldehyde dehydrogenase (acetylating; EC 1.2.1.10), glyceraldehyde-3-phosphate dehydrogenase (NADP⁺; EC 1.2.1.12), and pyruvate dehydrogenase (EC 1.2.4.1). Glutamate decarboxylase (EC 4.1.1.15) was negatively associated with Total score ($q = 0.0091$; Supplementary Table 9). Collectively, the consistent directionality and recurrence of negative EC coefficients suggest that greater severity of

positive and overall symptoms is associated with reduced representation of specific enzyme-level microbial cfRNA annotations in plasma-defining a preliminary, symptom-associated functional signature for future validation.

Integrating plasma cfRNA with fecal features improves discriminatory model performance

Across 16 machine learning classifiers, fecal-only models consistently outperformed plasma microbial cfRNA-only models (Fig. 5a). The best-performing plasma-only model was ridge regression (AUC = 0.873), while the top fecal-only model was PLS-DA (AUC = 0.964; accuracy = 0.901). Integration of plasma and fecal features improved AUC over plasma-only models for all 16 classifiers, with gains ranging from + 0.021 to + 0.207. The optimal combined model was ridge regression, with an exploratory internal AUC of 0.976 (accuracy = 0.904; sensitivity = 0.902; specificity = 0.907). Compared with fecal-only performance, the combined model yielded higher AUCs for 14 of 16 classifiers (Supplementary Table 10); only random forest and gradient boosting showed marginal decreases. Because feature selection was performed before the repeated cross-validation used for final classifier evaluation, these metrics may be optimistically biased and should be interpreted as internal, exploratory, proof-of-concept performance rather than externally validated diagnostic accuracy.

Twelve species ranked among the top five most important features in at least two classifiers under the combined-view framework (Fig. 5b). Ten were fecal-derived: *Blautia wexlerae* (nine models), *Fusicatenibacter saccharivorans* (eight), *Streptococcus salivarius* (seven), *Anthropogastromicrobium aceti* (six), *Hominilimicola fabiformis* (six), *Bifidobacterium pseudocatenulatum* (five), *GGB35068_SGB47850* (four), *Mesosutterella multiformis* (four), *Dorea hominis* (two), and *Dorea sp. AF36_15AT* (two). Two plasma cfRNA species recurred across models: *Aeribacillus pallidus* (six) and *Flavonifractor plautii* (four) (Supplementary Table 11). This predominantly fecal feature architecture, together with recurrent plasma contributions and consistent AUC improvements in combined models, supports a modest but non-redundant role for plasma bacterial cfRNA in distinguishing SCZ from HCs.

Discussion

In this study, we constructed the first contamination-aware, species-resolved plasma bacterial cfRNA landscape in SCZ and firstly demonstrated plasma-fecal microbial concordance was enhanced in SCZ at the group level but not at the individual level. Moreover, our data indicate that enzyme-level functional annotations of plasma microbial cfRNA, most notably glutamate decarboxylase tracked symptom severity, and integrating plasma features with fecal features modestly improved classification performance and fulfilled a strong SCZ discrimination (AUC 0.976). Unlike the notable gut microbiome reshaping in SCZ reported here and our previous study⁹, plasma bacterial cfRNAs showed modest alterations in SCZ, as diagnosis explained only 1.4% of microbial community variance, and just twelve species differed at an exploratory significance threshold ($q < 0.1$) between groups. Together, these

results position plasma bacterial cfRNA as a non-redundant, complementary microbial signal in SCZ rather than a surrogate of the gut microbiome.

After stringent removal of human reads and blank-derived contaminants, bacterial cfRNA was reliably detected and taxonomically resolved in plasma. Although cfRNA-sequencing has been applied to profile microbial communities in blood^{39,40}, interpretation of microbial signals in this compartment requires particular caution due to the unresolved question of whether detected microbial signals reflect a resident blood microbiota or transient translocation (atopobiosis e.g., from gut or oral mucosa)⁴¹. A population-scale analysis of 9,770 healthy individuals found no evidence for a shared blood microbiome, and microbial DNA was undetectable in 84% of samples³⁰. In light of this controversy, our analytical pipeline adopted a conservative design: (i) blank-control-informed filtering removed 47 candidate contaminant species; (ii) sequencing depth and global community dissimilarities was balanced across diagnostic groups. The residual cfRNA signal we report therefore reflects circulating microbial RNA fragments, likely shed from mucosal reservoirs and cleared via the bloodstream, rather than evidence of a viable, replicating blood microbiome. This interpretation is technically supported by prior validation of microbial cell-free DNA as a biomarker in infectious disease diagnostics⁴². Our findings extend this “liquid biopsy” framework to SCZ, representing the first application at species-level resolution in a psychiatric context.

The enhanced group-level concordance between plasma and fecal microbial communities in SCZ warrants cautious interpretation. As correctly paired plasma-fecal sample correlations did not exceed those observed in permuted pairs, our data do not support frequent, individually detectable gut-to-blood translocation. Instead, the pattern is consistent with a group-level reorganization of the gut-blood microbial axis, such as population-wide alterations in mucosal barrier integrity that elevate the baseline flux of microbial material into systemic circulation. This interpretation aligns with prior serological evidence of bacterial translocation in SCZ, including elevated plasma levels of soluble CD14 and lipopolysaccharide-binding protein¹⁸, as well as increased IgM and IgA antibody responses to commensal gut bacteria¹⁹, responses that were associated with negative symptoms and the deficit subtype. The underlying mechanism linking bacterial translocation and SCZ pathology is that the translocation of microbially driven enzymatic substrates and immunogenic components across a compromised epithelial barrier fuels peripheral low-grade inflammation, ultimately shaping SCZ-associated neural circuitry. Plasma microbial cfRNA may therefore reflect a molecular signature of these barrier abnormalities and could plausibly underlie the chronic, low-grade innate immune activation repeatedly documented in SCZ^{14,43}.

