Non-Intestinal Microbial Signatures in Stool as Predictors of Cancer Immunotherapy Outcome

[1,2]Vera A. Kanaeva, [1#]Evgenii I. Olekhnovich

Authors affiliations:
[1]Lopukhin Federal Research and Clinical Center of Physical-Chemical Medicine of Federal Medical Biological Agency, Moscow, Russian Federation;
[2]Moscow Institute of Physics and Technology, Dolgoprudny, Russian Federation.

#Address correspondence to Evgenii Olekhnovich jeniaole01@gmail.com.

Abstract
Despite advances in immunotherapy, clinical responses remain heterogeneous, underscoring the need to elucidate the gut microbiome's role in modulating therapeutic efficacy. We conducted a large-scale genome-resolved metagenomic analysis of 624 stool samples from 11 cohorts spanning five cancer types from three continents to identify conserved microbial biomarkers of immunotherapy response. Responders (R; n = 362) exhibited enrichment of *Blautia* species, linked to acetate production from CO_2 via the Wood-Ljungdahl (WL) pathway, glycoside hydrolases (GH), and amino acid biosynthesis. Non-responders (NR; n = 262) showed increased oral (e.g. *Veillonella parvula*, *Dialister invisus*) and food-derived taxa included opportunistic species e.g. *Citrobacter freundii*, *Klebsiella* spp., alongside depletion of gut-resident microbes and enriched lipopolysaccharide biosynthesis and flagellar assembly pathways. Interestingly, part of NR-associated microbes originated from non-intestinal niches (oral/food), which may correlate with reduced gut microbial load in the large intestine. These findings redefine dietary interventions as dual-edged: while carbon sources such as fiber processed by GH or CO_2 processed by WL pathway supports beneficial taxa and immune system, pathogen-free food sourcing is critical to avoid introducing opportunistic microbes. Our work establishes the resident-to-transient microbial ratio as a novel predictive biomarker and positions *Blautia* as a keystone genus for next-generation probiotics targeting carbon metabolism-immune crosstalk. By integrating taxonomic, functional, and ecological axes, this study provides a roadmap for microbiome-guided precision nutrition to optimize immunotherapy outcomes. While immunotherapy non-responders exhibit gut microbiota enriched with food/oral commensals, *Faecalibacterium prausnitzii* - shows context-dependent associations, with strain-level variations potentially linked to A blood group antigen N-acetylgalactosamine utilization, highlighting its role as a conditional biomarker and the prognostic value of metagenomics in personalized therapy.

Key words: cancer, immunotherapy, gastrointestinal microbiome, metagenomics, Blautia wexlerae

Introduction
Despite the increasing use of immunotherapy in treating advanced malignancies, clinical outcomes remain heterogeneous, with many patients exhibiting intrinsic resistance or eventual disease recurrence following immune checkpoint inhibitor (ICI) therapy [Robert et al., 2015]. While advancements in targeted therapies and immunotherapy have improved

survival outcomes [Switzer et al., 2022], ICIs targeting cytotoxic T-lymphocyte-associated antigen 4 (CTLA-4) and programmed cell death protein 1 (PD-1) demonstrate limited efficacy in approximately 50% of patients with metastatic melanoma [Larkin et al., 2015] and carry the risk of immune-related adverse events [Robert et al., 2019]. This highlights the critical need for predictive biomarkers to optimize treatment selection and personalize therapeutic strategies.

Accumulating evidence indicates a strong correlation between gut microbial composition and responsiveness to cancer immunotherapy [Peng et al., 2020]. Indeed, the gut microbiome is increasingly recognized as a critical modulator of anti-tumor immunity and, consequently, response to cancer immunotherapy. Initial investigations, utilizing both preclinical models [Sivan et al., 2015] and clinical cohorts, particularly those with melanoma [Routy et al., 2018], have demonstrated a compelling link between gut microbial composition and immunotherapy efficacy. This link has been further supported by fecal microbiota transplantation studies in both gnotobiotic mice [Gopalakrishnan et al., 2018] and human patients [Baruch et al., 2021]. While much of the early research focused on melanoma [Tsakmaklis et al., 2023], current efforts are shifting towards identifying conserved microbial biomarkers and mechanisms applicable across diverse cancers [Gunjur et al., 2024; Lin et al., 2025], predicated on the hypothesis that fundamental interactions between the intestinal microbes and the immune system are broadly relevant to immunotherapy response.

Despite numerous published studies, the biological mechanisms by which gut microbes regulate the immune system in the context of melanoma immunotherapy remain incompletely understood. A recent meta-analysis identified robust stool metagenomic markers associated with successful melanoma immunotherapy [Olekhnovich et al., 2023]. Expanding these findings, we employed genome-resolved metagenomics to gain a deeper understanding of the gut microbiota's role in modulating immunotherapy response in malignant tumors. This analysis revealed that reduced microbial capacity for key metabolic processes – including glycoside hydrolase activity, amino acid metabolism, and acetate synthesis – specifically through the Wood-Ljungdahl pathway – in the gut microbiome was associated with decreased immunotherapy efficacy in patients with metastatic melanoma [Zakharevich et al., 2024].

Building upon our initial findings, we expanded our application of genome-resolved metagenomics, analyzing 624 stool metagenomes from 11 published studies encompassing over five cancer types and representing diverse geographical regions across three continents. This analysis identified robust markers capable of differentiating patients based on immunotherapy response. These markers were validated using statistical analysis and machine learning techniques via logistic regression with Synthetic Minority Over-sampling Technique (SMOTE)-based class balancing focused on logarithmic ratios of identified features. Consistent with prior research [Zakharevich et al., 2024], our findings corroborate that disruptions in critical metabolic pathways – including amino acid metabolism and vitamin synthesis – are linked to systemic metabolic dysregulation observed in immunotherapy non-responders. Notably, the analysis revealed non-intestinal traits such as oral- and food-derived bacteria within the marker list predominantly associated with negative treatment outcomes, potentially reflecting the previously observed depletion of gut microbial load links with oral microbiota enrichment [Liao et al., 2024]. These findings underscore the importance of considering the metabolic landscape of the gut microbiome – and the surprising influence of non-intestinal microbial sources – when predicting and optimizing immunotherapy response. These results open up new possibilities for the development of novel therapeutic strategies targeting microbiome-immune interactions.

Materials and methods

Data acquisition and preprocessing

All 624 stool metagenomic samples from 11 published studies were downloaded from the NCBI/EBI databases using Kingfisher v0.4.1 [Woodcroft, 2022]. Data quality was assessed with FastQC v0.12.1 [Andrews et al., 2023], and raw reads were processed with fastp v0.23.4 to remove low-quality sequences [Chen et al., 2018]. Human-derived sequences were filtered out using HISAT2 v2.2 [Kim et al., 2019] with the GRCh37 human genome (Release 47) as a reference [Schneider et al, 2017].

Assembly and binning

Metagenomic assembly was performed using MEGAHIT v1.2.9 [Li et al., 2015], retaining contigs longer than 1000 bp. Filtered contigs were aligned to metagenomic reads using HISAT2 v2.2 [Kim et al., 2019]. Binning was conducted in two stages: (1) initial binning with MetaBAT2 v2.12.1 [Kang et al., 2019], MaxBin2 v2.2.7 [Wu et al., 2016], and Semibin2 v2.1.0 [Pan et al., 2023], followed by (2) refinement with DAS Tool v1.1.7 to improve bin quality [Sieber et al., 2018]. Bins were quality-checked with CheckM v1.2.3 [Parks et al., 2015] and dereplicated at 98% nucleotide identity using dRep v3.4.5 [Olm et al., 2017] to generate operational genomic units (OGUs) by analogy with operational taxonomic units (OTUs). Zhu et al. (2022) first introduced this term, but our approach to representing OGUs offers a key advantage: OGUs are generated directly from the analyzed data, rather than through a mapping process. This more faithfully corresponds to the core concept of the "operational unit". For the first time this approach was applied in the work of Zakharevich et. al. (2024).

Taxonomic profiling and analysis

Taxonomic annotation of bins was performed using GTDB-Tk v2.1.1 [Chaumeil et al., 2022] with the GTDB r207 database [Parks et al., 2022]. Additional data processing utilized Samtools v1.17 [Li et al., 2009], BEDTools v2.31.0 [Quinlan et al., 2010], and BBMap v39.06 [Bushnell et al., 2014]. Finally, OGU abundance profiles were generated by mapping reads to the OGU catalog using HISAT2 v2.2, followed by processing with InStrain v1.9.0 [Olm et al., 2021].

