UNIVERSITÉ DU LUXEMBOURG

PhD-FSTM-2024-088
Faculty of Science, Technology and Medicine

DISSERTATION

Defense held on 23 October 2024 in Esch-sur-Alzette

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG EN BIOLOGIE

by

Polina NOVIKOVA
Born on 2 September 1995 in Podolsk, Moscow Region (Russian Federation)

Systematic Characterization of Human Gut Microbiome in Relation to Parkinson's Disease

Dissertation defense committee

Dr. Paul Wilmes, Dissertation supervisor
Professor, Université du Luxembourg

Dr. Alexander Skupin, Chairman
Professor, University of Luxembourg, Luxembourg

Dr. Patrick May, Vice Chairman
Senior scientist, University of Luxembourg, Luxembourg

Dr. Mark Morrison
Professor, The University of Queensland, Australia

Dr. Almut Heinken
Professor, Université de Lorraine, France

Table of Contents

ABBREVIATIONS 4

ACKNOWLEDGMENTS 6

AFFIDAVIT 8

LIST OF PUBLICATIONS 9

MANUSCRIPTS INCLUDED IN THE THESIS 9

MANUSCRIPTS NOT INCLUDED IN THE THESIS 9

ABSTRACT 10

AIMS AND OBJECTIVES 12

MATERIALS AND METHODS 13

MATERIALS AND METHODS RELEVANT FOR PAPER 1 AND PAPER 2 13

COHORT DESCRIPTION 13

SEQUENCING DATA PROCESSING AND ANALYSIS 13

METABOLIC NETWORK CONSTRUCTION 14

CO-EXPRESSION NETWORK CONSTRUCTION 15

MATERIALS AND METHODS RELEVANT FOR PAPER 3 15

ARCHAEAL PROTEIN FUNCTIONAL ANNOTATION 16

SYNOPSIS: THE HUMAN MICROBIOME IN PD 17

STATE OF THE ART, PD AND MICROBIOME 17

MICROBIOME OF THE HUMAN GUT AND ITS IMPORTANCE IN PD 17

PARKINSON'S DISEASE PATHOGENESIS 24

PD ETIOLOGY, PRE-CLINICAL, PRODROMAL AND CLINICAL STAGES 26

DOPAMINE AND GLUTAMATE METABOLISM DYSREGULATION 28

Microbiome Links to PD: From Theory to Evidence 32
Meta-omics for Microbiome Research 35
Biological Networks **38**
Types of Biological Networks 40
Networks Topology and Metrics 41
Results and Discussion of Publications **45**

CONCLUSIONS AND PERSPECTIVES **159**

BIBLIOGRAPHY **169**

Abbreviations

PD	Parkinson’s disease
iRBD	idiopathic rapid-eye-movement sleep behavior disorder
REM	rapid eye movement
HC	healthy control
WGCNA	Weighted Gene Co-Expression Network Analysis
MG	metagenomic
MT	metatranscriptomic
MP	metaproteomic
MB	metabolomic
IMP	Integrated Meta-omic Pipeline
ORF	open-reading frame
KO	KEGG orthology
MAG	metagenome-assembled genome
TR	trRosetta
AF	AlphaFold
PF	ProFunc
ENS	enteric nervous system
CNS	central nervous system
SNpc	substantia nigra pars compacta
DOPAL	3,4-dihydroxyphenylacetaldehyde
ROS	reactive oxygen species

GABA	gamma-aminobutyric acid
SCFA	short-chain fatty acid
LPS	lipopolysaccharide
BA	bile acid
HDAC	histone deacetylase
BCAA	branched-chain amino acid
BCFA	branched-chain fatty acid
LBP	LPS-binding protein
CDCA	chenodeoxycholic acid
GCA	glycocholic acid
BMC	bacterial microcompartment
GSEA	gene set enrichment analysis
PTM	post-translational modification

Acknowledgments

First and foremost, I extend my deepest gratitude to my supervisors, Prof. Paul Wilmes and Dr. Patrick May, for the incredible opportunity to delve into microbiome research within the context of Parkinson's disease. Their guidance, expertise, and support throughout the past four years have deeply influenced both my research and my development as a person. I truly wouldn't be where I am today without them.

I also wish to thank the external jury members, Prof. Almut Heinken and Prof. Mark Morrison, for their participation in my defense committee and reviewing my thesis. My sincere thanks go to Prof. Alexander Probst for being on my CET committee and for the scientific discussions and encouragement throughout my PhD. Additionally, I am grateful to Prof. Alexander Skupin for kindly taking on the role of chairman for my defense at the University of Luxembourg.

To my colleagues in the Systems Ecology group, thank you for your practical support, engaging conversations, and the much-needed lunch breaks that provided a safe space to talk about life and science. I'm also deeply appreciative of my former colleagues – Velma Aho, Sofija Andric, Emna Bouhajja, Jordan Caussin, Ben Kunath, Francesco Delogu, Pedro Teixeira Queirós, Susheel Bhanu Busi, and Milena Despotovic – for their invaluable support and advice. Special thanks to everyone I've shared an office with during my four moves – each of you has contributed to making this journey memorable.

I am forever grateful to my dear friends who have been unwavering pillars of support throughout this PhD. Varya, Jane, Vova, and Denis, you've been there from the very beginning, reminding me of the significance of this journey – even in the toughest moments. Your words of encouragement and belief in me were priceless. To Charlotte De Rudder and Rémy Villette, I am thankful not only for the friendship that emerged during this time but also for our countless evenings filled with laughter. Rémy, you've been both a friend and mentor, and I've learned so much from you. A special shout-out to Marabu Lucchetti, my work bestie and cherished friend, for your endless support, advice, and warmth.

Спасибо моей любимой семье за веру в меня с самого начала. Спасибо за вашу неоценимую поддержку и вклад в мое будущее и достижения. Без вас ничего бы не получилось, и я надеюсь, что вы будете гордиться нашими общими успехами. Я вас люблю.

Lastly, to the most amazing man in my life, Kaan – this accomplishment is as much yours as it is mine. You stood by me in the darkest moments, reminding me of my purpose when I doubted myself. Thank you for wiping away my tears, for your calm presence, your wisdom, and for always knowing how to lift my spirits with your humour, boundless creativity and experimentalism. You remind me that together, "we can" overcome anything. I am eternally thankful for your boundless love, patience, and the deep faith you have in me.

Affidavit

I hereby declare that the dissertation entitled “Systematic Characterization of Human Gut Microbiome in Relation to Parkinson’s Disease” has been written only by the undersigned and without any assistance from third parties. Furthermore, I confirm that no sources have been used in the preparation of this thesis other than these indicated herein.

Luxembourg, October 2024

Polina Novikova

List of publications

Manuscripts included in the thesis

Published:

1. Novikova, P. V., Busi, S. B., Probst, A. J., May, P., & Wilmes, P. (2024). Functional prediction of proteins from the human gut archaeome. *ISME Communications, 4*(1), ycad014. https://doi.org/10.1093/ismeco/ycad014

In submission:

1. Villette, R.*, Ortís Sunyer, J.*, Novikova, P. V.*, Aho, V. T. E.*, Petrov, V. A., Hickl, O., ... Wilmes, P. (2024). Integrated multi-omics highlights alterations of gut microbiome functions in prodromal and idiopathic Parkinson's Disease.
2. Novikova P. V.*, Villette R.*, Laczny C. C., May P., Wilmes P. (2024) Microbiome Expression Network is Dysregulated in Parkinson's Disease Individuals.

Manuscripts not included in the thesis

1. Banas, I., Esser, S. P., Turzynski, V., Soares, A., Novikova, P., May, P., Moraru, C., Hasenberg, M., Rahlff, J., Wilmes, P., Klingl, A., & Probst, A. J. (2023). Spatio-functional organization in virocells of small uncultivated archaea from the deep biosphere. *The ISME Journal, 17*(10), 1789–1792. https://doi.org/10.1038/s41396-023-01474-1
2. Wilmes, P., Martin-Gallausiaux, C., Ostaszewski, M., Aho, V. T. E., Novikova, P. V., Laczny, C. C., & Schneider, J. G. (2022). The gut microbiome molecular complex in human health and disease. *Cell Host & Microbe, 30*(9), 1201–1206. https://doi.org/10.1016/j.chom.2022.08.016
3. Queirós, P., Novikova, P., Wilmes, P., & May, P. (2021). Unification of functional annotation descriptions using text mining. *Biological Chemistry, 402*(8), 983-990. https://doi.org/10.1515/hsz-2021-0125

Abstract

The human gut microbiome, which includes a diverse array of microorganisms such as bacteria, archaea, and viruses, plays a crucial role in maintaining overall health and influencing disease outcomes. This microbiome integrates environmental, genetic, and immune signals to support various physiological functions, including digestion, immune regulation, and detoxification. Dysregulation of the gut microbiome has been implicated in several diseases, including Parkinson's Disease (PD). Although PD is not traditionally associated with gut disorders, emerging evidence links microbial imbalances in the gut to disease onset and progression. This connection is supported by the observation that PD-related protein aggregations and gastrointestinal symptoms often precede motor symptoms. In PD, there is an elevated abundance of pro-inflammatory bacteria and a reduction in beneficial bacterial species. Furthermore, an increased presence of methanogenic archaea, particularly *Methanobrevibacter smithii*, has been observed, indicating a potential involvement in the gut-related symptoms frequently associated with PD. These findings underscore the importance of the gut-brain axis and highlight the need for further research into how gut microbiota may contribute to neurodegenerative diseases like PD.

This thesis investigates the role of the gut microbiome in PD through a range of computational and analytical methods. The primary objective of this work is to elucidate the complex interactions between microbial functions and PD using advanced meta-omics and network-based techniques. A key finding of this research is the alteration in microbial structure and function associated with PD, particularly the elevated levels of β-glutamate linked to specific microbial genera. This study identifies glutamate metabolism as a central process within the PD-associated microbiome, highlighting disruptions that correlate with decreased transcript abundances in chemotaxis and flagellar assembly among PD-related taxa. The reduction in flagellin transcription by certain bacteria in PD indicates intricate interactions between microbial changes and host immune responses. Further analysis utilizing Weighted Gene Co-Expression Network Analysis (WGCNA) revealed significant differences in co-expression patterns in PD. Notably, modules of co-expressed genes in healthy controls (HC) demonstrated greater functional diversity, while PD was characterized by reduced gene diversity and specific metabolic alterations, including glycerolipid metabolism, peptidoglycan biosynthesis, lipoic acid metabolism, and valine degradation. The network-based approach confirmed significant enrichment in flagellar assembly among HC, alongside the identification of secondary bile acid biosynthesis as an enriched process.

Additionally, our study revealed significant alterations in bacterial microcompartments (BMCs) within certain commensal bacteria, exhibiting a strong correlation with flagellar assembly genes, further underscoring their interconnected roles in PD. To our knowledge, this work is the first to establish a link between BMCs and flagellar assembly in the context of PD, revealing essential microbiome functions that are disrupted in this disease. These findings offer valuable insights into the microbial mechanisms contributing to PD pathogenesis and lay the groundwork for future experimental validation.

The study also explored the role of intestinal archaea, particularly *Methanobrevibacter smithii*, in PD. This archaeon, known for its involvement in gastrointestinal disorders, was found to have significant interactions with gut microbiome functions, with implications for gastrointestinal symptoms commonly seen in PD. Advanced protein structure prediction identified gut-specific archaeal proteins potentially involved in defense mechanisms, virulence, adhesion, and the degradation of toxic substances. Preliminary evidence also suggested interdomain horizontal gene transfer between *Clostridia* species and *M. smithii*, based on structure-based protein annotation.

In conclusion, this thesis underscores the significant role of the gut microbiome in the pathogenesis of PD. Through comprehensive computational and analytical methods, the research highlights critical alterations in microbial structure and function, particularly in glutamate metabolism and microbial diversity. The study also brings to light the involvement of intestinal archaea, such as *Methanobrevibacter smithii*, in microbiome function. These insights pave the way for future research aimed at understanding the gut-brain axis and developing microbiome-based interventions for neurodegenerative diseases like PD.

Aims and Objectives

This study aims to advance the field of microbiome research by deepening the understanding of microbial ecology and the interactions between human-associated microbiota and health. The following key objectives guided this research:

1. To utilize an integrated multi-omics approach to elucidate the functional activities and alterations of the gut microbiome in Parkinson's disease (PD) and its prodromal stage idiopathic rapid-eye-movement sleep behavior disorder (iRBD), by identifying and characterizing differential metabolite levels, gene expression patterns, and microbial functions between PD individuals, iRBD individuals, and healthy controls.
2. To investigate the dysregulation of microbial co-expression networks in PD, with the goal of identifying key regulatory mechanisms and altered biological processes that contribute to PD pathogenesis.
3. To advance the functional annotation of archaeal proteins, by utilizing advanced computational and structural methodologies to uncover their roles in the human gut microbiome and their potential implications for gut-related disorders.

Materials and Methods

Materials and Methods relevant for Paper 1 and Paper 2

The studies titled “Integrated Multi-omics Highlights Alterations of Gut Microbiome Functions in Prodromal and Idiopathic Parkinson’s Disease” and “Microbiome Expression Network is Dysregulated in Parkinson’s Disease Individuals” were conducted using the same cohort of patients and healthy individuals. This section will outline the primary methods and techniques employed in these investigations. For a comprehensive understanding, readers are encouraged to refer to the manuscripts available in the “Results and Discussion of Publications” chapter.

Cohort description

All subjects provided informed written consent, and the sample analysis was approved by the Comité National d’Ethique de Recherche of Luxembourg. The Kassel cohort involved a prospective, biannual follow-up study of PD patients at the Paracelsus-Elena Klinik, Germany, with fecal samples collected during a 4-year follow-up visit. The Marburg cohort recruited patients with polysomnography-confirmed idiopathic rapid-eye-movement sleep behavior disorder from the Department of Neurology, Philipps-University, Germany. Fecal samples were collected, flash-frozen, and stored at –80 °C. The initial set of samples included 50 PD patients, 30 idiopathic rapid-eye-movement sleep behavior disorder patients, and 50 healthy control subjects. However, some samples were excluded due to clinical reasons such as adjusted diagnoses, logistical issues, and the use of microbiome-altering medications like metformin, antidepressants, statins, and proton pump inhibitors. Additionally, samples were excluded due to missing values in metabolomics or a low number of identified analytes in metaproteomics. These exclusions led to the final sample numbers for various analyses: 49 HC, 27 iRBD, and 46 PD for metagenomics and metatranscriptomics; 42 HC, 22 iRBD, and 40 PD for metaproteomics; and 49 HC, 27 iRBD, and 41 PD for meta-metabolomics.

Sequencing data processing and analysis

Sequencing data utilized in the studies “Integrated Multi-omics Highlights Alterations of Gut Microbiome Functions in Prodromal and Idiopathic PD” and “The gut microbiome gene expression network is dysregulated in individuals with Parkinson’s disease” were obtained from fecal samples of PD patients, iRBD patients, and HC. Metagenomic (MG) data was

employed to investigate the functional potential and structure of microbial communities in these samples. Additionally, metatranscriptomic (MT) data was used to analyze gene expression, providing insights into microbial community function.

MG and MT data were processed and hybrid-assembled using the Integrated Meta-omic Pipeline (IMP) (https://git-r3lab.uni.lu/IMP/imp3, commit 8c1bd6fa443d064511909c9eede20703f45e6c69) (Narayanasamy et al., 2016). This pipeline encompasses several steps, including the trimming and quality filtering of reads, the exclusion of rRNA sequences from the MT data, and the elimination of human reads by mapping against the human genome (hg38). The pre-processed MG and MT reads were then assembled through the IMP-based iterative hybrid-assembly pipeline utilizing MEGAHIT (version 1.0.3) (D. Li et al., 2015). Post-assembly, structural features such as open-reading frames (ORFs) were predicted and annotated using a modified version of Prokka (Seemann, 2014), followed by functional annotation with Mantis (Queirós et al., 2021). Quantification of these structural features at both MG and MT levels was performed using featureCounts (Liao et al., 2014). Taxonomic annotation of reads and contigs was conducted using Kraken2 (Wood et al., 2019) with the GTDB (release 207) database (http://ftp.tue.mpg.de/ebio/projects/struo2/GTDB_release207/kraken2) and a confidence threshold of 0.5. Additionally, taxon abundances were estimated using mOTUs (version 2.5.1) (Milanese et al., 2019). These mOTU abundances were utilized to create abundance matrices for each taxonomic rank (phylum, class, order, family, genus, and species) by aggregating taxon marker read counts at the respective levels.

Metabolic network construction

In the study “Integrated Multi-omics Highlights Alterations of Gut Microbiome Functions in Prodromal and Idiopathic PD” we utilized metabolic network-based approach to highlight the importance of metabolites and related genes in the microbial metabolism. We reconstructed metabolic networks as previously described in the study by (Roume et al., 2015). The metabolic network analysis was performed by linking KEGG Orthology (KO) identifiers with corresponding ChEBI IDs and visualized using the NetworkX package (version 3.3) (Hagberg et al., 2008). In this network, nodes represented KEGG KOs, while edges corresponded to the associated metabolites. Such compounds as water, energy transporters and cofactors were removed, to only consider main compounds of a given reaction. The analysis focused on genes present in at least 50% of the samples. Gene expression effect size was quantified by normalizing gene expression, calculated as the

ratio of MT gene expression to MG gene abundance as follows: $\frac{Gene\ MT\ expression}{Gene\ MG\ abundance}$. To construct metabolite-specific networks, KEGG KOs associated with specific metabolites in KEGG were used. Subnetworks specific to glutamate, thymine, glycerol, serine, alanine, and glucuronate were extracted by searching for the given metabolite in the list of metabolites associated with KEGG KOs. The network topology metric 'betweenness centrality' was employed to highlight the significance of each metabolite in microbial metabolism.

Co-expression network construction

In our study "The gut microbiome gene expression network is dysregulated in individuals with Parkinson's disease" we attempted to decipher complex interactions within the human gut microbiome using the Weighted Gene Co-Expression Network Analysis (WGCNA) approach (Wan et al., 2018; B. Zhang & Horvath, 2005). This analysis was entirely unsupervised, with no prior filtering based on differential expression or function. Using this method, we were able to find correlation patterns of genes across a large set of PD and HC samples, and modules of co-expressed genes as well as their correlation with external traits such as disease state.

Co-expression patterns were analysed using the pyWGCNA package implemented in Python (Rezaie et al., 2023). Of the 8334 genes, 4879 genes were included in this analysis, only genes present in at least 50% of samples were kept. Gene abundance normalized expression as described previously was used as gene values. WGCNA function was run on power transformed data with *sklearn.preprocessing.PowerTransformer()* (https://scikit-learn.org/stable/api/sklearn.preprocessing.html) with the following parameters: minimum module size minModuleSize=20, dissimilarity threshold MEDissThres=0.18, networkType='signed'.

Materials and Methods relevant for Paper 3

Our study, titled "Functional Prediction of Proteins from the Human Gut Archaeome," extends the understanding of microbiome functions by specifically focusing on intestinal archaea. To achieve this, we utilized publicly available metagenomic data from the Genomes from Earth's Microbiomes (GEM) catalog (Nayfach et al., 2020) and the Unified Human Gastrointestinal Genome (UHGG) collection (Almeida et al., 2020), as well as bacterial metagenome-assembled genomes (MAGs) from the UHGG collection. Figure 1 ((Novikova et al., 2024)) depicts the workflow employed in this study, detailing the

integration and application of sequence and structure annotation methods for the functional annotation of archaeal proteins.

Archaeal protein functional annotation

In our study “Functional Prediction of Proteins from the Human Gut Archaeome” we utilized protein annotation based on both sequences and structures of proteins in question. Sequence annotation was performed with KEGG orthologs (KOs) using Mantis (1.5.4) (Queirós et al., 2021). We employed trRosetta (TR) (Du et al., 2021) and AlphaFold (AF) (Jumper et al., 2021) for structure prediction, annotating each predicted structure separately. TR-based models used high-identity and high-coverage templates, while AF models were annotated via the ProFunc (PF) web server (Laskowski et al., 2005). Only highly certain matches were used for functional assignments. DeepFri was used as an auxiliary tool to verify or refute annotations (Gligorijević et al., 2021).

Synopsis: The Human Microbiome in PD

State of the Art, PD and Microbiome

Microbiome of the Human Gut and Its Importance in PD

The human gut microbiome, consisting of commensal, symbiotic, and pathogenic microorganisms such as bacteria, archaea, microeukaryotes, and viruses, has emerged as a significant contributor to various diseases. The gastrointestinal tract, which contains the largest microbial biomass, encodes a genetic repertoire vastly exceeding that of human genes (Miyauchi et al., 2022). Thus, the gut microbiome acts as a central hub, integrating environmental inputs with genetic and immune signals to influence host physiology.

In a healthy state, the gut microbiome performs vital functions, including digesting dietary components, synthesizing vitamins, regulating the immune system, outcompeting pathogens, detoxifying harmful substances, and supporting intestinal function. These interconnected functions contribute to overall human metabolism, with microbial metabolites playing essential roles in immunomodulation (Fig. 1). The gut microbiome also interacts with other body systems through the circulatory, immune, endocrine, and nervous systems. The functional repertoire of the gut microbiome results in the production of diverse biomolecules that stimulate the host's immune system (Wilmes et al., 2022).

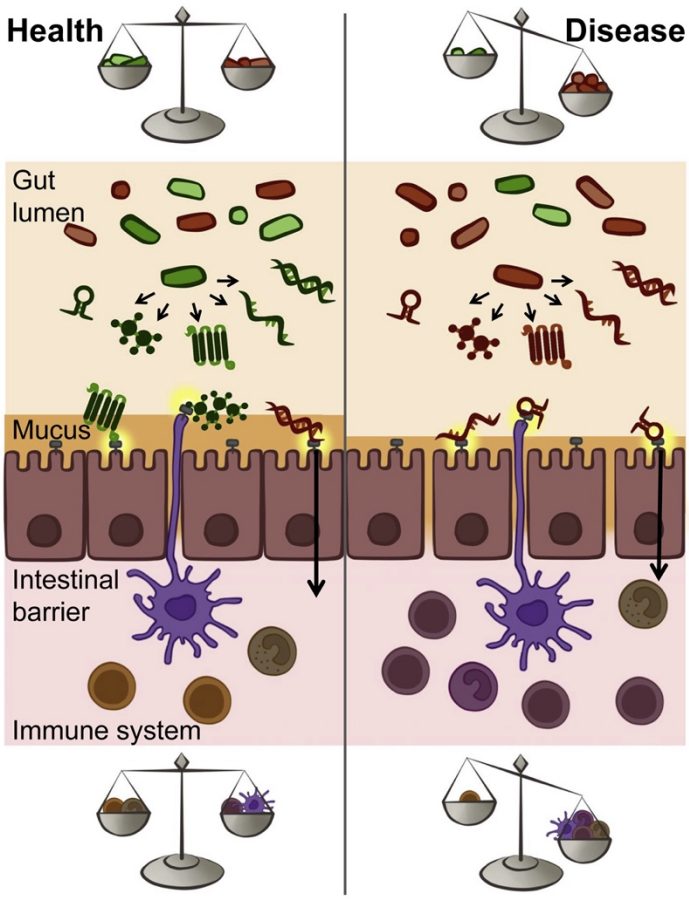


Figure 1. Microbiome-derived molecules that trigger inflammatory processes. In diseases with inflammatory signatures, the balance between cytotoxic and anti-inflammatory, pro-healing immune activation is dysregulated, reflecting microbial dysbiosis. Adapted from Wilmes et al., 2022.

