Decoding host-microbiome interactions through co-expression network analysis within the non-human primate intestine

Mika Uehara[1], Takashi Inoue[2], Sumitaka Hase[1], Erika Sasaki[2, 4], Atsushi Toyoda[3], Yasubumi Sakakibara[1*]

1) Department of Biosciences and Informatics, Keio University, Yokohama, Kanagawa 223-8522, Japan

2) Department of Marmoset Biology and Medicine, Central Institute for Experimental Animals, Kawasaki, Kanagawa 210-0821, Japan

3) Department of Genomics and Evolutionary Biology, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan

4) Laboratory for Marmoset Neural Architecture, RIKEN Center for Brain Science, 2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan.

*Corresponding Author

Yasubumi Sakakibara

3-14-1 Hiyoshi, Kohoku-ku, Yokohama, 223-8522, Japan

Phone/Fax: +81-45-566-1791

E-mail: yasu@bio.keio.ac.jp

ABSTRACT

The gut microbiome affects the health status of the host through complex interactions with the host's intestinal wall. These host-microbiome interactions may spatially vary along the physical and chemical environment of the intestine, but these changes remain unknown. This study investigated these intricate relationships through a gene co-expression network analysis based on dual transcriptome profiling of different intestinal sites—cecum, transverse colon, and rectum—of the primate common marmoset. We proposed a gene module extraction algorithm based on graph theory to find tightly interacting gene modules of the host and the microbiome from a vast co-expression network. The 27 gene modules identified by this method, which include both host and microbiome genes, not only produced results consistent with previous studies regarding the host-microbiome relationships, but also provided new insights into microbiome genes acting as potential mediators in host-microbiome interplays. Specifically, we discovered associations between the host gene *FBP1*, a cancer marker, and polysaccharide degradation-related genes (*pfkA* and *fucI*) coded by *Bacteroides vulgatus*, as well as relationships between host B-cell specific genes (*CD19*, *CD22*, *CD79B*, *PTPN6*) and a tryptophan synthesis gene (*trpB*) coded by *Parabacteroides distasonis*. Furthermore, our proposed module extraction algorithm surpassed existing approaches by successfully defining more functionally related gene modules, providing insights for understanding the complex relationship between the host and the microbiome.

IMPORTANCE We unveiled the intricate dynamics of the host-microbiome interactions along the intestinal tract by identifying closely interacting gene modules from a vast gene co-expression network, constructed based on simultaneous profiling of both host and microbiome transcriptomes. Our proposed gene module extraction algorithm, designed to interpret inter-species interactions, enabled the identification of functionally related gene modules

encompassing both host and microbiome genes, which was challenging with conventional modularity maximization algorithms. Through these identified gene modules, we discerned previously unrecognized bacterial genes that potentially mediate in known relationships between host genes and specific bacterial species. Our findings underscore the spatial variations in host-microbiome interactions along the colon, rather than displaying a uniform pattern throughout the colon.

INTRODUCTION

The mutual interaction between the intestinal microbiome and the host constitutes a complex biological process, with intertwined dependencies that are essential for the survival and health of both. The interactions between the intestinal microbiome and the host are formed by various factors such as the effects of metabolites produced by bacteria on the host, the binding of bacterial-derived proteins and host receptors, and the binding of bacterial cell surface structures and host receptors. Studies on host-microbiome interaction through bacterial-derived proteins have discovered that a novel protein P9 secreted by *Akkermansia muciniphila* binds to host ICAM-2, thereby increasing the secretion of glucagon-like peptide 1 (GLP-1), suggesting that the interaction between P9 and ICAM-2 could be a potential therapeutic target in type 2 diabetes patients (1). Excluding certain gram-negative bacteria, it has also been reported that bacterial lipopolysaccharides (LPS), which are components of the gram-negative bacterial cell wall, bind to the Toll-like receptor 4 (TLR4) complex and can induce innate immune responses (2). Among these interactions, the host-microbiome interaction mediated by metabolites is particularly well-studied. For instance, primary bile acids, produced in the host's liver cells, are converted into secondary bile acids by the microbiome in the colon. This microbiome-derived secondary bile

69 acid is known to regulate the homeostasis of regulatory T cells, which are immune-suppressive
70 cells in the host (3). Through these processes, the intestinal microbiome and host mutually
71 interact and establish a symbiotic relationship while forming a metabolic pathway or
72 physiological response.

73 To understand such complex interactions between the microbiome and the host, it is
74 beneficial to analyze combined profiles of both the microbiome and the host. Recent studies have
75 reported on dual profiling of host and microbiome gene expression levels to understand the
76 interactions between the microbiome and host (4, 5, 6). Previously, the focus was primarily on
77 identifying bacterial species related to host phenotypes (such as diseases) based on the relative
78 abundance of microbiome (7). However, profiling both host and microbiome gene expression
79 levels has provided further insights into which genes of which bacterial species are related to
80 host physiological functions and phenotypes. Gene co-expression network analysis is one of the
81 useful approaches for comprehensively deciphering complex interactions from the vast number
82 of gene expression profiles of the host and microbiome. A gene co-expression network is
83 constructed by representing genes as nodes and connecting genes with correlated expression
84 levels via edges. Representing gene relationships as a network enables the extraction of closely
85 interconnected gene modules (subgraphs) and the identification of hub genes connected to
86 multiple genes, thereby enhancing interpretability. For instance, a study (5) characterizing the
87 host-microbiome interaction during influenza virus infection identified modules from the host's
88 gene co-expression network. It discovered correlations between host gene modules involved in
89 interferon signal transduction and the sulfonamide antibiotic resistance genes of the microbiome.
90 In another study (6), network analysis was applied to the gene expression profiles of the host and
91 microbiome in the nasopharyngeal airway during bronchiolitis. Networks were constructed for

both the host and microbiome and modules were identified from these networks. The study revealed that host modules were involved in regulating host-T-cell activity. Additionally, microbiome modules involved in microbial-BCAA metabolism and oxidative stress response were associated with disease severity.

Despite such progress, several limitations remain in understanding the interactions between the host and the intestinal microbiome. First, there's a lack of understanding of the spatial variation in host-microbiome interactions. The colon, rich in microbiota, can be subdivided into the cecum, ascending colon, transverse colon, descending colon, sigmoid colon, and rectum. Each section possesses distinct functions and differs in its physicochemical environment, including variations in nutrient content, pH, and luminal flow (8, 9, 10) and the microbiome could vary as well. Indeed, studies using animal models have reported that the composition and functional fluctuation of the microbiome varies due to environmental changes depending on intestinal location (biogeographical location) (11, 12, 13). However, these studies focused solely on the microbiome, and the coordinated spatial variation of both the microbiome and host gut cells along the intestinal tract has yet to be clarified.

Second, there is a lack of computational methods for extracting significant modules (subgraphs) from the host-microbiome gene co-expression network. Analyzing a massive gene co-expression network integrating the gene expression profiles of both the host and microbiome helps interpret the interactions between the genes of the host and the microbiome. Especially in a huge gene co-expression network, it is crucial to extract densely connected subgraphs (modules) to divide the network into understandable units; however, existing algorithms for extracting modules from the network are insufficient for accurately capturing the interaction between the

114 host and microbiome because they were not designed with the purpose of understanding
115 interactions between organisms (14, 15, 16, 17).
116 Lastly, there are concerns regarding the applicability of the results of studies using animal
117 models to humans. While mouse models have been the basis of biological research for a long
118 time due to ethical constraints, experiments in controlled environments, and considerations of
119 reproducibility, there are several issues and limitations to their use. The physiology, anatomy, and
120 pharmacology of mice differ from those of humans, and there are differences in the composition
121 and function of the intestinal microbiome between mice and humans (13, 18). Particularly
122 considering the complexity of host-microbiome interactions, it is challenging to precisely
123 replicate the relationship between humans and their intestinal microbiome. In contrast, the
124 common marmoset, a small New World primate, is considered a useful model in preclinical trials
125 because it is genetically close to humans and has common physiological and anatomical
126 characteristics (19). Moreover, the common marmoset is the only non-human primate in which a
127 germ-free state has been successfully achieved (20), and it has a fecal microbiota similar to that
128 of humans compared to other animal models, including mice (13), potentially broadening the
129 scope of intestinal microbiome research.
130 To address these challenges, this study aimed to reveal the interactions between the host
131 and the microbiome along multiple sites of the common marmoset colon—cecum, transverse
132 colon, and rectum—by developing a module extraction method applicable to the gene co-
133 expression network that represents the complex interactions between the host and the
134 microbiome. Through simultaneous profiling of host and microbiome transcripts in the intestinal
135 sites of the common marmoset, we constructed an integrated gene co-expression network,
136 incorporating both host and microbiome genes. From this network, we further identified densely

137 covariant gene modules (subgraphs) to interpret the interactions between genes. As conventional
138 methods for extracting gene modules from the network were insufficient for interpreting the
139 interactions between organisms examined in this study, we proposed a module extraction
140 algorithm based on graph theory. Our proposed method successfully extracted more modules
141 containing functionally related genes of both the host and microbiome compared to existing
142 algorithms, underlining its suitability for predicting host-microbiome interactions. This method
143 extracted 27 gene modules containing both host and microbiome, indicating that the gut
144 microbiome is functionally closely involved in the variation of host gene expression along the
145 colon.
146

147 **RESULTS**

148 ***Overview of the analysis pipeline integrating host and microbiome transcriptomes:***
149 ***Identification of closely interacting gene modules from gene co-expression networks***

150 To investigate the host and microbiome interactions along the colon, we proposed an analysis
151 method integrating the host transcriptome and the microbiome metatranscriptome (Fig. 1). First,
152 we computed gene expression profiles from the host transcriptome and microbiome
153 metatranscriptome obtained via dual RNA-seq and created three types of gene co-expression
154 networks: host-host, host-microbiome, and microbiome-microbiome. By unifying these three
155 networks, we constructed a host and microbiome gene co-expression network, with the genes of
156 the host and microbiome serving as nodes, and covariant genes interconnected. Second, from the
157 gene co-expression network, we identified gene modules containing closely interacting genes
158 from both the host and microbiome. We aimed to interpret the interactions between host genes
159 and microbiome genes within modules extracted from a gene co-expression network, based on

160 the assumption that functionally similar genes are co-expressed (21). Conventional module
161 extraction methods are not designed to understand interspecies interactions (14, 15, 16, 17) and
162 do not distinguish between host gene and microbiome gene nodes in the network. Consequently,
163 modules composed solely of host genes or exclusively of microbiome genes can be extracted,
164 making it challenging to understand the interactions between the host and the microbiome when
165 such modules are present. Therefore, we proposed a clique-search-based module extraction
166 method that takes into account the structure of three networks: host-host, host-microbiome, and
167 microbiome-microbiome. This method begins by exploring cliques from the gene co-expression
168 network. In graph theory, a "clique" refers to a subgraph within the network where all nodes are
169 fully connected. For overlapping cliques, the method calculates the cliqueness in each of the
170 original three networks and merges them if the cliqueness of all networks exceeds a certain
171 threshold, thereby obtaining the final gene modules (refer to the MATERIALS AND METHODS
172 section for details).
173
174 ***Computing gene expression profiles of the host and the microbiome from dual RNA-seq data***
175 First, we extracted high-quality RNA (RNA integrity number (RIN) scores ≥8.6; Table S1) from
176 the intestinal sites (cecum, transverse colon, and rectum) of five common marmosets
177 (Supplementary Note Section 1) and carried out dual RNA-seq with NovaSeq 6000 v1.5 (100 bp
178 PE). The obtained RNA read data amounted to a total of 4,724 M reads (mean 157,450,033
179 paired-end reads/sample; Table S2; Supplementary Note Section 2).
180 Next, as the RNA read data obtained by dual RNA-seq contains a mixture of reads
181 derived from both the host and the microbiome, it is necessary to separate these read data in
182 silico (Table S3; Supplementary Note Section 3). The pipeline established in this study separates

