Integrative Analysis of the Mouse Cecal Microbiome Across Diet, Age, and Metabolic State in the Diverse BXD Population

Ziyun Zhou [1], Arianna Lamanna [1], Rashi Halder [1], Emeline Pansart [1], Besma Boussoufa [1], Thamila Kerkour [2], Paul Wilmes [1], and Evan Williams [1,✉]

[1]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette 4362, Luxembourg
[2]Department of Dermatology, Erasmus MC Cancer Institute, 3015 GD Rotterdam, The Netherlands

Abstract

Background

The gut microbiota both adapts to, and shapes, the host's metabolic state through metabolites and gene regulatory networks, influencing multiple organs via axes such as the gut–liver connection. While diet, age, and host genetics each modulate gut microbial composition, their combined effects and mechanistic links to host molecular phenotypes remain poorly understood. Integrated multi-omics approaches in genetically diverse populations offer an opportunity to dissect these interactions and identify predictive microbial and host features.

Results

We profiled the cecal metagenome, metatranscriptome, and host transcriptome from 232 mice spanning two diets (chow, high-fat), four adult ages (6–24 months), and 43 BXD strains. Diet exerted the strongest influence on microbiota composition, followed by genetic background and age, with high-fat feeding reducing diversity and altering >300 species. Machine learning models based on microbial profiles accurately predicted body weight (AUC = 0.92) and chronological age (AUC = 0.84), with accuracy rising to 0.95 when integrating top microbial features with liver proteomics. Network analyses revealed specific host–microbe links, including a negative association between cecal *Ido1* expression and short-chain fatty acid (SCFA)-producing *Lachnospiraceae*, suggesting dietary fat may modulate host tryptophan metabolism through microbiota shifts.

Conclusions

This multi-omics study maps how diet, age, and genotype jointly shape the cecal microbiome and host molecular responses, identifying robust microbial biomarkers of aging and metabolism. Using inbred mice sampled across time, we defined signature taxonomic sets with strong

predictive value for future metabolic outcomes, driven by microbial–host metabolic networks linking the cecum and liver. By advancing understanding of the gut–liver axis, our findings offer mechanistic hypotheses and compact biomarker panels for earlier, personalized diagnosis of metabolic disease, tailored to an individual's genetic background and environmental context.

Microbiota | Metagenome | Metatranscriptome | Aging | Machine Learning | GXE | Recombinant Inbred Strains | Multi-Omics

Correspondence: *evan.williams@uni.lu*

Introduction

The typical, individual mammalian intestinal tract is colonized by trillions of bacteria, coming from thousands to tens of thousands of species, collectively called the gut microbiota [1]. Acting as a crucial link between diet and host metabolism, the gut microbiota play an essential role in digestion and nutrient absorption and can impact overall metabolic health [2]. For instance, "germ-free" mice born and raised in sterile environments do not become obese when fed a high-fat and high-carbohydrate diet [3], but they also have immunodeficiencies [4], and with age they tend to suffer from a variety of metabolic disorders such as a dramatic hypertrophy of the cecum [5]. In parallel with diet, aging also alters the diversity and stability of gut microbiota [6]. These aging-associated changes are linked to increased inflammation and a reduction in beneficial metabolite production, which may further accelerate metabolic and immune decline [7].

The cecum, located at the junction between the small intestine and the colon, is a notable region for microbial activity, as it serves as a primary fermentation site for indigestible fibers, particularly for the production of short-chain fatty acids (SCFAs) such as butyrate [8]. The cecum is anatomically highly distinct, facilitating equivalent sample collection over long study periods (6 years in the case of our colony). The cecum is a "cul-de-sac" and has a relatively long residence time, with it filling by 6 hours after feeding and then emptying over the next 16 hours [9]. The relative ease of consistent sample collection over time and across variation in time since the last meal makes the cecum an ideal source for microbial sampling. Furthermore, and in contrast to collecting fresh feces from the rectum, cecal collection permits measurement of both the microbe content and the gene expression networks of immediately-surrounding mouse tissue, which can offer direct insights into systemic host-microbiota interactions that are essential for understanding the etiology of metabolic diseases [8, 10].

Given the high diversity of the microbiome and the colonial nature of microbe-microbe interactions, research in this field has broadly struggled to identify specific causal mechanistic relationships between particular microbial taxa to predictable phenotypic outcomes. However, numerous clinical studies over the past decade have shown that the gut microbiome as a whole can have a causal effect on both causing and curing complex metabolic diseases. For instance, mice with inflammatory bowel disease (IBD) can be cured by oral delivery of specific *Akkermansia* bacteria [11], and fecal microbiota transplants (FMT) have been widely proven to cure *Clostridioides difficile* infections in humans, which hitherto had been notoriously difficult to

treat [12]. However, successful clinical studies using FMT result in whole-scale shifts affecting hundreds of billions of bacteria, meaning that the specific causal bacteria and their mechanisms are exceedingly difficult to identify, validate, standardize, and reproduce at scale. The causal connection between the gut microbiome to diseases like IBD and colon cancer are unsurprising, yet in recent years, numerous pioneering clinical studies have further found the microbiome to be causally associated with a dizzying variety of diseases far more distal to intestinal health, from Parkinson's disease [13] to type I diabetes (T1D) [14] and type II diabetes (T2D) [15]. However, performing FMT to treat such systemic diseases, and metabolic phenotypes such as morbid obesity, have had unpredictable results [16]. More detailed mechanistic understanding of the relationship between the microbiome, host, and metabolic health is thus necessary to develop new, targeted treatments.

Therefore, in this study, we generated multi-omics data in the ceca from 232 mice varying according to the interactions between our three independent experimental variables: age (6 to 24 months), diet (low fat "chow" (CD) or high fat (HF)), and genotype (43 inbred strains from the BXD mouse genetic reference population (GRP)). In this cohort, we measured the metagenome (MG) and metatranscriptome (MT) from the cecal contents, and the transcriptome of the cecum tissue itself (abbreviated as "mRNA"). These newly generated data are then supplemented with our prior phenotype data on these mice, along with these individuals' liver transcriptome and proteomes [17]. We aim to first identify and quantify the abundance and activity of specific microbial taxa associated with diet and age, and then apply both traditional bioinformatics approaches and newer machine learning methods to calculate their relationships with key metabolic phenotypes like obesity and expression of key gene networks. We show that these multi-tissue, multi-omic data can be used to predict not only major phenotype outcomes, but also provide insight into the cellular activity of other tissues in the mouse. Thus, not only can we identify specific associations between the microbiome and metabolic health, but we can also infer the affected mechanistic pathways, and — by using the different ages of mice — suggest the causal direction, consistency, and timeline of the connection.

Results

Study Design

In 2021, we completed the phenotyping of an aging population of 2157 mice from 89 BXD strains [18], tested in two diets with calories obtained predominantly from carbohydrates (i.e. CD) or from fats (i.e. HF). 662 of these individuals were taken for tissue collection at four approximate intervals: 7, 12, 18, and 24 months of age. Our laboratory has recently completed a detailed molecular analysis of the paired transcriptome, proteome, and metabolome in 347 livers of these mice from a subset of 58 strains across diet and age [17]. In that study, we were able to identify predictive molecular signals of a mouse's future outcomes (such as lifespan), as well as regulatory genes affecting health (such as susceptibility to diet-induced obesity) and the control of molecular pathways affecting aging and diet related genes (such as cholesterol synthesis). From that aging study, we took the whole, flash-frozen cecum samples from 597 individuals in 56 different strains (all overnight fasted, CD n = 299, HF n = 298). For those 597 individuals, there are typically two biological replicates for the selected strain-diet-age matches. Thus, we selected cecum from a subset of 232 mice in order to study a wide range of interacting variables in a cost-conscious manner. For these 232 mice, we have sequenced the MG (n = 200 mice, of which 10 were performed in replicate), MT (n = 80), and the cecal mRNA (n = 80) (Figure 1 A). All MT and mRNA mice were the same, while 48 overlapped between MT and MG, and overall 155 of these samples overlapped with our prior liver transcript and protein expression data. To ensure consistent extraction across all samples, the entire frozen cecal tissue was homogenized in liquid nitrogen and a powdered aliquot of ~30 mg was taken both for DNA and RNA extraction. The MT and cecal mRNA samples were co-extracted using the same tissue aliquot. MG, MT, and cecal mRNA samples were extracted and sequenced in two temporally distinct batches (November 2022 and November 2023). An overview of the experimental and analytical workflow, including extraction, sequencing, quality control, and downstream processing, is presented in Figure 1 B. For the extracted DNA, samples were then cleaned up by size exclusion to remove large DNA fragments. As we are performing shotgun sequencing on the MG samples (rather than 16S), this step was necessary to remove most of the mouse DNA content of each sample, which was not of interest given that the BXDs are inbred and have recently been fully sequenced [19]. After cleanup, the resulting relative proportions of the microbial to host DNA content for each sample were estimated using RT-qPCR, and samples were subsequently ranked and filtered

based on these ratios (Figure S1 A-B). Samples that contained more than 40% microbe DNA were then retained for shotgun sequencing. For extracted total RNA, samples were first tested to determine the presence of four clear rRNA bands (i.e. indicating good quality abundance of both microbial and mouse RNA), and such samples were then selected to obtain mRNA (cecal) and MT measurements (Figure S1 C). These samples were then separated based on poly(A) enrichment, with the poly(A)-enriched sample being sequenced to obtain the cecal tissue mRNA. The fraction without poly(A) was then treated with restriction enzymes to deplete mouse ribosomal RNA before MT sequencing.

After read alignment, counting, and normalizing, we used ANOVA to systematically evaluate the relative contributions of the three independent variables — genotype ("strain"), diet, and age — to the observed variation in our multi-omics data, across the MG, MT, and cecal mRNA datasets. These results revealed that genotype, diet, age, and their interactions together explained, on average, 60-70% of the observed variation in MG (Figure 1 C), MT, and cecal mRNA datasets (Figure S1 D). While strain had by far the largest impact on variance, it is important to note the difference in degrees of freedom between the independent variables. The samples spanned 43 BXD strains, but only two diets, and the ages of sacrificed animals were clustered specifically around four timepoints in adulthood (i.e. around 6, 12, 18, and 24 months of age, with a margin of around 30 days). To address this imbalance in the degrees of freedom for each class of independent variable, we calculated the F-statistic, which is adjusted for that concern. This confirmed that diet individually exhibited the strongest adjusted association with gut microbiota composition, generally driving the greatest proportion of variance, although even once adjusted, strain still plays a major role (Figure 1 D). The major effect of diet on the gut microbiome is expected and well-documented, so to mitigate potential confounding effects arising from the strong diet-microbiota effect, subsequent analyses targeting diet utilized the full dataset, while age- or strain- related analyses were separately performed within the CD or HF dietary groups, as will be indicated for each analysis.

To further evaluate the internal consistency and reliability of the gut microbiome datasets, we examined genus-level correlations between MG and MT profiles for the 48 overlapping samples. The correlation distribution was strongly shifted toward positive values compared to a shuffled control (median correlation of around 0.75), indicating strong concordance between DNA- and RNA- based taxonomic estimates (Figure S1 E). In addition, principal component analysis (PCA) of species-level MG profiles precisely overlap between the two chronological extraction batches,

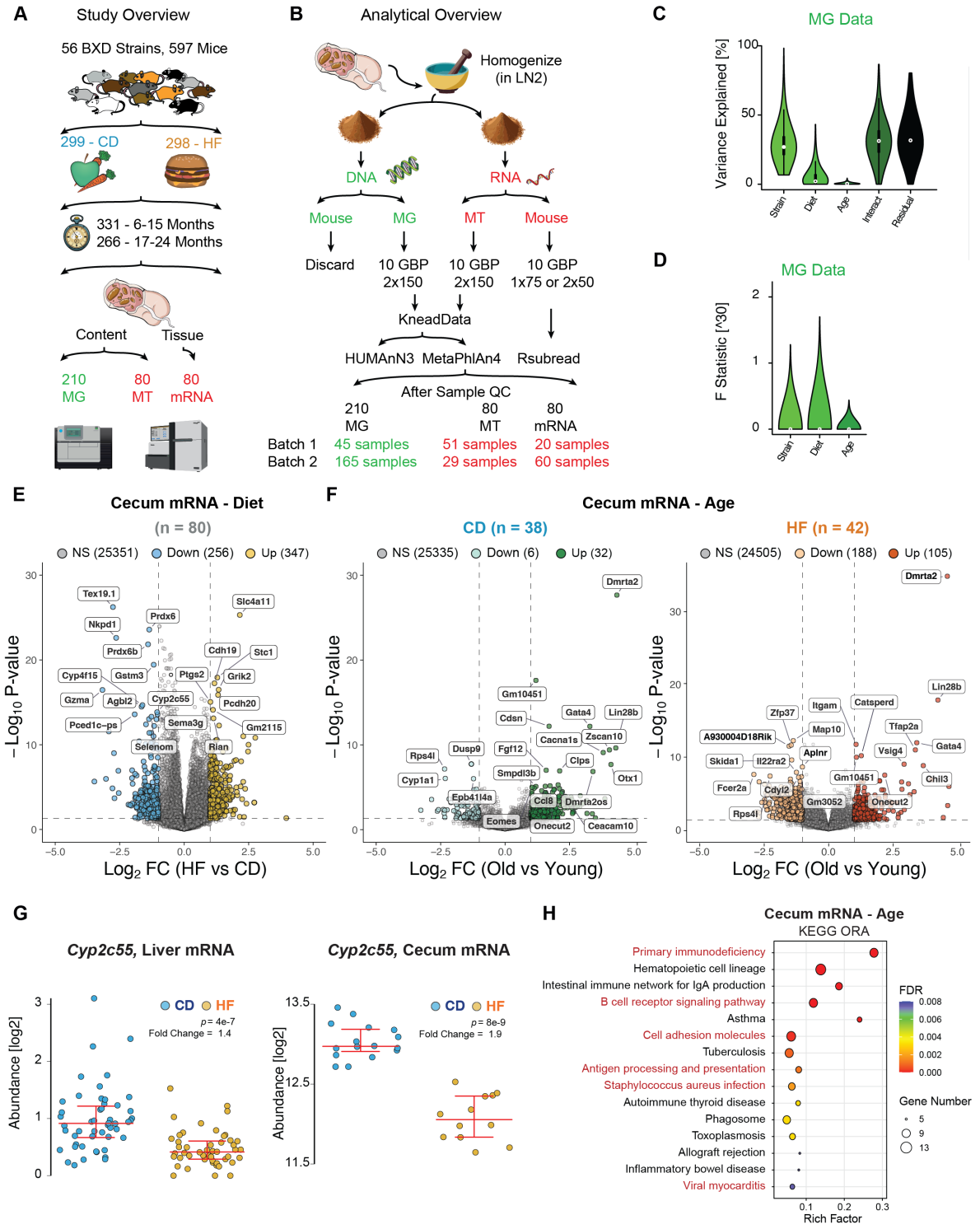

Figure 1. Experimental workflow and analysis of multifactorial cecal datasets.

(**A**) Study design for mouse population analysis and data generation of three different omics types from the cecum. (**B**) Experimental process for extraction, separation, sequencing, and analysis of the cecal samples. (**C**) ANOVA analysis of all MG measurements. "Interact" represents the combined effects of all interactions, including Strain*Diet, Strain*Age, Diet*Age, Strain*Diet*Age. (**D**) F statistics, computed based on species-level abundance data, for the effects of Strain, Diet, and Age. (**E-F**) Volcano plots of DEGs influenced by diet and age (separated by diet in panel F, due to its strong impact). (**G**) Dot plot showing the diet effect on *Cyp2c55* expression in liver and cecum, going in the same direction. (**H**) The 15 most significantly age-associated KEGG pathways, based on DEGs from the cecal mRNA dataset.

indicating that batch control was effective (Figure S1 F). Similarly, both batches of cecal mRNA showed no significant differences after both were transformed and variance stabilized (Figure S1 G). Together, these results demonstrate high internal agreement and minimal batch effects, supporting the reliability of the MG, MT, and cecal mRNA datasets for downstream integrative analyses.

Diet and Aging Alter the Cecal mRNA

Host cecal mRNA data were aligned and quantified using Rsubread, and normalized gene counts were used for downstream analysis (Figure 1 B). In this study, we aim to also connect the gene expression mechanisms to bacterial abundances and consequent metabolic outcomes, so we first analyzed the direct effect of our key independent variables of diet and age on the mRNA expression in the ceca (Figure 1 E-F). Differential expression analysis of cecal mRNA revealed distinct transcriptional responses to both diet and age. A total of 603 significant differentially expressed genes (DEGs) were observed (FDR < 0.05, $|\log_2 FC| > 1$) between CD and HF groups, including 347 upregulated and 256 downregulated genes by HF, Figure 1 E). Notably, genes such as *Slc4a11*, *Stc1*, and *Cdh19* were among the most strongly induced by HF feeding, while *Tex19.1*, *Nkpd1*, and *Prdx6b* were downregulated.