The shared differentially abundant taxa between fecal and plasma samples converged on three species: *Blautia obeum*, *Streptococcus salivarius*, and *Evtepiea gabavorous*. The bidirectional abundance pattern observed for *Blautia obeum* and *Streptococcus salivarius*, depletion in feces coupled with elevation in plasma in SCZ versus HC, suggests compartment-specific ecological dynamics rather than passive spillover from gut to blood. *Evtepiea gabavorous* is the only species exhibiting cross-compartment consistency in the direction of alteration, elevated in both fecal and plasma samples from SCZ patients,

is therefore prioritized for functional validation. Functional annotations provided the most direct mechanistic link of this species to SCZ neurobiology. Glutamate decarboxylase (EC 4.1.1.15), the enzyme catalyzing GABA synthesis from glutamate, was inversely associated with total PANSS symptom burden; moreover, the majority of enzyme-symptom associations were negative, suggesting that reduced abundance of microbial metabolic functions correlates with greater illness severity. This observation is consistent with animal studies showing that the SCZ gut microbiome modulates the glutamate-glutamine-GABA cycle and induces SCZ-relevant behavioral deficits in germ-free recipient mice¹⁰, and with evidence that specific gut bacteria can directly synthesize or catabolize GABA⁴⁴. The presence of *Evtepia gabavorous*, a GABA-catabolizing bacterium with experimentally validated activity⁴⁴, among the taxa altered in both compartments raises a testable hypothesis: that microbial GABA consumption contributes to systemic GABA dysregulation in SCZ. Supporting the causal plausibility of microbe-to-brain metabolic signaling, fecal microbiota transplantation from drug-free SCZ patients has been shown to dysregulate kynurenine pathway metabolism and induce behavioral abnormalities in colonized mice⁴⁵. We emphasize, however, that our functional signals derive from plasma cfRNA-based EC annotations at an exploratory statistical threshold and reflect predicted functional potential, rather than direct measurements of enzymatic activity or metabolite concentrations.

At the individual taxon level, clinical associations were weak. The sole fecal taxon-symptom association surviving FDR correction, *Veillonella rogosae* abundance negatively associated with General Psychopathology subscale score, was based on a low detection prevalence (16.7% of participants) and should therefore be considered hypothesis-generating. These modest taxon-level effects contrast with the comparatively stronger associations observed at the enzyme level, suggesting that microbial functional capacity may be more directly linked to symptom biology than taxonomic composition. Finally, although our classification models showed exploratory improvement when integrating plasma and fecal features, feature selection was performed before the repeated cross-validation used for final classifier evaluation rather than independently within each training fold, creating a potential source of feature-selection leakage. Therefore, the reported AUC values may be optimistic and require assessment using leakage-free nested cross-validation and independent external cohorts. Future studies aiming to assess the diagnostic utility of integrated fecal and plasma microbial biomarkers must implement leakage-free validation procedures and validate findings in independent cohorts. Because the present study was conducted at a single site, the observed microbial signatures and classification performance require validation in geographically, clinically, and technically independent cohorts.

Additional limitations warrant consideration. First, the cross-sectional design precludes causal inference and cannot determine whether observed microbial differences precede illness onset, emerge during disease progression, or reflect treatment effects. However, plasma cfRNA cannot establish direct microbial transport to the central nervous system or a causal gut-to-brain mechanism. Second, all primary downstream analyses were adjusted for age, sex, and BMI, whereas dietary characteristics were not systematically available for covariate adjustment. Detailed antipsychotic treatment information was also insufficiently standardized for quantitative adjustment. Therefore, residual confounding related to

diet and antipsychotic exposure cannot be excluded. Third, the anatomical origin of plasma microbial cfRNA remains unresolved: contributions from oral, upper respiratory, or intestinal mucosal sites cannot be distinguished based on sequence data alone. Finally, although we implemented stringent contamination controls, including four-criterion blank-control filtering and dual-library indexing, low-biomass analyses inherently carry residual risk of environmental or reagent-derived contamination; thus, our findings must be interpreted within this methodological constraint. Despite these limitations, our data provide the first evidence that plasma bacterial cfRNA captures a non-redundant, symptom-associated microbial signal in SCZ; further, they implicate plasma bacterial cfRNA in homeostatic disruption of the gut-brain axis in SCZ; and collectively, they support a model of systemic microecological reorganization, rather than patient-specific translocation events as the predominant feature of microbiome-host interaction in SCZ.

Declarations

Data availability

The raw fecal metagenomic and plasma cfRNA sequencing data have been deposited in the Genome Sequence Archive (GSA) database (<https://ngdc.cncb.ac.cn/gsa>) under accession numbers PRJCA071887 and PRJCA075935, respectively.

Code availability

All code is available on GitHub at (<https://github.com/anzhi525/scz-cfrna>).

Acknowledgements

This work was supported by grants from the National Natural Science Foundation of China (General Program, No. 82271572) and the National Key Research and Development Program of China (No. 2025YFC2511200).

Author information

Authors and Affiliations

Shaanxi Provincial Key Laboratory of Biological Psychiatry, Department of Psychiatry, The First Affiliated Hospital of Xi'an Jiaotong University, 277 Yanta West Road, Xi'an 710061, China.

Wen Lu, Ruina Liu, Cailing Wei, Yuan Gao, Ruijin Guo, Wei Wang, Xiancang Ma, Feng Zhu

Center for Translational Medicine, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710061, China.

Yudan Zhang, Yiyang Wang, Pu Lei, Feng Zhu

Center for Brain Science, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710061, China.

Chun'e Li, Yuan Gao

Shaan Probiomicros Co.,Ltd. Suzhou 215000, China.

