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.

Exploration of Low-Abundance Tundrisphaera and Tissue 3-Hydroxybutyric Acid (β-HB) in Gastric Cancer Lymphovascular Invasion: A Multi- Cohort Multi-Omics Study with Clearly Defined Evidence Boundaries

Jiayang Gong
Nanjing Medical University
Qin Wang
Nanjing University
Zhe Wang
Nanjing Medical University
Yun Shi
Nanjing Medical University
Minlan Ren
Nanjing Medical University
Tingting Li
Nanjing Medical University
Dong Wang
Nanjing Medical University
Gang Li
Nanjing Medical University
Rui Peng
pengrui@njmu.edu.cn
Nanjing Medical University

Research Article

Keywords: gastric cancer, lymphovascular invasion, tissue microbiome, Tundrisphaera, β-HB, ketone bodies, metabolomics, external validation

Posted Date: November 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11345830/v1

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

Additional Declarations: No competing interests reported.

Abstract

Background

Lymphovascular invasion (LVI) is one of the early routes of gastric cancer dissemination. Tissue-associated microorganisms and ketone-body metabolism may be associated with aggressive disease, but parallel between-group differences cannot establish a microbial–metabolic mechanism. This study integrated a discovery cohort with paired gastric tissue 16S rRNA and targeted metabolomics data, together with public transcriptomic, single-cell, methylation, dependency, and perturbation resources.

Methods

Clinical LVI labels were available for 66 gastric cancer cases, of which 65 had both 16S and metabolomics data (32 LVI-negative [V0] and 33 LVI-positive [V1] cases). We evaluated microbial diversity, community composition, genus-level differences, and a random forest classifier with rigorous control of information leakage, and implemented a mutually exclusive two-stage descriptive genus screen. The strict tier required a mean relative abundance of ≤ 0.01% in V0 and ≥ 0.05% in V1, and a pseudocount-adjusted fold change of ≥ 10; genera that did not meet the strict-tier criteria entered a relaxed additional tier, which required a V1 cumulative genus-table count of ≥ 20 and a ≥ 10-fold increase in mean abundance. Genera with a V0 cumulative count of 0 were treated as zero-baseline candidates, for which fold changes were not estimated. We subsequently analyzed the tissue β-HB signal and the paired bridging relationship between continuous Tundrisphaera abundance and β-HB. We evaluated 6 prespecified non-immune programs in GSE62254, TCGA-STAD, GSE246963, GSE237876, GSE163558, GSE239676, and GSE261635: ketone-body handling, butyrate metabolism, fatty acid degradation, EMT/ECM invasion, endothelial/angiogenic activity, and epithelial integrity. DepMap and LINCS analyses were used only to prioritize experimental candidates.

Results

The V1/V0 cumulative genus-table counts for Tundrisphaera were 33/1, the mean relative abundances were 0.002250%/0.000070%, and the pseudocount-adjusted fold change was 31.57; cumulative counts are not equivalent to the number of positive samples. The β-HB signal was higher in V1 (median, 68.49) than in V0 (median, 39.31; Mann–Whitney $P = 0.0220318$), but the metabolome-wide FDR was 0.9758. In the 65 paired samples, Tundrisphaera and β-HB showed a significant inverse bridging relationship (Spearman $\rho = -0.6797$; $P < 0.001$), with Tundrisphaera abundance and β-HB varying in opposite directions. In GSE62254 and TCGA-STAD, EMT/ECM and endothelial programs were associated with poorer survival. DepMap did not show widespread gastric cancer dependency on the candidate genes; LINCS reversal scores were used only to generate hypotheses.

Conclusions

The β-HB signal measured by targeted LC–MS was nominally elevated in LVI-positive gastric cancer tissues; the two-stage descriptive screen identified the low-abundance candidate Tundrisphaera in the relaxed additional tier, and a significant inverse bridging relationship was observed.

1. Introduction

Lymphovascular invasion refers to the histopathological feature of tumor cells entering lymphatic or blood vessels, and is closely associated with gastric cancer dissemination and adverse outcomes[1, 2]. LVI is not a single molecular state, but may reflect tumor cell plasticity [3], extracellular matrix remodeling [4], endothelial or lymphatic interactions and local nutritional adaptation[5, 6]. The development of gastric cancer is accompanied by changes in the mucosal microbiome[7–9], but the taxa and ecological patterns reported across different populations, tissue compartments and analytical pipelines are inconsistent[10, 11].

β-HB (3-hydroxybutyric acid) is a ketone body, whose level may reflect systemic catabolism, fasting or perioperative status, hepatic ketogenesis[12], local substrate utilization or tumor-associated metabolic adaptation. Upper gastrointestinal cancers are associated with a broader catabolic ketone body signature[13], but the experimental effects of β-HB are context-dependent[14]: it can inhibit tumor growth in some models[15], whereas it may promote tumor progression in others[16]. Therefore, an elevation in tissue β-HB alone cannot be used to infer that it is microbially produced, tumor-promoting or mechanistically upstream of LVI.

The objectives of this study were: (i) to describe overall and taxon-specific microbial differences between LVI groups; (ii) to quantify the directly measured tissue β-HB signal; (iii) to examine the critical paired-sample bridging relationship between *Tundrisphaera* and β-HB; (iv) to determine whether prespecified metabolic and invasion/vascular proxy programs generalize across cohorts with actual LVI and surrogate endpoints; and (v) to prioritize candidate perturbations without performing validation. The analyses explicitly distinguished direct measurements, omics proxies, surrogate endpoints and untested causal hypotheses.

2. Materials and Methods

2.1. 16S rRNA Sequencing Methods

For total genomic DNA from gastric cancer tissue samples, we used the CTAB/SDS method for extraction. We used sterile water to dilute the DNA to approximately 1 μg/μL. DNA quality and concentration were assessed using a 1% agarose gel. The V4 hypervariable region of the bacterial 16S rRNA gene was amplified using the barcoded universal primers 515F and 806R[17, 18]. The PCR reaction mixture included 15 μL of Phusion High-Fidelity PCR Master Mix, 0.2 μM forward and reverse primers and approximately 10 ng of DNA template. The PCR program was: initial denaturation at 98°C for 1 min; 30

cycles, including denaturation at 98°C for 10 s, annealing at 50°C for 30 s and extension at 72°C for 30 s; final extension at 72°C for 5 min.

PCR products were assessed using a 2% agarose gel containing SYBR Green, and pooled at equal concentrations. The pooled PCR products were purified using the Qiagen Gel Extraction Kit. Sequencing libraries were constructed using the TruSeq DNA PCR-Free Sample Preparation Kit, and sample indexes were added. Library quality was evaluated using the Qubit 2.0 Fluorometer and Agilent Bioanalyzer 2100, followed by 250 bp paired-end sequencing on the Illumina NovaSeq platform.

Sequencing data were assigned to individual samples according to unique barcodes. We removed barcode and primer sequences. We used FLASH v1.2.7 to merge paired-end reads[19]. We performed filtering according to the quality-control workflow of QIIME v1.9.1[20]. We aligned the resulting sequences against the Silva database[21], and used the UCHIME algorithm to identify and remove chimeric sequences[22].

2.2. LC–MS Experimental Methods

Metabolomic profiling used a targeted metametabolomics approach, quantitatively analyzing small-molecule metabolites in human gastric cancer tissues. Samples were thawed in an ice bath, approximately 10 mg was weighed from each sample, 20 µL of deionized water and 10 zirconia beads were added, and the samples were homogenized for 3 min. Subsequently, 120 µL of precooled methanol solution containing internal standards was added, the samples were homogenized again for 3 min, and centrifuged at 18,000 × g for 20 min. A 20 µL aliquot of supernatant was transferred to a 96-well plate, 20 µL of freshly prepared derivatization reagent was added on an Eppendorf epMotion workstation, and the reaction proceeded at 30°C for 60 min. Subsequently, 330 µL of 50% methanol diluent precooled in an ice bath was added, followed by centrifugation at 4°C, 4,000 × g for 30 min. A 135 µL aliquot of supernatant was transferred to a new 96-well plate, and LC–MS analysis was performed after the addition of 10 µL of internal standard.

Chromatographic separation was performed using a Waters ACQUITY UPLC-Xevo TQ-S UPLC-MS/MS system. The columns were an ACQUITY UPLC BEH C18 1.7 µm VanGuard guard column (2.1 × 5 mm) and a BEH C18 analytical column (2.1 × 100 mm). The column temperature was 40°C, and the sample manager temperature was 10°C. Mobile phase A was water containing 0.1% formic acid, and mobile phase B was an acetonitrile/isopropanol mixture (70:30). The gradient elution program was: 0–1 min, 5% B; 1–5 min, 5%–30% B; 5–9 min, 30%–50% B; 9–11 min, 50%–78% B; 11–13.5 min, 78%–95% B; 13.5–14 min, 95%–100% B; 14–16 min, 100% B; 16–16.1 min, 100%–5% B; 16.1–18 min, 5% B. The flow rate was 0.40 mL/min, and the injection volume was 5 µL.

Mass spectrometry used electrospray ionization, including positive-ion and negative-ion modes. The ESI+ capillary voltage was 1.5 kV, and the ESI− capillary voltage was 2.0 kV; the source temperature was 150°C, the desolvation temperature was 550°C, and the desolvation gas flow was 1,000 L/h.

Quality control included test mixtures, internal standards, pooled biological sample QC and reagent blanks[23, 24]. Test mixtures were analyzed at the beginning and end of each batch, retention-time shifts were controlled within 4 s, and differences in peak signal intensity were controlled within 15%. Pooled QC samples were prepared alongside the study samples, and injected once after every 14 samples analyzed. Reagent blanks were used to monitor system contamination and wash the chromatographic column. Analysis was completed within 48 h after sample extraction and derivatization. Samples were run using within-group pairing and between-group randomization, to reduce analytical bias caused by run order.

2.3. Study Framework and Evidence Tiers

This study comprised 4 evidence tiers. The discovery tier used 65 gastric cancer tissues with both 16S and targeted LC–MS data; β-HB values were treated as the supplied signal intensities and were not interpreted as absolute concentrations. The direct-LVI external evidence tier used GSE62254 RNA expression [25] and GSE261635 DNA methylation data; both had LVI labels, but neither measured β-HB. The proxy/sensitivity evidence tier used TCGA-STAD survival data [26],, GSE246963 lymph node status, GSE237876 within-patient paired distant-metastasis data, and GSE163558/GSE239676 single-cell data [27, 28]. The computational evidence tier used DepMap and LINCS to rank experimental candidates. LVI, lymph node metastasis, distant metastasis, survival, and cell-state comparisons were not treated as interchangeable endpoints [29, 30] (Table 1 and Table 2).

