Research Square

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

Stool consistency in IBS: a layer-wise multi-omics comparison

Yi Sui

yisui777@163.com

Chengdu No.7 High School, International Department https://orcid.org/0009-0006-3612-9633

Research Article

Keywords: irritable bowel syndrome, layer-wise multi-omics, Bristol stool consistency, short-chain fatty acids, subject-grouped cross-validation, label-blind transductive stability selection, held-out zero-shot validation, Mendelian randomization

Posted Date: November 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11496080/v2

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

Additional Declarations: The authors declare no competing interests.

ORIGINAL RESEARCH · MULTI-OMICS & GUT MICROBIOME

Stool consistency in IBS: a layer-wise multi-omics comparison

Yi Sui

ABSTRACT

Irritable bowel syndrome (IBS) is heterogeneous, and taxonomy-based subtyping has limited subject-level predictive value. In the Cell 2020 cohort (75 subjects; 441 Bristol-labelled samples) and an external LC-MS cohort, we compared multiple omics layers for Bristol stool consistency under subject-grouped cross-validation. In the primary measured-Bristol subset (n=235), the NMR metabolome had the highest point-estimate out-of-fold R^2 (0.247; 95% CI 0.097–0.327; ρ=0.510), but a paired comparison with genus on the measured common subset (n=127) did not establish superiority (ΔR^2 +0.186; 95% CI −0.010 to 0.404). A five-metabolite panel transferred externally (AUC 0.634), but added only +0.051 over age/sex/BMI (95% CI −0.029 to 0.132). Sample-wise centering increased NMR R^2 to 0.211–0.215, indicating a contribution from global solute concentration. Two-sample Mendelian randomization found no causal effects of 27 genera. Direct metabolic measurement captures stool consistency, but cross-layer superiority and incremental clinical value remain uncertain.

1 Introduction

1.1 Background

IBS is one of the most common disorders of gut–brain interaction [1], and its clinical subtypes (IBS-D, diarrhea-predominant / IBS-C, constipation-predominant) and symptom severity (IBS-SSS) are highly heterogeneous. Previous studies based on fecal microbiota have generally reported that IBS can be separated from healthy controls (AUC 0.7–0.9) [1,3], but whether this "separability" reflects a generalizable biological mechanism or is dominated by batch effects, subject effects, and repeated sampling has long been debated. For "precision biological subtyping," the core question is not "can we classify?" but which omics layer carries the most phenotypic information that generalizes across subjects and cohorts.

1.2 This study

We focused on Bristol stool consistency, a continuous phenotype with clear physiological meaning, performed a unified benchmark comparison of eight feature layers, further examined the coherence of the "functional potential–metabolite pool–phenotype" chain through cross-layer association analysis, and finally converged on transferable core features through external-cohort held-out zero-shot validation and label-blind transductive stability selection. Contributions:

- Constructed a continuous Bristol target validated in the clinical direction (IBS-D>healthy>IBS-C), covering 441 samples/67 subjects;
- Compared 8 omics layers (8 layers × 2 tasks) under a unified grouped cross-validation framework;
- Found that in Bristol regression, the NMR metabolome had a higher point-estimate OOF R^2 than taxonomy and functional layers; combined modeling showed no gain when unfiltered, and after filtering at the same prevalence was not significantly different from NMR;

Through cross-layer association and within-subtype robustness tests, provided mechanistic interpretation and the methodological conclusion “potential ≠ activity”;

Completed cross-cohort held-out zero-shot validation and label-blind transductive stability selection (source-domain closed-loop protocol), identifying 5 conditionally stable features (source-stabilized panel: lysine, tyrosine, uracil, glucose, leucine; under fully nested analysis, unconditional robustness concentrated on lysine/tyrosine)—not significantly different from the 11-shared-feature set in zero-shot discrimination, and significantly better than the 16-feature full model.

2 Materials and methods

2.1 Cohort and clinical labels

Source: Cell 2020 longitudinal multi-omics IBS cohort [2] (Mars *et al.*, Cell 2020; ENA PRJEB37924), with metagenomic abundances taken from GMrepo [11] processed data; sample–phenotype mapping was fully consistent with disease_stage (C↔IBS-C, D↔IBS-D, H↔healthy).

Scale: 75 subjects and 464 clinically labeled stool samples (IBS-D 169, IBS-C 142, healthy 153). Layer-specific availability differed: genus/species 457 (D169/C137/H151), KO/KEGG 459 (D169/C140/H150), NMR 445 (D168/C133/H144), bile acids 447, tryptophan 274, and NMR+genus 440.

Severity: IBS-SSS [7] available subset 250 samples/51 subjects, median 181.5 (binary threshold: median split).

Data hygiene: the genus matrix removed the unclassified artifact (mean 21.4%, non-zero in 99.5% of samples); the species matrix removed the -1 artifact (mean 11.9%). All previous genus-level conclusions were rerun on this basis and remained unchanged.

External validation cohort: Jacobs *et al.*, Microbiome 2023 [3] (PRJNA812699), 368 samples × 601 untargeted LC-MS metabolites, BH subtypes (Constipation 89 / Diarrhea 84 / Mixed 37 / Unspecified 19), no SCFA measurements and no matched taxonomic matrix.

Sample flow: 464 labeled samples → genus/species 457 (7 missing taxonomy), KO/KEGG 459 (5 missing the functional layer), and NMR 445 (19 missing NMR); Bristol was available for 441 (67 subjects) → NMR∩Bristol 424 (67 subjects). The 17 Bristol samples without NMR comprised w3 7, w7 4, and subject-mean fallback 6; another 21 NMR samples had no Bristol, and 2 samples had neither. KO/KEGG and genus/species are not the same subset (intersection 453); the three-class task does not require Bristol, so all 459 KO/KEGG samples were usable for classification. Jacobs 368 → main evaluation 210 (C 89/D 84/M 37), D vs C sensitivity 173 (C 89/D 84). See Figure 15.

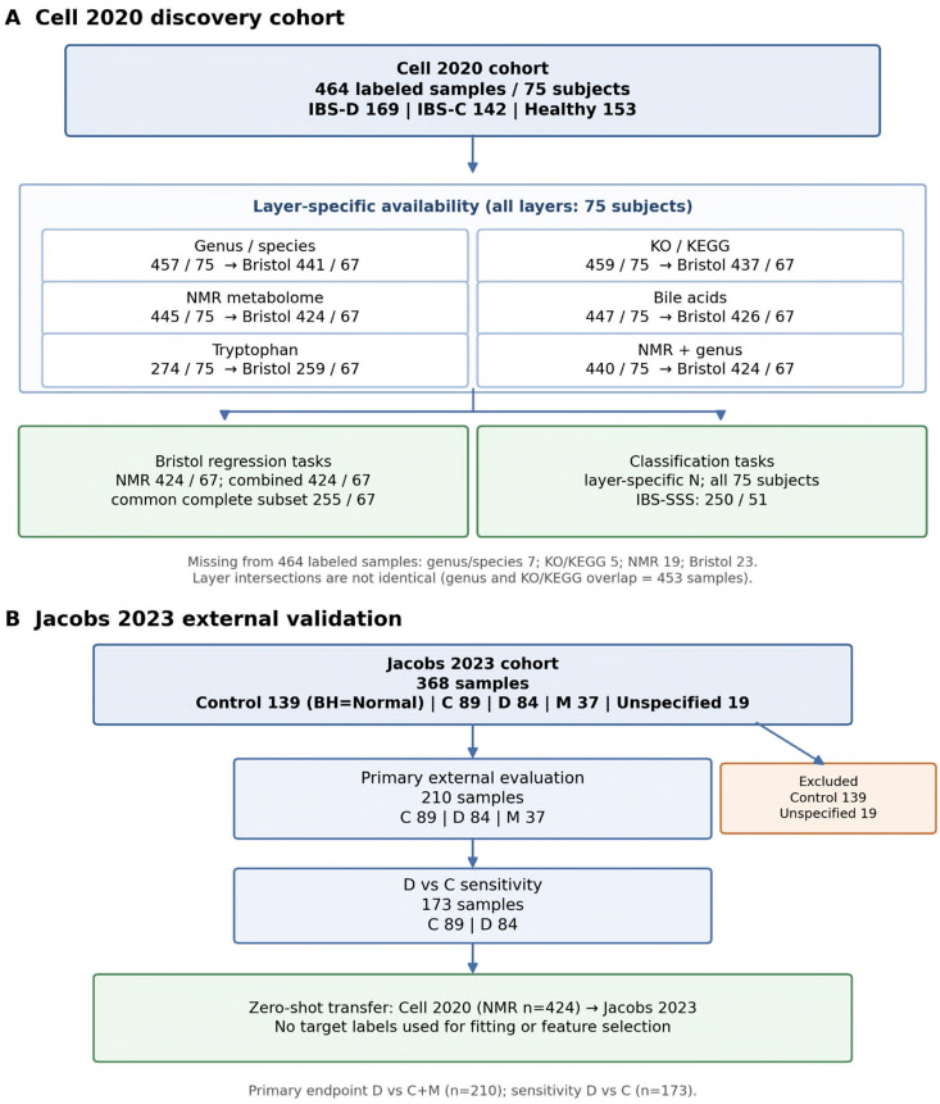


Figure 15. Sample flow and layer-specific availability. **A:** Cell 2020 discovery cohort (464 labeled samples; genus/species 457, KO/KEGG 459, NMR 445; Bristol 441/67, NMR∩Bristol 424/67; Bristol-missing 23 = 21 NMR samples without Bristol + 2 with neither); **B:** Jacobs 2023 external validation (368→210 main evaluation; 173 D vs C sensitivity).

2.2 Bristol target construction

The primary analysis uses the measured Bristol union (samples with a ±3-day or ±7-day diary match; n=235). The all-sample target including subject-mean fallback (n=424) is retained as a sensitivity analysis. Using stool diary collection dates (Excel serial format) for alignment, we constructed the Bristol score [6]: ±3-day window (110 samples) → ±7-day window (136 samples) → subject-specific mean fallback (195 samples), totaling 441 samples/67 subjects (strictly counted by sample-source hierarchy; stratified sensitivity in §3.4/Figure 9). Direction validation: IBS-D 4.92 (SD 0.79) > healthy 3.94 > IBS-C 2.93 (SD 1.10), fully consistent with clinical priors.

2.3 Omics layers

Table 1. Multi-omics feature layers, dimensionality, sample coverage, and preprocessing protocols.

Layer	Dimensions	Samples	Preprocessing
Genus taxonomy	144 (67–69 after filtering)	457	CLR (pseudo-count method consistent with the original pipeline)
Species taxonomy	328 (143–145 after filtering)	457	CLR
SHOGUN[9] KO	7332 (5298–5310 after prev≥0.3)	459	CLR
KEGG pathways	375 (aggregated; 327 after prev≥0.5)	459	KO assigned via rest.kegg.jp/link/pathway, summed, CLR
NMR metabolome	19	445	log1p
Bile acids	12	447	log1p (30.5% missing, min/2 imputation)
Tryptophan metabolites	12	274	log1p ('ND' treated as missing)
NMR+genus combined	163 (unfiltered) / 87 (same filtering, sensitivity)	440 / 424	pre-transform concatenation (transform=none); sensitivity filtered the genus layer at prevalence=0.10 (68 genera+19 NMR)

Functional layer data came from Mendeley repository [24] (SHOGUN KO table, NMR, bile acids, tryptophan, cytokines, and other processed matrices).

2.4 Modeling and validation

All cross-validation splits were grouped by subject because subjects contributed multiple longitudinal samples; this prevented subject-level leakage.

Regression (Bristol): Ridge/ElasticNet; outer 5-fold subject-grouped (GroupKFold), inner tuning ($\alpha \in \{0.1,\ldots,10^4\}$); metrics: OOF R^2, OOF Spearman ρ; univariate Spearman+BH-FDR [13].

Classification (IBS-D/C/healthy): L2 logistic regression (C tuning); StratifiedGroupKFold (5 folds, by subject); metrics: balanced accuracy, macro-F1; univariate tests+BH-FDR.