Marker OGU discovery

In this study, we intentionally deviated from conventional hypothesis-driven frameworks targeting individual microbial taxa to circumvent limitations intrinsic to metagenomic data analysis. Instead, we implemented a differential ranking approach to prioritize features based on their association strength with clinical response to immunotherapy, coupled with log-ratio transformations to quantify ecosystem-level restructuring of the gut microbiota [Morton et al., 2019]. This methodological decision was driven by the compositional nature of microbiome data, where log-ratios mitigate feature interdependencies while preserving ecological interpretability, and their inherent robustness to technical biases such as uneven Gram-negative/positive bias during DNA extraction or sequencing depth. This framework aligns with contemporary paradigms in microbiome research, where holistic community dynamics supersede isolated taxonomic signatures [Gloor et al., 2017].

The identification of OGUs associated with immunotherapy outcomes followed an established analytical framework from our previous works [Olekhnovich et al., 2023;

Zakharevich et al., 2024]. Differential rankings were first performed using Songbird v1.0.3 [Morton et al., 2019] to identify OGUs showing relative abundance variations between experimental groups, applying a conservative absolute differential value threshold of > 0.3. For candidate OGUs that met this criterion, we subsequently calculated log-ratio abundances according to the principle described in the works of Morton et al. (2019) and Fedarko et al. (2020). The statistical significance of the division of patients by response based on log ratios was determined through Wilcoxon rank-sum tests implemented in the R statistical environment v4.4.2. The biomarker selection process incorporated stringent cross-validation criteria to ensure robust identification of microbial signatures. OGUs demonstrating consistent positive associations with therapeutic response across multiple datasets were retained as potential beneficial biomarkers, while any evidence of negative association with treatment outcome in any dataset resulted in automatic exclusion regardless of other positive associations. This approach enabled simultaneous identification of two clinically meaningful biomarker categories: microbial taxa positively correlated with successful immunotherapy outcomes and those associated with adverse therapeutic responses. The methodology emphasizes reproducibility through multi-dataset validation and maintains rigorous standards for biomarker qualification by requiring consistent directional effects across independent cohorts.

Markers testing

Determined R/NR biomarkers were used to calculate log ratios, followed by statistical assessment using the lmer function from the lmerTest package v3.1.3 [Kuznetsova et al., 2017] in the R statistical environment. Post-hoc testing was performed using Tukey's method with the multcomp package v1.4.28 [Bretz et al., 2011]. Model residuals were tested for normality using the qqtest function from qqtest package v1.2.0. Additional testing of marker OGU included investigating the ability of calculated log ratios to separate experimental groups using logistic regression using the method of "leave-one group out" (LOGO) cross validation with SMOTE for class balance using the tidymodels package v1.3.0. Rahman et al. (2023) demonstrated the effectiveness of using log relations and class balanced log regression in machine learning models. Gene-set enrichment analysis (GSEA) implemented in the GSEA function from clusterProfiler package v4.16.0 [Xu et al., 2024] used to investigate statistical relationships of OGU marker sets across taxonomy and other attributes including food, body site or pathogens. GSEA analysis results were visualized using the GseaVis package v0.1.1 [Zhang et al., 2025].

Marker clustering with non-intestine databases

The presence of non-gut microbes in the stool microbiome, such as those frequently detected on the oral means, has been associated with a degradation of the gut microbiota [Liao et al., 2024]. This phenomenon frequently results in adverse outcomes, including various illnesses. The identification of non-gut microbes was conducted using published microbial metagenome-assembled genomes (MAGs) from body sites (oral, skin, vagina) [Pasolli et al., 2019] and food [Carlino et al., 2024]. The body site MAGs were retrieved from http://segatalab.cibio.unitn.it/data/Pasolli_et_al.html. A total of 9,412 genomes were selected based on specific criteria, including those from the "Body Site" category, such as the airways, nasal cavity, oral cavity, skin, and vagina. Genomes categorized in the "Age category" as "school age" (12–19 years), "adult" (19–65 years) and "senior" (65–70 years) were included in the analysis. Genomes from newborns (<1 year) or children (1–12 years) were excluded, as the gut microbiota typically reaches an adult-like phylogenetic

composition by 3 years of age [Yatsunenko et al., 2012]. The selected genomes were then clustered with our reconstructed catalog via dRep v3.4.5 [Olm et al., 2017] with 95% nucleotide identity and --ignoreGenomeQuality parameter. Genomes that passed the similarity cutoff to body site MAGs were labeled from the same source. The food MAGs were obtained from https://zenodo.org/records/13285428, where the cFMD_mags.tar.gz archive contains the folder cFMD_mags_prok with prokaryotic genomes. These 10,112 MAGs were pre-clustered among themselves using dRep v3.4.5 with 95% nucleotide identity and --ignoreGenomeQuality parameter, yielding a total of 983 genomes. The subsequent source identification process was consistent with that employed for body MAGs.

Functional profiling of marker OGU

Functional profiles of OGUs were generated using MetaCerberus v1.4.0 [Figueroa et al., 2024] with annotations from the Carbohydrate-Active EnZymes (CAZy) database [Drula et al., 2022] and the Kyoto Encyclopedia of Genes and Genomes (KEGG) [Kanehisa et al., 2025]. Statistical associations between functional features and experimental groups were assessed using logistic regression. GSEA was applied to evaluate links between gene sets and immunotherapy response, with results visualized using the GseaVis package v0.1.1. Genes were ranked for GSEA based on p-value multiplied by the sign of the odds ratio, where a negative sign indicates association with the NR group and a positive sign indicates association with the R group.

Additional analysis and visualization

For visualization and statistical analysis, we utilized the following R packages: ggplot2 v3.5.2 [Wickham et al., 2016] for graphics creation, ggpubr v0.6.0 [Kassambara, 2023] for publication-ready plots, hrbrthemes v0.8.7 [Rudis, 2020] for enhanced themes, and RColorBrewer v1.1.3 [Neuwirth, 2022] for color palette management. String data manipulation was performed using the stringr package v1.5.1 [Wickham, 2022]. The data analysis report was generated with Quarto v1.6.41 [Allaire et al., 2022], incorporating dynamic tables via DT v0.33 [Xie et al., 2023] and enhanced download functionality using downloadthis [Sidi, 2023].

Results

Data overview

The analysis incorporated 624 stool metagenomes obtained from 11 independent external datasets (see Table S1). Patients were stratified by immunotherapy response into two groups: responders (R group, n = 362; 58.2%) and non-responders (NR group, n = 262; 41.8%). Response assessment followed RECIST 1.1 criteria, with the R group including patients showing complete response (CR), partial response (PR), or stable disease (SD) at 6-month follow-up, while the NR group comprised exclusively progressive disease (PD) cases. The study cohort received various immunotherapy regimens, including anti-PD1, anti-CTLA4, or combination therapies. Cancer type distribution revealed melanoma predominance (n = 456; 72.7%) [Frankel et al. 2017; Gopalakrishnan et al. 2018; Matson et al. 2018; Spencer et al. 2021; Lee et al. 2022; McCulloch et al. 2022; Tsakmaklis et al. 2023], followed by gastrointestinal (GI) cancers (n = 82; 13.1%) [Peng et al. 2020; Heshiki et al. 2020; Gunjur et al. 2024], non-small cell lung cancer (n = 15; 2.4%) [Liu et al. 2022; Heshiki et al. 2020], breast cancer (n = 4; 0.6%) [Heshiki et al. 2020], ovarian cancer (n = 2;

0.3%) [Heshiki et al. 2020], and other malignancies (n = 68; 10.8%) [Gunjur et al. 2024]. The datasets are plotted on a world map as shown in Figure 1.

All samples were collected prior to treatment initiation to evaluate baseline microbiota status. Following quality filtering with fastp, aggregated quality metrics were compiled using MultiQC to summarize the FastQC reports (MultiQC are attached to the report below). Before filtering, approximately ~15 bn reads (MEAN ± SD: 24 ± 15 mln reads per sample) were retained for downstream analysis, with an average of 34 ± 18% of reads per sample removed during quality control (see Table S2 for detailed filtering statistics). A total of 3649 OGUs belonging to 12 bacterial phyla were detected in 624 stool metagenomes. The most numerous of these were Firmicutes (n = 2443), Actinobacteriota (n = 590), Bacteroidota (n = 394), Proteobacteria (n = 140), Desulfobacterota (n = 27) and others (n = 55). OGU abundance profiles of stool samples presented in Table S3.