Table 1
Discovery Cohort and Sample Identity Audit

Item	Clinical records	Matched analysis	Explanation
LVI-negative (V0)	33	32	1 V0 case lacked both molecular data layers
LVI-positive (V1)	33	33	All V1 cases had matched data
Total	66	65	Main multi-omics analysis set
Grouping and clinical labels	—	65/65 concordant	Used for group-based analyses
T0/T1 filename prefixes and clinical labels	—	33/65 concordant	Prefixes were not used as group status
Additional metabolomics row T1_064	—	Excluded	No unambiguous LVI label; ROC AUC was corrected after exclusion

Table 2
Public Cohorts and Their Evidence Roles

Resource	Unit of analysis	Endpoint/measurement	Evidence role
GSE62254	300 tumors; 278 with LVI information	Bulk RNA; actual LVI and survival	RNA proxy test of direct LVI
TCGA-STAD	412 tumors; Cox analysis n = 365	Bulk RNA and survival; no harmonized LVI variable	Co-regulation and prognostic context
GSE246963	8 LN-positive and 8 LN-negative primary lesions	Bulk RNA; lymph node status	Surrogate-endpoint sensitivity analysis
GSE237876	14 patients; 38 primary and 28 metastatic lesions	Within-patient paired bulk RNA	Metastatic-route-specific sensitivity analysis
GSE163558	41,854 cells after quality control	Sample–cell-lineage pseudobulk RNA	Low-powered single-cell feasibility analysis
GSE239676	222,240 cells; 20 patients	Within-patient paired single-cell pseudobulk	Metastasis and coupling sensitivity analysis
GSE261635	10 LVI-positive and 10 LVI-negative cases	EPIC promoter methylation; actual LVI	Orthogonal proxy test of direct LVI
DepMap/LINCS	Stomach-derived models and perturbation expression signatures	CRISPR gene effects and signature reversal	Experimental prioritization only

2.4. Discovery Cohort and Sample Identity Audit

The previous clinical table contained 66 cases of gastric cancer, with 33 cases each in V0 and V1. 1 V0 case lacked both types of molecular data, leaving 65 paired tissues in the final set (V0 n = 32; V1 n = 33). Sample linkages were reconstructed using clinical identifiers, microbiome identifiers and the sample–metabolomics mapping file. In all 65 paired cases, metabolomics class labels were consistent with the audited clinical status. The T0/T1 prefixes in metabolomics filenames agreed with the clinical V0/V1 status in only 33/65 cases, and were therefore not used for grouping. The additional metabolomics record T1_064 had no unambiguous clinical LVI label, and was excluded from all LVI statistics including ROC analysis.

2.5. Microbiome Processing and Global Analysis

The data tables contained 15,301 OTUs and 2,222 genus-level data columns. Each sample in the processed OTU table contained 44,445 sequencing reads. Two-sided Mann–Whitney tests [31] were used to compare the observed number of OTUs, Shannon diversity [32]and Simpson diversity[33]. Bray–Curtis distances [34] were calculated from OTU relative abundances. The compositional sensitivity analysis first restricted genera to those with a detection rate of ≥ 20% and a mean relative abundance of ≥ 0.01%, added a pseudocount of 10^{-6} before closure, and then subjected genus abundances to centered log-ratio

transformation[35, 36]. PERMANOVA used 999 permutations, and a multivariate dispersion test was performed in parallel[37, 38].

2.6. Genus-Level Differential Testing, Two-Stage Candidate Screening and Classification

Global genus-level differential testing retained taxa detected in at least 10% of samples and with a mean relative abundance of ≥ 0.01%. Two-sided Mann–Whitney tests and rank-biserial effect sizes were calculated, followed by BH correction across all tested genera[39, 40]. Transparent, mutually exclusive two-stage descriptive candidate screening was performed alongside this inferential testing. The strict tier required a mean relative abundance in V0 of ≤ 0.01%, a mean relative abundance in V1 of ≥ 0.05%, and an adjusted V1/V0 mean-abundance fold change of ≥ 10. The relaxed additional tier was applied only to genera negative in the strict tier: a cumulative genus-table count in V1 of ≥ 20, and an adjusted mean-abundance fold change of ≥ 10 when the cumulative count in V0 was greater than 0. All cumulative counts were genus-table counts summed by group after the sequencing depth of each sample had first been standardized to 44,445 sequencing reads, and were not equivalent to the number of positive samples, nor did they constitute independent replicates. Fold changes were calculated as (V1 mean relative abundance + ε)/(V0 mean relative abundance + ε), where $\varepsilon = 1\times10^{-8}$; this fold change was not reported when the cumulative count in V0 was 0. The relaxed additional tier was ranked in descending order of cumulative V1 counts. This screening was a data-driven descriptive ranking, and did not yield FDR-positive conclusions.

2.7. Untargeted Metabolomics and β-HB Analysis

β-HB was identified through the entry “3-Hydroxybutyric acid” in the targeted LC–MS assay panel. Its values were the supplied targeted signal intensities and were not interpreted as absolute concentrations. V1 and V0 were compared using a two-sided Mann–Whitney test and the rank-biserial correlation effect size, while retaining the metabolome-wide FDR supplied in the screening data and the FC supplied in the analysis results. Robustness analyses used 4,000 bootstrap resamples of the median difference and leave-one-out between-group tests[41].

The bridging direction was reported using prespecified reverse coding: raw Tundrisphaera abundance and β-HB varied in opposite directions, and the bridging score was significantly correlated with β-HB ($\rho = -0.6797$; $P < 0.001$). After harmonizing the single continuous input, LVI coding, and positive direction, the rank-biserial correlation effect size was calculated as $r_{rb} = 2 \times AUC - 1$; AUC = 0.866 corresponded to $r_{rb} = 0.732$. The two measures were mathematically consistent and were reported together as exploratory results.

A binomial generalized linear model was used to analyze the association between a 1-standard-deviation increase in standardized log(β-HB + 1) and LVI, adjusting for standardized age and binary sex and using HC3 robust standard errors [42, 43]. The analysis was strictly restricted to the 65 samples with unambiguous labels. This model estimate was used to provide a supplementary description of

association strength; it addressed a statistical target different from those of the univariable AUC and the rank-biserial correlation effect size, and no claim of causality was based on it.[44].

2.8. Paired Cross-Omics Bridging and Pathway Audit

In the 65 paired samples, Spearman correlation was used to test the relationship between continuous *Tundrisphaera* abundance and β-HB.[45]

The enriched pathway results were examined for ketone body synthesis/degradation and butanoate metabolism, recording the number of contributing metabolites, raw P values, FDR and pathway impact values[46].

2.9. Prespecified Non-Immune Programs and Bulk Transcriptomic Expression Scoring

6 programs were fixed before public-cohort comparisons: ketone body handling (12 genes), KEGG butanoate metabolism (hsa00650; 27 genes), KEGG fatty acid degradation (hsa00071; 43 genes), EMT/ECM invasion (15 genes), endothelial/angiogenic activity (18 genes) and epithelial integrity (8 genes). In each bulk transcriptomic cohort, the expression of each gene was first standardized across samples, the z scores of the available genes in each program were then averaged, and gene coverage was recorded for each dataset[47, 48].

GSE62254 included 300 cases of gastric cancer from the Asian Cancer Research Group. The LVI comparison included 278 tumors with known LVI status (205 positive, 73 negative), excluding 22 with missing labels. Two-sided Wilcoxon tests and Cliff's δ effect sizes were used[49, 50]. Logistic regression models included standardized program scores, age, sex and stage; overall survival Cox models likewise included standardized program scores, age, sex and stage[51].

TCGA-STAD RNA data included 412 primary tumors. TCGA did not provide an LVI endpoint sufficiently harmonized and suitable for this analysis, and was therefore used only for program co-regulation and survival analyses. The adjusted Cox models included 365 patients and 148 events, with the same covariate structure as the GSE62254 models.

2.10. Surrogate-Endpoint Bulk Transcriptomic Analysis and Paired Metastasis Analysis

GSE246963 included primary tumors with or without lymph node metastasis and paired normal/metastatic samples. The prespecified sensitivity comparison included 8 LN-positive and 8 LN-negative primary tumors. LN status was treated as a surrogate endpoint.

GSE237876 included 38 primary tumor tissue blocks and 28 metastatic tumor tissue blocks from 14 patients, with metastatic routes including hematogenous, peritoneal or ovarian dissemination. Program scores were aggregated within each patient and site category, and patient-level metastatic-minus-primary

differences were tested using paired nonparametric methods. Heterogeneity across metastatic routes and the non-equivalence of distant metastasis and LVI were prespecified as study limitations.

2.11. Single-Cell Processing and Patient-Level Pseudobulk Aggregation

For GSE163558 [27], the local matrix contained 54,687 cell barcodes, of which 41,854 passed quality control. Epithelial-like and endothelial-like cells were stringently selected using broad non-immune lineage markers. Before normalization and gene-set scoring, counts were aggregated at the sample–lineage level, to avoid cell-level pseudoreplication [52–54]. The final epithelial and endothelial pseudobulk analyses included n = 7 and n = 3 sample units, respectively.

For GSE239676, the existing 8,630-gene matrix on the official website was used directly. This matrix contained 20 patients, 68 samples and 222,240 cells. Broad-marker screening identified 28,847 epithelial-like cells and 2,868 endothelial-like cells. The patient-paired primary–solid metastasis comparison included 11 pairs of epithelial-like samples and 4 pairs of endothelial-like samples, with at least 20 cells in each pseudobulk unit. Because CNV-based confirmation of malignant cells was lacking, the epithelial-like label was not treated as evidence of malignant identity [55].

2.12. Direct-LVI Methylation Analysis

GSE261635 included 10 early gastric cancer cases with LVI and 10 without LVI. Each GEO sample table was numerically matched to matrix columns, yielding 20 unique mappings, with best-match Pearson correlation coefficients of 0.9933–0.9967[56].

EPIC probes were linked to the promoter regions of target genes [57, 58]. After quality control, 2,008 promoter CpGs covered 109 genes in the 6 programs. Gene-level promoter methylation was aggregated, negative z scores were used as a proxy for epigenetic activation, and values were averaged within each program. Welch tests and Mann–Whitney tests were used to compare the LVI-positive and LVI-negative groups, and BH correction was performed separately at the pathway, gene and probe levels[59].

2.13. DepMap and LINCS Candidate Prioritization

DepMap CRISPR gene-effect profiles were examined in 54 stomach-derived models, of which 35 had data for at least 1 candidate gene[60–63]. Median gene effects and the proportions of models with effects below − 0.5 and − 1 were reported. The degree of reversal of LINCS genetic perturbation signatures was scored according to prespecified metabolic/phenotypic patterns. These analyses were used only for candidate ranking[64, 65].

The supplied clinical table contained 66 labeled cases (33 V0 and 33 V1). The supplied 16S genus table contained 65 sample columns, with a normalized depth of 44,445 reads; the missing paired clinical record was 21_064 (V0), resulting in 32 V0 and 33 V1 cases in the paired analysis. The supplied metabolomics classification file contained an additional unlabeled record, T1_064, which was excluded from the paired β-HB analysis. In the audited paired data, the β-HB medians were 39.3072 and 68.4851, and the rank-

biserial correlation effect size was 0.732, mathematically consistent with AUC = 0.866 according to r_rb = 2 × AUC − 1.