Systematic studies of the gut microbiome-derived biomolecular complex have highlighted the uniqueness of extracellular biomolecular fractions (DNA, RNA, peptides, and metabolites) in terms of their taxonomic and functional affiliations within and between individuals (De Saedeleer et al., 2021). These microbiome-derived molecules are detected by epithelial, innate immune, and dendritic cells, which connect to adaptive immunity. Due to their continuous exposure to microbial molecules in the gut, epithelial and immune cells have adapted to tolerate beneficial microbiota while still protecting against harmful pathogens. Additionally, the gut microbiome modulates host responses to microbial molecules to stabilize its niche by stimulating antimicrobial peptide production and regulating mucosal homeostasis (Blander et al., 2017).

Microbial dysbiosis likely leads to differential enrichments in microbiome-derived molecules, either as a cause or consequence of disrupted microbiome-immune system interactions

(Fig. 1). Functional differences in the microbiome may more clearly distinguish healthy from diseased individuals than taxonomic changes alone (Heintz-Buschart & Wilmes, 2018). Our understanding of how variations in microbial taxa abundance affect diseases is still quite limited. This lack of knowledge hampers our ability to identify the crucial roles these microbes play in maintaining human health. Furthermore, we need to explore how changes in the microbiome's composition and function might lead to the onset and progression of diseases over a person's lifetime (Wilmes et al., 2022).

Parkinson's disease (PD) is not traditionally associated with the gut, yet studies reveal significant changes in microbial taxonomy in PD patients (Heintz-Buschart et al., 2018; Romano et al., 2021). PD is linked to the gut through several mechanisms. The primary pathogenic characteristics of PD include the progressive degradation of specific neurons in brain areas like the substantia nigra. Furthermore, aggregations of the protein α-synuclein (α-Syn), a hallmark of PD, have been observed in both the central and peripheral nervous systems, suggesting a potential origin in the gut (Heintz-Buschart et al., 2018). Epidemiological studies indicate a decreased risk of PD following complete truncal vagotomy, implicating the vagus nerve in the disease's progression from the gut to the brain (Heintz-Buschart et al., 2018; Sampson et al., 2016). Additionally, inflammation and increased permeability of the colonic mucosal lining are common in PD, creating an environment that promotes α-Syn aggregation and disease progression, which is influenced by the gut microbiome (Clairembault et al., 2015; Schwiertz et al., 2018). Studies have identified over 100 taxa with differing abundances between PD patients and controls, with some linking specific taxa to disease severity (Romano et al., 2021).

Therefore, studies on PD underscore the significant role of the gut microbiome and its functions in diseases not traditionally linked to the gut. This underscores the need to consider the microbiome's impact on human physiology alongside environmental, genetic, and immune factors.

Immuno-modulation of PD by Microbiome: Modern Approach to PD Etiology

In general, the gut microbiota comprises a complex and diverse community of bacteria, archaea, fungi, and viruses, which together maintain physiological homeostasis and influence various body functions. In PD, significant alterations in the gut microbiota composition have been observed, often characterized by a decrease in beneficial bacteria and an increase in potentially harmful bacteria (Keshavarzian et al., 2015; Scheperjans et

al., 2015). These alterations are not only biomarkers but also likely contributors to disease pathogenesis.

Recent research into PD has expanded beyond traditional neurocentric views, increasingly highlighting the role of gut microbiota and their metabolic products. These microbial metabolites and molecules play pivotal roles in the gut-brain axis, influencing neurological processes and potentially contributing to the pathogenesis of neurodegenerative diseases like PD. This section explores key molecular and cellular factors that are relevant in the context of the microbiome and PD, focusing on short-chain fatty acids (SCFAs), lipopolysaccharides (LPS), neurotransmitters, bile acids (BAs), and flagella.

Short-chain fatty acids

In PD, the production of SCFAs is often disrupted, reflecting broader alterations in gut microbiota composition. Studies have consistently shown that individuals with PD exhibit a reduction in SCFA-producing bacteria, such as those from the genera *Prevotella*, *Roseburia*, and *Faecalibacterium* (Keshavarzian et al., 2015). This decrease is important because SCFAs contribute to the maintenance of the intestinal barrier, which in turn prevents translocation of bacteria and their endotoxins, such as lipopolysaccharides, into the bloodstream. The translocation of these substances can trigger systemic inflammation and potentially exacerbate neuroinflammation, thereby promoting PD progression.

SCFAs play a crucial role in the regulation of the host's immune system, the maintenance of intestinal barrier integrity, and the modulation of energy metabolism. SCFAs exert their influence on multiple physiological systems, including gut barrier function, the vagus nerve, the enteric nervous system, immune function, and the integrity of the blood-brain barrier (Aho et al., 2021; Dalile et al., 2019; Liddle, 2018; Silva et al., 2020). Beyond their role as immune modulators, certain SCFAs, such as propionate and butyrate, have demonstrated the ability to inhibit neuroinflammation by suppressing cytokine storms and viral pathogenesis (Majumdar et al., 2023; McCarville et al., 2020). This anti-inflammatory action is particularly relevant in the context of PD, where neuroinflammation is a key pathological feature. Specifically, butyrate has been shown to reduce the permeability of the blood-brain barrier, mitigate microbial activation, and alleviate depressive symptoms that are commonly associated with PD (Xie et al., 2022).

By reducing systemic and neural inflammation, SCFAs like butyrate could potentially slow the progression of neurodegeneration in PD (Cryan & Dinan, 2012). Recent research has also explored the therapeutic potential of SCFAs in PD (J. Liu et al., 2024). Experimental

models of PD have shown that supplementation with SCFAs or SCFA-producing probiotics can restore gut microbial balance and reduce neuroinflammation (Macfarlane et al., 2008)

Acetate, the most abundant SCFA, is vital for maintaining energy balance and is involved in lipid metabolism and appetite regulation. It also serves as a substrate for the synthesis of other SCFAs, such as butyrate. Propionate, though less abundant, is crucial for hepatic gluconeogenesis and has immunomodulatory effects, particularly in promoting the generation of regulatory T cells, which are important for maintaining immune homeostasis. Additionally, propionate plays a significant role in neuroprotection by ameliorating motor deficits and dopaminergic neuronal loss in PD models, primarily through its interaction with the FFAR3 receptor in the enteric nervous system, which modulates gut microbiota and reduces neuroinflammation (Hou et al., 2021). Butyrate, well-known for its anti-inflammatory properties, supports gut health by serving as the primary energy source for colonocytes and by maintaining intestinal barrier integrity. It also inhibits histone deacetylases (HDACs), reducing inflammation and oxidative stress, which are all key factors in neurodegenerative diseases like PD (S. Wu et al., 2012).

Valerate, isobutyrate, and isovalerate, although less abundant than other SCFAs , play crucial roles in gut microbial ecology and host metabolism (J. Liu et al., 2024). Valerate has been shown to enhance gut barrier function and protect dopaminergic neurons (Jayaraj et al., 2020; Y. Li et al., 2020), contributing to overall gut health and brain function. Isobutyrate is produced through the fermentation of branched-chain amino acids (BCAAs) like valine in the colon. It has been shown to stimulate colonic sodium absorption, which is crucial for maintaining fluid and electrolyte balance in the gut (Rios-Covian et al., 2020). Additionally, isobutyrate contributes to the overall health of the gut microbiota by promoting the growth of beneficial bacteria (Peterson et al., 2022). Isovalerate is branched-chain saturated fatty acid (BCFA) derived from the fermentation of BCAAs, specifically leucine. It has been associated with the regulation of glucose and lipid metabolism, suggesting a role in maintaining metabolic health (Rios-Covian et al., 2020). High levels of isovalerate in feces have also been linked to human depression and elevated cortisol levels, indicating its potential impact on mental health. Furthermore, isovalerate, along with other BCFAs, has been found to increase the relative abundance of B vitamin-producing bacteria, which are essential for various metabolic functions (Peterson et al., 2022).

In conclusion, the disruption of SCFA production and balance in PD underscores the critical role of gut microbiota in maintaining intestinal barrier integrity and regulating systemic inflammation. The reduction in SCFA-producing bacteria in PD patients highlights the

importance of these metabolites in preventing neuroinflammation and supporting overall gut health. SCFAs such as acetate, propionate, and butyrate, along with less abundant BCFAs like valerate, isobutyrate, and isovalerate, play multifaceted roles in immune regulation, energy metabolism, and neuroprotection. Their therapeutic potential in restoring gut microbial balance and reducing neuroinflammation offers promising avenues for managing PD and enhancing overall neurological health.

Lipopolysaccharides

Lipopolysaccharides (LPS), components of the outer membrane of Gram-negative bacteria, play a crucial role in the immune response, eliciting potent inflammatory reactions by activating the innate immune system. In the context of PD, elevated levels of LPS-binding protein (LBP) have been correlated with an increased risk of developing the disease, suggesting that endotoxemia, characterized by the presence of LPS in the bloodstream, may contribute to PD pathogenesis, particularly through neuroinflammatory mechanisms (Y. Zhao et al., 2023).

Disruptions in gut microbiota composition have been linked to increased intestinal permeability, commonly referred to as a “leaky gut,” which allows bacterial products like LPS to enter the bloodstream and potentially reach the brain (X. Zhang et al., 2023). This compromised intestinal barrier is a critical factor in PD development. The increased permeability permits not only LPS but also other harmful substances and pathogens to translocate from the gut into the bloodstream, triggering systemic inflammation (Q. Li et al., 2023). This systemic inflammation exacerbates neuroinflammatory processes in the brain, significantly contributing to the progression of PD.

Once LPS enters the bloodstream, it can activate microglia, the resident immune cells in the brain, leading to chronic neuroinflammation (Y. Zhao et al., 2023). Additionally, the transmission of inflammatory signals from the gut to the brain via the vagus nerve amplifies neuroinflammation and neuronal damage. Chronic microglial activation results in the release of pro-inflammatory cytokines and reactive oxygen species, further exacerbating neuronal damage and accelerating the neurodegenerative process (Muzio et al., 2021; Woodburn et al., 2021).

In summary, LPS and disruptions in gut microbiota significantly contribute to PD through mechanisms involving neuroinflammation. Elevated LPS levels, coupled to a compromised intestinal barrier, allow harmful substances to enter the bloodstream and may culminate in

systemic inflammation. This inflammation, in turn, activates microglia in the brain, leading to chronic neuroinflammation and the degeneration of dopaminergic neurons.

Bile acids

Bile acids (BAs) are another significant class of microbiome-derived metabolites that have been increasingly associated with PD. These compounds, synthesized initially by the host in the liver from cholesterol, play a crucial role in the digestion and absorption of dietary fats. Upon their release into the gut, primary bile acids are subjected to further metabolism by the gut microbiota, transforming them into secondary bile acids with distinct biochemical properties (Ridlon et al., 2006). This microbial transformation is not merely a digestive process; it profoundly influences the gut's immunological landscape and systemic health. Emerging research has highlighted that secondary bile acids possess a range of cytotoxic and immunomodulatory activities, which can impact the progression of neurodegenerative diseases such as PD. For instance, alterations in the composition of gut microbiota in PD patients may lead to an imbalance in bile acid metabolism, which in turn can contribute to the disease's pathology through mechanisms such as inflammation and cellular stress (Hurley et al., 2022; Li et al., 2021). Studies have observed changes in bile acid profiles in the plasma, serum, and stool samples of individuals with PD, suggesting that these metabolites might serve as biomarkers for disease progression and severity (Figura et al., 2018; Hertel et al., 2019; Hirayama et al., 2016; P. Li et al., 2021b; Vascellari et al., 2020). The immunomodulatory capacities of secondary bile acids are of particular interest, as they can interact with various receptors in the gut and other tissues, influencing immune responses and potentially exacerbating or mitigating neuroinflammation (Wahlström et al., 2016). Understanding the precise role of bile acids in PD not only offers insights into the gut-brain axis's involvement in the disease but also opens up potential avenues for therapeutic intervention aimed at modulating bile acid metabolism and its downstream effects.

Flagella

Flagella, the whip-like appendages that facilitate bacterial motility, are highly immunogenic. Flagellin, the protein constituting flagella, is a potent antigen and pro-inflammatory agent in pathogens (Gram et al., 2021; F. Qian et al., 2015; Tran et al., 2019). These structures can elicit robust immune responses, thereby influencing the gut-brain axis. In mammalian hosts, flagellar motility enables bacteria to evade clearance mechanisms, access epithelial cells, and reach nutrient-rich niches. This mobility is associated with pathogenic activities such as

epithelial invasion, translocation across epithelial barriers, and biofilm formation, which disrupt homeostasis and reduce host reproductive fitness (Akahoshi & Bevins, 2022). Bacterial movement driven by flagella is directed by chemotaxis, allowing bacteria to move toward beneficial chemical gradients and away from harmful ones (Colin et al., 2021). Flagella and flagellin are involved in inflammatory responses by inducing pro-inflammatory cytokines such as IL-8 and TNF-α (Cruz-Córdova et al., 2012). Through the activation of Toll-like receptor 5 (TLR5), flagella can also trigger the MAPK and NF-κB signaling pathways, leading to further cytokine production (Batah et al., 2017). Conversely, flagella can stimulate the production of anti-inflammatory cytokines like IL-10, which helps modulate immune responses and maintain gut homeostasis (Cruz-Córdova et al., 2012). Flagellin has also demonstrated potential in anti-tumor and radioprotective therapies (Sfondrini et al., 2006; Vijay-Kumar et al., 2008). Also, metagenomic analyses have reported a decrease in flagellar assembly genes in PD patients (Boktor et al., 2023). The dual role of flagella in both pro- and anti-inflammatory responses underscores their complex involvement in PD, making them crucial for understanding the disease's etiology.

Parkinson's Disease Pathogenesis

PD is a progressively debilitating neurodegenerative disorder that primarily impairs the motor system, leading to a spectrum of clinical manifestations that severely affect an individual's quality of life. The hallmark motor symptoms of PD include involuntary or uncontrollable movements such as tremors, muscle rigidity, bradykinesia (slowness of movement), and postural instability, which collectively contribute to significant difficulties in balance, coordination, and overall motor function. These motor deficits typically emerge insidiously and tend to exacerbate as the disease advances, eventually culminating in profound impairments that hinder an individual's ability to walk, speak, and execute routine daily activities (Jankovic, 2008; Sveinbjornsdottir, 2016). In addition to these well-documented motor symptoms, PD is increasingly recognized for its wide array of non-motor symptoms, which often manifest long before the onset of motor impairments (Adams-Carr et al., 2016; Blesa et al., 2021). These non-motor symptoms encompass a broad spectrum of cognitive, psychiatric, and autonomic disturbances, including sleep disorders, depression, constipation, anxiety, cognitive decline, memory impairment, and pervasive fatigue (Roos et al., 2022). The early appearance of these non-motor symptoms – often preceding the motor symptoms by a decade or more – suggests that PD pathology may begin long before the clinical diagnosis is made, and that these symptoms are critical in understanding the full impact of the disease on patients' lives. The recognition of non-motor

symptoms as integral components of PD has profound implications for early diagnosis, patient management, and the development of therapeutic strategies aimed at mitigating the overall burden of the disease (Jankovic, 2008; Sveinbjornsdottir, 2016).

PD, while having the potential to impact any individual, demonstrates a notable epidemiological trend with a higher incidence observed in males compared to females (Cerri et al., 2019). Despite extensive research, the precise reasons underlying this sex disparity remain elusive, though hormonal, genetic, and environmental factors are often hypothesized as influencing factors (Savica et al., 2016). Age emerges as a significant risk factor for PD, with epidemiological data indicating that more than 1% of the population aged 60 and above is afflicted by the condition. However, it is crucial to note that a subset of PD cases, estimated between 5% and 10%, manifests before the age of 50. These early-onset cases are frequently attributed to genetic variants, underscoring the role of hereditary factors in disease etiology (Funayama et al., 2022; Kolicheski et al., 2022; Pitz et al., 2024). Nevertheless, the majority of PD cases occur sporadically, with only 10% of patients having causative genetic variants. Consequently, the remaining cases are classified as idiopathic PD (Tredici & Braak, 2013).

Despite this genetic predisposition, the pattern of inheritance in many cases remains unclear, suggesting a complex interplay between genetic susceptibility and other factors. In addition to genetic predisposition, environmental factors have been increasingly recognized as significant contributors to the onset and progression of PD. Epidemiological surveys have consistently shown that individuals residing in rural areas, particularly those engaged in agricultural work, exhibit a higher relative risk of developing PD compared to their urban counterparts (Corsini et al., 1985; Ngo et al., 2024; Perrin et al., 2021; Pouchieu et al., 2018; Seidler et al., 1996). This heightened risk is strongly suspected to be associated with chronic exposure to pesticides and herbicides, environmental toxins that have been linked to neurodegenerative processes. Such findings underscore the multifactorial nature of PD, where both genetic and environmental factors converge to influence disease pathogenesis (Kline et al., 2021; Tsalenchuk et al., 2023; Yuan et al., 2022).

Recent advancements in molecular genetics have underscored the intricate interplay between genetic predispositions, aging, and environmental exposures in the pathogenesis of PD. These studies illuminate how specific genetic variants, when combined with the natural aging process and environmental factors such as toxin exposure, can significantly increase the risk of developing PD. Among these contributing factors, an emerging and compelling body of research has highlighted the critical role of gut microbiome dysbiosis in

both the onset and progression of PD (Miyauchi et al., 2022; Nie & Ge, 2023; K. Zhang et al., 2022). This dysbiosis, or imbalance in the gut microbiota, has been increasingly recognized as a key player in PD pathogenesis, suggesting that the disease is not merely confined to the central nervous system (CNS) but also involves significant peripheral mechanisms. The gut-brain axis, a sophisticated and bidirectional communication network that links the gastrointestinal tract and the CNS, is central to this evolving understanding. This axis enables the gut microbiome to influence brain function and behavior, while also allowing the brain to affect gastrointestinal processes. Emerging evidence suggests that disruptions of the gut-brain communication pathway, often mediated by microbiome alterations, can contribute to the neuroinflammatory and neurodegenerative processes characteristic of PD (Boertien et al., 2022; Cryan & Dinan, 2012; Hashish & Salama, 2023; Z. Li et al., 2023). This growing recognition of the gut-brain axis in PD represents a paradigm shift, broadening the scope of PD research and treatment from a disorder traditionally viewed as being limited to the CNS to one that involves systemic interactions, particularly those involving the gastrointestinal tract. By considering the gut microbiome and its influence on the CNS, researchers and clinicians are beginning to appreciate the multifaceted nature of PD, which may pave the way for novel therapeutic strategies targeting not just the brain, but also the gut and its microbial inhabitants. This holistic approach could potentially offer new avenues for early intervention and a more comprehensive management of the disease, ultimately improving patient outcomes.

PD Etiology, Pre-clinical, Prodromal and Clinical stages

To frame the discussion of PD etiology in a broader context, it's important to start by addressing the concept of prodromal stages in the disease's progression. One of characteristic features of PD is the aggregation of α-synuclein, a protein that abnormally accumulates and forms Lewy bodies in the nervous system (Cheng et al., 2023). The pathological process begins years before the onset of motor symptoms, during what is known as the prodromal phase (Fig. 2). During this phase, aggregation of misfolded α-synuclein, which is thought to play a major role in nigral dopaminergic neuronal loss, often starts in the enteric nervous system (ENS), part of the peripheral nervous system located in the gastrointestinal tract (Beach et al., 2016; Fricova et al., 2020; W. Liu et al., 2022; Mahbub et al., 2024; Mu et al., 2015; Ortiz de Ora et al., 2024; Wakabayashi et al., 1990). This early involvement of the ENS is crucial, as it suggests that the pathological process of PD may begin in the gut and then spread to the central nervous system via the vagus nerve (Braak et al., 2003). The presence of α-synuclein in the ENS can lead to various

gastrointestinal symptoms, such as constipation, which are common in the prodromal stage of PD (Warnecke et al., 2022). As the disease progresses, α-synuclein pathology spreads to other parts of the nervous system, including the brainstem and the olfactory bulb, leading to non-motor symptoms such as olfactory dysfunction and sleep disturbances (Dodet et al., 2024; Gu et al., 2024).

One of the most notable non-motor symptoms that often occurs during the prodromal stage is rapid eye movement (REM) sleep behavior disorder. Isolated REM sleep behavior disorder (iRBD) is a parasomnia characterized by the loss of normal muscle atonia during REM sleep, which leads to the physical enactment of dreams. Typically, during REM sleep, the body is in a state of muscle paralysis – a mechanism that prevents individuals from acting out their dreams. However, iRBD is characterized by the loss of normal muscle atonia during REM sleep, leading to the enactment of vivid and often violent dreams. The clinical significance of iRBD extends far beyond its impact on sleep quality. Over the past few decades, a substantial body of evidence has emerged demonstrating that iRBD is not merely an isolated sleep disorder but is predictive of future neurodegenerative diseases, particularly synucleinopathies such as PD (Figorilli et al., 2023; Shrestha et al., 2021). Among these, PD is the most common disorder associated with iRBD, with longitudinal studies showing that a significant proportion of patients with iRBB – ranging from 40% to 80% – develop PD within ten years from the initial iRBD diagnosis (Galbiati et al., 2019; Iranzo et al., 2014; Postuma et al., 2009, 2015).

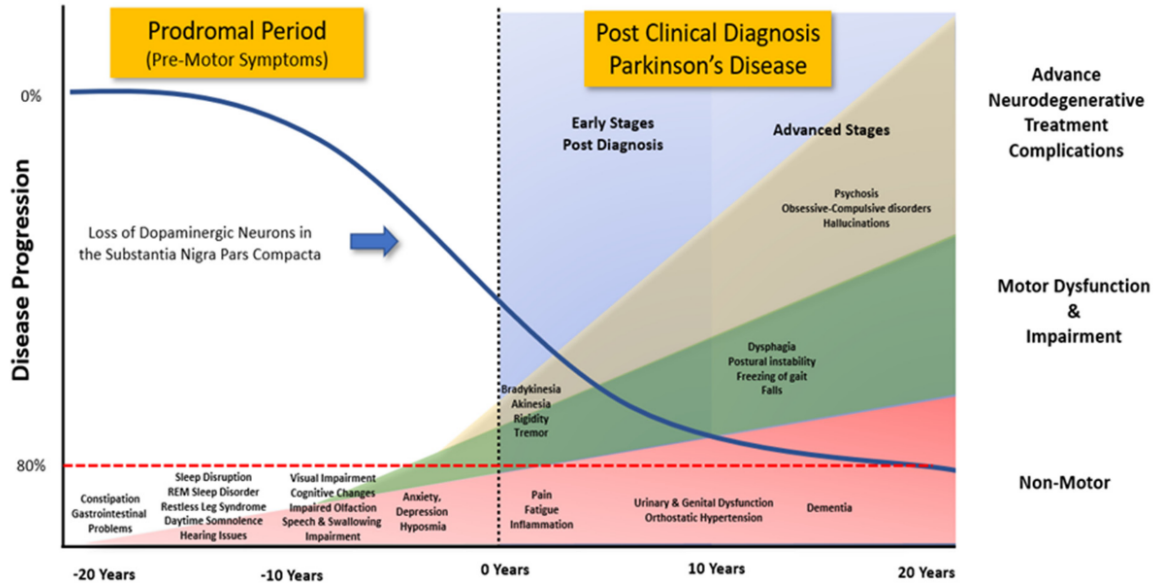


Figure 2. Progression of clinical symptoms from the initial prodromal phase to the confirmed diagnosis of PD. Adapted from Ravenhill et al., 2023.