Classification uncertainty: for each layer's OOF predictions, computed subject-cluster 2000× bootstrap 95% CI, subject-level label permutation p (NMR/genus/species/bile acids/tryptophan 200 permutations; KEGG 100; KO 20, with C fixed at 0.01 to control high-dimensional permutation cost), and 3×3 confusion matrices; the permutation p resolution floor is 1/(B+1) (≈0.005 for 200, ≈0.010 for 100, ≈0.048 for 20), and values at the floor should not be interpreted as exact p values; IBS-SSS was genus-only with the same 200-permutation suite.

Within-subtype: repeated Bristol regression separately within IBS-D (n=154) and IBS-C (n=128).

Bristol source sensitivity: recomputed the main model stratified by target source (±3d / ±7d / measured union / mean fallback) to test whether conclusions were driven by fallback samples (§3.4, Figure 9).

Cluster-level bootstrap: within IBS-D and IBS-C, ρ and R^2 were used to construct 95% CIs with subject-cluster 2000× bootstrap (§3.5, Figure 10).

External baselines: trained on Cell 2020 and evaluated held-out on Jacobs 2023, comparing R11, 200 random 11-feature sets, and age/BMI/sex demographic models; the primary AUC endpoint was D vs C+M, with D vs C as sensitivity; paired bootstrap 2000× tested ΔAUC.

Cross-layer association: 11 functional features (butyrate synthesis gene set, *bcd*, *adhE*, ko00650 butyrate metabolism, ko00280 BCAA degradation, ko00350 tyrosine metabolism, ko00310 lysine

degradation, ko00380 tryptophan metabolism, ko00120/ko00121 bile acid synthesis, BSH K01442) were Spearman-correlated with 16 significant NMR markers + 6 bile acid markers + Bristol (n=423), BH-FDR (denominator 253 tests).

Unified comparison protocol: cross-layer comparisons used the same preprocessing, subject-grouped folds, and Ridge/logistic tuning to produce OOF metrics. Unsupervised structural metrics were used only for cluster diagnostics and were not compared with supervised performance.

Software: Python 3.13.13, NumPy 2.4.6, pandas 3.0.3, SciPy 1.17.1, scikit-learn 1.8.0, Matplotlib 3.10.9; metagenomic annotation MetaPhlAn 3.0.14 [10] (GMrepo [11] processing pipeline).

Notation: ρ (rho) denotes the Spearman rank correlation coefficient, φ (phi) denotes the Shapley contribution, and p denotes the P value.

2.5 Source-domain closed-loop stability selection and held-out zero-shot evaluation (label-blind transductive)

Protocol overview (engineering hard isolation): the cross-cohort step was split into three unidirectional data flows: selector (reads only Cell features/labels and Jacobs unlabeled X, outputs a frozen selector) → held-out evaluation (reads only the frozen selector and Jacobs X; BH labels introduced only at the end of evaluation) → two-level nested sensitivity (pre-screening nested and fully nested). Target-domain labels had zero contact with any selection step (label-blind transductive feature selection).

Selection gates (prespecified): ① source-domain univariate stability π_uni_cell ≥ 0.8 (Cell 2020 grouped bootstrap, B=300, 50% subject sampling without replacement; frequency of Spearman $p<0.05$ with direction consistent with the full sample; π_multi is the Lasso non-zero selection frequency with fixed α within each grouped bootstrap resample, a secondary metric); ② platform shift KS < 0.15 (two-sample KS statistic between self-standardized log1p distributions of Cell and Jacobs; only X used, no labels. Note that domain shift is a composite shift—platform (NMR vs LC-MS), feature availability (Jacobs has no SCFA/no matched taxonomy), and sample processing/study protocol (longitudinal multiple time points vs cross-sectional, collection/storage/extraction differences) superimposed; KS can only provide an unlabeled diagnosis of composite magnitude, not decompose it into a single source, and "platform robustness" should be understood as robustness to composite shift); ③ the two groups were jointly measurable under the prespecified names. Features passing all three gates constituted the source-stabilized panel; the selector used GroupKFold(5) within Cell to select α and fit StandardScaler+Ridge, and the scaling parameters and coefficients were frozen and exported together with the feature set. An earlier label-informed determination is not used; all reported selection is label-blind.

Held-out zero-shot evaluation: the frozen selector was applied directly to 368 Jacobs samples (main evaluation subset n=210: C 89 / M 37 / D 84); metrics were AUC for Diarrhea vs C+M (same definition as previous analyses) and BH ordinal Spearman; sample-level paired bootstrap (B=2000) constructed 95% CIs for each model's AUC and the ΔAUC distribution, with p taken as the proportion of ΔAUC≤0; control models R11 (11 shared features) and F16 (16 features including imputed SCFAs) were both frozen from Cell 2020, without any target-domain tuning.

Two-level nested sensitivity: pre-screening nesting kept the gates fixed and re-ran univariate screening within each grouped resample; fully nested sensitivity recomputed the gates within resampling (outer subject bootstrap B_out=100; inner stability protocol B_in=100). Under fully nested resampling, all-gates selection frequencies were lysine 0.96, tyrosine 0.79, leucine 0.36, glucose 0.28,

and uracil 0.23. The panel is therefore described as conditionally stable; only lysine and tyrosine approach robust unconditional selection. No target-domain labels entered any selection step.

2.6 Two-sample Mendelian randomization (genus → IBS risk)

Exposure (instrumental variables): MiBioGen microbiome meta-analysis [4] (Kurilshikov *et al.*, Nat Genet 2021; 16–24 cohorts) full hit table MBG.allHits.p1e4.txt (122,110 rows / 211 taxa). We selected 27 candidate genera (covering study markers and GTDB homologous genera), with instrument threshold $p<10^{-5}$: 1,755 SNPs (F statistic 100% >10); because no individual-level LD panel was available, ±1 Mb distance pruning was used to approximate LD removal (retaining the lowest-p SNP within each window), yielding 313 independent SNPs.

Outcome: harmonized summary statistics from IBS GWAS GCST90016564 [5] (Eijsbouts *et al.* 2021; 486,601 European samples, 53,400 cases / 433,201 controls), matched by hm_rsid (305/313); allele harmonization included strand flipping (complementary pairs), and palindromic SNPs were conservatively removed; finally 259 SNPs entered estimation (4–16 per genus).

Estimation and sensitivity: primary analysis IVW [23] (random effects, delta-method SE); weighted median [22] (bootstrap 2000) and MR-Egger [21] (intercept test for horizontal pleiotropy) as sensitivity analyses; Cochran Q test for heterogeneity, leave-one-out robustness; BH-FDR on IVW p values for 27 genera. Separately, observational Spearman ρ between each genus and Bristol in this cohort/external cohort was summarized as a directional control.

STROBE-MR reporting: reported data sources, instrument selection, harmonization, primary analysis, sensitivity, and multiple correction according to the STROBE-MR [12] checklist; complete 27-genus results and item mapping in Appendix S2.

Data sources, result tables, and analysis workflow in §5.

3 Results

3.1 Clinical phenotype

The Bristol target aligned completely with subtypes (IBS-D 4.92 > healthy 3.94 > IBS-C 2.93), forming a credible continuous phenotype axis used as the unified target for all subsequent regression tasks.

3.2 Taxonomic layer: weak subtype signal, severity indistinguishable from chance

Table 2. Taxonomic layer out-of-fold classification performance for IBS subtype and symptom severity (IBS-SSS).

Task	Features	Result
IBS-D/C/healthy classification (genus)	69	balanced acc = 0.506±0.077; 95% CI [0.421, 0.595]; permutation p=0.005; macro-F1=0.508 (52/69 significant)
IBS-D/C/healthy classification (species)	145	balanced acc = 0.520±0.134; 95% CI [0.434, 0.611]; permutation p=0.005; macro-F1=0.519 (103/145 significant)
IBS-SSS binary (genus)	70	balanced acc = 0.500; 95% CI [0.420, 0.578]; permutation p=0.458; AUC=0.552; only 2/70 q<0.05 (*Dorea*, *Romboutsia*)

The taxonomic layer showed some signal relative to the chance baseline (0.333), but the magnitude was limited; the severity task was indistinguishable from chance (permutation p=0.458). Permutation resolution should be noted: with 200 permutations the p floor is 0.005, and the genus/species p=0.005 is at that floor; the KO p=0.095 is based on 20 permutations (floor ≈0.048) and is exploratory only. Eight-layer OOF performance, CIs, permutation p values, and confusion matrices are shown in Appendix S3.

3.3 Cross-layer comparison: point estimates favor the metabolome, but superiority over genus is not established (Figure 4)

Primary analysis (measured Bristol, n=235). The primary target uses only Bristol values measured within ±3 or ±7 days; subject-mean fallback samples are analysed separately. Per-layer measured-union estimates were:

Table 3. Primary measured-Bristol regression performance across omics layers (n = 235 union).

Layer	n	Features	OOF R^2	ρ
NMR	235	19	0.247	0.510
Genus	246	70	0.090	0.327
Species	246	145	0.107	0.373
Bile acids	237	12	0.047	0.268
Tryptophan	128	12	0.123	0.416
KO	244	5304	−0.029	0.166
KEGG	244	327	0.068	0.284
NMR+genus	235	163	0.180	0.447

Paired comparison on the measured common subset (n=127/59 subjects). Subject-cluster paired bootstrap (2,000 resamples) compared each layer with NMR on the same samples:

Table 4. Paired bootstrap comparison against the NMR metabolome on the measured common subset (n = 127 samples / 59 subjects).

Comparison	ΔR² (95% CI)	p(NMR > layer)
NMR − Genus	+0.186 (−0.010 to 0.404)	0.032
NMR − Species	+0.180 (0.028 to 0.330)	0.011
NMR − Bile acids	+0.205 (0.080 to 0.320)	0.003
NMR − Tryptophan	+0.078 (−0.125 to 0.272)	0.212
NMR − KO	+0.282 (0.119 to 0.422)	<0.001
NMR − KEGG	+0.310 (0.141 to 0.467)	<0.001
NMR − (NMR+genus)	−0.042 (−0.160 to 0.092)	0.754

NMR had the highest point estimate. The difference from genus was suggestive but not conventionally significant (one-sided p=0.032; two-sided p=0.064), whereas differences from species, bile acids, KO, and KEGG were clearer. The full-sample estimates below are retained as a sensitivity analysis.

Full-sample sensitivity (n=424; includes subject-mean fallback).

Table 5. Full-sample sensitivity analysis across omics layers (n = 424, including subject-mean fallback).

Feature layer	Dimensions	Bristol OOF R²	OOF ρ	Significant/total	Subtype balanced acc
NMR (19)	19	0.190	0.448	16/19	0.558±0.085
Genus (144)	67	0.089	0.309	22/67	0.506±0.077
Species (328)	143	0.043	0.261	54/143	0.520±0.134
Bile acids (12)	12	0.053	0.276	9/12	0.482±0.034
Tryptophan (12)	12	0.023	0.259	6/12	0.486±0.109
KO (7332)	5298	0.005	0.188	2349/5298	0.384±0.087
KEGG pathways (375)	327	0.036	0.235	274/327	0.436±0.088
NMR+genus combined (163)	163	0.171	0.438	62/163	0.503±0.099
NMR+genus combined (same filter, 87)	87	0.204	0.470	—	—

Interpretation: (i) the measured-union analysis supports the NMR layer as the strongest single-layer point estimate; however, the paired comparison with genus crossed zero, so the data do not establish that NMR is superior to taxonomy. (ii) The combined NMR+genus model was not significantly different from NMR alone. (iii) Functional-potential layers (KO/KEGG) were consistently weaker. (iv) The full-sample table is a sensitivity analysis and should not be presented as the primary result.