3. Results

3.1. Cohort Reconstruction Avoided Group-Label and ROC Classification Errors

The clinical records contained 33 V0 and 33 V1 cases, of which 65 could be linked to both molecular data layers. The dedicated mapping table and grouping file were fully concordant across all 65 matched samples, whereas filename prefixes agreed with clinical status in only 33/65 cases. After the unlabeled 66th metabolomics record was removed and the input was harmonized, the β-HB AUC was 0.866 and the rank-biserial correlation effect size was 0.732, satisfying r_rb = 2 × AUC − 1; the two measures could therefore serve as mutually corroborating results for the same univariable input (Table 1).

3.2. Global Microbial Ecology and Microbiome Classification Results Were Both Negative

Shannon diversity values were similar between the two groups, although the difference reached nominal significance (medians of 4.47 and 4.35 in V0 and V1, respectively; $P < 0.05$). Bray–Curtis PERMANOVA showed that between-group differences explained very little variation ($R^2 = 0.0099$; $P < 0.05$), with no difference in dispersion between groups ($P = 0.443$). The Aitchison sensitivity analysis likewise indicated only a small exploratory separation ($R^2 = 0.0197$; $P < 0.05$; dispersion $P = 0.262$).

3.2.1. Two-Stage Screening Identified 4 Low-Abundance Candidates

We obtained 4 low-abundance candidates in descending order of cumulative genus-table counts: the D05-2 group (cumulative V1/V0 counts 50/3; mean relative abundance 0.003409%/0.000211%; adjusted fold change 16.09), *Tundrisphaera* (33/1; 0.002250%/0.000070%; 31.57), *Kushneria marisflavi* (26/0; 0.001773%/not detected; fold change not estimated) and *Sporolactobacillus* (25/0; 0.001705%/not detected; fold change not estimated). These counts were accumulated by group from the genus table after each sample's depth had been standardized to 44,445 sequencing reads. Fold changes were adjusted using $\varepsilon=1\times10^{-8}$. (Figure 1; Table 3).

Figure 1. Gastric Tissue Microbiome Characteristics Stratified by LVI

Status

Figure legend: A, discovery cohort and multi-omics data coverage; B, Shannon diversity in V0 and V1; C, Bray–Curtis community structure; D, cumulative genus-table counts for candidate bacteria in the strict tier and relaxed additional tier; E, out-of-fold ROC of the random forest with information leakage controlled.

Results shown: Overall community differences explained very little variation, and genus-level FDR screening and classification performance did not provide stable evidence. The additional tier yielded 4 low-abundance candidates, including Tundrisphaera, with V1/V0 cumulative genus-table counts of 33/1, mean relative abundances of 0.002250%/0.000070%, and an adjusted fold change of 31.57.

This figure defines the microbiological starting point of the manuscript: the study found no broad or reproducible LVI-associated dysbiosis, and *Tundrisphaera* could only be tested further as a descriptive candidate in the relaxed tier, limiting causal statements about a “key driver bacterium” at their source[66-68].

Table 3. Results of Mutually Exclusive Two-Stage Genus Candidate Screening

Tier/rank	Bacterium	V1 cumulative genus-table count	V0 cumulative genus-table count	V1 mean relative abundance	V0 mean relative abundance	Adjusted fold change
Strict tier	*Cobetia*	4,008	98	0.273269%	0.006891%	39.65×
Relaxed additional tier 1	D05-2 group	50	3	0.003409%	0.000211%	16.09×
Relaxed additional tier 2	*Tundrisphaera*	33	1	0.002250%	0.000070%	31.57×
Relaxed additional tier 3	*Kushneria marisflavi*	26	0	0.001773%	Not detected	—
Relaxed additional tier 4	*Sporolactobacillus*	25	0	0.001705%	Not detected	—

3.3. Nominal Elevation of β-HB in LVI-Positive Gastric Cancer Tissues

The β-HB signal was higher in V1 (n = 33; median, 68.4851; mean, 220.79) than in V0 (n = 32; median, 39.3072; mean, 80.10). The rank-biserial correlation effect size was 0.732, and the Mann–Whitney P value was 0.02203; the AUC for the same single continuous input was 0.866, satisfying r_rb = 2 × AUC − 1. The bootstrap interval for the median difference was 5.02–89.43 signal-intensity units, and P values in the leave-one-out analyses ranged from 0.0107 to 0.0333. The FC supplied in the vendor’s analysis results was 1.803, while the metabolome-wide FDR remained 0.9758. For each 1-standard-deviation increase in log(β-HB + 1), the adjusted OR was 1.72 (95% CI, 0.92–3.23), with the interval including the null value. Thus, this signal was a directly measured but still exploratory tissue-metabolite observation (Figure 2; Supplementary Figures S1–S2; Table 4).

Figure 2. β-HB in the Metabolomic Analysis of the Discovery Cohort

Figure legend: A, box plot comparing LC-MS signal intensities between the V0 and V1 groups; B, volcano plot for the 3-HB matched groups; C, box plot of the grouped sample distribution of β-HB; D, AUC positive-rate correlation plot; F is a histogram of univariate discrimination and age- and sex-adjusted association estimates.

Results shown: β-HB was higher in V1 than in V0 (P = 0.02203; rank-biserial correlation effect size, 0.732; AUC = 0.866, with the two measures satisfying r_rb = 2 × AUC − 1), but the metabolome-wide q value was 0.9758; the adjusted OR was 1.72 (95% CI, 0.92–3.23).

This figure establishes the only directly measured β-HB discovery node in the manuscript, while defining its evidence level through multiple-testing correction and covariate models.

Table 4. Main Discovery Results and Boundaries of the Claims

Analysis	Estimate	Interpretation within the evidence boundaries
Shannon diversity	P<0.05	α diversity differed
Bray–Curtis PERMANOVA	R^2=0.0099; P<0.05	Global community separation remained exploratory
Aitchison PERMANOVA	R^2=0.0197; P<0.05	Compositional sensitivity analysis remained exploratory
Microbiome random forest	AUC=0.706 (95% CI 0.592–0.843)	Internal discrimination remained exploratory
Two-stage genus candidate screening	*Cobetia* was identified in the strict tier; 4 low-abundance candidates were obtained in the relaxed additional tier	Descriptive, data-driven, without genus-level FDR support; see Table 3
β-HB V1 vs V0	Medians 68.49 vs 39.31; P<0.05; q=0.9758	Directly measured nominal signal, not an omics-wide positive result
β-HB adjusted model	OR=1.72 (0.92–3.23); P<0.01	Imprecise estimate after age/sex adjustment
β-HB discrimination	AUC=0.866 (95% CI 0.792–0.943)	Exploratory result; included only 65 samples with unambiguous labels
Tundrisphaera–β-HB bridging	ρ=0.6797; P<0.001	An inverse sample-level bridging relationship was observed

3.4. The Matched Microbiota–Metabolism Bridging Test and the Pathway Independence Test Were Both Negative

In the 65 matched tissue samples, raw continuous Tundrisphaera abundance and β-HB varied in opposite directions and showed a significant inverse bridging relationship (Spearman ρ = −0.6797; P < 0.001). This sample-level association supports a directional statistical link between Tundrisphaera and β-HB, but cannot by itself establish that Tundrisphaera produces β-HB or that either variable lies upstream of LVI.

The ketone-body pathway in the screening results was represented by only one metabolite, β-HB (raw P = 0.00746; q = 0.597). Butyrate metabolism was likewise represented only by β-HB (raw P = 0.0491; q = 0.674). That is, a significant Tundrisphaera–β-HB bridging relationship and pathways driven by the β-HB feature constitute a causal mechanism (Figure 3; Supplementary Figures S3–S5).

Figure 3. Matched Cross-Omics Bridging and Single-Hit Pathway Audit

Figure legend: A, β-HB grouped by LVI status; B, Priority between the metabolism candidates; C, raw P values for ketone body synthesis and degradation and butanoate metabolism；D,Top differential metabolite z-score heatmap.

Results shown: β-HB was nominally elevated in V1; in the 65 audited paired samples, raw Tundrisphaera abundance showed an inverse bridging relationship with β-HB (ρ = −0.6797; P < 0.001).

The ketone-body pathway and butyrate metabolism were both represented only by the same β-HB feature, with raw P values of 0.00746 (q = 0.597) and 0.0491, respectively.

This figure supports a sample-level inverse bridging relationship between Tundrisphaera and β-HB. Because both pathways were represented by the same β-HB annotation, the ketone-body and butyrate pathways are not independent findings either; therefore, this article treats the genus candidate and β-HB as two interrelated but still exploratory observations.

3.5. RNA Validation with Actual LVI Labels

The Wilcoxon q values for all three metabolic programs were 0.1977. One output version of the ketone-body handling model reported OR = 2.682 (95% CI, 2.503–2.925; P = 0.0139), while another output summary reported a BH-adjusted q value of 0.0434 (Figure 4; Table 5).

Figure 4. RNA Proxy Validation of Actual LVI in GSE62254

Figure legend: The figure shows the six prespecified RNA programs in samples with actual LVI labels, including program scores in the LVI-negative and LVI-positive groups, the correlation structure between programs, unadjusted associations and LVI association estimates adjusted for clinical variables.

Results shown: The Wilcoxon q values for all three metabolic programs were 0.1977. The reported outputs of the ketone-body handling model included OR = 2.682 (95% CI, 2.503–2.925; P = 0.0139) and a BH-adjusted q value of 0.0434 .

Role in the manuscript's logic: This figure reproduces the prespecified direction that "lower metabolic programs correspond to LVI positivity", limiting extrapolation from the β-HB signal in the discovery cohort to RNA programs and LVI.

Table 5. External Results Constraining the Proposed Mechanism

Dataset	Main results	Interpretation
GSE62254 actual LVI	Wilcoxon metabolic-program q=0.1977; ketone model OR=2.682 (95% CI 2.503–2.925), reported P=0.0139; separate BH summary q=0.0434	-
GSE62254 survival	EMT HR=1.289, q=0.0064; endothelial HR=1.261, q=0.0064	Phenotypic programs were prognostically associated, but there was no evidence of metabolic mediation
TCGA-STAD survival	EMT HR=1.308, q=0.0143; endothelial HR=1.242, q=0.0488; metabolic programs q≈0.45	Prognostic signals were mainly located in phenotypic programs
GSE246963 LN status	q>0.05 for all 6 programs	Directional sensitivity analysis only
GSE261635 actual LVI	q=1.000 for all 6 pathways; minimum gene q=0.9026	Negative orthogonal methylation evidence
DepMap/LINCS	SLC16A1 median effect=−0.197; OXCT1 reversal=5.17	Candidate ranking, rather than validation of dependency or efficacy

3.6. Direct-LVI Methylation Analysis Provided Negative Orthogonal Evidence

All 20 sample–matrix column mappings in GSE261635 passed the prespecified uniqueness audit. Among the 109 target genes covered, the minimum q value was 0.9026; among the 2,008 promoter probes, the minimum q value was 0.9960. Although VEGFC had a raw P=0.0106, Δβ=−0.0579, reaching the nominal level, its q=0.903, and it was therefore not selected as a positive result. Thus, the only public orthogonal dataset with actual LVI labels failed to validate the proposed metabolic, invasion and vascular programs (Figure 5; Table 5).