Huiyao Zhao, Yanmei Ju,

Department of Psychiatry, the First Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan, China

Xueqin Song,

Contributions

F.Z. conceptualized the study, Y.Z. led data curation, supported by C.L, Y.W., R.L., W.L., H.Z., C.W., P.L., Y.G., Y.J., R.G., X.S. and W.W., Formal analysis was performed by Y.Z., C.L., and H.Z.. The investigation was conducted by W.L, C.W, P.L, Y.J., and R.G., Methodology development and resource management were jointly undertaken by Y.Z., H.Z., and C.L., Y.Z. and C.L. developed and maintained the software. , Validation was carried out by Y.Z., Y.W., R.L., and C.L. ,Visualization was performed by Y.Z. and H.Z. ,Y.Z. and F.Z. wrote the original draft. , Review and editing were conducted by Y.Z., F.Z., X.S., and X.M. ,Funding acquisition was led by F.Z. and Y.Z., with support from X.M. Project administration and supervision were led by X.M. and F.Z. All authors critically reviewed the manuscript and approved the final version.

Ethics declarations

Competing interests

The authors declare no competing interests.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Cryan, J. F. *et al.* The Microbiota-Gut-Brain Axis. *Physiol. Rev.* 99, 1877–2013 (2019).
2. Fan, Y. & Pedersen, O. Gut microbiota in human metabolic health and disease. *Nat. Rev. Microbiol.* 19, 55–71 (2021).
3. Morais, L. H., Schreiber, H. L. & Mazmanian, S. K. The gut microbiota–brain axis in behaviour and brain disorders. *Nat. Rev. Microbiol.* 19, 241–255 (2021).
4. Vasileva, S. S. *et al.* Associations of the Gut Microbiome With Treatment Resistance in Schizophrenia. *JAMA Psychiatry* 81, 292–302 (2024).

5. Thirion, F. *et al.* Alteration of Gut Microbiome in Patients With Schizophrenia Indicates Links Between Bacterial Tyrosine Biosynthesis and Cognitive Dysfunction. *Biol. Psychiatry Glob. Open Sci.* 3, 283–291 (2022).
6. Yuan, X. *et al.* Gut microbial biomarkers for the treatment response in first-episode, drug-naïve schizophrenia: a 24-week follow-up study. *Transl. Psychiatry* 11, 422 (2021).
7. Nguyen, T. T. *et al.* Gut microbiome in Schizophrenia: Altered functional pathways related to immune modulation and atherosclerotic risk. *Brain. Behav. Immun.* 91, 245–256 (2021).
8. Nguyen, T. T. *et al.* Differences in gut microbiome composition between persons with chronic schizophrenia and healthy comparison subjects. *Schizophr. Res.* 204, 23–29 (2019).
9. Zhu, F. *et al.* Metagenome-wide association of gut microbiome features for schizophrenia. *Nat. Commun.* 11, 1612 (2020).
10. Zheng, P. *et al.* The gut microbiome from patients with schizophrenia modulates the glutamate-glutamine-GABA cycle and schizophrenia-relevant behaviors in mice. *Sci. Adv.* 5, eaau8317 (2019).
11. Militaru, A.-M., Pietreanu, A. C., Trifu, S. & Popa, G. L. Gut Microbiota in Schizophrenia: Taxonomic Shifts, Beta- Diversity Alterations, and Biomarker Potential: A Systematic Review. *Int. J. Mol. Sci.* 27, 4606 (2026).
12. Murray, N. *et al.* Compositional and Functional Alterations in Intestinal Microbiota in Patients with Psychosis or Schizophrenia: A Systematic Review and Meta-analysis. *Schizophr. Bull.* 49, 1239–1255 (2023).
13. Neha, S. A., Hanson, J. D., Wilkinson, J. E., Bradley, R. D. & Phillips, C. D. Impacts of host phylogeny, diet, and geography on the gut microbiome of rodents. *PLOS ONE* 20, e0316101 (2025).
14. Gao, Y. *et al.* Systems biological assessment of altered cytokine responses to bacteria and fungi reveals impaired immune functionality in schizophrenia. *Mol. Psychiatry* 27, 1205–1216 (2022).
15. Braniste, V. *et al.* The gut microbiota influences blood-brain barrier permeability in mice. *Sci. Transl. Med.* 6, 263ra158 (2014).
16. Scheurink, T. A. W. *et al.* Association between gut permeability, brain volume, and cognition in healthy participants and patients with schizophrenia spectrum disorder. *Brain Behav.* 13, e3011 (2023).
17. Severance, E. G. *et al.* Gastrointestinal inflammation and associated immune activation in schizophrenia. *Schizophr. Res.* 138, 48–53 (2012).
18. Severance, E. G. *et al.* Discordant patterns of bacterial translocation markers and implications for innate immune imbalances in schizophrenia. *Schizophr. Res.* 148, 130–137 (2013).
19. Maes, M., Kanchanatawan, B., Sirivichayakul, S. & Carvalho, A. F. In Schizophrenia, Increased Plasma IgM/IgA Responses to Gut Commensal Bacteria Are Associated with Negative Symptoms, Neurocognitive Impairments, and the Deficit Phenotype. *Neurotox. Res.* 35, 684–698 (2019).
20. Larson, M. H. *et al.* A comprehensive characterization of the cell-free transcriptome reveals tissue- and subtype-specific biomarkers for cancer detection. *Nat. Commun.* 12, 2357 (2021).