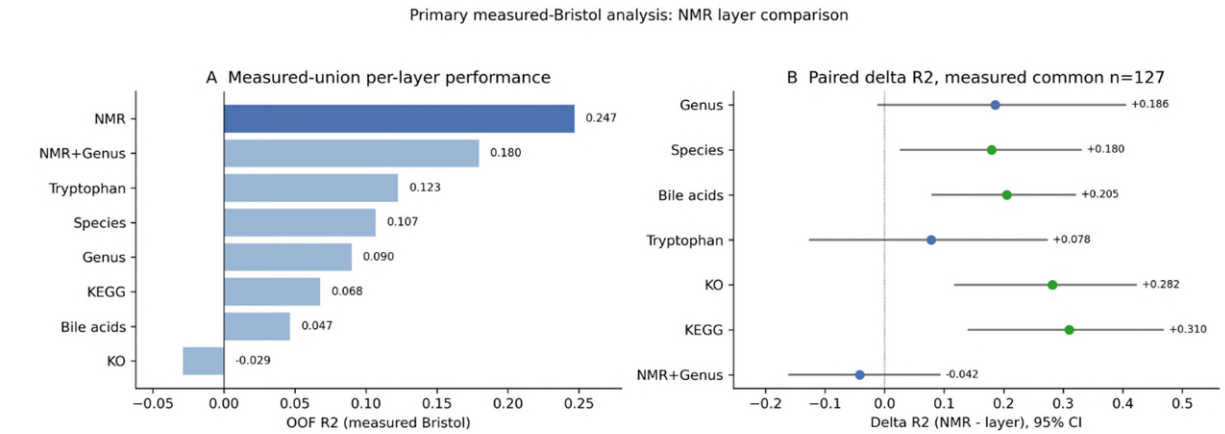


Figure 4. Primary measured-Bristol comparison and paired uncertainty. **A:** OOF R^2 for each layer on the measured Bristol union (n=235; per-layer n differs slightly by feature availability). **B:** paired ΔR^2 (NMR − layer) with 95% subject-cluster bootstrap CIs on the measured common subset (n=127/59). NMR has the highest point estimate; the difference from genus is not conventionally significant, while differences from species, bile acids, KO, and KEGG are clearer.

Common complete-subset sensitivity (Figure 14): On the common complete subset (268 samples/75 subjects; regression n=255/67), NMR remained the highest regression layer (OOF R^2=0.198; full-sample 0.190), followed by NMR+genus 0.095, tryptophan 0.069, bile acids 0.061, species 0.019, KEGG 0.006, genus 0.002, and KO −0.020. The relative ordering of mid-tier layers is composition-sensitive and should not be overinterpreted. In classification, the combined model was highest (0.587), followed by species 0.551, NMR 0.546, genus 0.546, bile acids 0.514, KEGG 0.487, KO 0.423, and tryptophan 0.408; NMR and genus were indistinguishable, so classification does not support a stable layer ranking.

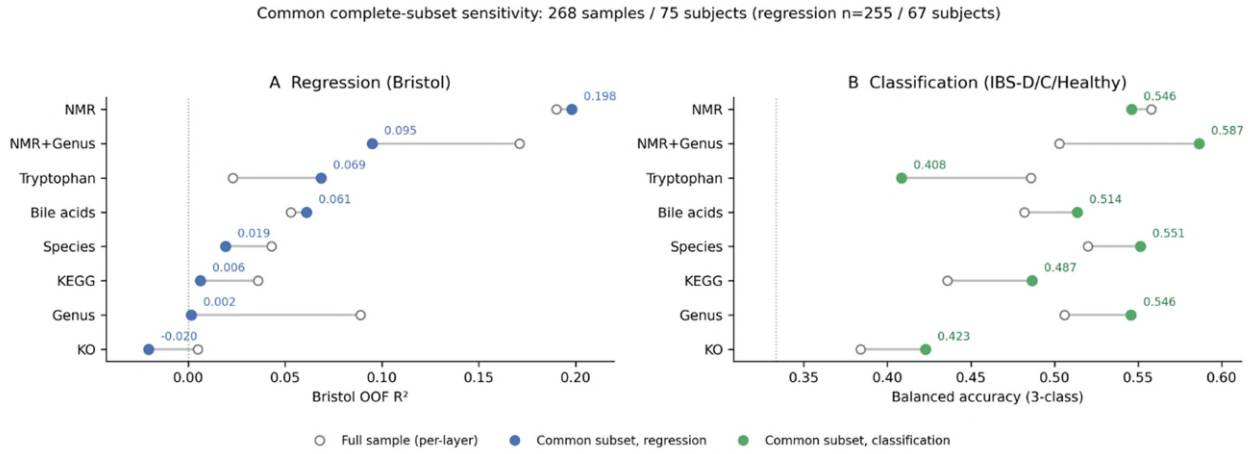


Figure 14. Common complete-subset sensitivity. **A:** Bristol regression OOF R^2 (open circles=full sample; filled circles=common complete subset n=255/67). **B:** three-class balanced accuracy (n=268/75). NMR remains highest for regression, but the paired measured-common analysis (§3.3) does not establish superiority over genus; classification ordering is not robust.

3.4 NMR markers: SCFA and amino acid profiles (Figures 1 and 2)

Sixteen of 19 NMR metabolites were significantly correlated with Bristol (BH-FDR), with a highly consistent direction spectrum (closer to loose stools, higher concentrations; except branched-chain SCFAs):

Table 6. Univariate Spearman correlations between NMR metabolites and Bristol stool consistency (BH-FDR corrected).

Marker	ρ	q	Interpretation
Acetate	+0.414	<1e-16	SCFA
Butyrate	+0.379	<1e-14	SCFA/FFAR2-3 pathway core
Lysine	+0.374	<1e-13	Proteolytic fermentation
Tyrosine	+0.311	<1e-12	Aromatic AA/tyramine precursor
Propionate	+0.297	<1e-11	SCFA
Succinate	+0.273	<1e-9	Facultative anaerobic fermentation marker
Uracil	+0.236	<1e-7	Nucleic acid degradation
Glucose	+0.234	<1e-7	Carbon-source delivery/absorption imbalance
Leucine	+0.232	<1e-7	BCAA
Valine	+0.198	1e-4	BCAA
Isoleucine	+0.187	2e-4	BCAA
Alanine, hypoxanthine, lactate (3 items)	0.12–0.16	<0.05	—
Isovalerate	−0.122	0.014	Branched SCFA
2-Methylbutyrate	−0.185	2e-4	Branched SCFA (only significant strong negative signal)

Note: The row "alanine, hypoxanthine, lactate (3 items)" is a merged row (univariate ρ 0.123–0.162, all q<0.05): alanine ρ=0.162, q=0.0012, assigned to the amino acid (AA) group in §3.4b (source of 6 significant features among 7 in that group); hypoxanthine ρ=0.156, q=0.0017 and lactate ρ=0.123, q=0.014 were assigned to the "other" group.

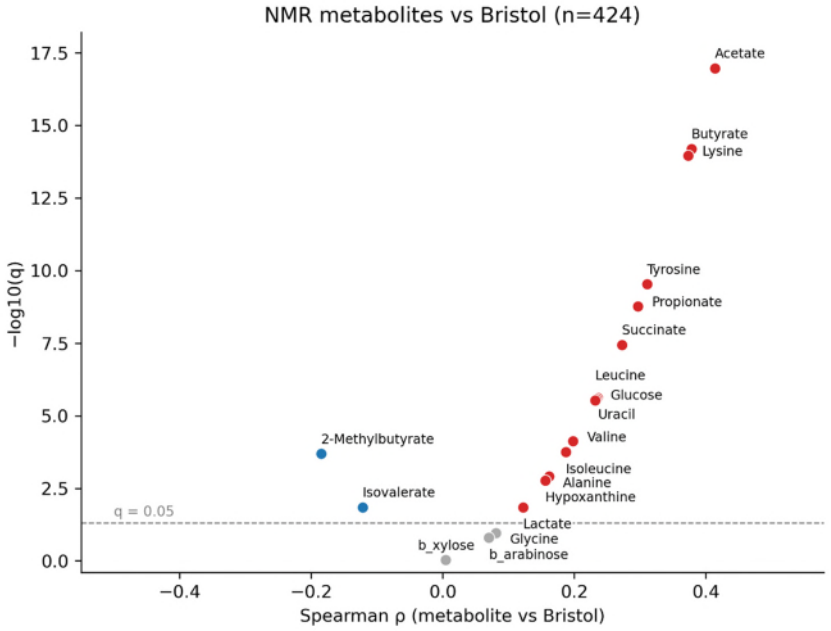


Figure 1. Volcano plot of NMR metabolite associations with Bristol (full-sample sensitivity, n=424). x-axis: Spearman ρ; y-axis: −log10(q); 16/19 metabolites had q<0.05; dashed line: q=0.05.

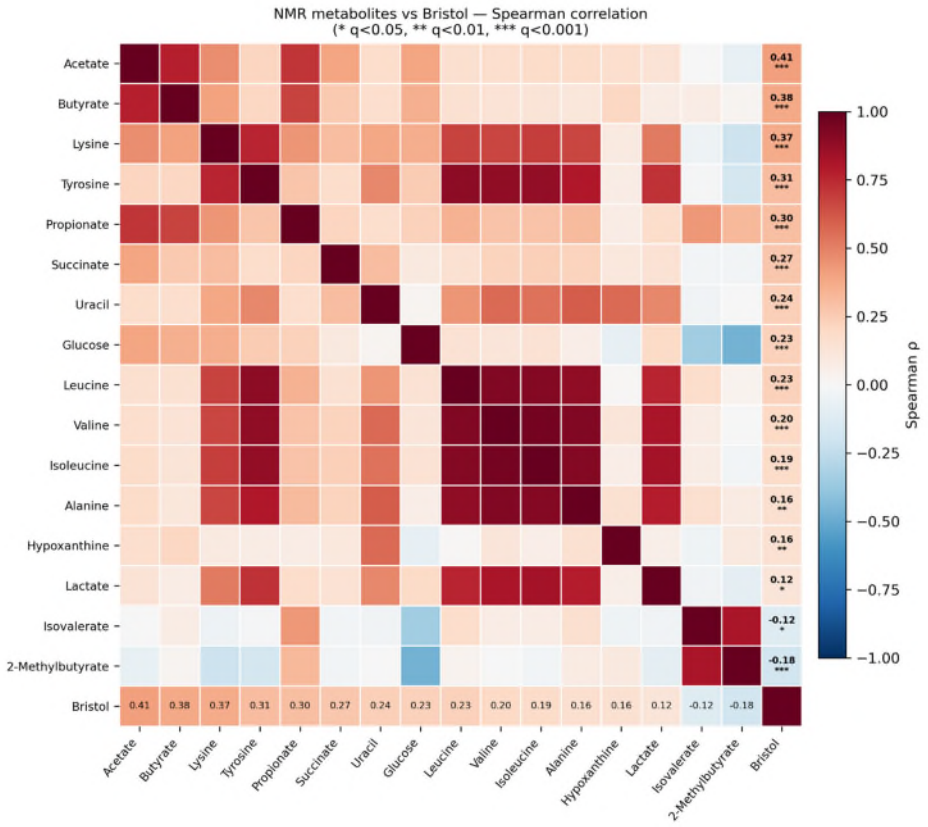


Figure 2. Spearman correlation heatmap of the 16 significant NMR markers × Bristol (full-sample sensitivity, n=424). Markers are ordered by correlation strength with Bristol; the Bristol column is annotated with ρ and significance (*q<0.05, **q<0.01, ***q<0.001).

Bile acid layer Top6 (all significant by BH-FDR): glycochenodeoxycholic acid GCDCA (+0.310), chenodeoxycholic acid CDCA (+0.309), deoxycholic acid-sulfate DCA-S (+0.305), cholic acid CA (+0.294), glycodeoxycholic acid GDCA (+0.272), and glycocholic acid GCA (+0.228)—primary bile acids and glycine-conjugated bile acids were positively correlated with loose stools, consistent with the continuous spectrum of bile acid diarrhea (BAD).

Sample-wise centering sensitivity: because the NMR matrix contains concentration-type measurements, we repeated the regression after sample-wise centering (subtracting each sample's mean log1p value) and after CLR transformation. OOF R^2 was 0.190 for log1p alone, 0.211 after sample-wise centering, and 0.215 after CLR (ρ=0.448, 0.483, and 0.484, respectively). The signal therefore is not solely a global solute-concentration effect, although the improvement after centering indicates that total concentration contributes to the log1p model.