183 the host-derived read data and the microbiome-derived read data by aligning the RNA reads to
184 the host reference genome (Fig. 1B). To avoid transcriptomic cross-contamination of read data
185 between host and microbiome read data, we adjusted the pipeline parameters using simulated
186 RNA read data from the host and the microbiome (Supplementary Note Section 4). As a result,
187 we achieved a precision of 1.0 in separating the read data of the host and the microbiome, and
188 our established pipeline completely separated the reads derived from the host and those from the
189 microbiome (Table S9 posted at https://doi.org/10.5281/zenodo.10425960).
190 After the above preprocessing, we calculated the gene expression levels (transcripts per
191 million (TPM)) for both the host and the microbiome. As a result, we detected 16,578 host
192 expressed genes and 8,533 microbiome expressed genes in total (Supplementary Note Section 5
193 and 6). To examine the expression patterns across different sites and across different individuals
194 for both host genes and microbiome genes, we applied Principal Component Analysis (PCA) to
195 the gene expression profiles for visualization. In the host profiles, distinct clusters were formed
196 for each site, while the clusters for each individual overlapped (Fig. 2A and B). In the
197 microbiome profiles, clusters were separated for the cecum and rectum, which are the beginning
198 and end of the colon, respectively, with the transverse colon positioned between the cecum and
199 rectum overlapping between the two clusters (Fig. 2C). The clusters for each individual also
200 overlapped in the microbiome profiles (Fig. 2D). These results suggest that both host and
201 microbiome gene expression profiles are more variable across different sites compared to those
202 across individuals.
203
204 ***Spatial variations in host gene expression among intestinal sites***

205 Of the 16,578 host genes detected in three regions of the common marmoset's intestinal tract—
206 cecum, transverse colon, and rectum—a total of 2,615 genes showed significant expression
207 variations among these sites (FDR < 0.01) (Table S4; Supplementary Note Section 5).
208 Furthermore, upon conducting functional pathway analysis, 72 pathways were identified, all of
209 which were significantly upregulated in the cecum compared to the transverse colon (normalized
210 enrichment score > 1.5, FDR < 0.1; Table S10 posted at
211 https://doi.org/10.5281/zenodo.10425960). In particular, pathways with high normalized
212 enrichment scores, such as Fc epsilon RI signaling pathway, T cell receptor signaling pathway,
213 leukocyte transendothelial migration, Yersinia infection, and B cell receptor signaling pathway
214 were detected, indicating that pathways related to immune functions were upregulated in the
215 cecum relative to the transverse colon (Fig. 3A). Additionally, a substantial number of genes
216 involved in these functional pathways were found to be significantly upregulated in the cecum
217 (FDR < 0.01). For example, among the genes specifically upregulated in the cecum, we found
218 genes related to B cell receptor signaling. These include *CD19*, which acts as a coreceptor for the
219 B cell antigen receptor complex and augments B cell activation (22); *CD79A* and *CD79B*,
220 components of the B cell antigen receptor that generate signals upon antigen recognition (23);
221 and *CD22* and *CD72*, which inhibit the signal transduction of the B cell antigen receptor (24,
222 25). In addition, genes associated with T cell receptor signaling were also found as being highly
223 expressed specifically in the cecum. These include *CD28*, a co-stimulatory receptor for T cell
224 receptors that activates T cells (26); *CD3D*, a component of CD3 that plays a role in signal
225 transduction during T cell activation (27); *CD40LG*, which is induced on the T cell surface
226 following T cell receptor (TCR) activation (28); and *CTLA4*, which attenuates T cell activation
227 and fine-tunes immune responses (29) (Fig. 3B).

228

229 ***Interpreting host-microbiome interactions through spatial gene co-expression network***

230 ***analysis within the intestine***

231 We constructed the unified gene co-expression network composed of 1,359 gene nodes, which

232 embodies co-expression information on host-host, host-microbiome, and microbiome-

233 microbiome interactions (Table S11, S12, and S13 posted at

234 https://doi.org/10.5281/zenodo.10425960; Supplementary Note Section 7). Subsequently, to

235 identify densely co-expressed gene sets within this network, we implemented the clique-based

236 module extraction algorithm proposed in this study. As a result, we then identified 27 gene

237 modules containing both host and microbiome genes (Table S14 and S15 posted at

238 https://doi.org/10.5281/zenodo.10425960). Furthermore, upon conducting a functional pathway

239 enrichment analysis on these gene modules, we found that 70.4% (19 out of 27) of gene modules

240 could be functionally characterized (Supplementary Note Section 8).

241 The modules we identified were characterized by functional pathways including the B

242 cell receptor signaling pathway, fructose and mannose metabolism, oxidative phosphorylation,

243 and biosynthesis of amino acids (Fig. 4). In addition, we provided visualizations of the bacterial

244 species that contribute to microbiome gene expression within these identified modules, together

245 with host and microbiome genes (Fig.5, 6, 7 and 8).

246 Module_07, as shown in Figure 5, was identified as a module with a positive correlation

247 between the host gene Fructose-1,6-bisphosphatase 1 (*FBP1*), which is highly expressed in the

248 transverse colon, and 40 microbiome genes. These 40 gut microbiome genes included those

249 associated with fructose-mannose metabolism, such as K01840 (*manB*), K00850 (*pfkA*), and

250 K01818 (*fucI*), as well as those related to oxidative phosphorylation, such as K00241 (*sdhC*,

251 *frdC*), K00426 (*cydB*), K02112 (*atpD*), and K02121 (*ntpE*, *atpE*). The most substantial
252 contributor to the expression of these microbiome genes was *Bacteroides vulgatus*.
253 *FBP1* is known to have reduced expression in various tumor tissues, including the esophagus
254 (30), liver (31), lung (32), and ovarian cancer (33). It has been confirmed that its expression
255 increases with fructose treatment in intestinal organoids (34). Previous studies have identified
256 *FBP1* as a differentially expressed gene in the small intestine of germ-free mice, suggesting that
257 the intestinal microbiome influences the expression of the host's *FBP1* gene (35), but the
258 bacterial species and genes affecting the expression of the *FBP1* gene have not been identified.
259 *Bacteroides vulgatus*, which contributed to the expression of fructose-mannose metabolism-
260 related microbiome genes observed in Module_07, has been reported to release fructose by
261 decomposing polysaccharides (36, 37). In this module, the gene K00850 (*pfkA*) involved in the
262 fructo-oligosaccharide degradation and the gene K01818 (*fucI*) involved in the 2'-fucosyllactose
263 degradation were detected, revealing a novel association between polysaccharide degradation
264 genes encoded by *Bacteroides vulgatus* and the host gene *FBP1*.
265 In Module_01, the B cell receptor signaling pathway was enriched (Fig. 6). This module featured
266 covariation between host B-cell specific genes, including *CD19*, *CD22*, *CD79B*, and *PTPN6*, which were
267 highly expressed in the cecum, and the tryptophan synthesis gene K01696 (*trpB*) from the microbiome.
268 The upregulation of these host B-cell specific genes indicates an increase in B cells within the cecum.
269 This tryptophan synthesis gene K01696 (*trpB*) expression was contributed to by *Parabacteroides*
270 *distasonis*. Tryptophan is an essential amino acid that, upon uptake by intestinal cells and
271 subsequent metabolism, can activate the Aryl hydrocarbon receptor and influence B cell immune
272 responses (38). Tryptophan is also known to ameliorate autoimmune diseases, such as
273 inflammatory bowel disease (39). *Parabacteroides distasonis* has been suggested to be involved
274 in various autoimmune conditions in several studies (40, 41). For instance, *Parabacteroides*

275 *distasonis* levels have been shown to be lower in multiple sclerosis (MS) patients compared to
276 healthy controls, and transplanting the intestinal microbiota from MS patients to germ-free mice
277 resulted in increased severity of autoimmune encephalomyelitis symptoms (40). While previous
278 studies have reported associations between tryptophan and autoimmune diseases via B-cell
279 immune responses, as well as between autoimmune diseases and *Parabacteroides distasonis*, this
280 study identifies a correlation between the tryptophan synthesis gene K01696 (trpB) from
281 *Parabacteroides distasonis* and B-cell specific genes (*CD19*, *CD22*, *CD79B*, and *PTPN6*).
282 In Module_21, a group consisting of three host genes and 40 microbiome genes was
283 identified. The three host genes included *MMP10*, two uncharacterized genes LOC103795407
284 and *CCDC162P*, all of which showed high expression in the cecum (Fig. 7). The microbiome
285 genes that showed a positive correlation with these genes included K01491 (*folD*), K00605
286 (*gcvT*), K01803 (*tpiA*), K00831 (*serC*), K15633 (*gpmI*), K00024 (*mdh*), all of which are involved
287 in carbon metabolism pathways. *Bacteroides vulgatus* was identified as the bacteria contributing
288 to the expression of these microbiome genes. The *MMP10* gene plays a crucial role in
289 maintaining epithelial barrier function and is necessary for resolving colonic epithelial damage
290 (42). A study examining the host transcriptome and microbiota composition of colorectal cancer
291 tumors found high expression of *MMP10* in tumor tissues and a significant increase in the genus
292 *Bacteroides*, indicating a link between the host gene *MMP10* and the genus *Bacteroides* (43).
293 These findings are consistent with our observations. Furthermore, in the identified Module_21,
294 not only the relationship between the host gene *MMP10* and *Bacteroides vulgatus*, but also a
295 group of microbiome genes related to carbon metabolism pathways showing a positive
296 correlation with the host gene *MMP10* was identified. One hypothesis explaining the observed
297 positive correlation between carbon metabolism-related genes encoded by *Bacteroides vulgatus*

