

Multi-level aggregation analysis of microbiome composition and host gene expression reveals associations with systemic and local immunity

Lev I. Shagam,¹ Anna Elizarova,² Yukihide Momozawa,¹ Julia Dmitrieva,¹ Rob Mariman,¹ Souad Rahmouni,¹ Edouard Louis,³ Michel Georges,^{1,4} Alexander V. Tyakht,² Natalia Klimenko²

AUTHOR AFFILIATIONS See affiliation list on p. 14.

ABSTRACT The human gut microbiome plays a critical role in immune regulation, yet the molecular links between microbiome composition and host gene expression remain incompletely understood. We analyzed associations between host gene expression and microbiome composition in a cohort of 315 healthy individuals, integrating microarray-based gene expression data from three intestinal sites (ileum, transverse colon, and rectum) and six immune cell types with microbiome sequencing data. Using a hierarchical feature aggregation strategy combining principal component analysis, clustering, and covariate correction, we discovered significant associations primarily related to immunity. While microbial profiles were similar across the three intestinal sites, the transverse colon yielded the most “microbiome-host gene expression” associations. Among the immune cell types, CD8⁺ cells showed the highest number of associations. The first principal component of microbiome composition, reflecting a gradient from commensals (e.g., *Ruminococcaceae* and *Christensenellaceae*) to proinflammatory taxa ([*Ruminococcus*] *gnavus* and *Lachnospirillum*), correlated with the expression of TNF- α -linked genes (*HMOX1*, *CPI17*, *HSD3B2*, and *SLC5A1*). Among individual genera, *Catenibacterium* abundance was associated with gene expression in both intestinal and immune cells, including negative associations with MRPS21 (related to mitochondrial function) in the transverse colon and with CD8⁺ gene programs related to T cell differentiation. These findings align with emerging evidence implicating mitochondrial dysfunction in intestinal inflammation. Our results identify multi-level associations between the gut microbiome and host gene expression, suggesting potential mechanisms by which microbiota shape local and systemic immunity and vice versa. The implicated genes and taxa represent candidates for experimental validation to improve understanding of host-microbiome homeostasis and its disruption in disease.

IMPORTANCE The gut microbiome and immune system are engaged in a complex interplay throughout human life. While most associative studies focus on case-control comparisons—typically examining patients with conditions such as inflammatory bowel disease or metabolic diseases—less is known about the molecular links between the microbiome and immune system in healthy individuals. In this study of a large cohort of healthy individuals, we addressed this gap by applying multiscale modeling to tackle the high dimensionality of host-microbiome data. We identified multi-level associations between microbiome composition and host gene expression in both intestinal tissues and immune cells. These findings offer a valuable reference for understanding baseline host-microbiome communication and highlight molecular candidates—such as TNF- α -related genes and mitochondrial pathways—for future experimental validation.

Editor Xue Zhang, China Agricultural University, Beijing, China

Address correspondence to Natalia Klimenko, natasha.klmnk@gmail.com.

Lev I. Shagam and Anna Elizarova contributed equally to this article. Lev I. Shagam took the lead in finalizing the manuscript text.

The authors declare no conflict of interest.

See the funding table on p. 14.

Received 5 November 2025

Accepted 10 July 2026

Published 10 August 2026

Copyright © 2026 Shagam et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

KEYWORDS gut microbiome, host gene expression, immunity, *Catenibacterium*, hierarchical feature aggregation

The gut microbiome plays a pivotal role in the functioning of the host organism. Its composition is associated with the host metabolome, as well as with risks, occurrence, and severity of various diseases. One particularly intriguing aspect is the relationship between the gut microbiome and the host's immune system. According to recent studies, these interactions influence not only the local structures of the immune system but also systemic immune response (1–4). Interestingly, despite being separated from myriads of live bacteria just by a single-cell-thick intestinal wall, a healthy human organism maintains a “tolerance-protection” balance via the fine-tuning of the immune system. How exactly the immune system manages this, despite the observed immense variability of microbiome compositions and absence of a reference “healthy” community, is still being clarified (5). Disruptions of this delicate equilibrium can contribute to the onset of severe pathologies, such as inflammatory bowel diseases (IBDs).

One direct approach to quantifying interactions at the “host-microbiome” interface and evaluating the local immune mechanisms is to compare host gene expression with microbiome composition. In mice, microbiota was shown to induce transcriptional response in intestinal epithelial cells affecting around 10% of the host's genes (6). Many human studies have explored similar interactions in various diseases marked by immune system dysfunction—and most of the identified associations were related to immunity. A longitudinal study of IBS patients (7) integrated multi-omics data, revealing a network of correlations among “host genes, microbial taxa, and metabolites”. The two major forms of IBD—ulcerative colitis and Crohn's disease (CD)—are known to manifest pronounced microbiome alterations and, besides clinical differentiation, can be distinguished by the microbiome profiles with an accuracy far better than chance. However, the colon mucosa transcriptomes were surprisingly similar between the two diseases (8). In a large multi-omics survey of IBD patients and their microbiome (9), it was observed that host gene expression in the gut was primarily determined by intestinal site (ileum vs rectum) rather than disease phenotype and inflammation. Associations between microbiome composition and host gene expression were site-specific, with several commensal taxa linked to human genes encoding antimicrobial proteins, including DUOX2, which is upregulated in IBD. Interestingly, DUOX2 expression was also associated with microbiome composition in another illustrative example of a disease with severely disrupted barrier function—cystic fibrosis (10). Another study investigating the disruption of host-microbiome interactions in IBD revealed tissue-specific links between intestinal microbes and inflammation-related genes. The findings included a correlation between the *Erysipelotrichaceae* family and collagen biosynthesis pathways, as well as a correlation between anaerobic butyrate-producing taxa and genes involved in transmembrane transport, host interleukin signaling, and metal ion response (11).

Furthermore, to increase the understanding of how gut microbiome affects systemic host homeostasis, it is important to examine the connections between microbiome composition and immune status parameters. In a longitudinal study of cancer patients undergoing hematopoietic cell transplantation, consistent correlations were found between the dynamics of immune cell counts and specific microbial genera in the gut (12). However, a deeper investigation of the immune transcriptome in humans has yet to be undertaken.

As highlighted by the disorder-focused studies mentioned above, the direction of host-microbiome links was clearly dictated by the nature of the condition. It is worth mentioning that these connections may have been confounded by secondary factors such as prior medical treatments, thus obfuscating the ability to identify the meaningful associations. The issue is further complicated by the high dimensionality of both host expression and microbiome community features, which, via the multiple comparison correction, reduces the sensitivity of discovery.

Here, we addressed this gap by comparing mucosal microbiome communities with matching host gene expression profiles at distinct sites along the intestinal tract (ileum, transverse colon, and rectum), as well as in multiple immune cell types within a large cohort of healthy subjects ($N = 315$). Utilizing a multi-level hierarchical feature aggregation approach to manage the high-dimensionality challenges, we discovered significant associations related to immunity and host metabolism that can shed light on local and systemic interactions between humans and their commensal microbes.

RESULTS

Establishing robust microbiome profiles based on multiple gut sites and 16S rRNA regions

Although stool samples are commonly used for human microbiome surveys due to their ease of collection, they provide a mixed signal, reflecting the luminal community of the distal colon. For a more accurate investigation of potential links between microbiome composition and human gene expression in the gut, we utilized biopsy samples from three distinct intestinal sites—the ileum, transverse colon, and rectum (previously published data [13]). Furthermore, since 16S rRNA amplicon sequencing is known to produce different profiles depending on the amplified variable region (14), we analyzed each sample using three regions (see Materials and Methods).

To assess the variability of observed microbiome composition across these target gene regions, we compared the abundance of microbial genera between the different amplicons in a pairwise manner. First, the amplicons varied by the presence of potential contaminants (contaminant detection strategy is described in reference 13, Table S1): the V34 region was less prone to contamination than V12 ($P = 2.76 \times 10^{-17}$, Mann-Whitney test) and V56 ($P = 1.85 \times 10^{-11}$), while no significant difference was observed between the latter two ($P = 0.17$) (Fig. S1B). However, the V34 region contained the highest proportion of reads mapping to human DNA (Fig. S1A). The observed variability in microbiome composition across amplicons was primarily due to the low-abundance genera (Fig. S1C). After filtering them out (see Materials and Methods), the median Spearman correlation between amplicons ranged from 0.73 to 0.78. Comparison of taxon abundance across amplicons showed borderline significance when comparing V1V2–V5V6 against V3V4–V5V6 (Mann-Whitney U test for Pearson correlation coefficients, $P = 0.058$), while no significant differences were observed for the other amplicon pair combinations (Fig. S1D). After applying centered-log ratio (clr) transformation, correcting for phenotypic and batch factors and filtering, the overall variability of microbiome composition explained by the intestinal site was relatively low (3.3% of the total variance, Fig. S2). Additionally, linear regression revealed that no feature was significantly associated with the intestinal site ($P > 0.05$). Consequently, we performed averaging across all intestinal sites and amplicons after correction using linear regression (see Materials and Methods). This refined microbiome composition profile for each subject was then used for the association analysis with the host gene expression.

Microbiome composition at different levels of hierarchical feature aggregation

The microbiome composition of the microbiome was dominated by the genera *Bacteroides* (34.2%), *Prevotella* (8.1%), and *Faecalibacterium* (5.5%) (Fig. 1A). These genera are typical members of the healthy human gut microbiome, as detected in both biopsy and stool sample studies (15–17).