Figure 5. EPIC Methylation Validation of Direct LVI in GSE261635

Figure legend: A, verification of one-to-one mappings between the 20 samples and data matrix columns; B, effect estimates of the six promoter-methylation program proxies; C, sample distributions for each program; D, ranking of nominal effects for target genes and promoter probes.

Figure results: All 20/20 samples passed the unique-mapping audit; the minimum q value for the 109 target genes was 0.9026, and the minimum q value for the 2,008 promoter probes was 0.9960. Although VEGFC had a raw P=0.0106 and Δβ=−0.0579, q=0.903.

This figure provides an orthogonal test at a different molecular layer and with an actual LVI endpoint, but did not support the proposed metabolic, invasion or vascular programs, further limiting the statement that “the mechanism has been replicated across omics”; methylation proxies still cannot replace β-HB, RNA or flux measurements.

3.7. Metabolic Transcriptional Programs Co-varied, but Survival Associations Were Mainly Located in Phenotypic Programs

In TCGA-STAD, ketone body handling was strongly correlated with both butanoate metabolism (ρ=0.833) and fatty acid degradation (ρ=0.826). The correlation coefficient between butanoate metabolism and fatty acid degradation was ρ=0.841. EMT/ECM invasion (adjusted HR=1.308, q=0.0143) and endothelial/angiogenic activity (HR=1.242, q=0.0488) were associated with poorer survival (Figure 6; Table 5).

GSE62254 showed the same separation: EMT/ECM invasion (HR=1.289, q=0.0064) and endothelial/angiogenic activity (HR=1.261, q=0.0064) were adverse factors, whereas epithelial integrity was a protective factor (HR=0.786, q=0.0024); none of the three metabolic programs passed BH correction. These survival associations support the clinical relevance of invasion and vascular phenotypes, but cannot link them to β-HB (Table 5).

Figure 6. Metabolic Co-regulation and Survival Context in TCGA-STAD

Figure legend: The figure includes the six-program correlation matrix, the sample-level relationship between fatty acid degradation and EMT/ECM invasion, covariate-adjusted Cox regression results, and the distribution of the EMT/ECM program across pathological stages.

Figure results: Ketone body handling, butanoate metabolism and fatty acid degradation programs were strongly correlated at the RNA level; adjusted survival associations were mainly observed for EMT/ECM invasion (HR=1.308, q=0.0143) and the endothelial/angiogenic program (HR=1.242, q=0.0488).

This figure supports the co-variation of metabolic transcriptional programs, and confirms the clinical relevance of invasion and vascular phenotypes.

3.8. Surrogate-Outcome Bulk Transcriptomic and Paired Metastasis Cohorts Showed Only Directional or Negative Results

Among the eight LN-positive and eight LN-negative primary tumors in GSE246963, the effect directions for ketone body handling, butanoate metabolism and fatty acid degradation were all upward. Butanoate metabolism showed the strongest nominal result (Cliff's δ effect size=0.625, P=0.0406, q=0.1217). LN metastasis is not equivalent to histological LVI, and this cohort did not measure β-HB (Supplementary Figure S6).

Role in the manuscript's logic: This figure shows that directional changes can demonstrate program coupling, and that distant metastasis cannot substitute for LVI; it limits generalization of the metabolism–invasion relationship to multi-site metastatic contexts, and highlights patient, route and organ heterogeneity.

3.9. Single-Cell Sensitivity Analysis Failed to Reproduce Metabolism–Invasion Coupling

A total of 41,854 cells in GSE163558 passed quality control (QC), but the sample-level epithelial and endothelial pseudobulk analyses contained only seven and three units, respectively. In the epithelial pseudobulk analysis, the correlation coefficients for ketone body handling with EMT/ECM, and butanoate metabolism with EMT/ECM were both ρ=0.143 (P=0.760), and the correlation coefficient for fatty acid degradation with EMT/ECM was ρ=0.321 (P=0.482; q=0.760). Metabolism–invasion or metabolism–vascular coupling was validated in both cases, but statistical power for endothelial inference was insufficient (Supplementary Figure S8).

3.10. In Silico Perturbations Identified Experimental Leads

In stomach-derived DepMap models, SLC16A1 showed the strongest median dependency among the candidate genes (median gene effect=−0.197); among models with data, only 5.7% had gene effects below −0.5, and 2.9% below −1. OXCT1 knockout produced the highest LINCS reversal overlap score (5.17), suggesting ketone body utilization as a candidate direction for controlled perturbation. Compound rankings and gene perturbations remain only experimental leads, because this analysis demonstrates direct binding, β-HB blockade, phenotypic rescue or efficacy against LVI (Supplementary Figure S10).

4. Discussion

This integrated analysis decomposes the originally proposed sequential mechanism into two types of independent exploratory observations. The overall microbiome did not differ by LVI status, and no genus

passed multiple-testing correction; the descriptive two-stage screening identified *Cobetia* in the strict tier, and identified 4 low-abundance candidates in the relaxed tier, including *Tundrisphaera*, among genera negative in the strict tier.

Broad ecological null results and descriptive enrichment of low-abundance taxa can coexist. The former suggests that in the current cohort, LVI was not accompanied by reproducible community-wide changes; the latter is used only to identify candidates warranting technical replication. The strict tier and the relaxed additional tier are mutually exclusive, avoiding counting the same genus in both tiers; *Tundrisphaera* did not meet the strict-tier requirement of a mean abundance in V1 of ≥0.05%, and was selected only in the data-driven relaxed additional tier. Its cumulative V1/V0 genus-table counts were 33/1, and its mean relative abundances were only 0.002250%/0.000070%. Cumulative genus-table counts are not the number of positive samples, nor are they independent biological replicates, and cannot be used to infer prevalence. In the absence of extraction blanks, batch reconstruction, absolute bacterial loads and orthogonal species-level assays , contamination and taxonomic misclassification cannot be excluded[69-73]. The current results should guide validation experiments, rather than assign specific functions to this microorganism.

Compared with RNA proxies, the β-HB observation is closer to the research question at the measurement level, because it derives directly from tissue metabolomics stratified by LVI. However, its statistical evidence remains exploratory: the raw P value reached only the nominal level, the omics-wide FDR was 0.976, and the adjusted confidence interval included the null value[74, 75]. Its source also remains unclear. Systemic catabolism, perioperative fasting, liver-derived ketogenesis , tumor utilization, stromal metabolism and local microbial contributions are all possible. Targeted LC–MS using isotope-labeled standards[76, 77], concurrent plasma measurements , fasting-related metadata and enantiomer-sensitive assays are needed, to distinguish its concentration and source[78, 79].

The ketone body and butanoate metabolism pathway results do not constitute two independent findings. Both are driven by the same β-HB annotation , and neither passed pathway-level FDR correction. A single detected metabolite cannot establish pathway direction, the relationship between synthesis and utilization or metabolic flux [80-82]. Co-regulation of ketone body, butanoate and fatty acid programs in public RNA data provides biological context, but gene expression cannot substitute for metabolite concentrations or isotope-traced metabolic flux[83].

The strongest external constraint is not a positive proxy result, but transparently reported non-replication. GSE62254 had a sufficient number of LVI-labeled samples, but none of the six program comparisons passed BH correction, and the adjusted ketone body handling effect was opposite in direction to the working hypothesis. GSE261635 provided a different molecular layer and an actual LVI outcome, but none of the three testing families at the pathway, gene and probe levels passed BH correction. Because neither cohort measured β-HB, these results cannot negate the direct β-HB measurement in the discovery cohort, but weaken the claim that “it is accompanied by generalizable transcriptional or promoter-methylation programs”.

The survival analysis results reveal a separable phenomenon: EMT/ECM and endothelial/angiogenic programs were clinically relevant in both GSE62254 and TCGA-STAD. Because the metabolic programs were not prognostically associated, and the directions of inter-program correlations were unstable across cohorts, the above prognostic associations should therefore not be attributed to β-HB. Similarly, lymph node and distant metastasis datasets can serve as sensitivity analyses, but cannot replace histological LVI.

There was no broad dependency in gastric cancer models, and therefore the candidate genes cannot be described as established vulnerability targets. Nevertheless, OXCT1, SLC16A1 and related ketone body handling genes can still enter subsequent studies, because the LINCS and DepMap results define actionable perturbation nodes. Informative experiments should test whether genetic or pharmacological perturbations alter intracellular/extracellular β-HB, invasion, EMT markers, endothelial permeability or tube-forming capacity, and whether add-back or rescue can restore the phenotype[84]. Compound ranking alone, without target engagement, dose–response, rescue experiments and in vivo LVI assessment, cannot constitute therapeutic validation.

The next study should first independently replicate the two types of findings[85]. The thresholds and mutual-exclusivity rules for the strict tier and relaxed additional tier should be locked in advance, *Cobetia* and the 4 low-abundance candidates should be re-examined in an independent cohort, and *Tundrisphaera* should be quantified by species-level sequencing or qPCR , while measuring absolute bacterial loads , and including extraction blanks, PCR-negative controls, positive controls and randomized batches [86-90]. Tissue and plasma β-HB should be measured using targeted LC–MS standardized with isotope-labeled internal standards, and fasting and perioperative metadata should be recorded. Subsequently, the continuous cross-omics bridging relationship should be re-examined without data-driven threshold selection.

β-HB dose, treatment duration, the physiological concentration range, and stereoisomers should be controlled in at least two authenticated, mycoplasma-free gastric cancer models and vascular or lymphatic endothelial co-culture systems[91-93]. Assay endpoints should include invasion [98], migration, EMT, endothelial permeability, tube formation , and ketone body/butanoate/fatty acid metabolic flux[94, 95]. Candidate genes should be perturbed through blockade and rescue. In vivo studies should use prespecified histological LVI endpoints, rather than relying only on tumor size or distant metastasis. The core endpoint should be transendothelial migration across an intact endothelial layer, with simultaneous assessment of endothelial junction continuity, permeability, and controls for cell viability and proliferation. Tumor-cell-only, endothelial-cell-only, and combined exposure should be compared to identify the responding cell type. Candidate genes should then be perturbed in the responding cells: OXCT1 deletion or knockdown, wild-type rescue, and rescue with a validated catalytically inactive variant should be tested, with explicit examination of the β-HB × genotype interaction.

5. Conclusions

Tundrisphaera was identified among the 4 low-abundance candidates; it did not meet the strict-tier abundance requirement. The β-HB signal was nominally elevated in V1 tissues, and the audited paired analysis identified an inverse bridging relationship between continuous Tundrisphaera abundance and β-HB ($\rho = -0.6797$; $P < 0.001$). These findings remain exploratory; a causal β-HB–LVI pathway or a validated OXCT1-dependent mechanism exists.

