

Cross-cohort cerebrospinal fluid multi-omics profiling reveals reproducible metabolic alterations and community-level microbial variation in amyotrophic lateral sclerosis

Yinmei Zhang

yinmeizhang@126.com

Peking University Third Hospital

Nan Yang

Peking University Third Hospital

Ang Li

Peking University Third Hospital

Jiajia Zheng

Peking University Third Hospital

Research Article

Keywords: Amyotrophic Lateral Sclerosis, Cerebrospinal fluid, 2bRAD-M, Microbiome, Metabolomics, Multi-omics

Posted Date: October 14th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11112530/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Cross-cohort cerebrospinal fluid multi-omics profiling reveals reproducible metabolic alterations and community-level microbial variation in amyotrophic lateral sclerosis

Yinmei Zhang^{1#}, Nan Yang², Ang Li¹, Jiajia Zheng^{1#}

¹Department of Laboratory Medicine, Peking University Third Hospital, Haidian District, No. 49 North Garden Road, Beijing 100191, People's Republic of China

²Department of Blood Transfusion, Peking University Third Hospital, Haidian District, No. 49 North Garden Road, Beijing 100191, People's Republic of China

#Correspondence:

Yinmei Zhang

Email: yinmeizhang@126.com Tel +86-010-82266215

Jiajia Zheng

Email: zhengjiajia@bjmu.edu.cn Tel +86-010-82266249

Abstract

Background: Amyotrophic lateral sclerosis (ALS) is accompanied by metabolic and inflammatory disorders, but whether cerebrospinal fluid (CSF)-associated microbial signals covary with metabolic alterations is unclear. We integrated CSF microbiome and metabolome profiling and evaluated the reproducibility of the identified patterns in an independent cohort.

Methods: The discovery cohort included 20 patients with ALS and 20 non-ALS neurological controls; microbiome data were available for all participants and untargeted metabolomics data for 18 participants per group. The validation cohort included 26 patients with ALS and 14 controls with paired datasets. We assessed microbial diversity and taxa, differential metabolites and pathway enrichment, associations with ALS Functional Rating Scale–Revised scores, and exploratory pooled microbiome–metabolome correlations.

Results: Alpha diversity did not differ between ALS and controls in either cohort. Bray–Curtis dissimilarity showed group-associated differences in the discovery (PERMANOVA $P = 0.023$) and validation cohorts (pseudo- $F = 6.089$, $R^2 = 0.138$, $P = 0.002$). No microbial taxon remained significant after false discovery rate correction in the validation cohort. Metabolic alterations were more reproducible: 21 annotated features were significant in both cohorts after correction, and all

changed in the same direction. Glycine and L-cysteine showed robust decreases, and 18 KEGG pathways were significant in both cohorts. Pooled cross-omic correlations were exploratory and potentially influenced by diagnostic group and negative-control signals.

Conclusions: This cross-cohort multi-omics study identifies a reproducible CSF metabolic phenotype in ALS, accompanied by recurring but more heterogeneous community-level microbial variation. Concordant alterations in glycine, L-cysteine, and amino acid-related pathways highlight coordinated ALS-associated biochemical changes. Exploratory microbiome–metabolome associations offer candidates for targeted investigation.

Key words: Amyotrophic Lateral Sclerosis; Cerebrospinal fluid; 2bRAD-M; Microbiome; Metabolomics; Multi-omics

Introduction

Amyotrophic lateral sclerosis (ALS) is a progressive neurodegenerative disorder characterized by degeneration of upper and lower motor neurons, resulting in progressive muscle weakness, atrophy, bulbar dysfunction, and ultimately respiratory failure [1,2]. Approximately 5–10% of ALS cases are familial and may involve pathogenic variants in genes such as SOD1 and C9orf72, whereas most cases are sporadic [1–3]. Despite substantial advances in the genetic and molecular characterization of ALS, its etiology remains incompletely understood, and marked clinical and biological heterogeneity continues to complicate disease classification, biomarker development, and therapeutic intervention [1–3]. Multiple interacting mechanisms have been implicated in ALS pathogenesis, including mitochondrial dysfunction, altered energy metabolism, glutamate excitotoxicity, oxidative stress, and neuroinflammation [4–7]. These processes suggest that ALS may involve broader disturbances in the molecular environments surrounding the nervous system rather than abnormalities confined to motor neurons alone.

The microbiome has emerged as a potentially important component of this broader disease environment because microbial communities can influence host metabolism, immune activity, inflammatory signaling, and the availability of biologically active metabolites. Studies in ALS have reported alterations in gut microbial composition and microbial metabolic functions, although the taxa and pathways identified have varied across cohorts and study designs [8–13]. This heterogeneity suggests that ALS-associated microbial variation may be more consistently expressed through community-level or functional patterns than through a single disease-specific microorganism. However, most available evidence has been derived from fecal or intestinal samples and therefore primarily reflects peripheral microbial ecosystems. Whether microbial signals can be detected in a biological compartment proximal to the central nervous system (CNS), and whether such signals covary with ALS-associated

biochemical alterations, remains insufficiently characterized.

Cerebrospinal fluid (CSF) is in direct contact with the CNS and provides a clinically accessible biofluid for investigating molecular changes occurring in proximity to the affected nervous system. Previous studies have identified alterations in inflammatory mediators and other CSF components in ALS, supporting the value of this compartment for studying disease-associated molecular processes [14]. Microbial profiling of CSF is nevertheless technically challenging because CSF is an extremely low-microbial-biomass and host-rich specimen. Microbial nucleic acids may be scarce relative to human DNA, and sequence-based profiles may be particularly susceptible to reagent-derived DNA, environmental background, stochastic sampling, and other contaminating signals [15–17]. Consequently, detected sequences should be interpreted as “CSF-associated microbial signals” rather than evidence of a viable or resident CSF microbiome.

Recent methodological advances have improved the feasibility of characterizing microbial DNA in low-biomass specimens. The 2bRAD-M approach extends the type IIB restriction enzyme-based 2b-RAD strategy to microbial profiling by generating iso-length reduced-representation tags and assigning microbial taxa through species-specific restriction-site markers [18,19]. Taxonomic assignments are generated computationally using microbial marker databases, together with filtering procedures designed to reduce host-derived and false-positive assignments [19,20]. More recent work has demonstrated that 2bRAD-M can be applied to host-rich total-DNA samples without prior host-DNA depletion [21]. These characteristics make 2bRAD-M a practical approach for investigating CSF-associated microbial profiles, although they do not eliminate the need for rigorous contamination control or establish the viability and biological origin of detected microbial DNA.

Metabolic dysregulation represents another prominent feature of ALS pathophysiology. Metabolomic studies of CSF and peripheral biofluids have identified abnormalities involving amino acid metabolism, mitochondrial and energy-related pathways, lipid metabolism, molecular transport, and redox homeostasis [22–25]. Untargeted metabolomics therefore provides a complementary approach for broadly characterizing biochemical alterations within the CSF environment. Integrating microbiome and metabolome data from the same CSF samples may further determine whether variation in detected microbial features occurs in parallel with reproducible metabolic changes. Such cross-omic associations can identify biologically informed candidates for subsequent targeted investigation, although correlation alone cannot establish the direction or causality of microbiome–metabolome relationships.

In this study, we combined 2bRAD-M-based microbial profiling with untargeted metabolomics to characterize the CSF molecular landscape of patients with ALS and non-ALS neurological controls. We first examined microbial community structure, taxonomic features, metabolic alterations, and selected clinical associations in a discovery cohort. We

then evaluated the reproducibility of these patterns in an independent validation cohort. Finally, paired microbiome and metabolome data were integrated to determine whether selected CSF-associated microbial features covaried with metabolic alterations across cohorts. Through this discovery-and-validation framework, we sought to establish a cross-cohort view of the ALS-associated CSF molecular environment, identify reproducible metabolic abnormalities, determine whether microbial variation showed consistency at the community or taxonomic level, and prioritize cross-omic relationships for future targeted validation.

Methods

Study design and sample collection

This study employed a discovery-and-validation design involving two independent cohorts of patients with amyotrophic lateral sclerosis (ALS) and non-ALS neurological controls. The discovery cohort comprised 20 patients with ALS and 20 controls. ALS was diagnosed according to the revised El Escorial criteria, with ICD-10 code G12.21 used for clinical coding. Controls were patients undergoing clinically indicated lumbar puncture for non-ALS neurological conditions, including non-inflammatory peripheral neuropathy, idiopathic facial nerve palsy, and idiopathic normal-pressure hydrocephalus. Cerebrospinal fluid (CSF) samples were obtained as residual specimens from routine clinical procedures.

An independent validation cohort comprising 26 patients with ALS and 14 non-ALS neurological controls was analyzed separately. Microbiome data were available for all participants in both cohorts. Untargeted metabolomics data were available for 18 patients with ALS and 18 controls in the discovery cohort and for all 26 patients with ALS and 14 controls in the validation cohort. Accordingly, paired microbiome–metabolome analyses included 36 participants in the discovery cohort and 40 participants in the validation cohort. Baseline demographic and clinical characteristics for both cohorts are summarized in Table 1. An overview of the study design, sample allocation, and analytical workflow is shown in Figure 1.

The inclusion criteria were: (1) a confirmed diagnosis of ALS or a non-ALS neurological condition; (2) a clinical indication for lumbar puncture; and (3) provision of written informed consent. Exclusion criteria included: (1) recent infection; (2) antibiotic or antifungal use within the preceding 3 months; (3) active malignancy; (4) immunosuppressive conditions; (5) severe hepatic or renal dysfunction; (6) other neurodegenerative diseases; and (7) major psychiatric disorders.

The study was approved by the Peking University Third Hospital Medical Science Research Ethics Committee (project approval number M20251225; ethics approval document No. 2025-1015-01). Written informed consent was obtained from all participants before sample

collection. All procedures were conducted in accordance with the ethical standards of the institutional research committee and the Declaration of Helsinki.