Aggregation of samples by applying enterotyping methodology revealed four enterotypes: three driven by *Bacteroides* and one by *Prevotella* (Fig. S3). The average abundances of genera in enterotypes were consistent with those observed in stool samples from the TwinsUK and MetaCardis cohorts (18, 19). There were two enterotypes with strongly dominating *Bacteroides* (>40%), distinct by their evenness of major taxon abundance; we denoted them as Bact1 (higher evenness) and Bact2 (lower evenness) in

Downloaded from <https://journals.asm.org/journal/spectrum> on 12 August 2026 by 139.165.31.34.Downloaded from <https://journals.asm.org/journal/spectrum> on 12 August 2026 by 139.165.31.34.Downloaded from <https://journals.asm.org/journal/spectrum> on 12 August 2026 by 139.165.31.34.Downloaded from <https://journals.asm.org/journal/spectrum> on 12 August 2026 by 139.165.31.34.Downloaded from <https://journals.asm.org/journal/spectrum> on 12 August 2026 by 139.165.31.34.Downloaded from <https://journals.asm.org/journal/spectrum> on 12 August 2026 by 139.165.31.34.Downloaded from <https://journals.asm.org/journal/spectrum> on 12 August 2026 by 139.165.31.34.

The resulting list of transcriptomic features included: (i) gPCs, (ii) gPC-corrected gene expression, and (iii) expression of gene clusters and singleton genes.

The most highly expressed genes varied across different intestinal sites and immune cell types (Fig. S4). In the intestinal tissues, among the top genes expressed were those involved in immune response regulation (two *HLA* genes in the rectum and ileum), maintenance of the mucosal layer (mucin genes *MUC1* in the rectum and *FCGBP* in the transverse colon), detoxification (two *UGT* genes in the transverse colon; metallothionein genes in both the transverse colon and rectum), as well as cell proliferation and differentiation (*OLFM4* and *regenerating* genes in the ileum). In the immune cells, the most highly expressed genes included those related to general processes such as transcription and translation (*RPS26*, *HNRNPH1*, *RPL8*, and *SRSF5*), as well as genes with immune-specific functions: *CD3G*, essential for T cell receptor signaling; *STAT1*, which mediates immune response to pathogen infections in T cells; and lysozyme-encoding genes important for antibacterial defense. In platelets, the top expressed genes were involved in secretion processes (*PIP4K2A* and *SRGN*) and associated with cytoskeleton formation and function (*TUBB1* and *ACTB*).

Associations between microbiome composition and gene expression

Having established features at various levels of hierarchical feature aggregation (aggregation levels), we used them to investigate associations between microbiome composition and host gene expression (three features for the transcriptome and four for the microbiome, resulting in nine comparisons, as shown in Fig. 2).

We discovered 20 significant associations ($q < 0.05$); the majority of them ($N = 17$) are shown in Fig. 3, and the three remaining associations (each involving one bacterial genus) are for compactness provided in Fig. S5. Additionally, we listed these and non-significant results with $0.05 < q < 0.15$ ($N = 32$) in Table S2 to offer insights for future studies.

Primary microbiome axis is mainly associated with the expression of immune-related genes in the transverse colon

The analysis of associations between microbiome principal components and gene expression data yielded eight significant results ($q < 0.05$, Fig. 3, top two rows). Half of them involved gene expression in the transverse colon and the first microbiome

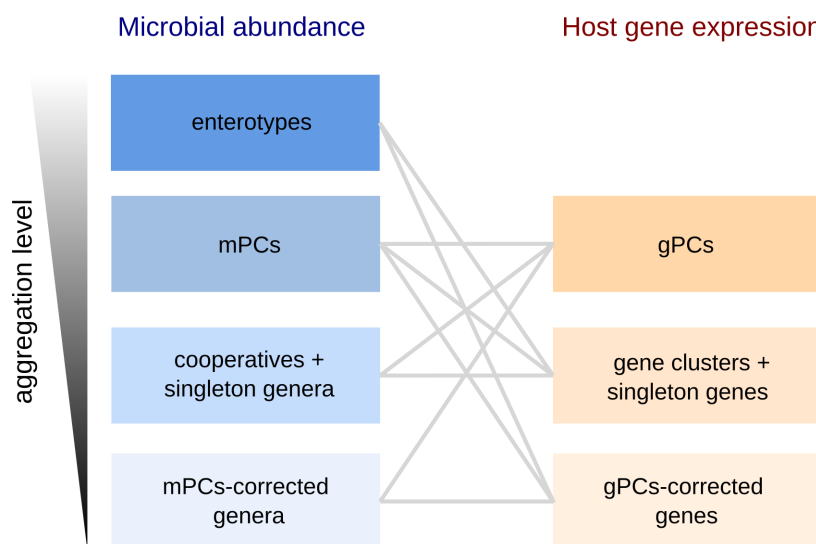


FIG 2 Outline of analyses conducted to identify associations between microbiome and host gene expression. Bluish boxes denote microbiome features, beige boxes denote host transcriptomic features, and gray lines denote comparisons conducted between levels of microbiome and host transcriptomic features.

principal component (mPC1). The mPC1 reflects a gradient of taxa from positive loadings dominated by unclassified *Christensenellaceae*, *Ruminococcaceae*, and [*Eubacterium*] *coprostanoligenes* group to negative loadings dominated by *Anaeroglobus*, *Brochothrix*, and [*Ruminococcus*] *gnavus* group (Fig. 4A) (here [*Ruminococcus*] in brackets refers to *Ruminococcus* from the *Lachnospiraceae* family later reclassified as *Mediterranibacter*). In the transverse colon, the expression of *HSD3B2* (raw $P = 9.8 \times 10^{-8}$, $q = 0.0079$) and *SLC5A1* ($P = 7.1 \times 10^{-8}$, $q = 0.0079$) was negatively associated with mPC1, while *HMOX1* ($P = 1.3 \times 10^{-6}$, $q = 0.0444$) and *CBX6* ($P = 4.1 \times 10^{-7}$, $q = 0.0166$) showed positive associations. Notably, *HSD3B2*, *SLC5A1*, and *HMOX1* are known to play roles in the intestinal immune response (see Discussion). No significant associations between mPC1 and gene expression were found in any other intestinal site or cell type. Interestingly, among top marginally significant associations for mPC1, all were also found in the transverse colon ($0.05 < q < 0.15$, $N = 5$: *OLFM4*, *SLC15A1*, *CPI17*, *ZDHHC14*, and *ZBTB11-AS1*, Fig. S6).

Despite the fact that the mPC1 reflects the main variation in microbiome data, similar to enterotypes, we found no significant differences in gene expression between the enterotypes ($q > 0.15$).

Catenibacterium abundance is related to systemic immunity and local inflammation

At the level of individual and clustered genera, *Catenibacterium* showed the highest number of significant associations ($n = 3$) with host gene expression data. Notably, it was linked to both intestinal (transverse colon biopsy) and blood immune cells (CD8+). In CD8+ immune cells, the genus abundance correlated positively with the fifth gene principal component (gPC5) ($P = 9.9 \times 10^{-8}$, $q = 0.0025$) and negatively with gene cluster