Marker OGUs discovery

Analysis of marker OGUs identified 637 significantly differentiated features between R and NR experimental groups. Of these, 298 were enriched in R, while 339 were enriched in NR. The most prominent microbial species enriched in R included *Blautia wexlerae* (n = 49), *Faecalibacterium prausnitzii* (n = 27), *Fusicatenibacter saccharivorans* (n = 15), *Gemmiger qucibialis* (n = 14), and *Blautia faecis* (n = 8). Conversely, NR exhibited enrichment of distinct microbial taxa, including *F. prausnitzii* (n = 17), *Veillonella parvula* (n = 7), and *Dialister invisus* (n = 7). Although *F. prausnitzii* was detected in both groups, its significantly higher abundance in responders (27 vs. 17 OGUs) suggests potential strain-level or functional variations contributing to treatment outcome. The robust association of *Blautia* species, particularly *B. wexlerae*, with positive response highlights its potential as a predictive biomarker. Notably, the enrichment of oral bacteria *Veillonella parvula* and *Dialister invisus* in NR suggests a potential link between oral microbiome and immunotherapy efficacy. Detailed descriptions of marker OGUs are provided in Table S4. Songbird coefficients for each OGU marker are presented in Figure S1.

Log ratio values from the metagenomic samples are detailed in Table S5, with their distribution visualized across experimental groups in Figure 2A. Statistical analysis utilizing linear mixed-effects models, accounting for dataset and response variables, revealed significant differences between R and NR groups ($p < 0.0001$). The inclusion of the dataset variable as a random effect to account for technical and biological heterogeneity between studies was statistically significant ($p < 0.0001$), justifying the employed modeling approach. Residual plot demonstrated an approximately symmetric, unimodal pattern with no extreme outliers (see Figure S2). This visual assessment, combined with the substantial effect size (Cohen's d = 0.91; 95% CI [0.74, 1.08]), supports the robustness of the model's inferences to mild non-normality, aligning with the central limit theorem's applicability to large-sample analyses. A predictive classifier, based on log ratios and logistic regression with SMOTE balancing and LOGO cross validation with dataset as grouping variable, demonstrated a receiver operating characteristic (ROC) area under the curve (AUC) of 0.72 ± 0.12 (Figure 2B), with other performance metrics detailed in Table S6 and Figure S2. The enhanced predictive accuracy of our model for identifying NR suggests that therapeutic success likely depends on a multifactorial interplay extending beyond the gut microbiome, potentially involving intrinsic host factors such as tumor biology (e.g., mutational burden, immune microenvironment) or patient-specific characteristics (e.g., genetic polymorphisms, metabolic status).

Marker OGUs characterization

As a summarizing tool gene set enrichment analysis (GSEA) was used which identified patterns of change related to microbial taxonomy (phylum, genus) and origin, distinguishing between microbes from non-intestinal body sites (oral cavity, airways, vagina, skin) and environment (food).

Taxonomy

OGUs with 98% nucleotide identity were grouped by taxonomy or other characteristics to facilitate enrichment analysis within specific groups. Results indicated that the Bacteroidetes phylum was upregulated in the R group and Proteobacteria in the NR group (abs. NES > 2, GSEA adj. $p < 0.01$). Obtained results presented in Figure 3A and Table S7. Bacteroidetes OGU included in core enrichment species presented in Table S8. Notably, enrichment of Proteobacteria OGUs included potential opportunistic species such as *Klebsiella quasipneumoniae*, *Klebsiella michiganensis*, *Enterobacter ludwigii*, *Enterobacter kobei*, *Citrobacter youngae*, *Citrobacter portucalensis*, *Citrobacter freundii*, and *Haemophilus* sp001815355 (see Table S9). At the genus level, *Blautia*, *Bacteroides*, *Fusicatenibacter*, *Gemmiger*, and *Faecalibacterium* were enriched in the R group, while *Veillonella*, *Prevotella*, *Dysosmobacter*, and *Acetatifactor* were enriched in the NR group (see Figure 3B and Table S10; GSEA adj. $p < 0.01$).

Origin

Clustering 3,816 OGUs from this study with 9,412 MAGs derived from different body sites and 10,112 food-derived MAGs revealed that 293 OGUs exhibited complementary to those from non-intestinal sources (see Table S11), and 91 were complementary to food-derived microbes, at 95% nucleotide similarity (Table S12), which corresponds to the species level. Based on the GSEA results, NR-linked OGUs were enriched by food-derived (see Figure 3C and Table S13) and oral (see Figure 3D and Table S14) microbial flora. Core food-derived OGUs comprised *Bifidobacterium animalis*, *Citrobacter freundii*, *Enterobacter kobei*, *Enterococcus faecalis*, *Enterococcus faecium*, *Escherichia coli*, *K. michiganensis*, *K. quasipneumoniae*, *Lactobacillus gasseri*, *Limosilactobacillus reuteri*, and *Megasphaera elsdenii*, showing substantial overlap with core enrichment patterns observed in Proteobacteria (Table S15). Core enrichment patterns within the oral cavity included OGU representatives of Veillonellales and Enterobacterales, specifically *Anaeroglobus micronuciformis*, *Escherichia coli*, *Veillonella parvula*, and *Haemophilus* sp001815355 (Table S16).

Functional analysis of marker OGU

Functional analysis of marker OGUs was performed using the MetaCerberus pipeline, CAZy and KEGG databases. Logistic regression was used to examine statistical relationships between functional groups and R/NR OGU sets. GSEA was then employed to identify links between CAZy/KEGG functional gene groups and R/NR OGU sets, utilizing p-values generated by the logistic regression.

CAZy

Glycoside hydrolase (GH) functional groups were linked to the R OGU set (NES = 1.4, adj. $p = 0.003$). Obtained results presented in Figure 4A and Table S17. The core enrichment set included 30 glycoside hydrolase families, such as GH39 (alpha-L-iduronidase), GH42 (beta-galactosidase), GH43 (beta-xylosidase), GH51 (endoglucanase), and GH73

(lysozyme) (see Table S18). The most common microbiota genera associated with these enriched GH families were *Blautia* (31 genomes), *Fusicatenibacter* (n = 11), *Bacteroides* (n=9), and *Parabacteroides* (n=5) (see Figure 4B).

KEGG

Analysis of enriched KEGG pathways revealed a significant association with the R OGU and NR OGU groups (abs. NES > 1, adj. $p < 0.01$). Obtained results presented in Figure S7 and Table S19. Specifically, the R OGU group exhibited enrichment for eight KEGG pathways, including: ko00340 (Histidine metabolism), ko01210 (2-Oxocarboxylic acid metabolism), ko00290 (Valine, leucine and isoleucine biosynthesis), ko00220 (Arginine biosynthesis), ko01230 (Biosynthesis of amino acids), ko00920 (Sulfur metabolism), ko00040 (Pentose and glucuronate interconversions), and ko01240 (Biosynthesis of cofactors). The NR OGU group, in contrast, showed enrichment for two KEGG pathways: ko00540 (Lipopolysaccharide biosynthesis) and ko02040 (Flagellar assembly). Further analysis revealed distinct bacterial genera associated with these functional pathways (see Figure S8). *Blautia* and *Fusicatenibacter* were primarily linked to the pathways enriched in the R OGU marker group, while *Veillonella*, *Dialister*, and *Enterobacter* were associated with ko00540 (Lipopolysaccharide biosynthesis) in the NR group. Similarly, *Acetifactor*, *Roseburia*, and *Ruminoclostridium* were linked to ko02040 (Flagellar assembly) enrichment in the NR group. These findings suggest potential mechanistic links between specific microbial taxa and the observed functional differences between the R OGU and NR OGU groups.

Our previous metagenomic analysis revealed significant enrichment of Wood-Ljungdahl (WL) pathway-encoding MAGs from *Blautia*, *Oliverpabstia*, *Fusicatenibacter*, and *Choladousia* genera in immunotherapy-responsive melanoma patients [Zakharevich et al., 2024]. In the current study, we identified eight core WL pathway gene clusters (fchA - K01500; cooS, acsA - K00198; acsE - K15023; acsB - K14138; cdhE, acsC - K00197; cdhD, acsD - K00194; rnfC2 - K25008; metV - K25007) that were consistently overrepresented in operational genomic units (OGUs) associated with positive treatment outcomes. Notably, these WL pathway genes showed strong associations with *Blautia* and *Fusicatenibacter* MAG sequences (see Figure S6). Functional divergence emerged between R- and NR-associated *F. prausnitzii* strains (see Table S20). NR-linked strains exhibited marked enrichment in galactose metabolism (ko00052), driven by core genes including galT (K00965), gatY/kbaY (K08302), agaS (K02082), and gatZ/kbaZ (K16371). These strains also overrepresented the N-acetyl-D-galactosamine utilization system (agaF - K02744; agaV - K02745, agaW - K02746, agaE - K02747).