Bristol source sensitivity (Figure 9): The Bristol target was stratified by source (NMR analysis set n=424; raw source counts 441 = w3 110 + w7 136 + fallback 195). Subject-cluster 2000× bootstrap: measured union n=235, R^2=0.247 [0.097, 0.327], ρ=0.510 [0.374, 0.617]; ±3d only n=103, R^2=0.257 [0.001, 0.390], ρ=0.536 [0.292, 0.700]; ±7d only n=132, R^2=0.217 [−0.035, 0.335], ρ=0.444 [0.217, 0.624]; fallback n=189, R^2=0.099 [−0.034, 0.175], ρ=0.358 [0.187, 0.505]; all n=424, R^2=0.190 [0.045, 0.283], ρ=0.448 [0.311, 0.558]. Interpretation: the measured-union subset supports the primary analysis; single-window estimates overlap, and the fallback subset is weaker and should be treated as sensitivity rather than primary evidence.

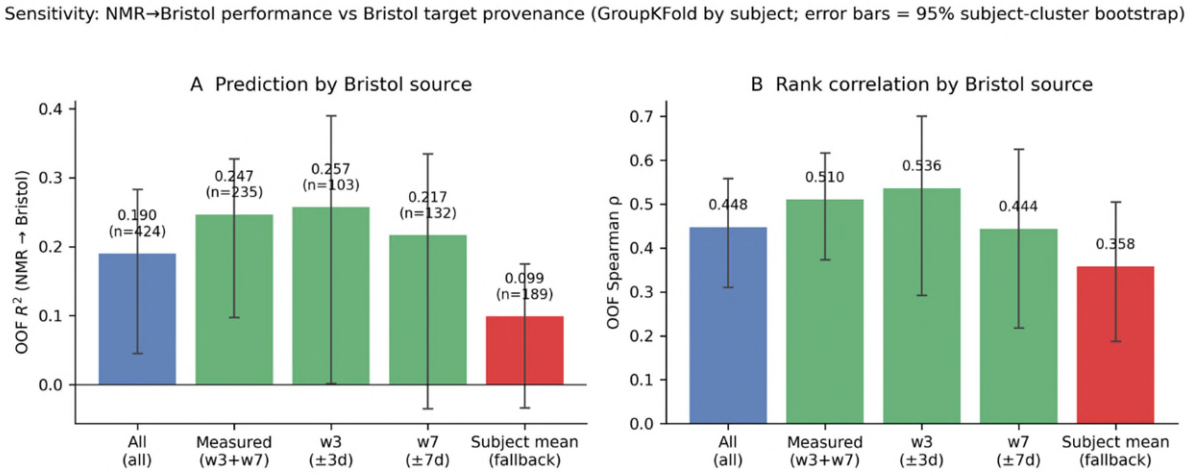


Figure 9. Bristol target source sensitivity. Measured union (n=235), ±3d (n=103), ±7d (n=132), subject-mean fallback (n=189), and all samples (n=424); error bars=subject-cluster 2000× bootstrap 95% CI. The primary analysis uses the measured union; fallback is sensitivity.

3.4b Internal decomposition of the NMR layer: Shapley contributions of SCFAs and amino acids (Figure 13)

Grouping and models: The 19 metabolites were divided into SCFA (5), amino acid (7), and other (7) groups. Seven OOF models were fitted under the unified protocol: full 0.190, SCFA-only 0.181, amino-acid-only 0.107, other-only 0.034, SCFA+amino-acid 0.208, SCFA+other 0.177, and amino-acid+other 0.111. A complete three-group Shapley decomposition gave φ_SCFA=0.128, φ_amino acid=0.057, φ_other=0.005. SCFA therefore remains the dominant within-NMR contribution, while the amino-acid contribution is smaller and the other group contributes little. The SCFA+AA model (0.208) was slightly higher than the full model (0.190), indicating that the other group adds little independent signal and may add modest noise. In the original full-sample permutation sensitivity, removing SCFA reduced R^2 by 0.076 (p=0.002) whereas removing amino acids reduced it by 0.016 (p=0.085); this permutation result is reported as supporting evidence, not as a substitute for the complete decomposition.

Measured-union sensitivity (n=235): the univariate direction spectrum was unchanged (Lysine ρ=0.424, Acetate 0.421, Butyrate 0.358, Tyrosine 0.355, Uracil 0.353; all q<0.001), and the complete Shapley decomposition gave φ_SCFA=0.107, φ_AA=0.063, φ_Other=0.076. SCFA remained the largest group, but the Other group was not negligible in this subset.

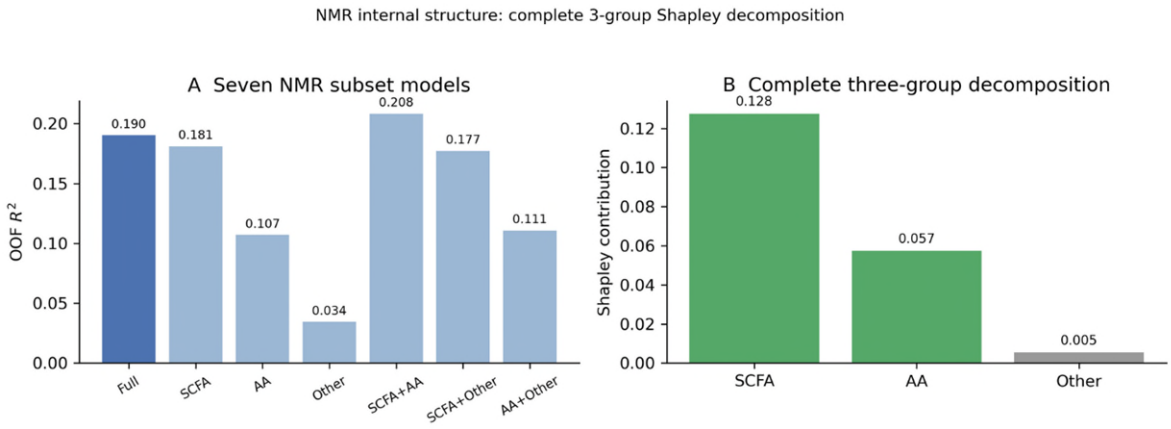


Figure 13. NMR group decomposition. The complete three-group Shapley decomposition is shown; the values are ϕ_SCFA=0.128, ϕ_AA=0.057, ϕ_other=0.005.

3.5 Predictive performance and within-subtype robustness (Figures 3 and 3b)

In the full-sample sensitivity analysis (including subject-mean fallback), the NMR layer contained 445 samples/75 subjects and Bristol regression used 424 samples/67 subjects. OOF predictions were R^2=0.190, ρ=0.448 (n=424; including IBS-D 154, IBS-C 128, healthy 142).

Within IBS-C: R^2=0.061, ρ=0.288 (5/19 significant; subject-cluster 2000× bootstrap: ρ 95% CI [+0.074, +0.483], P(ρ≥0)=0.996, R^2 95% CI [−0.180, +0.178])—the metabolite–Bristol coupling still held within a single subtype (rank correlation robust, magnitude CI wide).

Within IBS-D (exploratory): R^2=−0.048, ρ=−0.263. Subject-cluster 2000× bootstrap (n=154, 23 subjects): ρ 95% CI [−0.573, +0.128] crossing zero (P(ρ≥0)=0.083), and R^2 95% CI [−0.226, −0.001] entirely negative. Thus ρ was indistinguishable from zero, whereas the R^2 point estimate was negative with a CI on the negative side: the model had no predictive ability within this subtype (rather than a robust negative correlation). It is therefore registered only as an exploratory hypothesis-generating observation, and no unsupervised subgrouping was performed within IBS-D (DMM/hierarchical clustering tends to pseudo-converge in small samples) (Figure 10).

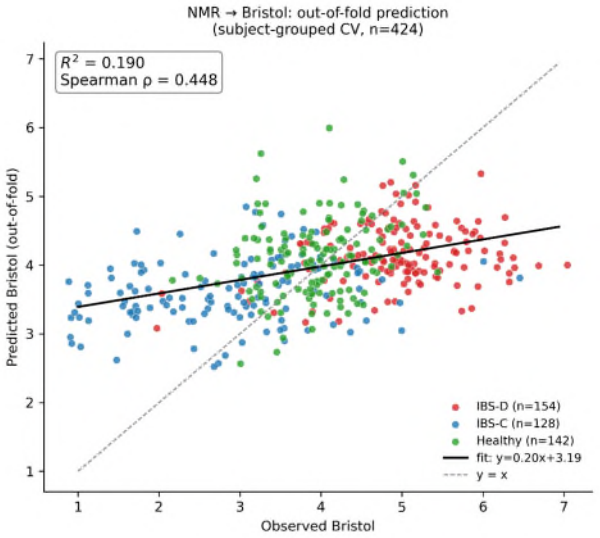


Figure 3. Subject-grouped cross-validated OOF predictions of the NMR metabolome for Bristol. R^2=0.190, ρ=0.448 (n=424); points colored by subtype (IBS-D red, IBS-C blue, healthy green); black line=linear fit, gray dashed line=y=x.

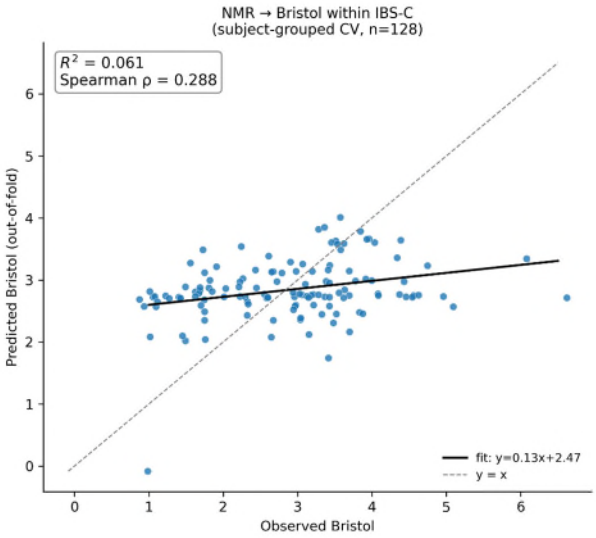


Figure 3b. Within-IBS-C robustness. n=128, R^2=0.061, ρ=0.288 (5/19 significant).

This asymmetry is only an exploratory observation (exploratory hypothesis-generating): it suggests (rather than proves) that the signal does not all come from subtype labels; under the power constraint of n≈150, sampling fluctuation cannot be excluded, and validation in larger within-subtype cohorts is needed (§4.7; bootstrap confidence intervals in Figure 10).

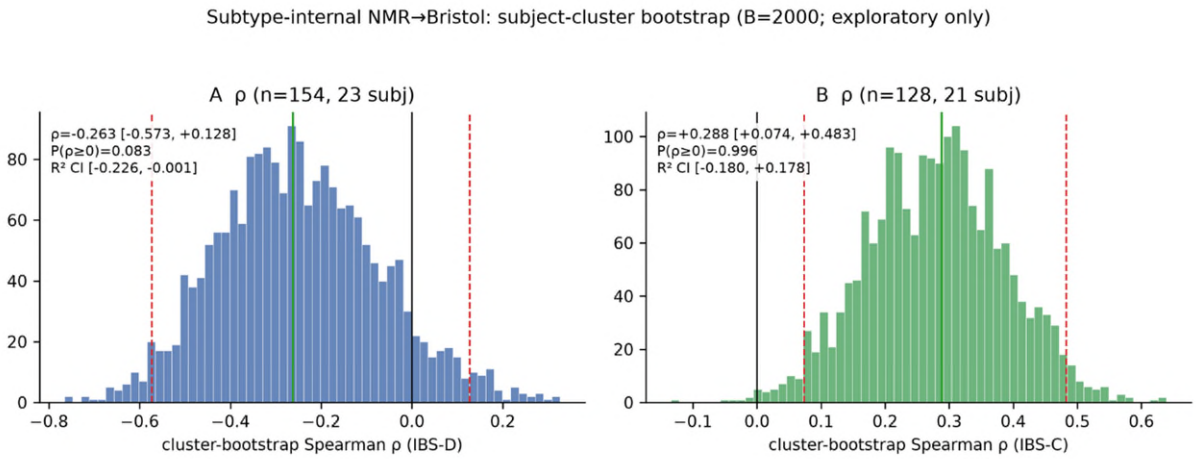


Figure 10. Subject-cluster bootstrap of within-subtype ρ (both subtypes). **A:** IBS-D (n=154, 23 subjects)—ρ point estimate −0.263, 95% CI [−0.573, +0.128] crossing zero (P(ρ≥0)=0.083), R^2 point estimate −0.048, 95% CI [−0.226, −0.001]; **B:** IBS-C (n=128, 21 subjects)—ρ point estimate +0.288, 95% CI [+0.074, +0.483] (P(ρ≥0)=0.996), R^2 CI [−0.180, +0.178]. The signals in both subtypes are only exploratory hypotheses.