21. Vorperian, S. K., Moufarrej, M. N. & Quake, S. R. Cell types of origin of the cell-free transcriptome. *Nat. Biotechnol.* 40, 855–861 (2022).
22. Koh, W. *et al.* Noninvasive in vivo monitoring of tissue-specific global gene expression in humans. *Proc. Natl. Acad. Sci. U. S. A.* 111, 7361–7366 (2014).
23. Wang, H. *et al.* Depletion-assisted multiplexed cell-free RNA sequencing reveals distinct human and microbial signatures in plasma versus extracellular vesicles. *Clin. Transl. Med.* 14, e1760 (2024).
24. Pan, W. *et al.* Simultaneously Monitoring Immune Response and Microbial Infections during Pregnancy through Plasma cfRNA Sequencing. *Clin. Chem.* 63, 1695–1704 (2017).
25. Olde Loohuis, L. M. *et al.* Transcriptome analysis in whole blood reveals increased microbial diversity in schizophrenia. *Transl. Psychiatry* 8, 96 (2018).
26. Karstens, L. *et al.* Controlling for Contaminants in Low-Biomass 16S rRNA Gene Sequencing Experiments. *mSystems* 4, e00290-19 (2019).
27. Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A. & Callahan, B. J. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* 6, 226 (2018).
28. Eisenhofer, R. *et al.* Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations. *Trends Microbiol.* 27, 105–117 (2019).
29. Salter, S. J. *et al.* Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* 12, 87 (2014).
30. Tan, C. C. S. *et al.* No evidence for a common blood microbiome based on a population study of 9,770 healthy humans. *Nat. Microbiol.* 8, 973–985 (2023).
31. Zozaya-Valdés, E. *et al.* Detection of cell-free microbial DNA using a contaminant-controlled analysis framework. *Genome Biol.* 22, 187 (2021).
32. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37, 907–915 (2019).
33. Blanco-Míguez, A. *et al.* Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nat. Biotechnol.* 41, 1633–1644 (2023).
34. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* 20, 257 (2019).
35. Beghini, F. *et al.* Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *eLife* 10, e65088 (2021).
36. Mallick, H. *et al.* Multivariable association discovery in population-scale meta-omics studies. *PLoS Comput. Biol.* 17, e1009442 (2021).
37. Kay, S. R., Fiszbein, A. & Opler, L. A. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr. Bull.* 13, 261–276 (1987).
38. Nickols, W. A. *et al.* MaAsLin 3: refining and extending generalized multivariable linear models for meta-omic association discovery. *Nat. Methods* 23, 554–564 (2026).

39. Ju, C.-W. *et al.* Modifications of microbiome-derived cell-free RNA in plasma discriminates colorectal cancer samples. *Nat. Biotechnol.* 44, 752–758 (2026).
40. Chen, S. *et al.* Cancer type classification using plasma cell-free RNAs derived from human and microbes. *eLife* 11, e75181 (2022).
41. Khan, I. *et al.* Blood microbiome signatures in systemic diseases: current insights, methodological pitfalls, and future horizons. *Front. Cell. Infect. Microbiol.* 15, 1616029 (2025).
42. Blauwkamp, T. A. *et al.* Analytical and clinical validation of a microbial cell-free DNA sequencing test for infectious disease. *Nat. Microbiol.* 4, 663–674 (2019).
43. Moussiopoulou, J. *et al.* Higher blood-brain barrier leakage in schizophrenia-spectrum disorders: A comparative dynamic contrast-enhanced magnetic resonance imaging study with healthy controls. *Brain. Behav. Immun.* 128, 256–265 (2025).
44. Strandwitz, P. *et al.* GABA-modulating bacteria of the human gut microbiota. *Nat. Microbiol.* 4, 396–403 (2019).
45. Zhu, F. *et al.* Transplantation of microbiota from drug-free patients with schizophrenia causes schizophrenia-like abnormal behaviors and dysregulated kynurenine metabolism in mice. *Mol. Psychiatry* 25, 2905–2918 (2020).