This prodromal phase, characterized by neurodegeneration without overt motor symptoms, is crucial for identifying potential disease-modifying interventions. Isolated REM sleep behavior disorder is a key prodromal marker for PD and other synucleinopathies, indicating an underlying neurodegenerative process that may lead to the full development of PD. As the disease advances, three major pathological features emerge: the progressive loss of dopaminergic neurons in the substantia nigra pars compacta (SNpc) (Mamelak, 2018), the formation of Lewy bodies composed primarily of aggregated α-synuclein (Tanei et al., 2021; Wakabayashi et al., 1990), and pervasive neuroinflammation (Muzio et al., 2021; Sampson et al., 2016). The loss of dopaminergic neurons results in a significant reduction in dopamine, a neurotransmitter essential for motor function regulation, leading to the classical motor symptoms of PD, such as bradykinesia, tremor, and rigidity. Lewy bodies, abnormal intracellular inclusions primarily consisting of aggregated α-synuclein, are a pathological hallmark of PD. The exact role of α-synuclein in PD, its genetic links, and its contribution to neurodegeneration remain active areas of research, with ongoing studies aiming to clarify how these aggregates contribute to the disease's progression (Braak et al., 2003; Spillantini & Goedert, 2018).

Dopamine and Glutamate Metabolism Dysregulation

Impaired dopamine metabolism is a critical aspect of PD pathogenesis, contributing significantly to the neurodegenerative processes observed in this disorder. (Masato et al., 2019; Scheffer et al., 2021) The dysregulation of dopamine homeostasis in PD leads to the

accumulation of toxic metabolites, particularly 3,4-dihydroxyphenylacetaldehyde (DOPAL) (Mattammal et al., 1995). DOPAL is a highly reactive aldehyde that can modify functional protein residues, leading to oxidative stress and neuronal cell death (Bisaglia et al., 2014). Furthermore, DOPAL interacts with α-synuclein, and this interaction promotes the oligomerization and aggregation of α-synuclein, further impairing neuronal function and survival (Masato et al., 2019; Mor et al., 2017, 2019). This metabolite is especially detrimental to dopaminergic neurons in the substantia nigra (Masato et al., 2019). The interplay between dopamine metabolism abnormalities and mitochondrial defects are also relevant to PD (Xu & Yang, 2022). The catabolism of dopamine involves mitochondrial processes that produce reactive oxygen species (ROS). Under physiological conditions, the rate of dopamine oxidation is slow, and the cellular antioxidant machinery can manage the formation of reactive products. However, in PD, the increased oxidative stress overwhelms these protective mechanisms, leading to mitochondrial damage and further impairing cellular energy metabolism (Xu & Yang, 2022). This interplay between dopamine metabolism and mitochondrial dysfunction creates a vicious cycle that exacerbates neuronal degeneration in PD.

While PD has traditionally been conceptualized as a disorder primarily driven by dopaminergic deficits, it is now increasingly recognized that disturbances in glutamatergic neurotransmission also play a critical role in the disease's progression. Research has shown that alterations in glutamate signaling are not merely secondary consequences but are integral to the pathological cascade that exacerbates neurodegeneration in PD (Blandini et al., 1996) (Fig. 3). In PD, the regulation of glutamate release and receptor activity is significantly impacted by the loss of dopamine. Dopamine typically inhibits the release of glutamate in certain brain regions. However, in PD, the degeneration of dopaminergic neurons in the substantia nigra and the resulting depletion of dopamine cause an overactivity of glutamate in the basal ganglia, which is neurotoxic (Blandini, 2010; J. Wang et al., 2020).

Glutamate, a pivotal neurotransmitter within the central nervous system, plays a significant role in the onset and progression of PD. As the primary excitatory neurotransmitter in the mammalian brain, glutamate is integral to many essential functions and metabolic processes. The precise regulation of excitatory and inhibitory neuronal activity is vital for maintaining normal brain function, and disruptions in this balance can have profound pathological consequences. Excessive activation of glutamate receptors, particularly in nigrostriatal neurons, may induce neuronal death through a mechanism known as glutamate-induced excitotoxicity (Bergman et al., 1994). This pathological condition arises

when there is an excessive release of glutamate from presynaptic terminals or a failure in the reuptake mechanisms, resulting in an abnormal accumulation of extracellular glutamate (Lin et al., 2012).

Elevated extracellular glutamate levels lead to the hyperactivation of NMDA receptors, which are permeable to calcium ions, culminating in calcium overload within the neurons. This calcium overload triggers excitotoxic damage, a process characterized by the activation of deleterious enzymatic pathways, the generation of toxic free radicals, and the disruption of cellular energy production. Such excitotoxicity is further aggravated by the activity of microglia and astrocytes, which, under pathological conditions, release additional glutamate, thereby exacerbating neuronal injury (Wetherington et al., 2008). The sustained elevation of intracellular calcium levels, driven by this glutamate surge, activates enzymes that further damage cellular structures, leading to oxidative stress and, ultimately, cell death. Beyond the receptor-mediated excitotoxicity, elevated glutamate levels can also provoke oxidative stress through non-receptor-mediated mechanisms, contributing to what is termed oxidative glutamate toxicity (Shirlee Tan et al., 2001).

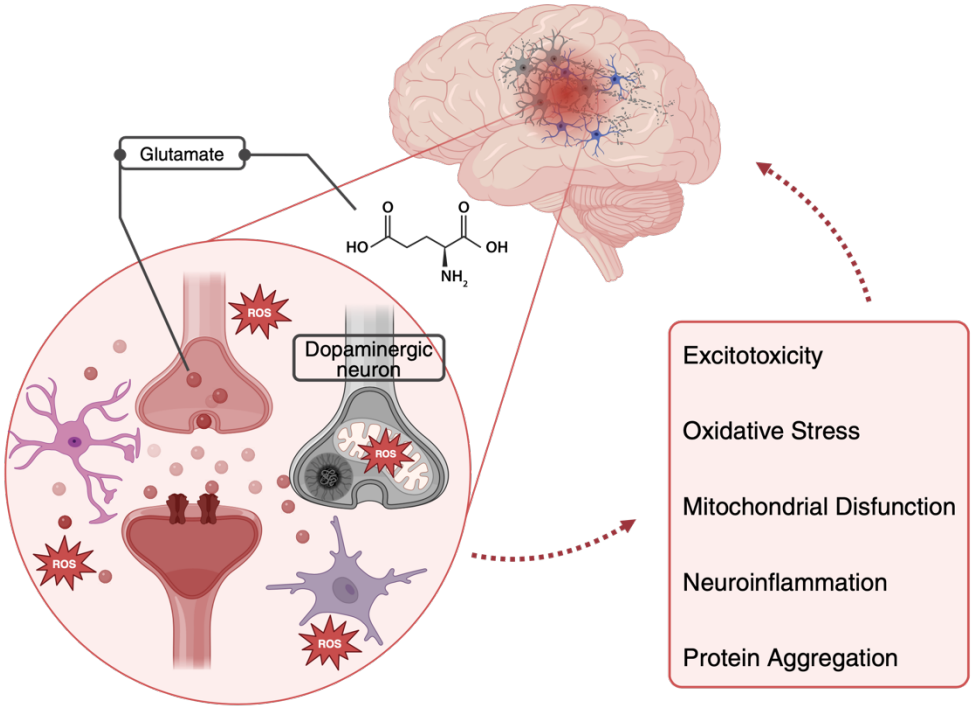


Figure 3. Mechanisms of glutamate toxicity in PD. Dysregulation of synaptic glutamate levels can result in excitotoxicity, where excessive glutamate accumulation leads to neurotoxicity and ultimately cell atrophy or death.

Previous studies have implicated glutamate excitotoxicity as a contributing factor in the degeneration of dopaminergic neurons, a hallmark of PD (Meredith et al., 2009). Clinical investigations have revealed subtle alterations in glutamate levels within the brains of PD patients, indicative of enhanced glutamate neurotransmission (Gröger et al., 2014; O'Gorman Tuura et al., 2018; Weingarten et al., 2015). While the exact causal relationship between these alterations in glutamate-related pathways and neurodegeneration in PD remains to be fully elucidated, it is clear that both inflammation and glutamate-induced excitotoxicity are central to the pathophysiology of PD. The resultant dopamine deficit linked to increased neurotoxicity of glutamate further exacerbates this pathological cycle, establishing a self-perpetuating loop that accelerates the neurodegenerative process, (dos-Santos-Pereira et al., 2018; J. Wang et al., 2020).

In the gastrointestinal tract, glutamate originates from several key sources: it is ingested through the diet, produced by the gut microbiota, and synthesized endogenously by gut epithelial cells (Reeds et al., 2000). This versatile neurotransmitter plays a pivotal role in various gastrointestinal functions, including the regulation of gut motility and secretion, as well as the maintenance of the intestinal barrier's integrity. These functions are mediated through interactions with specific glutamate receptors, such as NMDA and AMPA receptors, which are distributed along the gut lining (Hamnett et al., 2024). The influence of the gut microbiota on glutamate levels and metabolism is profound, with certain bacterial species actively participating in the synthesis and conversion of glutamate into other bioactive compounds, such as gamma-aminobutyric acid (GABA) and short-chain fatty acids (SCFAs) (Strandwitz, 2018). This microbial activity is essential not only for maintaining gut health but also for the proper functioning of the gut-brain axis – a sophisticated, bidirectional communication network that integrates gut physiology with central nervous system processes. Glutamate signaling within the gut has significant implications for brain function, primarily via peripheral neural pathways and the vagus nerve, which conveys signals directly between the gut and the brain. Additionally, glutamate or its metabolites may cross the blood-brain barrier through humoral pathways, subsequently modulating neurotransmission within the central nervous system (Baj et al., 2019). The delicate balance of glutamate within the gut, and its broader impact on brain health, highlights the importance of the gut-brain axis in the pathogenesis of PD. Disruptions in this balance, whether due to microbial dysbiosis or other factors, may contribute to the progression of PD, emphasizing the critical role of gut-derived glutamate in the disease's underlying mechanisms.

Glutamate, a key amino acid in neurotransmission, exists as two enantiomers: L-glutamate and D-glutamate. L-glutamate is a versatile amino acid that provides the umami taste in

foods and serves as a key player in metabolism and neurotransmission (Yamaguchi, 1991). It is crucial for amino acid and carbohydrate metabolism, impacting various cells like intestinal, liver, muscle, and immune cells (Kondoh et al., 2009). As the main excitatory neurotransmitter in the brain, it is vital for learning and memory (Meldrum, 2000). L-glutamate receptors are found in gut epithelial cells, which, when stimulated, activate vagal afferent nerve fibers and influence brain regions such as the cerebral cortex, basal ganglia, limbic system, and hypothalamus (Kondoh & Torii, 2008; M. T. Liu et al., 1997; Tsurugizawa et al., 2008). Therefore, L-glutamate has importance in both dietary and physiological functions. D-glutamate is less common and primarily found in the peptidoglycan cell wall of bacteria and certain fermented foods (Cava et al., 2011; Marcone et al., 2019). Unlike L-glutamate, D-glutamate is not commonly involved in human metabolic processes or neurotransmission, and its role in neurocognitive function remains unclear (Marcone et al., 2019). Another molecule related to L- and D-glutamate is β-glutamate. Although β-glutamate is structurally distinct from L- and D-glutamate due to the position of its amino group, it shares the fundamental glutamate backbone with the other two forms. Given the established implications of glutamate excitotoxicity in the pathogenesis of PD, the study of β-glutamate emerges as a particularly promising avenue of research. β-glutamate, which has been identified for its role as an osmolyte in archaea and its involvement in specific enzymatic reactions, presents several intriguing possibilities for enhancing our understanding of gut-brain interactions and neurodegenerative diseases. Its function as a substrate for glutamine synthetase, although with reduced efficiency compared to other substrates, underscores its biochemical versatility and potential implications for metabolic processes (Robinson et al., 2001a). Additionally, its established role in maintaining cellular integrity under conditions of osmotic stress highlights its importance in microbial metabolism and gut function (Robertson et al., 1990a).

Microbiome Links to PD: From Theory to Evidence

In the context of PD, the concept of the gut-brain axis offers a crucial physiological framework for understanding how gut microbiota might influence neurological health. This axis facilitates communication between the gut and brain through various pathways, including neural, endocrine, and immune routes. Gut microbes can modulate brain function and behavior by affecting immune responses, neurotransmitter production, and metabolism, all of which have implications for neurodegenerative conditions such as PD.

The involvement of the gut microbiome in PD is further supported by the Braak hypothesis, which posits that PD pathology might originate in the gut before spreading to the brain via

the vagus nerve (Braak et al., 2003; Morais et al., 2020; Schmit et al., 2023). One proposed mechanism involves the bacterial amyloid protein curli, which can enhance the misfolding and aggregation of α-synuclein in the gut. Curli, produced by certain gut bacteria, can cross-seed with α-synuclein, promoting its aggregation into pathogenic forms (Chen et al., 2016; C. Wang et al., n.d.). This misfolded α-synuclein can then propagate in a prion-like manner, traveling from the gut to the brain via the vagus nerve, contributing to the neurodegenerative processes observed in PD. This hypothesis is bolstered by clinical observations that gastrointestinal symptoms, such as constipation, often precede the onset of motor symptoms in PD by several years. These early non-motor symptoms suggest that alterations in the gut microbiome could serve as potential early biomarkers for PD, thus shifting the focus of research from solely the brain to include peripheral systems like the gastrointestinal tract.

A growing number of clinical studies have provided compelling evidence linking gut microbiome alterations to PD (Heintz-Buschart & Wilmes, 2018; Z. Li et al., 2023; X. Zhang et al., 2023). These studies consistently reveal patterns of dysbiosis in PD patients, with specific microbial taxa either increased or decreased in abundance compared to healthy controls. One of the most frequently reported findings is the increase in the family *Verrucomicrobiaceae*, particularly the genus *Akkermansia*, in PD patients. Several studies have documented elevated levels of *Akkermansia muciniphila* in individuals with PD, suggesting that this bacterium may play a significant role in the disease's pathology (Barichella et al., 2019; Bedarf et al., 2017; Heintz-Buschart et al., 2018). The increase in *Akkermansia* might reflect a compensatory response to the maintenance of gut barrier integrity, which is often compromised in PD, potentially leading to an enhanced permeability of the gut lining and subsequent systemic inflammation. Moreover, other microbial families such as *Lactobacillaceae* and *Enterobacteriaceae* have also been reported to be increased in PD. These taxa are known for their roles in immune modulation and have been implicated in inflammatory processes that may exacerbate neurodegeneration in PD (Barichella et al., 2019; Scheperjans et al., 2015; Tan et al., 2021).

Conversely, a consistent decrease in the family *Prevotellaceae*, especially the genus Prevotella, has been observed in multiple studies (Aho et al., 2019; Bedarf et al., 2017). Prevotella species are involved in the production of SCFAs, which have neuroprotective effects. The reduction in *Prevotellaceae* suggests a disruption in SCFA production, potentially contributing to the neurodegenerative processes observed in PD. Additionally, genera such as *Roseburia* and *Faecalibacterium*, which are crucial producers of the SCFA butyrate – a compound with well-documented anti-inflammatory properties – have also been

found to be decreased in PD patients (Aho et al., 2019; Cirstea et al., 2020). This decrease may be associated with the heightened inflammatory state that characterizes PD, further underscoring the potential link between gut dysbiosis and the progression of the disease.

Studies reveal that while there may not be significant differences in alpha-diversity – an indicator of the variety of microbial species within an individual – between PD patients and healthy controls, there is often a trend towards reduced alpha-diversity in PD, suggesting a potential loss of microbial richness and complexity (Y. Qian et al., 2018, 2020; K. Zhang et al., 2022). This reduction in diversity could imply a diminished resilience of the gut microbiome, making it more susceptible to imbalances and potentially contributing to disease progression. In terms of beta diversity, which measures the differences in microbial communities between individuals, several studies have reported significant differences between PD patients and healthy controls. These findings indicate that PD is associated with a distinct microbial composition, further supporting the notion that gut dysbiosis plays a role in the disease (W. Li et al., 2017; Scheperjans et al., 2015; Wallen et al., 2022).

The consistent alterations in specific microbial taxa, such as the increase in pro-inflammatory bacteria like *Akkermansia* and *Enterobacteriaceae*, and the decrease in beneficial, SCFA-producing bacteria such as *Prevotella*, *Roseburia*, and *Faecalibacterium*, underscore the potential role of the gut microbiome in fostering a pro-inflammatory environment that may contribute to the pathogenesis of PD. These microbial changes suggest that the gut microbiome could be a crucial factor in the development and progression of PD, potentially offering new avenues for early diagnosis and therapeutic interventions.

In addition to bacteria, the gut microbiome encompasses archaea, which have been reported to exhibit higher abundances in patients with PD. Notably, methanogenic archaea, such as *Methanobrevibacter smithii*, have been identified in increased numbers within the gut microbiota of individuals with PD (Cem Duru et al., 2024; Romano et al., 2021; Wallen et al., 2022b; F. Zhang et al., 2020). These methanogens are integral to methane production, a process that can influence gut motility and is associated with gastrointestinal symptoms frequently observed in PD, such as constipation (Sharma et al., 2020). Among the various methanogenic phylotypes, *M. smithii* is particularly significant, contributing to over 90% of methane production. The abundance and proportion of *M. smithii* in stool samples correlate strongly with the amount of breath methane in patients with irritable bowel syndrome (IBS) (Ghoshal et al., 2016; G. Kim et al., 2012).

The emerging evidence linking the gut microbiome to PD highlights the significant role that the gut-brain axis plays in neurodegeneration. The consistent patterns of microbial alterations observed in PD patients across multiple studies point to gut dysbiosis as a contributing factor to the disease's progression, possibly through mechanisms involving neuroinflammation, immune modulation, the disruption of SCFA production, and the role of bacterial amyloid protein curli in promoting the misfolding and aggregation of α-synuclein. As research continues to explore these connections, it becomes increasingly clear that understanding the gut microbiome's role in PD could lead to the development of microbiome-based biomarkers for early detection and novel therapeutic strategies aimed at modulating gut health to slow or prevent disease progression.

Meta-omics for Microbiome Research

The study of the microbiome has evolved significantly over the past few decades, driven by advancements in sequencing technologies and computational methods. Initially, microbiome research relied heavily on culture-based techniques, which were limited by the inability to grow many microbial species in the laboratory. The advent of metagenomics in the early 2000s marked a pivotal shift, allowing researchers to analyze microbial communities directly from environmental samples without the need for culturing (X. Zhang et al., 2019). This breakthrough enabled the identification and characterization of a vast array of previously unrecognized microbial species.

Meta-omics, encompassing metagenomics, metatranscriptomics, metaproteomics, and metabolomics, represents a comprehensive approach to studying the microbiome. This integrative methodology allows for a holistic understanding of microbial communities and their functional roles within various ecosystems, including the human gut. Leveraging multiple omics technologies allows to elucidate the complex interactions between microbes and their hosts, providing insights into health and disease states.

Recent advancements in meta-omics have significantly enhanced our ability to study microbial communities in a detailed and integrative manner. Meta-omics combines various omics approaches to provide a comprehensive view of the microbiome, encompassing genetic, transcriptomic, proteomic, and metabolomic data (T. Ma et al., 2019; L. Wang et al., 2022).

This integrative approach has been pivotal in uncovering the complex interactions between microbes and their hosts, particularly in the context of human health and disease. Meta-

omics has in particular been instrumental in identifying biomarkers for disease diagnosis and monitoring, as well as understanding the functional roles of microbial communities in different environments (T. Ma et al., 2019). By integrating meta-omics with other advanced technologies, researchers can achieve a more nuanced understanding of microbial ecosystems, ultimately leading to better diagnostic tools and therapeutic strategies.

Recent advances in MG have significantly enhanced our understanding of microbial communities (Lema et al., 2023; L. Zhang et al., 2021). These technologies enable rapid and detailed analysis of microbial diversity, community structure, genetic relationships, and functional potential. Metagenomics has been instrumental in identifying novel and uncultivable microorganisms, providing insights into microbial interactions and their roles in various environments. For instance, functional metagenomics allows for the screening of new bioactive substances and functional genes from microbial communities (L. Zhang et al., 2021). Despite these advancements, challenges remain, such as the inability to determine gene expression levels and the dynamic responses of microbial communities to environmental changes (Lema et al., 2023). By integrating MG with other omics approaches, such as MT, researchers can gain a more holistic understanding of microbial ecosystems, bridging the gap between genetic potential and actual functional activity.

To address these challenges, MT analyzes RNA transcripts in a sample, providing a dynamic view of gene expression within the microbiome. This approach identifies which microbial genes are actively transcribed under specific conditions, offering insights into the functional state of the microbial community. RNA sequencing (RNAseq) is employed to capture expressed transcripts within a microbiome at a particular time, allowing for a detailed view of active microbial members. Applications of metatranscriptomics include characterizing active microbes in a community, uncovering novel microbial interactions, detecting regulatory antisense RNA, and monitoring gene expression to elucidate the relationship between viruses and their hosts (Bao et al., 2015; Bashiardes et al., 2016; Bikel et al., 2015; Moniruzzaman et al., 2017).

Metaproteomics (MP) extends the analysis of microbial communities by examining their protein expression, thereby providing detailed functional information about the dynamic interactions between hosts and their microbiota. Proteins, as the functional molecules executing the biological activities encoded by genes, offer insights into the actual functional output of the microbiome. Consequently, metaproteomics has emerged as a crucial complementary approach to metagenomics, focusing on the large-scale characterization of proteins from environmental microbiota, such as those found in the human gut (Van Den

Bossche et al., 2021). This approach underscores the importance of integrating multiple omics techniques to capture the full spectrum of microbial activity, thereby providing a more comprehensive understanding of microbial functions and their impact on host health.

Metabolomics (MB), the study of small molecules and metabolites produced by microbial communities, provides insights into the biochemical interactions between microbes and their hosts. It is a key technology for detecting and identifying small molecules produced by the human microbiota and understanding their functional roles. By examining the complete set of metabolites produced by microbial communities, metabolomics reflects their metabolic activity and interaction with the host. This approach helps identify microbiome-derived metabolites that may serve as biomarkers or modulators of disease(David et al., 2014; Hanash et al., 2011; Lanpher et al., 2006). In PD studies, metabolomics can reveal alterations in key metabolites, such as short-chain fatty acids and bile acids, elucidating their roles in disease mechanisms and overall host health (Morrison & Preston, 2016; Zacharias et al., 2022).

The integration of these meta-omics approaches, as demonstrated in studies of the human gut microbiome, allows for a more comprehensive understanding of the microbial ecosystem. For example, in a study of familial type 1 diabetes, researchers employed metagenomic, metatranscriptomic, and metaproteomic analyses to investigate the gut microbiota’s taxonomic and functional attributes (Heintz-Buschart et al., 2016). The study demonstrated that gastrointestinal microbiome community structures are consistent across all omic levels, with each level showing individuality and family specificity. This work highlights the need for integrated multi-omic analyses to understand host-microbe interactions in health and disease. Moreover, meta-omics has been instrumental in advancing our understanding of drug-microbiome interactions. Recent studies have shown that the gut microbiome can influence drug metabolism, efficacy, and toxicity (Kolli et al., 2023; Wuyts et al., 2023; Q. Zhao et al., 2023). By integrating meta-omics data, researchers can identify microbial genes and pathways involved in drug metabolism, paving the way for personalized medicine approaches that consider an individual's microbiome structure and function.

The study of the human microbiome through meta-omics has revolutionized our understanding of the complex microbial ecosystems residing within us. Meta-omics encompasses a suite of high-throughput techniques that collectively analyze the genetic material, transcripts, proteins, and metabolites of microbial communities. This comprehensive approach allows researchers to capture the full breadth of microbial

diversity and function, providing insights into how these communities interact with their host and influence health and disease. By integrating data from metagenomics, metatranscriptomics, metaproteomics, and metabolomics, scientists can construct detailed models of microbial ecosystems, elucidating the roles of specific microbes and their metabolic pathways in various physiological processes. This holistic perspective is crucial for understanding the dynamic nature of the microbiome and its impact on human health.