298 and the host gene *MMP10* is that the microbiome may form colonies within the host's mucosal
299 layer. Mucin, the principal component of mucus secreted by the gastrointestinal mucosal
300 epithelium, contains galactose and promotes bacterial settlement and adhesion (44), serving as a
301 nutrient source for certain bacteria. *Bacteroides vulgatus* possesses mucosal adherence factors,
302 allowing access to host-derived carbon sources (45). The upregulation of *MMP10* in the cecum
303 might represent a protective response against colonization by *Bacteroides vulgatus*. The
304 detection of carbon metabolism-related genes of *Bacteroides vulgatus* in this module_21
305 suggests they might be utilizing carbon sources derived from mucin or dietary glucose.
306 Specifically, while *Bacteroides vulgatus* cannot degrade human mucin independently, it can
307 acquire carbon sources from mucin in co-culture with *Akkermansia muciniphila* (46), which
308 possesses genes detected in this module. When *Akkermansia muciniphila* converts galactose to
309 glucose, *Bacteroides vulgatus* might then activate these carbon metabolism-related genes to
310 utilize this glucose (Fig. S1). Our findings indicate a relationship between the carbon
311 metabolism-related genes of *Bacteroides vulgatus* and the host gene *MMP10*; however, further
312 experiments are required to ascertain whether the upregulation of *MMP10* serves a protective
313 role against colonization by *Bacteroides vulgatus* and *Akkermansia muciniphila* or if *Bacteroides*
314 *vulgatus* utilizes mucin as a carbon source.
315 In Module_18, three host genes and 54 microbiome genes were identified, and the
316 expression of these microbiome genes was further influenced by five bacterial species. The
317 relationships between these host genes and microbiome genes, as well as between the host genes
318 and bacterial species, were consistent with existing knowledge about bile acid synthesis and
319 degradation (Fig. 8). This module included the host gene *CYP27A1*, which is involved in bile
320 acid production (47), and *NR1H3*, a host gene that codes for a bile acid receptor (48). These

genes were upregulated in the cecum and transverse colon compared to the rectum. The microbiome genes interacting with these host genes included K03664 (*smpB*), a gene essential for bacterial survival when upregulated in the presence of secondary bile acid deoxycholic acid, K02835 (*prfA*), which positively regulates the expression of a gene coding for bile acid hydrolase, and K03527 (*ispH*, *lytB*), a gene involved in bile acid sensitivity. Additionally, *Bacteroides vulgatus* contributed to the expression of all these microbiome genes, which aligns with previous studies showing an increase in *Bacteroides* spp. correlating with the expression of the host gene *NR1H3* (49).

DISCUSSION

In this study, we extracted gene modules from a host and microbiome gene co-expression network based on dual transcriptome profiling. The module extraction method proposed in this study is based on clique finding and takes into account the structure of three types of networks: host-host, host-microbiome, and microbiome-microbiome. This approach enables the extraction of modules containing both host genes and microbiome genes. Conventional module extraction methods do not differentiate between host gene nodes and microbiome gene nodes within the network, potentially resulting in modules composed exclusively of either host genes or microbiome genes. Indeed, when conventional module extraction methods such as the Newman algorithm (15), the Louvain algorithm (16), the Leiden algorithm (17), and Weighted Correlation Network Analysis (WGCNA) (18) were applied to our dataset (Supplementary Note Section 9), we observed that these modules included those composed solely of host genes or microbiome genes (Table S5). Notably, about half of the genes in WGCNA modules were found in modules comprising only one species. While the purpose of module extraction is to interpret the

344 interactions among genes within a module, these single-species modules do not allow for an
345 interpretation of the relationship between host genes and microbiome genes. Furthermore, to
346 verify whether functionally similar genes were being extracted, we conducted enrichment
347 analysis on the extracted modules (Supplementary Note Section 8). The results showed that our
348 method had the highest proportion of functionally characterized modules (enriched modules)
349 (Table S6).

350 The gene modules extracted, which encompassed both host and microbiome genes,
351 frequently exhibited relationships where a small number of genes regulated a larger number of
352 genes between the host and microbiome genes. These observations are thought to reflect the
353 characteristics of the interactions between the host and the microbiome. Several studies have
354 reported the impact of host gene mutations on the intestinal microbiome. For instance, it has
355 been reported that the host gene *VDR*, which codes for the Vitamin D receptor, influences the
356 colonization of *Parabacteroides* (50), and that the host gene *ABO* is associated with
357 *Faecalicatena lactaris* (51). In another study investigating the relationship between the intestinal
358 microbiome and host gene loci through quantitative trait locus analysis, it was suggested that
359 *Akkermansia muciniphila* regulates the expression of *ATF3*, which codes for a transcription
360 factor that plays a crucial role in metabolic and immune regulation, and this bacterium is related
361 to only four genes (52). These studies highlight the fact that a single or a small number of host
362 genes are controlling, or are controlled by, the intestinal microbiome. Conversely, a small
363 number of microbiome genes can affect the host's metabolic (53), fitness (54), and longevity (55)
364 phenotypes. For example, structural mutations in the gene related to butyrate transport in
365 *Anaerostipes hadrus* are strongly associated with a reduction in the host's metabolic risk (53),
366 and the intestinal microbiome's thiamine biosynthesis genes, transferred by horizontal

367 transmission, influence the host's fitness (54). Considering these previous studies and our results,
368 it is suggested that the interaction between the host and the microbiome involves widespread
369 influence of one's genes on the activity of the other.

370 The common marmoset used in this study bears anatomical and pharmacological
371 similarities to humans (19), and the fecal microbiome of the common marmoset is more similar
372 to that of humans compared to other animal models (13). While the common marmoset's
373 cecum—one of the study's targeted intestinal sites—does not possess an appendage as
374 pronounced as the human appendix, it is known to share significant similarities with the human
375 appendix in terms of microanatomical structure and immune reactivity (56, 57). The results of
376 this study using common marmosets could therefore hold higher applicability to humans
377 compared to other animal models.

378 In conclusion, we generated host transcriptome and microbiome metatranscriptome
379 profiles of the common marmoset and performed gene co-expression network analysis. The
380 algorithm we proposed for extracting gene modules in the gene co-expression network enabled
381 the identification of dense, functionally related host-microbiome co-expression modules, which
382 was challenging using the conventional module extraction algorithms. The identified gene
383 modules provided new insights into relationships between host genes and intestinal microbiota
384 and microbial genes bridging interactions between the host and bacteria. Our findings underscore
385 that interactions between the host and microbiome are not uniform throughout the colon but are
386 localized, indicating spatial variation in host-microbiome interactions along the colon. These
387 insights could be crucial in understanding mechanisms behind localized gut diseases, such as
388 colorectal cancer and inflammatory bowel disease, and in developing localized therapies (58, 59,
389 60, 61).

390

391 **MATERIALS AND METHODS**

392 ***Animal experiments***

393 Common marmosets were housed at the Central Institute for Experimental Animals (Kawasaki,
394 Japan). Five marmosets were euthanized with intravenous administration of pentobarbital
395 overdose, and the digestive tract was isolated (Table S7, Supplementary Note Section 1). Three
396 gastrointestinal site samples (cecum, transverse colon and rectum), including the mucosal layer,
397 were used for dual RNA-seq. These three sites were targeted at the beginning, middle and end of
398 the colon, which has an abundant microbiome. The animal experiment protocol was approved by
399 the CIEA Institutional Animal Care and Use Committee (approval no. 17031, 18032, and 21084).
400 The study was conducted in accordance with the guidelines of CIEA that comply with the
401 Guidelines for Proper Conduct of Animal Experiments published by the Science Council of
402 Japan. Animal care was conducted in accordance with the Guide for the Care and Use of
403 Laboratory Animals (Institute for Laboratory Animal Resources, 2011).

404

405 ***RNA extraction and dual RNA sequencing of host and microbiome***

406 RNA libraries were prepared after removing rRNA from the extracted high-quality Total RNA
407 (RNA integrity number (RIN) scores ≥8.6; Table S1). Illumina NovaSeq 6000 v1.5 sequencing
408 yielded a total of 472 Gnt of paired-end reads (100 nt × 2). This dataset included an average of
409 314.9 M reads per sample (Table S2 and S8; Supplementary Note Section 2).

410

411 ***Construction of host-microbiome co-expression network and identification of gene modules***

412 In this study, we utilized gene co-expression network analysis to identify interacting gene
413 modules of host and microbiome (Supplementary Note Section 7). We established three types of
414 networks: host-host, host-microbiome, and microbiome-microbiome, wherein genes were
415 considered as nodes and co-expressing genes were interconnected (Fig. 1D). Networks were
416 defined by an adjacency matrix, where the adjacency a_{ij} between gene i and gene j was defined
417 as follows based on co-expression similarity s_{ij}:

418 $$a_{ij} = \begin{cases} 1\,, \text{if } s_{ij} > \alpha \\ 0\,, \text{otherwise} \end{cases}$$

419 The co-expression similarity s_{ij} was computed using Spearman's correlation coefficients for
420 host-host and microbiome-microbiome networks, and absolute values of Spearman's correlation
421 coefficients for host-microbiome networks. The adjacency threshold α was set at 0.80. These
422 three types of networks were then integrated into one.

423 We developed an algorithm to identify densely interacting gene modules from the host-
424 microbiome co-expression network. This algorithm, based on clique finding, progressively
425 merges cliques (Fig. 1E). In graph theory, a "clique" refers to a subgraph within the network
426 where all nodes are fully connected. In the context of a gene co-expression network, all genes
427 within a clique indicate a strong co-expression (correlation) amongst them. In this study, we
428 focus only on networks and cliques that include gene nodes from both the host and microbiome,
429 encompassing three types of networks: host-host, host-microbiome, and microbiome-
430 microbiome. Our algorithm identified modules of interacting genes in the gene co-expression
431 network as aggregates of overlapping cliques. It was implemented in three steps: (1) Cliques
432 within the network were found; (2) Beginning with the largest cliques, we examined their
433 overlap with any other cliques. If overlap occurred, we computed the cliqueness scores
434 independently for each of the three types of networks (host-host, host-microbiome, and

435 microbiome-microbiome networks) upon merging the overlapping cliques. The cliqueness scores
436 for the host-host, host-microbiome, and microbiome-microbiome networks were denoted as C_{hh},
437 C_{hm}, and C_{mm}, respectively; (3) Cliques were merged if they satisfied the condition: $C_{hh} > \tau$,
438 $C_{mm} > \tau$, and $C_{mm} > \tau$. The threshold τ for cliqueness was set at 0.60 based on parameter
439 determination (Supplementary Note Section 7). If there were multiple cliques with minimum
440 cliqueness scores exceeding the threshold, the clique with the highest cliqueness score was
441 selected. Steps (2) and (3) were repeated until no cliques could be merged.
442 For the host-host networks the cliqueness score C_{hh} between cliques A and B was defined as:

443 $$C_{hh} = \frac{E}{\frac{1}{2}N(N-1)}$$

444 Here, N is defined as the number of host gene nodes in the union of cliques A and B, represented
445 as $N = |N_A \cup N_B|$. N_A and N_B denote the number of host gene nodes within clique A and clique
446 B, respectively. E corresponds to the total number of edges between the host nodes included in
447 cliques A and/or B, formulated as $E = E_A + E_B + E_{AB}$. Here, E_A and E_B represent the number of
448 edges between host gene nodes within clique A and clique B, respectively, while E_{AB} denotes the
449 number of edges connecting across host gene nodes within clique A and clique B. The cliqueness
450 score for microbiome-microbiome networks C_{mm} was calculated in the same way.
451 For the host-microbiome network, the cliqueness score C_{hm}, was defined as:

452 $$C_{hm} = \frac{E}{N_{host}N_{microbiome}}$$

453 where N_{host} and $N_{microbiome}$ represent the number of host and microbiome gene nodes
454 respectively in the union of cliques A and B. E corresponds to the total number of edges between
455 the host or microbiome gene nodes included in cliques A and/or B.
456

457 ***Data and source code availability***

458 All raw sequence data have been submitted to the DDBJ under project PSUB020951 from the
459 Ministry of Education, Culture, Sports, Science and Technology of Japan. The gene expression
460 profiles of the host and microbiome are available at https://doi.org/10.5281/zenodo.8264866. The
461 algorithm for module extraction is available at https://github.com/MikaUhr/MEIGN.git.