We next examined age-associated transcriptional changes, separately for each dietary condition. Under long-term CD, only 38 genes (32 upregulated, 6 downregulated) were differentially expressed with age, suggesting that aging has a relatively limited or non-uniform impact on cecal mRNA expression in the absence of dietary challenge (Figure 1 F, left panel) - also in line with the lower contribution of age to variance in the ANOVA analysis. In contrast, the HF group exhibited a comparatively enhanced age-related transcriptomic shift, with 293 genes significantly altered (105 upregulated, 188 downregulated; Figure 1 F, right panel). To identify age-associated transcriptional signatures consistent across dietary conditions, we compared the 20 DEGs most affected by age from the CD and HF age subsets, including both upregulated and downregulated genes. Among the genes upregulated in old mice, five were shared in both diets: *Dmrta2*, *Lin28b*, *Gata4*, *Onecut2*, and *Gm10451*. In contrast, only *Rps4l* was consistently upregulated in young mice under both diets. Several of these overlapping genes are known to change with age, such as *Gata4* [20] and *Rps4l* [21], whereas the roles of others in aging remain uncharacterized. However, this consistent link with age across diet and numerous diverse genotypes suggests an underlying cellular mechanism. Regarding diet, which has been far more extensively studied,

many transcripts most affected by HF diet were in line with expectations from literature and from our prior liver dataset in these individuals. For example, the expression level of *Cyp2c55* is expected to decrease in the gut under HF feeding [22]. This trend was also observed in our dataset, which showed significantly lower *Cyp2c55* mRNA expression in both the liver (fold change = 1.4) and cecum (fold change = 1.9) in HF (Figure 1 G).

To further explore functional pathways associated with age and diet, we applied both over-representation analysis (ORA) based on DEGs and gene set enrichment analysis (GSEA) using ranked differential expression values. For diet-associated changes, using an FDR threshold of < 0.05, ORA identified 24 significantly enriched pathways, while GSEA revealed 99 pathways. 16 pathways were consistently detected by both approaches. The top 15 enriched pathways from each method are shown in Figure S2 A–B, with overlapping pathways highlighted in red. In addition, KEGG pathway analysis revealed that several metabolic pathways were significantly downregulated in the HF group, including "metabolism of xenobiotics by cytochrome P450", "chemical carcinogenesis – DNA adducts", "steroid hormone biosynthesis", and "retinol metabolism". In contrast, the "neuroactive ligand – receptor interaction" pathway was significantly upregulated in response to HF. Several pathways have been previously linked to HF intake. For instance, the "steroid hormone biosynthesis" pathway has been observed to be significantly suppressed under HF conditions [23]. This suppression may stem from HF-induced mitochondrial dysfunction that impairs cholesterol transport into mitochondria, the first step of steroid hormone biosynthesis [24]. This pathway was enriched with 15 DEGs (e.g., *Cyp2c55*, *Cyp3a11*, *Cyp1a1*, *Cyp2e1*, *Ugt1a9*) (Figure S2 C, upper panel). Moreover, the "neuroactive ligand–receptor interaction" pathway has also been reported to be upregulated by HF in the hypothalamic transcriptome of mice [25], potentially driven by the diet–gut–brain axis. In our data, this pathway was highly enriched (NES = 2.05, p = 2e-9) with key DEGs including *Lep*, *Grik1*, *Avpr2*, and *Grm7*, among others (Figure S2 C, lower panel).

Building on the same analytical framework, we next investigated how age influences pathway-level changes through both ORA and GSEA (Figure 1 H, Figure S2 D). ORA identified 27 significantly enriched pathways, while GSEA revealed 58, with 19 pathways consistently detected by both methods. Among these, the B cell receptor signaling pathway was significantly enriched in both analyses. This finding aligns with previous reports in humans and animal models, which have shown that aging is often accompanied by a reduction in B cell numbers and functional impairments within B cell populations [26, 27]. Furthermore, this pathway included 10 DEGs

(Figure S2 E, left panel). Conversely, several pathways positively associated with aging have also been extensively studied, such as the NOD-like receptor signaling pathway (Figure S2 E, right panel), which is upregulated with age but can be effectively suppressed by fecal microbiota transplantation from young donors [28]. These cecal mRNA results reveal distinct pathway alterations driven by diet and age, offering a potential framework for integration with microbial functional data.

MG and MT Analyses Reveal Cecal Microbiome Changes Associated with Dietary Change, Aging and Strains

We next examined the broad taxonomic profiles of the gut microbiota, using quality-controlled MG and MT reads processed with KneadData [29] and analyzed using MetaPhlAn4 [30] (Figure 1 B). Compared to the CD group, HF diet reduced the variety of identified taxa by around 30 to 50% across all taxonomic levels in both MG and MT data (Figure S3 A). In addition, α-diversity remained consistently and significantly decreased throughout life, with an equivalent reduction in both MG and MT measurements (Figure 2 A). These findings are consistent with previous widely reported observations that a HF diet reduces microbiome diversity, including at the earliest timepoint (roughly 90 days after the start of HF feeding), as the reduction in diversity is rapid, typically within less than a month [31]. However, the reduction in diversity with diet is an average. Given that our mouse population is highly genetically variable, we observed some overlap between CD mice with the lowest α-diversity and HF mice with the highest α-diversity, and this distinction remains stable over time. Examining how general patterns in our microbiome data align with established findings from the literature also served as a way to validate the reliability of the extraction, sequencing, and analytical procedures. In addition, β-diversity analyses (Bray-Curtis dissimilarity) demonstrated that both diet and age significantly influenced gut microbial community structure (Figure S3 B). Diet explained a larger proportion of the variance in community composition and was associated with more pronounced shifts across samples. In contrast, age accounted for a smaller portion of the variance but still exhibited statistically significant effects.

We then checked whether these taxonomic signatures in the ceca will reliably indicate general health outcomes, rather than only recapitulate the known dietary state of the mice. To initially test this, we checked the ratio between the *Firmicutes* and *Bacteroidetes* phyla, which is consistently reported in literature to track with obesity as well as HF diets [32]. Here, we again observed a

highly significant elevation in HF cohorts (Figure 2 B), and both *Firmicutes* and *Bacteroidetes* abundances were significantly correlated with body weight (Figure S3 C). However, when age was considered as a factor, no significant difference in the *Firmicutes / Bacteroidetes* ratio was observed in older mice (Figure 2 C). This may be due to the general decrease in gut microbial diversity and the increase in individual variation with aging, which could render the relationship between this ratio and body weight less stable. Moreover, while the overall trends are significant, there is substantial overlap for individual mice within the two diets, indicating that genetic background (i.e. strain) plays a substantial role in the strength of the connection between *Firmicutes* and *Bacteroidetes* abundances and dietary response. Moreover, both *Firmicutes* and *Bacteroidetes* are large and highly-diverse phyla of bacteria, and their lower taxonomic branches can have diverse effects. For instance, within the *Firmicutes* phylum, *Lactobacillus reuteri* has been associated with increased obesity, whereas *Lactobacillus gasseri SBT2055* exhibits anti-obesogenic properties [33]. This highlights the significance of generating MG data (or using long-read 16S sequencing), which allow precise analysis at more specific taxonomic rankings, down to the genus, species, and even strain level.

As a final quality control, we aimed to show that the specific taxonomies quantified also highlighted reliable, reproducible associations with host health. To this end, we analyzed genus-level relative abundance profiles derived from both MG and MT datasets (Figure 2 D) to dissect microbial composition at a deeper level. Samples were grouped by diet, and within each dietary group, individuals were ordered from left to right by increasing age. We visualize the 15 most abundant genera identified independently in the MG and MT datasets, along with the top 10 shared genera, resulting in a total of 34 genera (plus “Others”, which is the sum of all smaller genera). Here, we again observe marked differences in genus composition between the CD and HF groups, such as a large decrease in *Duncaniella* in HF, and a substantial increase in a large cluster of unclassified genera within the Lachnospiraceae family (bottom two bars in light blue and pale yellow, respectively, Figure 2 D). In contrast to diet, few patterns are visually evident when assessing based on the X-axis (age), consistent with the ANOVA results which showed that diet explains a greater proportion of the variance in microbial composition than age (Figure 1 C, Figure S1 D). Overall, 24 taxa were significantly regulated by diet in the MG data and 19 in the MT data, of which 18 were significant in both datasets.

As with the expected overall decrease in diversity with HF, specifically-affected taxonomic classifications also broadly matched expectations from literature. For instance,

Lachnospiraceae_unclassified, *Lactobacillus*, and *Bacteroides* exhibited increased relative abundance in the HF group across both MG and MT datasets [34, 35], while *Lactococcus* [36], *Prevotella*, and *Lactobacillus* were less abundant in HF across both measurement types. However, certain highly-consistent and variable genera, such as *GGB62427*, were less well-characterized and have not been extensively studied in the context of diet. Together, these results provide a valuable resource for identifying candidate genera consistently associated with obesity-promoting diets independently of genetic background.

Given the complex influences of diet, age, and strain, we next examined species-level profiles to uncover more specific microbial associations. Based on the results of MaAsLin3 [37], we identified a total of 274 gut microbial species at the MG level that were significantly associated (FDR < 0.05) with diet, and 31 species correlated with age. The 25 most strongly and significantly diet-associated features are displayed for the MG and MT datasets in Figure 2 E and Figure S3 D, respectively. The β-coefficients from the multivariable models indicate both the direction and magnitude of association with the HF diet. Since we used relative abundance data, both abundance and prevalence were considered in the analysis. Positive values indicate higher abundance or prevalence in HF, while negative values indicate depletion. Of the top 25 species selected, 12 were detected as significantly altered in both abundance and prevalence models. While the results do not always overlap, the discrepancies and alignments help provide a more comprehensive interpretation of the relationship between species and diet. For example, among the top features, *Lactococcus_cremoris* has both increased abundance and prevalence in the HF group, suggesting not only a higher overall abundance but also a broader presence across individuals in the cohort. In comparison, *Desulfovibrio_SGB41239* exhibits increased abundance but decreased prevalence under HF, indicating that while this taxon becomes more abundant in certain individuals, its occurrence became more restricted across the population. This discrepancy may reflect ecological heterogeneity within the HF group, for instance *Desulfovibrio_SGB41239* had a negative prevalence change across the entire cohort, but it had an increased abundance in a subset of mice which may have distinct gut environments or host responses. In contrast, species such as *Muribaculum_gordoncarteri* and *Duncaniella_dubosii* were consistently depleted in both abundance and prevalence by HF, highlighting their general reduction in both quantity and distribution. The observed species-level shifts mirrored those at the genus or family level, but provided more precise identification of diet-associated gut microbes. Notably, four species appeared among the top 25 diet-associated

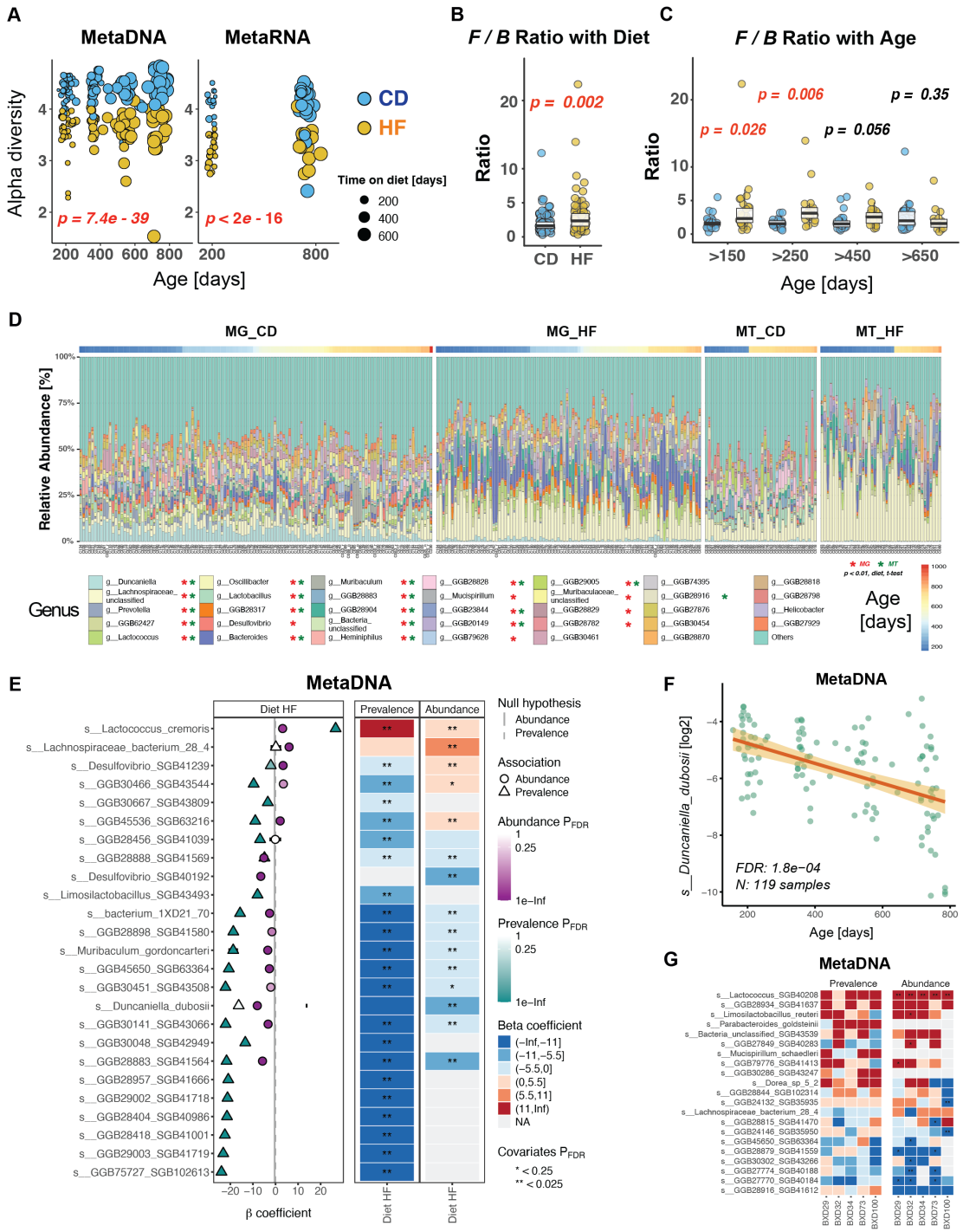

Figure 2. Microbial Diversity and Composition Are Shaped by Diet, Age, and Genotype in MG and MT Datasets.

(**A**) The α-diversity at the MG and MT levels, calculated using the Shannon index, showing significantly reduced diversity under HF. (**B**) The *Firmicutes* / *Bacteroidetes* ratio is significantly increased in HF cohorts. (**C**) Diet-associated differences in the F/B ratio diminish with age. (**D**) Stacked bar plots of the relative abundance of the 34 most abundant genera in the MG (left panels) and MT (right panels) datasets, ranked by diet-associated t-test significance. ⋆indicates genera with *p*-values < 0.01. (**E**) The top 25 species most significantly associated with diet in the MG dataset, identified by MaAsLin3. Direction and significance of associations are shown for prevalence and abundance. (**F**) *s__Duncaniella_dubosii*, the most strongly age-correlated taxon identified by MaAsLin3, shows a significant negative correlation with age. (**G**) Correlation matrix showing strain-level differences in diet associations for selected species, using B6 as the reference.

taxa in both the MG and MT datasets (Figure S3 D). For example, the abundance of *Duncaniella_dubosii* was significantly reduced by the HF in both the MG and MT datasets (Figure S3 E). Their consistent detection across MG and MT levels suggests that these taxa are not only present, but also transcriptionally active in response to dietary changes. In addition to well-characterized species, several unclassified taxa showed concordant patterns between MG and MT. For example, *bacterium_1XD21_70* consistently decreased in both abundance and prevalence under the HF diet (Figure S3 F-G), despite lacking functional annotation. These shared and reproducible species may serve as robust candidates for hypothesis generation and future functional validation.

To disentangle the effects of age and strain from dietary influences (Figure 1 C), we performed separate species-level analyses using only the CD samples from the MG dataset. Using the same MaAsLin3 method, we identified several species that were significantly associated with age, such as *Duncaniella_dubosii*, which not only decreased under HF but also showed a marked decline with aging (Figure 2 F). Additionally, to investigate strain-specific microbial patterns, we used C57BL/6 (B6) mice as the control group and included data from the 5 BXD strains with more than 10 measured individuals. Several species showed notable variation across different strains, highlighting genotype-specific microbial patterns that may reflect underlying host genetic influences on gut microbiota composition (Figure 2 G). Notably, *GGB27849_SGB40283* exhibited elevated abundance in the BXD32, while *GGB28815_SGB41470* was more enriched in BXD100 compared to other strains.

We next focused on functional profiling with HUMAnN3 using the same MG and MT datasets to quantify microbial pathway abundance and coverage across all samples. Prior studies have shown that MT-derived data, which capture microbial activities, are often more informative for linking microbial functions to host phenotypes than MG-derived data, which reflect functional genetic potentials [38]. Nevertheless, MG profiling serves as a valuable reference to account for potential biases introduced during MT data generation and interpretation. A total of 447 MetaCyc pathways were detected in the MG and 391 in the MT dataset. Among the 33 most abundant pathways across both datasets, 17 pathways in MG and 10 in MT showed significant differences between dietary groups (t-test, $p < 0.05$), with 8 pathways overlapping significantly between both datasets (Figure 3 A).