Discussion

Our analysis reveals a distinct microbial signature associated with positive response to immunotherapy, prominently featuring enrichment of *Blautia* species, particularly *Blautia wexlerae* and *Blautia faecis*. This finding reinforces our previous meta-analyses identifying *Blautia* as a significant biomarker for favorable outcomes in melanoma immunotherapy [Olekhnovich et al., 2023; Zakharevich et al, 2024], and aligns with the growing recognition of this genus's importance for human health [Liu et al., 2021]. The beneficial effects of *Blautia* may stem from its capacity to enhance gut barrier function through stimulation of mucus production and short-chain fatty acids (SCFA) e.g. acetate and propionate generation, activating the Ffar2 receptor [Holmberg et al., 2024]. Furthermore, *Blautia*'s unique metabolic capabilities, including acetate production via the Wood-Ljungdahl pathway

and vitamin B_{12} synthesis [Trischler et al., 2022; Zakharevich et al, 2024], may contribute to increased carbon availability and support microbial community function, ultimately bolstering host immunity. Interestingly, reduced *Blautia* abundance and acetate levels have been linked to diminished immune responses and accelerated cancer progression, particularly in contexts of chronic stress and depression [Ye et al., 2023]. These findings suggest that strategies to restore *Blautia* populations and acetate production hold promise as novel therapeutic interventions for enhancing immunotherapy efficacy and improving outcomes in patients.

Beyond *Blautia*, several other taxa – including species within the *Bacteroides* genus, *Fusicatenibacter saccharivorans*, and *Gemmiger qucibialis* – showed a positive correlation with response to immunotherapy, reinforcing our earlier findings identifying these species as biomarkers for favorable outcomes [Zakharevich et al, 2024]. While their specific contributions require further investigation, their co-occurrence with *Blautia* suggests a cooperative metabolic network centered on complex carbohydrate processing and SCFA production. Our data supports this, linking these species to key metabolic functions such as glycoside hydrolase activity, amino acid and vitamin synthesis, and sulfur metabolism. Elevated glycoside hydrolase levels likely reflect dietary fiber intake, facilitating the breakdown of polysaccharides into fermentable monosaccharides that fuel SCFA production – creating a positive feedback loop for gut health and immune modulation [Spencer et al., 2021]. Notably, *Blautia*'s capacity to generate acetate directly from CO_2 via the Wood-Ljungdahl pathway expands the metabolic flexibility of the gut microbiome, allowing carbon fixation independent of dietary inputs. This highlights the dynamic interplay between the microbiome, host metabolism, and environmental carbon sources. While the link between amino acid/vitamin synthesis and immune function is well-established, our findings also reveal a connection to sulfur metabolism. A critical microbial shift in Crohn's disease involves depletion of H_2S-consuming taxa like *Blautia* and *Gemmiger*, which disrupts sulfide metabolism and drives toxic H_2S accumulation [Marcelino et al., 2023]. This gasotransmitter, while physiologically regulating immunity, exhibits paradoxical roles: it exacerbates macrophage-driven inflammation [Cornwell et al., 2023] yet mitigates autoimmune hyperactivation [Pozzi et al., 2022]. Excess H_2S erodes gut barrier integrity, sustains pro-inflammatory cascades, and paradoxically bolsters pathogen resilience by shielding microbes from immune clearance [Toliver-Kinsky et al., 2019]. Such dysbiosis-linked H_2S dysregulation underscores its dual nature as both a metabolic byproduct and a central orchestrator of inflammatory tissue damage in Crohn's pathogenesis.

In contrast to the microbial signature associated with positive response, our analysis revealed a distinct composition in non-responding patients. Notably, the enrichment of taxa typically found in the oral microbiome – *Veillonella parvula* and *Dialister invisus* – was consistently observed in stool samples from those failing to respond to immunotherapy. We attribute this finding to a decrease in total microbial load in the large intestine [Liao et al., 2024], which correlates well with a decrease in functional capacity. Given the compositional nature of metagenomic data [Gloor et al., 2017], observed changes represent relative, not absolute, shifts in microbial abundance. These relative increases or decreases can arise from various scenarios, including an overall decline in the gut microbiota accompanied by a disproportionate loss in specific taxa, or an overall increase where one taxon expands more readily [Morton et al., 2019]. We hypothesize that a decline in the gut microbiota, coupled with the constant transit of oral microbes through the digestive system, creates a scenario where oral taxa serve as a relatively stable "reference" against which gut microbial representation is normalized. While we acknowledge the possibility that oral microbes may

also actively colonize vacated niches. Interestingly, alongside increased oral taxa, we observed a cohort of bacteria frequently sourced from food, likely facilitated by a decrease in overall gut microbial load and the consistent supply of these bacteria through diet. While this included beneficial species commonly found in fermented foods and dairy products – such as *Bifidobacterium animalis*, *Lactobacillus gasseri*, and *Limosilactobacillus reuteri* – it also encompassed opportunistic pathogens like *Citrobacter freundii*, *Enterobacter kobei*, *Enterococcus faecalis*, *Klebsiella michiganensis*, and *Klebsiella quasipneumoniae*. This dual presence highlights the complex interplay between diet and microbiome composition. While fermented foods and probiotics can be beneficial, our findings underscore the critical importance of food sources enhanced safety control for feeding patients. The introduction of opportunistic pathogens through ingested foods may contribute to dysbiosis and reduced immunotherapy efficacy. Therefore, both dietary composition and sourcing deserve careful consideration. This reveals a “double-edged sword” effect of dietary interventions: promoting a healthy microbiome through fiber intake is crucial, but equally important is minimizing the introduction of potentially pathogenic bacteria. Future studies should investigate the mechanisms underlying oral and food-borne microbial translocation and their impact on immunotherapy response. Targeted interventions focused on restoring gut barrier function, reducing microbial load, and ensuring pathogen-free food sources may offer promising strategies for enhancing immunotherapy efficacy in non-responding patients.

Interestingly, these observations align with the Anna Karenina principle applied to metagenomics [Zaneveld et al., 2017]. Healthy gut microbiomes are characterized by a conserved set of evolutionarily adapted bacteria – exhibiting a characteristic “sameness” (“all happy families are happy in the same way”) – whereas disruption of this core microbiome allows for increased detection of transient microbes from the oral cavity or environment, which constantly traverse the gut during ingestion. Consequently, greater degradation of the core gut microbiota results in a more variable microbial profile - “every unhappy family is unhappy in its own way”.

Distinct gut microbiota profiles in immunotherapy responders vs. non-responders underscore complex host-microbiome interplay in treatment outcomes. While *V. parvula* and oral commensals dominate non-responders, *F. prausnitzii* – a widely recognized beneficial bacterium linked to butyrate production and anti-inflammatory effects [Gao et al., 2023; Bredon et al., 2024, 2025] – shows inconsistent associations with response in our study. Our data, aligning with prior work [Olekhnovich et al., 2023; Zakharevich et al., 2024; Gunjur et al., 2024], reveal variable *F. prausnitzii* abundance patterns across both groups. These findings position *F. prausnitzii* as a “conditional biomarker,” effective only within specific ecological contexts. For instance, there is evidence that mucin polysaccharide structures may incorporate blood group antigen modifications, particularly N-acetylgalactosamine - A blood antigen [Jajosky et al., 2023]. Additionally, studies suggest a correlation between immunotherapy outcomes and blood group type, with patients bearing blood group A exhibiting poorer survival rates compared to those with blood group O [Chen et al., 2022]. Consequently, strain-level variations in *F. prausnitzii* linked to the N-acetylgalactosamine utilization system could underlie individual patient characteristics not directly linked to gut microbiota composition and to demonstrate the subtle ecological adaptation of this species to human biology. This finding underscores the potential prognostic value of stool metagenomic data in personalized medicine.

Conclusions

Our study elucidates a microbiota-dependent mechanism shaping immunotherapy efficacy through carbon source competition. We propose that dietary modulation of pathogen-free nutrient availability could optimize microbial metabolic networks to enhance therapeutic responses in cancer patients. Specifically, we identify *Blautia* species as key candidates for next-generation probiotics, leveraging their Wood-Ljungdahl pathway to convert CO_2 into immunomodulatory acetate, thereby potentiating anti-tumor immunity. Furthermore, our findings establish a novel framework for predicting immunotherapy outcomes using stool microbiome profiling. The resident-to-transient microbial ratio emerges as a robust normalization strategy, having potential to resolve inter-individual variability in gut microbial load while quantifying dietary impacts on therapeutic microenvironments. This dual approach - combining targeted probiotic development with machine learning-driven microbiota diagnostics - unlocks potential for personalized dietary-microbiome interventions. By functionally linking microbial carbon metabolism to host immune activation, our work provides actionable biomarkers for clinical translation and redefines nutritional guidelines in immuno-oncology.

Declarations

Ethics approval and consent to participate

For the analysis, sample identifiers from the SRA archive and publicly accessible metadata from published studies were employed.

Consent for publication

The data used in this study were sourced from SRA/NCBI (accession identifiers presented in Table S1, column 'sampleid'). All original studies obtained informed consent from participants for the publication of anonymized data. Our analysis did not require additional ethical approval, as only de-identified datasets were utilized from open sources.