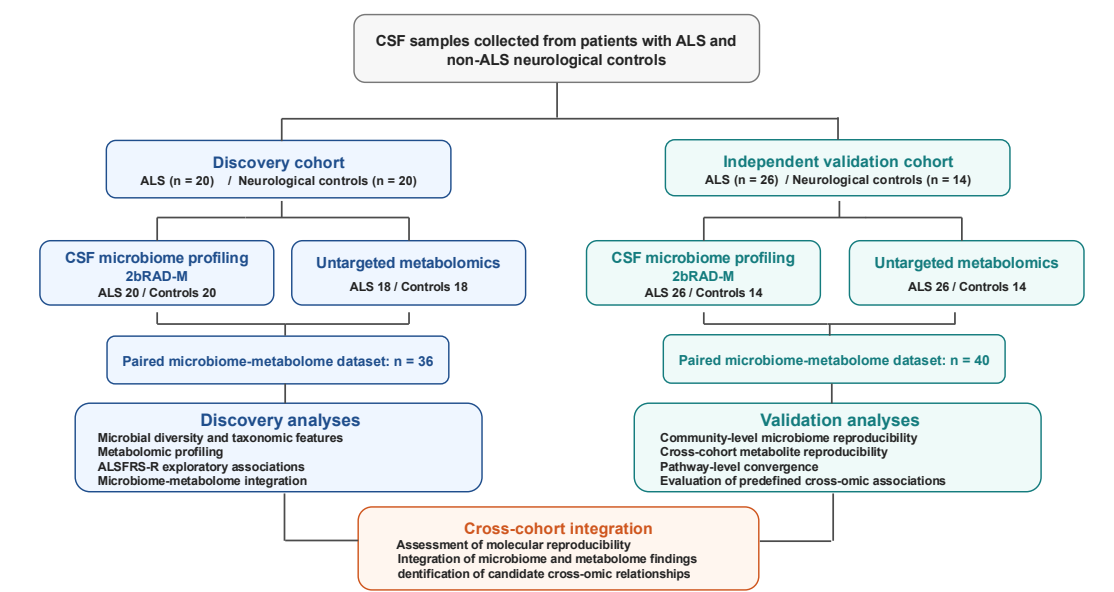


Figure 1. Study design and analytical workflow. Overview of the discovery and independent validation cohorts and the corresponding CSF microbiome, metabolomics, and integrated multi-omics analyses. The validation cohort was used to assess the reproducibility of community-level microbial patterns, metabolic alterations, and cross-omics associations.

Microbiome profiling using 2bRAD-M

CSF microbiome profiling was performed using the 2bRAD-M approach, which is designed for low-biomass and host-rich specimens [19,21]. Genomic DNA (1 pg–200 ng) was digested with 4 U BcgI (New England Biolabs) at 37°C for 3 h. Adapter ligation was performed at 4°C for 12 h using 0.2 μM of each adapter and 800 U T4 DNA ligase. Amplified products of approximately 100 bp were size-selected by 8% polyacrylamide gel electrophoresis, followed by gel extraction and indexing PCR using Phusion High-Fidelity DNA Polymerase. Final libraries were purified using the QIAquick PCR Purification Kit (Qiagen) and sequenced using an Illumina NovaSeq PE150 platform at OE Biotech (Qingdao, China).

Because a single read spans the 2bRAD insert, Read 1 was used for downstream analysis. Reads containing >8% ambiguous bases or in which >20% of bases had quality scores below Q30 were removed. Human-derived tags were subsequently removed using the 2bRAD-M Reads Level Decontamination (RLD) procedure. Four extraction/library-processing negative controls were analyzed in parallel. To further reduce background signals in these low-biomass samples, negative-control sequences were proportionally subsampled according to the relative numbers of clean reads in biological samples and negative controls. Sequences in biological

samples whose forward or reverse-complement sequences matched the subsampled negative-control library were considered background contaminants and removed before taxonomic assignment.

Taxonomic identification and relative abundance estimation

Taxonomic profiling was performed using a species-level 2bRAD-M reference database containing species-specific type IIB restriction tags. For the dataset used in this study, bacterial and archaeal genomes were derived from GTDB release r214 and fungal genomes from GenBank (release dated September 20, 2024). The reference database comprised 413,882 microbial genomes representing 88,890 species. Species-specific 2bRAD markers were defined as tags unique to a species-level taxon.

Quality-controlled, host-filtered, and negative-control-decontaminated reads were mapped to the reference marker database using the 2bRAD-M/MAP2B analysis pipeline. Candidate species were screened using the G score, which incorporates the number of reads assigned to species-specific markers and the number of detected species-specific markers. Species with a G score ≥ 5 were retained to control false-positive assignments. A sample-specific marker database was subsequently constructed from candidate species, and reads were remapped for quantitative estimation. For each species, abundance was estimated from marker coverage normalized by the theoretical number of species-specific markers, and relative abundance was calculated by normalizing this value across all detected species within each sample.

Prediction of microbial functional profiles

Where indicated, microbial functional potential was inferred using PICRUSt2 (v2.3.0b0) [26]. Predicted gene functions were summarized as Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology terms and KEGG pathway categories at levels 1–3, as well as Clusters of Orthologous Groups (COG) annotations. Between-group comparisons of predicted functional abundances were performed using Wilcoxon rank-sum tests, with FDR correction for multiple testing. Because PICRUSt2 infers functional potential from taxonomic composition rather than directly measuring microbial genes or activity, these analyses were treated as exploratory.

Untargeted metabolomics

Untargeted metabolomic profiling was performed using complementary gas chromatography–mass spectrometry (GC–MS) and ultra-performance liquid chromatography–tandem mass spectrometry (UPLC–MS/MS) platforms. CSF samples stored at -80°C were thawed on ice. A 100- μL aliquot of each sample was mixed with 200 μL of precooled methanol–acetonitrile (2:1, v/v) containing internal standards, vortexed for 1 min, sonicated in an ice-water bath for 10 min, and incubated overnight at -40°C . Samples were then centrifuged at 12,000 rpm for 20 min at 4°C . An aliquot of the supernatant was used directly for LC–MS analysis, while another aliquot was dried and subjected to methoxyamination and BSTFA derivatization before GC–MS analysis. Pooled quality-control (QC) samples were prepared by

combining equal volumes of extracts from all study samples.

For LC–MS/MS analysis, chromatographic separation was performed using a Waters ACQUITY UPLC I-Class plus system equipped with an ACQUITY UPLC HSS T3 column (100 mm × 2.1 mm, 1.8 μm) maintained at 45°C. The mobile phases consisted of water containing 0.1% formic acid (A) and acetonitrile (B), with a flow rate of 0.35 mL/min and an injection volume of 5 μL. Mass spectrometric detection was performed using a Thermo Q Exactive HF mass spectrometer with a heated electrospray ionization source in positive- and negative-ion modes. Data were acquired in data-dependent acquisition mode using Full MS/dd-MS2 (Top 10), with an MS1 mass range of m/z 70–1050, full-scan resolution of 60,000, MS/MS resolution of 15,000, and stepped normalized collision energies of 10, 20, and 40.

GC–MS analysis was performed using an Agilent 7890B gas chromatograph coupled to a 5977A mass spectrometer with a DB-5MS capillary column (30 m × 0.25 mm × 0.25 μm). Helium was used as the carrier gas at 1.0 mL/min. Samples were injected in splitless mode at an injection volume of 1 μL and an inlet temperature of 260°C. Electron-impact ionization was performed at 70 eV, with an ion-source temperature of 230°C and quadrupole temperature of 150°C. Full-scan mass spectra were acquired over m/z 50–500.

LC–MS data were processed using XCMS v4.5.4 [27] for baseline filtering, peak detection, integration, and retention-time correction. GC–MS data were processed using MS-DIAL v4.24 [28], including peak detection, deconvolution, metabolite annotation, peak alignment, filtering, and missing-value processing. LC–MS metabolite annotation was based on retention time, accurate mass, MS/MS fragmentation, and isotope-distribution information using HMDB, LipidMaps v2.3, METLIN, and the LuMet-Animal3.0 in-house database. GC–MS metabolites were annotated against the LuMet-GC 5.0 in-house spectral database.

Features with a relative standard deviation >30% in pooled QC samples were excluded. Features with missing or zero values in >50% of samples within a group were removed, and remaining zero values were replaced with one-half of the minimum non-zero intensity across all samples before log2 transformation. GC–MS peak intensities were additionally normalized using internal-standard-based segmented normalization. The processed LC–MS and GC–MS matrices were subsequently combined for downstream statistical analyses.

Metabolomic data analysis

Principal component analysis (PCA) was used for unsupervised visualization of global metabolic variation. Orthogonal partial least-squares discriminant analysis (OPLS-DA) was used as an exploratory supervised approach to assess group separation and to derive variable importance in projection (VIP) scores. OPLS-DA model robustness was evaluated using seven-fold cross-validation and 200 response permutation tests.

Differentially abundant metabolites were defined according to the original analytical pipeline as those meeting both a nominal statistical significance threshold of $P < 0.05$ and a

fold-change threshold of $FC \geq 1.20$ or $FC \leq 1/1.20$ (equivalent to $|\log_2 FC| > \text{approximately } 0.263$). Benjamini–Hochberg-adjusted q values were additionally calculated to assess robustness to multiple testing. VIP scores were used to rank and visualize representative metabolites but were not required for inclusion in the primary differential-metabolite set.

Metabolic pathway enrichment analysis

Differentially abundant metabolites were mapped to KEGG and Reactome pathway annotations. Over-representation analysis was performed using a hypergeometric test to determine whether metabolites assigned to individual pathways were enriched relative to the detected metabolite background. Nominal P values and Benjamini–Hochberg-adjusted q values were calculated, with $q < 0.05$ considered evidence robust to multiple-testing correction. Enrichment results were summarized using the number of mapped differential metabolites, pathway background size, and enrichment metrics. Because individual metabolites may map to multiple overlapping or hierarchically related pathways, pathway findings were interpreted as coordinated biological patterns rather than direct measurements of pathway activity.

Cross-cohort validation and multi-omics integration

The validation cohort was analyzed independently from the discovery cohort. For microbiome analyses, community-level reproducibility was assessed by comparing alpha-diversity patterns and Bray–Curtis beta diversity across cohorts, whereas taxon-level validation focused on the direction and statistical consistency of microbial features identified in the discovery cohort. Taxa newly identified only in the validation cohort were considered exploratory rather than independently validated findings. Because homogeneity of multivariate dispersion was not formally evaluated, significant PERMANOVA results were interpreted as group-associated differences in the overall dissimilarity structure rather than conclusive evidence of centroid separation.

For metabolomic validation, metabolite identities were harmonized between cohorts primarily using InChIKey. Concordant ionization mode was additionally required for quantitative cross-cohort analyses in which direct feature matching was necessary. Concordance was summarized among annotated features that met the same candidate criteria in both cohorts, and a more stringent core set was defined by FDR-adjusted $q < 0.05$ in each cohort. Because the first comparison was conditioned on candidate selection in both datasets, its direction and effect-size concordance estimates were treated as descriptive rather than as an unbiased out-of-sample validation measure. Integrated microbiome–metabolome analyses were restricted to participants with paired datasets. Spearman rank correlations were used to evaluate associations between selected microbial taxa, metabolites, and clinical variables. Microbiome–metabolome correlations were calculated across ALS and control participants combined.

Statistical analysis

Statistical analyses were performed using SPSS v26.0 and R v4.1.1, together with the

software implemented in the microbiome and metabolomics analysis pipelines. Continuous clinical variables were summarized as mean $\pm$ standard deviation when appropriate, and categorical variables were summarized as counts and percentages. Non-normally distributed microbial abundance and diversity data were analyzed using nonparametric methods.