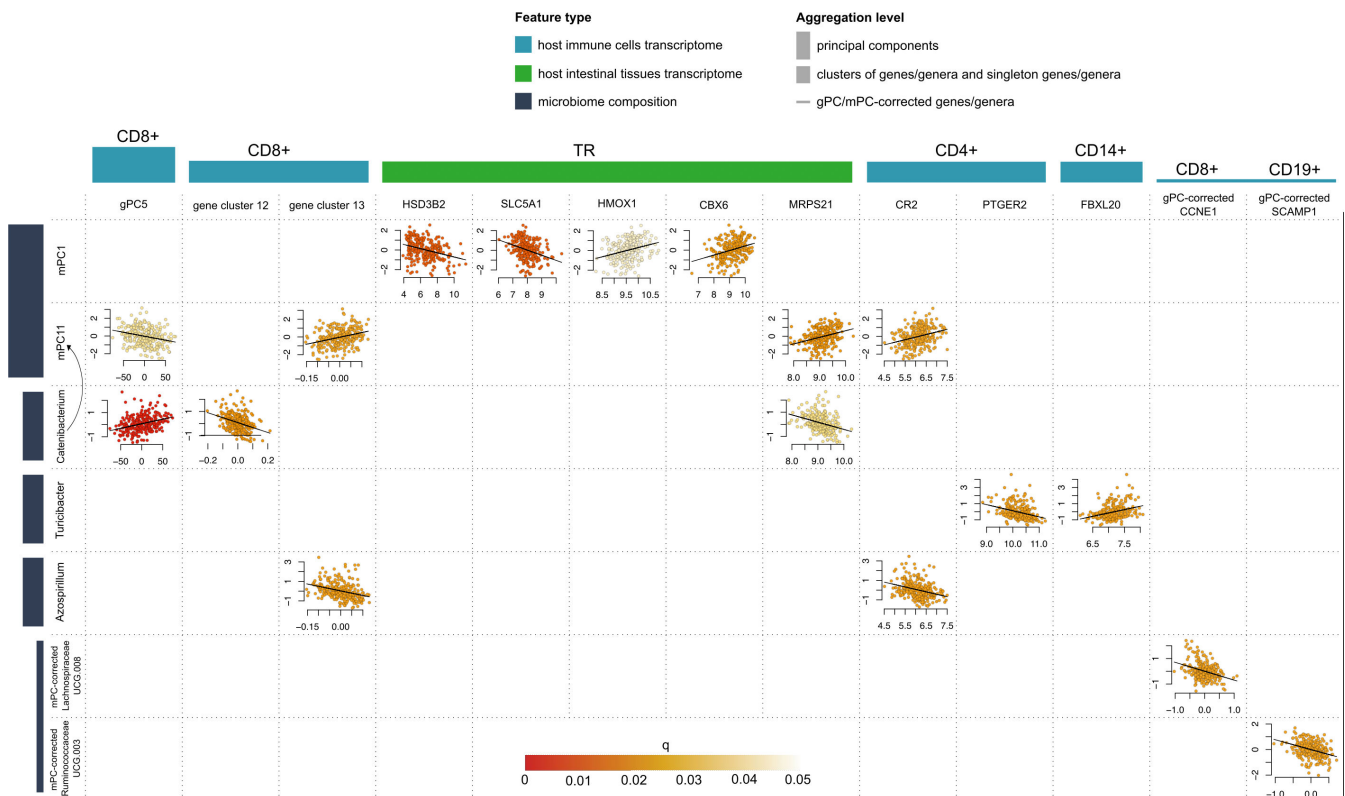
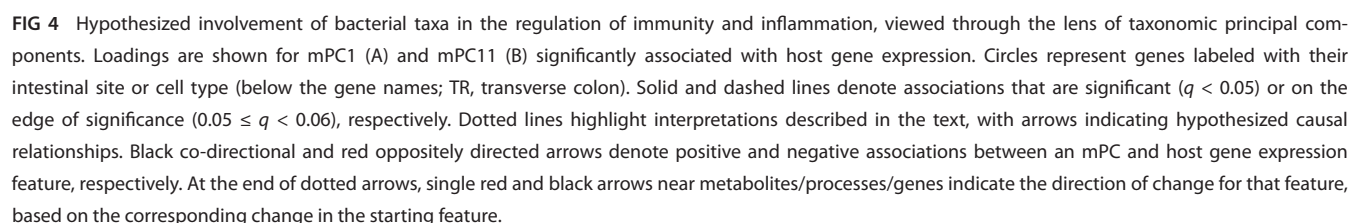


FIG 3 Associations between microbiome composition and host gene expression at different aggregation levels. The height of the top bars and the width of the left-side bars show the level of hierarchical feature aggregation, respectively; their colors correspond to feature type. Scatterplots show significant associations ($q < 0.05$) with the points colored by the q value. For compactness, among the uncorrected singleton genera, only those involved in ≥ 2 associations are shown (the rest three associations are provided in Fig. S5). TR, transverse colon.

Additionally, five bacterial genera manifested one to two associations with the host features. *Turicibacter* was positively associated with *FBXL20* expression in CD14+ cells and negatively associated with PTGER2 in CD4+ cells. *Azospirillum* 47_25 genus was negatively associated with *CR2* expression in CD4+ cells and with gene cluster 13 in CD8+ cells ($q = 0.019$ each). One immune cell-related association was discovered for each of the genera classified as *Clostridium innocuum* group, Clostridiales vadin BB60 group, and *Dielma* (Table S2 and Fig. S5).

After correcting for the main principal components (mPCs and gPCs), we identified two significant associations (Fig. 3). The mPC-corrected *Lachnospiraceae* UCG-008 abundance was negatively associated with gPC-corrected *CCNE1* expression in CD8+ T cells ($P = 3.5 \times 10^{-9}$, $q = 0.020$), and the mPC-corrected *Ruminococcaceae* UCG-003 abundance was negatively associated with gPC-corrected *SCAMP1* expression in CD19+ cells ($P = 3.9 \times 10^{-9}$, $q = 0.020$). *CCNE1* encodes cyclin E1, a positive cell cycle regulator (30, 31), while *SCAMP1* is involved in immune cell exocytosis (32) and endocytosis (33). Both *Lachnospiraceae* and *Ruminococcaceae* families have been frequently implicated



in studies exploring the connection between the microbiome and systemic immunity, though the specific immune cells associated with these taxa varied between studies (34–36). Interestingly, identified associations were not significant prior to the correction for the main mPCs and gPCs (raw P -values $> 10^{-5}$ before correction). This highlights that adjusting for principal components can help uncover hidden dependencies between microbiome composition and host gene expression.

DISCUSSION

The relationship between the human intestinal microbial community and the immune system is one of the most intriguing and complex examples of host-microbiome interactions. The immune system normally maintains a balance between tolerance to commensal taxa and defense against pathogens. The mechanisms underlying how the microbiome influences the host's intestinal immune system, as well as its role in inflammatory diseases, continue to be extensively studied (37). With an *in vitro* approach as one of the methodological solutions, it was shown that exposure of human colonocytes to gut microbiome resulted in widespread changes in the expression of genes involved in complex traits; as a proof-of-principle, a single microbe (*Collinsella*) was used to induce precise expression changes (38). In non-mammalian vertebrates (stickleback), observations of gut microbiome and host immune organ expression strongly suggest systemic impact of microbiome on immunity (39). Meanwhile, there remains a clear gap in robust analyses examining the “host-microbiome” interface in healthy human populations. In this study, we investigated these interactions by analyzing gut microbiome composition and host gene expression across intestinal tissues and immune cells within a large healthy cohort. Importantly, this cohort was particularly unique due to the exclusion of subjects with colon cancer, autoimmune conditions, or inflammatory diseases, providing a unique baseline for understanding microbiome-immune interactions in healthy individuals.

Linking the host transcriptome to the microbiome is inherently challenging due to the high dimensionality of the data. The sheer number of variables makes it difficult to distinguish true effects from random noise, often leading to few significant findings after multiple comparison corrections. In multi-omics studies, researchers may adopt a more relaxed significance threshold (e.g., $q < 0.25$ – 0.3) to capture suggestive associations. However, another effective strategy is to reduce the dimensionality of the analysis space, minimizing the number of statistical tests—for example, using PCA, cluster analysis, feature selection, and feature engineering (40–42). In this work, we employed an approach that incrementally increased the coarseness of microbial and host gene expression features. By aggregating features at higher levels, we reduced the number of tests and were able to detect broad, systemic effects, while lower-level aggregations allowed us to uncover more specific relationships between bacterial taxa and host gene expression. Our results revealed interesting associations at each level of the analysis, spanning both intestinal epithelial cells and blood immune cells. Some of these links obviously reflect the same interaction detected at different aggregation levels, while others may represent more level-specific phenomena.

One of the more prominent findings reproducible across different aggregation levels was the association between *Catenibacterium* and gene expression in immune cells, specifically CD8⁺ cells (Fig. 4B). The *Catenibacterium* genus belongs to the *Erysipelotrichaceae* family—an underexplored gut microbial family that has been linked to various aspects of host health. Considered rather immunogenic and related to host lipid metabolism (43), its members have been found to be overrepresented in the gut microbiome of HIV patients compared to healthy controls (44, 45). In these patients, *Erysipelotrichaceae* levels were positively associated with IL-1 β levels and negatively associated with a marker of microbial translocation (levels of IgM against lipopolysaccharide core antigen) (44). Furthermore, a study on germ-free mice showed that colonization with *Erysipelotrichaceae* increases the expression of genes responsible for antigen presentation and immune system education (46). The *Erysipelotrichaceae*

family was also associated with increased expression of extracellular matrix remodeling pathways in the intestinal tissue of IBD patients (11). Previously, we found that the proportions of *Catenibacterium*, alongside other highly immunogenic taxa, were increased in the gut of the indigenous Yakut population of Northeastern Siberia compared to the metropolitan population of Russia (47).

Interestingly, *Catenibacterium* was not only linked to host gene expression at the systemic level but also locally by showing a negative association with *MRPS21* expression in the transverse colon. The *MRPS21* gene encodes mitochondrial ribosomal protein S21, which plays a role in mitochondrial translation, and hence, its expression typically reflects overall mitochondrial metabolism. Mitochondrial dysfunction has been implicated in the pathogenesis of IBD (29, 48). A proteomic study on the onset of Crohn's disease showed that the majority of mitochondrial proteins are downregulated in CD patients compared to healthy controls (27). This dysfunction is thought to lead to the production of reactive oxygen species (ROS), increasing intestinal permeability and promoting endotoxemia (28, 29). Supporting this, an *ex vivo* study demonstrated that ROS initiates transcytosis of commensal microbes in human colon tissue, a process that can be inhibited by a mitochondria-targeted antioxidant (49). Intriguingly, *Catenibacterium* and mPC11 were the only microbiome features in our study to be associated with gene expression in both intestinal cells and blood immune cells, reinforcing the potential importance of this genus in host-microbiome interactions across systemic and local immune responses.