Data availability

In this study, we used open access data from the NCBI/EBI Sequence Read Archives, identified by the following BioProjects accession numbers: PRJNA397906, PRJEB22893, PRJNA399742, PRJNA770295, PRJEB43119, PRJNA762360, PRJNA1011235, PRJNA615114, PRJNA866654, PRJNA494824, PRJEB49516. OGUs catalog assembly pipeline was described in https://github.com/JeniaOle13/Cancer_MAGs. All initial MAGs sequences were deposited in NCBI under accession PRJNA1196825. Source code and Quarto report available at https://github.com/JeniaOle13/cancer-biomarkers.

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Fundings

Financial support for this study was kindly provided by the Russian Science Foundation under Grant No. 22-75-10029, which can be accessed at https://rscf.ru/project/22-75-10029/.

Authors' contributions

VAK conducted computational analyses, including metagenomic data processing and microbial feature identification. Taxonomic profiling and cross-database comparisons were

performed to characterize microbial origins. VAK contributed to manuscript editing. EIO designed the study, secured funding, and supervised all research stages. Statistical modeling and biomarker validation were implemented, followed by biological interpretation of immunotherapy-related microbial patterns. EIO wrote and revised the manuscript, approving the final version.

References

1) Andrews S. FastQC: A quality control tool for high throughput sequence data. Version 0.12.1, 2023.

2) Allaire, J. J., et al. Quarto. Version 1.3, 2022.

3) Baruch, Erez N., et al. "Fecal microbiota transplant promotes response in immunotherapy-refractory melanoma patients." *Science* 371.6529 (2021): 602-609. DOI: 10.1126/science.abb5920

4) Bolyen, Evan, et al. "Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2." *Nature biotechnology* 37.8 (2019): 852-857. DOI: 10.1038/s41587-019-0209-9

5) Bretz, F., Hothorn, T., & Westfall, P. (2011). Multiple comparisons using R. CRC Press.

6) Bredon, Marius, et al. "Faecalibacterium prausnitzii strain EXL01 boosts efficacy of immune checkpoint inhibitors." *Oncoimmunology* 13.1 (2024): 2374954. DOI: 10.1080/2162402X.2024.2374954

7) Bredon, Marius, et al. "Faecalibacterium prausnitzii is associated with clinical response to immune checkpoint inhibitors in patients with advanced gastric adenocarcinoma: Results of microbiota analysis of PRODIGE 59-FFCD 1707-DURIGAST trial." *Gastroenterology* 168.3 (2025): 601-603. DOI: 10.1053/j.gastro.2024.10.020

8) Bushnell, Brian. "BBMap: a fast, accurate, splice-aware aligner." (2014).

9) Carlino, Niccolò, et al. "Unexplored microbial diversity from 2,500 food metagenomes and links with the human microbiome." *Cell* 187.20 (2024): 5775-5795. DOI: 10.1016/j.cell.2024.07.039

10) Chaumeil, Pierre-Alain, et al. "GTDB-Tk v2: memory friendly classification with the genome taxonomy database." *Bioinformatics* 38.23 (2022): 5315-5316. DOI: 10.1093/bioinformatics/btac672

11) Chen, Ruifeng, et al. "Patient Blood Type Is Associated With Response to Immune Checkpoint Blockade in Metastatic Cancer." *The Oncologist* 27.9 (2022): e739-e747. DOI: 10.1093/oncolo/oyac128

12) Chen, Shifu, et al. "fastp: an ultra-fast all-in-one FASTQ preprocessor." *Bioinformatics* 34.17 (2018): i884-i890. DOI: 10.1093/bioinformatics/bty560

13) Cornwell, Alex, and Alireza Badiei. "From gasotransmitter to immunomodulator: the emerging role of hydrogen sulfide in macrophage biology." *Antioxidants* 12.4 (2023): 935. DOI: 10.3390/antiox12040935

14) Drula, Elodie, et al. "The carbohydrate-active enzyme database: functions and literature." *Nucleic acids research* 50.D1 (2022): D571-D577. DOI: 10.1093/nar/gkab1045

15) Fedarko, Marcus W., et al. "Visualizing ‘omic feature rankings and log-ratios using Qurro." *NAR genomics and bioinformatics* 2.2 (2020): lqaa023. DOI: 10.1093/nargab/lqaa023

16) Figueroa III, Jose L., et al. "MetaCerberus: distributed highly parallelized HMM-based processing for robust functional annotation across the tree of life." *Bioinformatics* 40.3 (2024): btae119. DOI: 10.1093/bioinformatics/btae119

17) Frankel, Arthur E., et al. "Metagenomic shotgun sequencing and unbiased metabolomic profiling identify specific human gut microbiota and metabolites associated with immune checkpoint therapy efficacy in melanoma patients." *Neoplasia* 19.10 (2017): 848-855. DOI: 10.1016/j.neo.2017.08.004
18) Gao, Yaqi, et al. "Faecalibacterium prausnitzii abrogates intestinal toxicity and promotes tumor immunity to increase the efficacy of dual CTLA4 and PD-1 checkpoint blockade." *Cancer Research* 83.22 (2023): 3710-3725. DOI: 10.1158/0008-5472.CAN-23-0605
19) Gloor, Gregory B., et al. "Microbiome datasets are compositional: and this is not optional." *Frontiers in microbiology* 8 (2017): 2224. DOI: 10.3389/fmicb.2017.02224
20) Gopalakrishnan, Vancheswaran, et al. "Gut microbiome modulates response to anti–PD-1 immunotherapy in melanoma patients." *Science* 359.6371 (2018): 97-103. DOI: 10.1126/science.aan4236
21) Gunjur, Ashray, et al. "A gut microbial signature for combination immune checkpoint blockade across cancer types." *Nature medicine* 30.3 (2024): 797-809. DOI: 10.1038/s41591-024-02823-z
22) Holmberg, Sandra M., et al. "The gut commensal Blautia maintains colonic mucus function under low-fiber consumption through secretion of short-chain fatty acids." *Nature Communications* 15.1 (2024): 3502. DOI: 10.1038/s41467-024-47594-w
23) Heshiki, Yoshitaro, et al. "Predictable modulation of cancer treatment outcomes by the gut microbiota." *Microbiome* 8 (2020): 1-14. DOI: 10.1186/s40168-020-00811-2
24) Jajosky, Ryan Philip, et al. "ABO blood group antigens and differential glycan expression: perspective on the evolution of common human enzyme deficiencies." *iScience* 26.1 (2023). DOI: 10.1016/j.isci.2022.105798
25) Kanehisa, Minoru, et al. "KEGG: biological systems database as a model of the real world." *Nucleic Acids Research* 53.D1 (2025): D672-D677. DOI: 10.1093/nar/gkae909
26) Kang, Dongwan D., et al. "MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies." *PeerJ* 7 (2019): e7359. DOI: 10.7717/peerj.7359
27) Kassambara A (2023). ggpubr: ‘ggplot2’ Based Publication Ready Plots. R package version 0.6.0
28) Kim, Daehwan, et al. "Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype." *Nature biotechnology* 37.8 (2019): 907-915. DOI: 10.1038/s41587-019-0201-4
29) Kuznetsova, Alexandra, Per B. Brockhoff, and Rune HB Christensen. "lmerTest package: tests in linear mixed effects models." *Journal of statistical software* 82 (2017): 1-26.
30) Larkin, James, et al. "Combined nivolumab and ipilimumab or monotherapy in untreated melanoma." *New England journal of medicine* 373.1 (2015): 23-34. DOI: 10.1056/NEJMoa1504030
31) Lee, Karla A., et al. "Cross-cohort gut microbiome associations with immune checkpoint inhibitor response in advanced melanoma." *Nature Medicine* 28.3 (2022): 535-544. DOI: 10.1038/s41591-022-01695-5
32) Li, Dinghua, et al. "MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph." *Bioinformatics* 31.10 (2015): 1674-1676. DOI: 10.1093/bioinformatics/btv033
33) Li, Heng, et al. "The sequence alignment/map format and SAMtools." *Bioinformatics* 25.16 (2009): 2078-2079. DOI: 10.1093/bioinformatics/btp352