For high-dimensional taxonomic, predicted functional, and metabolomic analyses, nominal P values and Benjamini-Hochberg FDR-adjusted q values were reported. A two-sided $P < 0.05$ was considered nominally significant, whereas FDR-adjusted $q < 0.05$ was regarded as evidence robust to multiple-testing correction. Spearman correlation coefficients were used for exploratory association analyses because of the limited sample size and non-normal distributions of several molecular variables. All statistical tests were two-sided unless otherwise specified.

Results

1. Study population characteristics

The discovery cohort included 20 patients with amyotrophic lateral sclerosis (ALS) and 20 non-ALS neurological controls, and the validation cohort included 26 patients with ALS and 14 controls (Table 1). Age, sex distribution, and body mass index did not differ significantly between ALS and controls in either cohort (all $P \geq 0.196$).

Among patients with ALS, mean ALSFRS-R score and disease duration were 39.6 ± 6.9 and 9.0 ± 7.5 months in the discovery cohort and 40.9 ± 5.6 and 13.4 ± 10.9 months in the validation cohort, respectively. Retrospective abstraction of available validation-cohort records indicated predominantly spinal onset (22/26, 84.6%); bulbar and combined spinal-and-bulbar onset each occurred in two patients (7.7%). The record-based stage distribution was Stage 1 in 10 patients (38.5%), Stage 2 in 9 (34.6%), Stage 3 in 5 (19.2%), and Stage 4 in 2 (7.7%).

Table 1. Demographic and clinical characteristics of the discovery and validation cohorts

[illegible]

Characteristic	Discovery cohort			Validation cohort		
	ALS (n = 20)	Controls (n = 20)	P value	ALS (n = 26)	Controls (n = 14)	P value
Spinal	12 (60.0%)	NA	—	22 (84.6%)	NA	—
Bulbar	6 (30.0%)	NA	—	2 (7.7%)	NA	—
Both	2 (10.0%)	NA	—	2 (7.7%)	NA	—
Disease stage, n (%)						
Stage 1	3 (15.0%)	NA	—	10 (38.5%)	NA	—
Stage 2	6 (30.0%)	NA	—	9 (34.6%)	NA	—
Stage 3	5 (25.0%)	NA	—	5 (19.2%)	NA	—
Stage 4	2 (10.0%)	NA	—	2 (7.7%)	NA	—
Missing/ unknown	4 (20.0%)	NA	—	0 (0.0%)	NA	—

Data are presented as mean $\pm$ SD or n (%), as appropriate. *P* values compare ALS with controls within each cohort. Validation-cohort *P* values were calculated using two-sided Welch's *t* tests for age and BMI and Fisher's exact test for sex. Validation-cohort ALSFRS-R score, disease duration, site of onset, and disease stage were retrospectively abstracted or reconstructed from available clinical documentation and are presented descriptively. Disease stage was summarized as Stages 1–4 for consistency with the discovery cohort. ALSFRS-R, Amyotrophic Lateral Sclerosis Functional Rating Scale–Revised; BMI, body mass index; NA, not applicable.

2. Alterations in the CSF microbial community in the discovery cohort

Alpha-diversity analysis showed no significant differences in microbial richness or diversity between patients with ALS and non-ALS neurological controls, as assessed by the Chao1, Shannon, and Simpson indices (all $P > 0.05$; Supplementary Figure S1). In contrast, beta-diversity analysis indicated a difference in the overall Bray–Curtis dissimilarity structure. Principal coordinates analysis showed partial group-level separation, although overlap remained between ALS and control samples (Figure 2A). PERMANOVA identified a significant group-associated difference ($P = 0.023$). Because homogeneity of multivariate dispersion was not formally evaluated, this result was interpreted as evidence of a modest group-associated community pattern.

Taxonomic comparisons identified several features with nominal between-group differences at the order, family, genus, and species levels (Figures 2B–E). Rhizobiales and Xanthomonadales were relatively more abundant in ALS samples. At lower taxonomic levels, candidate ALS-enriched features included members of Moraxellaceae, Xanthomonadaceae, Micrococcaceae, and Rhizobiaceae, as well as the genera *Acinetobacter*, *Pseudoxanthomonas*, *Corynebacterium*, *Micrococcus*, and *Achromobacter*. Representative species-level features

included *Achromobacter pulmonis*, *Micrococcus luteus*, and *Pseudoxanthomonas kaohsiungensis*. A low-abundance taxonomic feature assigned to *Mycobacterium tuberculosis* was enriched in control samples. Given the low microbial biomass of CSF and the lack of species-specific confirmation, this signal was interpreted cautiously and was not considered evidence of *M. tuberculosis* infection.

Exploratory LEfSe analysis further identified multiple group-associated taxonomic features, with a broader range of features enriched in the ALS group than in controls (Figure 2G). The taxonomic cladogram illustrated that these features were distributed across multiple lineages (Figure 2F). Given the multiple high-dimensional comparisons and the low microbial biomass of CSF, these taxa were regarded as candidate disease-associated signals. Overall, the discovery cohort suggested a modest shift in CSF-associated microbial community composition in ALS.

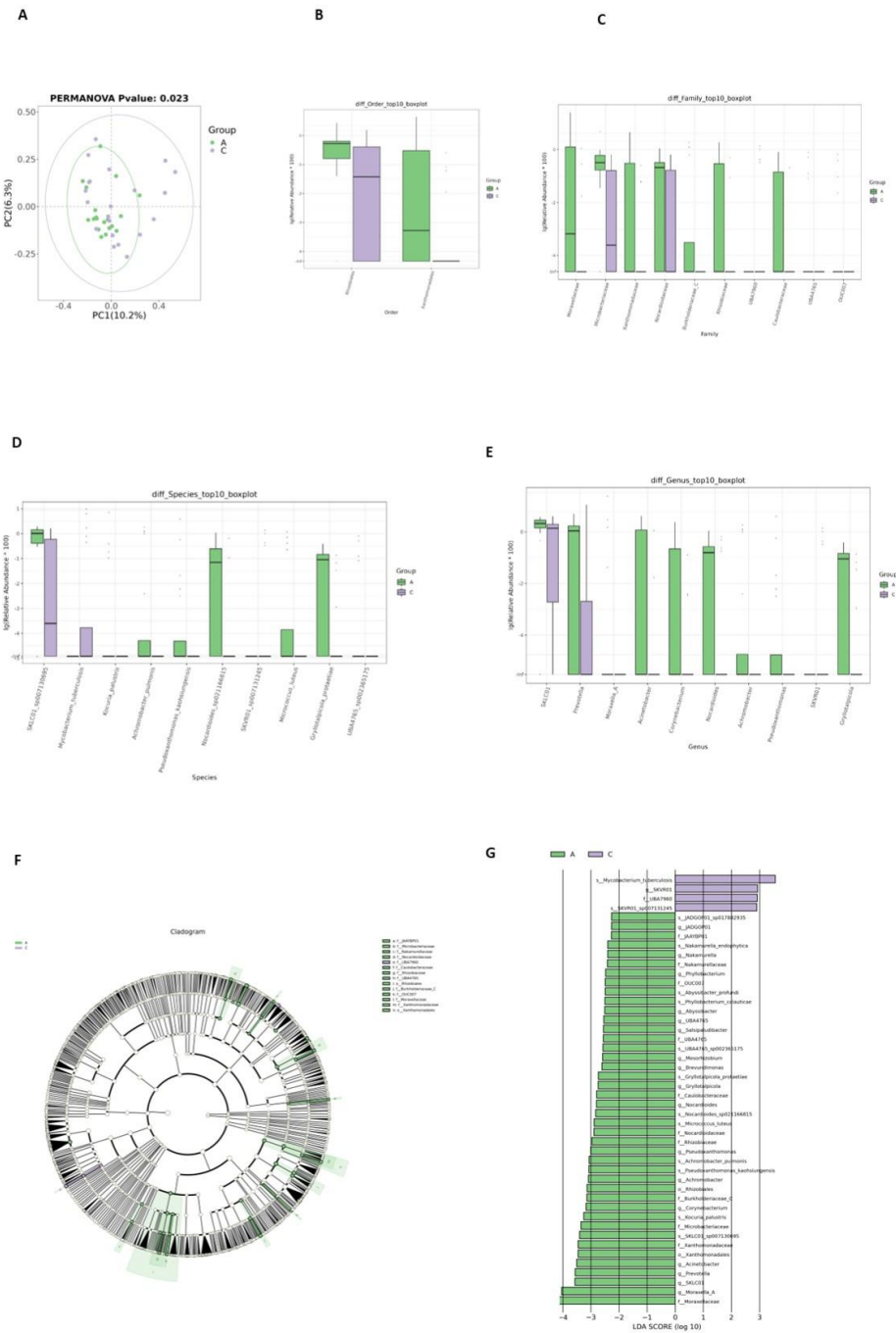


Figure 2. CSF microbial community differences between patients with ALS and non-ALS neurological controls in the discovery cohort.

(A) Principal coordinates analysis based on Bray–Curtis dissimilarity showing partial overlap with partial group-level separation between ALS and control samples. Group differences were assessed using PERMANOVA ($P = 0.023$). Because multivariate dispersion was not formally tested, the PERMANOVA result was interpreted as a group-associated difference in dissimilarity structure rather than definitive centroid separation.

(B–E) Relative abundances of representative taxa showing nominal between-group differences at the (B) order, (C) family, (D) species, and (E) genus levels. Boxes indicate the median and interquartile range, and comparisons were performed using the Wilcoxon rank-sum test.

(F) Taxonomic cladogram showing the distribution of group-associated microbial features across taxonomic lineages.

(G) Exploratory LEfSe analysis showing taxa associated with ALS or control samples using a nominal $P < 0.05$ and an LDA score threshold >2.0 . A, ALS; C, non-ALS neurological controls.

3. Metabolomic profiling reveals broad CSF metabolic alterations in the discovery cohort

Untargeted metabolomic profiling was performed in 36 participants from the discovery cohort with available metabolomics data, including 18 patients with ALS and 18 non-ALS neurological controls.

Principal component analysis (PCA) showed partial separation between ALS and control samples, with some overlap remaining, indicating that group-related metabolic variation is accompanied by inter-individual heterogeneity (Figure 3A). Supervised orthogonal partial least-squares discriminant analysis (OPLS-DA) showed more pronounced separation between the two groups (Figure 3B), with a goodness-of-fit value of $R^2 = 0.946$. Permutation testing using 200 permutations yielded a negative Q^2 intercept, providing no evidence of obvious model overfitting (Supplementary Figure S2).

Using the predefined differential-abundance criteria, a total of 505 metabolites differed between ALS and control samples, including 191 metabolites with higher abundance and 314 with lower abundance in ALS (Figure 3C). Thus, metabolites showing lower abundance in ALS outnumbered those showing higher abundance. Heatmap visualization further demonstrated group-associated abundance patterns among differential metabolites, while also revealing appreciable within-group heterogeneity (Figure 3D).