In conclusion, meta-omics provides a powerful framework for microbiome research, enabling the simultaneous analysis of microbial composition, gene expression, protein function, and metabolic activity. By integrating these diverse datasets, researchers can gain a deeper understanding of the microbiome's role in health and disease, paving the way for novel therapeutic strategies and personalized medicine.

Biological Networks

Biological networks provide a powerful and systematic framework for understanding the intricate interactions and functional relationships within microbial communities and various biological systems. By integrating diverse omics data with interactome data, including protein-protein interactions and gene-gene associations, network biology uncovers complex patterns that are often hidden in traditional, linear analyses. The application of network approaches has become indispensable across nearly all domains of science, particularly in the life sciences, where capturing the inherent complexity of biological systems is crucial (Bray, 2003; Koonin et al., 2002; Wall et al., 2004). These networks function at multiple spatial scales, from molecular interactions within cells to ecosystem-level interactions (Fang et al., 2020; Luck et al., 2020). In particular, on a broader ecological scale, networks that represent interactions between species within ecosystems have provided essential insights into keystone species and their roles in conservation efforts (Bascompte, 2010; Roume et al., 2015). Within cells, biological processes are intricately regulated at multiple levels, involving complex networks of transcriptional, post-transcriptional, and post-translational events. The network approach allows researchers to move beyond gene-centric analyses and adopt a more holistic, systems-oriented examination of biological data. This shift emphasizes the interactions and relationships between biological entities – such as genes, proteins, and metabolites – offering a more comprehensive understanding of how these elements coordinate to drive overall cellular functions (Barabási & Oltvai, 2004).

Traditional genome-wide omics studies generate lists of genes or their products that exhibit significant alterations under specific conditions. However, focusing solely on individual

genes or proteins in complex systems, such as environmental samples, can lead to the oversight of broader, unexpected relationships between genes or even entire pathways. In contrast, network biology recognizes that biological processes are not predominantly governed by isolated proteins or linear pathways but are controlled by intricate networks of molecular interactions at the system level (Gardy et al., 2009). Furthermore, network medicine theory expands on this concept by suggesting that disease-associated traits arise not from isolated single gene genetic variants, but from disruptions within the broader network context of genes (Charitou et al., 2022; Kennedy et al., 2020). Therefore, not all disease-related phenotypes are caused by genetic variants in single genes functioning independently, but instead stem from disturbances in the complex network of gene interactions . Understanding the connections, nodes, and patterns within these networks is crucial for understanding the broader context of gene function and the manifestation of complex phenotypes in both health and disease.

A network-based approach is a highly effective complement to traditional functional annotation methods, such as enrichment analyses. Techniques like gene set enrichment analysis (GSEA), which ranks gene sets based on factors like differential expression, rely heavily on well-annotated gene sets. While these methods are useful, they can miss important connections between functionally related genes, especially when gene sets share few overlapping genes but are involved in similar pathways or represent different subcomponents of the same biological process. Moreover, the sensitivity of these analyses is often constrained by the size of the gene set. In contrast, network-based methods model interactions between genes, proteins, and molecules rather than treating them as isolated entities. They capture relationships that enrichment analyses might overlook by linking well-annotated and poorly annotated elements, thus integrating diverse biological mechanisms. Examples include the use of protein-protein interaction data to predict protein functions based on the annotations of interacting partners (Deng et al., 2004), the application of pathway enrichment analysis with networks (L. Liu et al., 2017), and the prediction of protein functions via network-derived clusters (Song & Singh, 2009). Additionally, proximity-based methods in gene networks can reveal associations between gene sets and biological functions (Glaab et al., 2012). By incorporating expression data with cellular network information, like protein-protein and protein-DNA interactions, we can reveal the regulatory mechanisms behind these changes (Cline et al., 2007). Network analyses provide a broader and less biased view of genes and proteins, avoiding the overemphasis on well-characterized pathways. This approach allows for a more comprehensive understanding of biological systems by highlighting novel and significant components that might be

overlooked in traditional studies (Charitou et al., 2016). Therefore, networks are particularly valuable for predicting the functions of uncharacterized genes or proteins based on their interactions with known entities, aiding in the annotation of newly discovered or poorly understood components .

Types of Biological Networks

There are several distinct types of biological networks, each representing different aspects of cellular processes: protein-protein interaction networks, genetic interaction networks, regulatory networks, signaling networks, and metabolic networks. Despite their focus on different cellular functions, these networks share common organizational and functional principles. Protein-protein interaction networks illustrate the physical interactions between proteins, highlighting the collaborative nature of proteins in executing cellular functions. These networks are essentially representations of how proteins interact with one another within the cellular environment, encompassing all the proteins and their interactions (Vazquez, 2010). Through the study of protein-protein interaction networks, researchers can explore specific protein interactions, protein complexes, and signaling cascades. Genetic interaction networks describe the relationships between genes, particularly how the modification of one gene can influence the expression or function of another gene (Boucher & Jenna, 2013). Genetic interaction networks indicate that two genes have a functional relationship, which may manifest through their involvement in the same biological processes or pathways, or through compensatory mechanisms with functions that may not be immediately apparent. Gene regulatory networks focus on the interactions between transcription factors and their target genes, shedding light on the mechanisms that regulate gene expression. These networks integrate multiple elements of gene regulation, including transcription factors, splicing factors, long non-coding RNAs, microRNAs, and metabolites (Badia-i-Mompel et al., 2023). Cell signaling networks map out the pathways through which cells communicate and respond to external stimuli. These networks involve a series of molecular interactions that relay signals from the cell membrane to the nucleus, orchestrating cellular responses to environmental changes (Azeloglu & Iyengar, 2015). Metabolic networks are complex systems that outline the biochemical reactions occurring within a cell (Nikoloski et al., 2008). They illustrate the interconnectedness and regulation of various metabolic pathways, providing insights into cellular metabolism and its implications for health and disease. These networks encompass all reactions an organism can perform, the metabolites involved as substrates and products, and the genes encoding the enzymes that facilitate these reactions (Chalancon et al., 2013).

In the study of biological systems, large-scale networks such as networks inferred from gene expression data offer significant advantages over analyzing individual biological entities in isolation. Gene co-expression networks, for example, represent gene-gene interactions as undirected graphs, where nodes denote genes and edges reflect the strength of their co-expression. Within this framework, weighted gene co-expression network analysis (WGCNA) emerged as a key method in systems biology. WGCNA is instrumental in constructing these networks, identifying gene modules, and determining central hub genes (Chang et al., 2010; Wan et al., 2018; B. Zhang & Horvath, 2005). WGCNA enhances the understanding of gene interactions, highlights key regulatory genes, and aids in predicting the functions of previously unknown genes. This method provides a deeper insight into gene expression networks, complementing the broader biological networks discussed above. The distinct advantage of WGCNA lies in its ability to convert gene expression data or other omics data such as proteomic (J. X. Wu et al., 2023), metabolomic (Pei et al., 2017), 16S rRNA amplicon sequencing data (Jameson et al., 2023) into co-expression modules, thereby facilitating the exploration of signaling networks potentially associated with phenotypic traits of interest (Langfelder & Horvath, 2008).

Networks Topology and Metrics

Numerous mathematical and computational approaches have been developed to analyze large networks to identify features of interest. In the realm of biological networks, these analytical approaches are essential for uncovering the underlying architecture and identifying key elements that govern complex biological processes. The intricate web of interactions within these networks can be difficult to decipher, but by applying network analysis techniques, it is possible to pinpoint critical nodes and pathways that play significant roles in the functionality and stability of these biological systems.

One of the most powerful and widely used methods in this context is centrality analysis, which provides valuable insights into the importance of individual nodes within a given network (Bloch et al., 2023). Given that our analyses involve microbiome data, it's crucial to consider the topological features of ecological networks. These networks quantify the structure of ecological interactions, consisting of nodes (representing entities like genes or species) and edges (representing interactions or relationships between these entities). In our specific study, "The gut microbiome gene expression network is dysregulated in individuals with Parkinson's disease" we use KEGG KOs as nodes and co-expression ratios between genes as edges. To characterize the system at the level of individual nodes or edges, we employ various network centrality metrics. These metrics help us understand the

importance and influence of each node or edge within the network, drawing on established network centrality metrics (Bonacich, 1987; Costa et al., 2019; Lau et al., 2017; Wasserman & Faust, 1994).

Centrality analysis stands out as a crucial tool for investigating complex biological networks. It helps to identify key elements in biological processes (Koschützki & Schreiber, 2008). The most common centrality measures used in biological network analysis are degree centrality, closeness centrality, and betweenness centrality. Studies have shown that the mean centrality values for these measures are significantly higher for essential proteins compared to nonessential proteins (Hahn & Kern, 2005). In addition to these centrality measures, other metrics such as clustering coefficient, connectivity, and eigenvector centrality provide further insights into network topology and function. These metrics allow for a more nuanced analysis of how nodes interact, the formation of modules, and the overall flow of information within the network, thereby enriching our understanding of biological systems .

Degree centrality evaluates the importance of a node based on its number of connections (Ashtiani et al., 2018; Jeong et al., 2001a). It is one of the most widely used metrics and has been linked to various dynamic processes in diseases (Checco et al., n.d.; Opsahl et al., 2010). Although it provides insight into a node's connectivity, it can miss important aspects of network architecture, such as nodes that bridge different parts of the network but have a low degree of connections (Bloch et al., 2023).

Closeness centrality measures the average length of the shortest paths from a node to all other nodes, indicating how centrally located a node is within a network (Evans & Chen, 2022). It is used to assess how effectively information flows from one node to others and to identify ideal starting points for information propagation. This measure has been utilized to pinpoint important metabolites in genome-based, large-scale metabolic networks (Ashtiani et al., 2018). It has been found that the centrality measure is the most effective measures in terms of locating the network’s essential genes (Plaimas et al., 2010).

Betweenness centrality quantifies the proportion of shortest paths passing through a node, reflecting its role in communication flow within the network (Barthélemy, 2004). In protein interaction networks, high betweenness indicates a protein's potential to facilitate communication among various proteins (Bima et al., 2022). Nodes that bridge gaps between clusters tend to have high betweenness centrality, highlighting their importance in network structure and function (Ravasz et al., 2002). Hub genes are nodes with a high degree of connectivity within a network, often playing critical roles in maintaining the

structure and function of biological systems. These genes are typically involved in key regulatory processes and are essential for the stability of the network.

Clustering coefficient is a measure that quantifies the extent to which nodes in a network tend to cluster together. In a biological context, it indicates the likelihood that a gene's neighbors are also connected to each other, forming tightly knit groups or modules. A high clustering coefficient suggests that the network has a modular structure, whereby genes within the same module are more likely to be involved in related biological processes. This is particularly relevant in gene co-expression networks, where genes within a highly clustered module might share similar functions or be co-regulated (Watts & Strogatz, 1998). The clustering coefficient has been shown to be associated with network robustness and the ability to withstand perturbations, which is crucial in maintaining biological function under varying conditions (Pavlopoulos et al., 2011).

Connectivity in network analysis refers to the degree of interaction or linkage among nodes within the network. In gene co-expression networks, connectivity is often used to identify hub genes – genes with a high number of connections to other genes. These hub genes are generally considered to be critical regulators of biological processes, as they can influence the expression and activity of many other genes (B. Zhang & Horvath, 2005). Although these genes may not be vital in higher organisms, knock-out experiments in yeast have demonstrated that hub genes are crucial for survival (Carlson et al., 2006; Jeong et al., 2001). A high connectivity in a gene network often correlates with essentiality, meaning that the genes with the most connections are more likely to be crucial for the survival or proper functioning of the organism (Jeong et al., 2001). Moreover, changes in the connectivity of certain nodes can signal shifts in network dynamics, which may be associated with disease states or other biological changes (Nikoloski et al., 2010). Theoretical and empirical analyses have shown that intramodular connectivity, focusing on connections within specific gene modules in co-expression networks, is a valuable metric for identifying key genes (Fuller et al., 2011). This concept has been validated in various studies, including those on brain cancer and inflammatory responses (Gargalovic et al., 2006; Horvath et al., 2006), underscoring the importance of hub genes in understanding complex biological processes.

Eigenvector centrality is a measure that extends the concept of degree centrality by not only considering the number of connections a node has but also the importance of the nodes it is connected to. This measure assigns relative scores to all nodes in the network based on the principle that connections to highly connected nodes contribute more to the score of the

node being evaluated (Bonacich, 1987). In biological networks, eigenvector centrality can identify influential genes that, while perhaps not the most connected, are linked to other highly influential genes, thus playing a pivotal role in the overall network structure. This metric is particularly useful in identifying nodes that are central to the flow of information or regulation across the network (Golbeck, 2013).

Numerous mathematical and computational approaches have been developed to analyze large biological networks, uncovering their underlying architecture and identifying key elements that govern complex biological processes. Centrality analysis, including degree, closeness, and betweenness centrality, is a powerful method for pinpointing critical nodes and pathways within these networks. These measures help identify essential proteins and their roles in maintaining network stability. Additionally, metrics such as clustering coefficient, connectivity, and eigenvector centrality provide further insights into network topology, revealing modular structures, hub genes, and influential nodes. Together, these analytical techniques offer a comprehensive understanding of biological systems, aiding in the exploration of network dynamics and their implications for diseases like PD.

Results and Discussion of Publications

Paper 1: Integrated Multi-omics Highlights Alterations of Gut Microbiome Functions in Prodromal and Idiopathic Parkinson's Disease

Contribution

- As the shared first author of the paper, I participated in data preparation, bioinformatical and statistical analysis, data interpretation, visualization of results, and manuscript writing. In addition to preparing and curating all MG and MT data for the results, my specific contribution focuses on the metabolic network analysis. This includes the result section "Metabolites and metabolites-related genes associated with PD are central to microbiome metabolism" and the corresponding discussion.
- Responsible for the creation of visual elements of the manuscript, including figures 2G, 3D,E,F, 4, extended figures 2, 4:
    - Fig. 2G; 3D,E,F; 4; Ext. Fig. 2, 4 – handled the preparation and curation of MG and MT data for figure generation. Prepared tables with specific data for each plot to visualize the results. Participated in discussions and adjustments of results related to the aforementioned figures.
    - Fig. 5 – prepared and curated MG and MT data for the construction of the whole-community network, and mapped KEGG KOs to ChEBI identifiers for network construction. Visualized the network, highlighting key interactions between KEGG KOs corresponding to selected metabolites. Calculated and visualized network topology metrics for subnetworks associated with these metabolites.

Overview of the study

In our study “Integrated multi-omics highlights alterations of gut microbiome functions in prodromal and idiopathic Parkinson’s Disease” we delve into the intricate relationship between the gut microbiome and PD by employing an integrated multi-omics approach, which includes both meta-metabolomic and metagenomic analyses. The primary objective was to explore the functional ramifications of the distinct gut microbiome compositions observed in individuals with PD, those in the prodromal phase of the disease (specifically, iRBD), and HC. The research is particularly significant as it seeks to bridge the gap in understanding how alterations in the gut microbiome may contribute to the pathogenesis and progression of PD.

In the initial phase of the study, meta-metabolomic analyses were conducted to identify metabolites that exhibited differential abundance in individuals with PD or iRBD when compared to healthy controls. This analysis revealed 11 metabolites with significant differences in abundance, among which β-glutamate emerged as a particularly noteworthy compound. The levels of β-glutamate were markedly elevated in both PD and iRBD individuals, and this elevation was strongly correlated with the transcriptional activity of specific bacterial genera, including *Akkermansia*, *Methanobrevibacter*, and *Clostridium*. These findings suggest a potential link between gut microbial activity and disruptions in glutamate metabolism, which may play a critical role in the neurodegenerative processes underlying PD.

This study further explored the transcriptional activities associated with these metabolic changes. It was observed that transcripts related to glutamate metabolism were significantly linked to a decrease in the expression of genes involved in flagellar assembly and chemotaxis – biological processes essential for microbial motility and immune system interaction. This reduction in gene expression was particularly pronounced in PD patients, indicating that the dysregulation of these pathways may contribute to the altered immune responses and gut-brain communication observed in PD.

Interestingly, while no significant differences in overall metagenomic diversity were detected between PD and iRBD, the study highlighted substantial variations in the transcriptional activity of specific microbial taxa. Genera including *Roseburia*, *Blautia*, and *Eubacterium*, which are generally reduced in abundance in PD, exhibited altered gene expression profiles. These changes in microbial activity were further linked to increased levels of metabolites including β-glutamate, isovalerate, and isobutyrate – compounds previously

associated with disease severity and progression in PD. The findings emphasize the potential impact of these metabolites not only on gut health but also on the central nervous system, suggesting a broader role for microbial metabolism in influencing pathways involved in neurodegenerative diseases.

In addition to examining glutamate metabolism, the study also investigated the role of bile acids in PD. Bile acids are known to have wide-ranging effects on the immune system, metabolism, and CNS function. Our study found that specific bile acids, such as chenodeoxycholic acid (CDCA) and glycocholic acid (GCA), were decreased in individuals with PD and iRBD, respectively. This reduction in bile acid levels was accompanied by decreased expression of bile acid-related transcripts in PD patients, highlighting a potential disruption in bile acid metabolism that could contribute to disease progression. Given the immunomodulatory properties of bile acids, these findings suggest that alterations in bile acid metabolism may play a role in the pathogenesis of PD by influencing immune responses and gut-brain interactions.

The study's results underscore the importance of focusing on the functional aspects of the gut microbiome rather than merely analyzing its taxonomic composition. By integrating multi-omics data, the researchers were able to gain a more comprehensive understanding of the specific microbial activities and metabolic pathways that are disrupted in PD. The findings suggest that the future of microbiome research in the context of neurodegenerative diseases like PD may lie in identifying ways to modulate specific microbial functions, with the goal of improving our understanding of disease mechanisms and developing targeted interventions.

In conclusion, this study highlights the complex interplay between the gut microbiome, metabolic processes, and neurodegeneration in PD. The identification of β-glutamate and other metabolites as central players in this interaction, along with the altered transcriptional activities of key microbial taxa, provides new insights into the potential mechanisms driving PD. These findings pave the way for further research aimed at unraveling the microbiome-driven factors that contribute to the onset and progression of neurodegenerative diseases, with the ultimate goal of identifying novel therapeutic strategies.

Integrated multi-omics highlights alterations of gut microbiome functions in prodromal and idiopathic Parkinson's disease

Remy Villette[1,*], Júlia Ortís Sunyer [1,*], Polina V. Novikova [1,*], Velma T. E. Aho [1,*], Viacheslav A. Petrov [1], Oskar Hickl [1], Susheel Bhanu Busi [1,2], Charlotte De Rudder [1], Benoit Kunath [1], Anna Heintz-Buschart [1,3], Jean-Pierre Trezzi[1], Rashi Halder[1], Christian Jäger[1], Laura A. Lebrun[1], Annegrät Daujeumont[1], Sebastian Schade[5], Annette Janzen[6], Nico Jehmlich[7], Martin von Bergen[7], Cédric C. Laczny[1,], Patrick May[1,], Claudia Trenkwalder[8], Wolfgang Oertel[6], Brit Mollenhauer[4,5], Paul Wilmes [1,9,10 #]

1. Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg
2. UK Centre for Ecology and Hydrology, Wallingford, Oxfordshire, United Kingdom
3. Biosystems Data Analysis, Swammerdam Institute for Life Sciences, University of Amsterdam, The Netherlands
4. Department of Neurology, University Medical Center Göttingen, Göttingen, Germany
5. Paracelsus-Elena-Klinik, Kassel, Germany
6. Department of Neurology, Philipps-University Marburg, Marburg, Germany
7. Helmholtz-Centre for Environmental Research GmbH – UFZ, Department of Molecular Toxicology, Leipzig, Germany
8. Department of Neurosurgery, University Medical Center Göttingen, Göttingen, Germany
9. Department of Life Sciences and Medicine, Faculty of Science, Technology and Medicine, University of Luxembourg, Belvaux, Luxembourg
10. Departments of Genetics and Medicine (Hematology, Blood and Marrow Transplantation), Stanford University, Stanford, USA

Corresponding author

* Contributed equally

Abstract

Individuals with Parkinson's disease (PD) exhibit differences in their gut microbiomes' composition compared to healthy controls (HC). The functional consequences of these differences remain unclear. Here we use an integrated multi-omics approach to resolve the functional activities of the gut microbiome in prodromal PD (idiopathic REM sleep behavior disorder, iRBD) and PD compared to HC. Meta-metabolomic analyses identified 11 metabolites that were differentially abundant in PD or iRBD. Based on the robustness of these discriminant features, they guided our subsequent comparisons. Amongst the identified metabolites, β-glutamate was significantly increased in individuals with PD or iRBD, and correlated with the transcriptional activities of *Akkermansia, Methanobrevibacter* and *Clostridium* genera. We specifically identified differences in transcripts related to glutamate metabolism that in turn are linked to decreased transcript abundances in chemotaxis and flagellar assembly expressed by specific taxa in PD. Our integrated multi-omics data highlights multifactorial alterations of structure and function in PD with disrupted functions implicated in disease pathways.

Introduction

Parkinson's disease (PD), a neurodegenerative disease impacting movement due to dopaminergic neuron loss, is the second most prevalent neurodegenerative disease[1]. Individuals with PD are often characterized by an increase in gut permeability, inflammation and constipation which, together, suggest a link between the gut microbiome and PD etiology[2–4]. This potential link is supported by numerous studies reporting differences in the gut microbiome structure of individuals with PD compared to healthy individuals[5–11]. These findings have been further confirmed by recent meta-analyses[12,13]. Together, the studies highlight a decreased abundance for the genera *Roseburia*, *Blautia, Butyricoccus* and *Faecalibacterium* in PD while *Methanobrevibacter*, *Akkermansia, Lactobacillus*, *Bifidobacterium* and *Hungatella*, are typically enriched[5–13]. Similar changes in idiopathic REM sleep behavior disorder (iRBD), a prodromal stage of PD[14], have been reported[7,10]. Moreover, the taxa decreased in PD are known producers of short-chain fatty acids (SCFAs), which correspondingly have also been found to be decreased in concentration in fecal samples of PD individuals[6,15,16].

In addition to SCFAs, several microbiome-derived metabolites such as bile acids (BAs), glycine and glutamate have been associated with PD, either in plasma, serum[17,18] or stool[19–21]. BAs are produced by the host and metabolized by the gut microbiome into secondary bile acids with different cytotoxic capacities but also immunomodulatory capacities[22,23]. Glutamate is the major excitatory neurotransmitter and exerts toxic activity on neuronal cells[24]. Its levels in serum and cerebrospinal fluid have been reported as either increased[25,26], not different[27], or decreased[28], but decreased in the gut in PD compared to healthy controls (HC)[29].

Altogether, alterations of the gut microbiome are linked to PD, but less is known about iRBD or other prodromal stages of the disease. Moreover, most of the associations between PD and

the gut microbiome are based on taxonomic and metabolomic analyses. The resulting data, although insightful, lacks functional and systemic information that could better capture the complex crosstalk between the gut microbiome and the host in the context of PD. To obtain such information, we performed an integrated multi-omics study on a cross-sectional cohort comprised of individuals with iRBD and PD alongside HC. Metagenomics (MG), metatranscriptomics (MT), metaproteomics (MP) and meta-metabolomics (MM) were used to characterize taxonomic (taxMG, taxMT, taxMP) and functional (funMG, funMT, funMP, MM) differences between HC, iRBD and PD gut microbiomes. We identified substantial differences in gut microbiome functions and metabolites between the groups, including an increase in β-glutamate levels in PD that were related to a dysregulation of glutamate-related gene expression. Alterations in glutamate-related genes were linked with chemotaxis and flagellar assembly pathways, for which we identified strong and distinct taxonomic differences in transcription between PD and HC. Collectively, our data highlight the importance of multi-omics approaches for the identification of microbiome-mediated effects on neurological, and more broadly, complex human diseases involving host-microbiome interactions.

Results

Study cohort.