34) Liao, C., et al. "Oral bacteria relative abundance in faeces increases due to gut microbiota depletion and is linked with patient outcomes." *Nature Microbiology* 9 (2024): 1555–1565. DOI: 10.1038/s41564-024-01680-3
35) Lin, Yufeng, et al. "Effects of gut microbiota on immune checkpoint inhibitors in multi-cancer and as microbial biomarkers for predicting therapeutic response." *Med* 6.3 (2025). DOI: 10.1016/j.medj.2024.10.007
36) Liu, Ben, et al. "Exploring gut microbiome in predicting the efficacy of immunotherapy in non-small cell lung cancer." *Cancers* 14.21 (2022): 5401. DOI: 10.3390/cancers14215401
37) Liu, Xuemei, et al. "Blautia – a new functional genus with potential probiotic properties?" *Gut microbes* 13.1 (2021): 1875796. DOI: 10.1080/19490976.2021.1875796
38) Neuwirth, Erich. RColorBrewer: ColorBrewer Palettes. R package version 1.1-3, 2022.
39) Marcelino, Vanessa R., et al. "Disease-specific loss of microbial cross-feeding interactions in the human gut." *Nature Communications* 14.1 (2023): 6546. DOI: 10.1038/s41467-023-42112-w
40) Matson, Vyara, et al. "The commensal microbiome is associated with anti–PD-1 efficacy in metastatic melanoma patients." *Science* 359.6371 (2018): 104-108. DOI: 10.1126/science.aao3290
41) McCulloch, John A., et al. "Intestinal microbiota signatures of clinical response and immune-related adverse events in melanoma patients treated with anti-PD-1." *Nature Medicine* 28.3 (2022): 545-556. DOI: 10.1038/s41591-022-01698-2
42) Morton, James T., et al. "Establishing microbial composition measurement standards with reference frames." *Nature communications* 10.1 (2019): 2719. DOI: 10.1038/s41467-019-10656-5
43) Olekhnovich, Evgenii I., et al. "Consistent stool metagenomic biomarkers associated with the response to melanoma immunotherapy." *mSystems* 8.2 (2023): e01023-22. DOI: 10.1128/msystems.01023-22
44) Olm, Matthew R., et al. "dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication." *The ISME journal* 11.12 (2017): 2864-2868. DOI: 10.1038/ismej.2017.126
45) Olm, Matthew R., et al. "inStrain profiles population microdiversity from metagenomic data and sensitively detects shared microbial strains." *Nature Biotechnology* 39.6 (2021): 727-736. DOI: 10.1038/s41587-020-00797-0
46) Pan, Shaojun, Xing-Ming Zhao, and Luis Pedro Coelho. "SemiBin2: self-supervised contrastive learning leads to better MAGs for short-and long-read sequencing." *Bioinformatics* 39.Supplement_1 (2023): i21-i29. DOI: 10.1093/bioinformatics/btad209
47) Parks, Donovan H., et al. "CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes." *Genome research* 25.7 (2015): 1043-1055. DOI: 10.1101/gr.186072.114
48) Parks, Donovan H., et al. "GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy." *Nucleic acids research* 50.D1 (2022): D785-D794. DOI: 10.1093/nar/gkab776
49) Pasolli, Edoardo, et al. "Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle." *Cell* 176.3 (2019): 649-662. DOI: 10.1016/j.cell.2019.01.001
50) Peng, Zhi, et al. "The gut microbiome is associated with clinical response to anti–PD-1/PD-L1 immunotherapy in gastrointestinal cancer." *Cancer Immunology Research* 8.10 (2020): 1251-1261. DOI: 10.1158/2326-6066.CIR-19-1014

51) Pozzi, Giulia, et al. "Buffering adaptive immunity by hydrogen sulfide." *Cells* 11.3 (2022): 325. DOI: 10.3390/cells11030325
52) Quinlan, Aaron R., and Ira M. Hall. "BEDTools: a flexible suite of utilities for comparing genomic features." *Bioinformatics* 26.6 (2010): 841-842. DOI: 10.1093/bioinformatics/btq033
53) Rahman, Gibraan, et al. "BIRDMAn: A Bayesian differential abundance framework that enables robust inference of host-microbe associations." *bioRxiv* (2023). DOI: 10.1101/2023.01.30.526328
54) Robert, Caroline, et al. "Pembrolizumab versus ipilimumab in advanced melanoma." *New England Journal of Medicine* 372.26 (2015): 2521-2532. DOI: 10.1056/NEJMoa1503093
55) Robert, Caroline, et al. "Pembrolizumab versus ipilimumab in advanced melanoma (KEYNOTE-006): post-hoc 5-year results from an open-label, multicentre, randomised, controlled, phase 3 study." *The Lancet Oncology* 20.9 (2019): 1239-1251. DOI: 10.1016/S1470-2045(19)30388-2
56) Routy, Bertrand, et al. "Fecal microbiota transplantation plus anti-PD-1 immunotherapy in advanced melanoma: a phase I trial." *Nature medicine* (2023): 1-12. DOI: 10.1038/s41591-023-02453-x
57) Rudis, Bob. hrbrthemes: Additional Themes, Theme Components and Utilities for 'ggplot2'. R package version 0.8.0, 2020.
58) Schneider VA, et al. "Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly". *Genome Research* 2017;27(5):849-864. DOI: 10.1101/gr.213611.116
59) Sidi, Fahim. downloadthis: Implement Download Buttons in "R Markdown". R package version 1.0.0, 2023.
60) Sieber, Christian MK, et al. "Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy." *Nature microbiology* 3.7 (2018): 836-843. DOI: 10.1038/s41564-018-0171-1
61) Sivan, Ayelet, et al. "Commensal Bifidobacterium promotes antitumor immunity and facilitates anti–PD-L1 efficacy." *Science* 350.6264 (2015): 1084-1089. DOI: 10.1126/science.aac4255
62) Spencer, Christine N., et al. "Dietary fiber and probiotics influence the gut microbiome and melanoma immunotherapy response." *Science* 374.6575 (2021): 1632-1640. DOI: 10.1126/science.aaz7015
63) Switzer B, Puzanov I, Skitzki JJ, Hamad L, Ernstoff MS. Managing Metastatic Melanoma in 2022: A Clinical Review. *JCO Oncology Practice*. 2022; 18(5):335-351. DOI: 10.1200/OP.21.00686
64) Toliver-Kinsky, Tracy, et al. "H_2S, a bacterial defense mechanism against the host immune response." *Infection and Immunity* 87.1 (2019): 10-1128. DOI: 10.1128/IAI.00272-18
65) Trischler, Raphael, et al. "A functional Wood–Ljungdahl pathway devoid of a formate dehydrogenase in the gut acetogens Blautia wexlerae, Blautia luti and beyond." *Environmental microbiology* 24.7 (2022): 3111-3123. DOI: 10.1111/1462-2920.16029
66) Tsakmaklis, Anastasia, et al. "TIGIT+ NK cells in combination with specific gut microbiota features predict response to checkpoint inhibitor therapy in melanoma patients." *BMC Cancer* 23.1 (2023): 1160. DOI: 10.1186/s12885-023-11551-5
67) Wickham, Hadley. stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.5.0, 2022.
68) Wickham, H. (2016). ggplot2: Elegant graphics for data analysis (2nd ed.). Springer-Verlag.

69) Woodcroft, Ben J. Kingfisher: Rapid SRA Download Tool. Version 0.4.1, 2022.
70) Wu, Yu-Wei, Blake A. Simmons, and Steven W. Singer. "MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets." *Bioinformatics* 32.4 (2016): 605-607. DOI: 10.1093/bioinformatics/btv638
71) Xie, Yihui, Joe Cheng, and Xianying Tan. DT: A Wrapper of the JavaScript Library 'DataTables'. R package version 0.28, 2023.
72) Xu S, Hu E, Cai Y, Xie Z, Luo X, Zhan L, Tang W, Wang Q, Liu B, Wang R, Xie W, Wu T, Xie L, Yu G (2024). "Using clusterProfiler to characterize multiomics data." *Nature Protocols*, 19(11), 3292-3320. DOI: 10.1038/s41596-024-01020-z
73) Ye, Ling, et al. "Repressed Blautia-acetate immunological axis underlies breast cancer progression promoted by chronic stress." *Nature Communications* 14.1 (2023): 6160. DOI: 10.1038/s41467-023-41817-2
74) Yatsunenko, Tanya, et al. "Human gut microbiome viewed across age and geography." *Nature* 486.7402 (2012): 222-227. DOI: 10.1038/nature11053
75) Zhang, Jun, et al. "GseaVis: An R Package for Enhanced Visualization of Gene Set Enrichment Analysis in Biomedicine." *Med Research* (2025). DOI: 10.1002/mdr2.70000
76) Zhu, Qiyun, et al. "Phylogeny-aware analysis of metagenome community ecology based on matched reference genomes while bypassing taxonomy." *mSystems* 7.2 (2022): e00167-22. DOI: 10.1128/msystems.00167-22
77) Zakharevich, Natalia V., et al. "Systemic metabolic depletion of gut microbiome undermines responsiveness to melanoma immunotherapy." *Life Science Alliance* 7.5 (2024). DOI: 10.26508/lsa.202302480
78) Zaneveld, Jesse R., Ryan McMinds, and Rebecca Vega Thurber. "Stress and stability: applying the Anna Karenina principle to animal microbiomes." *Nature microbiology* 2.9 (2017): 1-8. DOI: 10.1038/nmicrobiol.2017.121

Figure captions

Figure 1. Global distribution of collected samples (n=624) across countries. Circle size represents sample count per region, while color indicates the proportion of patients R - responsive (or NR - non-responsive) to immunotherapy.