To highlight representative metabolic alterations, a lollipop plot integrating log₂ fold change and variable importance in projection (VIP) scores was generated (Figure 3E). Several metabolites showed marked decreases in ALS, including bilirubin, ethyl lysine, arabinonic acid, itaconic acid, behenic acid, and L-lyxulose. In contrast, threonylserine, selected ceramide-related lipids, N-carboxyethyl- γ -aminobutyric acid, Hymexazol N-glucoside, and PC(O-22:2/22:3) showed higher abundance in ALS samples. These representative features spanned multiple chemical classes, including amino acid-related metabolites, lipids, and other small molecules.

Collectively, the discovery-cohort metabolomic analysis identified a broad ALS-associated CSF metabolic signature characterized by a predominance of decreased metabolic features

together with selective increases in specific metabolites. These findings were subsequently evaluated for reproducibility in the independent validation cohort.

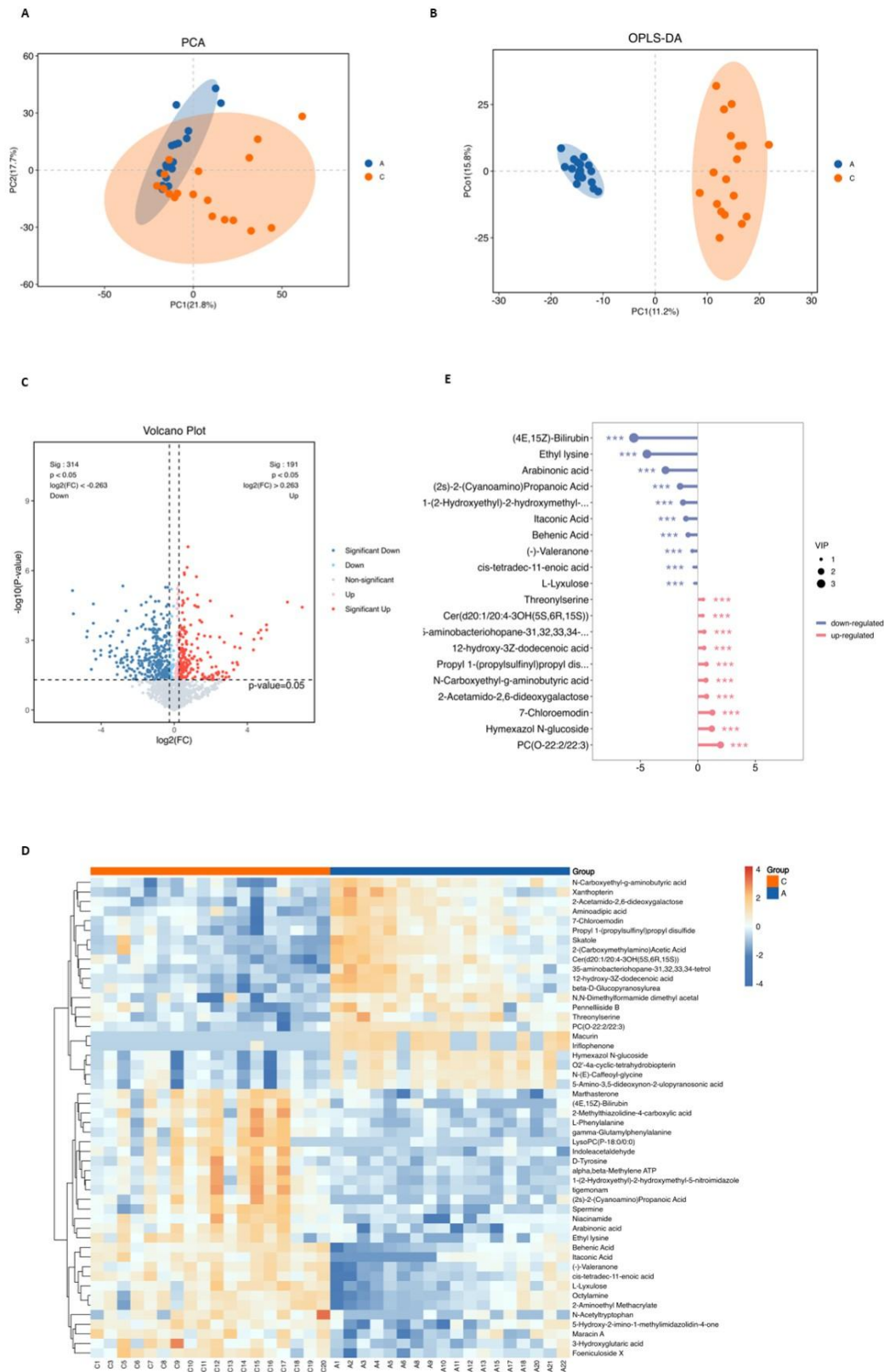


Figure 3. CSF metabolomic alterations in patients with ALS in the discovery cohort.

- (A) Principal component analysis (PCA) score plot showing partial separation between ALS patients and non-ALS neurological controls.
- (B) Orthogonal partial least-squares discriminant analysis (OPLS-DA) score plot showing group discrimination between ALS and control samples.
- (C) Volcano plot showing metabolites meeting the predefined differential-abundance criteria. Red and blue points indicate metabolites with higher and lower abundance, respectively, in ALS relative to controls.
- (D) Heatmap showing group-associated abundance patterns of representative differential metabolites across discovery-cohort samples.
- (E) Lollipop plot showing representative differential metabolites according to log₂ fold change and variable importance in projection (VIP) score. A, ALS; C, non-ALS neurological controls.

4. Exploratory associations between CSF molecular features and functional status in ALS

Exploratory associations between selected CSF molecular features and ALSFRS-R scores were evaluated using Spearman rank correlation in patients with ALS from the discovery cohort. The interval from symptom onset to diagnosis was not correlated with ALSFRS-R score ($\rho = -0.049$, $P = 0.836$).

JAAYBP01, JADGOP01, and JADGOP01_sp017882935 showed identical abundance profiles across ALS samples, consistent with a single lineage-related microbial signal rather than independent taxonomic features. This combined signal was negatively correlated with ALSFRS-R score ($\rho = -0.621$, $P = 0.0035$; $n = 20$), with its detection or higher abundance associated with lower functional scores (Figure 4A,B). Because the signal was detected in only six patients, the association may have been driven predominantly by differences between detected and undetected samples.

Among the metabolic features, a feature putatively annotated as Hymexazol N-glucoside was positively correlated with ALSFRS-R score ($\rho = 0.620$, $P = 0.0060$; $n = 18$), such that lower abundance was associated with greater functional impairment (Figure 4A,C). A feature annotated as threonylserine showed a weaker nominal positive association with ALSFRS-R score ($\rho = 0.493$, $P = 0.0377$; $n = 18$) and was considered a secondary exploratory finding.

Overall, selected microbial and metabolic features showed nominal associations with functional status in the ALS discovery cohort. Given the small sample size, sparse detection of the microbial signal, exploratory feature selection, and potential multiple-testing burden, these findings should be regarded as hypothesis-generating and require independent confirmation.

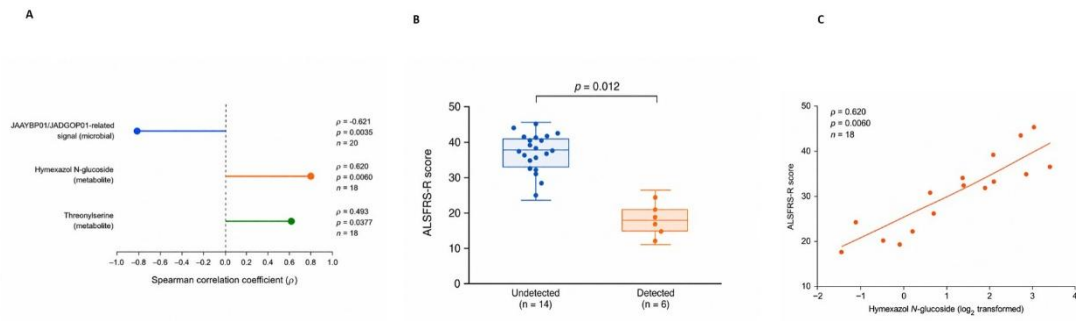


Figure 4. Exploratory associations between selected CSF molecular features and ALSFRS-R scores in patients with ALS.

(A) Summary of Spearman correlations between selected microbial and metabolic features and ALSFRS-R scores in the discovery cohort.

(B) ALSFRS-R scores in samples in which the JAAYBP01/JADG0P01-related signal was undetected or detected. Boxes show medians and interquartile ranges, and points represent individual patients. The secondary between-group comparison was performed using the exact Wilcoxon rank-sum test ($P = 0.012$). The primary association analysis used continuous abundance values and Spearman correlation.

(C) Scatter plot showing the association between Hymexazol N-glucoside abundance and ALSFRS-R score. The fitted line is shown for visualization only.

5. Independent validation supports reproducible CSF metabolic alterations but limited taxon-level microbial replication

An independent validation cohort comprising 26 patients with ALS and 14 non-ALS neurological controls was analyzed to assess the reproducibility of the molecular alterations observed in the discovery cohort.

At the microbial community level, alpha diversity remained comparable between ALS and control samples. No significant differences were observed in Chao1 richness ($P = 0.076$), Shannon diversity ($P = 0.266$), or Simpson diversity ($P = 0.318$) (Supplementary Figure S3A–C), consistent with the discovery-cohort findings. Bray–Curtis analysis detected a group-associated difference in community dissimilarities (PERMANOVA: pseudo- $F = 6.089$, $R^2 = 0.138$, $P = 0.002$; ANOSIM: $R = 0.229$, $P = 0.005$). Substantial overlap between groups and relatively broader dispersion among controls were evident in the ordination space (Supplementary Figure S3D). Because homogeneity

of multivariate dispersion was not formally evaluated, these results support recurring group-associated community variation but cannot distinguish centroid-location effects from dispersion effects.

Taxon-level reproducibility was comparatively modest. Differences between groups were observed at several taxonomic levels, yet none withstood FDR correction (Supplementary Figure S3E). Moreover, several microbial features highlighted in the discovery cohort were not consistently reproduced. *Acinetobacter* showed no significant difference in abundance distributions between ALS and control samples in the validation cohort. *Pseudoxanthomonas* and *Corynebacterium* remained directionally higher in ALS but did not reach statistical significance ($P = 0.167$ and $P = 0.103$, respectively), while *Nocardioides* did not show significant replication of the discovery-cohort finding ($P = 0.058$) (Supplementary Figure S3F). Thus, the validation data provided partial support for community-level microbial variation but did not establish a reproducible ALS-associated taxonomic signature.