The lack of significant associations with enterotypes suggests that the top-level microbiome features are not informative of host gene expression. However, a large-scale microbiome feature significantly associated with transverse colon expression for multiple host genes was the first microbiome principal component. Having analyzed its loadings, we discovered that it represents a gradient from common gut commensals (such as *Ruminococcaceae* and *Christensenellaceae*) to those that are either less typical for the colon (such as *Anaeroglobus*—usually found in the upper intestinal tract) and/or linked to inflammation ([*Ruminococcus*] *gnavus* from the *Lachnospiraceae* family and *Lachnoclostridium*) (50–52). The top negative loadings also included the *Brochothrix* genus, which, although not commonly found in the human gut, has occasionally been detected in stool samples (53), possibly as a transient microbe from food products (54). Interestingly, most taxa with positive loadings on mPC1 have recently emerged as biomarkers of metabolic and cardiovascular health, with some proposed mechanisms. For instance, the [*Eubacterium*] *coprostanoligenes* group is recognized for converting intestinal cholesterol to coprostanol, thus exerting serum cholesterol-lowering effects (55). Other important taxa, including the *Christensenellaceae* R7 group, *Ruminococcaceae* UCG-005, and *Ruminococcaceae* NK4A214 group, were identified among seven butyrate-producing taxa that were inversely associated with insulin resistance in a large cross-sectional study involving 2,166 participants (56).

The list of genes positively associated with mPC1 included *HMOX1*, *CBX6* ($q < 0.05$), and, at the edge of significance, *CPI17* (also known as *PPP1R14A*) ($q = 0.0585$). Both *HMOX1* and *CPI17* are thought to play protective roles in intestinal inflammation. *HMOX1* was shown to reduce inflammation through the induction of anti-inflammatory cytokine production (57). Experiments on germ-free animals showed that *HMOX1* expression is at least partially regulated by the gut microbiome (58). A by-product of *HMOX1*-encoded protein activity, carbon monoxide, was shown to ameliorate IBD symptoms, likely by reducing TNF- α expression (57, 59, 60). As for the *CPI17* (C-kinase potentiated protein phosphatase-1 inhibitor of 17 kDa), its expression is regulated by TNF- α (61). During intestinal inflammation, increased TNF- α levels result in reduced *CPI17* expression in intestinal smooth muscle (61). *CPI17* enhances smooth muscle contraction by inhibiting smooth muscle myosin phosphatase. A decrease in *CPI17* levels leads to decreased intestinal motility associated with IBD (62, 63).

On the other hand, *HSD3B2* and *SLC5A1* were among the genes negatively associated with the first microbiome principal component in the transverse colon. *HSD3B2* encodes the enzyme 3-Beta-HSD, typically found in hormone-producing tissues. However, recent

studies suggest that mucus cells can also produce this enzyme and other steroidogenic enzymes, with their expression upregulated by TNF- α (64, 65). Our findings indicate that many of the genes associated with mPC1 are linked to TNF- α in various ways (Fig. 4A). Moreover, one of the taxa with negative loadings on mPC1, [*Ruminococcus*] *gnavus*, is known to produce a pro-inflammatory polysaccharide that induces TNF- α production by dendritic cells (66). This leads us to speculate that some of the mPC1-gene associations are mediated through cytokines, particularly TNF- α .

SLC5A1 (*SGLT1*), another gene negatively associated with mPC1, encodes a transmembrane transporter responsible for dietary glucose and galactose absorption. Its expression is known to fluctuate with changes in lumen carbohydrate levels and is influenced by metabolic diseases (67) and inflammation (68). However, the direction of *SLC5A1* expression change in response to inflammation remains ambiguous (67, 69–71). Interestingly, in pigs, cecal microbiome composition was shown to explain 24% of the variation in serum glucose levels (72). Beyond their primary function in sugar uptake, glucose transporters are also involved in inflammation, malignancy, and the regulation of the gut microbiota (73).

Notably, the significant associations between microbiome features and gene expression were restricted to adaptive immune cells (T cells: CD8+ and CD4+; B cells: CD19+), while no associations were found with the examined components of the innate immune system (monocytes: CD14+; granulocytes: CD15+). For intestinal tissues, associations were observed only for the transverse colon host gene expression.

As a methodological limitation, we acknowledge that residualization may not fully account for correlations between predictors and covariates (age, sex, BMI, smoking status, collection date, intestinal site, and amplicon) and may slightly affect standard error estimates (74), though mathematically justified under certain conditions. However, because both microbiome and transcriptome descriptors were adjusted using the same strategy, the resulting associations represent covariate-independent relationships between the two data modalities, consistent with our goal of identifying intrinsic host-microbiome links. At the same time, residualization-based approaches are commonly used in microbiome association studies.

Our study demonstrates how a gradual dimensionality reduction approach can reveal novel associations between gut microbiome composition and host gene expression data. Most of the identified links, both in intestinal tissues and immune cells, were related to immunity, with some associations reproducible at multiple levels of hierarchical feature aggregation.

The identified associations provide promising candidates for experimental validation through methods such as culturomics, intestinal tract simulators with human colonocyte layers, or animal models. Establishing the specific effects of key microorganisms on host gene expression could lay the groundwork for precise microbiome-driven immune modulation using techniques such as fecal microbiota transplantation, live biotherapeutics, and other interventions. In the future, these insights may prove valuable for modulating systemic immunity.

MATERIALS AND METHODS

Study design and cohort enrollment

The study population consisted of 315 healthy individuals who visited the Academic Hospital of the University of Liège as part of a national colon cancer screening campaign. All participants were free of cancer, autoimmune diseases, or inflammatory diseases and were not taking corticosteroids or non-steroidal anti-inflammatory drugs, except for low-dose aspirin for thrombosis prevention (as described in reference 25).

Sample collection and preparation

Biopsy samples for microbiome and host gene expression analysis were collected from three regions of the intestine: the ileum, transverse colon, and rectum. Immune cells

were isolated from peripheral blood, following the protocol described in reference 25, and included six immune cell types from both the adaptive immune system (CD4+ T lymphocytes, CD8+ T lymphocytes, and CD19+ B lymphocytes) and the innate immune system (CD14+ monocytes and CD15+ granulocytes), as well as platelets.

Normalization of microbiome data

The microbiome sequencing data used in this study have been previously published (13); details on DNA extraction, sequencing, and raw data processing are also available in that publication. A summary of the data processing workflow is provided in Fig. S7B.

Similarly to the data processing in the abovementioned article, the original taxon abundances produced by the taxonomy classifier were further subjected to removal of the determined contaminants, downsampling, and centered log ratio transformation (13). However, the subsequent mixed model data correction was performed in a different way, as described in detail below.

The data were subjected to a mixed model approximation implemented in the lme4 R package (75) with the following covariates and variables, and the residuals were collected to represent the corrected abundance of genera:

$$\text{genus_abundance} \sim \text{age} + \text{sex} + \text{BMI} + \text{smoking_status} + (1 \mid \text{collection_date}) + \text{intestinal_site} + \text{Amplicon} \quad (1)$$

In the primary data analysis round, our aim was to determine the effects of the intestinal site and the 16S rRNA gene amplicon type on the observed taxon abundance measurements.

For the purpose of the amplicon (16S rRNA region) effect assessment, the abundances were corrected for all the variables of the model (equation 1) except for sample collection date, amplicon, and intestinal site. Then, the abundances of each taxon were compared across amplicon pairs (V1V2 to V3V4, V1V2 to V5V6, and V3V4 to V5V6) using Spearman correlation analysis. Further comparison of the amplicon pairs was performed using the Mann-Whitney *U* test.

In the process of determining the intestinal site effect on microbiome composition, the abundances were corrected for all the variables of the mixed model (equation 1) except for sample collection date and intestinal site. The homogeneity of variance was evaluated using the PERMDISP2 test (implemented via the betadisper function in the vegan R package (76)). To assess general differences between intestinal sites, a PERMANOVA test (via the adonis2 function in vegan) was conducted.

Since the analysis showed no significant association between sample intestinal site and microbiome composition and most taxa exhibited high correlation between amplicons, for further analysis, we updated the mixed model correction method by adding intestinal site and amplicon to the list of variables we correct for. Thereafter, we averaged the adjusted abundances for each genus-subject pair across different amplicons and sites. To avoid overfitting, for genera with <150 non-zero values (prior to clr transformation) across all “sample-intestinal site-amplicon” combinations, the correction for intestinal site and amplicon was substituted by subtracting the mean abundance across them.

After these steps, one abundance profile per subject was obtained.

Dimensionality reduction of microbiome data

Multiple strategies for dimensionality reduction were employed to derive microbiome features at various aggregation levels. After preprocessing and filtering, 145 genera were retained for analysis. The dimensionality reduction techniques applied included enterotyping, PCA, and clustering of co-abundant features.

Enterotyping was performed on taxonomic features after the removal of contaminants, without any transformations or corrections, using the DirichletMultinomial R package (77), with the optimal number of enterotypes ($n = 4$) selected using the Laplace

goodness-of-fit criterion. Prior to enterotyping, the total read counts per sample were rarefied to 10,000 (for the V34 amplicon, data were rarefied to 5,000 reads, and then the counts were multiplied by two due to lower coverage data available for the region V34) and averaged across intestinal sites and amplicons. Outliers with very high median Jensen–Shannon divergence from 50% of closest neighbors ($P < 0.01$) were removed.

The PCA method was applied to corrected genus abundances using the *ade4* R package (78). We selected the first 15 principal components of microbiome taxonomic abundances, which together explained 44.1% of the variance (further referred to as mPCs; Fig. S8).

To identify clusters of highly correlated genera, we constructed a correlation network based on corrected genus abundances with Spearman's $\rho \geq 0.6$. The Louvain algorithm, implemented in the *igraph* R package (79), was used to detect clusters within this network. The abundance of each cluster was calculated as the sum of abundances of its member taxa. Individual genera that did not belong to any of the identified clusters were retained as singletons.