In addition to these dominant pathways, several low-abundance pathways also exhibited significant associations with either diet or age, highlighting potentially subtle but biologically

relevant functional shifts. For example, the peptidoglycan biosynthesis III (mycobacteria) pathway, associated with bacterial cell wall formation and host immune stimulation, was significantly upregulated in the HF group at both the MG and MT levels (Figure 3 B). The components of this pathway are generated by a broad range of gut bacterial genera, including *Bacteroides*, *Clostridium*, *Lachnospiraceae_unclassified*, *Helicobacter*, and *Lactococcus*. In both datasets, the HF group exhibited higher total abundance of this pathway, indicating a potential upregulation of bacterial cell wall synthesis.

To comprehensively assess the associations between all pathways and host variables such as diet and age, we further performed detailed statistical comparisons using MaAsLin3, analyzing pathway abundance in MG and MT datasets (Figure 3 C-D) and pathway prevalence in the corresponding datasets (Figure S4 A-B) across different dietary groups. Overall, the pathway abundance analysis identified 210 diet-associated pathways in MG and 110 in MT, while the pathway prevalence analysis revealed 78 significant pathways in MG and 39 in MT. Of these, 91 pathways were shared between MG and MT in the abundance results, and 15 were shared in the prevalence results. Some pathways were primarily driven by specific species, such as “lactose and galactose degradation I” (Figure S4 C), whereas others, such as “4-deoxy-L-threo-hex-4-enopyranuronate degradation” (Figure S4 D), appeared to be largely attributable to unclassified species, making mechanistic investigation inherently more challenging. Notably, in both MG and MT, the “lactose and galactose degradation I” pathway was significantly upregulated in the HF, primarily driven by the marked increase of *s__Lactococcus lactis* under HF conditions. In contrast, other studies have reported a downregulation of this pathway in HF-fed mice, which was largely attributable to a significant decrease in *s__Lactobacillus murinus* [39]. Given that the species driving a pathway may also contribute to multiple functional modules, these results should be interpreted cautiously and are best considered hypothesis-generating. Further experimental validation will be required to delineate the precise mechanisms. By contrast, the downregulation of the “4-deoxy-L-threo-hex-4-enopyranuronate degradation” pathway under HF conditions was consistent with previous studies [40]. In addition, several pathways demonstrated robust results across both MG and MT (e.g., “superpathway of (R,R)-butanediol biosynthesis” and “UTP and CTP dephosphorylation I”), although no related studies have been reported to date. Together, these results provide complementary insights into microbial functional alterations associated with dietary change.

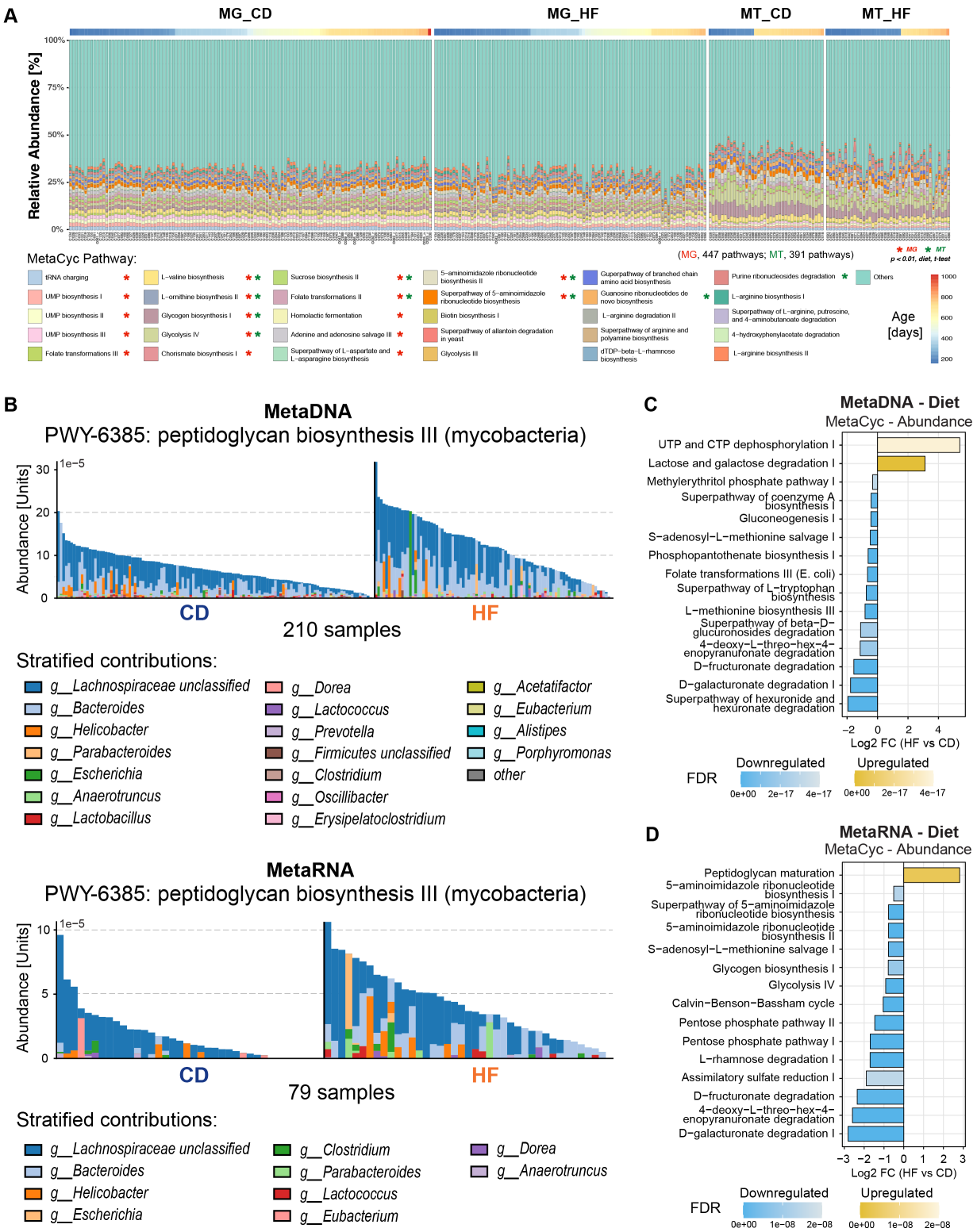

Figure 3. Diet-associated Changes in Microbial Functional Profiles.

(**A**) Stacked bar plots of the relative abundance of the 33 most abundant MetaCyc pathways in the MG (left) and MT (right) datasets, ranked by diet-associated t-test significance. ⋆ indicates pathways with p-values < 0.01. (**B**) Dietary differences in the peptidoglycan biosynthesis III pathway based on MG (top) and MT (bottom) data. Contributions from "unclassified" bacterial genera were excluded from both panels. (**C-D**) Top 15 MetaCyc pathways with significant dietary differences identified by MaAsLin3. Abundance from MG (**C**) and MT (**D**) between CD and HF diets are shown. Bars indicate $\log_2$FC (HF vs CD).

Beyond diet, age also influenced microbial functional profiles (Figure S5 A-B). In the pathway abundance analysis (Figure S5 A), the "L-histidine biosynthesis" and "L-lysine biosynthesis III" pathways were negatively associated with aging, consistent with observations from studies across different human age groups [41]. Conversely, the "fucose and rhamnose degradation" superpathway showed a significant positive association with aging in the pathway prevalence analysis (Figure S5 B), aligning with previous findings of its marked upregulation in older adults with sarcopenia [42]. Taken together, these results highlight how diet and age jointly modulate the gut microbial landscape at both taxonomic and functional levels, and illustrate the utility of integrating MG and MT datasets to capture microbial metabolic outputs.

Machine Learning-based Cecal Age Modeling to Identify Microbial Biomarkers Associated with Aging

Beyond conventional statistical analyses of taxonomic and functional profiles described in the previous section, we further employed supervised machine learning to model and predict host outcomes based on gut microbiome composition, starting with chronological age. This approach allows us to first evaluate whether microbial profiles contain sufficient predictive information about host aging, and then to identify key microbial species that contribute most to age classification. By training and validating multiple classifiers, we aimed to complement previous age-associated analyses and prioritize robust microbial biomarkers that may not be easily identified by traditional univariate methods. Given that ANOVA results (Figure 1 C) indicated a much stronger influence of diet than age on microbial composition in the MG dataset, we trained age-prediction models separately within each dietary group. Moreover, the age distribution within our cohort was not continuous (Figure S5 C), making classification more suitable than regression for downstream modeling. To ensure balanced group sizes, we defined samples younger than 350 days as the "Young" group and those older than 450 days as the "Old" group. After filtering, a total of 100 samples (Young n = 40, Old n = 60) were retained for modeling.

We trained four supervised learning models including Random Forest (RF), XGBoost (XGB), logistic regression (Logistic), and deep neural network (DNN) using 5-fold cross-validation to evaluate their ability to classify age groups based on MG species-level data. Among these, the RF classifier achieved the highest performance with an average area under the curve (AUC) of 0.84. To ensure that these selected features reflect biological aging rather than technical artifacts, such as cage- or facility-associated contamination during the extended sample collection period,

we generated a combined plot of the relative abundances of the top 20 RF-ranked species across all samples. No evidence of batch-related shifts was observed (Figure S5 D), supporting their relevance to aging rather than environmental confounders. These overlapping features may serve as prioritized candidates for generating aging-related hypotheses and for future experimental validation. In our initial test of age prediction, we achieved significant predictive capacity (AUC 0.84) within the training dataset (Figure 4 A, left panel). To further assess the generalizability of these models, we conducted an independent batch-wise evaluation in which one MG data batch was used for training and the other for testing. Both RF and XGB models achieved AUCs exceeding 0.80 in this setting (Figure 4 A, right panel), supporting their robustness across biological replicates. To improve this, we incorporated feature interaction (FI) terms to enhance model expressiveness when trained on smaller datasets. Inclusion of interacting features yielded modest improvements in prediction accuracy, suggesting potential benefits (e.g. AUC improved from 0.80 to 0.88 with RF) in capturing non-linear relationships between microbial features and host age. In parallel, we also trained age classification models using MG data from the HF group and MT data from the CD group (Figure S6 A–B). In both cases, the RF classifier outperformed other models, with cross-validated AUCs exceeding 0.80. Even with the smaller sample size in the MT dataset, the AUC of age prediction remained similar with RF (0.81 in both cases).

Building on the strong performance of the classification model for age, we next ranked all microbial features by their importance scores derived from the trained classifiers. To evaluate the contribution of top-ranked features, we retrained 5-fold cross-validation models using stepwise subsets of the top 50 ranked species, increasing the feature set in increments of five. The RF model achieved an AUC of 0.85 using only the top 5 features, and exceeded 0.90 when using the top 20 features. No improvement was observed with larger numbers of features selected, indicating that a small subset of microbial taxa carries sufficient information for accurate age prediction (Figure 4 B). XGB models showed similar trends, although their predictive performance was slightly lower than that of RF. By ranking features according to the sum of SHAP value magnitudes across all samples, we found that 10 species overlapped among the top 20 features identified by both models (Figure 4 C), highlighting their robustness as age-associated candidates, e.g. *GGB27866_SGB40300*, *Streptococcus danieliae*, *Paramuribaculum intestinale*, and *Duncaniella dubosii*.

To compare microbial features identified by different analytical strategies, we next examined the overlap between statistical significance from MaAsLin3 linear models and feature importance

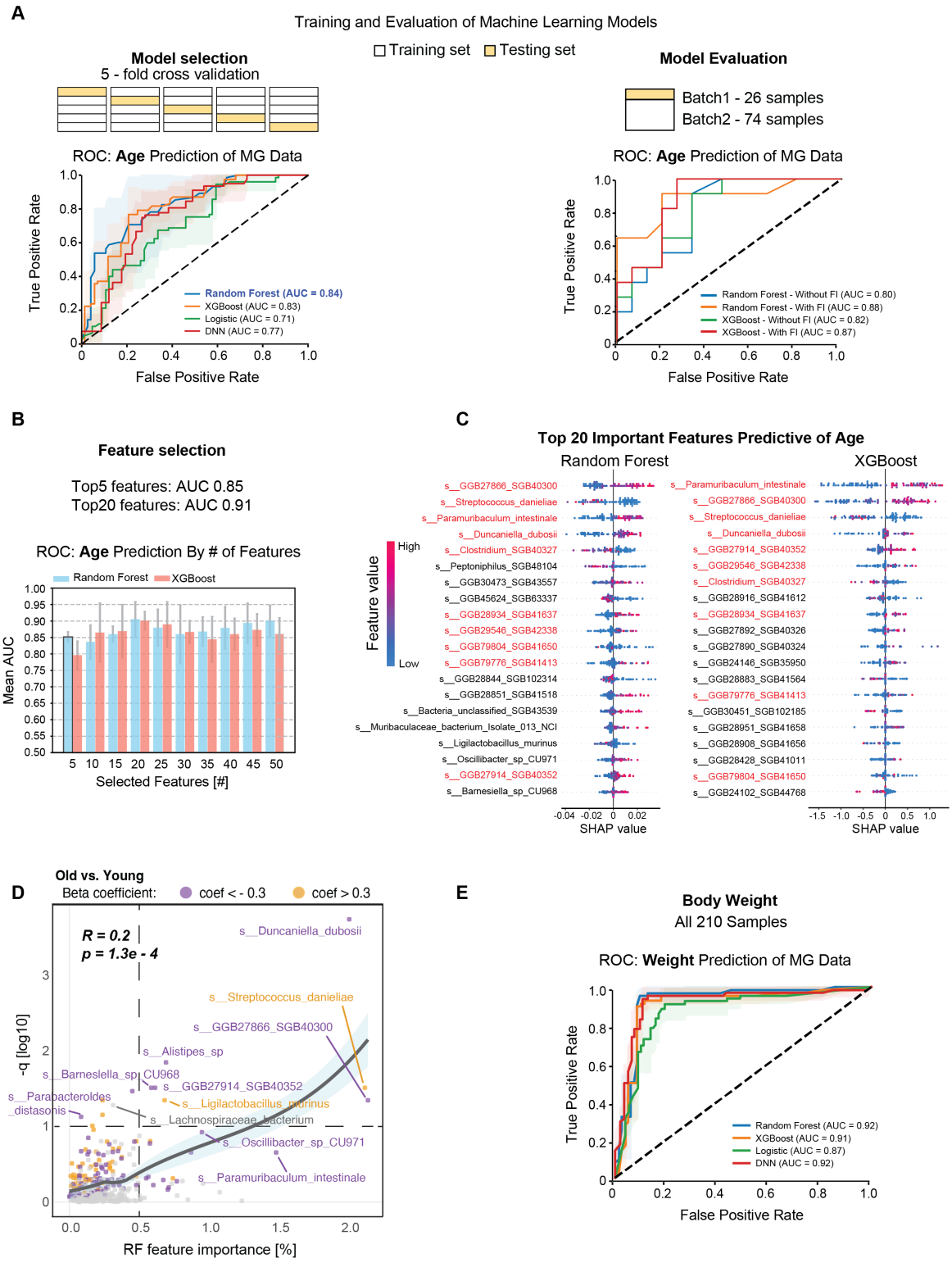

Figure 4. Machine Learning Models Identify Microbial Species Predictive of Host Age.

(**A**) Five-fold cross-validation for model training and evaluation of age prediction using MG data. Left: RF achieved the highest AUC (0.84) among tested models. Right: Evaluation on two independent batches demonstrated consistent model performance. (**B**) AUC values of models retrained using different numbers of top-ranked features with 5-fold cross-validation. (**C**) Feature importance was calculated using RF and XGB models. (**D**) Comparison of species associated with age identified by MaAsLin3 and RF. Each point represents one species. The scatter plot shows a positive correlation between RF feature importance and MaAsLin3 statistical significance. (**E**) MG data used to train the body weight prediction models have shown high accuracy.

scores from RF classifiers. While there was substantial variation between the two approaches, a subset of species was consistently highlighted by both methods (Figure 4 D). A moderate but significant positive correlation was observed between model-derived feature importance and the negative log-transformed q-values from MaAsLin3 (Spearman: *Rho* = 0.2, *p* = 1.3e-4), suggesting that species strongly associated with age are more likely to be prioritized across analytical frameworks. Of particular note, *Streptococcus danieliae* exhibited the strongest positive correlation with aging, whereas *Duncaniella dubosii* showed the strongest negative association. To further explore the phylogenetic distribution of age- and diet- associated microbial features, we annotated a taxonomic tree of all detected species with significance results from the MaAsLin3 analysis (Figure S6 C). This visualization enabled us to examine whether microbes responsive to host variables were clustered within specific taxonomic clades. However, no clear aggregation patterns were observed, suggesting that microbial responses to diet and aging are shaped by complex host–microbe interactions rather than being driven by a single functional pathway or metabolite class. Interestingly, several species highlighted by both machine learning methods and traditional bioinformatic approaches exhibited low mean abundance, reinforcing the importance of using complementary analytical strategies to identify robust features. This integrative approach allowed us to uncover microbial signatures that may otherwise be overlooked — such as low-abundance bacteria with only sequential identifiers rather than formal species names — and offers a framework for identifying and selecting candidates for mechanistic experiments.