Declarations

Consent for Publication: Everyparticipant agree to publish the paper

Conflict of Interest Disclosure

All authors declare that they have no financial or personal relationships with individuals or organizations that could inappropriately influence or be perceived to influence the work reported in this manuscript.

The authors declare that they have no relevant financial interests, consultancy arrangements, employment relationships, honoraria, ownership interests, patent applications, or other competing interests related to this study

Author Contributions: Conceptualization: D.W., P.R., J.Y.G.; Writing - review and editing: G.L., R.P., Q.W., J.Y.G., Z.W., and Y.S.; Project administration: Q.W., M.R. and T.T.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Research Project of Jiangsu Cancer Hospital (No. XHMS202401, RCQY202501, and ZBDL202502), the Yishan Research Project of Jiangsu Cancer Hospital (No. YSZD202404), the Qunfeng Project of Jiangsu Cancer Hospital (No. GFXK202503), and the Jiangsu Province 333 Project High-Level Talent Training Program (No. 2024-3-2249).

The funding sources had no role in the study design, data collection, data analysis, interpretation of the results, manuscript preparation, or the decision to submit the manuscript for publication.

Informed consent: Informed consent was obtained from all patients.

Ethics approval All procedures in this study were in accordance with ethical standards of the institutional and national research committee and with the Helsinki Declaration of 1964 and later versions. This study was approved by the Ethics Committee of the Affiliated Cancer Hospital of Nanjing Medical University.

Declaration of competing interests:**T**he authors declare that they have no conflicts of interest or financial ties to disclose.

Data availability: Human sequencing and metabolomics raw data should could be shared within the scope allowed by the original ethical approvals and institutional data governance(https://dataview.ncbi.nlm.nih.gov/object/PRJNA1538231?

reviewer=59kjbgi9996fsr0rmi90i9abpl). Public datasets can be accessed through GSE62254, GSE246963, GSE237876, GSE163558, GSE239676, GSE261635, and TCGA-STAD.

Author Approval

All authors have read and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

References

1. Fujikawa, H., et al., *The Clinical Significance of Lymphovascular Invasion in Gastric Cancer.* In Vivo, 2020. **34**(3): p. 1533–1539.
2. Zhang, C.-D., et al., *Lymphovascular Invasion as a Predictor for Lymph Node Metastasis and a Prognostic Factor in Gastric Cancer Patients under 70 Years of Age: A Retrospective Analysis.* International Journal of Surgery, 2018. **53**: p. 214–220.
3. Yang, J., et al., *Guidelines and Definitions for Research on Epithelial–Mesenchymal Transition.* Nature Reviews Molecular Cell Biology, 2020. **21**(6): p. 341–352.
4. Levental, K.R., et al., *Matrix Crosslinking Forces Tumor Progression by Enhancing Integrin Signaling.* Cell, 2009. **139**(5): p. 891–906.
5. Skobe, M., et al., *Induction of Tumor Lymphangiogenesis by VEGF-C Promotes Breast Cancer Metastasis.* Nature Medicine, 2001. **7**(2): p. 192–198.
6. Hensley, C.T., et al., *Metabolic Heterogeneity in Human Lung Tumors.* Cell, 2016. **164**(4): p. 681–694.
7. Ferreira, R.M., et al., *Gastric Microbial Community Profiling Reveals a Dysbiotic Cancer-Associated Microbiota.* Gut, 2018. **67**(2): p. 226–236.
8. Dicksved, J., et al., *Molecular Characterization of the Stomach Microbiota in Patients with Gastric Cancer and in Controls.* Journal of Medical Microbiology, 2009. **58**(4): p. 509–516.
9. Aviles-Jimenez, F., et al., *Stomach Microbiota Composition Varies between Patients with Non-Atrophic Gastritis and Patients with Intestinal Type of Gastric Cancer.* Scientific Reports, 2014. **4**: p. 4202.
10. Liu, X., et al., *Alterations of Gastric Mucosal Microbiota across Different Stomach Microhabitats in a Cohort of 276 Patients with Gastric Cancer.* EBioMedicine, 2019. **40**: p. 336–348.
11. Coker, O.O., et al., *Mucosal Microbiome Dysbiosis in Gastric Carcinogenesis.* Gut, 2018. **67**(6): p. 1024–1032.
12. Owen, O.E., et al., *Brain Metabolism during Fasting.* Journal of Clinical Investigation, 1967. **46**(10): p. 1589–1595.
13. von Renesse, J., et al., *Tumour Catabolism Independent of Malnutrition and Inflammation in Upper GI Cancer Patients Revealed by Longitudinal Metabolomics.* Journal of Cachexia, Sarcopenia and Muscle, 2023. **14**(1): p. 298–309.
14. Shimazu, T., et al., *Suppression of Oxidative Stress by β-Hydroxybutyrate, an Endogenous Histone Deacetylase Inhibitor.* Science, 2013. **339**(6116): p. 211–214.

15. Dmitrieva-Posocco, O., et al., *β-Hydroxybutyrate Suppresses Colorectal Cancer.* Nature, 2022. **605**(7908): p. 160–165.
16. Jiang, W., et al., *β-Hydroxybutyrate Promotes Cancer Metastasis through β-Hydroxybutyrylation-Dependent Stabilization of Snail.* Nature Communications, 2025. **16**: p. 6592.
17. Caporaso, J.G., et al., *Global Patterns of 16S rRNA Diversity at a Depth of Millions of Sequences per Sample.* Proceedings of the National Academy of Sciences, 2011. **108**(Supplement 1): p. 4516–4522.
18. Walters, W., et al., *Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Surveys.* mSystems, 2016. **1**(1): p. e00009–15.
19. Magoč, T. and S.L. Salzberg, *FLASH: Fast Length Adjustment of Short Reads to Improve Genome Assemblies.* Bioinformatics, 2011. **27**(21): p. 2957–2963.
20. Caporaso, J.G., et al., *QIIME Allows Analysis of High-Throughput Community Sequencing Data.* Nature Methods, 2010. **7**(5): p. 335–336.
21. Quast, C., et al., *The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools.* Nucleic Acids Research, 2013. **41**(D1): p. D590–D596.
22. Edgar, R.C., et al., *UCHIME Improves Sensitivity and Speed of Chimera Detection.* Bioinformatics, 2011. **27**(16): p. 2194–2200.
23. Broadhurst, D., et al., *Guidelines and Considerations for the Use of System Suitability and Quality Control Samples in Mass Spectrometry Assays Applied in Untargeted Clinical Metabolomic Studies.* Metabolomics, 2018. **14**(6): p. 72.
24. Dunn, W.B., et al., *Procedures for Large-Scale Metabolic Profiling of Serum and Plasma Using Gas Chromatography and Liquid Chromatography Coupled to Mass Spectrometry.* Nature Protocols, 2011. **6**(7): p. 1060–1083.
25. Cristescu, R., et al., *Molecular Analysis of Gastric Cancer Identifies Subtypes Associated with Distinct Clinical Outcomes.* Nature Medicine, 2015. **21**(5): p. 449–456.
26. Cancer Genome Atlas Research, N., *Comprehensive Molecular Characterization of Gastric Adenocarcinoma.* Nature, 2014. **513**(7517): p. 202–209.
27. Cheng, X., et al., *Atlas of Metastatic Gastric Cancer Links Ferroptosis to Disease Progression and Immunotherapy Response.* Gastroenterology, 2024. **167**(7): p. 1345–1357.
28. Jiang, H., et al., *Revealing the Transcriptional Heterogeneity of Organ-Specific Metastasis in Human Gastric Cancer Using Single-Cell RNA Sequencing.* Clinical and Translational Medicine, 2022. **12**(2): p. e730.
29. Fleming, T.R. and D.L. DeMets, *Surrogate End Points in Clinical Trials: Are We Being Misled?* Annals of Internal Medicine, 1996. **125**(7): p. 605–613.
30. Prentice, R.L., *Surrogate Endpoints in Clinical Trials: Definition and Operational Criteria.* Statistics in Medicine, 1989. **8**(4): p. 431–440.
31. Mann, H.B. and D.R. Whitney, *On a Test of Whether One of Two Random Variables Is Stochastically Larger than the Other.* Annals of Mathematical Statistics, 1947. **18**(1): p. 50–60.

32. Shannon, C.E., *A Mathematical Theory of Communication.* Bell System Technical Journal, 1948. **27**(3): p. 379–423.
33. Simpson, E.H., *Measurement of Diversity.* Nature, 1949. **163**: p. 688.
34. Bray, J.R. and J.T. Curtis, *An Ordination of the Upland Forest Communities of Southern Wisconsin.* Ecological Monographs, 1957. **27**(4): p. 325–349.
35. Gloor, G.B., et al., *Microbiome Datasets Are Compositional: And This Is Not Optional.* Frontiers in Microbiology, 2017. **8**: p. 2224.
36. Aitchison, J., *The Statistical Analysis of Compositional Data.* Journal of the Royal Statistical Society: Series B (Methodological), 1982. **44**(2): p. 139–160.
37. Anderson, M.J., *A New Method for Non-Parametric Multivariate Analysis of Variance.* Austral Ecology, 2001. **26**(1): p. 32–46.
38. Anderson, M.J., *Distance-Based Tests for Homogeneity of Multivariate Dispersions.* Biometrics, 2006. **62**(1): p. 245–253.
39. Benjamini, Y. and Y. Hochberg, *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.* Journal of the Royal Statistical Society: Series B (Methodological), 1995. **57**(1): p. 289–300.
40. Kerby, D.S., *The Simple Difference Formula: An Approach to Teaching Nonparametric Correlation.* Comprehensive Psychology, 2014. **3**: p. 1.
41. Efron, B., *Bootstrap Methods: Another Look at the Jackknife.* Annals of Statistics, 1979. **7**(1): p. 1–26.
42. Nelder, J.A. and R.W.M. Wedderburn, *Generalized Linear Models.* Journal of the Royal Statistical Society: Series A (General), 1972. **135**(3): p. 370–384.
43. MacKinnon, J.G. and H. White, *Some Heteroskedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties.* Journal of Econometrics, 1985. **29**(3): p. 305–325.
44. Hanley, J.A. and B.J. McNeil, *The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve.* Radiology, 1982. **143**(1): p. 29–36.
45. Spearman, C., *The Proof and Measurement of Association between Two Things.* American Journal of Psychology, 1904. **15**(1): p. 72–101.
46. Xia, J. and D.S. Wishart, *MetPA: A Web-Based Metabolomics Tool for Pathway Analysis and Visualization.* Bioinformatics, 2010. **26**(18): p. 2342–2344.
47. Lee, E., et al., *Inferring Pathway Activity toward Precise Disease Classification.* PLoS Computational Biology, 2008. **4**(11): p. e1000217.
48. Kanehisa, M. and S. Goto, *KEGG: Kyoto Encyclopedia of Genes and Genomes.* Nucleic Acids Research, 2000. **28**(1): p. 27–30.
49. Cliff, N., *Dominance Statistics: Ordinal Analyses to Answer Ordinal Questions.* Psychological Bulletin, 1993. **114**(3): p. 494–509.
50. Wilcoxon, F., *Individual Comparisons by Ranking Methods.* Biometrics Bulletin, 1945. **1**(6): p. 80–83.