To correct for the major signals in the microbiome data, we adjusted genus abundances using a linear model based on the first 15 mPCs and calculated the residuals for further analysis. These residuals, referred to as mPC-corrected abundances, were used in subsequent association analyses.

Overall, the following features were used to analyze microbiome associations with host gene expression:

1. Enterotypes ($N = 4$);
2. Microbial principal components ($N = 15$);
3. Microbial clusters and singleton genera ($N = 7$ clusters and 118 singletons);
4. mPC-corrected genera ($N = 145$).

Host gene expression analysis

Human transcriptome data had been generated for the biopsy samples taken from the ileum, transverse colon, and rectum, as well as from the following immune cell types: CD4+ T lymphocytes, CD8+ T lymphocytes, CD19+ B lymphocytes, CD14+ monocytes, CD15+ granulocytes, and platelets using Illumina HT12 arrays as described in a previous publication (25). The data had been preprocessed, filtered, and corrected for covariates (sex, age, smoking status, and Sentrix ID). A summary of the data processing steps is shown in Fig. S7A.

Similar to the microbiome features, we employed several strategies for dimensionality reduction of the host transcriptome data. First, we utilized the classical principal components of gene expression data (referred to as gPCs), which were obtained from the same study (25). The number of gPCs varied slightly between different immune cell types, as the original paper selected PCs that maximized the number of *cis*-eQTLs ($N = 27$ –59). Second, we performed a weighted gene co-expression network analysis (WGCNA) on the gene expression data corrected for covariates (from reference 25); parameters: `minModuleSize = 30`, soft thresholding power was evaluated by plotting scale-free topology fit index for each tissue/cell type and resulted in a range from 6 to 9. WGCNA identifies modules (clusters) of highly correlated genes, summarizing their expression values into module eigengenes (the first principal component of a given module) (80). Additionally, highly variable unclustered genes (those in the top quartile of expression variance distribution) were included as singletons. Finally, we obtained gene expression values corrected for the main principal components from reference 25.

In summary, the following transcriptomic features were used to analyze associations with microbiome features:

1. Host gene expression principal components ($N = 27$ –59 gPCs depending on cell type and intestinal site, see reference 25);

2. Gene clusters and singleton genes ($N = 14\text{--}47$ clusters and 1,703–6,719 singletons depending on cell type and intestinal site);
3. Gene expression values corrected for the first 27–59 gPCs (gPC-corrected gene expression) ($N = 6,638\text{--}15,401$ depending on cell type and intestinal site).

Analysis of associations between microbiome features and host gene expression

We explored potential associations between the microbiome and host transcriptomic features across different aggregation levels. The following relationships were modeled (Fig. 2):

1. gPCs, gene clusters + singleton genes, and gPC-corrected gene expression—with mPCs;
2. gPCs and gene clusters + singleton genes—with microbial clusters + singleton genera;
3. gPCs and gPC-corrected gene expression—with mPC-corrected genus abundances;
4. gPC-corrected gene expression and gene clusters + singleton genes—with enterotypes.

To detect associations, we used a linear regression model for each of the mentioned contrasts, except for the comparison between host genes and enterotypes: microbial feature ~ host gene expression feature.

Both microbial features and host transcriptomic features had been independently corrected for biological and technical covariates as described above. Before regression analysis, outliers were removed. Values were considered outliers if they fell outside the range of $[Q1 - 5 \times IQR; Q3 + 5 \times IQR]$ for the feature distribution. The proportion of removed outliers was low ($0.6\% \pm 1.4\%$ per comparison; see Table S5), and in most comparisons (Table S6), no samples were removed. Multiple testing correction was applied for each type of comparison (e.g., gPC ~ mPC) separately for intestinal and immune cells using the Benjamini-Hochberg method. Associations with $q < 0.05$ were considered significant. To address the potential presence of non-independent comparisons being subject to multiple comparison adjustment, for each of our nine comparison blocks, we implemented an advanced procedure based on Sidak's equation (81) that performs multiple comparison adjustment taking into account the effective number (Neff) of independent comparisons that is estimated from a P -value distribution derived from a large number of random permutations of the compared data ($k = 1,000$); after the Neff is calculated from Sidak slope, it is used for Sidak adjustment of the raw P -values. We also applied a global multiple testing procedure across all types of comparisons and all intestinal sites or immune cell types to highlight the most significant associations (Table S2).

Associations between enterotypes and gene expression data were assessed using the Kruskal-Wallis test. For comparison involving high dimensionality, the q values were additionally corrected for inflation using the genomic inflation factor (82). Significant associations were further validated using robust MM-type regression (*lmrob* function, method = "MM") to mitigate the influence of residual outliers.

Pathway enrichment analysis

To functionally characterize gene clusters of interest, we conducted over-representation analysis (ORA) using the Reactome Pathway Knowledgebase (83) via its embedded ORA online tool. For the gPCs of interest, we performed ORA on the 200 genes with the highest and the 200 genes with the lowest Spearman correlation coefficients per gPC. To mitigate potential bias caused by the overrepresentation of certain pathways in specific immune cell types (which could affect all principal components and result in the

identification of cell type rather than cluster-/PC-specific pathways), we implemented a permutational control for all identified hits. Specifically, all transcripts were divided into bins of 100 elements each, based on their expression levels within the immune cell type of interest. We subsequently generated gene sets where each of the genes of interest was replaced with a random gene from the same bin. ORA was then applied to these permuted gene sets, and we retained as truly overrepresented only those pathways that were not overrepresented in the permuted gene sets ($N = 1,000$).

ACKNOWLEDGMENTS

We thank Maria Pashkova, Viktoriya Voroshilova, and Evgenii Levin for the help with the pilot data analysis, and Dmitry Alexeev for the contribution to the initiation of the project's pilot stage.

This project was conducted with funding from the SYSCID H2020 grant (ref. 733100), the MyQuant (ref. 30770923), and BRIDGE (O.0006.22–RG3124) projects from the Excellence of Science (EOS) program (FNRS, Fédération Wallonie-Bruxelles and FWO, Flemish Community), the CLIMAX (WELBIO-CR-2022 A) project from WELBIO (Walloon Region), the RHEAQT (T.0096.19) and IBD-GI-Seq (T.0190.19) projects from the FNRS (Fédération Wallonie-Bruxelles), the ARC RHEACT WITH HSPC project from the University of Liège. Computational resources have been provided by the Consortium des Équipements de Calcul Intensif (CÉCI), funded by the Fonds de la Recherche Scientifique de Belgique (F.R.S.-FNRS) under Grant No. 2.5020.11 and by the Walloon Region. Part of the work on the manuscript by A.E., N.K. and A.V.T. was supported by the Ministry of Science and Education of the Russian Federation, project No. 075-15-2025-475.

AUTHOR AFFILIATIONS

¹Unit of Animal Genomics, GIGA Institute and Faculty of Veterinary Medicine, Avenue de l'Hôpital, University of Liège, Liège, Belgium

²Institute of Gene Biology, Russian Academy of Sciences, Moscow, Russia

³Department of Gastroenterology, Faculty of Medicine, CHU & GIGA Institute, Avenue de l'Hôpital, University of Liège, Liège, Belgium

⁴WELBIO Department, WEL Research Institute, Wavre, Belgium

AUTHOR ORCIDs

Lev I. Shagam  <http://orcid.org/0000-0003-1325-3085>

Yukihide Momozawa  <http://orcid.org/0000-0001-5638-3504>

Souad Rahmouni  <http://orcid.org/0000-0003-0956-0242>

Edouard Louis  <http://orcid.org/0000-0002-1287-7075>

Michel Georges  <http://orcid.org/0000-0003-4124-2375>

Alexander V. Tyakht  <http://orcid.org/0000-0002-7358-2537>

Natalia Klimenko  <http://orcid.org/0000-0001-9640-0102>

FUNDING

Funder	Grant(s)	Author(s)
Horizon 2020 (EU)	SYSCID (733100)	Julia Dmitrieva
Excellence of Science (EOS) (Belgium)	MyQuant (30770923)	Michel Georges Rob Mariman Lev I. Shagam
Excellence of Science (EOS) (Belgium)	Bridge (O.0006.22–RG3124)	Michel Georges Souad Rahmouni
WELBIO (Walloon Region, Belgium)	CAUSIBD	Michel Georges Julia Dmitrieva

Funder	Grant(s)	Author(s)
WELBIO (Walloon Region, Belgium)	CLIMAX (CR-2022A)	Michel Georges
Fonds de la Recherche Scientifique (FNRS) (Wallonia-Brussels Federation, Belgium)	RHEAQT (T.0096.19)	Michel Georges
Fonds de la Recherche Scientifique (Wallonia-Brussels Federation, Belgium)	IBD-GI-Seq (T.0190.19)	Souad Rahmouni Lev I. Shagam
University of Liège (Wallonia-Brussels Federation, Belgium) - ARC RHEACT WITH HSPC		Michel Georges
Ministry of Science and Education of the Russian Federation	075-15-2025-475	Anna Elizarova Alexander V. Tyakht Natalia Klimenko

AUTHOR CONTRIBUTIONS

Lev I. Shagam, Data curation, Formal analysis, Methodology, Software, Writing – original draft, Writing – review and editing | Anna Elizarova, Data curation, Formal analysis, Methodology, Software, Writing – original draft, Writing – review and editing | Yukihide Momozawa, Data curation, Formal analysis, Methodology | Julia Dmitrieva, Data curation, Formal analysis, Methodology | Rob Mariman, Data curation, Formal analysis, Methodology | Souad Rahmouni, Data curation, Formal analysis, Methodology | Edouard Louis, Data curation, Formal analysis, Methodology | Michel Georges, Conceptualization, Project administration, Resources, Supervision, Writing – original draft, Writing – review and editing | Alexander V. Tyakht, Conceptualization, Project administration, Resources, Supervision, Writing – original draft, Writing – review and editing, Formal analysis | Natalia Klimenko, Data curation, Formal analysis, Methodology, Software, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

The host gene expression data were previously published (47) and are available in the Array Express database (<https://www.ebi.ac.uk/arrayexpress/>) under accession number E-MTAB-6667. The microbiome data for 194 subjects were previously published (59) and are available in the European Nucleotide Archive (ENA) under BioProject accession PRJNA814419. The newly presented microbiome data for 121 subjects are available under the accession PRJEB111129. The essential data analysis code is available at https://github.com/Lev-genetik/CEDAR_mt.