To further evaluate the reliability and biological relevance of our modeling framework, we applied the same machine learning pipeline to predict mouse body weight using MG data (Figure 4 E). Given that diet exerts a much stronger effect than age in the MG dataset and is directly associated with body weight, the body weight prediction model — as expected — performed exceptionally well (RF: AUC = 0.92). We also compared the top 20 features ranked by RF and XGB models (Figure S6 D), and observed a high degree of agreement, with 10 overlapping species — equivalent to what was observed in the age prediction models. This result not only highlights the predictive capacity of microbiome-based features for host metabolic traits but also indirectly supports the robustness of our age prediction models, and that the approach can be applied to both dependent (body weight) or independent (age) variables.

Integrated Multi-omics Analyses Reveal Host-Microbiome Interactions Linked to Metabolism and Aging

Having performed a series of single-omics analyses on individual datasets in the previous sections, we identified multiple microbial and host features associated with diet, age, and strain. To further leverage the multi-omics datasets generated from the same BXD mouse cohort [17], we conducted integrative analyses across microbiome, cecum, and liver datasets. This enabled us to explore host–microbiome interactions more comprehensively and evaluate the added value of multi-omics integration in predicting biological traits and identifying functionally relevant microbial associations. We first examined the relationship between gut microbial composition and host metabolic gene expression by integrating MG species-level profiles with liver transcriptomic data. In particular, we tested the expression of *Fasn* (fatty acid synthase), a key enzyme in hepatic lipogenesis, in relation to the F/B ratio. In the CD group, *Fasn* expression showed a significant positive correlation with the F/B ratio (r = 0.52, p = 0.001), whereas no such association was observed in the HF group (r = 0.26, p = 0.16) (Figure 5 A). These findings suggest that the relationship between microbial composition and hepatic lipid metabolism may be modulated by dietary context, with stronger microbiota–host interactions under CD conditions.

To further assess whether integrating multi-omics data could further enhance the performance of microbiome-based age classifiers, we focused on two datasets — MG species-level data and liver proteomics (LP) — which had the largest number of overlapping samples (n = 139) within our multi-omics collection. As a first step, we evaluated the predictive capacity of LP data alone. Using the same classification framework applied to the MG dataset, we trained classifiers on LP data from CD mice, binarized into Young (< 350 days) and Old (> 450 days) age groups. The RF model achieved an average AUC of 0.81 in 5-fold cross-validation (Figure S7 A, left panel). Notably, several models reached AUCs of approximately 0.85 using only the top five features, indicating that a small subset of liver proteins was sufficient for robust age prediction (Figure S7 A, right panel). Consequently, we next selected the overlapping samples between the MG and LP datasets and stratified them by age, retaining 61 samples that fell into distinct Young (< 350 days, n = 25) or Old (> 450 days, n = 36) categories. These samples were then used to independently train age-prediction classifiers on each dataset (Figure 5 B). When using all available features from each dataset, model performance was modest (MG: AUC = 0.59, LP: AUC = 0.76), likely due to the small sample size relative to high feature dimensionality. To address this, we applied prior feature selection based on the top-performing models from earlier analyses,

retaining only the top 10 features from each dataset. This reduction led to a marked improvement in performance for both MG (AUC = 0.78) and LP (AUC = 0.92), demonstrating the utility of targeted feature selection in small-sample, high-dimensional settings. Finally, we combined the top five MG features with the top five LP features to train an integrated classifier. This resulted in a compact 10-feature model, allowing direct comparison with the earlier models trained using the top 10 features from each individual dataset. This minimal multi-omics model achieved a striking AUC of 0.95, suggesting that low-dimensional integration of key features across omics layers can substantially improve predictive performance. These results highlight the potential of compact, cross-modal biomarker panels to overcome limitations in multi-omics sample overlap, enabling more flexible study designs. The most predictive microbial and proteomic features identified here also offer promising targets for future mechanistic studies into host–microbiome interactions during aging.

Given the central role of the gut–liver axis in host metabolic and immune regulation, we next explored potential interactions between microbial and hepatic protein features identified through our machine learning models. Specifically, we used the HAllA framework [43] to assess associations between the top 30 LP features and the top 20 MG features. The number of LP features was expanded relative to MG to reflect the greater feature dimensionality in the proteomic dataset, with the top 10 ranked associations highlighted (Figure 5 C). We observed that *Streptococcus danieliae*, one of the most predictive MG features for age, exhibited strong positive correlations with three liver proteins — *Por*, *Gsta2*, and *Cyp4a10*. These cross-omics associations provide preliminary evidence for microbiome–host molecular crosstalk and highlight candidate gut–liver pathways for future mechanistic investigation.

Lastly, to further investigate cecum–microbiome interactions at the systems level, we constructed cross-omics co-expression networks (Figure S7 B-C, Figure 5 D) using Weighted Gene Correlation Network Analysis (WGCNA) based on cecal mRNA and metagenomic species data. This analysis identified several modules containing tightly correlated host genes and microbial taxa. For example, *Ido1* expression in the cecum strongly associated with a group of HF-enriched species in the family Lachnospiraceae, such as *Lachnospiraceaebacterium M18-1*. Previous studies have shown that *Ido1* knockout in mice alters gut microbial composition, and that mice lacking *Ido1* exhibit significantly reduced white adipose tissue when fed a HF diet [44]. These findings support the intriguing possibility that, under HF conditions, *Ido1* may influence host metabolic outcomes through its interactions with specific *Lachnospiraceae*

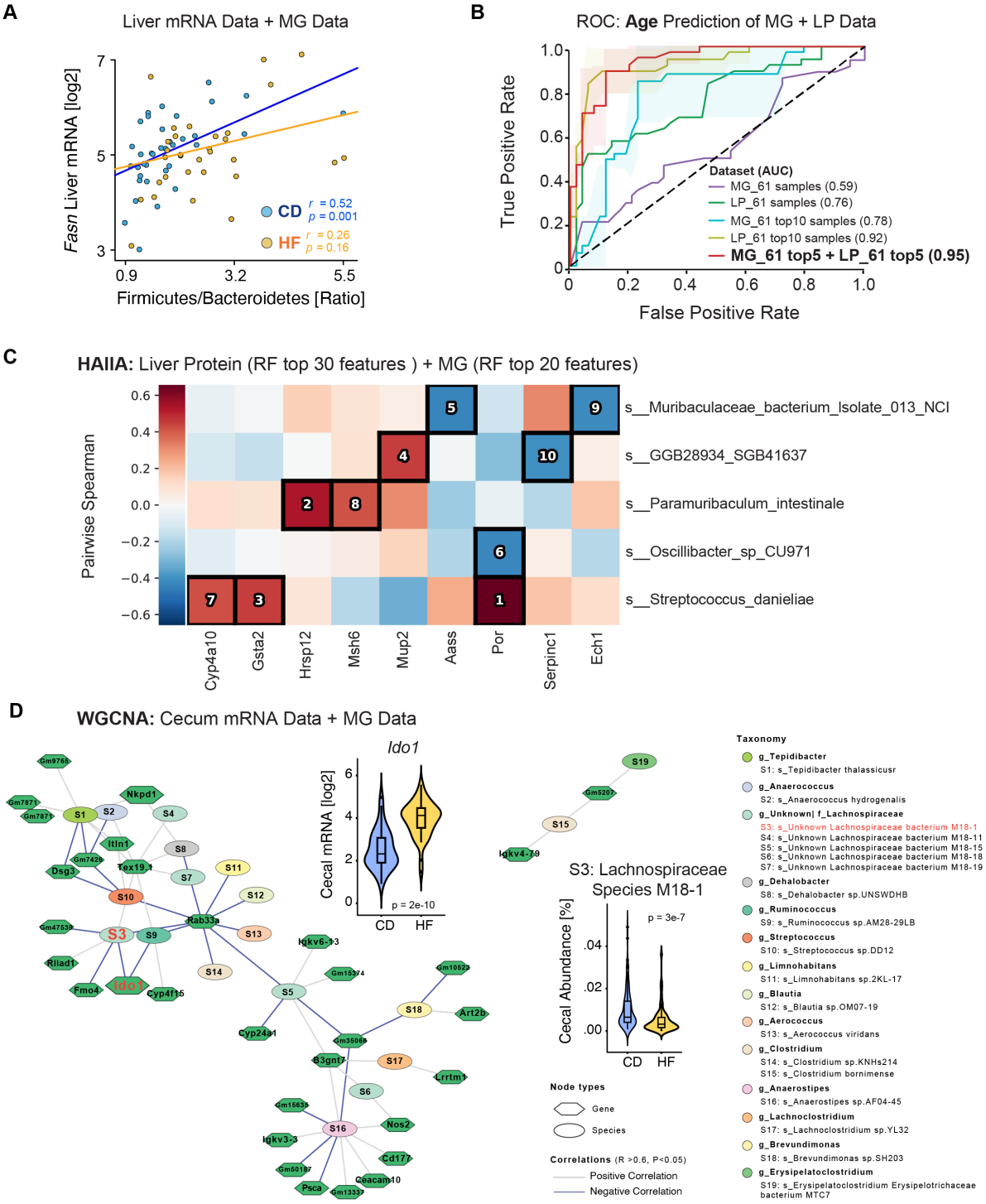

Figure 5. Integrated Multi-omics Analysis Reveals Age Prediction Performance and Microbial-host Interactions.

(**A**) Correlation between *Firmicutes / Bacteroidetes* ratio and liver mRNA expression of *Fasn* stratified by diet group. (**B**) ROC curves for age prediction models based on MG and liver proteomic (LP) datasets, showing improved performance with integrated top features (AUCs up to 0.95). (**C**) Pairwise Spearman correlation heatmap between liver proteins and MG species. Black boxes mark significant correlations, with numbers showing their significance ranks. (**D**) Network visualization of host–microbiome correlations. Nodes represent either genes (hexagons) or species (ovals); edges indicate significant correlations (Rho > |0.6|, p < 0.05). Violin plots show the expression of *Ido1* and relative abundance of *Lachnospiraceae_species_M18-1*.

species. Consequently, this integrated network approach allowed us to link multi-layered biological signals across host and microbial compartments, highlighting candidate gene–microbe relationships that warrant further mechanistic exploration. Together, these results underscore the value of combining statistical modeling, machine learning, and network-based strategies to extract meaningful biological insights from multi-omics data. They also establish a foundation for future hypothesis-driven studies aimed at dissecting causal microbiome–host interactions in the context of diet and aging.

Discussion

Comprehensive studies analyzing the gut microbiota are showing new methodological approaches and diagnostic insights that researchers and clinicians can use to diagnose individuals' current health concerns, and for prognosticating their future metabolic health. Furthermore, as with blood and urine, gut microbiome samples can be collected in the clinic longitudinally and minimally invasive. In this study, we have focused on the cecum rather than fecal samples due both to the fasting state of the aging mice, the larger sample size compared to fecal pellets, to maintain timing consistency with the previously-collected liver samples, and the ability to ensure consistent sample collection across an aging colony of several years. Over the past decade, rapid improvements in sequencing quality and orders of magnitude of cost decreases have opened the door for large-scale studies on the composition and activity of the microbiome. These technological gains now allow for studies, like this one, which can accurately identify and quantify hundreds to thousands of taxa reproducibly and accurately across hundreds of individuals. Here, we generated and systematically integrated MG, MT, and host cecal mRNA data from a genetically heterogeneous mouse cohort to investigate how diet, aging, and genetic background collectively shape gut microbial communities and host molecular responses. We first showed that age-associated transcriptional changes in the ceca were relatively modest under long-term CD but became substantially more pronounced under HF conditions, highlighting a strong diet-by-age interaction at the host transcriptome level. At the taxonomic level, we quantified 614 microbial species in the MG dataset. Using supervised machine learning, we demonstrated that these species-level profiles can accurately predict key host phenotypes: body weight (AUC = 0.92) and age (AUC = 0.84), underscoring the predictive capacity of the gut microbiome. Furthermore, through integrative modeling and prior feature selection between this dataset and our recently-collected liver transcriptome and proteome in the same mice, we were able to perform cross-tissue multi-omics analyses to obtain more reliable results with fewer selected features. Given the dynamic responsiveness of the microbiome to environmental shifts and the convenience of sampling, this integrative approach offers powerful predictive capabilities for individualized assessments of future health trajectories, reflecting genetic and environmental contexts.

Through an integrated analytical approach utilizing both MaAsLin3 and RF modeling, several microbial species consistently associated with aging emerged, notably *Streptococcus danieliae* and *Duncaniella dubosii*. *S. danieliae*, identified as strongly positively correlated with age in

our analysis, aligns with prior findings from fecal 16S rRNA sequencing studies indicating that increased abundance of the *Streptococcus* genera accelerates aging [45]. These studies suggest that certain *Streptococcus* species could causally promote cellular aging, such as by elevating oxidative stress levels. Our findings raise the possibility that *S. danieliae* may represent a species-level candidate contributing to the observed age-associated microbial shifts. Conversely, *D. dubosii* showed a significant negative correlation with aging, highlighting its potential protective role. Prior studies have linked microbial modulation of tryptophan metabolism to longevity and age-related disease resistance [46], and intriguingly, the survival benefits conferred by tryptophan were found to be microbiota-dependent. While its precise role in aging remains to be determined, our findings nominate *D. dubosii* as a candidate taxon for further investigation into microbiome-based mechanisms supporting healthy aging.

Our network analysis using WGCNA identified significant associations between cecal expression of *Ido1* and specific Lachnospiraceae, particularly *Lachnospiraceae bacterium M18-1*. Notably, under HF conditions, we observed a significant increase in *Ido1* expression alongside a marked reduction in the abundance of several Lachnospiraceae species. This inverse pattern was reflected in a negative correlation within the co-expression network. Although no prior studies have directly examined interactions between *Ido1* and individual species within the Lachnospiraceae family-level taxon, existing literature provides a plausible mechanistic link. Members of Lachnospiraceae are core constituents of the gut microbiota and are major producers of SCFAs such as butyrate [47]. SCFAs have been shown to inhibit the expression of *Stat1*, thereby downregulating IFNγ-induced transcription of *Ido1* [48]. Our findings are consistent with this model, suggesting that a HF-induced decline in SCFA-producing microbes may relieve inhibitory control over *Ido1*, leading to its upregulation. Furthermore, *Ido1* knockout has been associated with altered microbiota composition under HF conditions [44], pointing to a bidirectional feedback between host tryptophan metabolism and microbial ecology. These observations underscore a potential regulatory axis between microbial SCFA production and host immune-metabolic responses via the *Ido1* pathway.

Our multi-omics predictive modeling, integrating the top five MG species and top five liver proteomic features, significantly enhanced predictive accuracy (AUC = 0.95), demonstrating the efficacy of compact multi-omics biomarker panels. Beyond improved predictive performance, this strategy effectively addresses the challenge posed by limited overlapping samples common in multi-omics studies, especially those with extended and heterogeneous sampling periods.

By initially identifying robust biomarkers within individual omic datasets, researchers can subsequently leverage smaller, overlapping sample subsets for integrative modeling. This approach not only enhances study flexibility and resilience but is also highly applicable to human studies, where sample heterogeneity and logistical constraints often limit integrative analyses [49].

Several limitations of our study should be acknowledged. Firstly, the sample size per strain restricted detailed analysis of the genetic impact across diverse BXD strains. Only 5 strains had more than 10 individuals analyzed, thus limiting insights into microbiome interactions unique to individual outlier strains. Secondly, key identified microbial taxa, such as *"GGB28815 SGB41470"*, remain poorly characterized and their cultivation conditions are unknown, complicating downstream functional validations through traditional microbiological methods. Further development in culturomics and advanced molecular techniques are required to overcome these barriers. Additionally, our study relied on cross-sectional sampling at discrete age points, precluding longitudinal analyses that could offer deeper insights into microbiome dynamics and causal inference.

In conclusion, our integrative multi-omics approach highlights the influence of dietary patterns and aging on the gut microbiome composition and functional landscape, identifying key microbial taxa and metabolic pathways intricately linked to host metabolic health. Our results further underscore the remarkable utility of microbiome profiling combined with machine learning for accurate host age prediction, underscoring its potential application as a reliable biomarker for biological aging and personalized health monitoring. These findings provide a robust foundation for future mechanistic explorations and translational applications aiming to harness microbiome insights for metabolic health interventions and age-related health assessments.

Methods

Mice Sample Collection and Processing

All animal handling and tissue collection were done in the scope of our prior study [17], but it is briefly recapitulated here. 662 animals from 58 strains of the BXD family were sacrificed at specific ages (7, 12, 18, or 24 months of age, plus or minus one month), of which cecal tissue was retained for 597 individuals from 56 strains. The aging colony was 95% females due to methodological challenges of aging large groups of males (e.g. fighting during regroupment). Diets were either Harlan Teklad 2018 (CD; 24% calories from protein, 18% from fat, 58% from carbohydrates) or Harlan Teklad 06414 (HF; 18.3% calories from protein, 60.3% from fat, 21.4% from carbohydrates). Animals were removed from the aging colony the night before sacrifice but retained access to food and water. Animals were anesthetized between 9am and noon with tribromoethanol and perfused with PBS. Cecal tissue containing fecal matter was harvested within about 7 minutes of perfusion, then immediately frozen in liquid nitrogen and stored in -80℃ freezers.