462

463 **ACKNOWLEDGEMENTS**

464 We thank the scientists of the Central Institute for Experimental Animals for providing support for
465 animal experiments and the members of the Sakakibara Lab at Keio University for helpful
466 discussions.

467 This work was supported by grants from the Japan Agency for Medical Research and Development
468 (AMED PRIME) Grant Number JP19gm6010006 and JST, CREST Grant Number JPMJCR20S3,
469 Japan. M. Uehara has received funding from JSPS KAKENHI Grant Numbers JP20J21477.

470

471 M.U. performed the experiments, conducted the bioinformatics analysis and cowrote the paper;
472 T.I. and E.S. provided common marmoset samples; M.U. and S.H. performed RNA extraction; A.T.
473 performed deep sequencing with high-throughput sequencer; Y.S. designed and supervised the
474 research, analysed the data, and cowrote the paper. All authors have read and approved the
475 manuscript.

476

477 The authors declare no competing interests.

478

REFERENCES

1. Yoon HS, Cho CH, Yun MS, Jang SJ, You HJ, Kim JH, et al. 2021. Akkermansia muciniphila secretes a glucagon-like peptide-1-inducing protein that improves glucose homeostasis and ameliorates metabolic disease in mice. Nat Microbiol 6:563–73.
2. Vatanen T, Kostic AD, D'Hennezel E, Siljander H, Franzosa EA, Yassour M, et al. 2016. Variation in Microbiome LPS Immunogenicity Contributes to Autoimmunity in Humans. Cell 165:842–53.
3. Song X, Sun X, Oh SF, Wu M, Zhang Y, Zheng W, et al. 2020. Microbial bile acid metabolites modulate gut RORγ+ regulatory T cell homeostasis. Nature 577:410–5.
4. Zhang L, Forst CV, Gordon A, Gussin G, Geber AB, Fernandez PJ, et al. 2020. Characterization of antibiotic resistance and host-microbiome interactions in the human upper respiratory tract during influenza infection. Microbiome 8:1–12.
5. Fujiogi M, Raita Y, Pérez-Losada M, Freishtat RJ, Celedón JC, Mansbach JM, et al. 2022. Integrated relationship of nasopharyngeal airway host response and microbiome associates with bronchiolitis severity. Nat Commun 13:4970.
6. Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon TW, et al. 2019. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature 569:655–62.
7. Priya S, Burns MB, Ward T, Mars RAT, Adamowicz B, Lock EF, et al. 2022. Identification of shared and disease-specific host gene–microbiome associations across human diseases using multi-omic integration. Nat Microbiol 7:780–95.
8. Cremer J, Arnoldini M, Hwa T. 2017. Effect of water flow and chemical environment on microbiota growth and composition in the human colon. Proc Natl Acad Sci 114:6438-43.

502 9. Dupont HL, Jiang ZD, Dupont AW, Utay NS. 2020. The intestinal microbiome in human
503 health and disease. Trans Am Clin Climatol Assoc 131:178.
504 10. Hu S, Wang Y, Lichtenstein L, Tao Y, Musch MW, et al. 2010. Regional differences in
505 colonic mucosa-associated microbiota determine the physiological expression of host
506 heat shock proteins. Am J Physiol Gastrointest Liver Physiol 299:G1266-75.
507 11. Yasuda K, Oh K, Ren B, Tickle TL, Franzosa EA, Wachtman LM, et al. 2015.
508 Biogeography of the intestinal mucosal and lumenal microbiome in the rhesus macaque.
509 Cell Host Microbe 17:385–91.
510 12. Kozik AJ, Nakatsu CH, Chun H, Jones-Hall YL. 2019. Comparison of the fecal, cecal,
511 and mucus microbiome in male and female mice after TNBS-induced colitis. PLoS One
512 14:1–14.
513 13. Uehara M, Inoue T, Kominato M, Hase S, Sasaki E, Toyoda A, et al. 2022. Intraintestinal
514 Analysis of the Functional Activity of Microbiomes and Its Application to the Common
515 Marmoset Intestine. mSystems 7:e00520-22.
516 14. Newman ME. 2004. Detecting community structure in networks. Eur Phys J B 38:321–
517 30.
518 15. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. 2008. Fast unfolding of
519 communities in large networks. J Stat Mech 10008.
520 16. Traag VA, Waltman L, Van Eck NJ. 2019. From Louvain to Leiden: guaranteeing well-
521 connected communities. Sci Rep 9:5233.
522 17. Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network
523 analysis. BMC bioinformatics 9:1-3.

524 18. Hugenholtz F, de Vos WM. 2018. Mouse models for human intestinal microbiota
525 research: a critical evaluation. Cell Mol Life Sci 75:149–60.
526 19. Kishi N, Sato K, Sasaki E, Okano H. 2014. Common marmoset as a new model animal
527 for neuroscience research and genome editing technology. Dev Growth Differ 56:53–62.
528 20. Inoue T, Okahara N, Okahashi N, Sato K, Ueda M, Atarashi K, et al. 2021. Generation of
529 germ-free common marmosets. Research Square. Preprint.
530 https://assets.researchsquare.com/files/rs-428622/v1_covered.pdf?c=1631871455
531 21. Stuart JM, Segal E, Koller D, Kim SK. 2003. A gene-coexpression network for global
532 discovery of conserved genetic modules. Science 302:249–255.
533 22. Wang K, Wei G, Liu D. 2012. CD19: a biomarker for B cell development, lymphoma
534 diagnosis and therapy. Exp Hematol Oncol 1:1–7.
535 23. Huse K, Bai B, Hilden VI, Bollum LK, Våtsveen TK, Munthe LA, et al. 2022.
536 Mechanism of CD79A and CD79B support for IgM+ B cell fitness through B cell
537 receptor surface expression. J Immunol 209:2042–53.
538 24. Clark EA, Giltiay N V. 2018. CD22: a regulator of innate and adaptive B cell responses
539 and autoimmunity. Front Immunol 9:2235.
540 25. Pan C, Baumgarth N, Parnes JR. 1999. CD72-deficient mice reveal nonredundant roles of
541 CD72 in B cell development and activation. Immunity 11:495–506.
542 26. Alegre ML, Frauwirth KA, Thompson CB. 2001. T-cell regulation by CD28 and CTLA-
543 4. Nat Rev Immunol 1:220–8.
544 27. Yang Y, Zang Y, Zheng C, Li Z, Gu X, Zhou M, et al. 2020. CD3D is associated with
545 immune checkpoints and predicts favorable clinical outcome in colon cancer.
546 Immunotherapy 12:25–35.

547 28. Hubbard N, Hagin D, Sommer K, Song Y, Khan I, Clough C, et al. 2016. Targeted gene
548 editing restores regulated CD40L function in X-linked hyper-IgM syndrome. Blood
549 127:2513–22.
550 29. Sobhani N, Tardiel-Cyril DR, Davtyan A, Generali D, Roudi R, Li Y. 2021. CTLA-4 in
551 regulatory T cells for cancer immunotherapy. Cancers 13:1–18.
552 30. He Y, Hua R, Li B, Gu H, Sun Y, Li Z. 2021. Loss of FBP1 promotes proliferation,
553 migration, and invasion by regulating fatty acid metabolism in esophageal squamous cell
554 carcinoma. Aging 13:4986–8.
555 31. Hirata H, Sugimachi K, Komatsu H, Ueda M, Masuda T, Uchi R, et al. 2016. Decreased
556 expression of fructose-1,6-bisphosphatase associates with glucose metabolism and tumor
557 progression in hepatocellular carcinoma. Cancer Res 76:3265–76.
558 32. Dong Y, Huaying S, Danying W, Chihong Z, Ruibin J, Xiaojiang S, et al. 2018.
559 Significance of methylation of FBP1 gene in non-small cell lung cancer. Biomed Res Int
560 2018:13–9.
561 33. Li H, Qi Z, Niu Y, Yang Y, Li M, Pang Y, et al. 2021. FBP1 regulates proliferation,
562 metastasis, and chemoresistance by participating in C-MYC/STAT3 signaling axis in
563 ovarian cancer. Oncogene 40:5938–49.
564 34. Hou Y, Wei W, Guan X, Liu Y, Bian G, He D, et al. 2021. A diet-microbial metabolism
565 feedforward loop modulates intestinal stem cell renewal in the stressed gut. Nat Commun
566 12:271.
567 35. Mardinoglu A, Shoaie S, Bergentall M, Ghaffari P, Zhang C, Larsson E, et al. 2015. The
568 gut microbiota modulates host amino acid and glutathione metabolism in mice. Mol Syst
569 Biol 11:1–15.

570 36. Medina DA, Pinto F, Ovalle A, Thomson P, Garrido D. 2017. Prebiotics mediate
571 microbial interactions in a consortium of the infant gut microbiome. Int J Mol Sci
572 18:2095.
573 37. Cheng J, Hu J, Geng F, Nie S. 2022. Bacteroides utilization for dietary polysaccharides
574 and their beneficial effects on gut health. Food Sci Hum Wellness 11:1101–10.
575 38. Rosser EC, Piper CJM, Matei DE, Blair PA, Rendeiro AF, Orford M, et al. 2020.
576 Microbiota-derived metabolites suppress arthritis by amplifying aryl-hydrocarbon
577 receptor activation in regulatory B cells. Cell Metab 31:837-851.
578 39. Ding X, Bin P, Wu W, Chang Y, Zhu G. 2020. Tryptophan metabolism, regulatory T cells,
579 and inflammatory bowel disease: a mini review. Mediators Inflamm 2020.
580 40. Ezeji JC, Sarikonda DK, Hopperton A, Erkkila HL, Cohen DE, Martinez SP, et al. 2021.
581 Parabacteroides distasonis: intriguing aerotolerant gut anaerobe with emerging
582 antimicrobial resistance and pathogenic and probiotic roles in human health. Gut
583 Microbes 13:1–27.
584 41. Cekanaviciute E, Yoo BB, Runia TF, Debelius JW, Singh S, Nelson CA, et al. 2017. Gut
585 bacteria from multiple sclerosis patients modulate human T cells and exacerbate
586 symptoms in mouse models. Proc Natl Acad Sci USA 114:10713–8.
587 42. Łaniewski P, Herbst-Kralovetz MM. 2021. Bacterial vaginosis and health-associated
588 bacteria modulate the immunometabolic landscape in 3D model of human cervix. NPJ
589 Biofilms Microbiomes 7:1–17.
590 43. Tomkovich S, Gharaibeh RZ, Dejea CM, Pope JL, Jiang J, Winglee K, et al. 2020.
591 Human colon mucosal biofilms and murine host communicate via altered mRNA and
592 microRNA expression during cancer. mSystems 5:e00451-19.