3.6 Cross-layer validation: selective weak coupling between functional potential and metabolite pools (Figure 5)

Among 253 associations across 11 functional features × 23 targets (16 NMR + 6 BA + Bristol, n=423):

Table 7. Significant cross-layer Spearman correlations between microbial functional potential and metabolite pools / Bristol score (n = 423).

Functional feature	Target	ρ	q
bcd (K00248)	Butyrate	+0.265	8×10^{-6}
bcd	Hypoxanthine	+0.209	9×10^{-4}
bcd	Lysine / succinate	+0.189	0.003
bcd	Acetate	+0.182	0.005
bcd	Uracil	+0.183	0.005
bcd	Bristol	+0.175	0.007
adhE (K04072)	DCA-sulfate	−0.231	2×10^{-4}
adhE	Bristol	−0.203	0.001
ko00380 tryptophan metabolism	Bristol	−0.215	7×10^{-4}
ko00280 BCAA degradation	Bristol	−0.181	0.005
Butyrate synthesis gene set (7 KOs)	Lysine	+0.161	0.017
Butyrate synthesis gene set	2-Methylbutyrate	−0.144	0.050

Key observations: (i) the single core gene marker *bcd* (K00248, a genomic abundance marker of butyrate-producing communities, not the rate-limiting enzyme of butyrate synthesis) showed consistent weak-to-moderate coupling with the major fecal SCFA pools and Bristol; (ii) however, aggregation diluted the signal—neither the 7-gene sum nor the ko00650 pathway aggregation showed significant associations, possibly reflecting isoenzyme heterogeneity, annotation noise, and the absence of expression regulation; (iii) bile acid-related functional genes (BSH K01442, *bai*-related ko00121) showed no significant association with the bile acid pool—a typical case of "functional potential ≠ enzyme activity ≠ metabolite pool"; (iv) ko00280/ko00380 (BCAA/tryptophan degradation potential) were negatively correlated with Bristol and positively correlated with fecal free AAs, forming a "low potential, high product" combination, suggesting that when transit accelerates, the host absorption window shortens, and the fate of AAs in the lumen is determined by competition between absorption and microbial activity, rather than by degradation gene abundance alone.

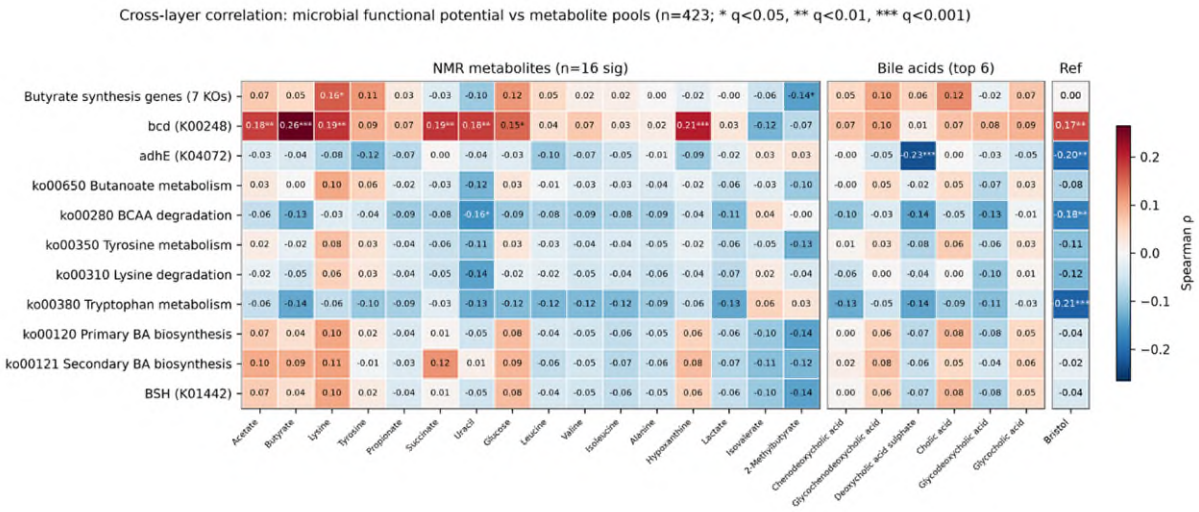


Figure 5. Cross-layer association heatmap: functional potential vs metabolite pools. 11 functional features × (16 NMR + 6 bile acids + Bristol), n=423; global BH-FDR correction (253 tests). Core findings: *bcd*↔butyrate (ρ=0.265), *adhE*↔DCA-sulfate (−0.231), ko00380/ko00280↔Bristol (−0.215/−0.181).

3.7 External reference: Microbiome 2023 [3] (Jacobs et al., PRJNA812699)

In the independent cohort (368 samples × 601 metabolites; IBS-C 89 / IBS-D 84 / mixed 37), the metabolome distinguished IBS from controls with AUC≈0.85; however, classification by BH subtype (D/C/M) was weak (balanced acc≈0.42–0.52), consistent with the cohort's own report of "limited metabolic differences between subtypes, with elevated bile acids and polyamines in the D subtype," and also consistent with the positive bile acid–Bristol direction in our cohort. Both cohorts point to the same conclusion: the metabolic layer is good at capturing "disease vs healthy" and continuous severity axes, rather than discrete subtype labels.

3.8 Zero-shot cross-cohort domain generalization: Cell 2020 → Jacobs 2023 (Figure 6)

Feature alignment: The 16 Bristol-significant markers from Cell 2020 were matched by prespecified names to the 601 LC-MS panel of Jacobs 2023. The Jacobs platform (untargeted LC-MS) does not cover SCFAs; the transferable intersection was 11/16 (Lysine, Tyrosine, Succinate, Uracil, Glucose, Leucine, Valine, Isoleucine, Alanine, Hypoxanthine, Lactate); the 5 missing items (acetate/butyrate/propionate/isovalerate/2-methylbutyrate) were neutrally imputed at the training mean in z-space.

Zero-shot blind test: The model was trained only on Cell 2020 (subject-grouped CV for α); weights, standardization statistics, and intercept were frozen and then applied directly to 368 Jacobs samples—no training, tuning, or feature selection was performed in the target cohort. BH groups (Constipation 89 / Mixed 37 / Diarrhea 84) were used as a proxy endpoint for intestinal motility (prespecified hypothesis: predicted score D>M>C).

Results: The shared 11-feature model (R11) performed well zero-shot—predicted score Diarrhea vs Constipation+Mixed AUC=0.647 (MWU $p=3.2\times10^{-5}$; D vs C sensitivity AUC=0.676), with BH ordinal Spearman ρ=+0.278 ($p=4.4\times10^{-5}$), Kruskal-Wallis $p=3.1\times10^{-4}$; consistent across batches (Batch One AUC=0.679; Batch Two=0.586); no significant confounding by age/BMI/sex (p>0.05). Within the training domain, this submodel had OOF R^2=0.115, ρ=0.370.

Failure-mode attribution (key): The full 16-feature model including 5 imputed SCFAs (F16) collapsed to AUC=0.558 (ρ=+0.136, p=0.049) on the same task—when the SCFAs with the largest variance and strongest weights in the training domain were unmeasurable in the target domain, the

effective transferable variance of the full model was absorbed by them; signal recovered after shrinking to the measurable intersection. This pattern complements the conclusions of DANN domain-adversarial experiments: the core constraint on cross-center generalization is not only distribution shift but also feature transferability—non-transferable high-weight features systematically poison transfer.

Directional replication: 11/11 shared features had the same correlation direction with BH ordinal in Jacobs as with Bristol in Cell 2020 (one-sided binomial sign test $p=4.9\times10^{-4}$).

Random and demographic baselines: Under the same Cell→Jacobs zero-shot pipeline, the mean AUC of 200 random 11-feature models was 0.506 (95% quantile 0.434–0.589), and R11's 0.647 exceeded this null (p=0.005). The demographics-only model (age/BMI/sex) had AUC=0.593 (95% CI 0.509–0.675). Adding R11 to demographics increased AUC to 0.656 (ΔAUC=+0.062, 95% CI −0.021–0.148, p=0.080); adding the five-metabolite panel increased it to 0.645 (ΔAUC=+0.051, 95% CI −0.029–0.132, p=0.112). Thus the metabolite panel does not demonstrate statistically significant incremental value over demographics alone; the external results should be interpreted as exploratory transferability, not clinical utility.

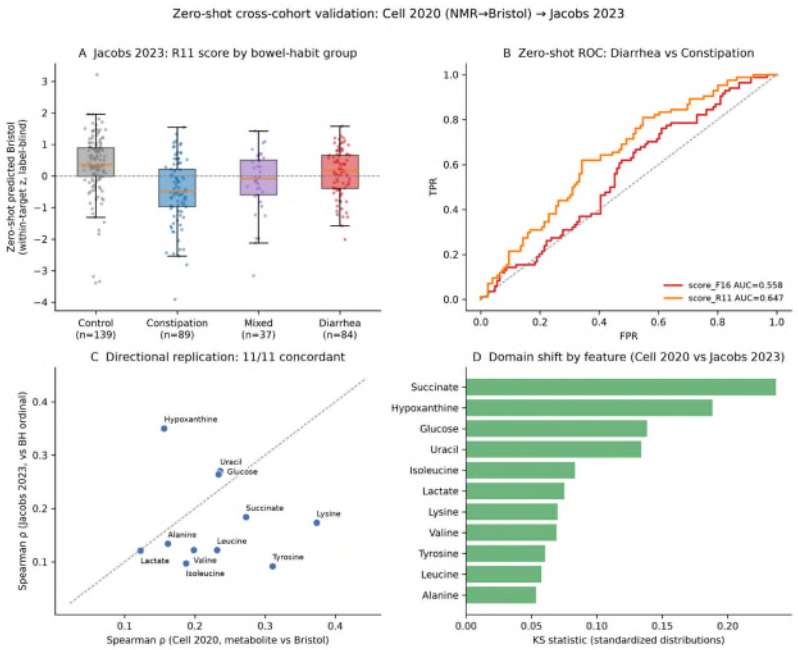


Figure 6. Zero-shot cross-cohort domain generalization (Cell 2020 → Jacobs 2023). **A:** R11 predicted scores by Jacobs motility group (z-standardized within the target domain, label-blind; healthy controls (Control; BH=Normal; n=139) included as reference, controls not included in ROC)—a clear C<M<D monotonic trend; **B:** Diarrhea vs Constipation+Mixed zero-shot ROC (R11 AUC=0.647; D vs C sensitivity 0.676; F16 collapsed to 0.558 because of imputed SCFAs); **C:** directional replication of the 11 shared features (11/11 consistent); **D:** domain-shift diagnostics (self-standardized KS statistics).

3.9 Source-domain closed-loop stability selection and held-out zero-shot evaluation: the source-stabilized metabolite panel (Figure 7)

Cross-cohort stability selection was upgraded to a source-domain closed-loop protocol (label-blind transductive feature selection; criteria in §2.5): inclusion decisions used only ① Cell 2020 grouped bootstrap stability π_uni_cell; ② the platform-shift KS of target-domain unlabeled X; ③ measurability in both groups—target-domain labels had zero contact during selection. The selector (features, scaling parameters, Ridge coefficients) was frozen before zero-shot evaluation on the fully held-out target cohort (2000× paired bootstrap).

Source-domain stability (π_uni_cell): In Cell 2020, acetate/butyrate/lysine had π_uni_cell=1.0, tyrosine 0.990, uracil 0.910, glucose 0.883, and leucine 0.883; succinate 0.957, valine 0.753, 2-methylbutyrate 0.700, isoleucine 0.680, alanine 0.597, hypoxanthine 0.570, and lactate 0.403—those below 0.8 or borderline were excluded by the stability gate.