For the sample preprocessing, a subset of 232 of the above samples from 43 strains were selected, of which 225 were from female mice and the remaining 7 from males, similar to the sex distribution in the original aging colony. For these samples, the mixed cecum and cecal contents tissue samples were transferred from a -80℃ freezer to dry ice. Prior to grinding, the mortar and pestle was pre-cooled with liquid nitrogen. The tissues were ground using the pre-cooled mortar and pestle until fine powders were obtained. After the sublimation of the liquid nitrogen, the powder was transferred to new tubes using a round spatula scraper, all while remaining at around -196℃. The mixed cecum and cecum content powder was stored in a -80℃ freezer for subsequent DNA and RNA extraction.

Optimization of DNA extraction protocol

For DNA extraction, a modified extraction method employing the Wizard Genomic DNA Purification Kit was used to simultaneously extract DNA from both the host and the gut bacteria in mixed samples of cecum and feces. ~30 mg of tissue powder was suspended in 480 µL of 50 mM EDTA and homogenized with stainless steel milling beads using an Oscillating Mill MM400 (Retsch) at 25 Hz for 1 min, then incubated at 37℃ for 60 min. After the homogenates were centrifuged for 2 min at 500 g, the supernatant was transferred into a fresh tube. This

supernatant was then centrifuged for 2 min at 16,000 g. After centrifugation, the supernatant was removed and the pellets were collected. The pellets were lysed by adding 600 µL of chilled Nuclei Lysis Solution, gently mixed by pipetting, incubated at 80℃ for 5 min, and subsequently cooled to room temperature. Next, 3 µL of RNase Solution was added, and the mixture was incubated at 37℃ for 45 min before cooling to room temperature. Protein precipitation involved adding 200 µL of Protein Precipitation Solution (from the same Wizard kit), followed by vortexing and a 5 min incubation on ice. The sample was then centrifuged at 13,000 × g for 3 min. Subsequently, the supernatant was transferred to a fresh tube with 600 µL of isopropanol for DNA precipitation and mixed gently by inversion. This was followed by a centrifugation at 16,000 × g for 2 min to remove the supernatant. Next, 600 µL of 70% ethanol was added, mixed, and the sample was centrifuged again for 2 min at 16,000 g. Ethanol was removed by aspiration, and the pellet was air-dried completely, which took roughly 15 min. The DNA pellet was resuspended in 100 µL of Rehydration Solution for 1 hour at 65℃. For DNA purification, the Wizard DNA Clean-up System was utilized following the manufacturer's protocol.

For DNA quality assessment, the concentration and purity of each DNA sample were measured using a spectrophotometer (Nanodrop 2000C, Thermo Scientific). We established a method for evaluating the proportion of gut bacterial DNA using RT-qPCR with the ratio of *Gapdh* primers (Forward: CAAGGAGACCTCAAGGTCAG; reverse: GTATGCACCTCACAACGCCATG) for the eukaryotic content compared to the universal prokaryotic 16S rRNA primers (Eurogentec, Forward: AAACTCAAAKGAATTGACGG; reverse: CTCACRRCACGAGCTGAC). Host DNA and intestinal bacterial DNA were extracted separately from liver and true fecal samples (i.e. using excreted feces, rather than cecal content) and then diluted to a concentration of 10 ng/µL. Subsequently, different ratios of host DNA and gut microbiota DNA were mixed to serve as controls. The concentration of all samples was adjusted to 10 ng/µL for use in RT-qPCR assays. RT-qPCR was performed using the LightCycler 480 Instrument II (Roche). Each reaction had 2.5 µL of cDNA, 2.5 µL of primer mixture (forward and backward primers) (2 µM) and 5 µL of the iQ SYBR Green Supermix (BIO-RAD). The PCR reaction conditions were as follows: an initial denaturation step at 95℃ for 10 min, followed by 40 cycles of denaturation at 95℃ for 15 seconds and annealing/extension at 60℃ for 1 min. The reaction was subsequently cooled to 20℃ for 1 hour. Based on the Crossing Point (Cp) values obtained from the RT-qPCR results, we defined a %Expression index to estimate gut microbiota-to-host DNA ratios and used it to filter samples with higher bacterial DNA content. The %Expression index was defined as

$$\Delta C_p(16S - \textit{Gapdh}) = (\text{16S_average_}C_p) - (\textit{Gapdh}\text{_average_}C_p)$$

$$\Delta\Delta C_p(\Delta C_p\text{_sample} - \Delta C_p\text{_6L4F}) = \Delta C_p(\text{16S - }\textit{Gapdh}\text{)_sample} + \Delta C_p(\text{16S - }\textit{Gapdh}\text{)_6L4F}$$

$$\%\text{Expression} = 100 \times 2^{-\Delta\Delta C_p(\Delta C_p\text{_sample} - \Delta C_p\text{_6L4F})}$$

Where 16S_average_Cp is the average Cp value from three technical replicates of each DNA sample extracted from mouse cecal contents, using universal prokaryotic 16S rRNA primers; and *Gapdh*_average_Cp is the average Cp value from the same samples using primers targeting the mouse reference gene *Gapdh*. The value of ΔC$_p$_6L4F refers to the Cp difference calculated from a reference DNA mixture composed of 40% fecal DNA and 60% liver DNA (mixed at equal concentrations), which serves as a normalization control across samples. Samples with a gut microbiota DNA proportion estimated to be less than 40%, based on the Cp values of the controls, were excluded from sequencing (Figure S1 A-B).

RNA Extraction Protocol

For RNA extraction, ~30 mg of tissue powder was again suspended in 1 mL chilled TRIzol (Thermo Scientific) and homogenized with stainless steel milling beads in an Oscillating Mill MM400 (Retsch) at 25 Hz for 30 seconds, followed by a phase separation using 200 µL of chloroform, and the mixture was vortexed before being left at room temperature for another 2 min. The mixture was centrifuged at 12,000 g for 15 min at 4℃, the upper aqueous phase was then transferred to a new RNase-free tube. For RNA precipitation, 400 µL of isopropanol was added and thoroughly mixed by pipetting, followed by incubated at room temperature for 2-5 min. This was followed by centrifugation at 12,000 g for 10 min at 4℃, with the supernatant was discarded. The RNA pellet was washed with 1 mL of 75% ethanol, centrifuged at 7,500 g for 5 min at 4℃, and the supernatant was subsequently discarded. Finally, the RNA pellet was air-dried at room temperature for 10 min before being resuspended in 50 µL of RNase-free water. For RNA purification, the RNeasy MinElute Cleanup Kit was utilized following the manufacturer's protocol (Qiagen, March 2016 version). For RNA quality assessment, the concentration and purity of each RNA sample were measured using a spectrophotometer (Nanodrop 2000C, Thermo Scientific). RNA quality was evaluated using an Agilent RNA 6000 Nano Kit and Agilent 2100 Bioanalyzer. Given that the RNA samples contain both host RNA and gut bacterial RNA, the RNA Integrity Number (RIN) is not applicable for these mixed samples. Only samples exhibiting four distinct peaks were selected for subsequent sequencing (Figure S1 C).

Sequencing and Profiling

Sequencing

The mRNA sequencing libraries were prepared using the TruSeq Stranded mRNA Library Prep Kit (Illumina). Sequencing was performed on the Illumina NextSeq 500 system, utilizing single-end sequencing with a read length of 75 base pairs. MT sequencing was conducted after library preparation using the Zymo-Seq RiboFree Total RNA Kit, and sequencing was carried out on the Illumina NextSeq 2000 system with 2×150 bp paired-end reads. MG sequencing used the Illumina HiSeq 1000, with roughly 25 million reads per sample, using 2×150 bp paired-end reads. Quality assessment of the MG and MT sequenced reads, along with their corresponding bases, was conducted using FastQC v0.12.1. In our case, all reads and bases were within an acceptable quality range, with Phred scores consistently exceeding 30, indicating high-quality sequencing data.

Cecal mRNA Data Processing

For gene expression analyses, the cecal mRNA sequencing data were aligned and quantified to generate per-gene unique read counts using the R package `Rsubread` v2.12.3 [50], resulting in a raw gene count matrix. This raw count matrix was used for differentially expressed gene (DEG) analysis in `DESeq2` v1.44.0 to identify transcriptional changes associated with diet and age, and the resulting $\log_2(\text{fold changes})$ were used for GSEA. In parallel, raw counts were also converted into a TPM (Transcripts Per Million) matrix using the package `tximport` v1.26.1 [51], followed by log transformation as $\log_2(\text{TPM}+1)$ for use in machine learning models and WGCNA. This dual processing strategy enabled robust DEG identification while facilitating pathway-level interpretation and comparison across samples.

MG and MT Data Processing

All raw paired-end MG and MT sequence reads were quality-processed using KneadData v0.10.0 [29]. Sequence trimming was performed using KneadData, which ran Trimmomatic v0.39 to trim regions where base quality fell below Q20 within a 4-nucleotide sliding window, and to remove reads that were truncated by more than 50% (parameters: SLIDINGWINDOW:4:20, MINLEN:50). Additionally, KneadData was used to run Tandem Repeat Finder (TRF) v4.09.1, to locate and remove reads characterized by tandem repeats. Bowtie2 v2.4.2, also employed within KneadData, was used to deplete host-derived sequences from the MG data using the mouse (C57BL/6) reference database, and from the MT data using both the mouse reference

database and the SILVA Ribosomal RNA reference database. After processing with KneadData, the remaining non-host reads were used as input to obtain the taxonomic profiles of the samples using MetaPhlAn v4.0.6 and the CHOCOPhlAn (mpa_vOct22_CHOCOPhlAnSGB_202212) database. Functional profiling was subsequently obtained through the HUMAnN v3.8 workflow.

Miscellaneous Bioinformatics and Statistical Analysis

All analyses were performed using R 4.2.2. For exploratory data visualization, differential expression analysis, statistical modeling, and multivariate analyses (e.g., ANOVA and PCA), we used standard R packages including `ggplot2` and `vegan`. To identify modules of co-varying host genes and microbial taxa, we performed WGCNA using the R package: `WGCNA`. Cross-omics edges were defined by Spearman correlation coefficients $|R| > 0.6$ and FDR-adjusted $p < 0.05$. Significant gene–microbe interactions were exported as edge lists and visualized using Cytoscape (v3.10.3) [52].

Multivariable and Cross-Omics Association Analysis

To identify microbial features associated with metadata variables such as diet, age, and strain, we applied MaAsLin3 (v1.10.0) [37], a multivariable linear modeling framework suitable for compositional microbiome data. Relative abundance tables (species- or genus- level) from MetaPhlAn4 [30] were used as input. We applied arcsine square root transformation to the abundance data and used diet, age, and strain as fixed effects. False discovery rate (FDR) was controlled using the Benjamini-Hochberg procedure, and features with FDR < 0.05 were considered statistically significant.

To explore cross-modal associations between microbial species and host liver proteomic features, we used the HAllA framework [43]. HAllA identifies statistically significant associations between high-dimensional datasets by combining hierarchical clustering and association testing across feature blocks. Spearman correlations were calculated for each species–protein pair, and the significance of associations was determined using a permutation-based FDR threshold of 0.05.

Supervised Learning and Model Evaluation

All prediction models were implemented using Python 3.9.18 with standard open-source libraries, including `pandas` (v2.1.4), `numpy` (v1.23.5), `scikit-learn` (v1.3.0), and `matplotlib` (v3.8.0).

We trained multiple supervised classifiers including RF, XGB, Logistic, and DNN to predict host phenotypes (e.g., age and body weight) based on gut microbial composition. Model performance was evaluated using five-fold cross-validation and ROC curves. Feature importance scores were extracted from tree-based models (RF and XGB) to rank the contribution of individual microbial species or protein features. To mitigate overfitting and improve interpretability, we applied iterative feature selection based on model-derived importance scores. To further capture potential nonlinear or combinatorial effects between microbial features, polynomial feature interaction terms were generated using the `PolynomialFeatures` module from `scikit-learn`. In multi-omics integration analyses, classifiers were trained using top-ranked features from each omics layer and evaluated using a compact, 10-feature integrated model. Model optimization was guided by cross-validation accuracy and AUC metrics across independent data batches.

Data availability

The raw sequence data are available in the NCBI Bioproject repository under the accession number PRJNA953833 (data will be released upon article publication). All processed data tables for the MG, MT, and cecal mRNA are available in eight supplemental tables (S1-S8).

Acknowledgements

We would like to thank Kurt Lamour for assistance with DNA sequencing. We are grateful to Xiaowei Zhan, Susheel Bhanu, and Julien Schleich for their valuable advice on data processing and analysis. We also thank Jack Haverty for his help during the early stages of the project. Rob Williams and Suheeta Roy from the University of Tennessee Health Science Center are acknowledged for hosting the aging mouse colony and their support in the organization of cecal sample collection. The mRNA and MT sequencing experiments presented in this paper were carried out at the LCSB sequencing platform (RRID: SCR_021931) at the University of Luxembourg. The processing of sequencing raw data presented in this paper were carried out using the HPC facilities of the University of Luxembourg [53].

Funding

This project is implemented with the support of the Fondation du Pélican de Mie et Pierre Hippert-Faber, under the aegis of the Fondation de Luxembourg. This project has been

Supported by the Luxembourg National Research Fund (PRIDE21/16749720/NEXTIMMUNE2).

Author Contributions Statement

E.W. and P.W. were responsible for the initial conceptualization and study design. E.W. supervised and provided guidance throughout the research process. Z.Z. and E.W. wrote the main manuscript text. Z.Z. prepared all figures and tables, with individual panel input and figure editing from E.W. Z.Z. optimized protocols for simultaneous extraction of host cecal and gut microbial DNA and RNA for sequencing. Z.Z. prepared all samples used in this manuscript. R.H. performed RNA sequencing. A.A., E.P., B.B., and T.K. assisted with various practical aspects of the experiments and exploratory analyses. All authors reviewed the manuscript.

Ethics declarations

Ethics approval and consent to participate

The mice from this study were approved by the University of Tennessee Health Science Center's committee for animal care and use, following AAALAC guidelines and with regular inspections by independent veterinarian oversight.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Code Availability

Codes and scripts developed in this study are all available at the GitHub repository (https://github.com/zzhougithub/GEM_Microbiome).

Supplemental Information

Supplemental Information contains seven figures and eight supplemental tables with the metadata for the cecal samples, and the sequence result data for the MG, MT, and cecal mRNA.

References

1. Benjamin S. Beresford-Jones, Samuel C. Forster, Mark D. Stares, George Notley, Elisa Viciani, Hilary P. Browne, Daniel J. Boehmler, Amelia T. Soderholm, Nitin Kumar, Kevin Vervier, Justin R. Cross, Alexandre Almeida, Trevor D. Lawley, and Virginia A. Pedicord. The Mouse Gastrointestinal Bacteria Catalogue enables translation between the mouse and human gut microbiotas via functional mapping. *Cell Host & Microbe*, 30(1):124–138.e8, January 2022. ISSN 19313128. doi: 10.1016/j.chom.2021.12.003.
2. Juliana Durack and Susan V. Lynch. The gut microbiome: Relationships with disease and opportunities for therapy. *Journal of Experimental Medicine*, 216(1):20–40, October 2018. ISSN 1540-9538. doi: 10.1084/jem.20180448.
3. Sylvie Rabot, Mathieu Membrez, Aurélia Bruneau, Philippe Gérard, Taoufiq Harach, Mireille Moser, Frederic Raymond, Robert Mansourian, and Chieh J. Chou. Germ-free c57bl/6j mice are resistant to high-fat-diet-induced insulin resistance and have altered cholesterol metabolism. *The FASEB Journal*, 24(12):4948–4959, December 2010. ISSN 1530-6860. doi: 10.1096/fj.10-164921.
4. June L. Round and Sarkis K. Mazmanian. The gut microbiota shapes intestinal immune responses during health and disease. *Nature Reviews Immunology*, 9(5):313–323, May 2009. ISSN 1474-1741. doi: 10.1038/nri2515.
5. Elizabeth A. Kennedy, Katherine Y. King, and Megan T. Baldridge. Mouse Microbiota Models: Comparing Germ-Free Mice and Antibiotics Treatment as Tools for Modifying Gut Bacteria. *Frontiers in Physiology*, 9, October 2018. ISSN 1664-042X. doi: 10.3389/fphys.2018.01534. Publisher: Frontiers Media SA.
6. Maria Elisa Caetano-Silva, Akriti Shrestha, Audrey F. Duff, Danica Kontic, Patricia C. Brewster, Mikaela C. Kasperek, Chia-Hao Lin, Derek A. Wainwright, Diego Hernandez-Saavedra, Jeffrey A. Woods, Michael T. Bailey, Thomas W. Buford, and Jacob M. Allen. Aging amplifies a gut microbiota immunogenic signature linked to heightened inflammation. *Aging Cell*, 23(8), August 2024. ISSN 1474-9718, 1474-9726. doi: 10.1111/acel.14190. Publisher: Wiley.
7. Tarini Shankar Ghosh, Fergus Shanahan, and Paul W. O'Toole. The gut microbiome as a modulator of healthy ageing. *Nature Reviews Gastroenterology & Hepatology*, 19(9):565–584, September 2022. ISSN 1759-5045, 1759-5053. doi: 10.1038/s41575-022-00605-x.
8. Ara Koh, Filipe De Vadder, Petia Kovatcheva-Datchary, and Fredrik Bäckhed. From Dietary Fiber to Host Physiology: Short-Chain Fatty Acids as Key Bacterial Metabolites. *Cell*, 165(6):1332–1345, June 2016. ISSN 00928674. doi: 10.1016/j.cell.2016.05.041.
9. Parasuraman Padmanabhan, Johannes Grosse, Abu Bakar Md Ali Asad, George K Radda, and Xavier Golay. Gastrointestinal transit measurements in mice with 99mTc-DTPA-labeled activated charcoal using NanoSPECT-CT. *EJNMMI Research*, 3(1):60, 2013. ISSN 2191-219X. doi: 10.1186/2191-219x-3-60. Publisher: Springer Science and Business Media LLC.
10. Giorgia Greter, Claudia Moresi, Stefanie Oswald, Alice De Wouters d'Oplinter, Daria Künzli, Elisa Cappio Barazzone, Jiayi Lan, Emma Slack, and Markus Arnoldini. Acute targeted induction of

gut-microbial metabolism affects host clock genes and nocturnal feeding, September 2024.