593 44. Cornick S, Tawiah A, Chadee K. 2015. Roles and regulation of the mucus barrier in the
594 gut. Tissue Barriers 3:1–2.
595 45. González-Morelo KJ, Vega-Sagardía M, Garrido D. Molecular insights into O-linked
596 glycan utilization by gut microbes. Frontiers in Microbiology. 2020 Nov 5;11:591568.
597 46. Png CW, Lindén SK, Gilshenan KS, Zoetendal EG, McSweeney CS, Sly LI, McGuckin
598 MA, Florin TH. Mucolytic bacteria with increased prevalence in IBD mucosa augmentin
599 vitroutilization of mucin by other bacteria. Official journal of the American College of
600 Gastroenterology| ACG. 2010 Nov 1;105(11):2420-8.
601 47. Zurkinden L, Sviridov D, Vogt B, Escher G. 2020. Downregulation of Cyp7a1 by cholic
602 acid and chenodeoxycholic acid in Cyp27a1/ApoE double knockout mice: differential
603 cardiovascular outcome. Front Endocrinol 11:1–13.
604 48. Biagioli M, Marchianò S, Carino A, Di Giorgio C, Santucci L, Distrutti E, et al. 2021.
605 Bile acids activated receptors in inflammatory bowel disease. Cells 10:1281.
606 49. Badi SA, Motahhary A, Bahramali G, Masoumi M, Khalili SFS, Ebrahimzadeh N, et al.
607 2020. The regulation of Niemann-Pick C1-Like 1 (NPC1L1) gene expression in opposite
608 direction by Bacteroides spp. and related outer membrane vesicles in Caco-2 cell line. J
609 Diabetes Metab Disord 19:415–22.
610 50. Wang J, Thingholm LB, Skiecevičie J, Rausch P, Kummen M, Hov JR, et al. 2016.
611 Genome-wide association analysis identifies variation in Vitamin D receptor and other
612 host factors influencing the gut microbiota. Nat Genet 48:1396–406.
613 51. Qin Y, Havulinna AS, Liu Y, Jousilahti P, Ritchie SC, Tokolyi A, et al. 2022. Combined
614 effects of host genetics and diet on human gut microbiota and incident disease in a single
615 population cohort. Nat Genet 54:134–42.

616 52. Zhang Q, Linke V, Overmyer KA, Traeger LL, Kasahara K, Miller IJ, et al. 2023. Genetic
617 mapping of microbial and host traits reveals production of immunomodulatory lipids by
618 Akkermansia muciniphila in the murine gut. Nat Microbiol 8:424–40.
619 53. Zeevi D, Korem T, Godneva A, Bar N, Kurilshikov A, Lotan-Pompan M, et al. 2019.
620 Structural variation in the gut microbiome associates with host health. Nature 568:43–8.
621 54. Wang Y, Baumdicker F, Schweiger P, Kuenzel S, Staubach F. 2021. Horizontal gene
622 transfer-mediated bacterial strain variation affects host fitness in Drosophila. BMC Biol
623 19:1-12.
624 55. Han B, Sivaramakrishnan P, Lin CCJ, Neve IAA, He J, Tay LWR, et al. 2017. Microbial
625 genetic composition tunes host longevity. Cell 169:1249-1262.
626 56. Zinnen AD, Vichich J, Metzger JM, Gambardella JC, Bondarenko V, Simmons HA, et al.
627 2022. Alpha-synuclein and tau are abundantly expressed in the ENS of the human
628 appendix and monkey cecum. PLoS One 17:e0269190.
629 57. Fisher RE. The primate appendix: A reassessment. Vol. 261, Anatomical Record. 2000. pp
630 228–36.
631 58. Debelius JW, Engstrand L, Matussek A, Brusselaers N, Morton JT, Stenmarker M,et al.
632 2023. The local tumor microbiome is associated with survival in late-stage colorectal
633 cancer patients. Microbiol Spectr. 2023 11:e05066-22.
634 59. Drewes JL, White JR, Dejea CM, Fathi P, Iyadorai T, Vadivelu J, et al. 2017. High-
635 resolution bacterial 16S rRNA gene profile meta-analysis and biofilm status reveal
636 common colorectal cancer consortia. NPJ Biofilms Microbiomes 3:34.

60. Čipčić Paljetak H, Barešić A, Panek M, Perić M, Matijašić M, Lojkić I, et al. 2022. Gut microbiota in mucosa and feces of newly diagnosed, treatment-naïve adult inflammatory bowel disease and irritable bowel syndrome patients. Gut Microbes 14:2083419.
61. Yu LC. 2018. Microbiota dysbiosis and barrier dysfunction in inflammatory bowel disease and colorectal cancers: exploring a common ground hypothesis. J Biomed Sci. 25:1-4.

Fig. 1 Overview of the main analytical approach used in this study.

This method involves the simultaneous acquisition of the host and microbiome transcriptome (metatranscriptome) from intestinal sites and conducts a gene co-expression network analysis of the host and microbiome community. The procedure is outlined from step A to E. Perform dual RNA-seq of host and microbiome transcriptomes in internal site samples. (A) Separate host and microbiome RNA reads in silico using the host reference genome. (B) Compute the gene expression profiles of the host and microbiome, respectively. (C) Unify gene co-expression networks between host-host, host-microbiome, and microbiome-microbiome. The host-host and microbiome-microbiome gene co-expression networks are connected when there is a positive correlation between genes, while the host-microbiome gene co-expression network is connected when there is either a positive or negative correlation between genes. (D) Identify closely interacting gene modules from the gene co-expression network using a proposed clique-based algorithm. (E Step 1) This process begins with searching for cliques. (E Step 2) If overlaps exist between cliques, the cliqueness score is computed in three gene co-expression networks (host-host, host-microbiome, and microbiome-microbiome) when overlapping cliques are merged. (E Step 3) If the cliqueness score exceeds the merging threshold τ in all three networks, the cliques

660 are merged into a single module. If the score does not exceed the merging threshold τ in at least
661 one network, the original cliques are retained without merging. This process is repeated until no
662 further modules can be merged.
663
664 Fig. 2 PCA analysis of gene expression data for host and microbiome
665 PCA was applied to the host gene expression profiles to visualize groupings according to site (A)
666 and individual (B). PCA was similarly applied to the microbiome gene expression profiles to
667 visualize groupings according to site (C) and individual (D).
668
669 Fig. 3 Variations in host gene expression and functional pathways among intestinal sites
670 (A) The top 20 functional pathways significantly upregulated in the cecum compared to the
671 transverse colon (normalized enrichment score > 1.5, FDR < 0.1), as revealed by gene set
672 enrichment analysis, are presented alongside their normalized enrichment score and FDR. (B)
673 Heatmap of differentially expressed genes belonging to the top 5 enriched functional pathways.
674 The heatmap on the left represents the z-scores of gene expression levels, with genes
675 significantly upregulated in the cecum compared to the transverse colon and rectum (FDR $<$
676 0.01) colored in pink on the middle heatmap. The heatmap on the right indicates the functional
677 pathways to which the genes belong, represented in green.
678
679 Fig. 4 Functional pathways characterized in modules extracted from the host-microbiome gene
680 co-expression network
681 This figure displays the identified modules along with the results of their functional pathway
682 enrichment analysis (Supplementary Section 8). To enhance clarity, the background of the bar

683 graphs alternates between gray and white for each module. Modules were characterized using
684 KEGG pathways, with those meeting an FDR < 0.05 as determined by Fisher's exact test. The
685 dashed line represents an FDR < 0.05 (-log10(FDR) > 1.30). All pathways with an FDR < 0.1
686 are represented, and modules with no identified function are denoted as "None".
687
688 Fig. 5 Metabolic pathway-enriched Module_07
689 A module showing a positive correlation between the host gene *FBP1*, which is highly expressed
690 in the transverse colon, and 40 microbiome genes, including those involved in fructose-mannose
691 metabolism and oxidative phosphorylation, was extracted. These genes were visualized
692 alongside bacterial species nodes that contribute to the expression of the microbiome genes. Note
693 that the bacterial species nodes were added after the module detection analysis. Edges among
694 genes indicate an absolute Spearman's correlation coefficient greater than 0.8. Edges connecting
695 microbiome genes to bacterial species are shown when the contribution rate of bacterial species
696 to microbiome gene expression exceeds 0.2. The thickness of the edge increases with higher
697 contribution rates.
698
699 Fig. 6 B cell receptor signaling pathway-enriched Module_01
700 Host genes involved in the immune response of B cells, including *CD19*, *CD22*, *CD79B* and
701 *PTPN6*, which were highly expressed in the cecum, co-expressed with the K01696 (*trpB*). These
702 genes were visualized alongside bacterial species nodes that contribute to the expression of the
703 microbiome genes. K01696 (*trpB*) expression was contributed to by *Parabacteroides distasonis*.
704 Edges among genes indicate an absolute Spearman's correlation coefficient greater than 0.8.
705 Edges connecting microbiome genes to bacterial species are shown when the contribution rate of

706 bacterial species to microbiome gene expression exceeds 0.2. Note that the bacterial species
707 nodes were added after the module detection analysis. The thickness of the edge increases with
708 higher contribution rates.
709
710 Fig. 7 Carbon metabolism pathway-enriched Module_21
711 Host genes, including *MMP10*, which is specifically highly expressed in the cecum, and 40
712 microbiome genes involved in carbon metabolism, were identified. These genes were visualized
713 alongside bacterial species nodes that contribute to the expression of the microbiome genes. Note
714 that the bacterial species nodes were added after the module detection analysis. Edges among
715 genes indicate an absolute Spearman's correlation coefficient greater than 0.8. Edges connecting
716 microbiome genes to bacterial species are shown when the contribution rate of bacterial species
717 to microbiome gene expression exceeds 0.2. The thickness of the edge increases with higher
718 contribution rates.
719
720 Fig. 8 Module_18 containing host genes and microbiome genes related to the production and
721 breakdown of bile acids. These genes were visualized alongside bacterial species nodes that
722 contribute to the expression of the microbiome genes. Note that the bacterial species nodes were
723 added after the module detection analysis. Edges among genes indicate an absolute Spearman's
724 correlation coefficient greater than 0.8. Edges connecting microbiome genes to bacterial species
725 are shown when the contribution rate of bacterial species to microbiome gene expression exceeds
726 0.2. The thickness of the edge increases with higher contribution rates.