Metabolic alterations showed substantially stronger cross-cohort reproducibility. In the validation cohort, unsupervised PCA suggested partial separation between ALS and control metabolic profiles, although appreciable overlap remained (Figure 5A). A separately fitted supervised OPLS-DA model showed clearer group separation (Figure 5B). Among 1,498 detected metabolic features, 344 met the same nominal criteria used in the discovery cohort ($P < 0.05$ and $|\log_2 \text{fold change}| > 0.263$), including 91 features with higher abundance and 253 with lower abundance in ALS (Figure 5C). The predominance of decreased features was consistent with the discovery cohort, in which 314 of 505 candidate differential features were lower in ALS.

Cross-cohort matching provided descriptive evidence of concordance at the annotated-feature level. Among candidate differential features matched by InChIKey, 80 annotated features were shared between cohorts, of which 75 (93.8%) changed in the same direction. Requiring both an identical InChIKey and concordant ionization mode yielded 70 shared features, of which 68 (97.1%) were directionally concordant; their effect sizes were correlated (Pearson $r = 0.804$, $P = 4.98 \times 10^{-17}$; Figure 5D). Because these comparisons were restricted to features meeting candidate criteria in both cohorts, the concordance estimates may be optimistic and do not represent unselected out-of-sample performance. More stringently, 21 matched features remained FDR-

significant ($q < 0.05$) in both cohorts, and all 21 changed in the same direction (Figure 5E).

Representative features illustrated different levels of cross-cohort replication (Figure 5F). A feature putatively annotated as Hymexazol N-glucoside was increased in ALS in both the discovery ($\log_2$ fold change = 1.216, $q = 0.00061$) and validation cohorts ($\log_2$ fold change = 0.905, $q = 0.00557$). Features annotated as L-cysteine and glycine were decreased in both cohorts and achieved FDR-level replication (L-cysteine: discovery $\log_2$ fold change = -0.506 , $q = 0.0168$; validation $\log_2$ fold change = -0.812 , $q = 2.59 \times 10^{-7}$; glycine: discovery $\log_2$ fold change = -1.586 , $q = 0.0122$; validation $\log_2$ fold change = -1.446 , $q = 0.0429$). L-kynurenine, L-phenylalanine, and (4E,15Z)-bilirubin also showed concordant decreases, although their FDR-adjusted values exceeded 0.05 in the validation cohort and therefore represented directional rather than FDR-level replication.

Overall, the validation cohort reproduced the predominance of decreased CSF metabolic features and identified a stringent core of 21 annotated features with FDR-significant, directionally concordant changes in both cohorts. In contrast, microbial community differences were more heterogeneous and showed limited reproducibility at the individual-taxon level.

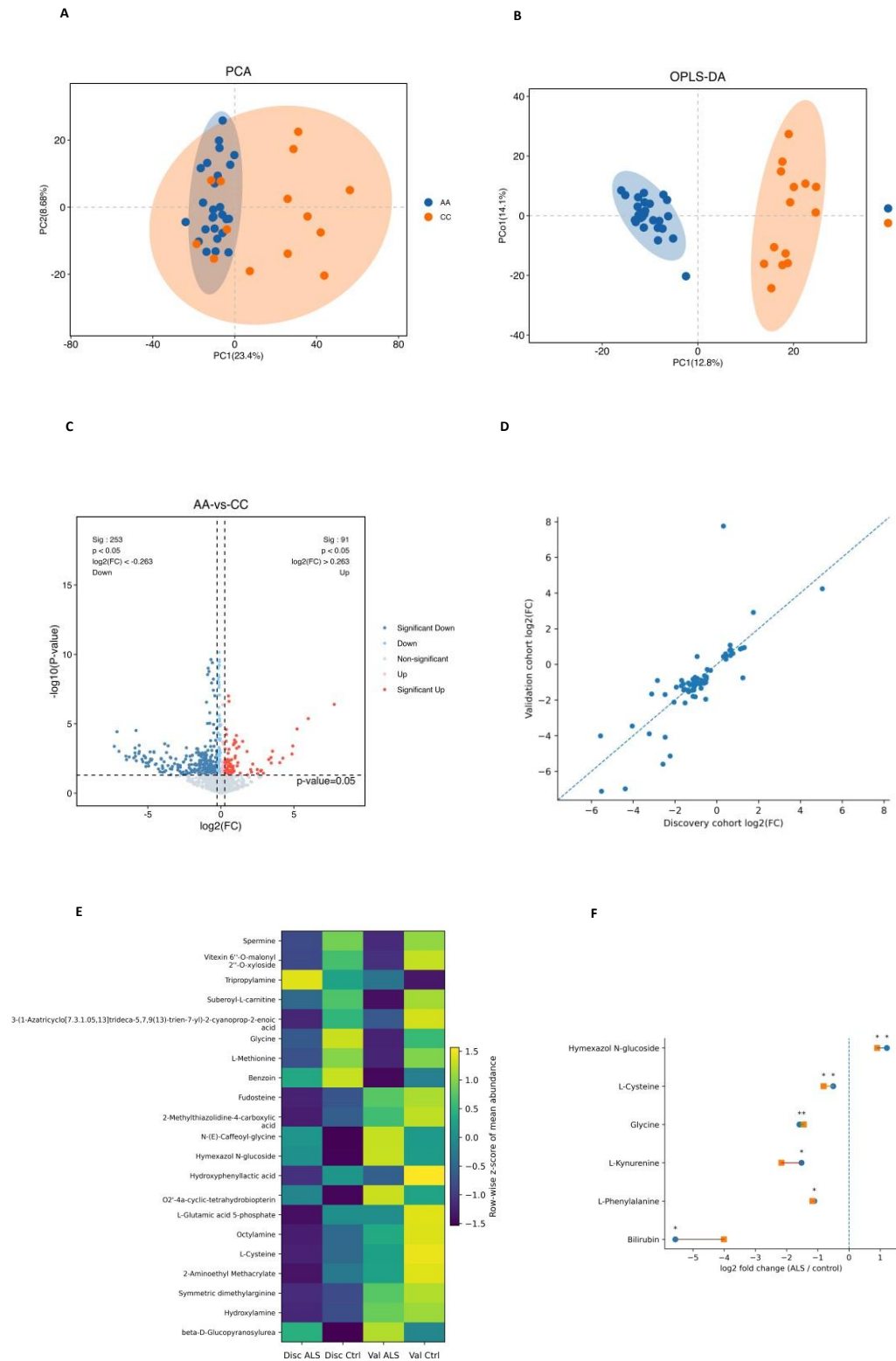


Figure 5. Independent validation supports reproducible CSF metabolic alterations in ALS.

(A) Principal component analysis (PCA) of the validation-cohort metabolomic profiles from patients with ALS

(AA) and non-ALS neurological controls (CC), showing partial group separation with overlap.

(B) Orthogonal partial least-squares discriminant analysis (OPLS-DA) of the validation cohort, showing clearer group separation.

(C) Volcano plot of differential metabolites in the validation cohort. Using the same predefined criteria as in the discovery cohort ($P < 0.05$ and absolute $\log_2$ fold change > 0.263), 344 metabolites were identified as differential, including 91 with higher and 253 with lower abundance in ALS.

(D) Cross-cohort comparison of fold changes for candidate differential metabolites matched by both InChIKey and ionization mode. Each point represents one of 70 shared metabolites; 68 of 70 (97.1%) showed concordant directions of change. The dashed diagonal indicates equal fold changes (Pearson $r = 0.804$, $P = 4.98 \times 10^{-17}$). Because the displayed features met candidate criteria in both cohorts, this concordance estimate is descriptive and does not represent unselected out-of-sample performance.

(E) Heatmap of the 21 metabolites that remained FDR-significant ($q < 0.05$) in both cohorts and showed concordant directions of change. Values represent row-wise standardized mean metabolite abundance for ALS and control groups in the discovery and validation cohorts.

(F) Cross-cohort effect sizes for representative replicated metabolites. Circles and squares denote discovery- and validation-cohort $\log_2$ fold changes, respectively, with connecting lines linking the same metabolite across cohorts. Positive values indicate higher abundance in ALS and negative values indicate lower abundance. Asterisks indicate FDR-adjusted $q < 0.05$. Representative metabolites include a feature putatively annotated as Hymexazol N-glucoside, L-cysteine, glycine, L-kynurenine, L-phenylalanine, and bilirubin. AA, ALS; CC, non-ALS neurological controls.

6. Pathway-level analyses identify convergent metabolic alterations across cohorts

To characterize the biological processes represented by the CSF metabolic changes, KEGG and Reactome pathway representation analyses were performed separately using annotated differential features from the discovery and validation cohorts.

In the discovery cohort, 28 KEGG pathways remained significant after FDR correction. The representative pathways displayed in Figure 6A were predominantly related to amino acid metabolism, energy metabolism, neuroactive signaling, and molecular transport. These included alanine, aspartate and glutamate metabolism, GABAergic synapse, aminoacyl-tRNA biosynthesis, beta-alanine metabolism, the citrate cycle, taurine and hypotaurine metabolism, glycolysis/gluconeogenesis, neuroactive ligand–receptor interaction, D-amino acid metabolism, glyoxylate and dicarboxylate metabolism, lysine degradation, and ABC transporters.

The validation cohort showed a broadly similar pattern, with 28 KEGG pathways remaining significant after FDR correction. Representative pathways included aminoacyl-tRNA biosynthesis, taurine and hypotaurine metabolism, alanine, aspartate and glutamate

metabolism, D-amino acid metabolism, arginine biosynthesis, one-carbon pool by folate, glutathione metabolism, neuroactive ligand–receptor interaction, glycine, serine and threonine metabolism, cysteine and methionine metabolism, sulfur metabolism, phenylalanine metabolism, and ABC transporters (Figure 6B).

Cross-cohort comparison identified 18 KEGG pathways that were FDR-significant in both cohorts. Recurrent pathways included aminoacyl-tRNA biosynthesis, alanine, aspartate and glutamate metabolism, D-amino acid metabolism, taurine and hypotaurine metabolism, neuroactive ligand–receptor interaction, ABC transporters, glyoxylate and dicarboxylate metabolism, arginine biosynthesis, and one-carbon pool by folate. These results indicated substantial pathway-level concordance, particularly in amino acid-related metabolism, cellular transport, and neuroactive signaling.

Reactome analysis provided complementary findings. In the discovery cohort, significant enrichment was concentrated in transport-related processes (Figure 6C). SLC-mediated transmembrane transport; transport of nucleosides and free purine and pyrimidine bases across the plasma membrane; transport of vitamins, nucleosides, and related molecules; transport of small molecules; and organic cation, anion, and zwitterion transport remained significant after FDR correction (all $q < 0.05$).

In the validation cohort, Reactome analysis identified nominal enrichment of carnitine synthesis, endogenous sterol metabolism, plasma lipoprotein assembly, remodeling and clearance, phase I functionalization of compounds, bile acid and bile salt metabolism, and biological oxidations (Figure 6D). However, none of these terms remained significant after FDR correction, and the validation-cohort Reactome findings were therefore considered exploratory rather than independently reproduced.