Figures

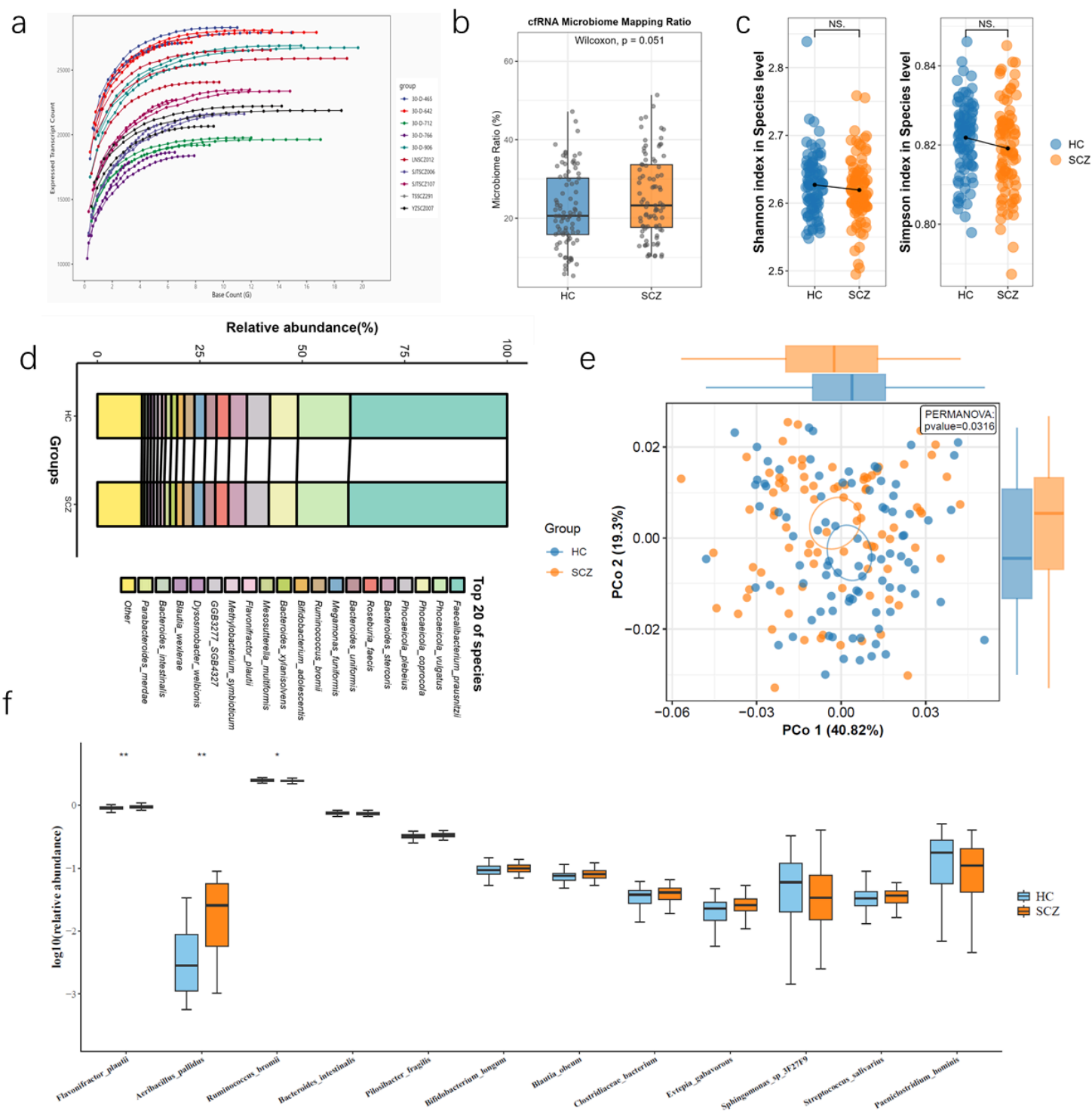


Figure 1

Plasma microbial cfRNA profiles in SCZ and HC participants

a) Sequencing-depth saturation curves for ten representative samples (five SCZ, five HC), plotting the cumulative number of uniquely detected transcripts against increasing sequencing depth (targeted: ~10, ~20, and ~30 Gb). Saturation was reached near 10 Gb, supporting adoption of this depth for primary analysis.

- b) Fraction of post-quality-control clean reads classified as bacterial by Kraken2. Each dot represents one participant (Wilcoxon rank-sum test, $P = 0.051$); median values are indicated by horizontal lines.
- c) Alpha diversity (Shannon and Simpson indices) in SCZ and HC groups. Points represent individual participants; black diamonds and error bars denote group means $\pm$ standard error of the mean (SEM). No significant intergroup differences were observed (Shannon: $P = 0.321$; Simpson: $P = 0.417$, Wilcoxon rank-sum tests).
- d) Group-level relative-abundance composition of the 20 most abundant plasma bacterial species. All remaining species were aggregated into the “Other” category.
- e) Principal-coordinate analysis (PCoA) of Bray-Curtis dissimilarities. PCo1 and PCo2 explain 40.82% and 19.30% of total variation, respectively. Ellipses enclose 95% confidence regions of group centroids; marginal boxplots display coordinate distributions. PERMANOVA confirmed statistically significant compositional divergence between groups ($P = 0.0316$, 9999 permutations, adjusted for age, sex, and BMI).
- f) Relative abundances of the 12 plasma species identified as differentially abundant between the SCZ and HC groups by MaAsLin2 among the 109 species passing the prespecified abundance and prevalence threshold. Models were adjusted for age, sex and BMI, and an exploratory BH-adjusted threshold of $q < 0.10$ was applied. Relative abundances are displayed on a $\log_{10}$ scale for visualisation; blue and orange denote HC and SCZ. Asterisks indicate BH-adjusted significance levels: ** $q < 0.01$ and * $q < 0.05$. Species with $0.05 \leq q < 0.10$ are shown without asterisks.