11. Maëva Meynier, Valentine Daugey, Geoffroy Mallaret, Sandie Gervason, Mathieu Meleine, Julie Barbier, Youssef Aissouni, Stéphane Lolignier, Mathilde Bonnet, Denis Ardid, Willem M. De Vos, Matthias Van Hul, Peter Suenaert, Amandine Brochot, Patrice D. Cani, and Frédéric A. Carvalho. Pasteurized *Akkermansia muciniphila* improves irritable bowel syndrome-like symptoms and related behavioral disorders in mice. *Gut Microbes*, 16(1):2298026, December 2024. ISSN 1949-0976, 1949-0984. doi: 10.1080/19490976.2023.2298026.
12. Colleen R. Kelly, Eugene F. Yen, Ari M. Grinspan, Stacy A. Kahn, Ashish Atreja, James D. Lewis, Thomas A. Moore, David T. Rubin, Alison M. Kim, Sonya Serra, Yanina Nersesova, Lydia Fredell, Dea Hunsicker, Daniel McDonald, Rob Knight, Jessica R. Allegretti, Joel Pekow, Imad Absah, Ronald Hsu, Jennifer Vincent, Sahil Khanna, Lyn Tangen, Carl V. Crawford, Mark C. Mattar, Lea Ann Chen, Monika Fischer, Razvan I. Arsenescu, Paul Feuerstadt, Jonathan Goldstein, David Kerman, Adam C. Ehrlich, Gary D. Wu, and Loren Laine. Fecal Microbiota Transplantation Is Highly Effective in Real-World Practice: Initial Results From the FMT National Registry. *Gastroenterology*, 160(1):183–192.e3, January 2021. ISSN 0016-5085. doi: 10.1053/j.gastro.2020.09.038. Publisher: Elsevier BV.
13. Dayoon Kwon, Keren Zhang, Kimberly C. Paul, Aline D. Folle, Irish Del Rosario, Jonathan P. Jacobs, Adrienne M. Keener, Jeff M. Bronstein, and Beate Ritz. Diet and the gut microbiome in patients with Parkinson's disease. *npj Parkinson's Disease*, 10(1):89, April 2024. ISSN 2373-8057. doi: 10.1038/s41531-024-00681-7.
14. Xiaoxiao Yuan, Ruirui Wang, Bing Han, ChengJun Sun, Ruimin Chen, Haiyan Wei, Linqi Chen, Hongwei Du, Guimei Li, Yu Yang, Xiaojuan Chen, Lanwei Cui, Zhenran Xu, Junfen Fu, Jin Wu, Wei Gu, Zhihong Chen, Xin Fang, Hongxiu Yang, Zhe Su, Jing Wu, Qiuyue Li, Miaoying Zhang, Yufeng Zhou, Lei Zhang, Guang Ji, and Feihong Luo. Functional and metabolic alterations of gut microbiota in children with new-onset type 1 diabetes. *Nature Communications*, 13(1):6356, October 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-33656-4.
15. Ge Yang, Jinlong Wei, Pinyi Liu, Qihe Zhang, Yuan Tian, Guowen Hou, Lingbin Meng, Ying Xin, and Xin Jiang. Role of the gut microbiota in type 2 diabetes and related diseases. *Metabolism*, 117: 154712, April 2021. ISSN 00260495. doi: 10.1016/j.metabol.2021.154712.
16. Zhang, Mocanu, Cai, Dang, Slater, Deehan, Walter, and Madsen. Impact of fecal microbiota transplantation on obesity and metabolic syndrome—a systematic review. *Nutrients*, 11(10):2291, September 2019. ISSN 2072-6643. doi: 10.3390/nu11102291.
17. Evan G. Williams, Niklas Pfister, Suheeta Roy, Cyril Statzer, Jack Haverty, Jesse Ingels, Casey Bohl, Moaraj Hasan, Jelena Čuklina, Peter Bühlmann, Nicola Zamboni, Lu Lu, Collin Y. Ewald, Robert W. Williams, and Ruedi Aebersold. Multiomic profiling of the liver across diets and age in a diverse mouse population. *Cell Systems*, 13(1):43–57.e6, January 2022. ISSN 2405-4712. doi: 10.1016/j.cels.2021.09.005.
18. S. Roy, M.B. Sleiman, P. Jha, J.F. Ingels, C.J. Chapman, M.S. McCarty, J.D. Ziebarth, M. Hook, A. Sun, W. Zhao, J. Huang, S.M. Neuner, L.A. Wilmott, T.M. Shapaker, A.G. Centeno, D. G. Ashbrook, M. K. Mulligan, C.C. Kaczorowski, L. Makowski, Y. Cui, R.W. Read, R.A. Miller, K. Mozhui, E.G.

Williams, S. Sen, L. Lu, J. Auwerx, and R.W. Williams. Gene-by-environmental modulation of lifespan and weight gain in the murine bxd family. *Nature Metabolism*, 2021.

19. David G. Ashbrook, Danny Arends, Pjotr Prins, Megan K. Mulligan, Suheeta Roy, Evan G. Williams, Cathleen M. Lutz, Alicia Valenzuela, Casey J. Bohl, Jesse F. Ingels, Melinda S. McCarty, Arthur G. Centeno, Reinmar Hager, Johan Auwerx, Lu Lu, and Robert W. Williams. A platform for experimental precision medicine: The extended bxd mouse family. *Cell Systems*, 12(3):235–247.e9, March 2021. ISSN 2405-4712. doi: 10.1016/j.cels.2020.12.002.
20. Chanhee Kang, Qikai Xu, Timothy D. Martin, Mamie Z. Li, Marco Demaria, Liviu Aron, Tao Lu, Bruce A. Yankner, Judith Campisi, and Stephen J. Elledge. The DNA damage response induces inflammation and senescence by inhibiting autophagy of GATA4. *Science*, 349(6255):aaa5612, September 2015. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.aaa5612.
21. Juri Kazakevych, Elena Stoyanova, Anke Liebert, and Patrick Varga-Weisz. Transcriptome analysis identifies a robust gene expression program in the mouse intestinal epithelium on aging. *Scientific Reports*, 9(1):10410, July 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-46966-3.
22. Jose Martinez-Lomeli, Poonamjot Deol, Jonathan R Deans, Tao Jiang, Paul Ruegger, James Borneman, and Frances M. Sladek. Impact of Various High Fat Diets on Gene Expression and the Microbiome Across the Mouse Intestines, October 2023.
23. Evert M. Van Schothorst, Nicole Franssen-van Hal, Mirjam M. Schaap, Jeroen Pennings, Barbara Hoebee, and Jaap Keijer. Adipose Gene Expression Patterns of Weight Gain Suggest Counteracting Steroid Hormone Synthesis. *Obesity Research*, 13(6):1031–1041, June 2005. ISSN 1071-7323, 1550-8528. doi: 10.1038/oby.2005.121. Publisher: Wiley.
24. Zheng-Mei Lv, Chao Liu, Ping Wang, and Yuan-Hua Chen. Dysregulation of mitochondrial dynamics and mitophagy are involved in high-fat diet-induced steroidogenesis inhibition. *Journal of Lipid Research*, 65(10):100639, October 2024. ISSN 0022-2275. doi: 10.1016/j.jlr.2024.100639. Publisher: Elsevier BV.
25. Y. Sun, D. Li, L. Zhao, X. Liu, K. Guan, Y. Ma, R. Wang, and Q. Li. PYY-mediated appetite control and obesity alleviation through short-chain fatty acid–driven gut-brain axis modulation by Lacticaseibacillus rhamnosus HF01 isolated from Qula. *Journal of Dairy Science*, 108(8):7960–7978, August 2025. ISSN 0022-0302. doi: 10.3168/jds.2024-26193. Publisher: American Dairy Science Association.
26. Daniela Frasca and Bonnie B Blomberg. Effects of aging on B cell function. *Current Opinion in Immunology*, 21(4):425–430, August 2009. ISSN 0952-7915. doi: 10.1016/j.coi.2009.06.001. Publisher: Elsevier BV.
27. Jiahao Wang, Menghua Zhang, Qingyao Zhao, Siqian Chen, Yongjie Tang, Quanzhen Chen, Lei Xu, Dan Wang, Xiaoping Guo, Kai Xing, Yachun Wang, ChuduanWang, Xixia Huang, and Ying Yu. Characterizing age-related features for assessing biological age and characteristics in Xinjiang Brown cattle. *BMC Genomics*, 26(1), May 2025. ISSN 1471-2164. doi: 10.1186/s12864-025-11430-2. Publisher: Springer Science and Business Media LLC.
28. Xinwei Jiao, Xiaoting Pei, Dingli Lu, Di Qi, Shenzhen Huang, Siyu He, and Zhijie Li. Microbial

Reconstitution Improves Aging-Driven Lacrimal Gland Circadian Dysfunction. *The American Journal of Pathology*, 191(12):2091–2116, December 2021. ISSN 0002-9440. doi: 10.1016/j.ajpath.2021.08.006. Publisher: Elsevier BV.

29. Francesco Beghini, Lauren J McIver, Aitor Blanco-Míguez, Leonard Dubois, Francesco Asnicar, Sagun Maharjan, Ana Mailyan, Paolo Manghi, Matthias Scholz, Andrew Maltez Thomas, Mireia Valles-Colomer, George Weingart, Yancong Zhang, Moreno Zolfo, Curtis Huttenhower, Eric A Franzosa, and Nicola Segata. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *eLife*, 10:e65088, May 2021. ISSN 2050-084X. doi: 10.7554/eLife.65088.
30. Aitor Blanco-Míguez, Francesco Beghini, Fabio Cumbo, Lauren J. McIver, Kelsey N. Thompson, Moreno Zolfo, Paolo Manghi, Leonard Dubois, Kun D. Huang, Andrew Maltez Thomas, William A. Nickols, Gianmarco Piccinno, Elisa Piperni, Michal Punčochář, Mireia Valles-Colomer, Adrian Tett, Francesca Giordano, Richard Davies, Jonathan Wolf, Sarah E. Berry, Tim D. Spector, Eric A. Franzosa, Edoardo Pasolli, Francesco Asnicar, Curtis Huttenhower, and Nicola Segata. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nature Biotechnology*, 41(11):1633–1644, November 2023. ISSN 1087-0156, 1546-1696. doi: 10.1038/s41587-023-01688-w. Publisher: Springer Science and Business Media LLC.
31. Yuan Liu, Kangni Yang, Yuqian Jia, Jingru Shi, Ziwen Tong, Dan Fang, Bingqing Yang, Chengrui Su, Ruichao Li, Xia Xiao, and Zhiqiang Wang. Gut microbiome alterations in high-fat-diet-fed mice are associated with antibiotic tolerance. *Nature Microbiology*, 6(7):874–884, May 2021. ISSN 2058-5276. doi: 10.1038/s41564-021-00912-0.
32. Xianping Li, Yuanming Huang, Liqiong Song, Yuchun Xiao, Shan Lu, Jianguo Xu, Jianguo Li, and Zhihong Ren. Lactobacillus plantarum prevents obesity via modulation of gut microbiota and metabolites in high-fat feeding mice. *Journal of Functional Foods*, 73:104103, October 2020. ISSN 17564646. doi: 10.1016/j.jff.2020.104103.
33. Spase Stojanov, Aleš Berlec, and Borut Štrukelj. The Influence of Probiotics on the Firmicutes/Bacteroidetes Ratio in the Treatment of Obesity and Inflammatory Bowel disease. *Microorganisms*, 8(11):1715, November 2020. ISSN 2076-2607. doi: 10.3390/microorganisms8111715. Publisher: MDPI AG.
34. María Bailén, Carlo Bressa, Sara Martínez-López, Rocío González-Soltero, Maria Gregoria Montalvo Lominchar, Celia San Juan, and Mar Larrosa. Microbiota Features Associated With a High-Fat/Low-Fiber Diet in Healthy Adults. *Frontiers in Nutrition*, 7:583608, December 2020. ISSN 2296-861X. doi: 10.3389/fnut.2020.583608.
35. Rebecca H. Gellman, Matthew R. Olm, Nicolas Terrapon, Fatima Enam, Steven K. Higginbottom, Justin L. Sonnenburg, and Erica D. Sonnenburg. Hadza Prevotella require diet-derived microbiota-accessible carbohydrates to persist in mice. *Cell Reports*, 42(11):113233, November 2023. ISSN 22111247. doi: 10.1016/j.celrep.2023.113233.
36. Maria Elisa Perez-Munoz, Autumn M. McKnite, Evan G. Williams, Johan Auwerx, Robert W. Williams, Daniel A. Peterson, and Daniel C. Ciobanu. Diet modulates cecum bacterial diversity and

physiological phenotypes across the bxd mouse genetic reference population. *PLOS ONE*, 14(10): e0224100, October 2019. ISSN 1932-6203. doi: 10.1371/journal.pone.0224100.

37. Himel Mallick, Ali Rahnavard, Lauren J. McIver, Siyuan Ma, Yancong Zhang, Long H. Nguyen, Timothy L. Tickle, George Weingart, Boyu Ren, Emma H. Schwager, Suvo Chatterjee, Kelsey N. Thompson, Jeremy E. Wilkinson, Ayshwarya Subramanian, Yiren Lu, Levi Waldron, Joseph N. Paulson, Eric A. Franzosa, Hector Corrada Bravo, and Curtis Huttenhower. Multivariable association discovery in population-scale meta-omics studies. *PLOS Computational Biology*, 17(11):e1009442, November 2021. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1009442.
38. Fuyong Li, Thomas C. A. Hitch, Yanhong Chen, Christopher J. Creevey, and Le Luo Guan. Comparative metagenomic and metatranscriptomic analyses reveal the breed effect on the rumen microbiome and its associations with feed efficiency in beef cattle. *Microbiome*, 7(1), December 2019. ISSN 2049-2618. doi: 10.1186/s40168-019-0618-5. Publisher: Springer Science and Business Media LLC.
39. Boping Li, Juanjuan Chen, Xiaobin Ou, Xiuli Liu, Zaoxu Xu, Xuesong Xiang, Yan Yang, and Qi Wang. In-depth multiomic characterization of the effects of obesity in high-fat diet-fed mice. *FEBS Open Bio*, 14(5):771–792, May 2024. ISSN 2211-5463, 2211-5463. doi: 10.1002/2211-5463.13788.
40. XuWen Mao, Guliruoyi Paerhati, Yuche Wu, and Lu Feng Cheng. Modulation of gut microbiota, up-regulation of ZO-1, and promotion of metabolism as therapeutic mechanisms of indole-3-carbinol against obesity in mice. *Frontiers in Pharmacology*, 15:1499142, January 2025. ISSN 1663-9812. doi: 10.3389/fphar.2024.1499142.
41. Jingjing Wang, Jinlong Qie, Danrong Zhu, Xuemei Zhang, Qingqing Zhang, Yuyu Xu, Yipeng Wang, Kai Mi, Yang Pei, Yang Liu, Guozhong Ji, and Xingyin Liu. The landscape in the gut microbiome of long-lived families reveals new insights on longevity and aging – relevant neural and immune function. *Gut Microbes*, 14(1):2107288, December 2022. ISSN 1949-0976, 1949-0984. doi: 10.1080/19490976.2022.2107288.
42. Yangli He, Weipeng Cui, Tuanyu Fang, Zeng Zhang, and Min Zeng. Metabolites of the gut microbiota may serve as precise diagnostic markers for sarcopenia in the elderly. *Frontiers in Microbiology*, 14: 1301805, December 2023. ISSN 1664-302X. doi: 10.3389/fmicb.2023.1301805.
43. Andrew R Ghazi, Kathleen Sucipto, Ali Rahnavard, Eric A Franzosa, Lauren J McIver, Jason Lloyd-Price, Emma Schwager, George Weingart, Yo Sup Moon, Xochitl C Morgan, Levi Waldron, and Curtis Huttenhower. High-sensitivity pattern discovery in large, paired multiomic datasets. *Bioinformatics*, 38(Supplement_1):i378–i385, June 2022. ISSN 1367-4803, 1367-4811. doi: 10.1093/bioinformatics/btac232.
44. Ludivine Laurans, Nicolas Venteclef, Yacine Haddad, Mouna Chajadine, Fawaz Alzaid, Sarvenaz Metghalchi, Bruno Sovran, Raphael G. P. Denis, Julien Dairou, Marina Cardellini, Jose-Maria Moreno-Navarrete, Marjolene Straub, Sarah Jegou, Claire McQuitty, Thomas Viel, Bruno Esposito, Bertrand Tavitian, Jacques Callebert, Serge H. Luquet, Massimo Federici, José Manuel Fernandez-Real, Remy Burcelin, Jean-Marie Launay, Alain Tedgui, Ziad Mallat, Harry Sokol, and Soraya Taleb. Genetic deficiency of indoleamine 2,3-dioxygenase promotes gut microbiota-mediated

metabolic health. *Nature Medicine*, 24(8):1113–1120, August 2018. ISSN 1078-8956, 1546-170X. doi: 10.1038/s41591-018-0060-4.