Figure 2. (A) The boxplot shows log ratio (LR) values obtained using marker OGU. LR is highest in the R group (median = 3.0) and lowest in NR group (median = 1.4). (B) Receiver operating characteristic (ROC) curve for log regression performance using "leave-one group out" (LOGO) cross-validation.

Figure 3. (A) Gene Set Enrichment Analysis (GSEA) at the phylum level comparing responders and non-responders. (B) GSEA at the genus level showing differential enrichment patterns between groups. (C-D) Origin-based stratification of microbial communities: (C) food-derived taxa and (D) oral cavity-associated species. Significance levels were adjusted via Bendjamini-Hochberg correction for the multiple hypothesis testing (threshold adj. $p < 0.05$).

Figure 4. (A) Gene Set Enrichment Analysis (GSEA) of carbohydrate-active enzymes (CAZy) reveals differential representation of glycoside hydrolase families between responders and non-responders. (B) Pathway-level analysis using KEGG database highlights distinct metabolic networks enriched in each group. Significance thresholds account for multiple comparisons (Bendjamini-Hochberg adj. $p < 0.05$).

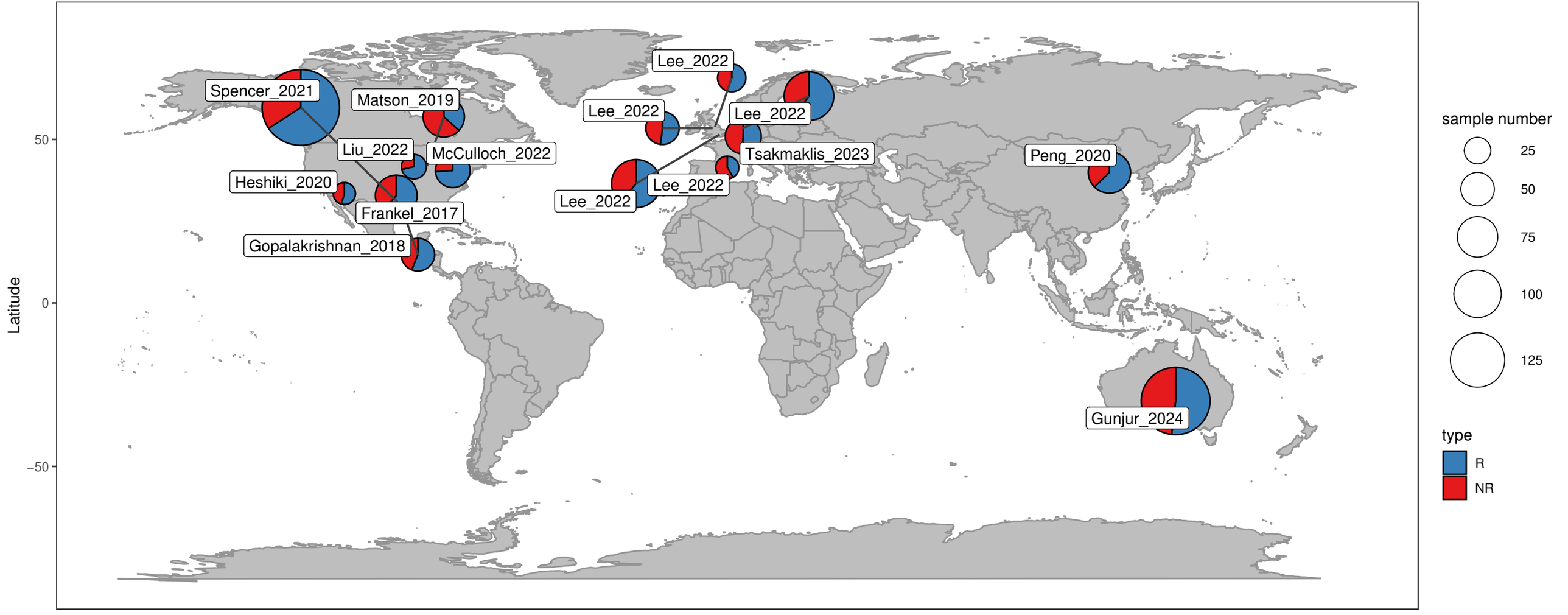
Supplementary materials

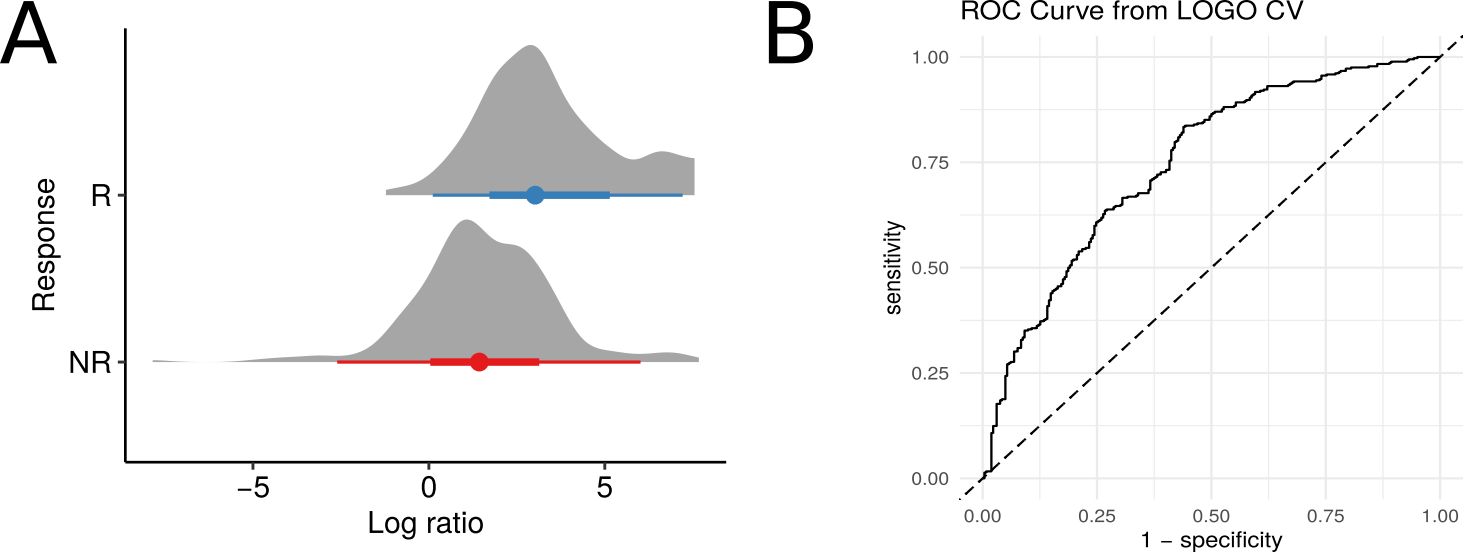
Tables

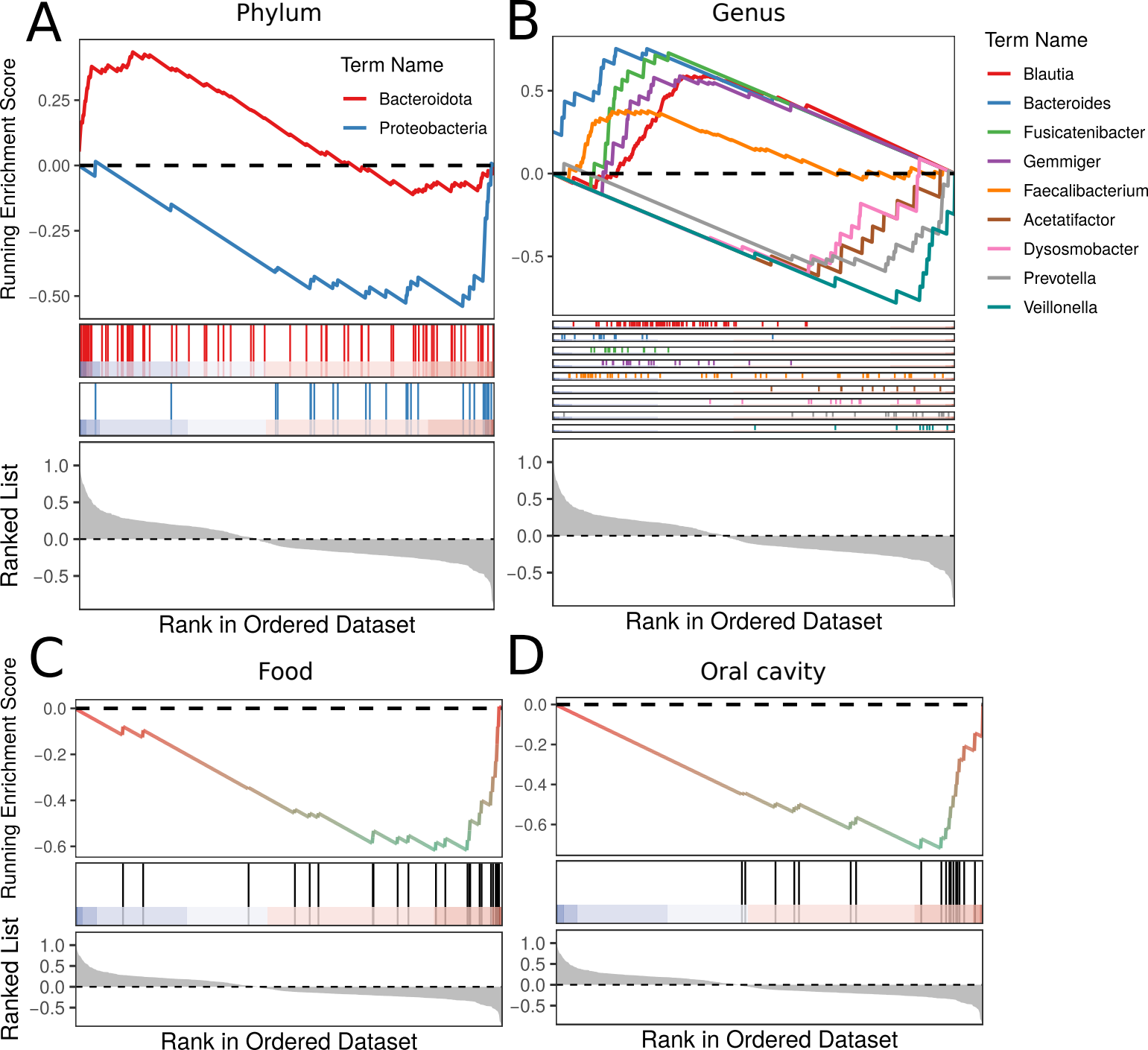
Table S1. Sample metadata.
Table S2. Quality control statistics.
Table S3. Operational genomic unit (OGU) abundance profile by samples.
Table S4. Marker OGU with Sondbird coefficients (MEAN ± SD).
Table S5. Log ratio (LR) values of marker OGU across samples.
Table S6. Log regression prediction metrics.
Table S7. GSEA results at the phylum level.
Table S8. GSEA results at the phylum level: Bacteroidetes core enrichment.
Table S9. GSEA results at the phylum level: Proteobacteria core enrichment.
Table S10. GSEA results at the genus level.
Table S11. OGUs clustered with food-associated MAGs at 95% nucleotide similarity.
Table S12. OGUs clustered with different bodysite-associated MAGs at 95% nucleotide similarity.
Table S13. GSEA results at food-related OGUs.
Table S14. GSEA results at bodysite-related OGUs.
Table S15. GSEA results at food-related OGUs: core enrichment.
Table S16. GSEA results at bodysite-related OGUs: core enrichment.
Table S17. GSEA results with CAZy database.
Table S18. GSEA results with CAZy database: core enrichment.