Our initial set of subjects consisted of 50 individuals with PD, diagnosed according to United Kingdom Parkinson's Disease Society Brain Bank (UKPDSBB) clinical diagnostic criteria[30], and 30 people with polysomnography-confirmed iRBD as well as 50 healthy control (HC) subjects. The data from 4 PD and 3 iRBD as well as 1 HC were subsequently excluded (see Methods), leading to a final data set of 46 PD, 27 iRBD and 49 HCs. The subjects in the three groups were of similar age but had slightly different gender distributions, with males overrepresented in the iRBD and PD groups (Table 1, Fisher test, $p = 0.004$), as is typical for these conditions[31,32]. Constipation, a prevalent non-motor symptom of PD[33], was also more common in the iRBD and PD groups compared to HC (Fisher test, $p < 0.001$).

Microbiome function is altered in PD.

Alpha diversity comparisons revealed no statistically significant differences between the three groups when considering taxMG, taxMT, funMG, taxMP and MM (Fig. 2A, Wilcoxon test, $p > 0.05$). However, funMT and funMP showed a statistically significant increase in alpha diversity in PD compared to HC and iRBD compared to PD, respectively (Fig. 2A, Wilcoxon test, $p < 0.05$). We then analyzed beta diversity for all omics layers. TaxMG, funMG and taxMP revealed no statistically significant differences between the three groups (Extended fig. 1 A, B and D, PERMANOVA, $p= 0.2$, 0.8 and 0.35, respectively), while taxMT, funMT, funMP and MM showed a statistically significant separation of the groups, especially for funMT (Fig. 2B and C, Extended fig. 1 C, PERMANOVA $p = 0.001$, 0.001, 0.005 and $p = 0.008$, respectively). Permutation analysis revealed that funMT, funMP and MM resulted in the best separation of the three groups while taxMG, funMG and taxMP exhibited the lowest separation capacity

(Extended fig. 1D, PERMANOVA R^2 of 0.043, 0.041, 0.038, 0.018, 0.014 and 0.017, respectively). Pairwise comparisons using PERMANOVA demonstrated that most differences were found between either HC and PD or HC and iRBD, with only funMP and funMT showing statistically significant differences between PD and iRBD (Fig. 2E). Based on these differences, we next assessed how the confounding factors sex, age and constipation may impact beta diversity. Sex and constipation were found to have a significant association with beta diversity for taxMT and funMT, while age was associated with taxMG (Fig. 2F). Importantly, MM and funMP were not found to be associated with any confounders (Fig. 2F).

We next looked at differential abundances on taxonomic layers to highlight the compositional differences between the groups. No significant differences were found in taxMG at the genus and species levels between any of the three groups (Fig. 2G and Extended fig. 2C, $q > 0.05$, SIAMCAT). For taxMT, SIAMCAT highlighted an increase in *Alistipes obesi* and *Ruthenibacterium lactatiformans* in iRBD vs HC, *Roseburia incertae sedis*, *Blautia massiliensis, B. obeum* and *Clostridium* sp. were decreased in PD vs HC and no significant differences were found between PD and iRBD (Fig. 2G, $q < 0.05$, $q < 0.05$ and $q > 0.05$, respectively). We identified the genus *Eubacterium* as being decreased in PD compared to HC in taxMT (Extended fig. 2C, $q < 0.05$, SIAMCAT). The ALDEx2 algorithm highlighted only *Roseburia incertae sedis* to be depleted in PD after FDR correction (Extended fig. 2B).

Subsequently, we investigated overall differences in gene abundances and expressions of the microbial genes linked to the observed differences in metabolites using the KEGG database. We observed no statistically significant differences in gene abundances between HC vs PD or HC vs iRBD (funMG, Extended fig. 4A and B, $q > 0.05$). Gene expression highlighted one transcript upregulated in HC vs iRBD, but 145 transcripts downregulated in HC vs PD (funMT, Extended fig. 4C and D, $q < 0.05$ and $q > 0.05$, respectively). No genes or transcripts were

significantly higher in iRBD or PD compared to HC (Extended fig. 4A-D, $q > 0.05$). Finally, we found no statistically significant differences in gene or transcript abundances between PD and iRBD (data not shown).

Altered metabolome of PD patients is linked to microbial abundance and activity.

Considering that MM is associated with PD and iRBD but not with confounders, we chose MM as a robust guide for further statistical comparisons. In addition, MM can be considered as one of the final outputs of microbial activity and an important driver of microbial effects on the host. For this purpose, we removed unidentified compounds from the statistical testing, because we cannot associate microbial genes with them. Our analyses revealed 11 statistically significantly different compounds between the groups, including alanine, β-glutamate, serine, and glycerol (Fig. 3A, $q < 0.05$). We found a significant increase in isovalerate, isobutyrate and valerate in PD patients (Fig. 3A, $q < 0.05$) but no differences for butyrate, acetate, formate, propionate or total SCFAs (data not shown). Primary bile acids glycocholic and chenodeoxycholic acids were decreased in PD and in iRBD and PD, respectively (Fig. 3A, $q < 0.05$). Glutamate was not differentially abundant between the three groups (data not shown). To unravel the effect of confounding factors, we measured the variance explained by each factor on the compounds' abundances. Diagnosis explained more variance than constipation or sex (Fig. 3B, diagnosis: mean= 4.18% ; median = 3.33%, sex: mean = 2.08%, median = 1.3% ; constipation: mean = 3.7%, median = 2.68%). We found only malic acid as being significantly associated with sex and no compound to be associated with constipation (Extended fig. 3A and B, $q < 0.05$, $q > 0.05$). Based on the PERMANOVA results and variance analysis, the differences observed in MM were most strongly associated with the disease status and, importantly, not with the confounding factors.

We next checked correlations between metabolite abundances and microbial abundances alongside activities. Metabolite abundances were linked to microbial abundances for taxMG and taxMT. TaxMG exhibited fewer significant correlations with the abundances of metabolites compared to taxMT, except for the genera *Akkermansia* and *Methanobrevibacter,* which showed the same associations to metabolites in taxMT (Extended fig. 3B, Spearman test, $q < 0.001$). Additional correlations seen in taxMT included positive correlations with β-glutamate, isovalerate, and isobutyrate with *Akkermansia* and *Clostridium* alongside *Methanobrevibacter,* and negative correlations with *Faecalibacterium* (Fig. 3C, Spearman test, $q < 0.05$). In addition*, Roseburia* and *Eubacterium* were positively correlated with glycerol (Fig. 3C, Spearman test, $q < 0.05$). Interestingly, *Akkermansia*, *Clostridium* and *Methanobrevibacter* were positively correlated with compounds increased in PD while *Roseburia*, *Faecalibacterium* and *Eubacterium* were negatively correlated, highlighting groups of bacteria being linked with either individuals with PD and iRBD or HC (Fig. 3C).

Expression of genes linked to glutamate and flagella is dysregulated in the PD gut microbiome, but not in iRBD.

To reinforce the results of the differential analysis, we acquired all the orthologs related to the metabolites which were statistically significant between the three groups. More specifically, we used regular expression matching to retrieve all orthologous genes that are linked to the above-mentioned metabolites in KEGG. Because the KEGG database only has one metabolite entry annotated as β-glutamate, we selected all glutamate-related genes instead (linked to both L- and D-glutamate). We found no statistically significant differences in gene abundances after correction between any of the three groups (funMG, Fig. 3D and Extended fig. 3D, $q > 0.05$, Wilcoxon test). However, we did find a decrease in transcripts in PD for the three known glutamate synthase genes (funMT, GLT1:K00264, GLU:K00284 and gltB:K00265, Fig. 3D,

Wilcoxon test, $q = 0.004$, $q = 0.004$ and $p = 0.006$, respectively) alongside a decrease in cheB in PD (K03412, protein-glutamate methylesterase/glutaminase, Fig. 3D, $q < 0.01$, Wilcoxon test). Furthermore, we found an increase in cofE, mainly found in Archaea and involved in methanogenesis (K12234, coenzyme F420-0:L-glutamate ligase, $q < 0.05$, Wilcoxon test). In addition, we found a decrease in BA-related transcripts in PD while transcripts related to serine and isovalerate were increased in PD (Fig. 3D, $q < 0.05$). HC vs iRBD comparisons revealed significant differences in alanine-related transcripts but not for the other metabolites (Extended fig. 3D, $q < 0.05$). β-glutamate abundance was negatively correlated with the transcripts of the three glutamate synthases and carbamoyl-phosphate synthases, but positively correlated with methyl aspartate mutase, methylamine-glutamate N-methyltransferase and glutaminase (Extended fig. 5C, Spearman correlation test, $q < 0.01$).

Since cheB is part of the chemotaxis gene family, we further inspected the chemotaxis and flagellar assembly pathways (two pathways with overlap in orthologs) to assess the microbial capacity for motility. We found no statistically significant differences in gene abundances (funMG) for either flagellar assembly or chemotaxis pathways between the three different groups (data not shown, $q > 0.05$). In contrast however, based on funMT, the chemotaxis pathway (14/26 transcripts downregulated, 0/26 transcripts up-regulated, Fig. 3E, $q < 0.05$) and the flagellar assembly pathway (30/46 transcripts downregulated, 1/46 transcripts up-regulated in PD, Fig. 3E, $q < 0.05$) were strongly downregulated in PD but not in iRBD vs HC (data not shown). Moreover, eight transcripts showed a decrease in alpha diversity for chemotaxis and flagellin assembly pathways (Fig. 3E, $p < 0.01$). Finally, we observed a decrease in alpha diversity for GLU, flagellar assembly transcripts (fliJ, fliQ and fliC) and cheD in PD compared to the other groups (Fig. 3F, $q < 0.05$, Dunn test) while dltD and aaaT were elevated in iRBD (Fig. 3F, Dunn test, $p < 0.05$).

Flagellin and chemotaxis are differentially expressed depending on taxonomy and disease status.

We next assessed the taxa expressing flagellar assembly and chemotaxis pathway genes. Interestingly, flagella and chemotaxis genes were expressed differentially according to disease and taxonomy. More specifically, we identified different clusters of taxa expressing flagella and chemotaxis genes. The first cluster was composed of microbes expressing these genes principally in PD, including *Ruminiclostridium*, *Enterocloster*, *Dysosmobacter* and *Butyvibrio*. A second cluster was composed of microbes expressing these genes principally in HC, including *Roseburia*, *Agathobacter* and *Eubacterium* (Fig. 4). Strikingly, we found that the third cluster is composed of taxa expressing flagellin or chemotaxis genes only in PD, including *Escherichia*, *Cellulosilyticum*, *Citrobacter* or *Eisenbergiella*; a fourth cluster was composed only of taxa expressing in HC including *Flavonifractor*, *Succinivibrio*, *Eisenbergiella* or *CAG-603* (Fig. 4). We subsequently investigated the expression levels of genes of the extracellular parts of flagella in the cluster wherein *Roseburia* and most *Lachnospiraceae* were located (Cluster 2). Overall, we found a decrease in flagellin (fliC), filament cap (fliD, fliS) and hook-filament junction genes (flgK, flgL) in PD for *Roseburia*, *CAG-115* and *Agathobacter* (Extended fig. 6, $p < 0.05$, Wilcoxon test).

Metabolites and metabolites-related genes associated with PD are central to microbiome metabolism.

To quantify the importance of glutamate derivatives and related genes in microbial metabolism, we next reconstructed microbiome-wide metabolic networks as previously described[34] . We mapped genes related to the compounds identified earlier as significantly different between the different groups. The metabolic network highlighted glutamate-related genes as compact and placed in the middle of the network while glycerol was more scattered across the network (Fig.

5A). Strikingly, glutamate-related genes formed a subnetwork central to the overall network with a betweenness centrality measure (BC) of 93.0 compared to 0.0009 for the whole community (Fig. 5B). Crucially, L-glutamate was the most central non-cofactor metabolite in the network and 2-oxoglutarate (L-glutamate derivative) was the second most central metabolite (Fig. 5C). Considering this, it is apparent that glutamate and glutamate-related genes are central to microbiome metabolism and that modifications in the levels of these metabolites or transcripts reflect profound modifications of microbial metabolism.

Multi-Omics Factor Analysis validates β-glutamate and flagella links with PD.

To validate our findings, we used an unsupervised method with the Multi-Omics Factor Analysis (MOFA). The resulting MOFA model included 10 factors (F1-10 hereafter) whereby F1 showed a strong association with disease status ($p < 0.05$, ANOVA, Fig. 6A). A complete description of the MOFA model is provided in the Extended information. We found that F1 mainly explained the variance of funMT, taxMT, and MM (17.6%, 9.6%, and 5.7%, respectively, Extended fig. 6A-B) and showed separation of HC and PD, but not iRBD versus other groups ($p < 0.05$, ANOVA, Fig. 6B). Specifically, the microbiome of PD was characterized by the joint increase in abundance of *M. smithii*, archaeal proteins and genes (based on MP and MG data), *A. muciniphila* (based on taxMT data), β-glutamate, isovalerate, isobutyrate, hexadecanoic, and hyocholic acids, whereas the abundance of *R. hominis* (taxMG data), flagellin (funMT), GCA and glycerol were decreased in PD (Fig. 6C). Overall, the MOFA results are strongly consistent with the per omic layer analysis and validate these findings.

Discussion

Here we investigate the links between the gut microbiome and PD using an integrated multi-omics approach based on standardized sample collection and extraction. We include individuals with iRBD as a prodrome of PD to compare early and later stages of the disease but do not find statistically significant differences between iRBD and PD, especially in comparison to the more pronounced differences found between HC and PD. Previous studies have shown differences in the different early stages of PD, but these studies were performed using 16S rRNA gene amplicon data[7,10]. In contrast to the amplicon-based results, no differentiation between iRBD and PD are found using metagenomic data[5]. More specifically, and contrary to previous findings[5] including our own work[7], we do not find significant differences between HC and PD individuals in the metagenome with respect to alpha or beta diversity. However, we show differences in transcriptional activity for *Roseburia*, *Blautia* and *Eubacterium* which are known to be decreased in PD[5,6,8–13]. Moreover, we find several genera, including *Akkermansia*, *Methanobrevibacter* and *Clostridium,* correlating positively with the abundance of compounds increased in PD such as β-glutamate, isovalerate or isobutyrate, while other genera are inversely correlated with these compounds (*Faecalibacterium*, *Blautia*, *Eubacterium* and *Roseburia*). Collectively, these findings are in line with findings associating those genera with PD[5,6,8–13].

We identify metabolites that are associated with PD, amongst which BAs, alanine, serine and β-glutamate highlighted linked differences in gene expression. The BA chenodeoxycholic acid (CDCA) is decreased in PD and iRBD whereas glycocholic acid (GCA) is decreased only in iRBD. We also detect a decrease in transcripts related to BA in PD but not iRBD. BAs have a wide spectrum of effects on the immune system[22,23], metabolism[35], hormones[36] and also on the CNS[37]. GCA has been found to be increased in the CNS of PD and associated with disease

duration[37], stressing the importance of analyzing BAs in PD and prodromal stages of PD. In contrast with previous studies[6,15], we do not find a decrease in butyrate, propionate, acetate or overall levels of SCFAs. However, we find an increase in isovalerate, isobutyrate and valerate as described previously[15]. Previous studies have highlighted the correlation of fecal concentrations of isobutyrate and isovalerate with PD severity[38] and concentrations of valerate with disease duration[39]. There currently is an apparent lack of knowledge regarding isobutyrate or isovalerate with respect to host-microbiome interactions and very few related genes annotated in the KEGG database hindering interpretations concerning potential links between these metabolites and PD.

β-glutamate and glutamate-related genes are of particular interest in the context of PD because of the reported toxic effects of L-glutamate on neurons[24] and because of its association with microbial activity. In addition, glutamate levels have been reported to be increased in PD individuals' blood sera[17,18]. β-glutamate has strikingly only one reaction described in the KEGG database, in contrast, L-glutamate has 213 reactions and 54 pathways, while D-glutamate has 12 reactions and 2 pathways. We find that glutamate and glutamate-related genes are central to microbial metabolism, which underpins the notion that the highlighted differences reflect a pronounced impact on gut microbiome function. Of note, glutamate likely has local effects on enteric neurons with subsequent influences on the CNS[40]. Although we show significant differences in β-glutamate and glutamate-related genes in the stool of PD patients, we detect no significant differences in the glutamate levels (both enantiomers) between the three groups. We highlight significant differences in β-glutamate for which we currently cannot evaluate the effects on the ENS and CNS. Overall, based on our results, microbiome-driven glutamate metabolism and its impact on the glutamatergic system must be comprehensively studied in the future to disentangle its link with PD.

Glutamate-related genes are further involved in chemotaxis and flagellar assembly pathways, highlighting a modification of the latter genes' expression with a majority being decreased in PD individuals but not in iRBD. A decrease in flagellar assembly gene abundances has been previously reported in a metagenomic analyses of PD[9]. We do not see significant differences in the linked gene abundances, but their expression levels are significantly different, highlighting altered regulation of transcription in PD compared to HC. Flagellar assembly and chemotaxis genes are also differentially expressed by specific microbes; the genera *Escherichia* and *Cellulosilyticum* for instance are expressing flagellar assembly genes only in the context of PD without being statistically differentially abundant.

Flagellin is a known immunogenic molecule, a potent pro-inflammatory compound in pathogens[41–43] and is thus targeted by secretory IgA[44]. However, flagellin in commensals, especially in the *Lachnospiraceae* family, has been shown to be either 'silently recognised'[45] or elicit anti-inflammatory effects[46–48]. Amongst the *Lachnospiraceae* family, the genus *Roseburia* shows a decrease in the transcription of flagellin in the gut microbiome of PD. This in turn may be linked to immune system dysregulation and exert indirect effect on the CNS, particularly in microglia as shown in a previous study using a murine model of PD[49]. Previous studies have shown an increased inflammatory state in PD, with increased pro-inflammatory circulating immune cells[50], cytokines[51], and activated microglia[52]. In addition, microglia may be activated by α-synuclein via NLRP3 following TLR2 (lipopolysaccharide (LPS) sensing) and TLR5 (flagellin sensing) activation[53]. Therefore, the activation or inhibition of TLRs by distinct flagellins may modulate microglia activation by competing with α-synuclein and affect PD progression. Bacterial antigens such as flagellin are usually regarded as inflammatory agents that elicit a strong immune response. Our findings, which reveal elevated expression of flagellin in HC, suggest the need for a re-evaluation of the impact of common antigens on the immune system. The gut microbiome produces a wide range of immune-modulating

compounds, the impact of which must be re-evaluated through detailed *in vivo* and *ex vivo* studies.

Overall, our work clearly highlights the importance of studying microbiome functions rather than restricting microbiome analyses to taxonomic structure. Specifically, the combination of MT and MM provides clear insight into the activity of specific microbial taxa in relation to disease. In our present work, MT reveals that disease association is not solely determined by gene expression levels; the diversity of microbes capable of expressing a specific function and the specific taxa expressing those functions are also of immediate interest and relevance. The future of microbiome research might lie in understanding how we can modulate, re-activate or shut down specific microbial functions *in vivo* in order to improve knowledge and later improve or functionally tailor microbiome interventions.

Methods

Patient cohorts and sampling

All subjects from both cohorts provided informed written consent, and the sample analysis was approved by the Comité National d'Ethique de Recherche of Luxembourg (reference no.: 140174_ND).

Kassel Cohort

The DeNoPa cohort represents a prospective, biannual follow-up study of (initially *de novo)* Parkinson's disease (PD) patients at the Paracelsus-Elena Klinik, Kassel, Germany. Fecal samples from PD patients (46) and healthy controls (29) were collected during the 4-year follow-up visit for the cohort. Details on inclusion and exclusion criteria and ancillary investigations have been published previously[54,55]. Subjects with idiopathic rapid-eye-movement sleep behaviour disorder (iRBD, 13) were recruited at the same clinic, diagnosed according to consensus criteria of the International RBD study group[56] using video-assisted polysomnography, and were included only if they showed no signs of a neurodegenerative disorder. DeNoPa subjects were required to have a 4-week antibiotic free interval before fecal sample collection. As additional control subjects, we collected fecal samples from (20) neurologically healthy subjects living in the same household as the DeNoPa participants. Samples of *de novo* PD patients from a cross-sectional cohort at the same clinic were included if subjects were recently diagnosed, drug-naïve and met United Kingdom Parkinson's Disease Society Brain Bank (UKPDSBB) clinical diagnostic criteria[30]. All subjects except household HC were interviewed and examined by an expert in movement disorders. The study conformed to the Declaration of Helsinki and was approved by the ethics committee of the Physician's Board Hessen, Germany (FF 89/2008). The DeNoPa trial is registered at the German Register for Clinical trials (DRKS00000540).

Marburg Cohort

We also added samples from 14 patients with polysomnography-confirmed iRBD which were recruited from the outpatient clinic of the Department of Neurology, Philipps-University, Marburg, Germany, between November 2015 and November 2016. iRBD was diagnosed according to the guidelines of the American Academy of Sleep Medicine (AASM ICSD-3)[57]. A detailed medical history was recorded, and a complete neurological examination performed to verify the subjects' suitability. Inclusion criteria were age above 18 years, no dopamimetic therapy, and no diagnosis of PD, MSA, DLB or PSP. Exclusion criteria were smoking, antibiotic therapy in the last 24 months, history of other neurological diseases or disorders of the gastrointestinal tract. Non-motor and autonomic symptoms were evaluated with the SCOPA-AUT[58] and PD-NMS[59] questionnaires. Motor function was evaluated with the UPDRS[60]. Additionally, patients were asked to complete the RBD-Sleep questionnaire[61]. The study conformed to the Declaration of Helsinki and was approved by the ethics committee of the Medical Faculty of the Philipps-University, Marburg, Germany (46/14).

Fecal sample collection

Fecal samples were collected at the clinics via a stool specimen collector (MedAuxil) and collection tubes (Sarstedt), as previously described[7]. Samples were immediately flash-frozen on dry ice after collection. Samples were subsequently stored at –80 °C and shipped on dry ice.

Sample exclusions

The initial set of samples consisted of 50 PD, 30 iRBD and 50 healthy control subjects (HC). Three PD and two iRBD cases were subsequently excluded for clinical reasons (adjusted diagnosis), one iRBD and one PD subject for logistical reasons, and one control due to a combination of microbiome-altering medications (metformin, antidepressants, statins, and proton pump inhibitors). Additional samples were excluded due to missing values (metabolomics) or a low amount of identified analytes (metaproteomics), leading to the final numbers of samples summarized below:

- Metagenomics (MG) & metatranscriptomics (MT): 49 HC, 27 iRBD, 46 PD
- Metaproteomics (MP): 42 HC, 22 iRBD, 40 PD
- Meta-metabolomics: 49 HC, 27 iRBD, 41 PD

Metagenomic and metatranscriptomic sequencing

Extractions from fecal samples were performed according to a previously published protocol[62], conducted on a customized robotic system (Tecan Freedom EVO 200). After extraction, DNA and RNA were purified prior the sequencing analysis by using the following commercial kits respectively: Zymo DNA Clean&Concentrator-5 (D4014) and Zymo RNA Clean&Concentrator-5 (R1014). RNA quality was assessed and quantified with an Agilent 2100 Bioanalyser (Agilent Technologies) and the Agilent RNA 6000 Nano kit, and genomic DNA and RNA fractions with a NanoDrop Spectrophotometer 1000 (Thermo Scientific) as well as commercial kits from Qubit (Qubit ds DNA BR Assay kit, Q32850; Qubit RNA BR Assay kit, Q10210). All DNA samples were subjected to random shotgun sequencing. Following DNA isolation, 200-300 ng of DNA was sheared using a Bioruptor NGS (Diagenode) with 30s ON and 30s OFF for 20 cycles. Sequencing libraries were prepared using the TruSeq Nano DNA library preparation kit (Illumina) following the manufacturer's protocol, with 350 bp average insert size. For MT, 1 µg of isolated RNA was rRNA-depleted using the RiboZero kit (Illumina, MRZB12424). Library preparation was performed using the TruSeq Stranded mRNA library preparation kit (Illumina) following the manufacturer's protocol, apart from omitting the initial steps for mRNA pull down. MG and MT analyses, the qualities of the libraries were checked using a Bioanalyzer (Agilent) and quantified using Qubit (Invitrogen). Libraries were sequenced on an Illumina NextSeq500 instrument with 2x150 bp read length.