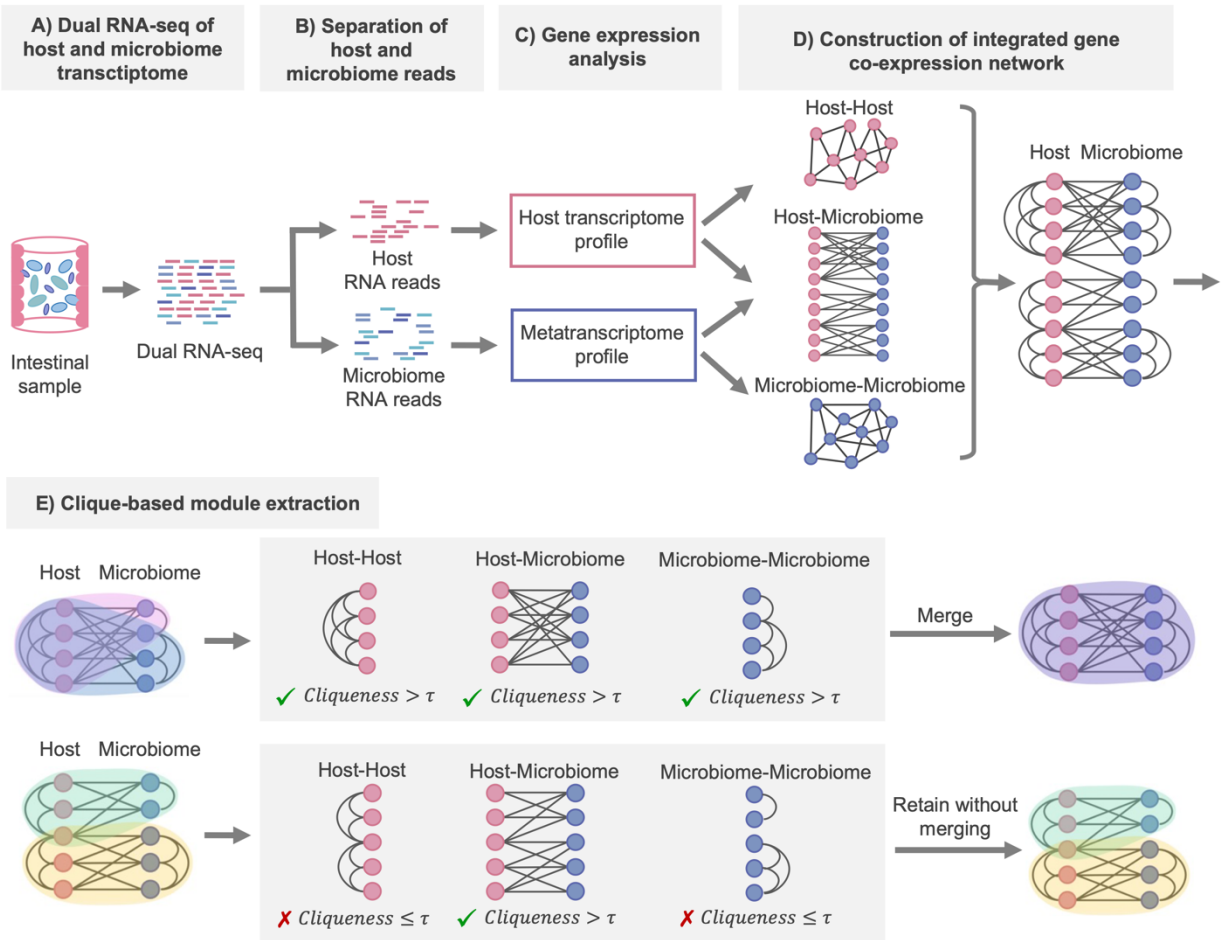

Fig. 1 Overview of the main analytical approach used in this study.
This method involves the simultaneous acquisition of the host and microbiome transcriptome (metatranscriptome) from intestinal sites and conducts a gene co-expression network analysis of the host and microbiome community. The procedure is outlined from step A to E. Perform dual RNA-seq of host and microbiome transcriptomes in internal site samples. (A) Separate host and microbiome RNA reads in silico using the host reference genome. (B) Compute the gene expression profiles of the host and microbiome, respectively. (C) Unify gene co-expression networks between host-host, host-microbiome, and microbiome-microbiome. The host-host and microbiome-microbiome gene co-expression networks are connected when there is a positive correlation between genes, while the host-microbiome gene co-expression network is connected when there is either a positive or negative correlation between genes. (D) Identify closely interacting gene modules from the gene co-expression network using a proposed clique-based algorithm. (E Step 1) This process begins with searching for cliques. (E Step 2) If overlaps exist between cliques, the cliqueness score is computed in three gene co-expression networks (host-host, host-microbiome, and microbiome-microbiome) when overlapping cliques are merged. (E Step 3) If the cliqueness score exceeds the merging threshold τ in all three networks, the cliques are merged into a single module. If the score does not exceed the merging threshold τ in at least one network, the original cliques are retained without merging. This process is repeated until no further modules can be merged.

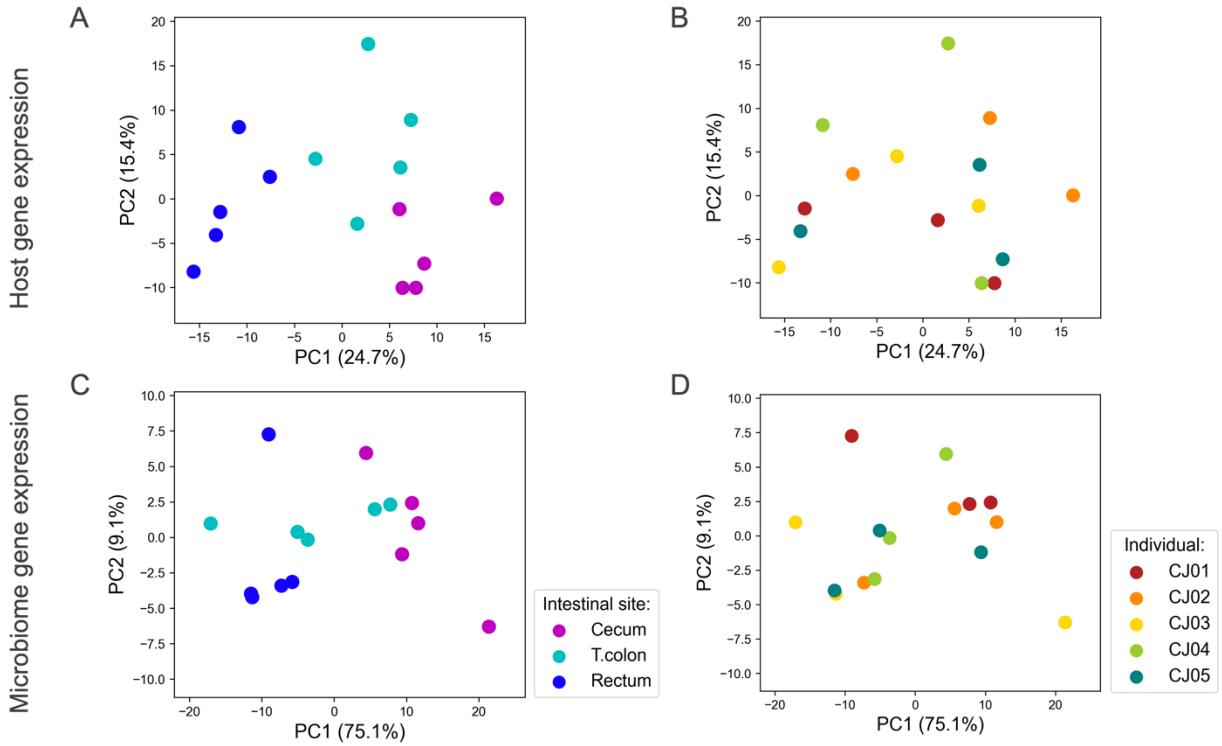

Fig. 2 PCA analysis of gene expression data for host and microbiome
PCA was applied to the host gene expression profiles to visualize groupings according to site (A) and individual (B). PCA was similarly applied to the microbiome gene expression profiles to visualize groupings according to site (C) and individual (D).

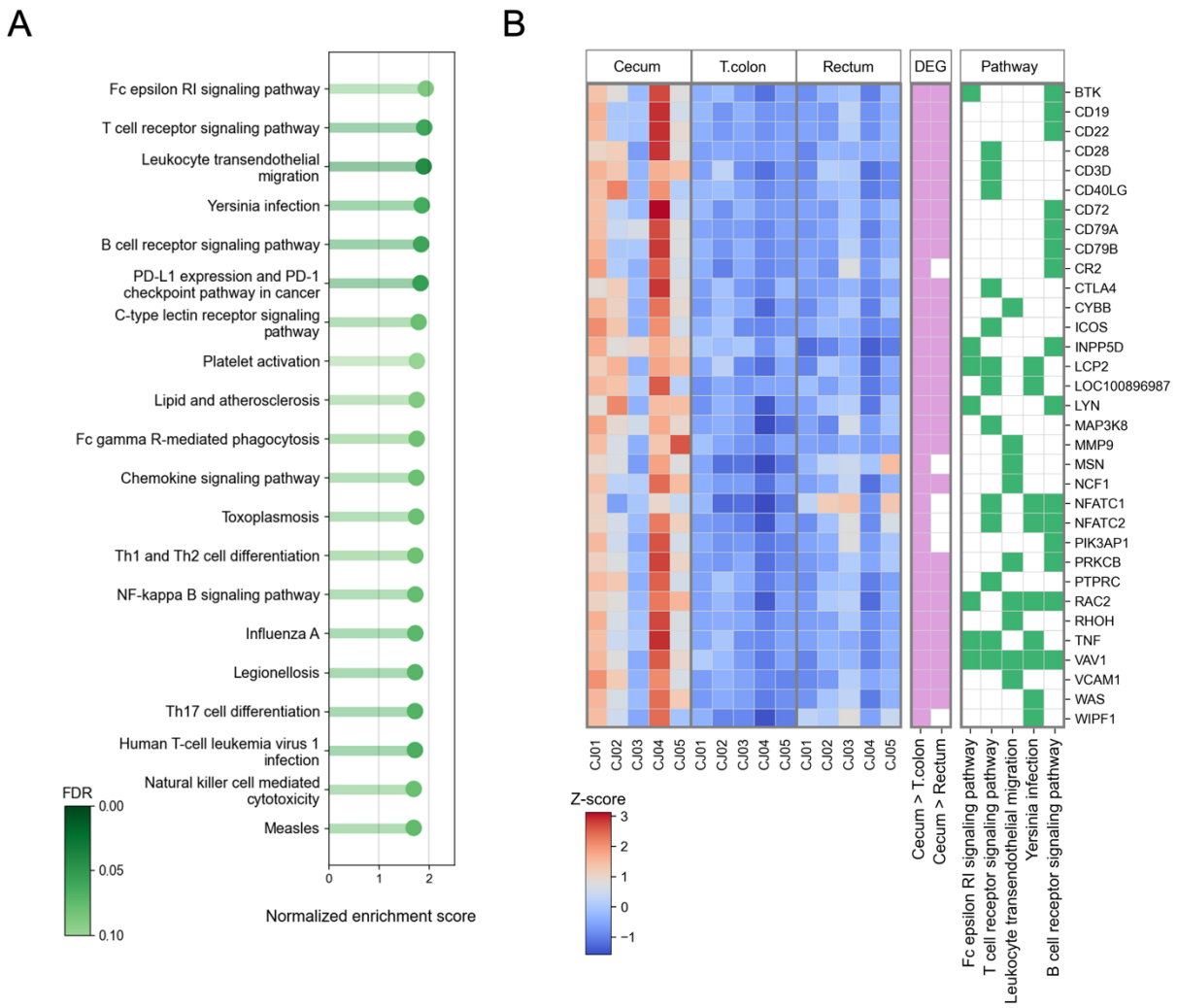

Fig. 3 Variations in host gene expression and functional pathways among intestinal sites
(A) The top 20 functional pathways significantly upregulated in the cecum compared to the transverse colon (normalized enrichment score > 1.5, FDR < 0.1), as revealed by gene set enrichment analysis, are presented alongside their normalized enrichment score and FDR. (B) Heatmap of differentially expressed genes belonging to the top 5 enriched functional pathways. The heatmap on the left represents the z-scores of gene expression levels, with genes significantly upregulated in the cecum compared to the transverse colon and rectum (FDR < 0.01) colored in pink on the middle heatmap. The heatmap on the right indicates the functional pathways to which the genes belong, represented in green.