ETHICS APPROVAL

The study was reviewed and approved by the Ethics Committee of the University of Liège Academic Hospital. The patients/participants provided their written informed consent to participate in this study.

ADDITIONAL FILES

The following material is available [online](#).

Supplemental Material

Supplemental figures (Spectrum03593-25-s0001.docx). Fig. S1 to S8.

Supplemental tables (Spectrum03593-25-s0002.xlsx). Tables S1 to S6.

REFERENCES

- Brown EM, Kenny DJ, Xavier RJ. 2019. Gut microbiota regulation of T cells during inflammation and autoimmunity. *Annu Rev Immunol* 37:599–624. <https://doi.org/10.1146/annurev-immunol-042718-041841>
- Lo BC, Chen GY, Núñez G, Caruso R. 2021. Gut microbiota and systemic immunity in health and disease. *Int Immunol* 33:197–209. <https://doi.org/10.1093/intimm/dxaa079>
- Wang J, Zhu N, Su X, Gao Y, Yang R. 2023. Gut-microbiota-derived metabolites maintain gut and systemic immune homeostasis. *Cells* 12:793. <https://doi.org/10.3390/cells12050793>
- Wiertsema SP, van Bergenhenegouwen J, Garssen J, Knipfels LMJ. 2021. The Interplay between the gut microbiome and the immune system in the context of infectious diseases throughout life and the role of nutrition in optimizing treatment strategies. *Nutrients* 13:886. <https://doi.org/10.3390/nu13030886>
- Clasen SJ, Bell MEW, Borbón A, Lee D-H, Henseler ZM, de la Cuesta-Zuluaga J, Parys K, Zou J, Wang Y, Altmannova V, Youngblut ND, Weir JR, Gewirtz AT, Belkhabir Y, Ley RE. 2023. Silent recognition of flagellins from human gut commensal bacteria by toll-like receptor 5. *Sci Immunol* 8:eabq7001. <https://doi.org/10.1126/sciimmunol.abq7001>
- Sommer F, Nookaew I, Sommer N, Fogelstrand P, Bäckhed F. 2015. Site-specific programming of the host epithelial transcriptome by the gut microbiota. *Genome Biol* 16:62. <https://doi.org/10.1186/s13059-015-0614-4>
- Mars RAT, Yang Y, Ward T, Houtti M, Priya S, Lekatz HR, Tang X, Sun Z, Kalari KR, Korem T, et al. 2020. Longitudinal multi-omics reveals subset-specific mechanisms underlying irritable bowel syndrome. *Cell* 182:1460–1473. <https://doi.org/10.1016/j.cell.2020.08.007>
- Granlund A van B, Flatberg A, Østvik AE, Drozdov I, Gustafsson BJ, Kidd M, Beisvag V, Torp SH, Waldum HL, Martinsen TC, Damås JK, Espevik T, Sandvik AK. 2013. Whole Genome gene expression meta-analysis of inflammatory bowel disease colon mucosa demonstrates lack of major differences between Crohn's disease and ulcerative colitis. *PLoS One* 8:e56818. <https://doi.org/10.1371/journal.pone.0056818>
- Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon TW, Andrews E, Ajami NJ, Bonham KS, Brislawn CJ, et al. 2019. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* 569:655–662. <https://doi.org/10.1038/s41586-019-1237-9>
- Dayama G, Priya S, Niccum DE, Khoruts A, Blekhan R. 2020. Interactions between the gut microbiome and host gene regulation in cystic fibrosis. *Genome Med* 12:12. <https://doi.org/10.1186/s13073-020-0710-2>
- Hu S, Bourgonje AR, Gacesa R, Jansen BH, Björk JR, Bangma A, Hidding IJ, van Dullemen HM, Visschedijk MC, Faber KN, Dijkstra G, Harmsen HJM, Festen EAM, Vich Vila A, Spekhorst LM, Weersma RK. 2024. Mucosal host-microbe interactions associate with clinical phenotypes in inflammatory bowel disease. *Nat Commun* 15:1470. <https://doi.org/10.1038/s41467-024-45855-2>
- Schluter J, Peled JU, Taylor BP, Markey KA, Smith M, Taur Y, Niehus R, Staffas A, Dai A, Fontana E, Amoretti LA, Wright RJ, Morjaria S, Fenelus M, Pessin MS, Chao NJ, Lew M, Bohannon L, Bush A, Sung AD, Hohl TM, Perales M-A, van den Brink MRM, Xavier JB. 2020. The gut microbiota is associated with immune cell dynamics in humans. *Nature* 588:303–307. <https://doi.org/10.1038/s41586-020-2971-8>
- Petrov VA, Sharapov SZ, Shagun L, Nostaeva AV, Pezer M, Li D, Hanić M, McGovern D, Louis E, Rahmouni S, Lauc G, Georges M, Aulchenko YS. 2022. Association between human gut microbiome and N-glycan composition of total plasma proteome. *Front Microbiol* 13:811922. <https://doi.org/10.3389/fmicb.2022.811922>
- Rintala A, Pietilä S, Munukka E, Eerola E, Pursiheimo J-P, Laiho A, Pekkala S, Huovinen P. 2017. Gut microbiota analysis results are highly dependent on the 16S rRNA gene target region, whereas the impact of DNA extraction is minor. *J Biomol Tech* 28:19–30. <https://doi.org/10.7171/jbt.17-2801-003>
- Li G, Yang M, Zhou K, Zhang L, Tian L, Lv S, Jin Y, Qian W, Xiong H, Lin R, Fu Y, Hou X. 2015. Diversity of duodenal and rectal microbiota in biopsy tissues and luminal contents in healthy volunteers. *J Microbiol Biotechnol* 25:1136–1145. <https://doi.org/10.4014/jmb.1412.12047>
- Tyakht AV, Kostyukova ES, Popenko AS, Belenikin MS, Pavlenko AV, Larin AK, Karpova IY, Selezneva OV, Semashko TA, Ospanova EA, et al. 2013. Human gut microbiota community structures in urban and rural populations in Russia. *Nat Commun* 4:2469. <https://doi.org/10.1038/ncomms3469>
- Zhernakova A, Kurilshikov A, Bonder MJ, Tigchelaar EF, Schirmer M, Vatanen T, Mujagic Z, Vila AV, Falony G, Vieira-Silva S, et al. 2016. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science* 352:565–569. <https://doi.org/10.1126/science.1233669>
- Holmes I, Harris K, Quince C. 2012. Dirichlet multinomial mixtures: generative models for microbial metagenomics. *PLoS One* 7:e30126. <https://doi.org/10.1371/journal.pone.0030126>
- Vieira-Silva S, Falony G, Belda E, Nielsen T, Aron-Wisniewsky J, Chakaroun R, Forslund SK, Assmann K, Valles-Colomer M, Nguyen TTD, et al. 2020. Statin therapy is associated with lower prevalence of gut microbiota dysbiosis. *Nature* 581:310–315. <https://doi.org/10.1038/s41586-020-2269-x>
- Klimenko NS, Tyakht AV, Popenko AS, Vasiliev AS, Altukhov IA, Ischenko DS, Shashkova TI, Efimova DA, Nikogosov DA, Osipenko DA, Musienko SV, Selezneva KS, Baranova A, Kurilshikov AM, Toshchakov SM, Korzhnikov AA, Samarov NI, Shevchenko MA, Tepliuk AV, Alexeev DG. 2018. Microbiome responses to an uncontrolled short-term diet intervention in the frame of the citizen science project. *Nutrients* 10:576. <https://doi.org/10.3390/nu10050576>
- van den Bogert B, Meijerink M, Zoetendal EG, Wells JM, Kleerebezem M. 2014. Immunomodulatory properties of *Streptococcus* and *Veillonella* isolates from the human small intestine microbiota. *PLoS One* 9:e114277. <https://doi.org/10.1371/journal.pone.0114277>
- Volokh O, Klimenko N, Berezhnaya Y, Tyakht A, Nesterova P, Popenko A, Alexeev D. 2019. Human gut microbiome response induced by fermented dairy product intake in healthy volunteers. *Nutrients* 11:547. <https://doi.org/10.3390/nu11030547>
- Dubois PCA, Trynka G, Franke L, Hunt KA, Romanos J, Curtotti A, Zhernakova A, Heap GAR, Adány R, Aromaa A, et al. 2010. Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet* 42:295–302. <https://doi.org/10.1038/ng.543>
- Fairfax BP, Humburg P, Makino S, Naranbhai V, Wong D, Lau E, Jostins L, Plant K, Andrews R, McGee C, Knight JC. 2014. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* 343:1246949. <https://doi.org/10.1126/science.1246949>
- Momozawa Y, Dmitrieva J, Théâtre E, Deffontaine V, Rahmouni S, Charleatoux B, Crins F, Docampo E, Elansary M, Gori A-S, et al. 2018. IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes. *Nat Commun* 9:2427. <https://doi.org/10.1038/s41467-018-04365-8>
- Araki K, Morita M, Bederman AG, Konieczny BT, Kissick HT, Sonenberg N, Ahmed R. 2017. Translation is actively regulated during the differentiation of CD8⁺ effector T cells. *Nat Immunol* 18:1046–1057. <https://doi.org/10.1038/ni.3795>
- Mottawea W, Chiang C-K, Mühlbauer M, Starr AE, Butcher J, Abujamel T, Deeke SA, Brandel A, Zhou H, Shokralla S, Hajibabaei M, Singleton R, Benchimol EI, Jobin C, Mack DR, Figeys D, Stintzi A. 2016. Altered intestinal microbiota-host mitochondria crosstalk in new onset Crohn's disease. *Nat Commun* 7:13419. <https://doi.org/10.1038/ncomms13419>
- Novak EA, Mollen KP. 2015. Mitochondrial dysfunction in inflammatory bowel disease. *Front Cell Dev Biol* 3:62. <https://doi.org/10.3389/fcell.2015.00062>
- Özsoy M, Stummer N, Zimmermann FA, Feichtinger RG, Sperl W, Weghuber D, Schneider AM. 2022. Role of energy metabolism and mitochondrial function in inflammatory bowel disease. *Inflamm Bowel Dis* 28:1443–1450. <https://doi.org/10.1093/ibd/izac024>
- Wang F, Fu X-D, Zhou Y, Zhang Y. 2009. Down-regulation of the cyclin E1 oncogene expression by microRNA-16-1 induces cell cycle arrest in human cancer cells. *BMB Rep* 42:725–730. <https://doi.org/10.5483/bmbrep.2009.42.11.725>
- Zhang L, Yu X, Zheng L, Zhang Y, Li Y, Fang Q, Gao R, Kang B, Zhang Q, Huang JY, Konno H, Guo X, Ye Y, Gao S, Wang S, Hu X, Ren X, Shen Z, Ouyang W, Zhang Z. 2018. Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. *Nature* 564:268–272. <https://doi.org/10.1038/s41586-018-0694-x>
- Fernández-Chacón R, Alvarez de Toledo G, Hammer RE, Südhof TC. 1999. Analysis of SCAMP1 function in secretory vesicle exocytosis by means of gene targeting in mice. *J Biol Chem* 274:32551–32554. <https://doi.org/10.1074/jbc.274.46.32551>