Metaproteomics

20 µL protein extract were processed using the paramagnetic bead approach with SP3 carboxylate coated beads[63,64]. Briefly, the protein samples were reduced with 2µL 25 mM DTT in 20 mM ammonium bicarbonate (Sigma-Aldrich) for 1 h at 60°C. Subsequently, 4 µL 100

410 mM iodoacetamide (Merck) in 20 mM ammonium bicarbonate was added and incubated for
411 30 min at 37°C in the dark. Next, 5 µL of 10% formic acid was added as well as 70 µL 100%
412 acetonitrile (ACN) to reach a final organic content higher than 50% (v/v). 2 µL SP3 beads per
413 sample were washed with water three times with subsequent addition of the sample. After
414 protein binding to the beads, the supernatant was discarded. The beads were washed twice with
415 200 µL 70% (v/v) ethanol, and once with 200 µL ACN. The protein lysates were proteolytically
416 cleaved using trypsin (1:50) over night at 37 °C. Since trypsin is added in aqueous solution to
417 the samples, the proteins are not bound to the beads during enzymatic cleavage. ACN was
418 added to each sample to reach a final organic content higher than 95% (v/v). After peptide
419 binding to the beads, the samples were washed with pure ACN on the magnetic rack. Finally,
420 the peptides were eluted in two steps. First, with 200 µL 87% ACN (v/v) containing 10 mM
421 ammonium formate (pH 10), and next with two times adding 50 µL water containing 2 % (v/v)
422 DMSO and combination of the two aqueous supernatants. Thus, two fractions of peptides were
423 generated, which were evaporated and re-dissolved in water containing 0.1 % formic acid (20
424 µL) and analyzed on a Q Exactive HF instrument (Thermo Fisher Scientific) equipped with a
425 TriVersa NanoMate source (Advion) in LC chip coupling mode. Peptide lysates were injected
426 on a trapping column (Acclaim PepMap 100 C18, 3 µm, nanoViper, 75 µm x 2 cm, Thermo
427 Fisher Scientific) with 5 µL/min by using 98% water/2% ACN 0.5% trifluoroacetic acid, and
428 separated on an analytical column (Acclaim PepMap 100 C18, 3 µm, nanoViper, 75 µm x 25
429 cm, Thermo Fisher Scientific) with a flow rate of 300 nL/min. Mobile phase was 0.1% formic
430 acid in water (A) and 80 % ACN/0.08 % formic acid in water (B). Full MS spectra (350–1,550
431 *m/z*) were acquired in the Orbitrap at a resolution of 120,000 with automatic gain control (AGC)
432 target value of 3×10^6 ions.

Meta-metabolomics

Untargeted GC-MS as well as targeted measurements (SCFA GC-MS/MS and bile acids LC-MS/MS) from fecal samples were performed according to a previously published protocol[65]. All GC-MS chromatograms were processed using MetaboliteDetector, v3.220190704[66] while LC-MS chromatogram were acquired with Thermo Xcalibur software (version 4.1.31.9) and analyzed with TaceFinder (Version 4.1). Compounds were initially annotated by retention time and mass spectrum using an in-house mass spectral library. Internal standards were added at the same concentration to every medium sample to correct for uncontrolled sample losses and analyte degradation during metabolite extraction. The data was normalized by using the response ratio of the integrated peak area of the analyte and the integrated peak area of the internal standard.

Bioinformatics and statistical analysis

Sequencing data processing and analysis

For all samples, MG and MT sequencing data were processed and hybrid-assembled using the Integrated Meta-omic Pipeline (IMP)[67] (https://git-r3lab.uni.lu/IMP/imp3, commit 8c1bd6fa443d064511909c9eede20703f45e6c69). It includes steps for the trimming and quality filtering of the reads, the filtering of rRNA from the MT data, and the removal of human reads after mapping against the human genome (hg38). Pre-processed MG and MT reads were assembled using the IMP-based iterative hybrid-assembly pipeline using MEGAHIT[68] 1.0.3. After assembly, the prediction and annotation of structural features such as open-reading frames (ORFs) was performed using a modified version of Prokka[69] and followed by functional annotation of those using Mantis[70]. Structural features were quantified on MG and MT level using featureCounts[71]. Taxonomic annotation of reads and contigs was performed using Kraken2[72] with a GTDB release207 database (http://ftp.tue.mpg.de/ebio/projects/struo2/GTDB_release207/kraken2) and a 0.5 confidence

458 threshold. Additionally, taxon abundances were estimated using mOTUs 2.5.1[73]. The mOTU
459 abundances were used to generate abundance matrices for each taxonomic rank (phylum, class,
460 order, family, genus and species) by summing up taxon marker read counts at the respective
461 levels.

462 **Metaproteomics prediction and annotation**

463 For each sample, the predicted proteins were concatenated with a cRAP database of
464 contaminants and the human UniProtKB Reference Proteome prior to the MP search. In
465 addition, reversed sequences of all protein entries were added to the databases for the
466 estimation of false discovery rates. The search was performed using Sipros v1.1[74] as search
467 engine with the following parameters: trypsin was used as the digestion enzyme and a
468 maximum of two missed cleavages was allowed. The tolerance levels for matching to the
469 database were 1 Da for MS1 and 0.01 Da for MS2. Peptides with large errors for parent ions
470 were later filtered out by setting the Filter Mass Tolerance Parent Ion parameter to 0.05 Da.
471 Carbamidomethylation of cysteine residues was set as a fixed modification and oxidation of
472 methionines was allowed as a variable modification. Peptides with length between 7 and 60
473 amino acids, with a charge state composed between +2 and +4 and a maximum missed
474 cleavages of 3 were considered for identification. The results from all identifications were
475 filtered by Sipros using at least one unique peptide per protein and peptide false discovery rate
476 (FDR) was dynamically set to achieve a 1% of protein FDR.

477 Data analysis was performed on all samples with at least 2000 proteins identified. A summary
478 matrix of all selected samples consisting of the KO annotations from the integrated MG and
479 MT analysis and the spectral count from the MP identification was then generated and used for
480 statistical analysis.

481

Dimensionality reduction and ordination

Beta diversity for MG and MT was assessed using the Bray-Curtis dissimilarity and subjected to a Non-Metric MultiDimensionnal Scaling (NMDS) for both the taxonomic and functional levels, using the *metaMDS()* function from the vegan package (2.6.2). Principal Component Analysis (PCA) was performed for MP and MM using the *rda()* function in the vegan package (2.6.4). PERMANOVA was used to assess statistical differences between groups using the Bray-Curtis dissimilarity and conducted in the vegan R package with the *adonis2()* function.

Differential abundance analysis and correlations

Differential abundance was done in two different approaches. The first approach consisted in using SIAMCAT[75] and ALDEx2[76] algorithm to find all the taxa and genes differentially expressed between groups without prior assumption. We used two different algorithms to have a sensitive algorithm (SIAMCAT, less prone to have false negative) and a more conservative one (ALDEx2, less prone to have false positive). The second approach consisted in using the MM significant compounds to drive the analysis on the functional level for MG and MT. Therefore, differential abundance tests and multiple correlation tests were conducted with a classical approach. We used Mann-Whitney or Kruskal-Wallis followed by a Dunn test (depending on the number of groups) and Spearman correlation tests. We applied FDR correction using the Benjamini & Hochberg method[77]. We depicted both FDR corrected as q-values and non-FDR corrected p-values to represent most of the differences found in our datasets. All statistical tests were done using the *rstatix* package (0.7.2).

Variance analysis

Variance analysis was used to assess the importance of each clinical factor on MM. To verify the covariance of factors and to assess which factors explained the most variance in our datasets, we computed the total variance for each clinical factor (removing the NAs for each

506 factor) and the variance explained by each group within a clinical factor. Explained variance
507 was calculated as follows: $var.explained = \frac{1-variance.group}{variance.total}$.

508 **Microbiome-wide metabolic network analysis**

509 The microbiome-wide metabolic network analysis was conducted by establishing an
510 association between KEGG KOs and corresponding ChEBI IDs. The networks were visualized
511 utilizing the NetworkX package (release 3.3)[78]. In this network, the nodes were represented by
512 KEGG KOs, while the edges were denoted by the corresponding metabolites (either products
513 or reactants)[34]. The analysis was restricted to genes that were present in a minimum of 50% of
514 the samples. To construct metabolite-specific networks, we used KEGG KOs which have either
515 a reactant or product in KEGG. Glutamate-, thymine-, glycerol-, serine-, alanine- and
516 glucuronate-specific subnetworks were composed of 146, 9, 66, 43, 70, 18 genes, respectively.
517 The network topology metric 'Betweenness centrality' was used to underscore the importance
518 of a metabolite in microbiome-wide metabolism[34].

519 **Integrated multi-omics analysis using MOFA2**

520 Integrative analysis for the seven omics layers was conducted with the Multi-Omics Factor
521 Analysis (MOFA) 2 R package (version 1.10.0)[79]. Before the analysis, data were preprocessed
522 as follows: (I) funMG, funMT, funMP, taxMG, taxMT, and taxMP data were filtered based on
523 the number of non-zero features, a feature was kept if it was present in at least in 25% of
524 samples in each group (PD, HC, iRBD) or at least in 75% in any of the groups; (II) funMG,
525 funMT, taxMG, and taxMT count data were separately residualized in a linear model to remove
526 variance explained by differences in sequencing depth; (III) funMP and taxMP data were
527 residualized by the sum of protein counts per sample and information on the number of high-
528 quality proteins recovered per sample; (IV) regression residuals were cubic-root transformed
529 to account for heteroscedasticity; (V) MM data were transformed using a centered log-ratio

transformation. Each dataset was then additionally filtered to retain the features with the largest variance for the subsequent analysis. For funMG and funMT, we included features with variances equal to or larger than 90% feature variance for a dataset, for the other datasets we included features with variance equal to or larger than the median feature variance for a given dataset. In the results, the feature size for omics layers was as follows: 759 for funMG, 657 for funMT, 410 for funMP, 109 for taxMG, 71 for taxMT, 115 for taxMP, and 34 for MM. MOFA analysis was run on scaled omics data with fifteen initial factors. All factors that explained less than 2% of the variance were excluded from the model. The remaining factors were tested for differential abundance between the groups studied using the linear regression followed by ANOVA type II controlling the participants' sex, age, and recruitment cohort.

Data availability

The datasets generated by this study are available in the following repositories: metagenomic and metatranscriptomic data at the NCBI BioProject collection with the ID PRJNA782492 (http://www.ncbi.nlm.nih.gov/bioproject/782492), metaproteomic data at the Proteomics Identifications (PRIDE) database with accession number PXD031457 (https://www.ebi.ac.uk/pride/archive/projects/PXD031457), and metabolomic data at MetaboLights with ID MTBLS5092 (https://www.ebi.ac.uk/metabolights/MTBLS5092). Due to privacy restrictions, clinical and demographic data are available on request from the corresponding author.

Code availability

The IMP pipeline, which was used for analysis of metagenomic and metatranscriptomic data, is available at https://gitlab.lcsb.uni.lu/IMP/imp3. The R and python code used for statistical analyses and visualizations is available at https://gitlab.lcsb.uni.lu/ESB/[TBA].

Acknowledgments

We thank the staff of the Luxembourg Centre for Systems Biomedicine (LCSB), particularly the sequencing platform and the metabolomics platform, for running the sequencing and metabolomic analyses. The bioinformatics presented in this paper were carried out using the HPC facilities of the University of Luxembourg[80].

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 863664), and was further supported by the Luxembourg National Research Fund (FNR) CORE/16/BM/11333923 (MiBiPa), CORE/15/BM/10404093 (microCancer/MUST),

PRIDE/11823097 (MICROH DTU), and CORE/19/BM/13684739 (metaPUF), the Michael J. Fox Foundation under grant IDs 14701 (MiBiPa-PLUS) and MJFF-019228 (PARKdiet), the Parkinson's Foundation (MiBiPa Saliva), the Institute for Advanced Studies of the University of Luxembourg through an AUDACITY grant (ref. no. MCI-BIOME_2019), as well as an "Espoîr en tête" grant from the Rotary Club Luxembourg to P.W. This work was also supported by a Fulbright Research Scholarship from the Commission for Educational Exchange between the United States, Belgium and Luxembourg to P.W. Additional funding was provided by the FNR under INTERMOBILITY/23/17856242. C.d.R. received funding from the European Union's Horizon 2020 Widening Fellowships (N° 101026381).

Contributions

Conceptualization: J.-P.T., A.H.-B., W.O., B.M., P.W. Patient recruitment, clinical coordination, and sampling: S.S., A.J., C.T., W.O., B.M. Multi-omic data generation: C.D.R., J.-P.T., C.J., L.A.L, A.D, N.J, M.v.B. Bioinformatics, statistics and data visualization: R.V, J.O.S., P.N., V.T.E.A, V.P, O.H, B.K, P.M. Initial manuscript draft: R.V, J.O.S, P.N, V.T.E.A, V.P. Review and editing: C.C.L., S.B.B., P.M. and P.W. Funding acquisition: C.C.L, W.O, B.M, P.W. All authors read and approved of the submitted version.

Rights Retention Statement

This research was funded in whole, or in part, by the Luxembourg National Research Fund (FNR), grant references CORE/16/BM/11333923 (MiBiPa), CORE/15/BM/10404093 (microCancer/MUST), PRIDE/11823097 (MICROH DTU), and CORE/19/BM/13684739 (metaPUF). For the purpose of open access, and in fulfilment of the obligations arising from

Figure 1. Schematic representation of the analytical workflow. Metagenomic (MG), metatranscriptomic (MT), metaproteomic (MP) and meta-metabolomic (MM) data were generated for each sample. Pre-processed MG and MT reads were sample-wise assembled using the iterative hybrid assembly pipeline of the Integrated Meta-omics Pipeline (IMP). After assembly, taxonomic annotation was performed at the read and contig levels, followed by gene prediction and functional annotation on the assembled contigs. Expressed proteins (MP) were identified using the predicted genes from the MG/MT hybrid assembly. For these three omics levels, we generated taxonomic and functional profiles that are referred to as taxMG, taxMT and taxMP for the taxonomic level, and funMG, funMT and funMP for the functional level, respectively. Community-based networks were reconstructed from gene annotations. Finally, the meta-metabolome (MM) was integrated with the other omics data at the network level. The integrated multi-omics analysis was performed with the available clinical metadata.

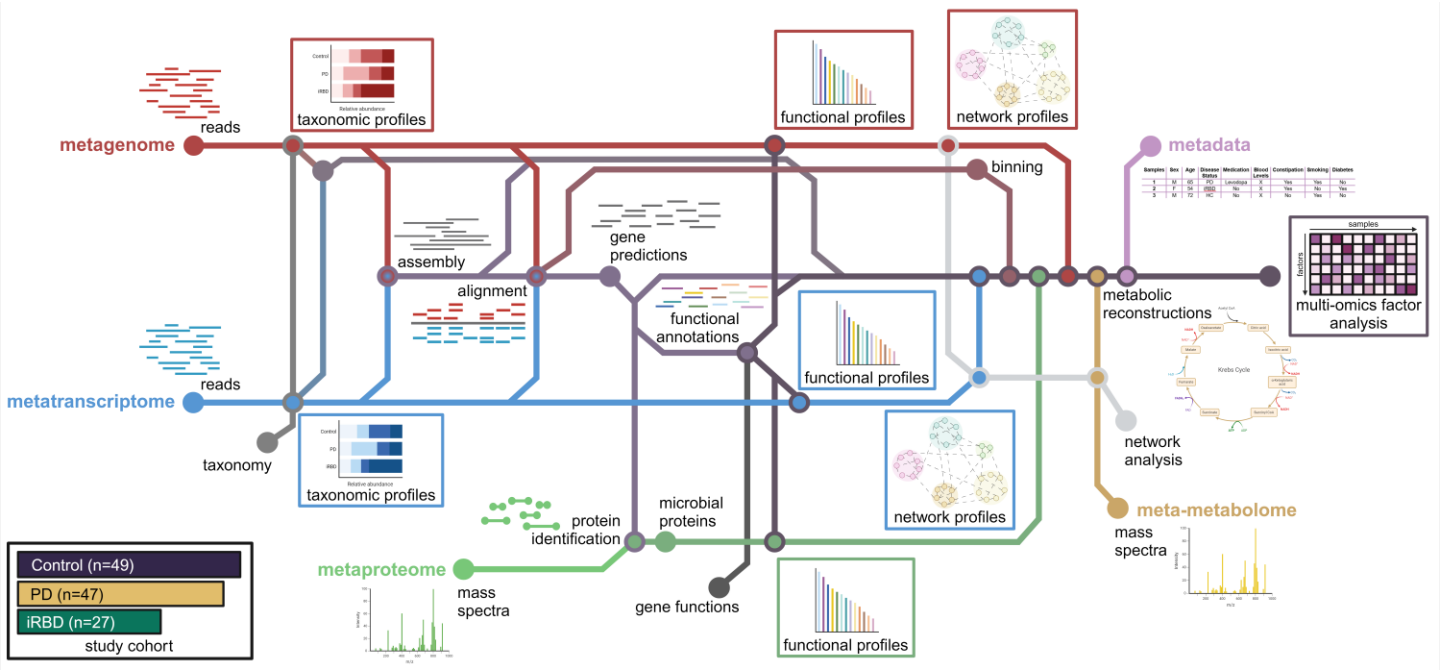


Figure 2. Microbiome structure is altered in PD and iRBD vs HC. A. Shannon index for different omics between Healthy Controls (HC), idiopathic REM sleep behaviour disorder (iRBD) and Parkinson's Disease (PD). P-values are based on pairwise Mann-Whitney tests. Non-metric Multidimensional Scaling (NMDS) based on Bray-Curtis dissimilarity for taxonomic annotation of metatranscriptomic (taxMT) (**B**) and functional metatranscriptomic (funMT) data (**C**). **D.** Principal Component Analysis (PCA) for meta-metabolomic (MM) data based on untargeted, targeted SCFA and targeted bile acids abundance. All three quantifications have been sum-normalised before any merging. PCA was then computed on the merged matrix. All tests are based on PERMANOVA with 1000 permutations. **E.** Pairwise PERMANOVA between groups for each omics. **F.** PERMANOVA analysis for "Age", "Sex" and "Constipation" for each omics. Size of rectangle is based on –log10(p-value) and colour on R^2 value. **G.** Differential abundance analysis using SIAMCAT for taxMG and taxMT. Values are pseudo fold changes for pairwise comparison between groups and size is based on –log10(p-value). Shape is based on significance before and after FDR correction.

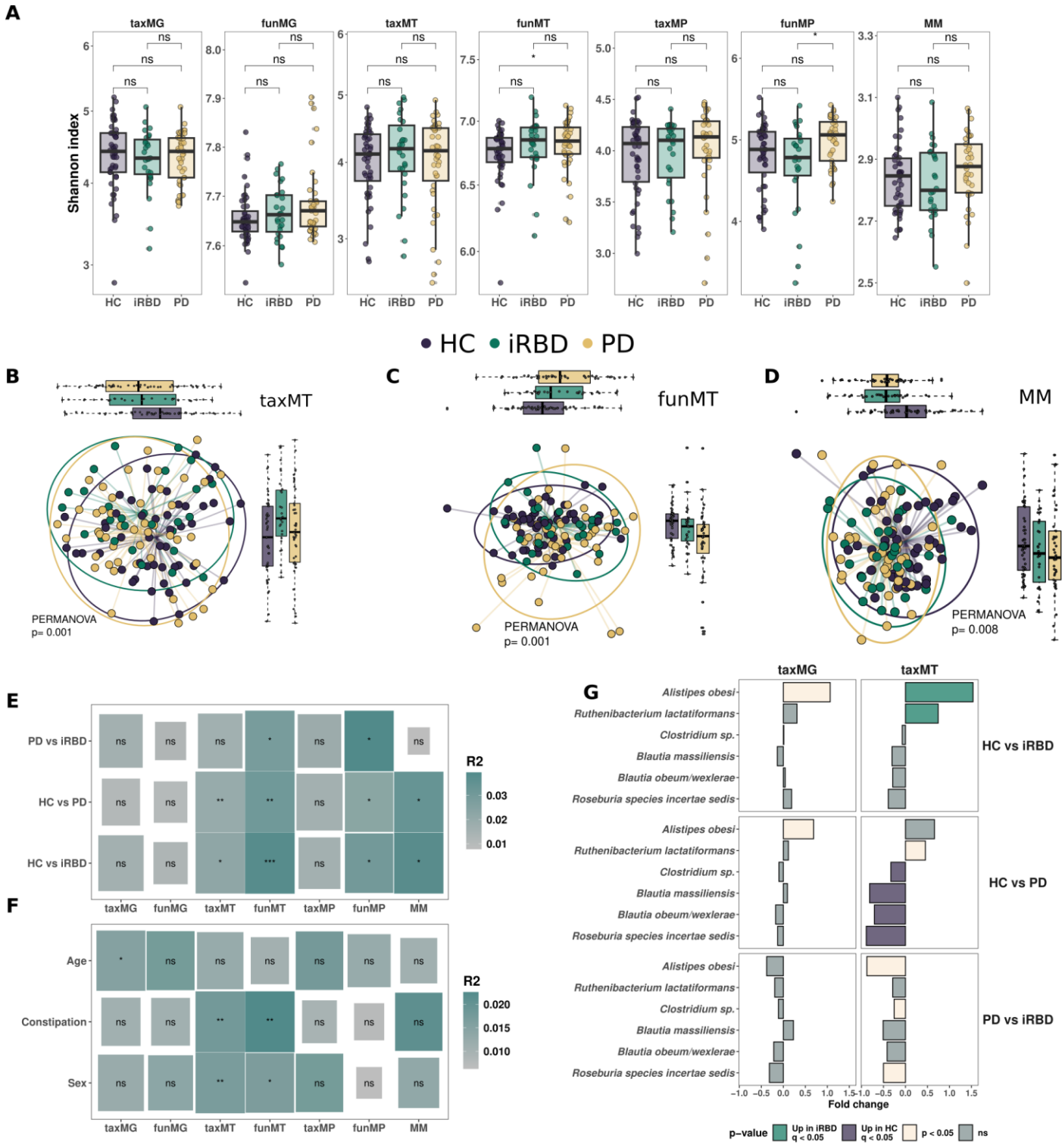


Figure 3 Altered metabolome is linked with microbial activity and transcripts. A. Metabolite relative abundances for significant compounds. Metabolomic data from untargeted meta-metabolomics, targeted SCFA and targeted bile acids were combined after normalization by sum for each. Dunn test, FDR corrected. **B.** Variance for each metabolite associated to the clinical factors “Diagnosis”, “Sex” and “Constipation”. **C.** Spearman correlation between taxMT at the genus level with significant metabolites. All p-values are FDR corrected. Genera are selected based on differential abundance or relevance in the literature. **D.** Absolute log2 fold change between HC and PD for funMG and funMT associated to significant compounds. Dots are scaled by the –log10(p-value), colorized and shaped according to p-value significance before (triangle shape) and after FDR correction (round shape). **E.** Chemotaxis and flagellin assembly pathway genes expression and Shannon index fold change between HC and PD group. Dots are colorized and shaped according to p-value significance before (triangle shape) and after FDR correction (round shape). Wilcoxon test, FDR corrected. **F.** Shannon index for significantly different expressed genes found in pairwise differential analysis in Fig. 2D and Extended fig. 2A. Shannon index was calculated for each KO for funMG and funMT. Only genes significantly different after FDR correction on a Kruskal-Wallis test are plotted. P-values are calculated using Dunn post hoc test.