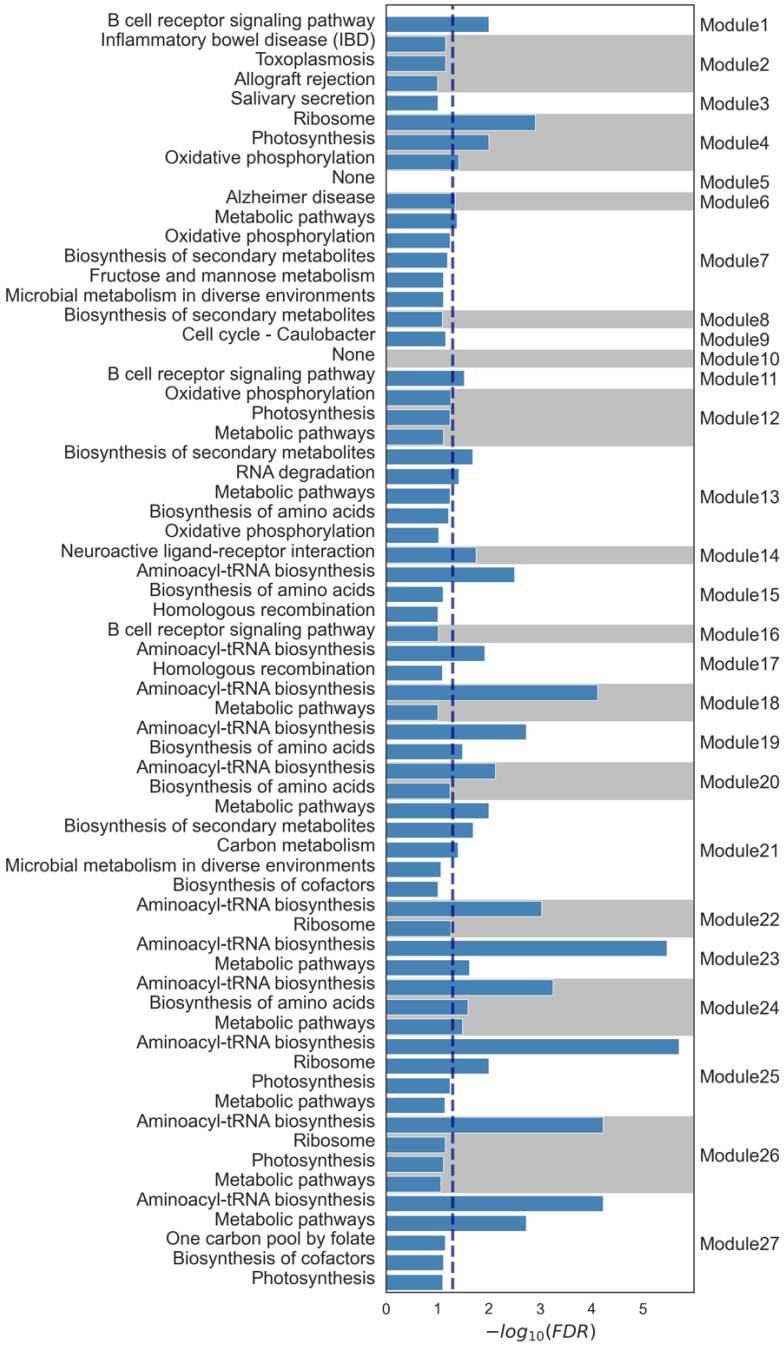

Fig. 4 Functional pathways characterized in modules extracted from the host-microbiome gene co-expression network

This figure displays the identified modules along with the results of their functional pathway enrichment analysis (Supplementary Section 8). To enhance clarity, the background of the bar graphs alternates between gray and white for each module. Modules were characterized using KEGG pathways, with those meeting an FDR < 0.05 as determined by Fisher's exact test. The dashed line represents an FDR < 0.05 (-log10(FDR) > 1.30). All pathways with an FDR < 0.1 are represented, and modules with no identified function are denoted as “None”.

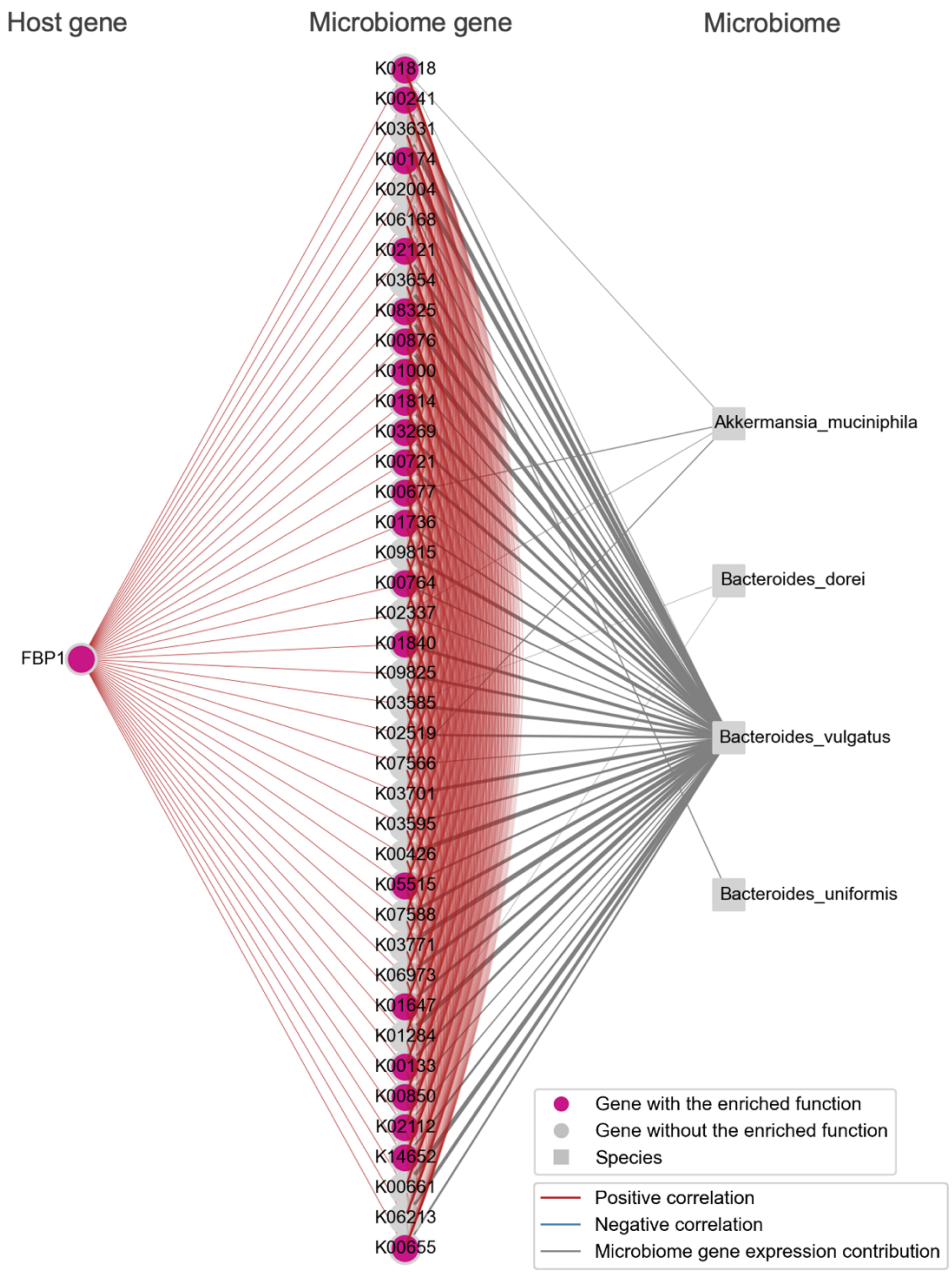

Fig. 5 Metabolic pathway-enriched Module_07

A module showing a positive correlation between the host gene *FBP1*, which is highly expressed in the transverse colon, and 40 microbiome genes, including those involved in fructose-mannose metabolism and oxidative phosphorylation, was extracted. These genes were visualized alongside bacterial species nodes that contribute to the expression of the microbiome genes. Note that the bacterial species nodes were added after the module detection analysis. Edges among genes indicate an absolute Spearman's correlation coefficient greater than 0.8. Edges connecting microbiome genes to bacterial species are shown when the contribution rate of bacterial species to microbiome gene expression exceeds 0.2. The thickness of the edge increases with higher contribution rates.