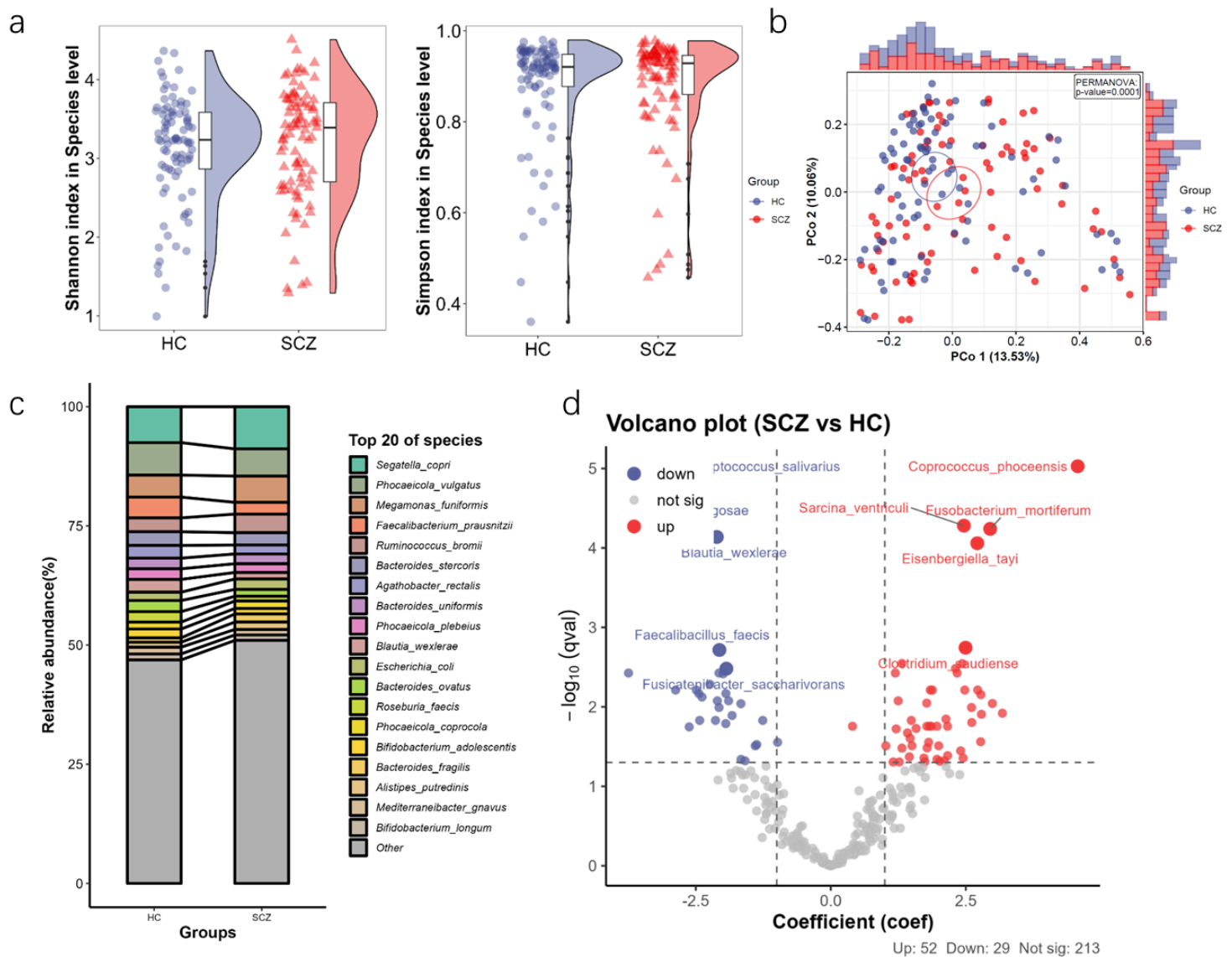


Figure 2

Fecal microbial diversity, taxonomic structure, and differential abundance

a) Alpha diversity (Shannon and Simpson indices) in SCZ and HC groups. Symbols represent individual participants; boxplots show median and interquartile range (IQR); half-violin plots depict kernel density estimates.

b) PCoA of Bray-Curtis dissimilarities. PCo1 and PCo2 account for 13.53% and 10.06% of total variation, respectively. Ellipses indicate 95% confidence regions; marginal histograms show distribution of sample coordinates along each axis. PERMANOVA revealed highly significant group separation ($P = 0.0001$, 9,999 permutations, adjusted for age, sex, and BMI).

c) Relative-abundance composition of the 20 most abundant fecal bacterial species. Remaining species were grouped as "Other."

d) Volcano plot of MaAsLin2-derived species-level differential abundance (adjusted for age, sex, and BMI). The x-axis shows the (SCZ vs. HC) coefficient; the y-axis displays $-\log_{10}(q)$, where q is the Benjamini-Hochberg-adjusted false discovery rate. Red and blue points denote 52 SCZ-enriched and 29 SCZ-depleted species ($q < 0.05$), respectively; grey points represent 213 non-significant species ($q \geq 0.05$).

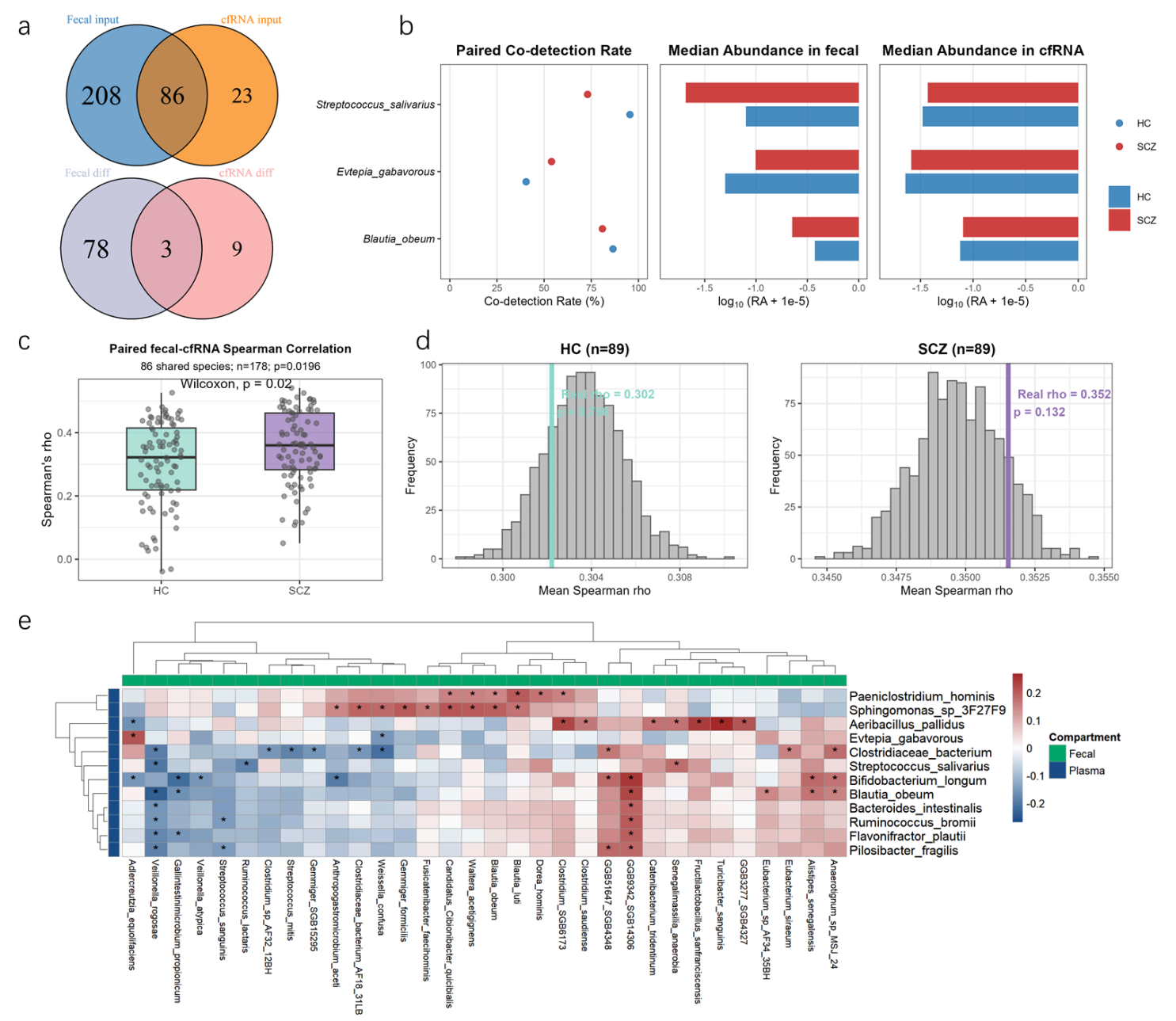


Figure 3

Cross-compartment taxonomic correspondence and co-variation

a) Venn diagram illustrating overlap between retained fecal (n = 294) and plasma (n = 109) species, and between their respective differentially abundant subsets. A total of 86 species were shared across

compartments; among these, three exhibited differential abundance in both compartments, with compartment-specific directions: *Blautia obeum* and *Streptococcus salivarius* were depleted in fecal samples but increased in plasma cfRNA from participants with SCZ, whereas *Evtepia gabavorous* was increased in both fecal samples and plasma cfRNA.