Overall, pathway-level analyses demonstrated substantial cross-cohort reproducibility at the KEGG level, with 18 pathways reaching FDR significance in both cohorts. The convergent pathways predominantly involved amino acid metabolism, molecular transport, and neuroactive signaling. Reactome analysis additionally supported transport-related alterations in the discovery cohort, whereas the oxidative and lipid-related processes observed in the validation cohort did not withstand multiple-testing correction.

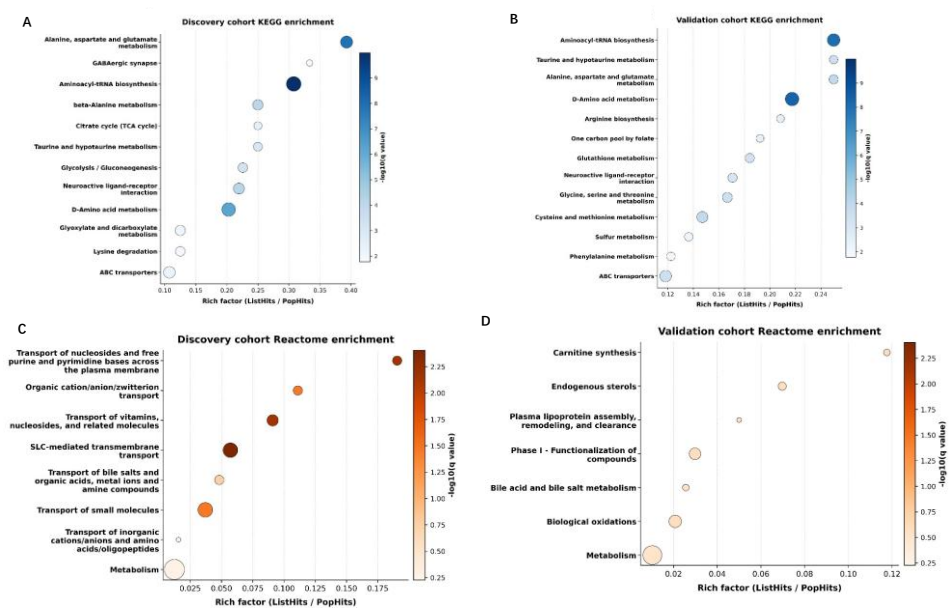


Figure 6. Pathway-level characterization of CSF metabolic alterations across the discovery and validation cohorts.

(A, B) KEGG pathway enrichment analyses of differential metabolites in the discovery (A) and independent validation (B) cohorts. Recurrent pathways included aminoacyl-tRNA biosynthesis, D-amino acid metabolism, alanine/aspartate/glutamate metabolism, taurine and hypotaurine metabolism, neuroactive ligand–receptor interaction, and ABC transporters.

(C, D) Reactome enrichment analyses in the discovery (C) and validation (D) cohorts. Discovery-cohort enrichment was dominated by molecular transport-related processes, whereas the validation cohort showed nominal enrichment of oxidative, sterol, bile acid, carnitine, and lipoprotein-related processes. No Reactome term in the validation cohort remained significant after FDR correction. In all panels, the x-axis represents the rich factor (ListHits/PopHits), bubble size represents the number of mapped metabolites, and bubble color represents $-\log_{10}(\text{FDR-adjusted } q \text{ value})$.

7. Integrated multi-omics analysis identifies exploratory cross-cohort patterns of microbiome–metabolome associations in ALS

Integrated analysis of paired microbiome and metabolome data was performed to examine whether CSF-associated microbial features covaried with selected metabolic alterations. In the discovery cohort, paired data were available for 36 participants, including 18 patients with ALS and 18 non-ALS neurological controls. Spearman correlations were calculated across the two diagnostic groups combined between six selected microbial genera and ten metabolic features,

resulting in 60 genus–metabolite comparisons. Sixteen associations reached nominal significance ($P < 0.05$), although none remained significant after Benjamini–Hochberg correction (all $q \geq 0.085$; Figure 7A). Because diagnostic group was not adjusted for, common case–control shifts may have contributed to these pooled correlations.

Nesterenkonia showed the broadest nominal association profile, with positive correlations with L-kynurenine, pyruvic acid, propionylcarnitine, bilirubin, and L-cysteine, and a negative correlation with L-aspartic acid. *Halomonas* was positively correlated with glycine, L-kynurenine, fumaric acid, and sphingosine. *Corynebacterium* was negatively correlated with bilirubin and glycine, whereas *Pelomonas* was negatively correlated with bilirubin. *Pseudoxanthomonas* was negatively correlated with L-phenylalanine and pyruvic acid, and *Nocardioideis* was negatively correlated with L-phenylalanine. Given that these associations did not withstand multiple-testing correction, they were considered exploratory rather than independently significant cross-omic relationships.

The same 60 genus–metabolite combinations were examined in the validation cohort comprising 40 paired samples. Of the 16 nominal associations identified in the discovery cohort, 14 showed the same direction in validation and eight reached nominal $P < 0.05$. Five associations remained FDR-significant within the validation-cohort 60-pair matrix: *Halomonas* with glycine, L-kynurenine, fumaric acid, and sphingosine, and *Corynebacterium* with glycine. However, the corresponding genera did not show robust differential abundance between ALS and control samples. *Halomonas* and *Corynebacterium* were also detected in negative-control samples at relative abundances comparable to or greater than those in study samples. Accordingly, correlations involving these genera were not considered microbiologically validated and were retained only as exploratory observations that may include background or technical contributions.

Several metabolic nodes represented in the discovery network also showed reproducible ALS-associated abundance changes across cohorts (Figure 7C). L-cysteine and glycine were decreased in ALS and remained FDR-significant in both cohorts. L-kynurenine, L-phenylalanine, and bilirubin also showed concordant decreases, although their validation-cohort q values were above 0.05. Pyruvic acid showed a concordant decrease and was FDR-significant in the validation cohort but not in the discovery cohort. Propionylcarnitine and sphingosine showed concordant decreases but did not remain FDR-significant in both datasets. In contrast, L-aspartic acid changed in

opposite directions between cohorts, whereas fumaric acid showed little evidence of a stable cross-cohort effect.

Exploratory PICRUST2 analyses identified nominal differences in predicted COG, KO, and KEGG functions, but no predicted microbial functional feature remained significant after FDR correction in either cohort. Because these functions were inferred from taxonomic composition rather than directly measured, they were considered supplementary and were not interpreted as evidence of microbial functional activity.

Overall, the paired analyses identified candidate microbiome–metabolome associations with partial directional consistency across cohorts, alongside reproducible ALS-associated changes in selected metabolic nodes. These cross-omic patterns provide a focused starting point for targeted investigation, but the microbial associations remain exploratory because correlations were pooled across diagnostic groups and some implicated genera overlapped with negative-control signals. They do not establish a directional microbiome-to-metabolome mechanism.

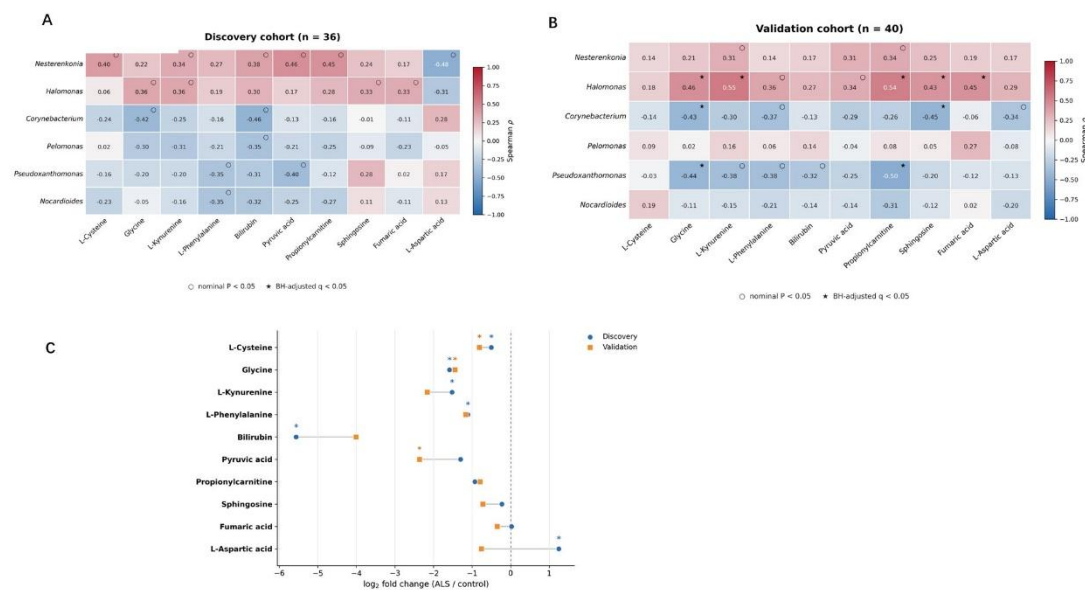


Figure 7. Cross-cohort comparison of CSF microbiome–metabolome associations and metabolic changes in ALS.

(A, B) Heatmaps of Spearman correlations between six selected CSF-associated genera and ten representative metabolites in paired samples from the discovery (A; n = 36) and validation (B; n = 40) cohorts. Cell colors and numbers indicate Spearman correlation coefficients (rho). Open circles indicate nominal $P < 0.05$, and stars

indicate BH-adjusted $q < 0.05$ after correction for the 60 genus–metabolite tests within each cohort. (C) Cross-cohort comparison of log₂ fold changes for the ten metabolites included in the correlation analysis. Circles and squares represent discovery- and validation-cohort estimates, respectively, and connecting lines link the same metabolite across cohorts. Positive values indicate higher abundance in ALS and negative values indicate lower abundance. Asterisks indicate BH-adjusted $q < 0.05$ within the corresponding cohort. Correlations were calculated across ALS and control participants combined and were not adjusted for diagnostic group; they may therefore partly reflect common case–control differences. Correlations involving taxa also detected in negative-control samples were considered exploratory.

Discussion

In this study, we integrated CSF-associated microbial profiling and untargeted metabolomics across two independent ALS cohorts, identifying a reproducible metabolic phenotype alongside recurring group-associated community-level microbial variation. The independent validation design revealed an informative distinction between molecular layers: 21 annotated metabolic features were FDR-significant and directionally concordant in both cohorts, and 18 KEGG pathways were significant in both, whereas microbial variation was more consistent at the community level than at the level of individual taxa. Exploratory microbiome–metabolome analyses further identified candidate relationships for targeted study. Together, these findings extend the molecular characterization of ALS-associated CSF abnormalities while distinguishing the well-replicated metabolic results from more heterogeneous and technically challenging microbial signals.