33. Fernández-Chacón R, Achiriloaie M, Janz R, Albanesi JP, Südhof TC. 2000. SCAMP1 function in endocytosis. *J Biol Chem* 275:12752–12756. <https://doi.org/10.1074/jbc.275.17.12752>
34. Ingham AC, Kielsen K, Cilieborg MS, Lund O, Holmes S, Aarestrup FM, Müller KG, Pamp SJ. 2019. Specific gut microbiome members are associated with distinct immune markers in pediatric allogeneic hematopoietic stem cell transplantation. *Microbiome* 7:131. <https://doi.org/10.1186/s40168-019-0745-z>
35. Lu D, Zhang J-B, Wang Y-X, Geng S-T, Zhang Z, Xu Y, Li S-Y, Wang K-H, Kuang Y-Q. 2021. Association between CD4⁺ T cell counts and gut microbiota and serum cytokines levels in HIV-infected immunological non-responders. *BMC Infect Dis* 21:742. <https://doi.org/10.1186/s12879-021-06491-z>
36. Pang R, Wang J, Xiong Y, Liu J, Ma X, Gou X, He X, Cheng C, Wang W, Zheng J, Sun M, Bai X, Bai L, Zhang A. 2022. Relationship between gut microbiota and lymphocyte subsets in Chinese Han patients with spinal cord injury. *Front Microbiol* 13:986480. <https://doi.org/10.3389/fmicb.2022.986480>
37. Visekruna A, Slavova N, Dullat S, Gröne J, Kroesen A-J, Ritz J-P, Buhr H-J, Steinhoff U. 2009. Expression of catalytic proteasome subunits in the gut of patients with Crohn's disease. *Int J Colorectal Dis* 24:1133–1139. <https://doi.org/10.1007/s00384-009-0679-1>
38. Richards AL, Muehlbauer AL, Alazizi A, Burns MB, Findley A, Messina F, Gould TJ, Cascardo C, Pique-Regi R, Blekhan R, Luca F. 2019. Gut microbiota has a widespread and modifiable effect on host gene regulation. *mSystems* 4:e00323-18. <https://doi.org/10.1128/mSystems.00323-18>
39. Fuess LE, den Haan S, Ling F, Weber JN, Steinel NC, Bolnick DI. 2021. Immune gene expression covaries with gut microbiome composition in stickleback. *mBio* 12:e00145-21. <https://doi.org/10.1128/mBio.00145-21>
40. Eicher T, Kinnebrew G, Patt A, Spencer K, Ying K, Ma Q, Machiraju R, Mathé AEA. 2020. Metabolomics and multi-omics integration: a survey of computational methods and resources. *Metabolites* 10:202. <https://doi.org/10.3390/metabo10050202>
41. Iablukov SN, Klimenko NS, Efimova DA, Shashkova T, Novichkov PS, Rodionov DA, Tyakht AV. 2020. Metabolic phenotypes as potential biomarkers for linking gut microbiome with inflammatory bowel diseases. *Front Mol Biosci* 7:603740. <https://doi.org/10.3389/fmolb.2020.603740>
42. Morgan XC, Kabackchiev B, Waldron L, Tyler AD, Tickle TL, Milgrom R, Stempak JM, Gevers D, Xavier RJ, Silverberg MS, Huttenhower C. 2015. Associations between host gene expression, the mucosal microbiome, and clinical outcome in the pelvic pouch of patients with inflammatory bowel disease. *Genome Biol* 16:67. <https://doi.org/10.1186/s13059-015-0637-x>
43. Kaakoush NO. 2015. Insights into the role of *Erysipelotrichaceae* in the human host. *Front Cell Infect Microbiol* 5:84. <https://doi.org/10.3389/fcimb.2015.00084>
44. Dinh DM, Volpe GE, Duffalo C, Bhalchandra S, Tai AK, Kane AV, Wanke CA, Ward HD. 2015. Intestinal microbiota, microbial translocation, and systemic inflammation in chronic HIV infection. *J Infect Dis* 211:19–27. <https://doi.org/10.1093/infdis/jiu409>
45. Lozupone CA, Li M, Campbell TB, Flores SC, Linderman D, Gebert MJ, Knight R, Fontenot AP, Palmer BE. 2013. Alterations in the gut microbiota associated with HIV-1 infection. *Cell Host & Microbe* 14:329–339. <https://doi.org/10.1016/j.chom.2013.08.006>
46. Han G, Luong H, Vaishnav S. 2022. Low abundance members of the gut microbiome exhibit high immunogenicity. *Gut Microbes* 14:2104086. <https://doi.org/10.1080/19490976.2022.2104086>
47. Kuznetsova V, Tyakht A, Akhmadishina L, Odintsova V, Klimenko N, Kostyukova E, Vakhitova M, Grigoryeva T, Malanin S, Vladimirov V, Nikitina R, Volok V, Osakovskiy V, Sivtseva T, Platonov F, Alexeev D, Karganova G. 2020. Gut microbiome signature of Viliuisk encephalomyelitis in Yakuts includes an increase in microbes linked to lean body mass and eating behaviour. *Orphanet J Rare Dis* 15:327. <https://doi.org/10.1186/s13023-020-01612-4>
48. Rath E, Moschetta A, Haller D. 2018. Mitochondrial function - gatekeeper of intestinal epithelial cell homeostasis. *Nat Rev Gastroenterol Hepatol* 15:497–516. <https://doi.org/10.1038/s41575-018-0021-x>
49. Wang A, Keita AV, Phan V, McKay CM, Schoultz I, Lee J, Murphy MP, Fernando M, Ronaghan N, Balce D, Yates R, Dicay M, Beck PL, MacNaughton WK, Söderholm JD, McKay DM. 2014. Targeting mitochondria-derived reactive oxygen species to reduce epithelial barrier dysfunction and colitis. *Am J Pathol* 184:2516–2527. <https://doi.org/10.1016/j.ajpath.2014.05.019>
50. Hall AB, Yassour M, Sauk J, Garner A, Jiang X, Arthur T, Lagoudas GK, Vatanen T, Fornelos N, Wilson R, Bertha M, Cohen M, Garber J, Khalili H, Gevers D, Ananthakrishnan AN, Kugathasan S, Lander ES, Blainey P, Vlamakis H, Xavier RJ, Huttenhower C. 2017. A novel *Ruminococcus gnavus* clade enriched in inflammatory bowel disease patients. *Genome Med* 9:103. <https://doi.org/10.1186/s13073-017-0490-5>
51. Png CW, Lindén SK, Gilshenan KS, Zoetendal EG, McSweeney CS, Sly LI, McGuckin MA, Florin THJ. 2010. Mucolytic bacteria with increased prevalence in IBD mucosa augment *in vitro* utilization of mucin by other bacteria. *Am J Gastroenterol* 105:2420–2428. <https://doi.org/10.1038/ajg.2010.281>
52. Qiu Z, Yang H, Rong L, Ding W, Chen J, Zhong L. 2017. Targeted metagenome based analyses show gut microbial diversity of inflammatory bowel disease patients. *Indian J Microbiol* 57:307–315. <https://doi.org/10.1007/s12088-017-0652-6>
53. Wang A-H, Li M, Li C-Q, Kou G-J, Zuo X-L, Li Y-Q. 2016. Human colorectal mucosal microbiota correlates with its host niche physiology revealed by endomicroscopy. *Sci Rep* 6:21952. <https://doi.org/10.1038/srep21952>
54. Holley RA. 2014. *Brochothrix*. Encyclopedia of Food Microbiology. <https://doi.org/10.1016/B978-0-12-384730-0.00048-3>
55. Kenny DJ, Plichta DR, Shungin D, Koppel N, Hall AB, Fu B, Vasan RS, Shaw SY, Vlamakis H, Balskus EP, Xavier RJ. 2020. Cholesterol metabolism by uncultured human gut bacteria influences host cholesterol level. *Cell Host & Microbe* 28:245–257. <https://doi.org/10.1016/j.chom.2020.05.013>
56. Chen Z, Radjabzadeh D, Chen L, Kurilshikov A, Kavousi M, Ahmadizar F, Ikram MA, Uitterlinden AG, Zernakova A, Fu J, Kraaij R, Voortman T. 2021. Association of insulin resistance and type 2 diabetes with gut microbial diversity: a microbiome-wide analysis from population studies. *JAMA Netw Open* 4:e2118811. <https://doi.org/10.1001/jamanetworkopen.2021.18811>
57. Sebastián VP, Salazar GA, Coronado-Arrázola I, Schultz BM, Vallejos OP, Berkowitz L, Álvarez-Lobos MM, Riedel CA, Kalergis AM, Bueno SM. 2018. Heme oxygenase-1 as a modulator of intestinal inflammation development and progression. *Front Immunol* 9:1956. <https://doi.org/10.3389/fimmu.2018.01956>
58. Onyiah JC, Sheikh SZ, Maharshak N, Steinbach EC, Russo SM, Kobayashi T, Mackey LC, Hansen JJ, Moeser AJ, Rawls JF, Borst LB, Otterbein LE, Plevy SE. 2013. Carbon monoxide and heme oxygenase-1 prevent intestinal inflammation in mice by promoting bacterial clearance. *Gastroenterology* 144:789–798. <https://doi.org/10.1053/j.gastro.2012.12.025>
59. Takagi T, Naito Y, Mizushima K, Akagiri S, Suzuki T, Hirata I, Omatsu T, Handa O, Kokura S, Ichikawa H, Yoshikawa T. 2010. Inhalation of carbon monoxide ameliorates TNBS-induced colitis in mice through the inhibition of TNF-α expression. *Dig Dis Sci* 55:2797–2804. <https://doi.org/10.1007/s10620-009-1112-x>
60. Zuckerbraun BS, Otterbein LE, Boyle P, Jaffe R, Upperman J, Zamora R, Ford HR. 2005. Carbon monoxide protects against the development of experimental necrotizing enterocolitis. *Am J Physiol Gastrointest Liver Physiol* 289:G607–13. <https://doi.org/10.1152/ajpgi.00055.2005>
61. Ohama T, Hori M, Momotani E, Iwakura Y, Guo F, Kishi H, Kobayashi S, Ozaki H. 2007. Intestinal inflammation downregulates smooth muscle CPI-17 through induction of TNF-α and causes motility disorders. *Am J Physiol Gastrointest Liver Physiol* 292:G1429–38. <https://doi.org/10.1152/ajpgi.00315.2006>
62. Ohama T, Hori M, Fujisawa M, Miyosue M, Hashimoto M, Ikenoue Y, Jinno Y, Miwa H, Matsumoto T, Murata T, Ozaki H. 2008. Downregulation of CPI-17 contributes to dysfunctional motility in chronic intestinal inflammation model mice and ulcerative colitis patients. *J Gastroenterol* 43:858–865. <https://doi.org/10.1007/s00535-008-2241-2>
63. Ozaki H, Hori M, Kinoshita K, Ohama T. 2005. Intestinal dysmotility in inflammatory bowel disease: mechanisms of the reduced activity of smooth muscle contraction. *Inflammopharmacology* 13:103–111. <https://doi.org/10.1163/156856005774423773>
64. Bouguen G, Dubuquoy L, Desreumaux P, Brunner T, Bertin B. 2015. Intestinal steroidogenesis. *Steroids* 103:64–71. <https://doi.org/10.1016/j.steroids.2014.12.022>
65. Noti M, Corazza N, Mueller C, Berger B, Brunner T. 2010. TNF suppresses acute intestinal inflammation by inducing local glucocorticoid synthesis. *J Exp Med* 207:1057–1066. <https://doi.org/10.1084/jem.20090849>
66. Henke MT, Kenny DJ, Cassilly CD, Vlamakis H, Xavier RJ, Clardy J. 2019. *Ruminococcus gnavus*, a member of the human gut microbiome