Platform shift (KS, unlabeled diagnostic): Unmeasurable SCFAs (acetate/butyrate/propionate/isovalerate/2-methylbutyrate) had no KS definition and were treated as non-transferable; among measurable items, succinate (0.237) and hypoxanthine (0.188) exceeded the 0.15 threshold and were excluded; the 5 core features all had KS < 0.15 (0.058–0.138).

Source-stabilized panel determination (three gates): Features simultaneously satisfying ① π_uni_cell ≥ 0.8, ② KS < 0.15, and ③ measurability in both groups totaled 5 items: lysine, tyrosine, uracil, glucose, and leucine.

Table 8. Source-domain closed-loop stability selection gates and inclusion decisions for the transferable metabolite panel.

Feature	π_uni_cell	π_multi_cell	KS	Decision
Lysine	1.000	0.840	0.070	**Included ✓**
Tyrosine	0.990	0.990	0.060	**Included ✓**
Uracil	0.910	0.950	0.134	**Included ✓**
Glucose	0.883	0.963	0.138	**Included ✓**
Leucine	0.883	0.913	0.058	**Included ✓**
Succinate	0.957	—	0.237	Platform-shift exclusion
Hypoxanthine	0.570	—	0.188	Platform-shift + source-stability exclusion
Valine	0.753	—	0.069	Source-stability exclusion
Isoleucine	0.680	—	0.083	Source-stability exclusion
Lactate	0.403	—	0.075	Source-stability exclusion
Acetate/butyrate/propionate/isovalerate /2-methylbutyrate	—	—	Unmeasurable	Non-jointly-measurable exclusion

Held-out zero-shot evaluation: The frozen 5-item model (within-domain OOF R^2=0.072 under the unified primary protocol) was applied directly to Jacobs (D vs C+M): AUC=0.634 (95% CI 0.559–0.705), BH ordinal ρ=+0.243; not significantly different from R11 (0.647, CI 0.572–0.716; paired ΔAUC=−0.013, 95% CI −0.044–+0.017, p=0.805), and significantly better than F16 (0.558, CI 0.478–0.630; ΔAUC=+0.075, CI +0.011–+0.140, p=0.013). Two-level nested sensitivity: under pre-screening nesting (B=300, gates fixed), the 5-item selection survival rates were 0.997/0.977/0.867/0.857/0.803, and the core was fully reproduced; under fully nested analysis (gates recomputed within resampling, outer full bootstrap B_out=100 × inner π protocol B_in=100), the per-feature all-gates-off selection frequencies were lysine 0.96, tyrosine 0.79, leucine 0.36, glucose 0.28, and uracil 0.23—the screening layer (0.94–1.00) and KS gate (amino acids 1.00; uracil 0.78, glucose 0.63) were highly stable, and the bottleneck was estimation noise of the π gate for intermediate values (0.883–0.910, near the 0.80 gate boundary); the conservative composite half-sample variant (outer 50% subject subsampling, inner another 50% of the outer, simulating a "sample size halved" composite conservative lower bound) was overall lower (lysine 0.75, tyrosine 0.58, no feature reached 0.80), consistent with the power loss from halving the sample. The frozen core should be read as a transferable set "conditional on source-domain gate estimation": conditional stability fully holds, unconditional robustness concentrates on lysine/tyrosine, and the other three await external replication. That is: source-domain stability shrinkage (16→5) repairs F16 migration poisoning and maintains discrimination not significantly different from R11 with fewer features—the number of platform-robust features is determined by the data.

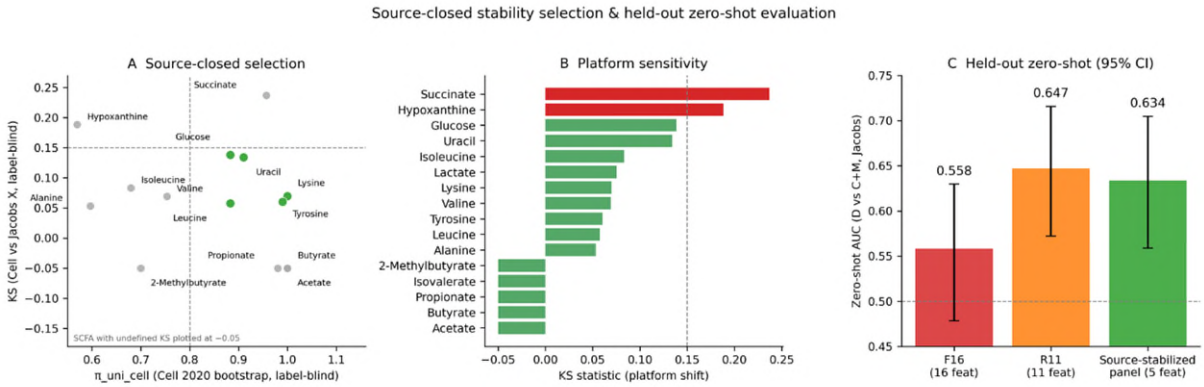


Figure 7. Source-domain closed-loop stability selection and held-out zero-shot evaluation (source-stabilized panel). **A:** source-domain stability × platform-shift scatter (x=π_uni_cell, y=KS; neither uses target-domain labels; green points=5 core items; dashed lines=0.8/0.15 thresholds); **B:** platform-shift diagnostics (self-standardized distribution KS; red=above the 0.15 threshold: Succinate 0.237, Hypoxanthine 0.188); **C:** held-out zero-shot AUC (D vs C+M; error bars=2000× paired bootstrap 95% CI)—source-stabilized panel (5 items, 0.634 [0.559, 0.705]) was not significantly different from R11 (11 items, 0.647 [0.572, 0.716]; ΔAUC=−0.013, p=0.805) and was significantly better than F16 (16 items, 0.558 [0.478, 0.630]; ΔAUC=+0.075, p=0.013).

Permutation significance of the within-domain fit (Figure 12): A subject-level permutation test (1,000 permutations, unified primary protocol) was performed on the within-domain OOF fits: full 19 items observed R^2=0.190, p=0.001; source-stabilized panel 5 items within-domain R^2=0.072, p=0.002—both within-domain fits were significantly better than the permutation null (null mean ≈−0.040). By design, the primary evidence for source-stabilized panel remains cross-domain transferability (held-out zero-shot AUC=0.634 and paired bootstrap), and within-domain significance does not replace cross-domain evidence.

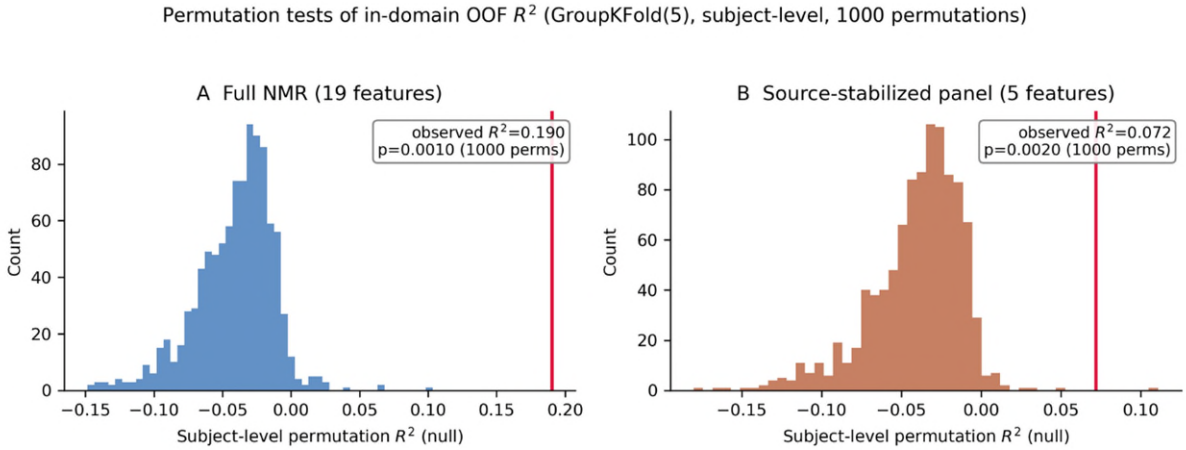


Figure 12. Subject-level permutation test of within-domain R^2. **A:** full NMR (19 items, observed R^2=0.190, p=0.001); **B:** source-stabilized panel (5 items, observed R^2=0.072, p=0.002); null distribution=1,000 subject-level permutations (unified primary protocol, null mean ≈−0.040), solid red line=observed.

Protocol note: an earlier label-informed two-feature selection is not used in this manuscript; all reported selection is source-domain closed-loop and label-blind.

Threshold robustness (Figure 11): The prespecified gates (π_uni_cell ≥ 0.80, KS < 0.15) lie on a plateau—threshold scanning showed that within πt∈[0.80, 0.85]×KSt∈[0.15, 0.20] the core was always the same 5 items; the core shrank only when πt=0.90 (to 3 items: Lysine/Tyrosine/Uracil) or KSt≤0.12 (the baseline row shrank to 3 items, with only amino acids remaining); relaxing to πt≤0.75 increased it to 6 items (+Valine)—the qualitative conclusion of core composition was unchanged within ±1 step (see §4.7).

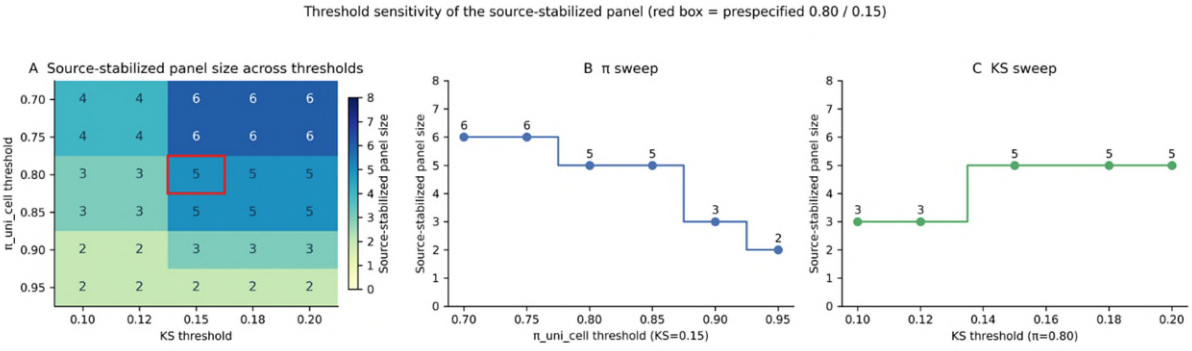


Figure 11. Threshold sweep for stability selection. In the figure, π_uni_cell (abbreviated πt in the text) is source-domain univariate stability; KSt is the KS threshold. **A:** source-stabilized panel size on the grid πt∈[0.70,0.95]×KSt∈[0.10,0.20] (number in each cell=number of core items, color=size; red box=prespecified 0.80/0.15, located on the plateau; πt∈[0.80,0.85]×KSt∈[0.15,0.20] always yielded the same 5 items); **B:** π sweep with KSt=0.15 fixed (πt≤0.75→6 items, +Valine; πt=0.90→3 items); **C:** KS sweep with πt=0.80 fixed (KSt≤0.12→3 items, Uracil/Glucose cut, only amino acids remained).

3.10 Exploratory two-sample MR (supplementary)

In an exploratory two-sample MR, 27 candidate genera were tested against overall IBS risk. No genus survived FDR correction (minimum q=0.835), and nominal signals were weak and directionally inconsistent with observational associations. Because the analysis used a lenient instrument threshold and distance-based pruning rather than an LD reference panel, it is best interpreted as a negative screening analysis, not causal evidence. Full results and STROBE-MR items are provided in Appendix S2.

4 Discussion

Three mechanistic axes (SCFA–5-HT, proteolytic fermentation, bile acid–TGR5) share a more parsimonious common upstream: intestinal transit time. Faster transit simultaneously produces: (i) more carbohydrate/protein substrate escaping to distal fermentation (SCFA, BCAA, lysine, and tyrosine rise in parallel); (ii) more pronounced bile acid escape (BAD spectrum); (iii) higher mucosal turnover and nucleic acid degradation (uracil); (iv) a shorter water/salt absorption window (higher Bristol, increased glucose delivery). Within IBS-C, positive correlations remained, and all 5 clean core features under the closed-loop protocol point to transit–substrate exposure processes, consistent with the parsimonious model of "transit time as the main effect and each layer's signal as its mechanistic branch"; by Occam's principle, this paper does not prioritize three independent parallel mechanisms (distinguishing them would require direct measurement of orocecal transit time, which was not available here).