b) Paired co-detection rates (proportion of participants with non-zero abundance in both compartments) and median relative abundances ($\log_{10}(\text{relative abundance} + 1 \times 10^{-5})$) for the three shared differentially abundant species, stratified by group and compartment.

c) Participant-level Spearman's rank correlations (ρ) between paired plasma and fecal relative-abundance vectors across the 86 shared species. Each point represents one participant; group medians are indicated by horizontal lines (HC: $\rho = 0.291$; SCZ: $\rho = 0.347$; Wilcoxon rank-sum test, $P = 0.0196$).

d) Empirical null distributions generated by 10,000 random permutations of plasma-fecal pairings within each diagnostic group. Vertical dashed lines indicate observed mean correlations for correctly paired samples (HC: $\rho = 0.302$, permutation $P = 0.796$; SCZ: $\rho = 0.352$, permutation $P = 0.132$), confirming that observed concordance exceeds chance expectation only at the group level—not per individual pairing.

e) Spearman correlation matrix of covariate-adjusted (age, sex, BMI) centered-log-ratio (CLR) residuals between 12 plasma and 81 fecal differentially abundant species. Only associations with nominal significance ($P < 0.05$) are displayed (384 of 972 pairs), arranged in a 12×32 matrix after hierarchical clustering (Ward's linkage). Color intensity reflects correlation strength (blue: negative; red: positive); A single asterisk indicates nominal significance ($P < 0.05$); no association survived Benjamini–Hochberg correction (minimum $q = 0.286$).

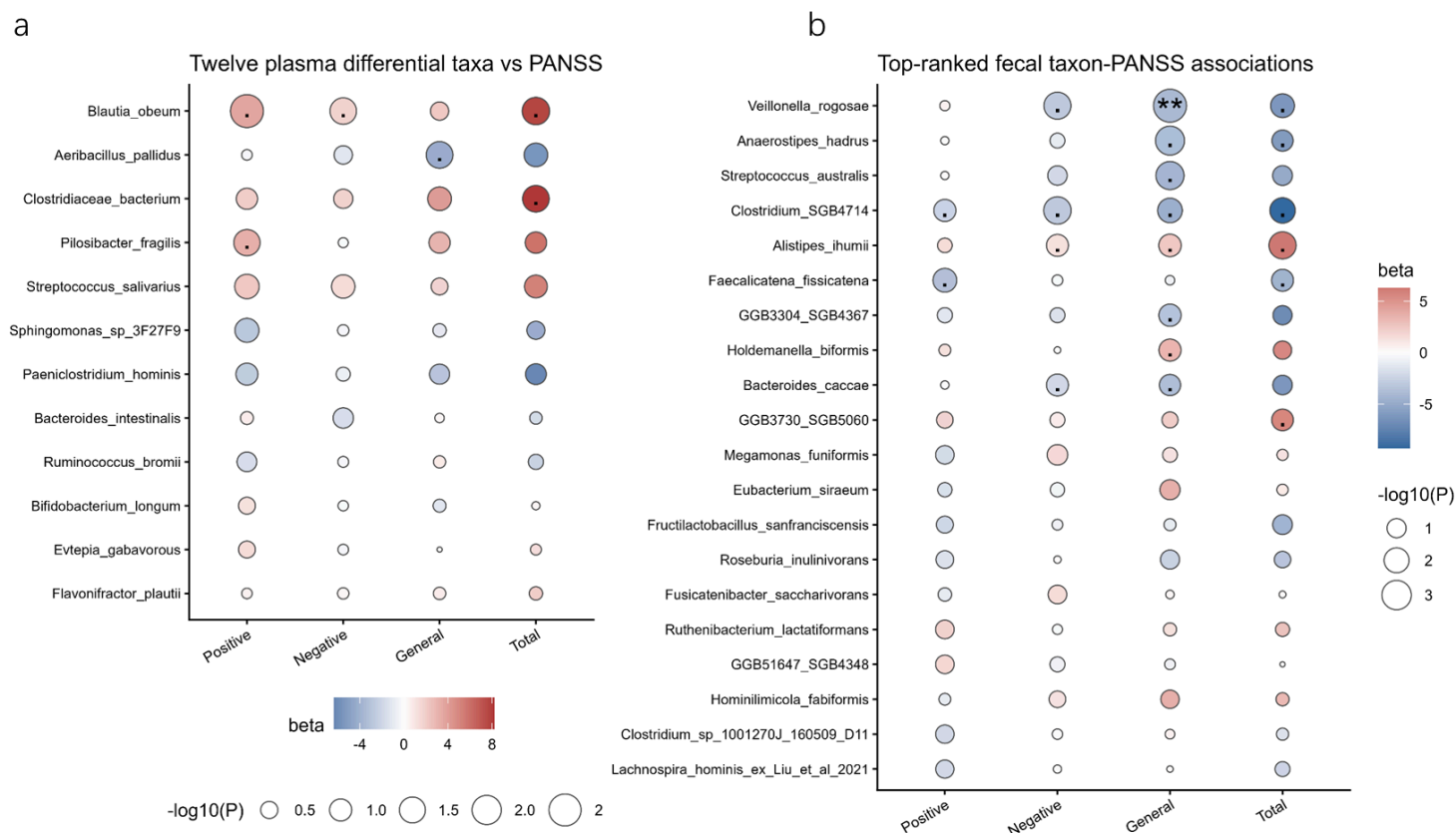


Figure 4

Microbial taxon-PANSS associations in plasma and fecal compartments.