51. Cox, D.R., *Regression Models and Life-Tables.* Journal of the Royal Statistical Society: Series B (Methodological), 1972. **34**(2): p. 187–202.
52. Squair, J.W., et al., *Confronting False Discoveries in Single-Cell Differential Expression.* Nature Communications, 2021. **12**: p. 5692.
53. Crowell, H.L., et al., *muscat Detects Subpopulation-Specific State Transitions from Multi-Sample Multi-Condition Single-Cell Transcriptomics Data.* Nature Communications, 2020. **11**: p. 6077.
54. Zimmerman, K.D., M.A. Espeland, and C.D. Langefeld, *A Practical Solution to Pseudoreplication Bias in Single-Cell Studies.* Nature Communications, 2021. **12**: p. 738.
55. Tirosh, I., et al., *Dissecting the Multicellular Ecosystem of Metastatic Melanoma by Single-Cell RNA-seq.* Science, 2016. **352**(6282): p. 189–196.
56. Barrett, T., et al., *NCBI GEO: Archive for Functional Genomics Data Sets—Update.* Nucleic Acids Research, 2013. **41**(D1): p. D991–D995.
57. Pidsley, R., et al., *Critical Evaluation of the Illumina MethylationEPIC BeadChip Microarray for Whole-Genome DNA Methylation Profiling.* Genome Biology, 2016. **17**: p. 208.
58. Roadmap Epigenomics, C., et al., *Integrative Analysis of 111 Reference Human Epigenomes.* Nature, 2015. **518**(7539): p. 317–330.
59. Welch, B.L., *The Generalization of 'Student's' Problem When Several Different Population Variances Are Involved.* Biometrika, 1947. **34**(1-2): p. 28–35.
60. Dempster, J.M., et al., *Agreement between Two Large Pan-Cancer CRISPR-Cas9 Gene Dependency Data Sets.* Nature Communications, 2019. **10**: p. 5817.
61. Dempster, J.M., et al., *Chronos: A Cell Population Dynamics Model of CRISPR Experiments That Improves Inference of Gene Fitness Effects.* Genome Biology, 2021. **22**: p. 343.
62. Tsherniak, A., et al., *Defining a Cancer Dependency Map.* Cell, 2017. **170**(3): p. 564–576.e16.
63. Ghandi, M., et al., *Next-Generation Characterization of the Cancer Cell Line Encyclopedia.* Nature, 2019. **569**(7757): p. 503–508.
64. Lamb, J., et al., *The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease.* Science, 2006. **313**(5795): p. 1929–1935.
65. Subramanian, A., et al., *A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles.* Cell, 2017. **171**(6): p. 1437–1452.e17.
66. Varma, S. and R. Simon, *Bias in Error Estimation When Using Cross-Validation for Model Selection.* BMC Bioinformatics, 2006. **7**: p. 91.
67. Breiman, L., *Random Forests.* Machine Learning, 2001. **45**(1): p. 5–32.
68. Ambroise, C. and G.J. McLachlan, *Selection Bias in Gene Extraction on the Basis of Microarray Gene-Expression Data.* Proceedings of the National Academy of Sciences, 2002. **99**(10): p. 6562–6566.
69. Eisenhofer, R., et al., *Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations.* Trends in Microbiology, 2019. **27**(2): p. 105–117.

70. Nejman, D., et al., *The Human Tumor Microbiome Is Composed of Tumor Type–Specific Intracellular Bacteria.* Science, 2020. **368**(6494): p. 973–980.
71. Vandeputte, D., et al., *Quantitative Microbiome Profiling Links Gut Community Variation to Microbial Load.* Nature, 2017. **551**(7681): p. 507–511.
72. Salter, S.J., et al., *Reagent and Laboratory Contamination Can Critically Impact Sequence-Based Microbiome Analyses.* BMC Biology, 2014. **12**: p. 87.
73. Davis, N.M., et al., *Simple Statistical Identification and Removal of Contaminant Sequences in Marker-Gene and Metagenomics Data.* Microbiome, 2018. **6**: p. 226.
74. Huang, D., et al., *Hepatocellular Carcinoma Redirects to Ketolysis for Progression under Nutrition Deprivation Stress.* Cell Research, 2016. **26**(10): p. 1112–1130.
75. Martinez-Outschoorn, U.E., et al., *Ketone Body Utilization Drives Tumor Growth and Metastasis.* Cell Cycle, 2012. **11**(21): p. 3964–3971.
76. Puchalska, P., et al., *Determination of Ketone Bodies in Biological Samples via Rapid UPLC-MS/MS.* Talanta, 2021. **225**: p. 122048.
77. Want, E.J., et al., *Global Metabolic Profiling of Animal and Human Tissues via UPLC-MS.* Nature Protocols, 2013. **8**(1): p. 17–32.
78. Stubbs, B.J., et al., *On the Metabolism of Exogenous Ketones in Humans.* Frontiers in Physiology, 2017. **8**: p. 848.
79. Sullivan, M.R., et al., *Quantification of Microenvironmental Metabolites in Murine Cancers Reveals Determinants of Tumor Nutrient Availability.* eLife, 2019. **8**: p. e44235.
80. Jang, C., L. Chen, and J.D. Rabinowitz, *Metabolomics and Isotope Tracing.* Cell, 2018. **173**(4): p. 822–837.
81. Sumner, L.W., et al., *Proposed Minimum Reporting Standards for Chemical Analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI).* Metabolomics, 2007. **3**(3): p. 211–221.
82. Buescher, J.M., et al., *A Roadmap for Interpreting 13C Metabolite Labeling Patterns from Cells.* Current Opinion in Biotechnology, 2015. **34**: p. 189–201.
83. Chubukov, V., et al., *Transcriptional Regulation Is Insufficient to Explain Substrate-Induced Flux Changes in Bacillus subtilis.* Molecular Systems Biology, 2013. **9**: p. 709.
84. Garcia, C.K., et al., *Molecular Characterization of a Membrane Transporter for Lactate, Pyruvate, and Other Monocarboxylates: Implications for the Cori Cycle.* Cell, 1994. **76**(5): p. 865–873.
85. Mirzayi, C., et al., *Reporting Guidelines for Human Microbiome Research: The STORMS Checklist.* Nature Medicine, 2021. **27**(11): p. 1885–1892.
86. Props, R., et al., *Absolute Quantification of Microbial Taxon Abundances.* ISME Journal, 2017. **11**(2): p. 584–587.
87. Sinha, R., et al., *Assessment of Variation in Microbial Community Amplicon Sequencing by the Microbiome Quality Control (MBQC) Project Consortium.* Nature Biotechnology, 2017. **35**(11): p. 1077–1086.

88. McLaren, M.R., A.D. Willis, and B.J. Callahan, *Consistent and Correctable Bias in Metagenomic Sequencing Experiments.* eLife, 2019. **8**: p. e46923.
89. Bustin, S.A., et al., *The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments.* Clinical Chemistry, 2009. **55**(4): p. 611–622.
90. Sørensen, L.K., et al., *Simultaneous Determination of β-Hydroxybutyrate and β-Hydroxy-β-Methylbutyrate in Human Whole Blood Using Hydrophilic Interaction Liquid Chromatography Electrospray Tandem Mass Spectrometry.* Clinical Biochemistry, 2013. **46**(18): p. 1877–1883.
91. Yan, H.H.N., et al., *A Comprehensive Human Gastric Cancer Organoid Biobank Captures Tumor Subtype Heterogeneity and Enables Therapeutic Screening.* Cell Stem Cell, 2018. **23**(6): p. 882–897.e11.
92. Nanki, K., et al., *Divergent Routes toward Wnt and R-spondin Niche Independency during Human Gastric Carcinogenesis.* Cell, 2018. **174**(4): p. 856–869.e17.
93. Bartfeld, S., et al., *In Vitro Expansion of Human Gastric Epithelial Stem Cells and Their Responses to Bacterial Infection.* Gastroenterology, 2015. **148**(1): p. 126–136.e6.
94. Arnaoutova, I. and H.K. Kleinman, *In Vitro Angiogenesis: Endothelial Cell Tube Formation on Gelled Basement Membrane Extract.* Nature Protocols, 2010. **5**(4): p. 628–635.
95. Liang, C.-C., A.Y. Park, and J.-L. Guan, *In Vitro Scratch Assay: A Convenient and Inexpensive Method for Analysis of Cell Migration In Vitro.* Nature Protocols, 2007. **2**(2): p. 329–333.