4.1 SCFA–serotonin–intestinal motility axis

Acetate (ρ=0.414), butyrate (0.379), and propionate (0.297) were positively correlated with Bristol (higher=looser). Mechanistically, SCFAs activate FFAR2 (GPR43)/FFAR3 (GPR41) [14] on enterochromaffin cells (EC cells), upregulate tryptophan hydroxylase TPH1, and promote 5-HT release (Reigstad *et al.*, FASEB J 2015 [15]; Yano *et al.*, Cell 2015 [16] confirmed that microbiota–SCFA drive colonic 5-HT synthesis); 5-HT amplifies the peristaltic reflex via 5-HT3/5-HT4 receptors [17] on intrinsic primary afferent neurons → faster transit → higher fecal water content. In addition, SCFAs are poorly absorbed anions, and their accumulation in the distal colon produces an osmotic water-retaining effect. Together, the two pathways explain the directional consistency of "high SCFA in IBS-D ↔ loose stools." Succinate (+0.273), as a SUCNR1 (GPR91) [18] agonist, can also promote 5-HT release from EC cells, in the same direction as the main effect. Branched SCFAs (isovalerate, 2-methylbutyrate: fermentation products of BCAAs) were negatively correlated, suggesting a contrasting picture of a different fermentation niche (slow transit, carbohydrate-replete environment). This mechanistic narrative is consistent with the Shapley decomposition in §3.4b: the SCFA group's marginal predictive contribution (φ=0.128) was greater than that of the amino acid group (φ=0.057); removing SCFAs decreased the full-model R^2 by 0.076 (p=0.002), whereas removing amino acids decreased it by only 0.016 (p=0.085, not significant)—the SCFA axis is not only the strongest in univariate correlation but also the group with the largest multivariate marginal contribution.

4.2 Proteolytic fermentation fingerprint: BCAAs and aromatic amino acids

Leucine/valine/isoleucine (0.19–0.23), lysine (0.374), and tyrosine (0.311) were synchronously positive, consistent with the substrate-delivery model of "fast transit → proximal carbohydrate depletion → enhanced distal proteolytic fermentation." Microbial decarboxylation products further amplify motility effects: tyrosine → tyramine (TAAR1 agonist, promotes 5-HT release); lysine →

cadaverine/polyamines (polyamines are associated with IBS-D; Jacobs *et al.*, Microbiome 2023 [3] also reported elevated polyamines and bile acids in the D subtype, cross-validating the direction in our cohort). 2-Methylbutyrate (−0.185) was the only significant negative marker, complementary to the positive BCAA profile, suggesting that the "free state vs fermentation consumption" of branched-chain acids reflects opposite transit ecologies. Unlike §4.1, the group-level marginal contribution of proteolytic fermentation-related amino acids (lysine, tyrosine, leucine) did not reach permutation significance (φ_AA=0.057, p=0.085, §3.4b); the narrative in this section is based on the directional consistency of univariate correlations (§3.4), which is not equivalent to an independent group-level predictive contribution—the independent increment of the amino acid profile was not confirmed at the current sample size.

4.3 Bile acid–TGR5 secretomotor axis

Primary bile acids (CA, CDCA), glycine-conjugated forms (GCA, GCDCA, GDCA), and DCA-sulfate were consistently positively correlated (ρ 0.23–0.31). Mechanism: bile acids escaping ileal reabsorption reach the colon and, via TGR5 (GPBAR1) [19] on EC cells/enteric neurons → cAMP → 5-HT release + CFTR-dependent Cl^-/water secretion, and accelerate colonic motility—the classic BAD (bile acid diarrhea) pathway [20]. About one-third of IBS-D patients have BAD [20]; this study suggests that the mechanism exists with weak-to-moderate strength across the full continuous cohort spectrum, not only in the D subtype.

4.4 "Potential ≠ activity": methodological implications of cross-layer validation

Cross-layer associations show: (i) *bcd* (a genomic abundance marker of butyrate-producing communities, not the catalytic rate-limiting step) was significantly associated with the butyrate pool (ρ=0.265), demonstrating that metagenomic butyrate-producing capacity can partially reflect the metabolite pool at the community level; (ii) however, the signal disappeared after aggregation (7-gene sum, ko00650), indicating that community-level aggregation systematically dilutes biological signal; (iii) BSH/*bai* showed zero association with the bile acid pool, warning that functional gene abundance cannot replace enzyme activity/metabolite measurement. Direct implications for "multimodal fusion": the metabolic layer should be a first-tier feature layer, and the metagenomic functional layer has conditional value only at single-gene/single-enzyme resolution; simple concatenation (NMR+genus) showed no gain when unfiltered (0.171 vs 0.190), and after filtering at the same prevalence was comparable to NMR (0.204 vs 0.190, ΔR^2 not significant); fusion strategies still require layer-wise weighting or causal structure, which warrants further investigation in structured multimodal modeling.

4.5 Interpretability of the stable panel: predictive magnitude is not transferability

The stable panel contains two kinds of features. Uracil and glucose reflect luminal turnover, carbon delivery, and transit exposure; their source-domain stability is high, but their fully nested selection frequencies are lower (0.23–0.28), so they are conditionally rather than unconditionally stable. Lysine, tyrosine, and leucine are proteolytic-fermentation markers that are measurable on both platforms; lysine and tyrosine have the strongest source-domain stability and the smallest distribution shift. This distinction matters: the panel was selected for cross-platform measurability and source-domain stability, not because it is the strongest within-domain predictor. The external AUC should therefore be read as transferability evidence, not as proof that these five metabolites are the dominant biological drivers.

4.6 Two-sample MR: exploratory negative evidence

The two-sample MR was exploratory. No genus survived FDR correction (minimum q=0.835), nominal signals were weak, and their directions were not consistently aligned with observational associations. The analysis used a lenient instrument threshold and ±1 Mb distance pruning without an LD reference panel, so it should not be read as a definitive causal test; MVMR was not attempted because conditional instrument strength and an inter-exposure LD matrix were unavailable. The MR results are supplementary and do not test the genus→metabolite→IBS mediation chain.

4.7 Limitations

The study is cross-sectional and observational, and the Cell 2020 cohort is modest in size (75 subjects; 67 with Bristol). Within-subtype estimates are underpowered, so the IBS-D result is exploratory and does not establish absence of an effect; the same constraint applies to the measured common subset (n=127/59) used for paired cross-layer comparisons. Reverse causation between transit time and microbial/metabolite profiles cannot be excluded.

The Bristol target is a diary-window aggregate with partial ±3-day/±7-day matching and subject-mean fallback in the sensitivity set. Faecal water content and dry weight were unavailable, so concentration effects, dilution, and true concentration changes cannot be fully separated; the sample-wise centering sensitivity partly addresses this but does not replace direct water-content measurement.

Cross-cohort validation is limited to metabolites measurable on both platforms; SCFAs and matched taxonomy were unavailable in Jacobs. The five-metabolite panel transferred externally, but its increment over age/BMI/sex was not statistically significant (ΔAUC +0.051, 95% CI −0.029 to 0.132), and the panel is conditionally rather than unconditionally stable. The demographic comparison is not a substitute for a full clinical decision model, and the external endpoint is a BH motility proxy rather than Bristol.

The MR analysis used a lenient instrument threshold and distance-based pruning without an LD reference panel, and the outcome was overall IBS rather than subtype-specific; it should be read as exploratory negative evidence. MVMR and mediation analyses were not attempted because conditional instrument strength and an inter-exposure LD matrix were unavailable.

4.8 Conclusions and outlook

In the primary measured-Bristol analysis, the NMR metabolome had the highest point-estimate OOF R^2 (0.247) and the strongest rank correlation (ρ=0.510), and the SCFA group dominated the within-NMR decomposition (φ=0.128). However, a paired comparison with genus on the measured common subset did not establish superiority (ΔR^2 +0.186, 95% CI −0.010 to 0.404), and the combined NMR+genus model was not significantly better than NMR alone. The five-metabolite panel transferred externally (AUC 0.634) but did not show a statistically significant increment over age/BMI/sex (ΔAUC +0.051, 95% CI −0.029 to 0.132). The results therefore support the metabolome as a strong single-layer correlate of stool consistency, not as a proven superior or clinically incremental biomarker. Future work should validate the panel in independent cohorts, measure faecal water content and transit time, use LD-aware MR and metabolite instruments, and test combined demographic–metabolite models in larger samples.

5. Data and code availability

Data: raw sequencing data at ENA PRJEB37924 [2]; processed multi-omics matrices at Mendeley Data (10.17632/29n2z5r5ph.3) [24]; external validation cohort at PRJNA812699 [3]; taxonomic and functional annotation abundances from GMrepo [11].

MR data sources: MiBioGen full hit table [4] (122,110 rows / 211 taxa) and IBS GWAS summary statistics GCST90016564 [5] (53,400 cases / 433,201 controls; Eijsbouts *et al.*, 2021), both public.

Analysis code and result tables: the analysis code for eight-layer feature construction, unified comparison protocol, classification uncertainty, external baselines, cross-cohort validation, and two-sample MR, together with primary result tables and derived data, are provided as Supplementary Code S1 (IBS_Supplementary_Code_S1_v1.0.zip) / Supplementary Data S1 (IBS_Supplementary_Data_S1_v1.0.zip) with this submission. The analysis code is publicly available at https://github.com/toby27227-dev/IBS-multilayer-comparison; a permanent Zenodo DOI will be added once archiving is complete. Primary result tables are no longer described as available only from the author on request; supplementary materials S1–S3 are in the appendices.

6. Declaration

Ethics: This study is a secondary analysis of publicly available de-identified data (Cell 2020: ENA PRJEB37924; Jacobs 2023: PRJNA812699); no new participant recruitment or sample collection was performed. The original studies obtained the relevant ethics approvals and informed consent, and data use followed the original data licenses (the Cell 2020 processed matrices are CC BY 4.0). Under the authors' institutional and national regulations for secondary analysis of de-identified public data, no additional ethics approval was required. The study was conducted in accordance with the Declaration of Helsinki.

References

[1] Enck P, Aziz Q, Barbara G, Farmer AD, Fukudo S, Mayer EA, *et al.* Irritable bowel syndrome. *Nat Rev Dis Primers*. 2016;2:16014. doi:10.1038/nrdp.2016.14.

[2] Mars RAT, Yang Y, Ward T, Houtti M, Priya S, Lekatz HR, *et al.* Longitudinal multi-omics reveals subset-specific mechanisms underlying irritable bowel syndrome. *Cell*. 2020;182(6):1460-1473.e17. doi:10.1016/j.cell.2020.08.007.

[3] Jacobs JP, Lagishetty V, Hauer MC, Labus JS, Dong TS, Toma R, *et al.* Multi-omics profiles of the intestinal microbiome in irritable bowel syndrome and its bowel habit subtypes. *Microbiome*. 2023;11:5. doi:10.1186/s40168-022-01450-5.

[4] Kurilshikov A, Medina-Gomez C, Bacigalupe R, Radjabzadeh D, Wang J, Demirkan A, *et al.* Large-scale association analyses identify host factors influencing human gut microbiome composition. *Nat Genet*. 2021;53(2):156-165. doi:10.1038/s41588-020-00763-1.

[5] Eijsbouts C, Zheng T, Kennedy NA, Bonfiglio F, Anderson CA, Moutsianas L, *et al.* Genome-wide analysis of 53,400 people with irritable bowel syndrome highlights shared genetic pathways with mood and anxiety disorders. *Nat Genet*. 2021;53(11):1543-1552. doi:10.1038/s41588-021-00950-8.

[6] Lewis SJ, Heaton KW. Stool form scale as a useful guide to intestinal transit time. *Scand J Gastroenterol*. 1997;32(9):920-924. doi:10.3109/00365529709011203.