a) Plasma taxon-PANSS associations. Heatmap displays standardized regression coefficients (β) for the 12 diagnosis-associated plasma species regressed against PANSS Positive, Negative, General Psychopathology, and Total scores. Rows: taxa; columns: PANSS domains. Color intensity indicates magnitude and direction of β (red: positive; blue: negative); bubble area scales with $-\log_{10}(P)$, where P values derive from linear models with HC3 heteroscedasticity-robust standard errors (adjusted for age, sex, BMI). Significance annotations: $**q < 0.05$ (FDR-corrected), $*q < 0.10$ (exploratory), $\cdot P < 0.05$ but $q \geq 0.10$.

b) Top 20 fecal taxon-PANSS associations ranked by absolute β value across all PANSS domains. Visualization follows identical conventions as panel (a).

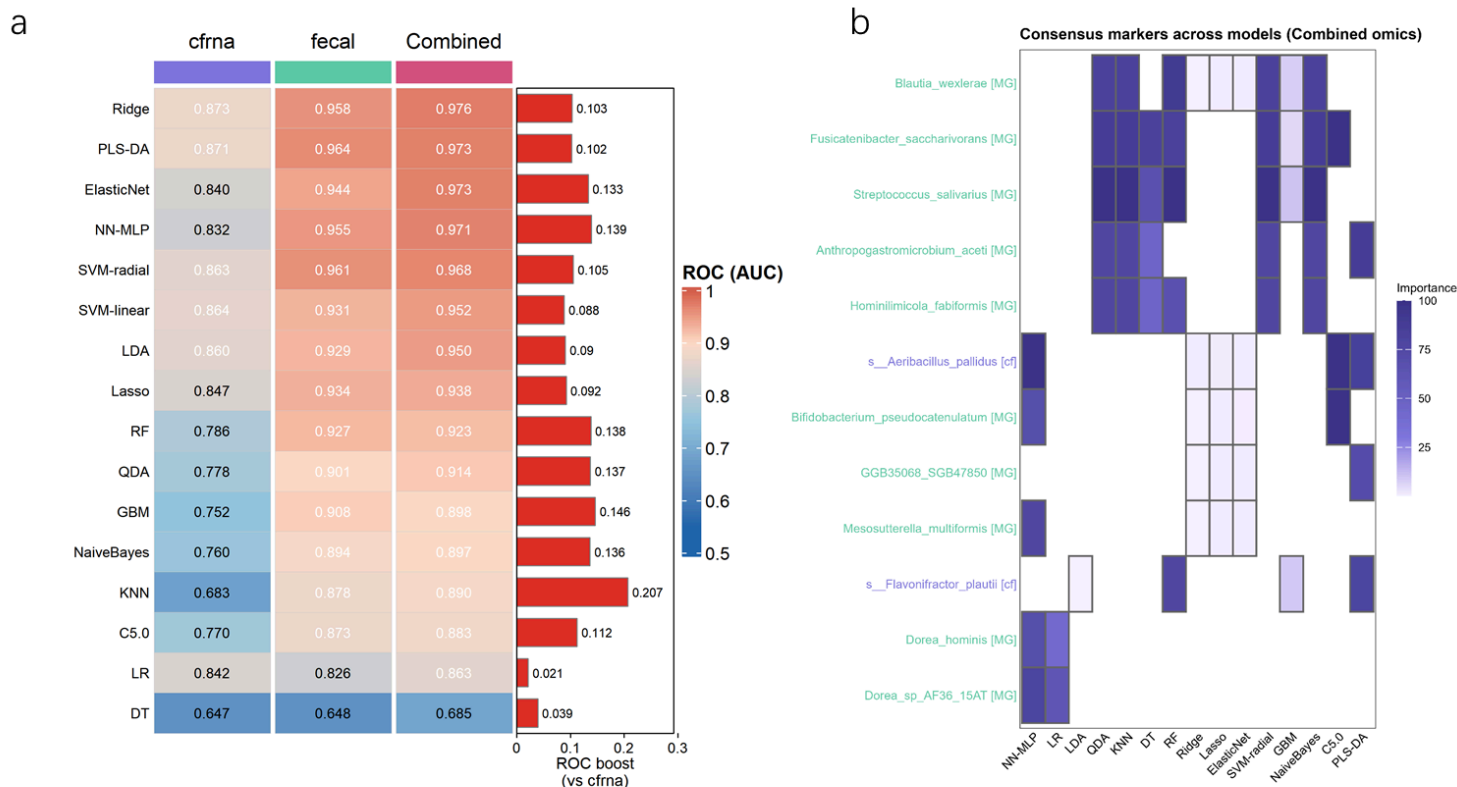


Figure 5

Classification performance and cross-model consensus feature importance.

a) Cross-validated AUCs for 16 classifiers trained on three feature modalities: plasma cfRNA species (left), fecal metagenomic species (center), and their concatenated combined set (right). Heatmap colors reflect AUC magnitude (blue → red). Adjacent bar plots quantify AUC gain of the combined model over its plasma-only counterpart. Peak performance was achieved by Ridge regression (plasma: AUC = 0.873), PLS-DA (fecal: AUC = 0.964), and Ridge regression (combined: AUC = 0.976).

b) Cross-model consensus features—species appearing among the top five most important features in ≥ 2 classifiers within the combined analysis. Rows: species; columns: classifiers. Cell color intensity reflects classifier-specific importance (darker purple = higher importance); blank cells indicate exclusion from top-five ranking. Compartment origin is encoded by label color: green = fecal metagenomic [MG], blue = plasma cfRNA [cf].

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFigures.docx](#)
- [SupplementaryTable.xlsx](#)