Table S19. GSEA results with KEGG database.
Table S20. Differential KEGG pathway enrichment in *Faecalibacterium prausnitzii* strains stratified by host response status.

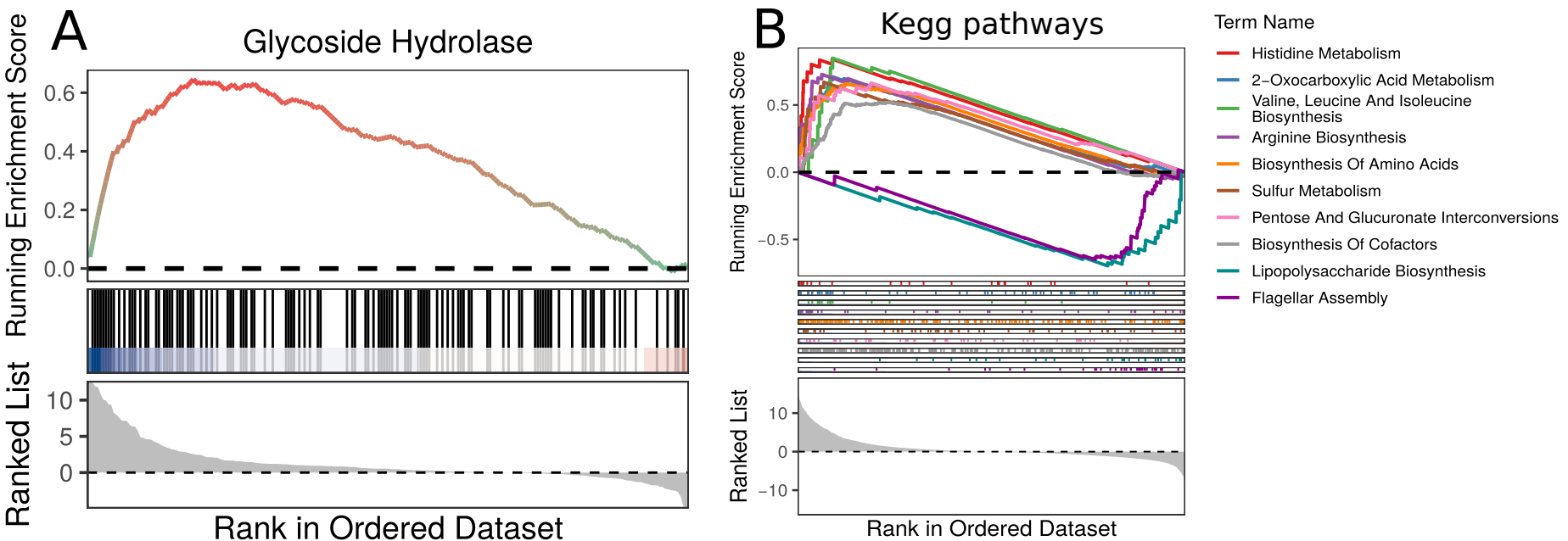
Figures

Figure S1. The barplot depicts Songbird differential abundance coefficients (MEAN ± SD) for marker Operational Genomic Units (OGUs), with the x-axis representing individual OGUs and the y-axis showing the magnitude and direction of association coefficients.
Figure S2. Distribution of log regression prediction metrics across datasets obtained by LOGO cross-validation.
Figure S3. Quantile-quantile diagram of mixed linear model residuals.
Figure S4. Marker OGU contains CAZy core enrichment glycoside hydrolase within the genome.
Figure S5. Marker OGU contains KEGG pathways core enrichment gene groups within the genome.
Figure S6. Marker OGU contains overrepresented R group WL-pathway genes within the genome.


Spencer_2021
Matson_2019
Liu_2022
McCulloch_2022
Heshiki_2020
Frankel_2017
Gopalakrishnan_2018
Lee_2022
Lee_2022
Lee_2022
Tsakmaklis_2023
Lee_2022
Lee_2022
Peng_2020
Gunjur_2024
Latitude
50
0
−50
sample number
25
50
75
100
125
type
R
NR


A
Response
R
NR
−5
0
5
Log ratio
B
ROC Curve from LOGO CV
sensitivity
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
1 − specificity


A
Phylum
Term Name
Bacteroidota
Proteobacteria
Running Enrichment Score
0.25
0.00
−0.25
−0.50
Ranked List
1.0
0.5
0.0
−0.5
Rank in Ordered Dataset
B
Genus
Term Name
Blautia
Bacteroides
Fusicatenibacter
Gemmiger
Faecalibacterium
Acetatifactor
Dysosmobacter
Prevotella
Veillonella
0.5
0.0
−0.5
1.0
0.5
0.0
−0.5
Rank in Ordered Dataset
C
Food
Running Enrichment Score
0.0
−0.2
−0.4
−0.6
Ranked List
1.0
0.5
0.0
−0.5
Rank in Ordered Dataset
D
Oral cavity
0.0
−0.2
−0.4
−0.6
1.0
0.5
0.0
−0.5
Rank in Ordered Dataset


A
Glycoside Hydrolase
Running Enrichment Score
0.6
0.4
0.2
0.0
Ranked List
10
5
0
Rank in Ordered Dataset
B
Kegg pathways
Running Enrichment Score
0.5
0.0
−0.5
Ranked List
10
0
−10
Rank in Ordered Dataset
Term Name
Histidine Metabolism
2−Oxocarboxylic Acid Metabolism
Valine, Leucine And Isoleucine Biosynthesis
Arginine Biosynthesis
Biosynthesis Of Amino Acids
Sulfur Metabolism
Pentose And Glucuronate Interconversions
Biosynthesis Of Cofactors
Lipopolysaccharide Biosynthesis
Flagellar Assembly