Figures

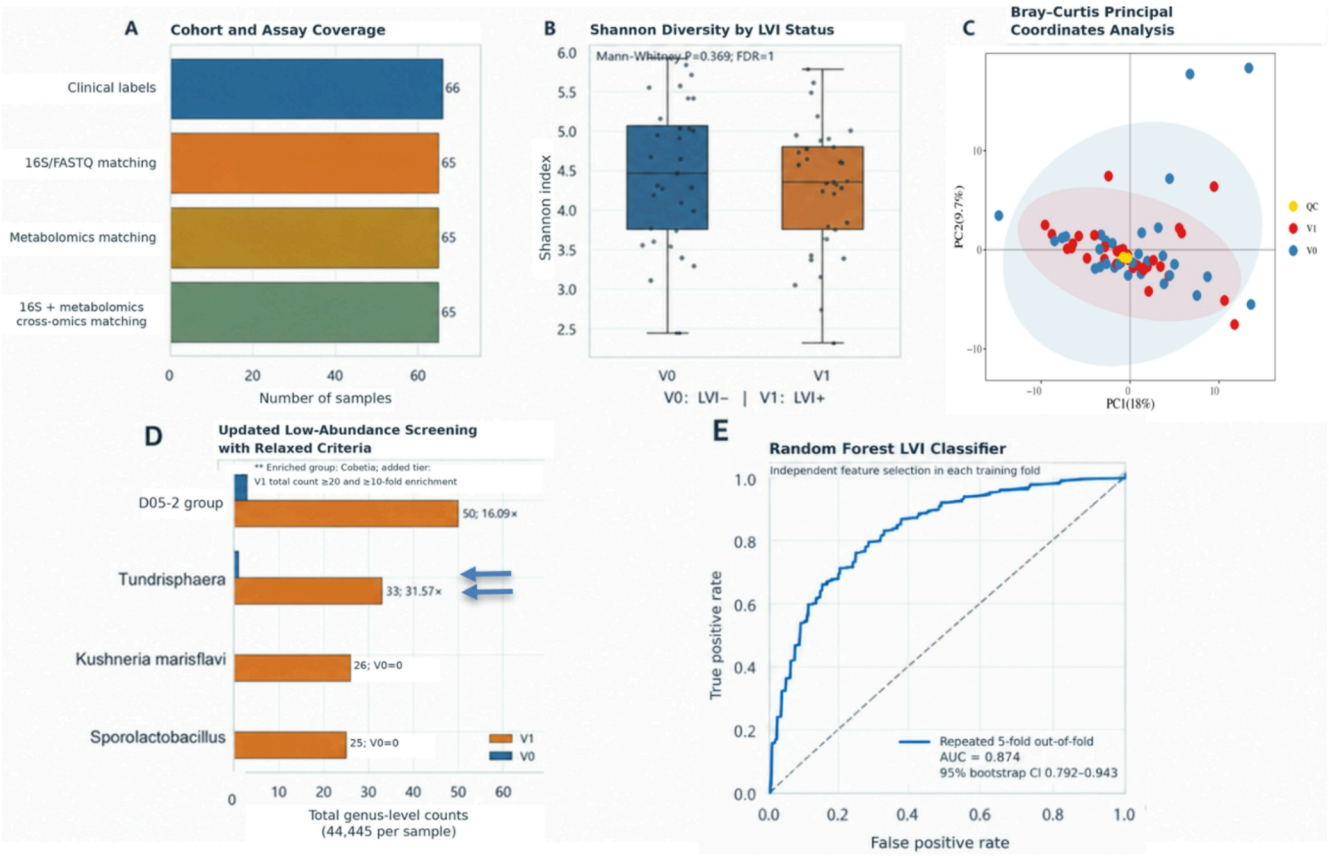


Figure 1

Gastric Tissue Microbiome Characteristics Stratified by LVI

Figure legend: A, discovery cohort and multi-omics data coverage; B, Shannon diversity in V0 and V1; C, Bray–Curtis community structure; D, cumulative genus-table counts for candidate bacteria in the strict tier and relaxed additional tier; E, out-of-fold ROC of the random forest with information leakage controlled.

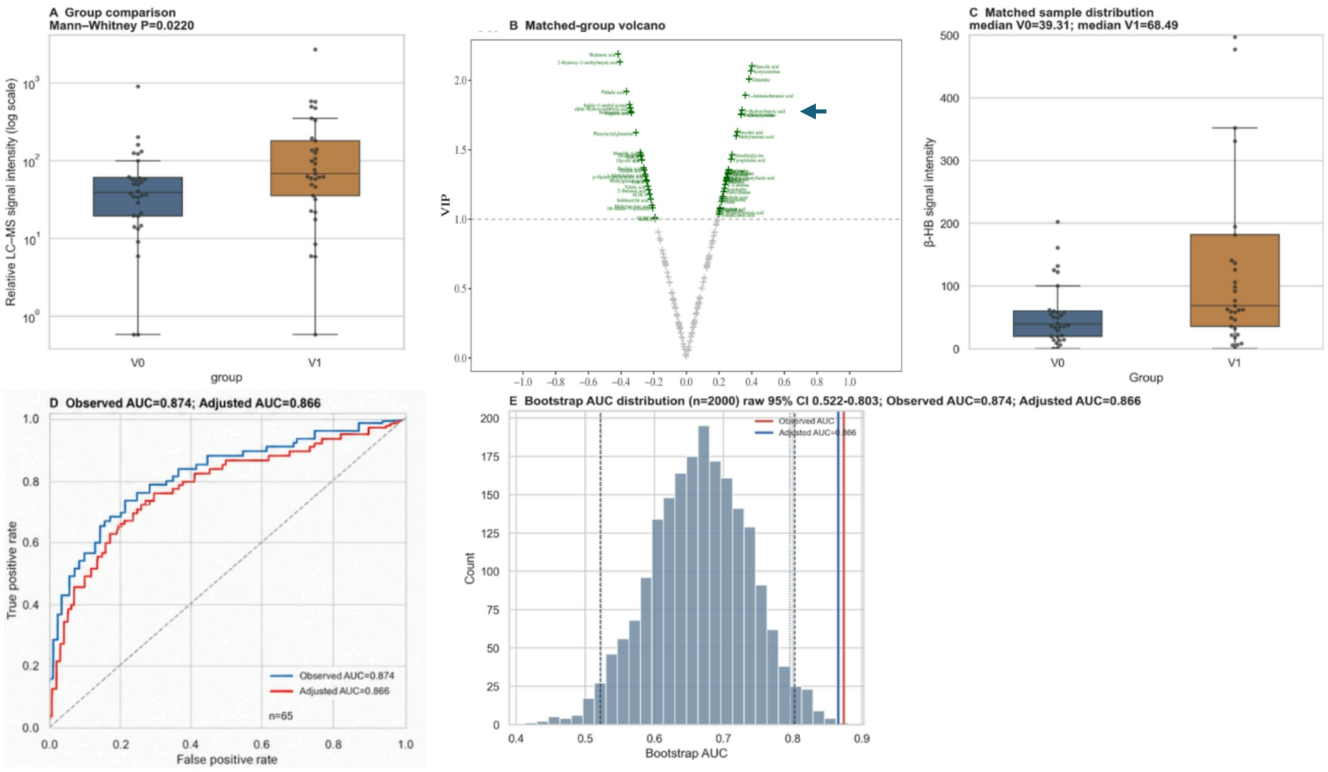


Figure 2

β-HB in the Metabolomic Analysis of the Discovery Cohort

Figure legend: A, box plot comparing LC-MS signal intensities between the V0 and V1 groups; B, volcano plot for the 3-HB matched groups; C, box plot of the grouped sample distribution of β-HB; D, AUC positive-rate correlation plot; F is a histogram of univariate discrimination and age- and sex-adjusted association estimates.

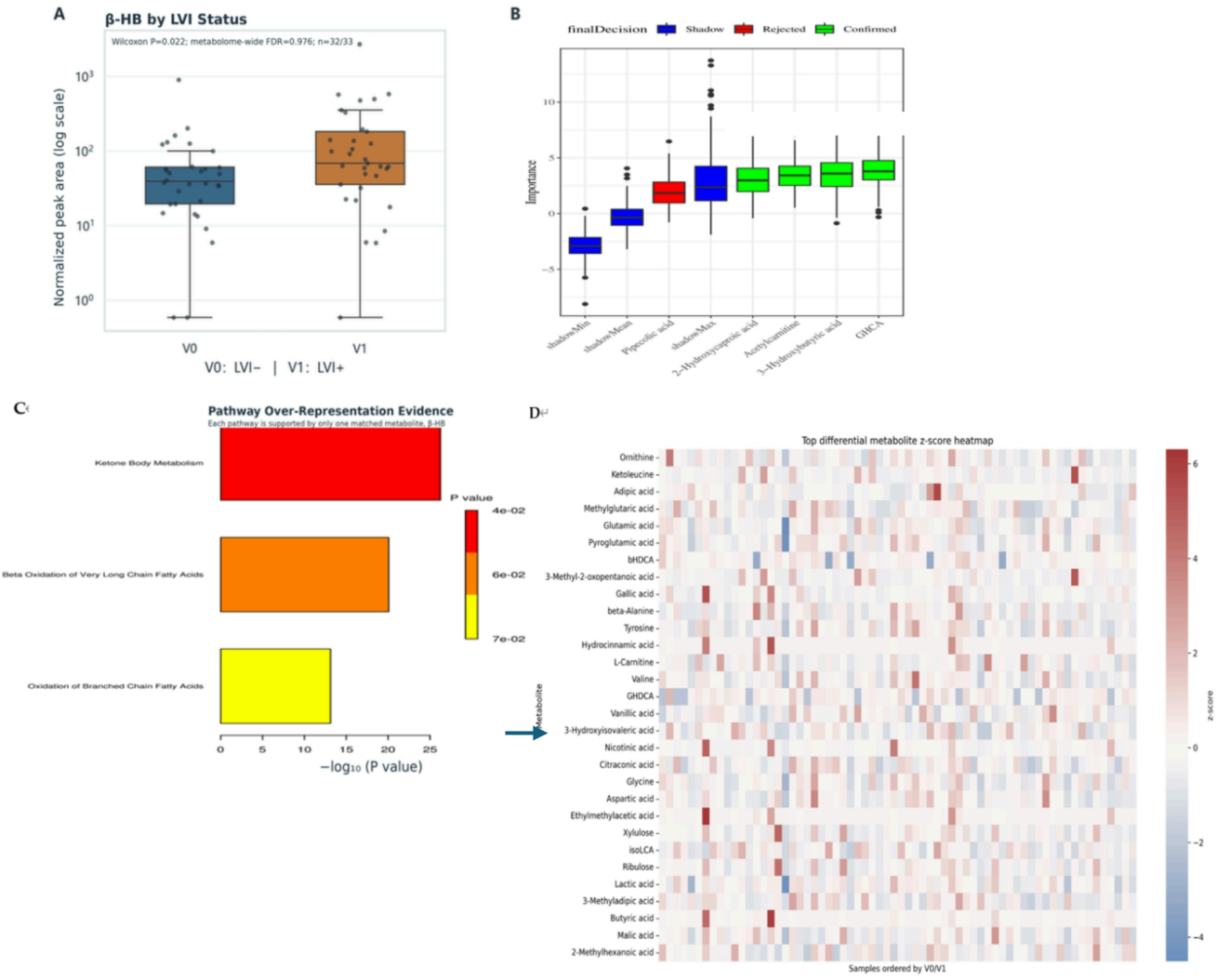


Figure 3

Matched Cross-Omics Bridging and Single-Hit Pathway Audit

Figure legend: A, β-HB grouped by LVI status; B, Priority between the metabolism candidates; C, raw P values for ketone body synthesis and degradation and butanoate metabolism；D,Top differential metabolite z-score heatmap.