[7] Francis CY, Morris J, Whorwell PJ. The irritable bowel severity scoring system: a simple method of monitoring irritable bowel syndrome and its progress. *Aliment Pharmacol Ther*. 1997;11(2):395-402. doi:10.1046/j.1365-2036.1997.142318000.x.

[8] Holmes I, Harris K, Quince C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. *PLoS One*. 2012;7(2):e30126. doi:10.1371/journal.pone.0030126.

[9] Hillmann B, Al-Ghalith GA, Shields-Cutler RR, Zhu Q, Knight R, Knights D. SHOGUN: a modular, accurate and scalable framework for microbiome quantification. *Bioinformatics*. 2020;36(13):4088-4090. doi:10.1093/bioinformatics/btaa277.

[10] Beghini F, McIver LJ, Blanco-Míguez A, Dubois L, Asnicar F, Maharjan S, *et al.* Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *eLife*. 2021;10:e65088. doi:10.7554/eLife.65088.

[11] Dai D, Zhu J, Sun C, Li M, Liu J, Wu S, *et al.* GMrepo v2: a curated human gut microbiome database with special focus on disease markers and cross-dataset comparison. *Nucleic Acids Res*. 2022;50(D1):D777-D784. doi:10.1093/nar/gkab1019.

[12] Skrivankova VW, Richmond RC, Woolf BAR, Yarmolinsky J, Davies NM, Swanson SA, *et al.* Strengthening the reporting of observational studies in epidemiology using Mendelian randomization: the STROBE-MR statement. *JAMA*. 2021;326(16):1614-1621. doi:10.1001/jama.2021.18236.

[13] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol*. 1995;57(1):289-300. doi:10.1111/j.2517-6161.1995.tb02031.x.

[14] Brown AJ, Goldsworthy SM, Barnes AA, Eilert MM, Tcheang L, Daniels D, *et al.* The orphan G protein-coupled receptors GPR41 and GPR43 are activated by propionate and other short chain carboxylic acids. *J Biol Chem*. 2003;278(13):11312-11319. doi:10.1074/jbc.M211609200.

[15] Reigstad CS, Salmonson CE, Rainey JF 3rd, Szurszewski JH, Linden DR, Sonnenburg JL, *et al.* Gut microbes promote colonic serotonin production through an effect of short-chain fatty acids on enterochromaffin cells. *FASEB J*. 2015;29(4):1395-1403. doi:10.1096/fj.14-259598.

[16] Yano JM, Yu K, Donaldson GP, Shastri GG, Ann P, Ma L, *et al.* Indigenous bacteria from the gut microbiota regulate host serotonin biosynthesis. *Cell*. 2015;161(2):264-276. doi:10.1016/j.cell.2015.02.047.

[17] Gershon MD, Tack J. The serotonin signaling system: from basic understanding to drug development for functional GI disorders. *Gastroenterology*. 2007;132(1):397-414. doi:10.1053/j.gastro.2006.11.002.

[18] He W, Miao FJ, Lin DC, Schwandner RT, Wang Z, Gao J, *et al.* Citric acid cycle intermediates as ligands for orphan G-protein-coupled receptors. *Nature*. 2004;429(6988):188-193. doi:10.1038/nature02488.

[19] Kawamata Y, Fujii R, Hosoya M, Harada M, Yoshida H, Miwa M, *et al.* A G protein-coupled receptor responsive to bile acids. *J Biol Chem*. 2003;278(11):9435-9440. doi:10.1074/jbc.M209706200.

[20] Walters JR, Pattni SS. Managing bile acid diarrhoea. *Therap Adv Gastroenterol*. 2010;3(6):349-357. doi:10.1177/1756283X10377126.

[21] Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol*. 2015;44(2):512-525. doi:10.1093/ije/dyv080.

[22] Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol*. 2016;40(4):304-314. doi:10.1002/gepi.21965.

[23] Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol*. 2013;37(7):658-665. doi:10.1002/gepi.21758.

[24] Mars R. Mars et al, IBS multi-omics [Data set]. *Mendeley Data*. 2020. doi:10.17632/29n2z5r5ph.3.

Appendix S1 Supplementary methodological note

An earlier exploratory selection procedure used target-domain labels and is not used in this manuscript. The reported panel is based only on source-domain stability, unlabeled platform-shift diagnostics, and cross-platform measurability. Full results for the current protocol are reported in §3.8–§3.9 and Appendix S2.

Appendix S2 Key STROBE-MR items and complete results for 27 genera

Table S1. STROBE-MR reporting checklist and study implementation summary.

STROBE-MR item	Corresponding content in this study
Exposure data source	MiBioGen microbiome GWAS [4] (Kurilshikov *et al.*, 2021; 16–24 cohorts; 211 taxa)
Outcome data source	IBS GWAS GCST90016564 [5] (Eijsbouts *et al.*, 2021; 486,601 European samples, 53,400 cases/433,201 controls)
Instrument selection	$p<10^{-5}$; F statistic 100% >10; ±1 Mb distance pruning to approximate LD removal (retain the lowest-p SNP within the window)
Harmonization	rsID matching 305/313; strand flipping and complementary pair handling; palindromic SNPs conservatively removed; 259/313 entered estimation
Primary analysis	Random-effects IVW (delta-method SE)
Sensitivity	Weighted median (bootstrap 2000), MR-Egger intercept, Cochran Q, leave-one-out
Multiple correction	BH-FDR of IVW p values for 27 genera
Software	Python 3.13.13; NumPy 2.4.6; SciPy 1.17.1; pandas 3.0.3; scikit-learn 1.8.0
Limitations	Distance pruning is not full LD; overall IBS phenotype; sample overlap unknown; MVMR not reported

Table S2. Complete two-sample MR results for 27 candidate genera (sorted by IVW p ascending; OR>1 indicates the risk direction).

Genus	nSNP	IVW OR (95% CI)	IVW p	BH-FDR	Cochran Q p	Egger intercept p	LOO max shift
Bifidobacterium	12	1.070 (1.004–1.141)	0.039	0.835	0.717	0.868	0.024
Eubacterium hallii group	12	1.074 (0.996–1.157)	0.062	0.835	0.894	0.937	0.020
Barnesiella	12	1.051 (0.980–1.128)	0.165	0.911	0.646	0.787	0.021
Coprococcus1	11	1.060 (0.976–1.151)	0.168	0.911	0.226	0.647	0.026
Roseburia	11	0.930 (0.828–1.044)	0.219	0.911	0.070	0.607	0.033
Anaerostipes	11	0.936 (0.839–1.044)	0.236	0.911	0.089	0.848	0.035
Desulfovibrio	8	0.965 (0.899–1.036)	0.325	0.911	0.585	0.940	0.019
Romboutsia	11	0.973 (0.907–1.044)	0.450	0.911	0.712	0.603	0.021
Faecalibacterium	10	1.028 (0.955–1.108)	0.459	0.911	0.582	0.014	0.020
Coprococcus2	9	1.025 (0.950–1.107)	0.521	0.911	0.449	0.870	0.020
Adlercreutzia	7	0.971 (0.883–1.067)	0.535	0.911	0.135	0.917	0.033
Alistipes	9	0.964 (0.844–1.101)	0.591	0.911	0.070	0.097	0.038
Dorea	9	1.036 (0.910–1.181)	0.592	0.911	0.030	0.090	0.044
Bacteroides	8	1.038 (0.906–1.189)	0.593	0.911	0.080	0.437	0.043
Lachnospira	5	0.966 (0.852–1.096)	0.595	0.911	0.344	0.785	0.043
Gordonibacter	11	1.011 (0.969–1.054)	0.623	0.911	0.351	0.335	0.013
Dialister	9	0.975 (0.883–1.077)	0.623	0.911	0.128	0.771	0.029
Clostridium sensu stricto 1	7	1.019 (0.941–1.104)	0.636	0.911	0.493	0.333	0.028
Parabacteroides	7	1.023 (0.929–1.125)	0.647	0.911	0.754	0.177	0.025
Blautia	11	0.987 (0.917–1.062)	0.728	0.911	0.916	0.861	0.027
Methanobrevibacter	4	1.009 (0.948–1.075)	0.772	0.911	0.616	0.988	0.022
Lachnoclostridium	11	1.015 (0.887–1.162)	0.828	0.911	0.009	0.709	0.034
Oscillibacter	11	1.007 (0.943–1.076)	0.833	0.911	0.264	0.554	0.020
Fusicatenibacter	16	0.991 (0.912–1.078)	0.836	0.911	0.122	0.111	0.017
Coprococcus3	9	1.009 (0.925–1.101)	0.844	0.911	0.764	0.364	0.021
Eubacterium rectale group	6	1.003 (0.901–1.117)	0.951	0.976	0.919	0.547	0.020
Akkermansia	12	0.999 (0.938–1.064)	0.976	0.976	0.659	0.133	0.016

Note: Minimum BH-FDR=0.835; no genus passed FDR 0.05; only *Bifidobacterium* had nominal p<0.05, indistinguishable from the global null expectation of 1.35.

Appendix S3 Classification uncertainty and OOF confusion matrices

Table S3. Eight-layer three-class OOF performance and uncertainty (subject-cluster 2000× bootstrap 95% CI; subject-level label permutation).

Layer	n	Features	balanced acc (fold mean±SD)	95% CI	macro-F1	Permutation p (permutations)
NMR	445	19	0.558±0.085	[0.477, 0.638]	0.553	0.005 (200)
Bile acids	447	12	0.482±0.034	[0.419, 0.539]	0.470	0.005 (200)
Tryptophan	274	12	0.486±0.109	[0.401, 0.564]	0.477	0.005 (200)
Genus	457	69	0.506±0.077	[0.421, 0.595]	0.508	0.005 (200)
Species	457	145	0.520±0.134	[0.434, 0.611]	0.519	0.005 (200)
NMR+Genus	440	163	0.503±0.099	[0.424, 0.581]	0.501	0.010 (100)
KEGG	459	327	0.436±0.088	[0.347, 0.520]	0.428	0.030 (100)
KO	459	5310	0.384±0.087	[0.313, 0.465]	0.384	0.095 (20)

Table S3b. IBS-SSS binary classification (genus layer, 70 features, n=250/51 subjects): balanced acc=0.500 (fold mean 0.500±0.033), 95% CI [0.420, 0.578], macro-F1=0.500, AUC=0.552, permutation p=0.458 (200 permutations)—indistinguishable from chance.

Confusion matrices (rows=true, columns=predicted; class order H/IBS-C/IBS-D):

NMR (n=445)

True\Predicted	H	IBS-C	IBS-D
H	66	42	36
IBS-C	33	86	14
IBS-D	40	33	95

Bile acids (n=447)

True\Predicted	H	IBS-C	IBS-D
H	52	55	39
IBS-C	31	86	16
IBS-D	40	54	74

Tryptophan (n=274)

True\Predicted	H	IBS-C	IBS-D
H	28	36	26
IBS-C	22	43	17
IBS-D	17	22	63

Genus (n=457)

True\Predicted	H	IBS-C	IBS-D
H	71	37	43
IBS-C	40	65	32
IBS-D	41	30	98

Species (n=457)

True\Predicted	H	IBS-C	IBS-D
H	65	45	41
IBS-C	30	76	31
IBS-D	38	33	98

NMR+Genus (n=440)

True\Predicted	H	IBS-C	IBS-D
H	64	36	43
IBS-C	34	75	20
IBS-D	56	31	81

KEGG (n=459)

True\Predicted	H	IBS-C	IBS-D
H	34	49	67
IBS-C	42	68	30
IBS-D	42	26	101

KO (n=459)

True\Predicted	H	IBS-C	IBS-D
H	38	51	61
IBS-C	48	54	38
IBS-D	49	32	88

Note: Confusion matrices are pooled results of 5-fold OOF predictions; CIs and permutation p values are based on pooled OOF predictions, and fold means are the average of 5 folds; H=Healthy. The permutation p resolution floor is 1/(B+1); values of 0.005/0.010 are at that floor, and the KO value of 0.095 (20 permutations) is exploratory only. Full-sample cross-layer differences in the classification task should be read as point estimates, not strong ranking claims (see §3.3 and §4.7).

Supplementary Files

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

- IBSSupplementaryCodeS1v1.0.zip
- IBSSupplementaryDataS1v1.0.zip