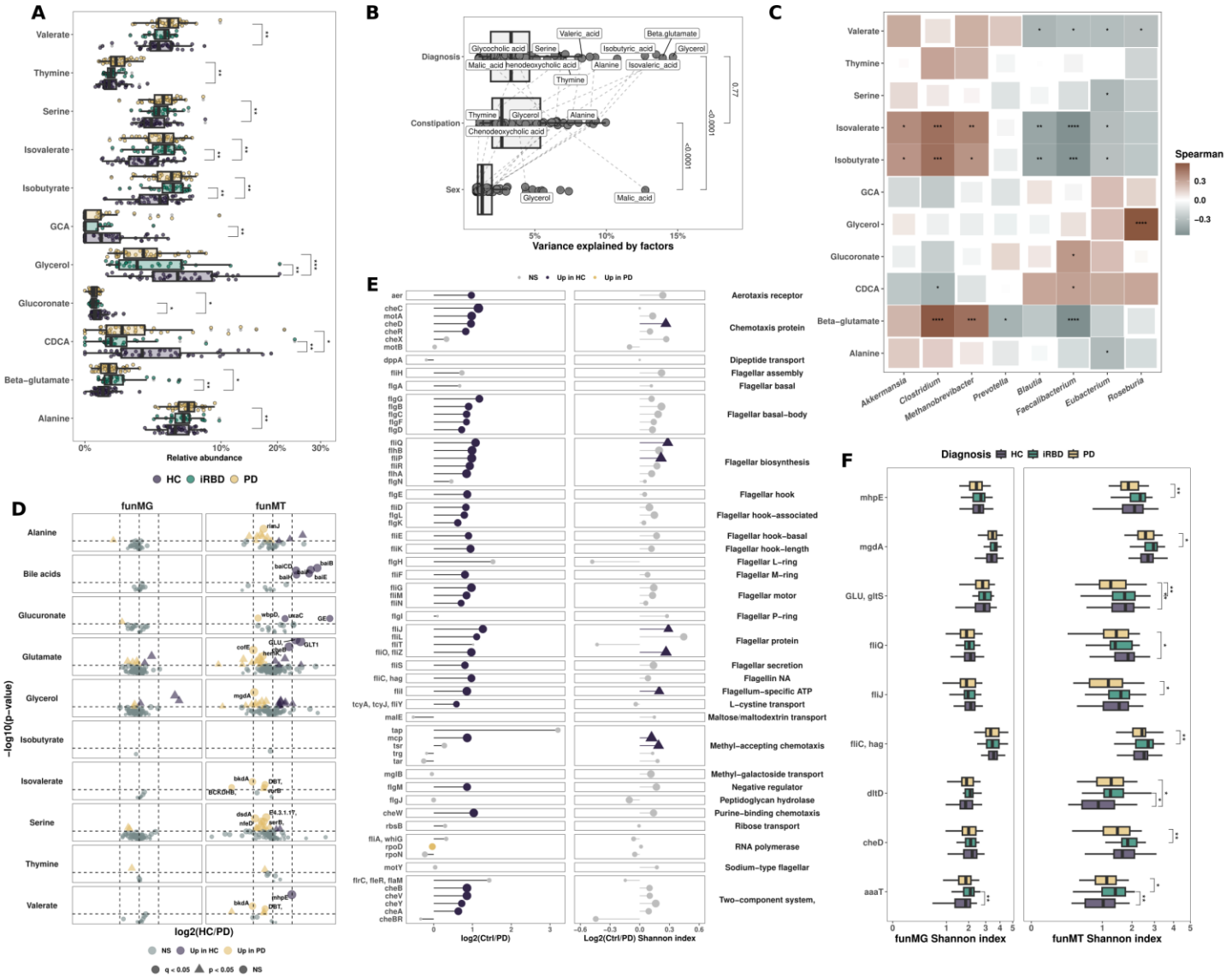


Figure 4. Flagellar assembly and chemotaxis gene expressions according to taxonomy. Top 50 genera expressing flagellar assembly and/or chemotaxis related genes are depicted by the mean of transcripts per million per disease status. Values are depicted for each genus and genes on the left panel and the sum of all genes per genus on the right panel with a square root transformation. Genera are clustered based on the log2FC(HC/PD) for each genes using a Canberra distance and *hclust()* using "ward.D2" method.

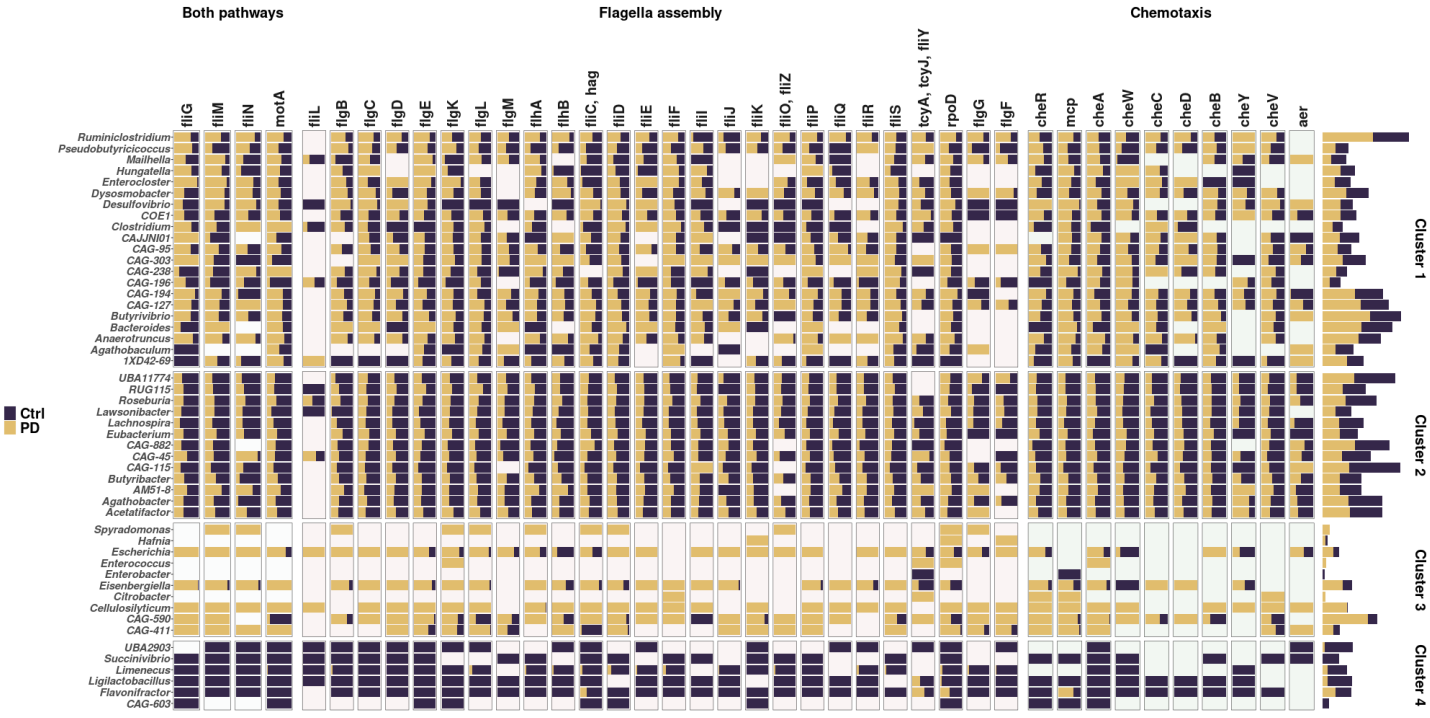


Figure 5. Metabolic network of whole community interactions. A. Metabolic network of whole community interactions, with KEGG KOs represented as nodes and associated metabolites as edges. Node sizes reflect MT/MG ratio of normalized read counts for each KO. Highlighted is the overlap between glutamate-, thymine-, glycerol-, serine-, alanine- and glucuronate-associated subnetworks mapped on the whole community network. **B.** Betweenness Centrality calculated for the key metabolites highlighted in the whole-community network based on genes as nodes. **C.** The network was inverted to calculate Betweenness centrality for metabolites, here metabolites are nodes and genes are edges.

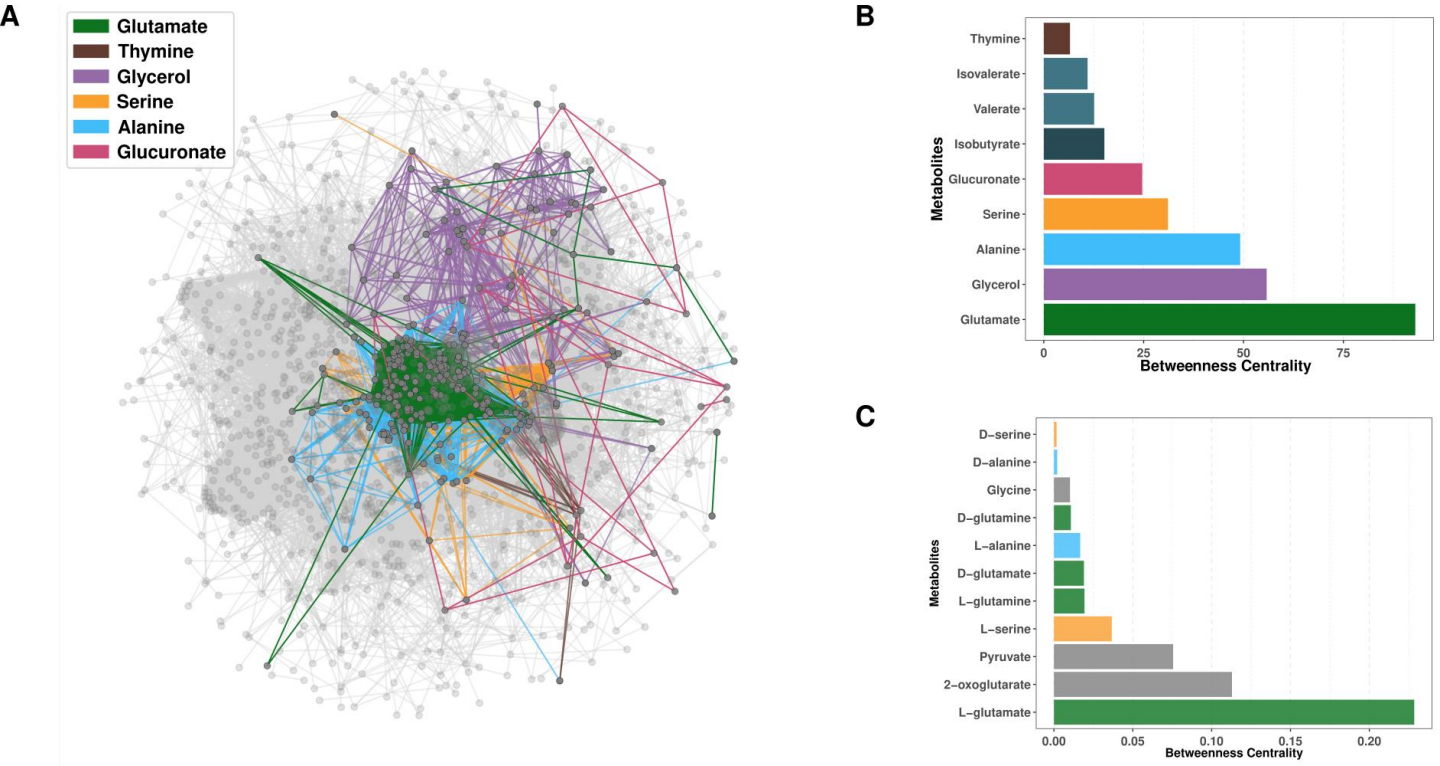
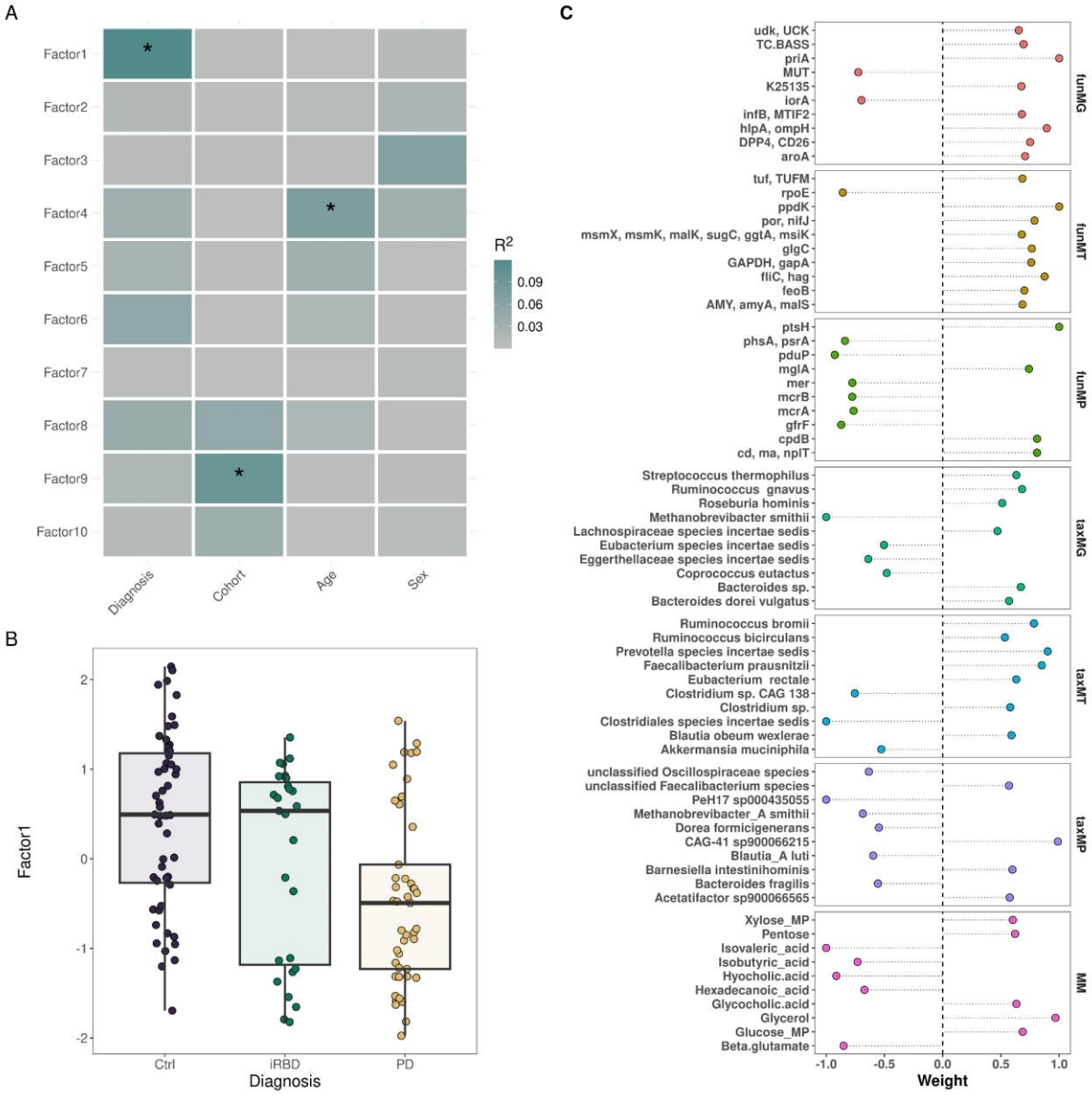
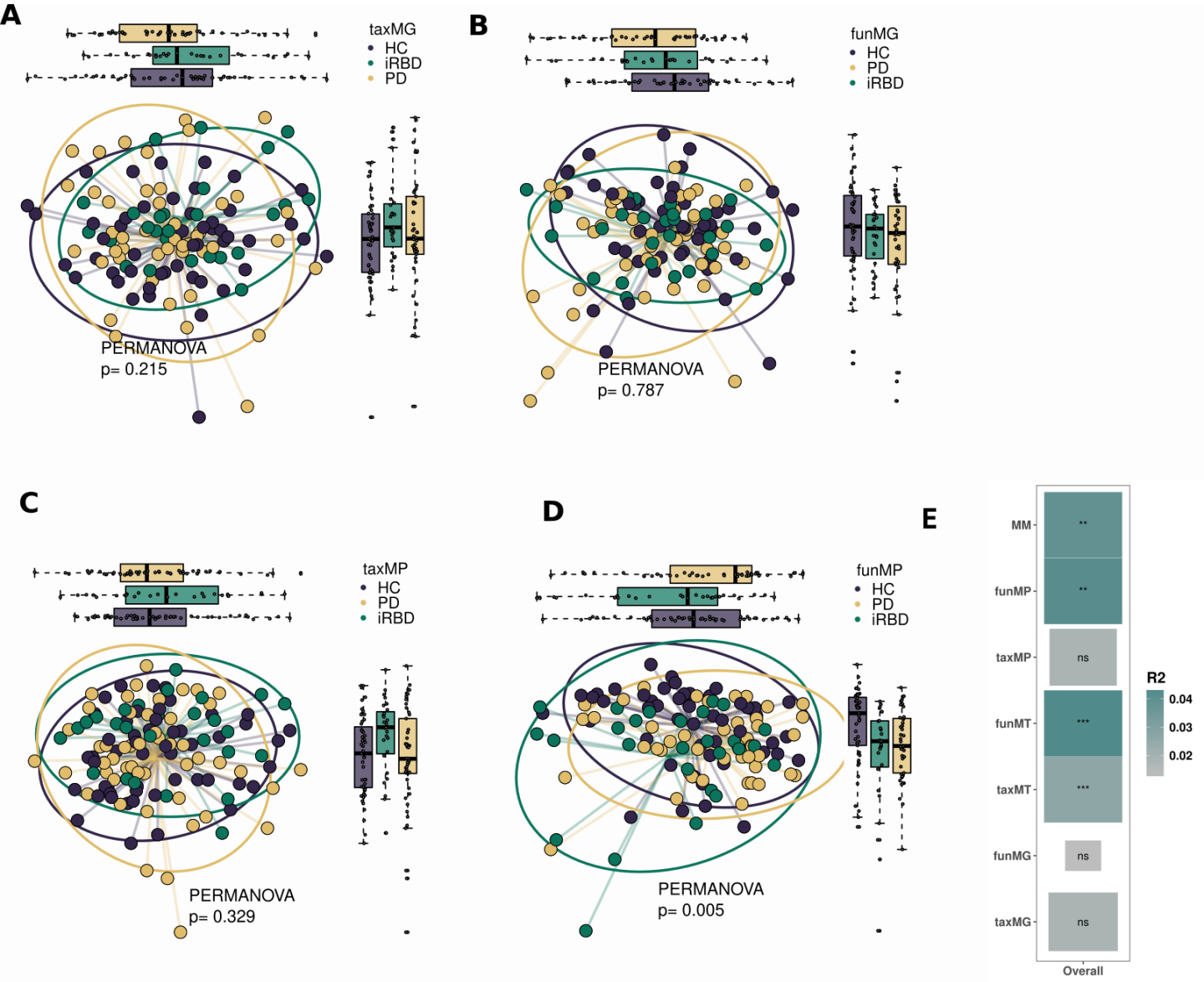


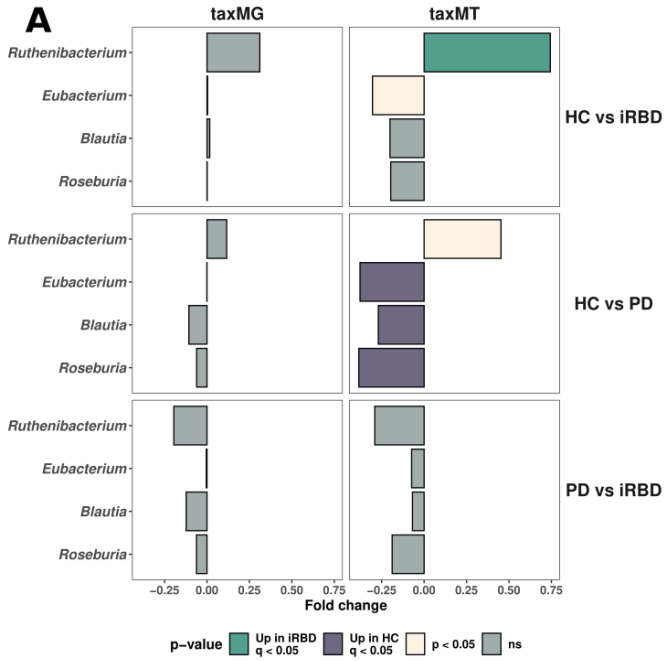
Figure 6 MOFA analysis validates the findings of per omics layer analyses. A. Associations of MOFA factors with the diagnosis and confounders, the colour of rectangles represents partial R^2 values, significant associations (FDR-adjusted p-value < 0.05) marked with an asterisk. **B**. Abundance of the Factor 1 in studied groups. **C.** Min-max scaled weights of top 10 features per omics layer contributing to Factor 1. The sign of the weight indicates the direction of the effect, the abundance of features with positive weights is positively associated with the Factor 1 level, and the abundance of features with negative weights is negatively associated with the Factor 1 level.

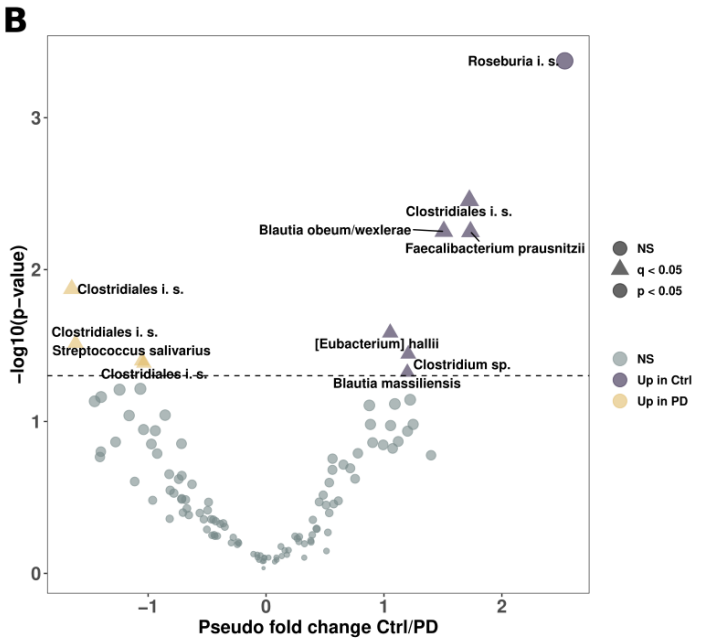


Extended figure 1. Beta diversity analysis of remaining omics. NMDS analysis of **A.** metagenomic taxonomic composition, **B.** metagenomic functions and **C.** meta-proteomic taxonomic composition, using a Bray-Curtis dissimilarity matrix. **D.** PCA analysis of metaproteomic functions. **E.** PERMANOVA analysis for the three groups and all omics. Colour represents R^2 values and size is –log10(p-value). All PERMANOVA analysis were run using 1000 permutations using a Bray-Curtis dissimilarity matrix.