Our microbiome findings add a CSF-associated perspective to the heterogeneous literature linking ALS to microbial alterations. Most human studies have examined the gastrointestinal microbiome, reporting differences in microbial composition, butyrate-producing taxa, and predicted microbial functions [8–13,29]. Experimental manipulation of intestinal microbiota and microbiota-associated metabolites has also modified disease phenotypes in ALS-prone mice, supporting the biological plausibility of microbiome–metabolite interactions [30]. Human taxon-level findings, however, vary across cohorts [10,11,29]. In our two CSF cohorts, alpha diversity was unchanged, whereas Bray–Curtis analyses repeatedly detected group-associated differences in dissimilarity structure. These results motivate investigation of broader CSF-associated microbial profiles rather than a single disease-specific taxon; because multivariate dispersion was not formally assessed, they do not establish a difference in group centroids.

The validation cohort clarified the level at which the microbial findings were reproducible. *Acinetobacter*, *Pseudoxanthomonas*, *Corynebacterium*, and *Nocardioide*s did not replicate as statistically significant individual genera, and no taxon remained significant after FDR correction in the validation cohort. *Pseudoxanthomonas* and *Corynebacterium* nevertheless retained the same direction of difference, providing limited directional consistency for part of the discovery pattern. The distinction between recurring community-level variation and less stable taxon-level findings is itself informative for designing future studies of CSF-associated microbial signals in ALS.

This taxonomic instability is particularly relevant because CSF is an extremely low-biomass matrix. Sequence-based analyses of such specimens are susceptible to reagent-derived DNA, environmental background, stochastic sampling, and differences in effective microbial template abundance [15–17]. Some genera detected here, including *Acinetobacter* and *Corynebacterium*, have been reported among possible reagent- or environment-associated background taxa [15–17]. In the present data, *Halomonas* and *Corynebacterium* were also detected in negative-control samples at comparable or higher relative abundance than in study samples, precluding strong biological attribution of correlations involving these genera. The 2bRAD-M strategy was selected because it can profile host-rich, low-biomass samples without prior host-DNA depletion, with taxonomic assignment based on computational matching of microbial-specific restriction tags [19,21]. Nevertheless, sequence detection cannot demonstrate that microbial DNA originated from viable or resident microorganisms. We therefore use “CSF-associated microbial signals” rather than imply a stable endogenous CSF microbiome. Within these constraints, the recurring community-level pattern supports further study using absolute microbial quantification, more extensive process controls, and targeted confirmation.

The most consistent molecular finding was the CSF metabolic phenotype. Previous NMR and high-resolution mass spectrometry studies have identified ALS-associated CSF metabolic signatures [22,23], and plasma studies have reported abnormalities in mitochondrial, oxidative-stress, amino acid, and lipid-related metabolism [24,25]. Here, decreased candidate metabolic features predominated in both independent cohorts. Although the strong directional and effect-size concordance among features selected as candidates in both datasets is descriptive and potentially optimistic, a stringent core of 21 annotated features independently achieved FDR-adjusted $q < 0.05$ in each cohort and changed in the same direction. Convergence across 18 FDR-significant KEGG pathways further supports a coordinated ALS-associated CSF metabolic phenotype rather than a finding driven by one isolated molecular marker.

Several replicated metabolites warrant particular attention. Glycine and L-cysteine were decreased in ALS in both cohorts and remained significant after multiple-testing correction.

Glycine participates in inhibitory neurotransmission, one-carbon metabolism, and broader amino acid homeostasis. Abnormal glycine handling in motor neuron disease has been reported previously, although earlier physiological studies did not necessarily show the same direction of steady-state change observed here, emphasizing the potential influence of analytical platform, disease stage, and sampling context [31]. L-cysteine is closely connected to glutathione synthesis and cellular redox buffering. The cysteine–glutathione axis is biologically relevant to ALS given the established involvement of oxidative stress and redox dysregulation in motor neuron degeneration [6,7]. A pharmacokinetic study further demonstrated that the cysteine prodrug Procysteine can enter and accumulate in CSF after intravenous and oral administration [32]. Thus, the replicated reduction in L-cysteine, together with enrichment of cysteine and methionine metabolism and glutathione-related pathways in the validation dataset, provides a coherent biochemical signal. Nevertheless, these observations do not establish impaired glutathione synthesis or demonstrate that augmentation of cysteine availability would modify ALS progression.

L-kynurenine, L-phenylalanine, and bilirubin also showed concordant decreases across cohorts, although their FDR-adjusted values in the validation cohort were above 0.05. The kynurenine pathway is particularly relevant to neurodegenerative biology because tryptophan-derived kynurenine metabolites participate in immune regulation, glutamatergic signaling, and redox balance. However, the biological consequences of changes in individual kynurenine-pathway metabolites depend strongly on pathway branch, cellular source, and biological compartment, and the present data do not allow pathway flux to be inferred from L-kynurenine abundance alone. Bilirubin likewise has antioxidant properties but may also reflect systemic, hematological, or hepatic influences, illustrating the difficulty of attributing a CSF metabolite change to a single CNS-specific mechanism. These metabolites should therefore be regarded as directionally replicated components of the ALS-associated metabolic phenotype rather than direct mechanistic markers.

A feature putatively annotated as Hymexazol N-glucoside was notable because it showed increased abundance in ALS in both cohorts and was positively associated with ALSFRS-R score in the discovery cohort. The combination of cross-cohort case–control significance and an association with functional status makes this feature potentially interesting for follow-up. Nevertheless, its biological origin in human CSF is uncertain, and the ALSFRS-R association was not independently replicated. Confirmation with an authentic chemical standard and targeted mass spectrometry is essential before this annotation is interpreted biologically or considered as a candidate biomarker. More broadly, statistical reproducibility of an untargeted feature does not substitute for analytical confirmation of metabolite identity.

Pathway analysis strengthened the evidence for coordinated metabolic dysregulation. Both the discovery and validation datasets contained 28 KEGG pathways that remained significant

after FDR correction, with 18 pathways significant in both cohorts. Among the biologically relevant recurrent terms were D-amino acid metabolism, alanine, aspartate and glutamate metabolism, taurine and hypotaurine metabolism, aminoacyl-tRNA biosynthesis, ABC transporters, glyoxylate and dicarboxylate metabolism, arginine biosynthesis, one-carbon metabolism, and neuroactive ligand–receptor interaction. The validation cohort additionally showed enrichment of cysteine and methionine metabolism, glycine, serine and threonine metabolism, glutathione metabolism, and sulfur metabolism. This pattern is consistent with previous metabolomic evidence that ALS involves broad perturbations of amino acid, bioenergetic, redox, and lipid-related metabolism rather than disruption of a single isolated biochemical pathway [22–25]. Reactome analysis provided complementary but less uniformly reproducible findings. Five transport-related Reactome terms remained FDR-significant in the discovery cohort, whereas the validation cohort showed nominal enrichment of biological oxidation, sterol, bile acid, carnitine, cytochrome P450-associated, and lipoprotein-related processes, none of which survived FDR correction. Accordingly, the KEGG results provide stronger evidence for cross-cohort pathway convergence, whereas the validation-cohort Reactome findings should remain exploratory.

The exploratory analyses of functional status provide an additional clinical perspective. The JAAYBP01/JADGOP01-related microbial signal was negatively associated with ALSFRS-R score, whereas Hymexazol N-glucoside showed a positive association. Because lower ALSFRS-R scores reflect greater functional impairment, these findings suggest that selected CSF molecular features may vary with functional status within ALS. However, these analyses were based on small numbers of patients and were exploratory rather than prespecified prognostic analyses. The JAAYBP01/JADGOP01-related feature was not identifiable in the validation dataset, and threonylserine, another nominally associated metabolite, was unavailable for direct cross-cohort validation. Furthermore, a cross-sectional ALSFRS-R measurement cannot distinguish whether a molecular feature reflects current functional status, disease duration, rate of progression, or another correlated clinical characteristic. Longitudinal sampling with prospectively defined progression outcomes will therefore be required before these associations can be interpreted as prognostic.

Integrated analysis of paired microbiome and metabolome data identified 16 nominal cross-omic associations in the discovery cohort, although none remained significant after BH correction across the 60 tested genus–metabolite pairs. These pooled analyses linked *Nesterenkonia*, *Halomonas*, *Corynebacterium*, *Pelomonas*, *Pseudoxanthomonas*, and *Nocardioide*s with metabolites related to amino acid, energy, redox, lipid, and carnitine metabolism. Biological plausibility is supported in principle by experimental evidence that microbial perturbations can influence host metabolic phenotypes in ALS models and by human studies demonstrating concurrent microbiome and metabolic alterations [11,30].

Nevertheless, the absence of FDR-significant discovery-cohort correlations, the pooled diagnostic groups, and negative-control overlap for some genera emphasize that these findings are exploratory.

Cross-cohort comparison provided partial directional and statistical support for selected cross-omic associations. Of the 16 nominal discovery-cohort associations, 14 showed the same direction in validation, eight remained nominally significant, and five were BH-significant within the validation 60-pair matrix. Across all 60 validation comparisons, 18 were nominally significant and nine were BH-significant. However, the complete association pattern was not stable, and the five discovery-to-validation associations that were BH-significant in validation involved *Halomonas* or *Corynebacterium*, both detected in negative controls. At the metabolite level, L-cysteine and glycine showed the strongest case–control replication; L-kynurenine, L-phenylalanine, and bilirubin were directionally concordant with weaker validation significance, whereas L-aspartic acid changed in opposite directions. The results therefore nominate a limited set of relationships for targeted investigation rather than validate a unified microbiome–metabolome network.

The partial consistency of selected correlations does not establish causality or a directional microbiome-to-metabolome mechanism. Because correlations were calculated across ALS and controls combined without adjustment for diagnostic group, shared case–control shifts may produce or amplify an apparent association even when no comparable within-group relationship exists. Correlations may also reflect systemic metabolism, immune activation, blood–CSF barrier function, medication exposure, nutrition, technical background, or compositional dependencies in relative-abundance data. PICRUSt2 does not resolve these issues because it infers functional potential from taxonomic composition rather than directly measuring microbial genes or activity; no predicted function remained significant after FDR correction in validation. The present results therefore define testable hypotheses concerning covariation between CSF-associated microbial signals and metabolic alterations.

The study's central strength is its independent assessment of two molecular layers measured in the same CSF compartment. Paired microbial and metabolomic profiles allowed exploratory cross-omic analysis, while the second cohort provided an opportunity to test whether observations recurred beyond the discovery samples. The 2bRAD-M strategy enabled profiling of host-rich, low-biomass material [19,21], and the validation design identified reproducible metabolic features and pathway-level convergence while revealing the greater heterogeneity of microbial signals. Applying the same selected genus–metabolite matrix across cohorts also allowed its directional consistency and microbiological limitations to be assessed explicitly.