45. Chenglin Ye, Zhiqiang Li, Chun Ye, Li Yuan, Kailang Wu, and Chengliang Zhu. Association between Gut Microbiota and Biological Aging: A Two-Sample Mendelian Randomization Study. *Microorganisms*, 12(2):370, February 2024. ISSN 2076-2607. doi: 10.3390/microorganisms12020370.
46. Hyeon Cheol Kim, Hyun-Jin Kim, Jeongwoo La, Won Hyung Park, Sang Hee, Byeong Hoon Kang, Yumin Kim, and Heung Kyu Lee. 1 Title: Gut microbiota dysbiosis induced by brain tumor modulates the 2 efficacy of immunotherapy. *bioRxiv*, 2024.
47. Diana Ćesić, Liborija Lugović Mihić, Petar Ozretić, Ivana Lojkić, Marija Buljan, Mirna Šitum, Mario Zovak, Dinko Vidović, August Mijić, Nada Galić, and Arjana Tambić Andrašević. Association of Gut Lachnospiraceae and Chronic Spontaneous Urticaria. *Life*, 13(6):1280, May 2023. ISSN 2075-1729. doi: 10.3390/life13061280.
48. Yingjian Hou, Jing Li, and Shuhuan Ying. Tryptophan Metabolism and Gut Microbiota: A Novel Regulatory Axis Integrating the Microbiome, Immunity, and Cancer. *Metabolites*, 13(11):1166, November 2023. ISSN 2218-1989. doi: 10.3390/metabo13111166.
49. Yehudit Hasin, Marcus Seldin, and Aldons Lusis. Multi-omics approaches to disease. *Genome Biology*, 18(1):83, December 2017. ISSN 1474-760X. doi: 10.1186/s13059-017-1215-1.
50. Yang Liao, Gordon K Smyth, and Wei Shi. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Research*, 47(8): e47–e47, May 2019. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkz114.
51. Charlotte Soneson, Michael I. Love, and Mark D. Robinson. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research*, 4:1521, December 2015. ISSN 2046-1402. doi: 10.12688/f1000research.7563.1.
52. Paul Shannon, Andrew Markiel, Owen Ozier, Nitin S. Baliga, Jonathan T. Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, 13(11):2498–2504, November 2003. ISSN 1088-9051. doi: 10.1101/gr.1239303.
53. S. Varrette, H. Cartiaux, S. Peter, E. Kieffer, T. Valette, and A. Olloh. Management of an Academic HPC & Research Computing Facility: The ULHPC Experience 2.0. In *Proc. of the 6th ACM High Performance Computing and Cluster Technologies Conf. (HPCCT 2022)*, Fuzhou, China, July 2022. Association for Computing Machinery (ACM). ISBN 978-1-4503-9664-6.

Supplementary Information

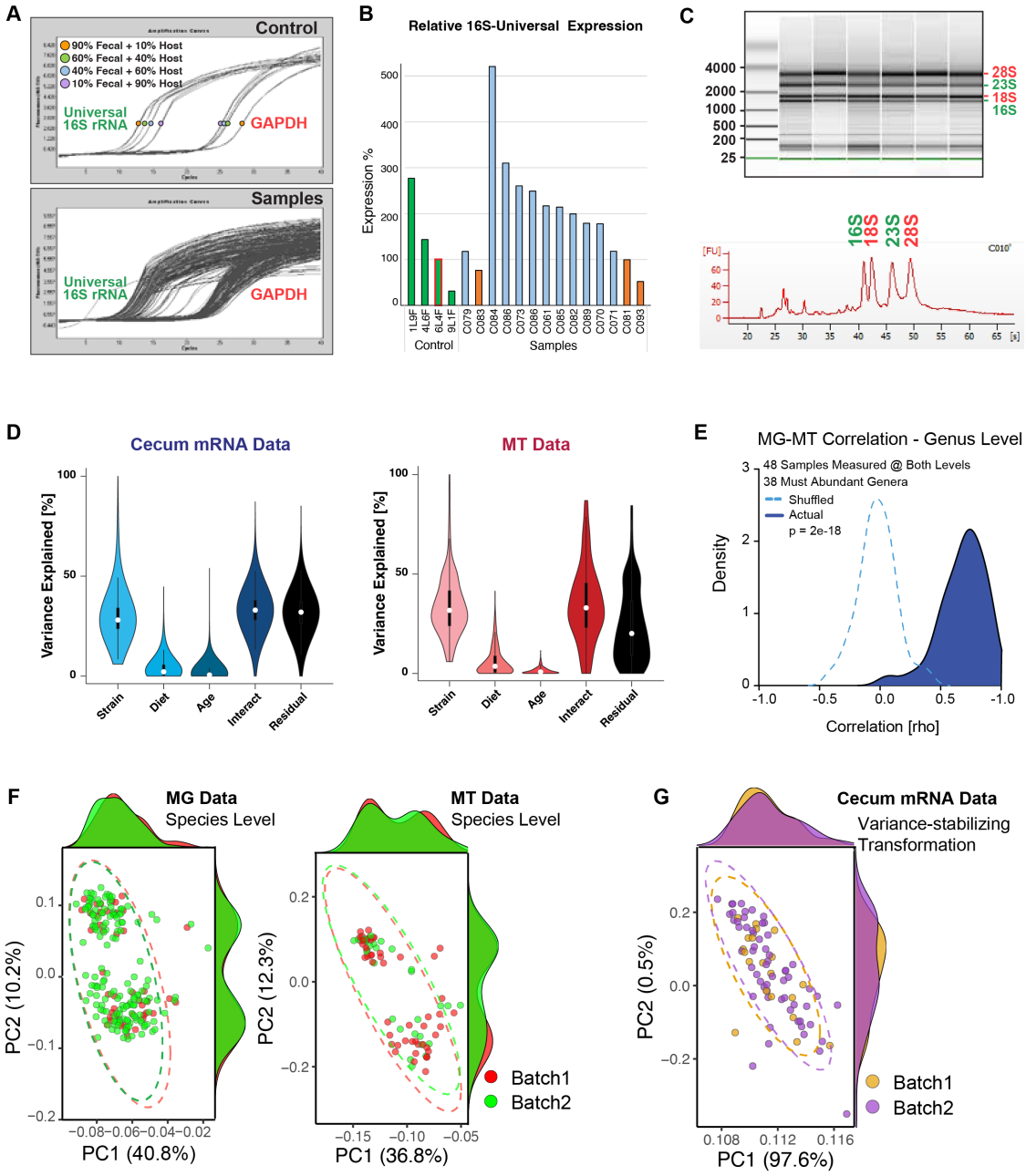

Figure S1. Establishment of Experimental Methods and Quality Control.

(**A**) A method based on RT-qPCR was established to assess the ratio of host to gut microbiota DNA in samples. (**B**) Samples with gut microbial DNA content estimated to be above ~40% were selected based on gradient controls. (**C**) Bioanalyzer results of RNA samples were used to select samples with four clear bands and no significant degradation for subsequent sequencing. (**D**) ANOVA analysis of cecal mRNA and MT data showing variation explained. (**E**) Correlation analysis indicates a strong correlation between the MG and MT data. (**F**) PCA plots of MG and MT at the species level show no dominant batch effect; the two distinct clouds within MG and MT data are the dietary batches. (**G**) PCA based on variance-stabilized cecal mRNA expression data shows that samples do not cluster by batch, indicating no evident batch effect. In this data, the impact of diet is not immediately evident from the data cloud.

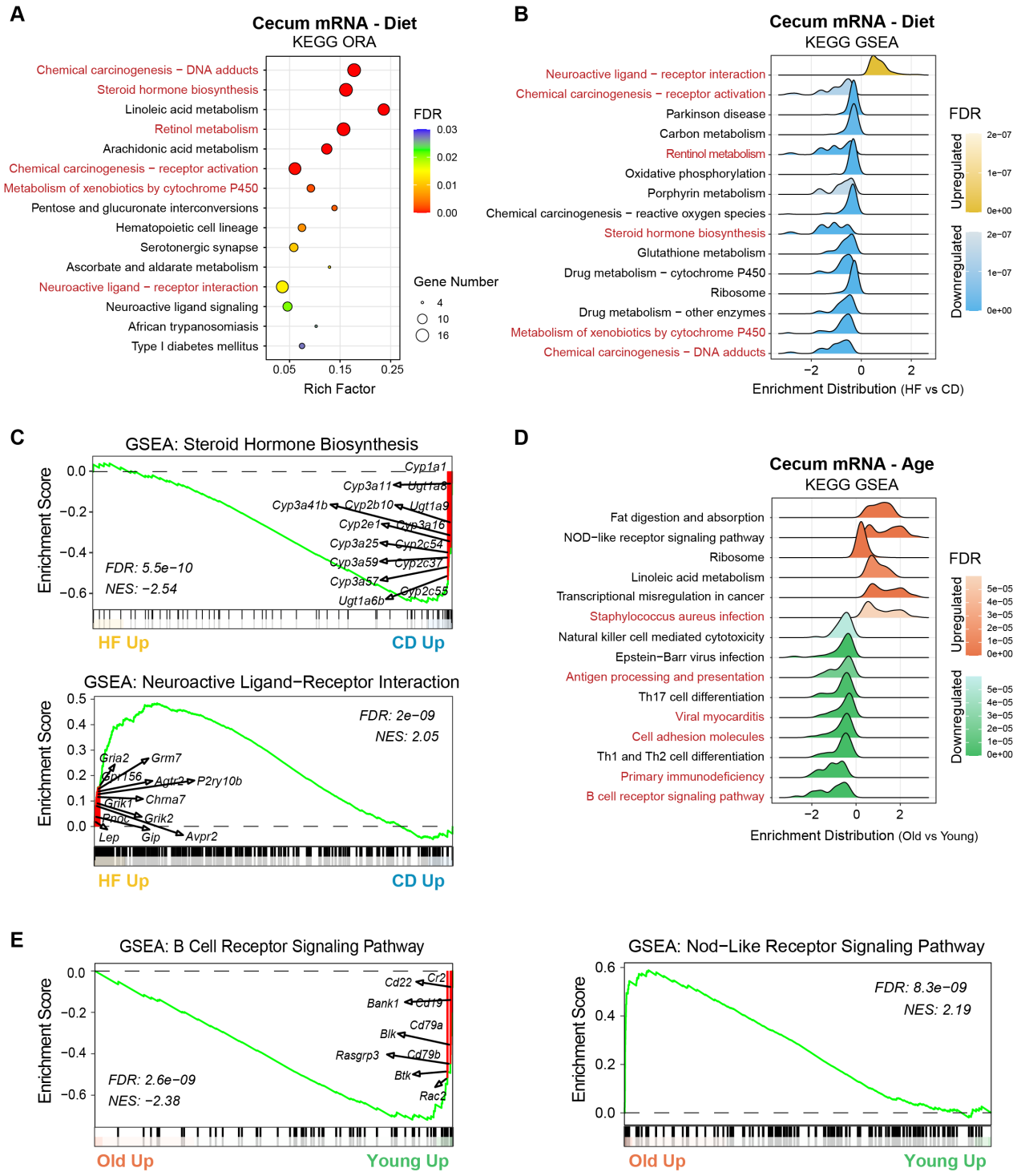

Figure S2. Differential Pathway Enrichment in Cecal Transcriptomes by Diet and Age.

(**A**) KEGG ORA results showing the top 15 enriched pathways. Bubble size indicates the number of DEGs per pathway; bubble color represents FDR; the x-axis shows Rich Factor. (**B**) KEGG GSEA for the same diet comparison, displaying the top 15 pathways ranked by FDR. (**C**) Representative GSEA enrichment plots for diet: Steroid hormone biosynthesis (upper; enriched in CD) and Neuroactive ligand–receptor interaction (lower; enriched in HF). Marked genes are DEGs present in the respective pathways. (**D**) KEGG GSEA results showing the top 15 age-associated pathways. (**E**) Two GSEA plots for age: B cell receptor signaling (left; enriched in Young) and NOD-like receptor signaling (right; enriched in Old).

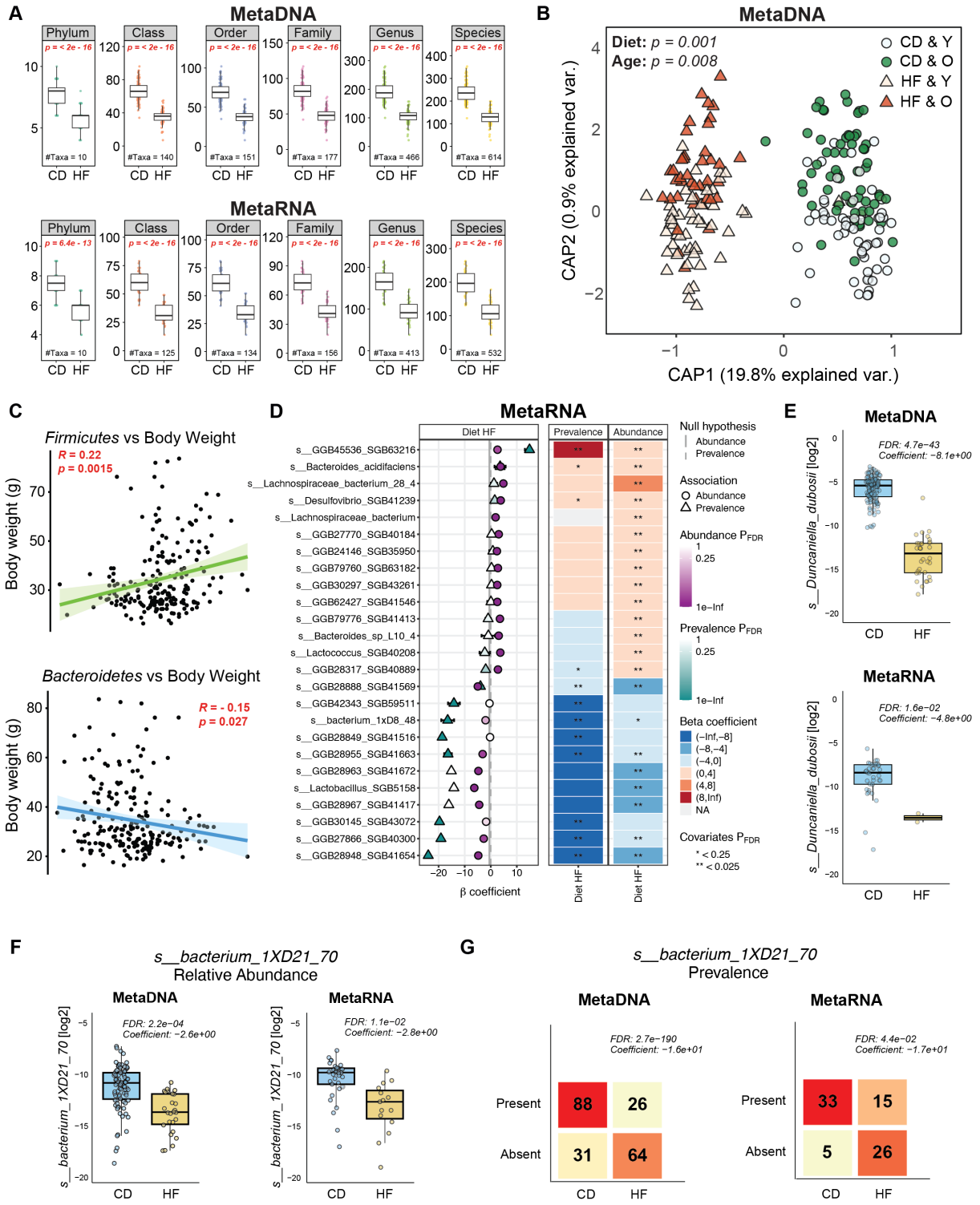

Figure S3. Supplementary Analysis of MG and MT Data.

(**A**) Total number of detected gut microbial taxa across all taxonomic levels. HF significantly reduced microbial richness at each taxonomic level. (**B**) Distance-based redundancy analysis(dbRDA) based on Bray–Curtis dissimilarity showing gut microbial community structure across diet and age groups (CD-Young, CD-Old, HF-Young, HF-Old). *P*-values from PERMANOVA (999 permutations). (**C**) *Firmicutes* and *Bacteroidetes* phyla are significantly associated with body weight. (**D**) Coefficient plot and heatmap of the most significant associated species in MT data between different diets. (**E**) *s__Duncaniella_dubosii*. The relative abundance (**F**) and prevalence (**G**) of *s__bacterium_1XD21_70* across different diets.