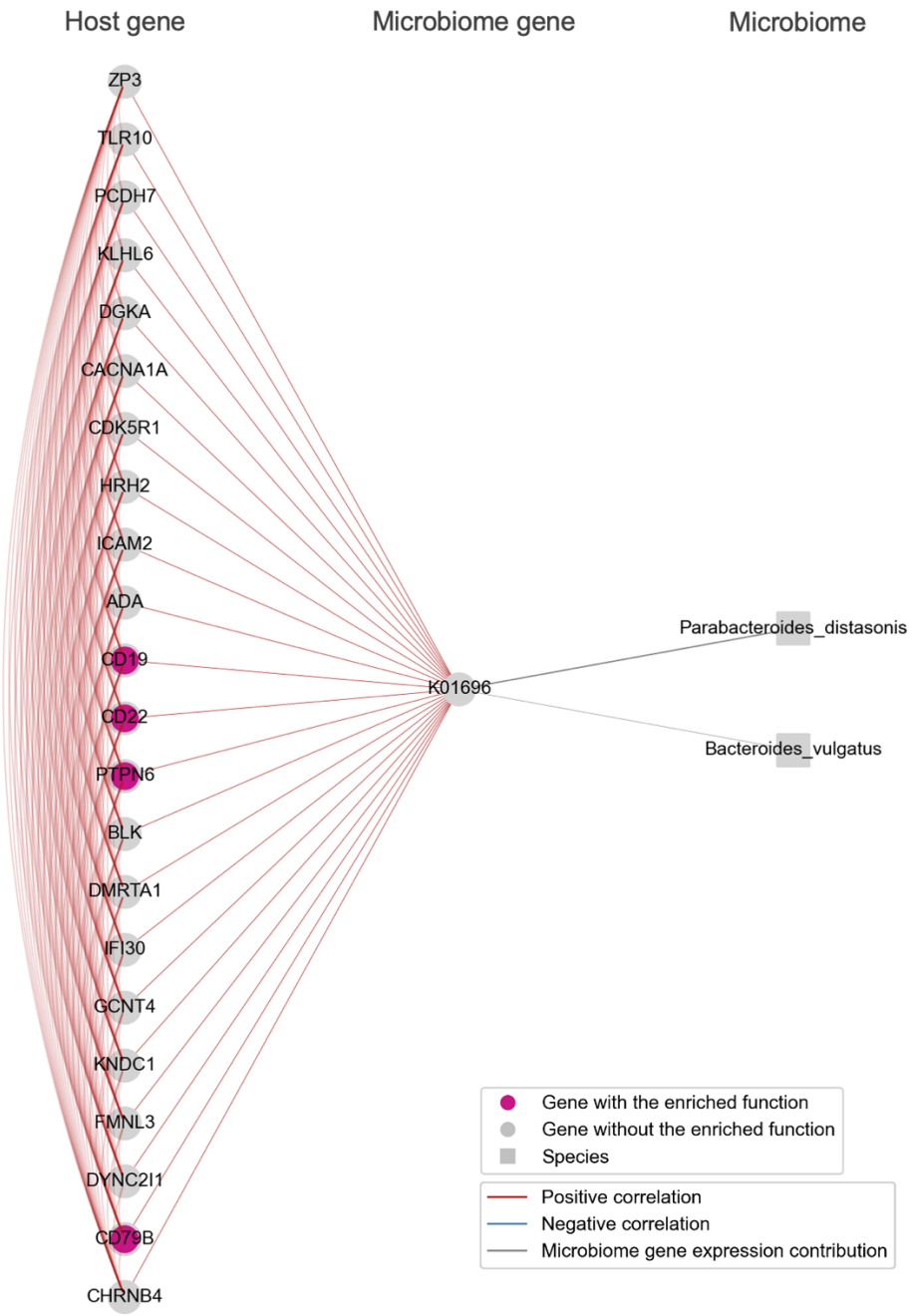

Fig. 6 B cell receptor signaling pathway-enriched Module_01
Host genes involved in the immune response of B cells, including *CD19*, *CD22*, *CD79B* and *PTPN6*, which were highly expressed in the cecum, co-expressed with the K01696 (*trpB*). These genes were visualized alongside bacterial species nodes that contribute to the expression of the microbiome genes. K01696 (*trpB*) expression was contributed to by *Parabacteroides distasonis*. Edges among genes indicate an absolute Spearman's correlation coefficient greater than 0.8. Edges connecting microbiome genes to bacterial species are shown when the contribution rate of bacterial species to microbiome gene expression exceeds 0.2. Note that the bacterial species nodes were added after the module detection analysis. The thickness of the edge increases with higher contribution rates.

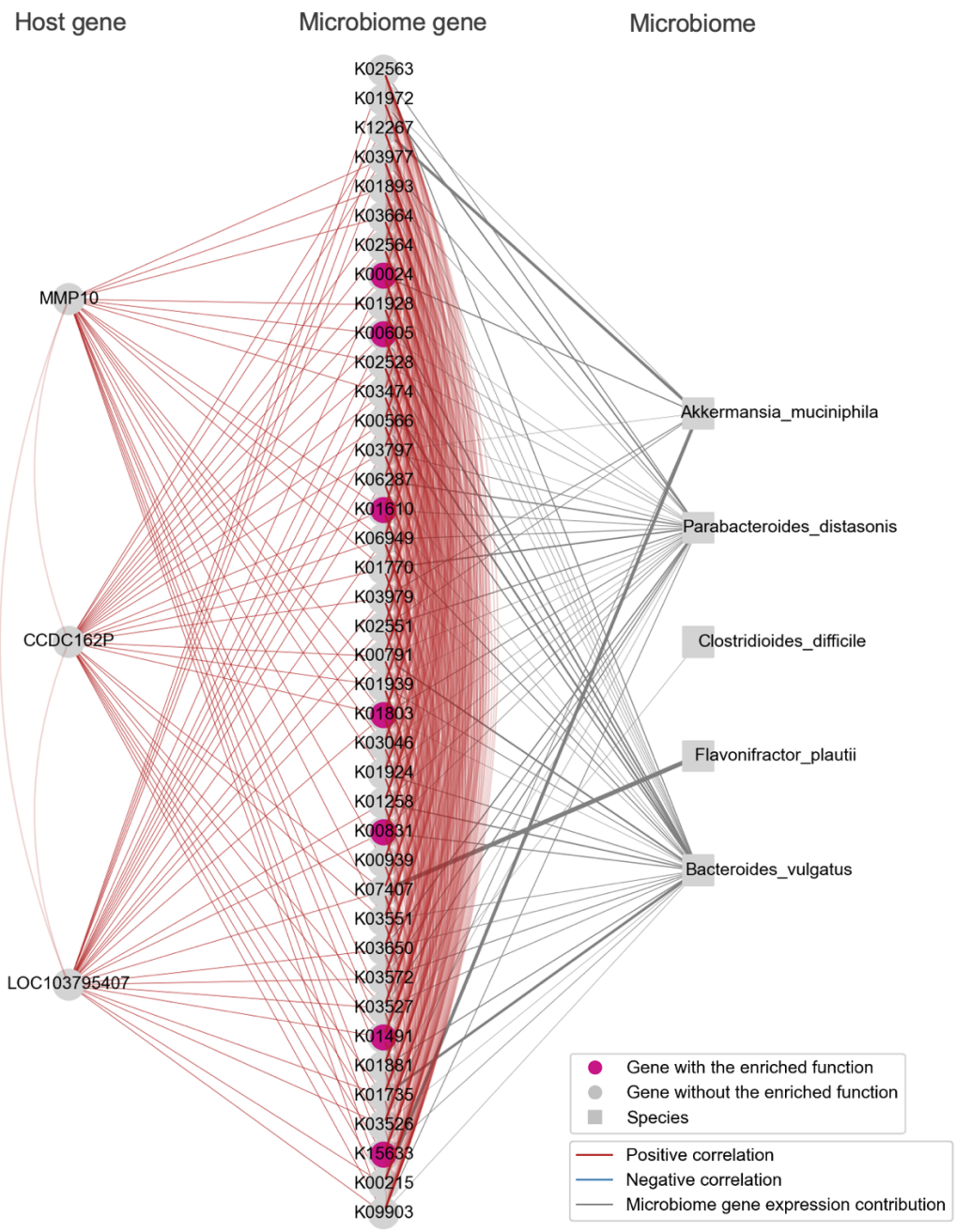

Fig. 7 Carbon metabolism pathway-enriched Module_21

Host genes, including *MMP10*, which is specifically highly expressed in the cecum, and 40 microbiome genes involved in carbon metabolism, were identified. These genes were visualized alongside bacterial species nodes that contribute to the expression of the microbiome genes. Note that the bacterial species nodes were added after the module detection analysis. Edges among genes indicate an absolute Spearman's correlation coefficient greater than 0.8. Edges connecting microbiome genes to bacterial species are shown when the contribution rate of bacterial species to microbiome gene expression exceeds 0.2. The thickness of the edge increases with higher contribution rates.

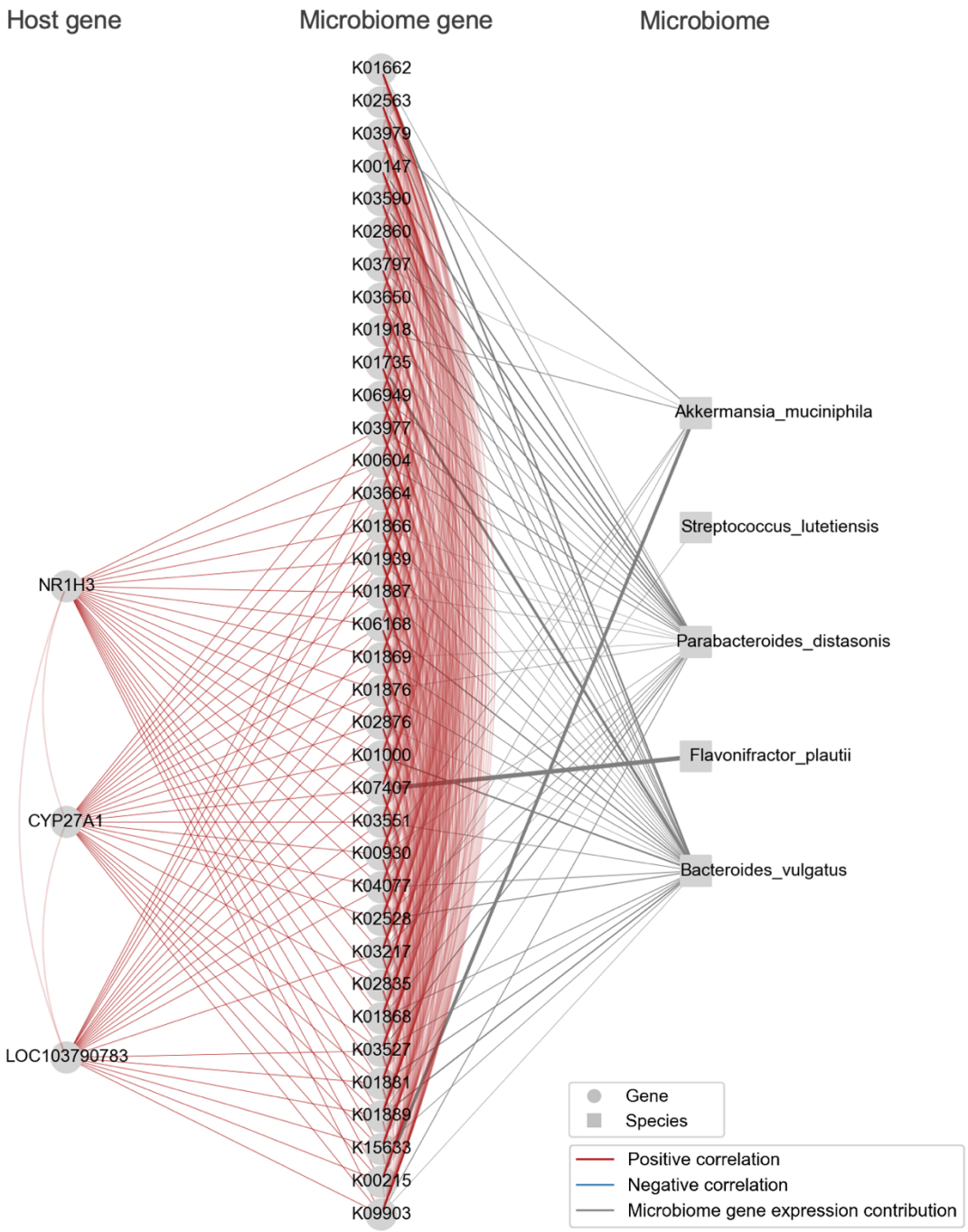

Fig. 8 Module_18 containing host genes and microbiome genes related to the production and breakdown of bile acids. These genes were visualized alongside bacterial species nodes that contribute to the expression of the microbiome genes. Note that the bacterial species nodes were added after the module detection analysis. Edges among genes indicate an absolute Spearman's correlation coefficient greater than 0.8. Edges connecting microbiome genes to bacterial species are shown when the contribution rate of bacterial species to microbiome gene expression exceeds 0.2. The thickness of the edge increases with higher contribution rates.