Several limitations should be acknowledged. Both cohorts were relatively small and originated from a single center, limiting power and generalizability. Controls had

heterogeneous non-ALS neurological disorders rather than being healthy volunteers; this clinically pragmatic design may add molecular heterogeneity. CSF is an extremely low-biomass matrix, detection of microbial DNA does not demonstrate viability or colonization, and absolute microbial load was not independently quantified [15–17]. Some taxa in the cross-omic results were also present in negative controls. Homogeneity of multivariate dispersion was not formally assessed, so the PERMANOVA findings cannot distinguish centroid-location effects from dispersion effects. Cross-omic correlations were calculated across diagnostic groups without group adjustment and may partly reflect common case–control differences. The cross-sectional design precluded temporal inference. Some metabolite annotations could not be matched between cohorts and several require confirmation with authentic standards and targeted methods. Diet, detailed medication exposure, nutritional status, systemic metabolic disease, and preanalytical variation could not be comprehensively modeled. Finally, the feature-concordance analysis was conditioned on candidate selection in both cohorts and should not be interpreted as unbiased out-of-sample performance.

In conclusion, this cross-cohort multi-omics study identifies a reproducible ALS-associated CSF metabolic phenotype, including 21 directionally concordant, FDR-significant annotated features and 18 convergent KEGG pathways. The repeated decreases in glycine and L-cysteine, together with recurring amino acid-related and other metabolic pathway alterations, highlight coordinated biochemical changes relevant to the molecular characterization of ALS. Community-level microbial variation recurred across cohorts despite unchanged alpha diversity, while individual taxa and microbiome–metabolome relationships remained more heterogeneous and exploratory. By distinguishing robust metabolic findings from microbial signals requiring further confirmation, these results provide a focused basis for targeted biochemical and microbiological studies of the ALS-associated CSF environment.

Ethics approval and consent to participate

This study was performed in accordance with the Declaration of Helsinki. The Peking University Third Hospital Medical Science Research Ethics Committee reviewed and approved the study protocol (project approval number M20251225; ethics approval document No. 2025-1015-01). We obtained written informed consent from all participants before inclusion in the study.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests related to this study.

Availability of data and materials

De-identified data supporting the findings of this study are available from the corresponding author upon reasonable request and with approval from the institutional ethics committee. Due to patient privacy and data protection regulations, raw data are not publicly available.

Author contributions

Yinmei Zhang conceived and designed the study, performed the microbiome and metabolomics data analyses, interpreted the results, and drafted the manuscript. Nan Yang extracted data from the clinical databases and laboratory information system (LIS) and conducted statistical analyses. Ang Li was mainly responsible for sample collection and data collection. Jiajia Zheng critically reviewed and revised the manuscript. All authors approved the final version.

Funding

This study was supported by the Clinical Key Project Incubation Program of Peking University Third Hospital (Grant No. BYSYZD2025021).

Acknowledgements

Not applicable.

References

1. Hardiman O, Al-Chalabi A, Chio A, Corr EM, Logroscino G, Robberecht W, et al. Amyotrophic lateral sclerosis. *Nat Rev Dis Primers*. 2017;3:17071. doi:10.1038/nrdp.2017.71.
2. Kiernan MC, Vucic S, Talbot K, McDermott CJ, Hardiman O, Shefner JM, et al. Improving clinical trial outcomes in amyotrophic lateral sclerosis. *Nat Rev Neurol*. 2021;17(2):104–118. doi:10.1038/s41582-020-00434-z.
3. Taylor JP, Brown RH Jr, Cleveland DW. Decoding ALS: from genes to mechanism. *Nature*. 2016;539(7628):197–206. doi:10.1038/nature20413.
4. Brown RH Jr, Al-Chalabi A. Amyotrophic lateral sclerosis. *N Engl J Med*. 2017;377(2):162–172. doi:10.1056/NEJMr1603471.
5. van Es MA, Hardiman O, Chio A, Al-Chalabi A, Pasterkamp RJ, Veldink JH, et al. Amyotrophic

- lateral sclerosis. *Lancet*. 2017;390(10107):2084–2098. doi:10.1016/S0140-6736(17)31287-4.
6. Muyderman H, Chen T. Mitochondrial dysfunction in amyotrophic lateral sclerosis—a valid pharmacological target? *Br J Pharmacol*. 2014;171(8):2191–2205. doi:10.1111/bph.12476.
 7. D’Ambrosi N, Cozzolino M, Carri MT. Neuroinflammation in amyotrophic lateral sclerosis: role of redox (dys)regulation. *Antioxid Redox Signal*. 2018;29(1):15–36. doi:10.1089/ars.2017.7271.
 8. Sun J, Huang T, Debelius JW, Fang F. Gut microbiome and amyotrophic lateral sclerosis: a systematic review of current evidence. *J Intern Med*. 2021;290(4):758–788. doi:10.1111/joim.13336.
 9. Boddy SL, Giovannelli I, Sassani M, Cooper-Knock J, Snyder MP, Segal E, et al. The gut microbiome: a key player in the complexity of amyotrophic lateral sclerosis (ALS). *BMC Med*. 2021;19(1):13. doi:10.1186/s12916-020-01885-3.
 10. Brenner D, Hiergeist A, Adis C, Mayer B, Gessner A, Ludolph AC, et al. The fecal microbiome of ALS patients. *Neurobiol Aging*. 2018;61:132–137. doi:10.1016/j.neurobiolaging.2017.09.023.
 11. Zeng Q, Shen J, Chen K, Zhou J, Liao Q, Lu K, et al. The alteration of gut microbiome and metabolism in amyotrophic lateral sclerosis patients. *Sci Rep*. 2020;10(1):12998. doi:10.1038/s41598-020-69845-8.
 12. Wright ML, Fournier C, Houser MC, Tansey M, Glass J, Hertzberg VS. Potential role of the gut microbiome in ALS: a systematic review. *Biol Res Nurs*. 2018;20(5):513–521. doi:10.1177/1099800418784202.
 13. Niccolai E, Di Pilato V, Nannini G, Baldi S, Russo E, Zucchi E, et al. The gut microbiota-immunity axis in ALS: a role in deciphering disease heterogeneity? *Biomedicines*. 2021;9(7):753. doi:10.3390/biomedicines9070753.
 14. Chen X, Hu Y, Cao Z, Liu Q, Cheng Y. Cerebrospinal fluid inflammatory cytokine aberrations in Alzheimer’s disease, Parkinson’s disease and amyotrophic lateral sclerosis: a systematic review and meta-analysis. *Front Immunol*. 2018;9:2122. doi:10.3389/fimmu.2018.02122.
 15. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol*. 2014;12:87. doi:10.1186/s12915-014-0087-z.
 16. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome*. 2018;6(1):226. doi:10.1186/s40168-018-0605-2.
 17. Zinter MS, Mayday MY, Ryckman KK, Jelliffe-Pawlowski LL, DeRisi JL. Towards precision quantification of contamination in metagenomic sequencing experiments. *Microbiome*. 2019;7(1):62. doi:10.1186/s40168-019-0678-6.
 18. Wang S, Meyer E, McKay JK, Matz MV. 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nat Methods*. 2012;9(8):808–810. doi:10.1038/nmeth.2023.
 19. Sun Z, Huang S, Zhu P, Lam T, Zhao H, Lv J, et al. Species-resolved sequencing of low-biomass or degraded microbiomes using 2bRAD-M. *Genome Biol*. 2022;23(1):36. doi:10.1186/s13059-021-02576-9.

20. Sun Z, Liu J, Zhang M, Wang T, Huang S, Weiss ST, et al. Removal of false positives in metagenomics-based taxonomy profiling via targeting Type IIB restriction sites. *Nat Commun*. 2023;14(1):5321. doi:10.1038/s41467-023-41099-8.
21. Jiang Y, Liu J, Zhang Y, Zhou L, Kao E, Hou S, et al. High-resolution microbiome analysis of host-rich samples using 2bRAD-M without host depletion. *NPJ Biofilms Microbiomes*. 2025;11(1):223. doi:10.1038/s41522-025-00851-2.
22. Blasco H, Corcia P, Pradat PF, Bocca C, Gordon PH, Veyrat-Durebex C, et al. Metabolomics in cerebrospinal fluid of patients with amyotrophic lateral sclerosis: an untargeted approach via high-resolution mass spectrometry. *J Proteome Res*. 2013;12(8):3746–3754. doi:10.1021/pr400376e.
23. Gray E, Larkin JR, Claridge TDW, Talbot K, Sibson NR, Turner MR. The longitudinal cerebrospinal fluid metabolomic profile of amyotrophic lateral sclerosis. *Amyotroph Lateral Scler Frontotemporal Degener*. 2015;16(7–8):456–463. doi:10.3109/21678421.2015.1053490.
24. Lawton KA, Cudkowicz ME, Brown MV, Alexander D, Caffrey R, Wulff JE, et al. Biochemical alterations associated with ALS. *Amyotroph Lateral Scler*. 2012;13(1):110–118. doi:10.3109/17482968.2011.619197.
25. Goutman SA, Guo K, Savelieff MG, Patterson A, Sakowski SA, Habra H, et al. Metabolomics identifies shared lipid pathways in independent amyotrophic lateral sclerosis cohorts. *Brain*. 2022;145(12):4425–4439. doi:10.1093/brain/awac025.
26. Douglas GM, Maffei VJ, Zaneveld JR, Yurgel SN, Brown JR, Taylor CM, et al. PICRUSt2 for prediction of metagenome functions. *Nat Biotechnol*. 2020;38(6):685–688. doi:10.1038/s41587-020-0548-6.
27. Smith CA, Want EJ, O’Maille G, Abagyan R, Siuzdak G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem*. 2006;78(3):779–787. doi:10.1021/ac051437y.
28. Tsugawa H, Cajka T, Kind T, Ma Y, Higgins B, Ikeda K, et al. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat Methods*. 2015;12(6):523–526. doi:10.1038/nmeth.3393.
29. Nicholson K, Bjornevik K, Abu-Ali G, Chan J, Cortese M, Dedi B, et al. The human gut microbiota in people with amyotrophic lateral sclerosis. *Amyotroph Lateral Scler Frontotemporal Degener*. 2021;22(3–4):186–194. doi:10.1080/21678421.2020.1828475.
30. Blacher E, Bashiardes S, Shapiro H, Rothschild D, Mor U, Dori-Bachash M, et al. Potential roles of gut microbiome and metabolites in modulating ALS in mice. *Nature*. 2019;572(7770):474–480. doi:10.1038/s41586-019-1443-5.
31. Lane RJ, Bandopadhyay R, de Bellerocche J. Abnormal glycine metabolism in motor neurone disease: studies on plasma and cerebrospinal fluid. *J R Soc Med*. 1993;86(9):501–505.
32. Cudkowicz ME, Sexton PM, Ellis T, Hayden DL, Gwilt PR, Whalen J, et al. The pharmacokinetics and pharmacodynamics of Procysteine in amyotrophic lateral sclerosis. *Neurology*. 1999;52(7):1492–1494. doi:10.1212/WNL.52.7.1492.

Supplementary Files

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

- [Supplementarymaterials.docx](#)