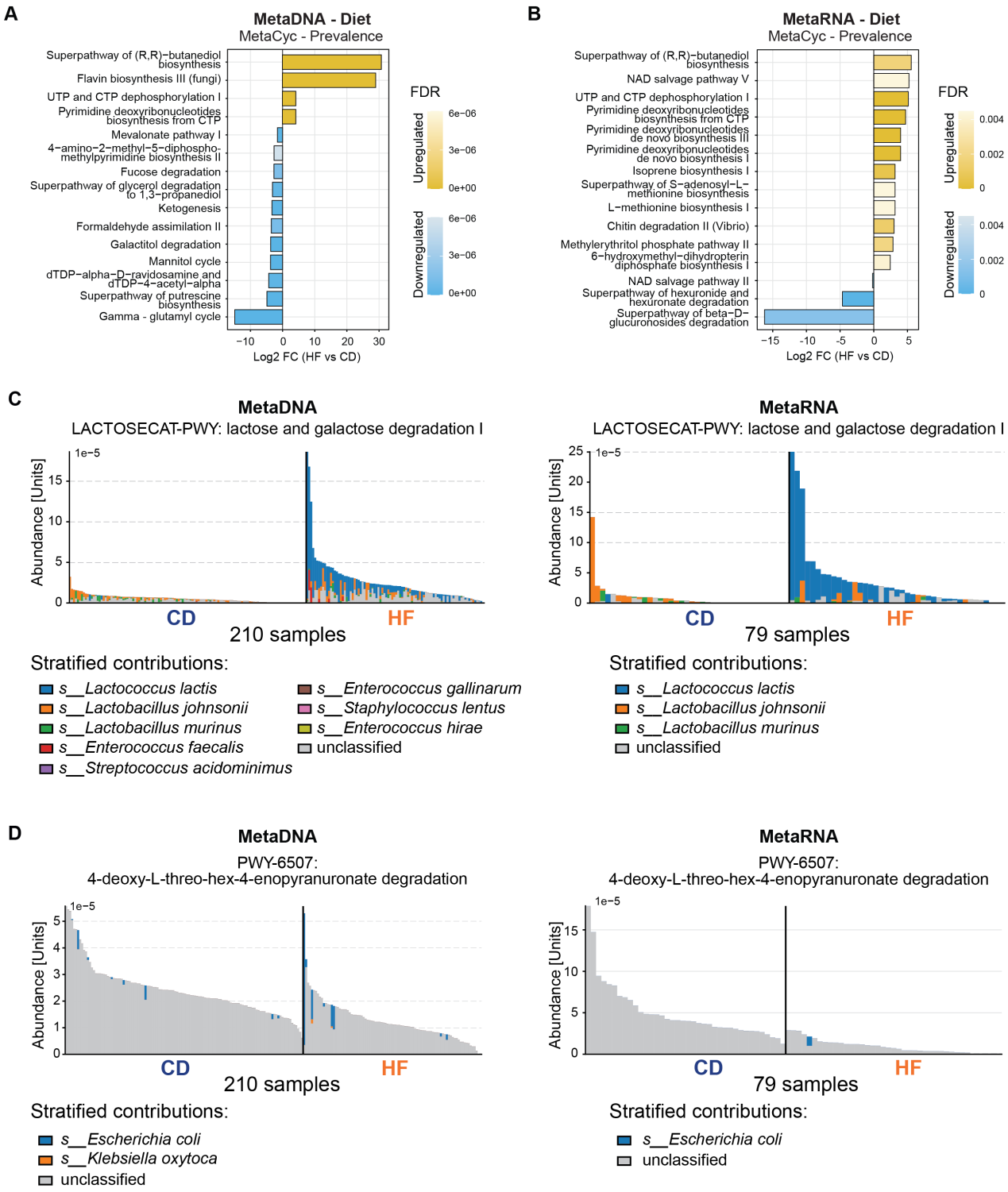

Figure S4. Supplementary Details of Diet-associated Changes in Microbial Functional Profiles.

(**A-B**) Top 15 MetaCyc pathways with significant dietary differences in pathway prevalence from MG data(**A**) and MT data(**B**). (**C-D**) Relative abundance distributions of the lactose and galactose degradation I pathway (LACTOSECAT-PWY) (**C**) and the 4-deoxy-L-threo-hex-4-enopyranuronate degradation pathway (PWY-6507) (**D**) in the MG and MT datasets, stratified by major contributing species — few of which have been identified for this pathway.

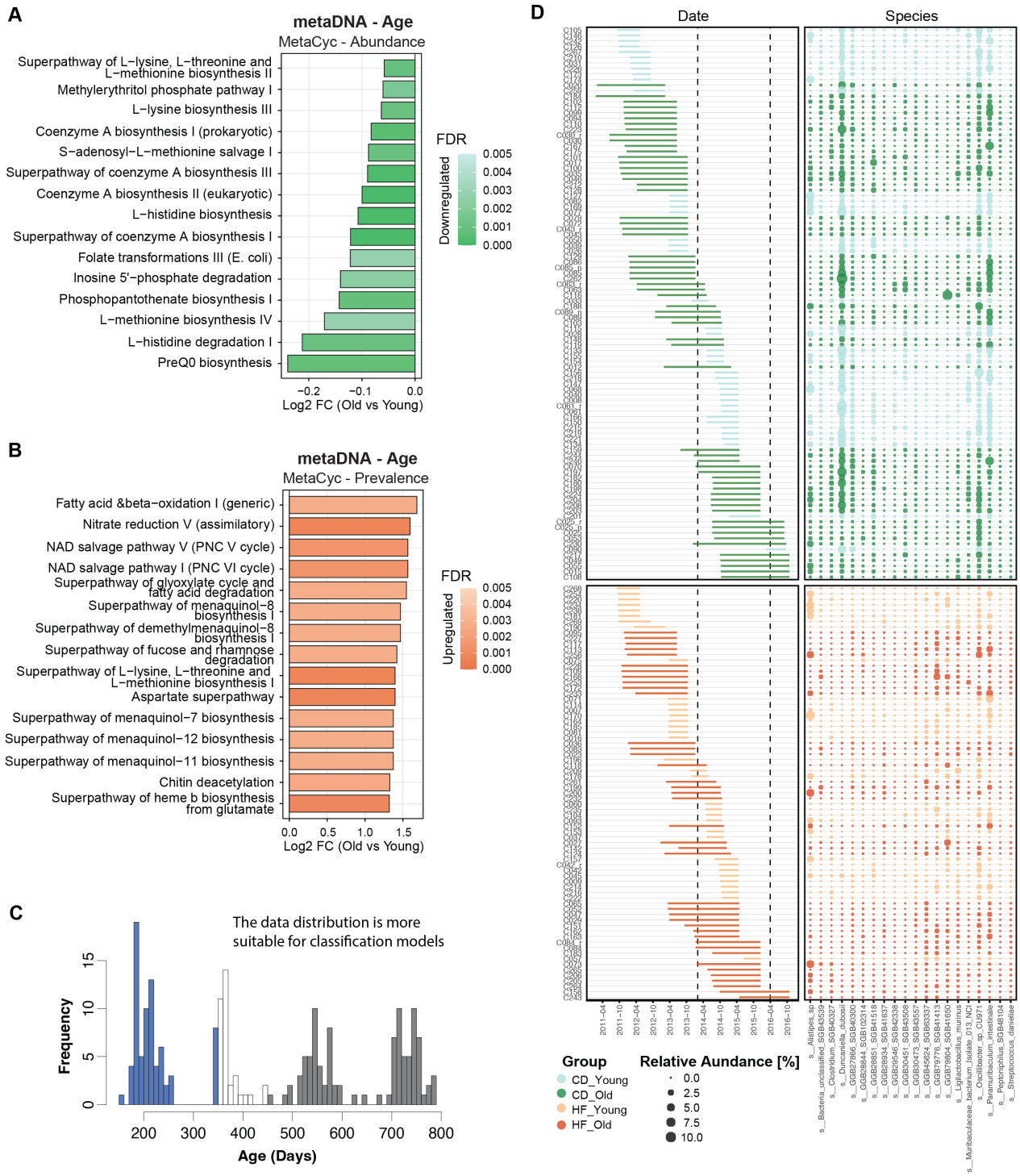

Figure S5. Supplementary Details Related to Age-Associated Analyses.

(**A-B**) Abundance and prevalence of the top 15 age-associated MetaCyc pathways in MG data. Bars show $\log_2$FC (Old vs Young). (**C**) Distribution of sample ages across the MG cohort. Young and old groups were defined to ensure comparable sample sizes. (**D**) Combined plot of the relative abundances of the top 20 species identified by Random Forest model and the experimental time points of the samples.

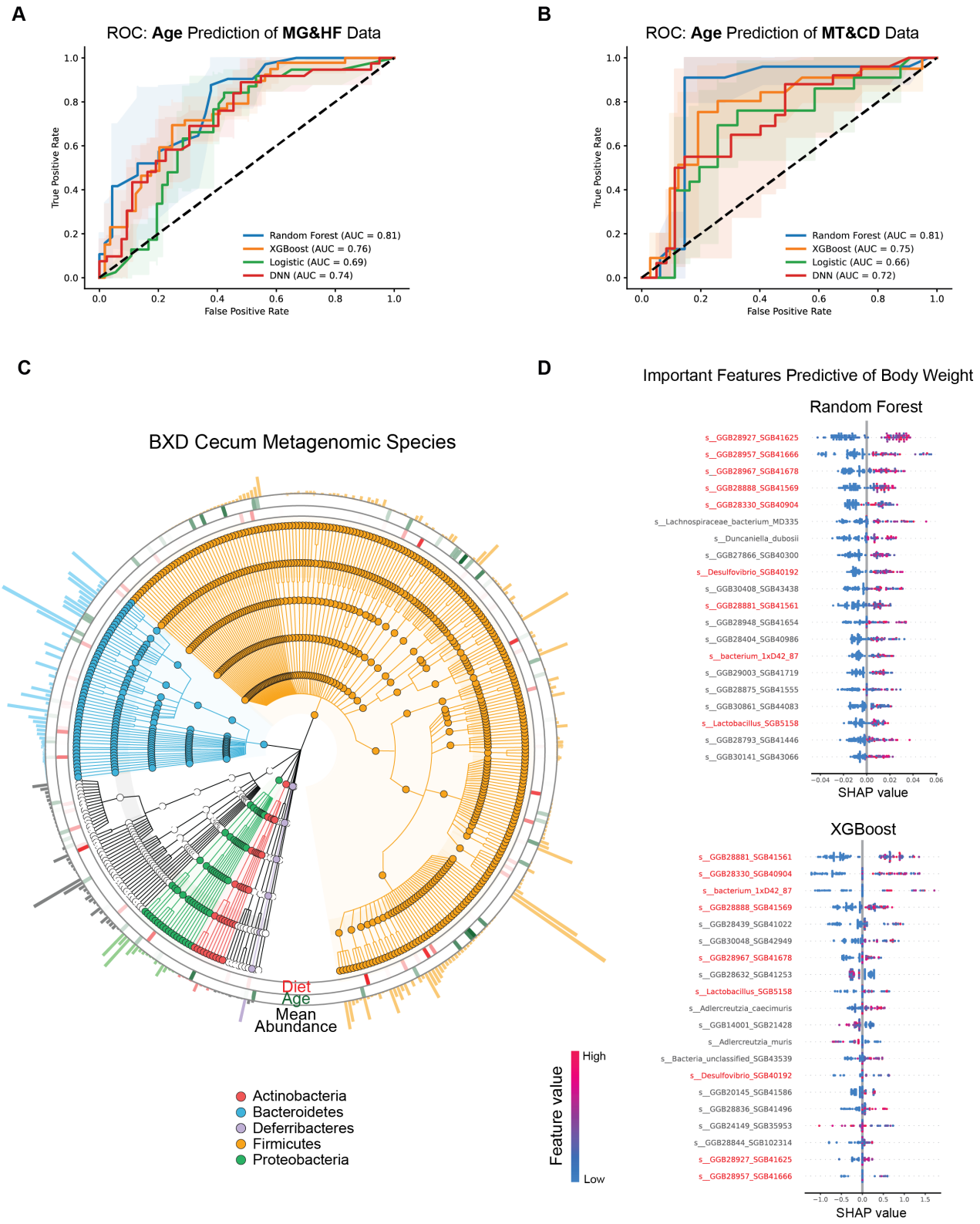

Figure S6. Machine Learning Models Predicting Age or Body Weight Using Multi-Omics Data.

(**A**) ROC curve of the age prediction model for the HF group in the MG data. (**B**) ROC curve of the age prediction model for the CD group in the MT data. (**C**) Cladogram of the MG. Outer ring 1 represents species with significant differences across different diets. Outer ring 2 represents species with significant differences across different ages. Outer ring 3 indicates the mean relative abundance of each species. (**D**) Top 20 biomarkers selected by RF and XGB models for body weight prediction using MG data.

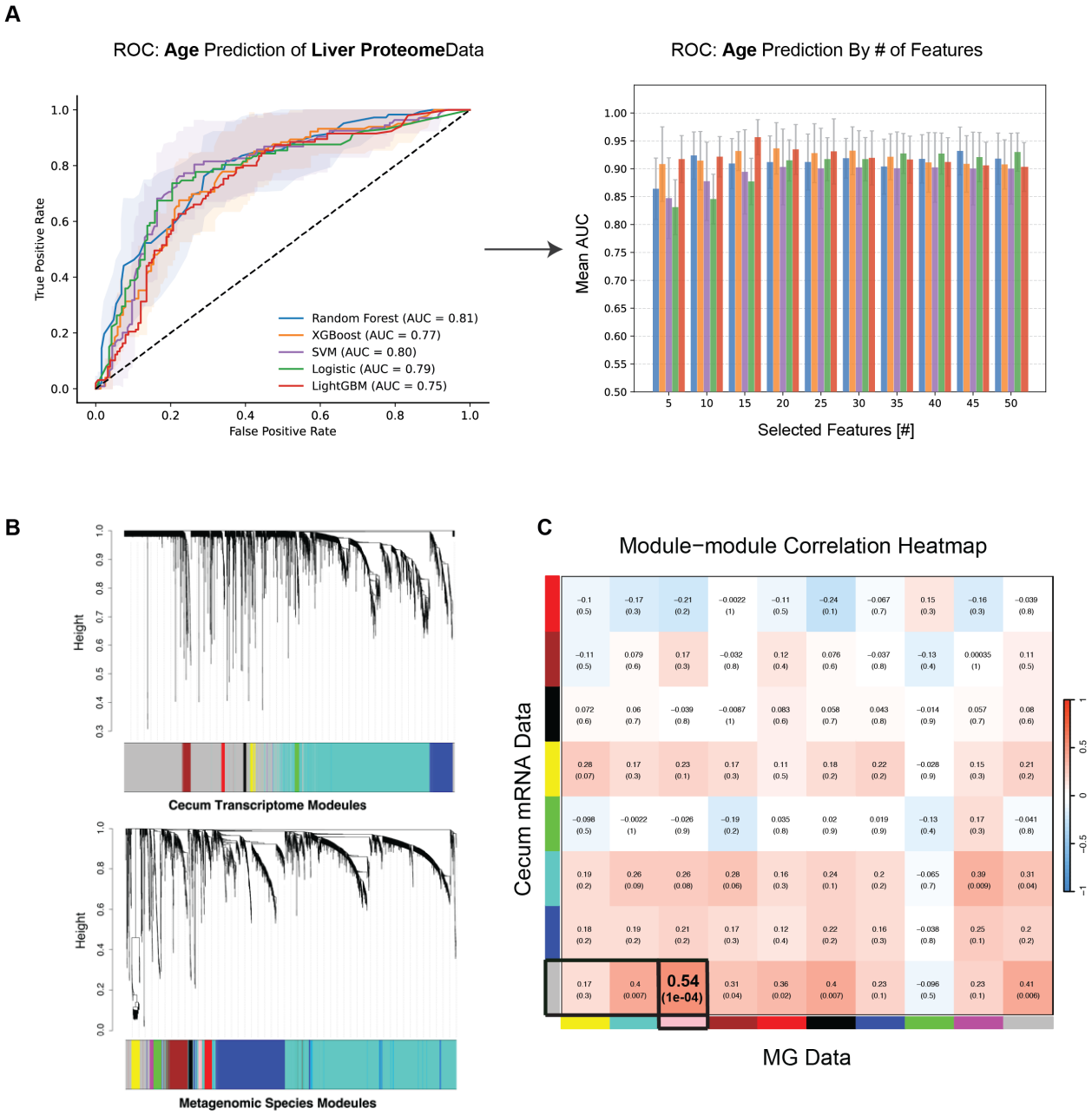

Figure S7. Supporting Results from Integrated Multi-omics and Network Analyses.

(**A**) ROC curve and feature selection bar plot of the age prediction model for the CD group based on liver proteomics data. (**B**) Hierarchical clustering dendrograms of cecal mRNA (left panel) and MG (right panel), showing co-expression modules identified by WGCNA. (**C**) Module–module correlation heatmap between cecal mRNA modules and MG species modules. The strongest positive correlation was observed between MEgrey (cecum) and MEpink (microbiome) modules (r = 0.54, FDR = 1e-04).