- associated with Crohn's disease, produces an inflammatory polysaccharide. *Proc Natl Acad Sci USA* 116:12672–12677. <https://doi.org/10.1073/pnas.1904099116>
67. Lehmann A, Hornby PJ. 2016. Intestinal SGLT1 in metabolic health and disease. *Am J Physiol Gastrointest Liver Physiol* 310:G887–98. <https://doi.org/10.1152/ajpgi.00068.2016>
 68. Park J, Lee I-S, Kim K-H, Kim Y, An E-J, Jang H-J. 2019. GI inflammation increases sodium-glucose cotransporter SglT1. *IJMS* 20:2537. <https://doi.org/10.3390/ijms20102537>
 69. Amador P, García-Herrera J, Marca MC, de la Osada J, Acín S, Navarro MA, Salvador MT, Lostao MP, Rodríguez-Yoldi MJ. 2007. Inhibitory effect of TNF- α on the intestinal absorption of galactose. *J Cell Biochem* 101:99–111. <https://doi.org/10.1002/jcb.21168>
 70. Kekuda R, Saha P, Sundaram U. 2008. Role of Sp1 and HNF1 transcription factors in SGLT1 regulation during chronic intestinal inflammation. *Am J Physiol Gastrointest Liver Physiol* 294:G1354–61. <https://doi.org/10.1152/ajpgi.00080.2008>
 71. Koepsell H. 2020. Glucose transporters in the small intestine in health and disease. *Pflugers Arch - Eur J Physiol* 472:1207–1248. <https://doi.org/10.1007/s00424-020-02439-5>
 72. Huang X, Fang S, Yang H, Gao J, He M, Ke S, Zhao Y, Chen C, Huang L. 2017. Evaluating the contribution of gut microbiome to the variance of porcine serum glucose and lipid concentration. *Sci Rep* 7:14928. <https://doi.org/10.1038/s41598-017-15044-x>
 73. Merigo F, Brandolese A, Facchin S, Missaggia S, Bernardi P, Boschi F, D'Incà R, Savarino EV, Sbarbati A, Sturniolo GC. 2018. Glucose transporter expression in the human colon. *World J Gastroenterol* 24:775–793. <https://doi.org/10.3748/wjg.v24.i7.775>
 74. York R. 2012. Residualization is not the answer: rethinking how to address multicollinearity. *Soc Sci Res* 41:1379–1386. <https://doi.org/10.1016/j.ssresearch.2012.05.014>
 75. Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4. *J Stat Softw.* <https://doi.org/10.18637/jss.v067.i01>
 76. Jari Oksanen FGB, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Peter Solymos M, Szoecs E, Wagner H. 2019. Vegan: community ecology package. <https://CRAN.R-project.org/package=vegan>.
 77. DirichletMultinomial. Available from: <https://bioconductor.org/packages/release/bioc/html/DirichletMultinomial.html>. Retrieved 2 Nov 2021. Accessed November 2, 2021
 78. Bougeard S, Dray S. 2018. Supervised multiblock analysis in R with the ade4 package. *J Stat Softw* 86:1–17. <https://doi.org/10.18637/jss.v086.i01>
 79. Antonov M, Csárdi G, Horvát S, Müller K, Nepusz T, Noom D, Salmon M, Traag V, Welles BF, Zanini F. 2023. Igraph enables fast and robust network analysis across programming languages. <https://doi.org/10.48550/arXiv.2311.10260>
 80. Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:9. <https://doi.org/10.1186/1471-2105-9-559>
 81. Sidak Z, Hajek J, Sen PK. 1999. Theory of rank tests. In *Probability and mathematical statistics*. academic press, 2nd ed. Academic Press.
 82. Bacanu SA, Devlin B, Roeder K. 2000. The power of genomic control. *Am J Hum Genet* 66:1933–1944. <https://doi.org/10.1086/302929>
 83. Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, Sidiropoulos K, Cook J, Gillespie M, Haw R, Loney F, May B, Milacic M, Rothfels K, Sevilla C, Shamovsky V, Shorser S, Varusai T, Weiser J, Wu G, Stein L, Hermjakob H, D'Eustachio P. 2020. The reactome pathway knowledgebase. *Nucleic Acids Res* 48:D498–D503. <https://doi.org/10.1093/nar/gkz1031>