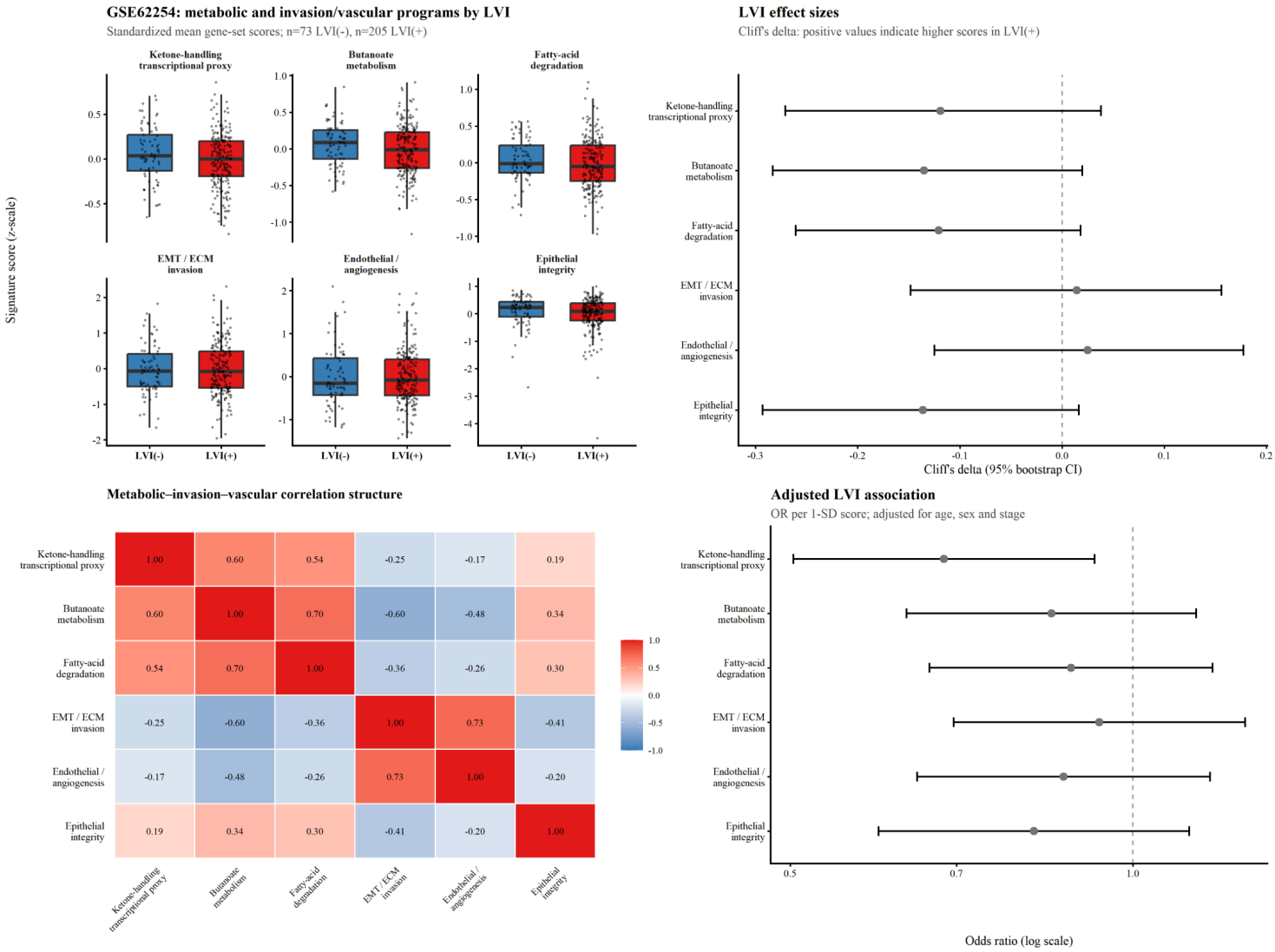


Figure 4

RNA Proxy Validation of Actual LVI in GSE62254

Figure legend: The figure shows the six prespecified RNA programs in samples with actual LVI labels, including program scores in the LVI-negative and LVI-positive groups, the correlation structure between programs, unadjusted associations and LVI association estimates adjusted for clinical variables.

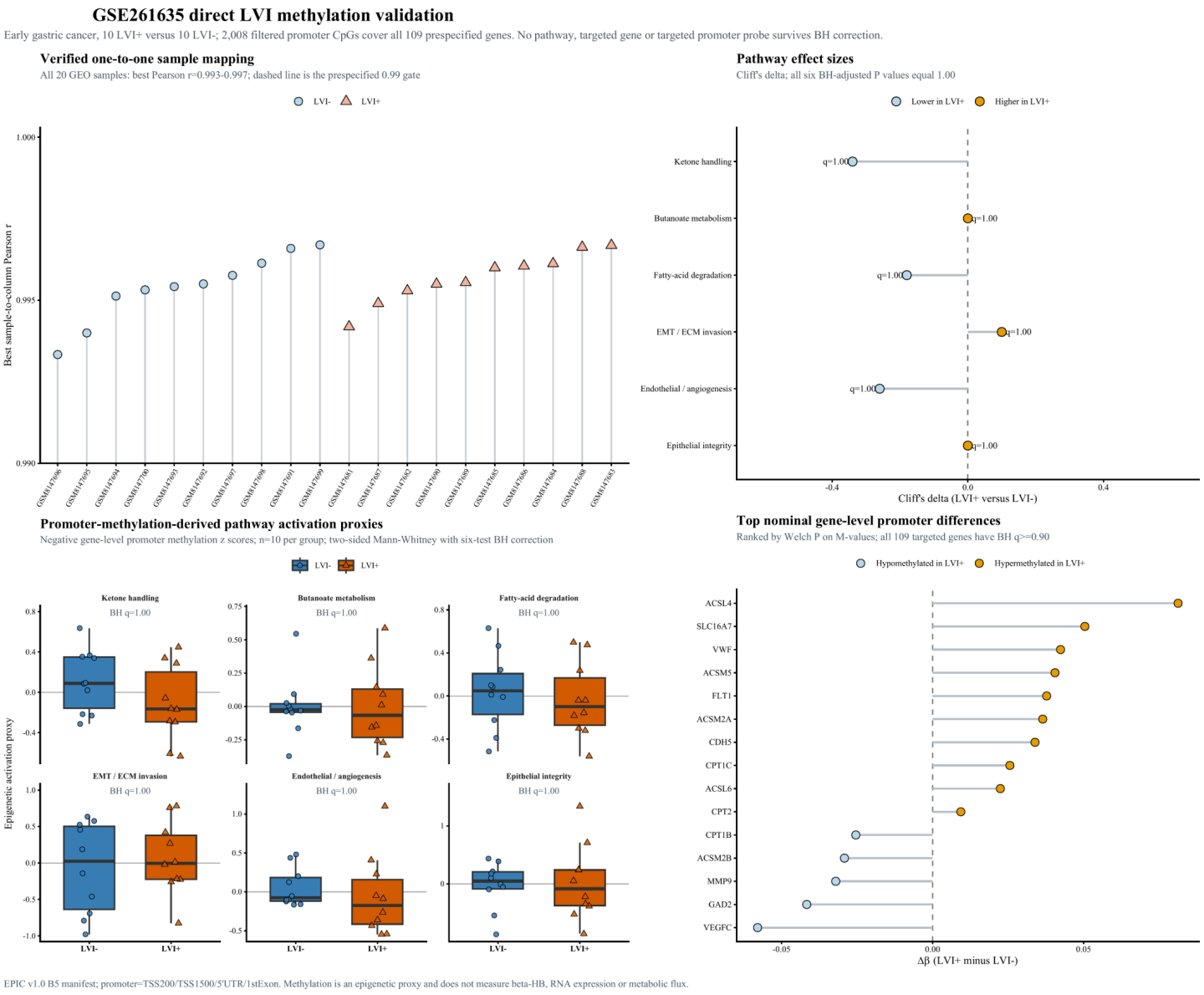


Figure 5

EPIC Methylation Validation of Direct LVI in GSE261635

Figure legend: A, verification of one-to-one mappings between the 20 samples and data matrix columns; B, effect estimates of the six promoter-methylation program proxies; C, sample distributions for each program; D, ranking of nominal effects for target genes and promoter probes.

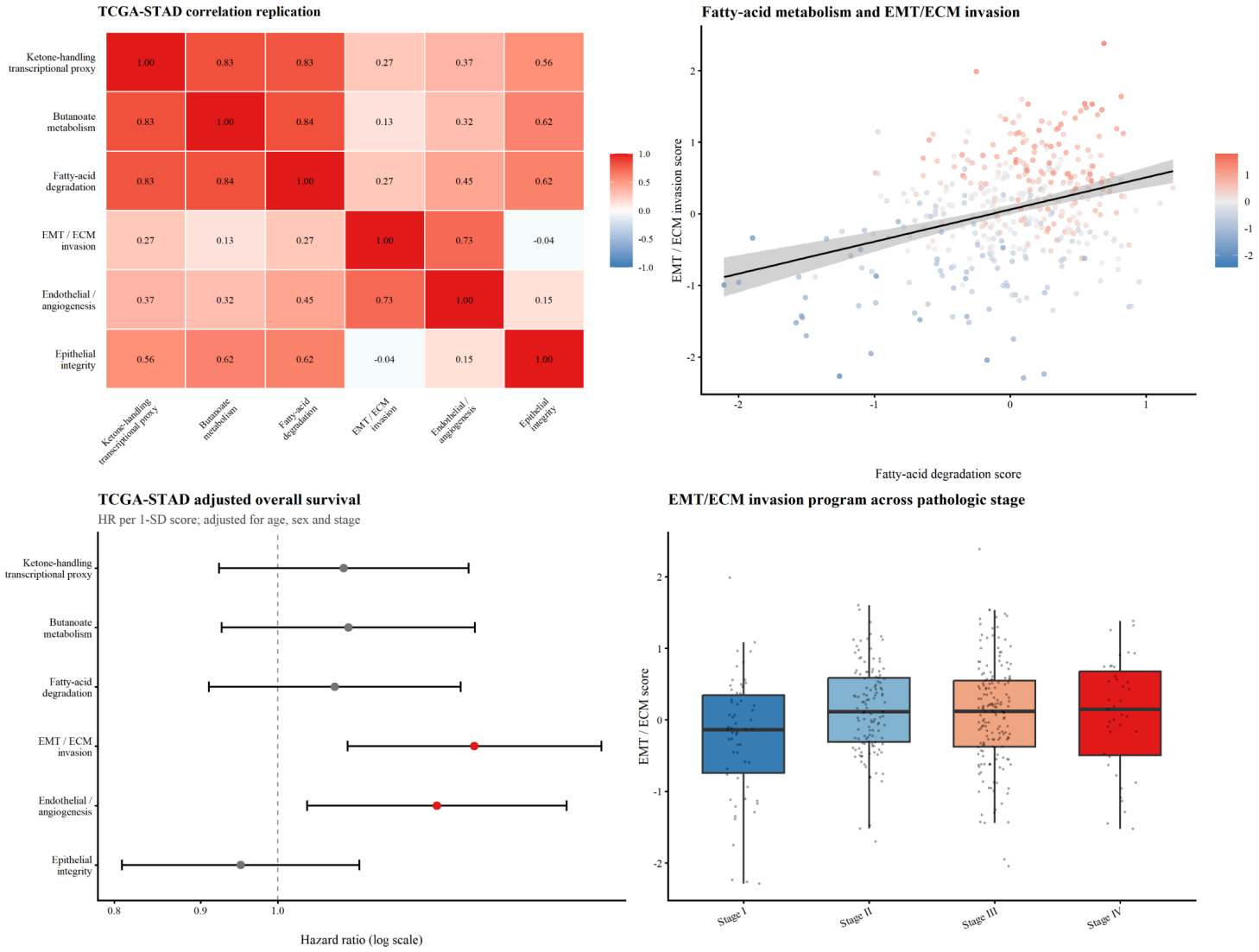


Figure 6

Metabolic Co-regulation and Survival Context in TCGA-STAD

Figure legend: The figure includes the six-program correlation matrix, the sample-level relationship between fatty acid degradation and EMT/ECM invasion, covariate-adjusted Cox regression results, and the distribution of the EMT/ECM program across pathological stages.

Supplementary Files

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

- externalpublicdatabaseanalysis.r
- SupplementaryFigures.docx