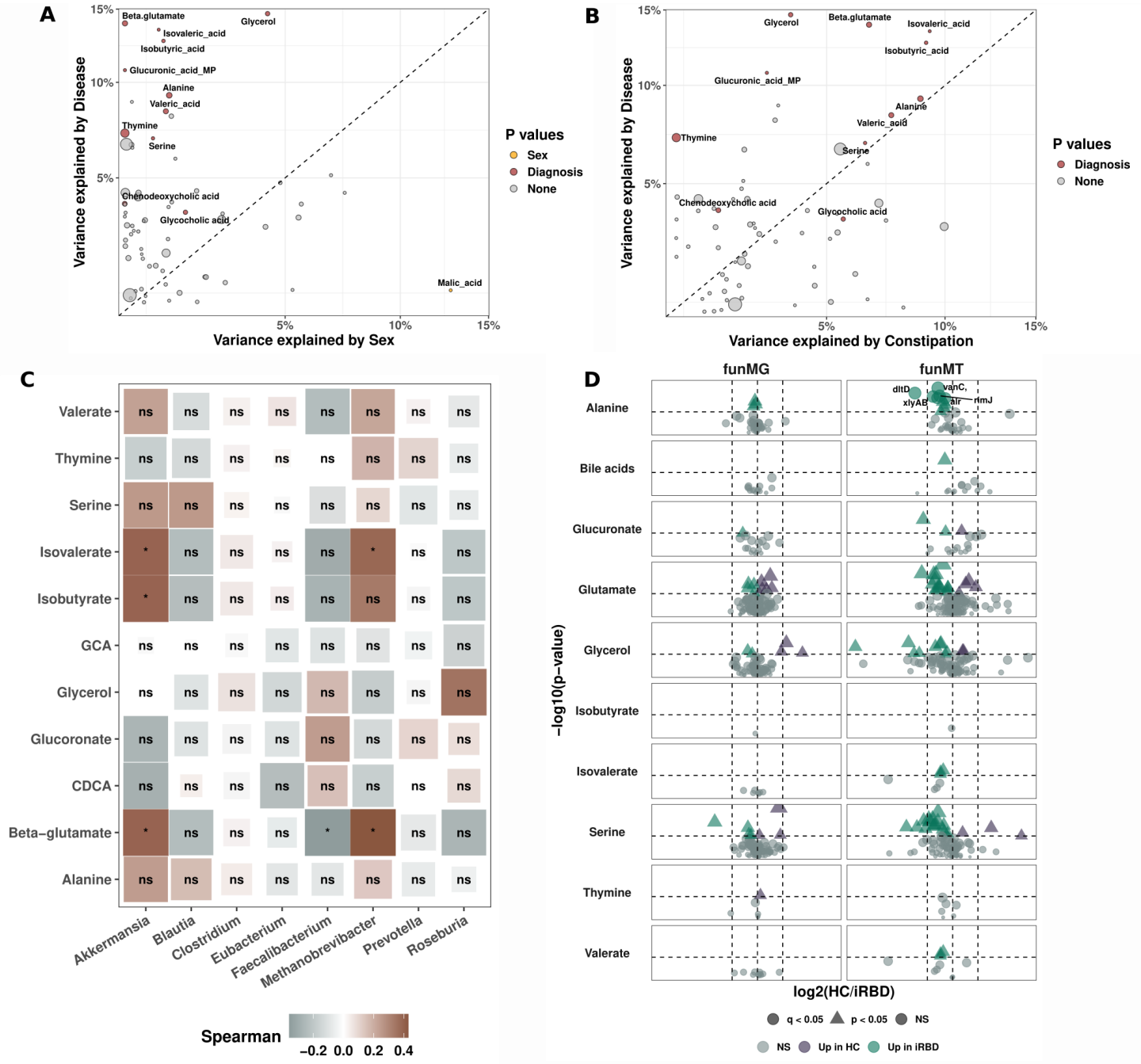


Extended figure 2. A. Differential abundance analysis at the genus level using SIAMCAT algorithm. **B.** Differential abundance analysis using ALDEx2 algorithm at the species level. Values are pseudo fold changes for HC/PD and size is based on –log10(p-value). Shape is referring to level of significance, triangular shape for p-value significance before and round shape for p-values < 0.05 after FDR correction.

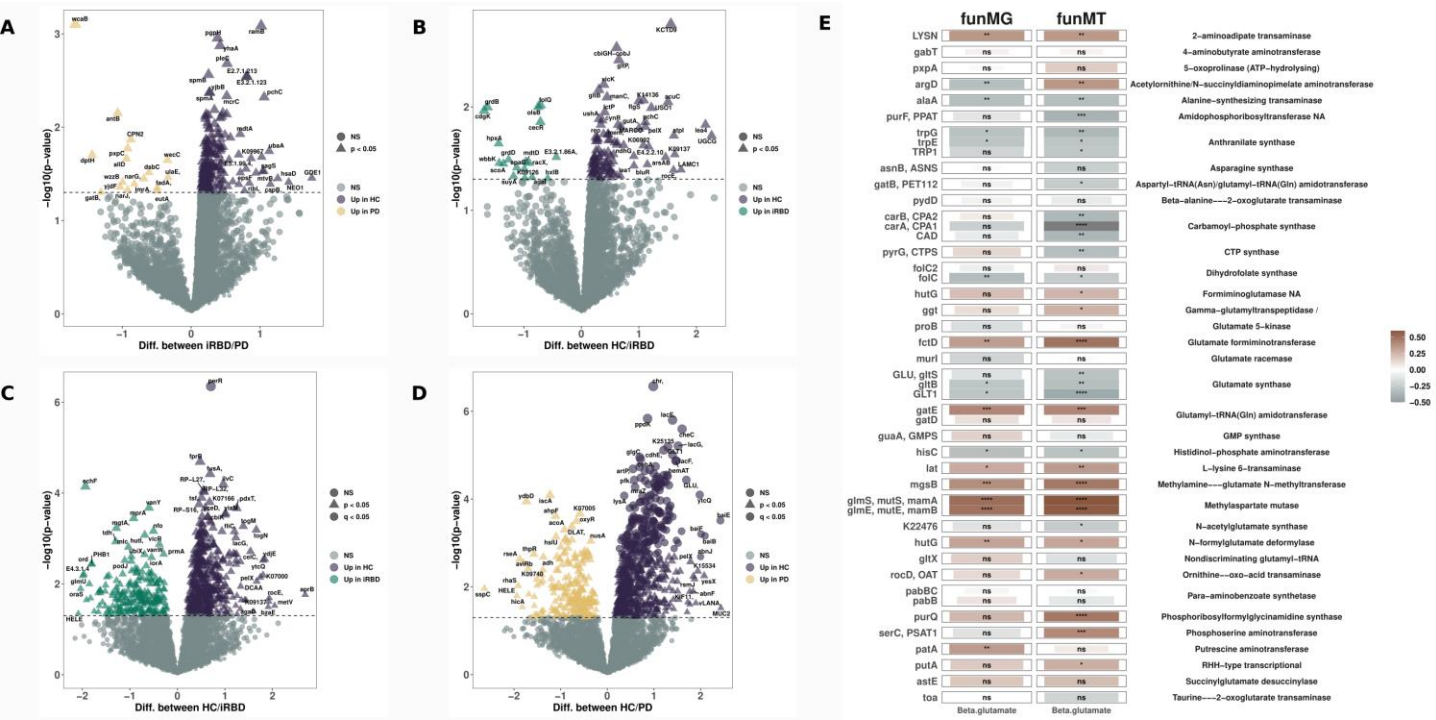




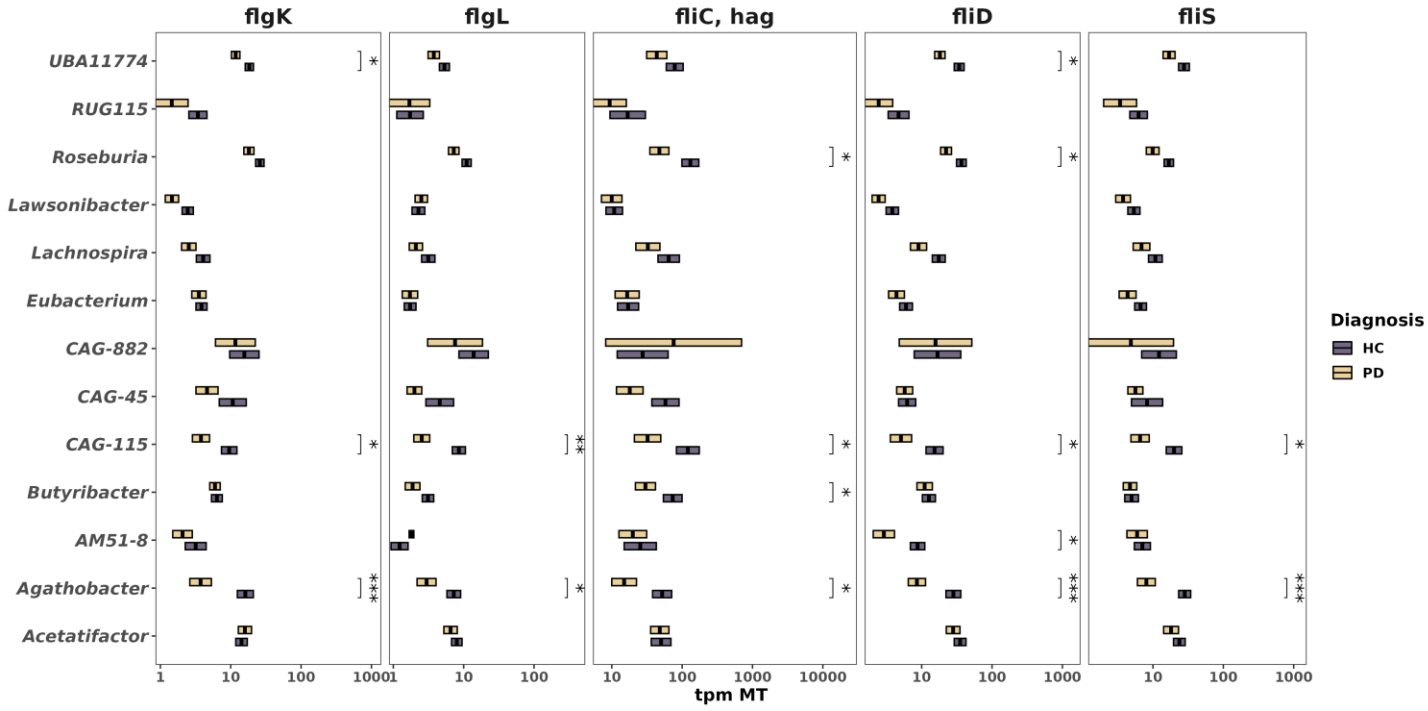
Extended figure 3. A. Percentage of variance explained for each metabolite for "Sex" and "Diagnosis". **B.** Percentage of variance explained for each metabolite for "Constipation" and "Diagnosis". Metabolites are including untargeted meta-metabolomics, targeted SCFA and targeted bile acids, normalized by sum before merging and variance quantification. **C.** Spearman correlation between metabolites and taxMG. P-values are FDR corrected. **D.** Absolute log2 fold change between HC and iRBD for funMG and funMT associated to significant compounds. Dots are scaled by the –log10(p-value), colorized and shaped according to p-value significance before (triangle shape) and after FDR correction (round shape).



Extended figure 4. A to D. ALDEx2 differential abundance analysis on funMG for HC vs PD (**A.**) and HC vs iRBD (**B.**); funMT for HC vs PD (**C.**) and HC vs iRBD (**D.**). All genes and transcripts are colorized and shaped according to p-value significance before (triangle shape) and after FDR correction (round shape) **E.** Spearman correlation between beta-glutamate relative abundance and funMG-funMT KEGG orthologs related to glutamate species. Only genes with at least one significant correlation are plotted. All p-values are FDR corrected.



Extended figure 5. Flagellar assembly transcripts encoding for extracellular component of the flagella for the genus present in Cluster 2. All tests are Wilcoxon tests.



Extended figure 6. Multiomics variance explained by MOFA factors. **A**. Variance explained by the MOFA factors across different omics layers, total. **B**. Variance explained by the MOFA factors across different omics layers, splitted by factors.

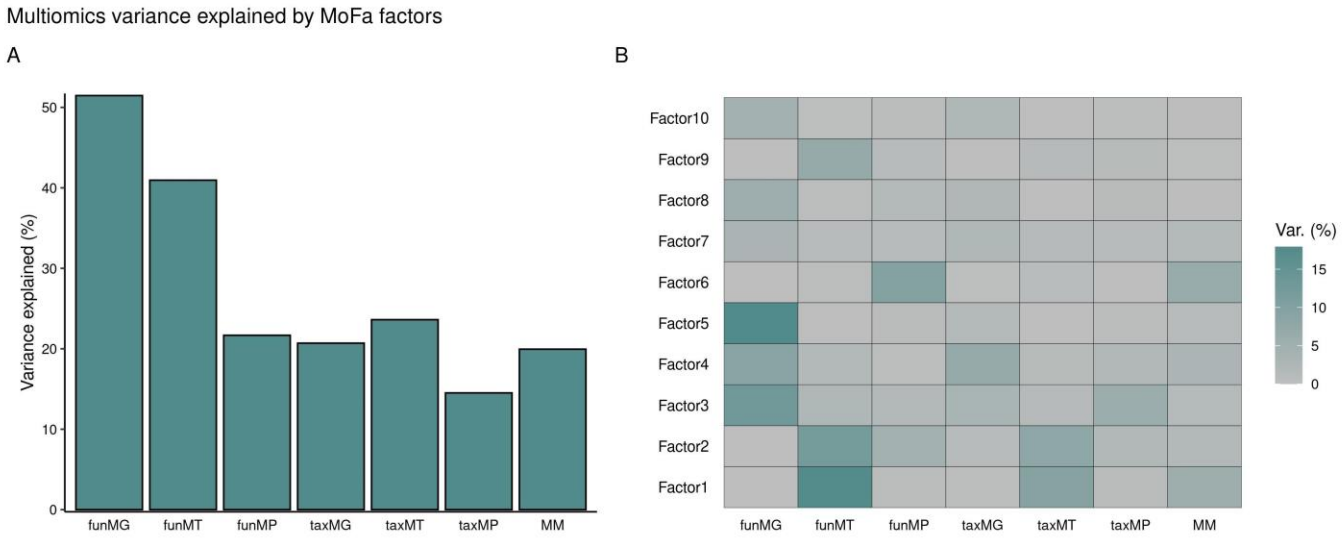


Table 1

Variable[1]	Ctrl	iRBD	PD	p-value[2]
n	49	27	46	
Sex (f / m)	26 / 23	4 / 23	19 / 27	0.004
Constipation	26 / 3 / 20	13 / 14 / 0	21 / 22 / 3	<0.001
Smoking	28 / 1 / 20	12 / 1 / 14	34 / 2 / 10	0.821
Diabetes	24 / 5 / 20	25 / 2 / 0	43 / 3 / 0	0.328
Metformin medication	27 / 2 / 20	25 / 2 / 0	42 / 2 / 2	0.878
Antidepressant medication	26 / 3 / 20	23 / 4 / 0	43 / 1 / 2	0.115
Statin medication	27 / 2 / 20	22 / 5 / 0	39 / 5 / 2	0.416
PPI medication	27 / 2 / 20	13 / 1 / 13	37 / 7 / 2	0.534
Levodopa medication		13 / 1 / 13	12 / 32 / 2	<0.001
Agonist medication		13 / 1 / 13	18 / 26 / 2	<0.001
Entacapone medication		27 / 0 / 0	39 / 5 / 2	0.056
Age at sampling (years)	68.1 ± 6.15	65.96 ± 7.83	65.8 ± 9.85	0.327
Disease duration (months)		76.37 ± 56.3	63.46 ± 42.52	0.271
PD-NMS sum	4.52 ± 3.05	8.31 ± 3.53	7.83 ± 3.65	<0.001
UPDRS I-III sum	3.31 ± 4.26	6.08 ± 4.94	31.47 ± 17	<0.001
Scopa-AUT sum	8.76 ± 6.35	10.17 ± 4.91	13.09 ± 6.21	0.018
Sniffin Sticks Identification	12.03 ± 2.03	7.19 ± 3.88	6.67 ± 3.88	<0.001
Hoehn and Yahr stage	0 ± 0	0 ± 0	2.01 ± 0.77	<0.001

[1] For categorical variables other than sex, values given as “no / yes / not available”; for continuous variables, values given as mean ± standard deviation.

[2] Categorical variables: Fisher’s exact test (missing data excluded); continuous variables: one-way ANOVA.

Ctrl = control subject, PD = Parkinson's disease patient, iRBD = patient with idiopathic REM sleep behaviour disorder, PPI = proton pump inhibitor, PD-NMS = Non-Motor Symptoms questionnaire, UPDRS = Unified Parkinson’s Disease Rating Scale, Scopa-AUT = Scales for Outcomes in Parkinson’s Disease - Autonomic Dysfunction.

Extended information

Multi-omics data overview

Using our previously developed methodological framework[84,85], we performed a systematic multi-omic analysis of DNA, RNA, protein, and metabolite fractions isolated from flash-frozen fecal samples. We used MG, MT, MP and MM data to find biomarkers associated with the PD phenotype (Fig. 1). We generated a mean of 7.5 (std 1.7) Gbps and 7.5 (std 1.4) Gbps of sequencing data for MG and MT, respectively. After trimming and filtering, we retained a mean of 6.8 (std 1.7) Gbps and 3.2 (std 1.3) Gbps for MG and MT, respectively. The mean assembly size was 0.4 (std 0.1) Gbps, with on average $5.9x10^5$ (std $1.7x10^5$) genes predicted. Finally, protein databases contained a mean of $7.2x10^5$ (std $1.8x10^5$) proteins, an average of $4.1x10^4$ (std $0.6x10^4$) MS spectra per sample were acquired, and a mean of $3.4x10^3$ (std $1.7x10^3$) proteins were identified.

MOFA model description

MOFA is an unsupervised machine learning approach for the integration of multi-omics data sets[79]. It allows for the identification of highly informative features across multiple omics. It has previously been used in the study of the gut microbiome in several diseases, giving critical insights into the link between the gut microbiome, health, and disease[81–83]. The biggest proportion of variance was explained by funMG and funMT, followed by the taxMT and funMP datasets (Extended fig. Aa). F1-2 incorporated most of the variance related to the funMT and taxMT, whereas funMG and taxMG variance was predominantly covered by F3-F5 (Extended fig. 8B). The funMP variance was explained mostly by F6, and MM variance was explained by F1 and F6. MOFA factors were tested in a linear model followed by ANOVA with disease status, as well as confounders including patients' sex, age, and recruitment cohort.

References

1. Elbaz, A., Carcaillon, L., Kab, S. & Moisan, F. Epidemiology of Parkinson's disease. *Revue Neurologique* **172**, 14–26 (2016).
2. Schwiertz, A. *et al.* Fecal markers of intestinal inflammation and intestinal permeability are elevated in Parkinson's disease. *Parkinsonism & Related Disorders* **50**, 104–107 (2018).
3. Clairembault, T. *et al.* Structural alterations of the intestinal epithelial barrier in Parkinson's disease. *Acta Neuropathologica Communications* **3**, 12 (2015).
4. Scd, van Ij. & P, D. The Intestinal Barrier in Parkinson's Disease: Current State of Knowledge. *Journal of Parkinson's disease* **9**, (2019).
5. Palacios, N. *et al.* Metagenomics of the Gut Microbiome in Parkinson's Disease: Prodromal Changes. *Annals of Neurology* **94**, 486–501 (2023).
6. Aho, V. T. E. *et al.* Relationships of gut microbiota, short-chain fatty acids, inflammation, and the gut barrier in Parkinson's disease. *Mol Neurodegeneration* **16**, 6 (2021).
7. Heintz-Buschart, A. *et al.* The nasal and gut microbiome in Parkinson's disease and idiopathic rapid eye movement sleep behavior disorder. *Movement Disorders* **33**, 88–98 (2018).
8. Cirstea, M. S. *et al.* Microbiota Composition and Metabolism Are Associated With Gut Function in Parkinson's Disease. *Mov Disord* **35**, 1208–1217 (2020).
9. Boktor, J. C. *et al.* Integrated Multi-Cohort Analysis of the Parkinson's Disease Gut Metagenome. *Movement Disorders* **38**, 399–409 (2023).
10. Huang, B. *et al.* Gut microbiome dysbiosis across early Parkinson's disease, REM sleep behavior disorder and their first-degree relatives. *Nat Commun* **14**, 2501 (2023).
11. Wallen, Z. D. *et al.* Metagenomics of Parkinson's disease implicates the gut microbiome in multiple disease mechanisms. *Nat Commun* **13**, 6958 (2022).
12. Romano, S. *et al.* Meta-analysis of the Parkinson's disease gut microbiome suggests alterations linked to intestinal inflammation. *npj Parkinsons Dis.* **7**, 1–13 (2021).
13. Toh, T. S. *et al.* Gut microbiome in Parkinson's disease: New insights from meta-analysis. *Parkinsonism & Related Disorders* **94**, 1–9 (2022).

14. Howell, M. J. & Schenck, C. H. Rapid Eye Movement Sleep Behavior Disorder and Neurodegenerative Disease. *JAMA Neurology* **72**, 707–712 (2015).
15. Unger, M. M. *et al.* Short chain fatty acids and gut microbiota differ between patients with Parkinson's disease and age-matched controls. *Parkinsonism & Related Disorders* **32**, 66–72 (2016).
16. Chen, S.-J. & Lin, C.-H. Gut microenvironmental changes as a potential trigger in Parkinson's disease through the gut–brain axis. *J Biomed Sci* **29**, 54 (2022).
17. Hirayama, M., Tsunoda, M., Yamamoto, M., Tsuda, T. & Ohno, K. Serum Tyrosine-to-Phenylalanine Ratio is Low in Parkinson's Disease. *Journal of Parkinson's Disease* **6**, 423–431 (2016).
18. Figura, M. *et al.* Serum amino acid profile in patients with Parkinson's disease. *PLOS ONE* **13**, e0191670 (2018).
19. Vascellari, S. *et al.* Gut Microbiota and Metabolome Alterations Associated with Parkinson's Disease. *mSystems* **5**, 10.1128/msystems.00561-20 (2020).
20. Li, P. *et al.* Gut Microbiota Dysbiosis Is Associated with Elevated Bile Acids in Parkinson's Disease. *Metabolites* **11**, 29 (2021).
21. Hertel, J. *et al.* Integrated Analyses of Microbiome and Longitudinal Metabolome Data Reveal Microbial-Host Interactions on Sulfur Metabolism in Parkinson's Disease. *Cell Reports* **29**, 1767-1777.e8 (2019).
22. Hang, S. *et al.* Bile acid metabolites control T H 17 and T reg cell differentiation. *Nature* **576**, 143–148 (2019).
23. Song, X. *et al.* Microbial bile acid metabolites modulate gut RORγ + regulatory T cell homeostasis. *Nature* **577**, 410–415 (2020).
24. Iovino, L., Tremblay, M. E. & Civiero, L. Glutamate-induced excitotoxicity in Parkinson's disease: The role of glial cells. *Journal of Pharmacological Sciences* **144**, 151–164 (2020).
25. Iwasaki, Y., Ikeda, K., Shiojima, T. & Kinoshita, M. Increased plasma concentrations of aspartate, glutamate and glycine in Parkinson's disease. *Neurosci Lett* **145**, 175–177 (1992).

643 26. Jiménez-Jiménez, F. J. *et al.* Neurotransmitter amino acids in cerebrospinal fluid of patients with
644 Parkinson's disease. *J Neurol Sci* **141**, 39–44 (1996).
645 27. Mally, J., Szalai, G. & Stone, T. W. Changes in the concentration of amino acids in serum and
646 cerebrospinal fluid of patients with Parkinson's disease. *J Neurol Sci* **151**, 159–162 (1997).
647 28. Yuan, Y.-S. *et al.* Change in Plasma Levels of Amino Acid Neurotransmitters and its Correlation
648 with Clinical Heterogeneity in Early Parkinson's Disease Patients. *CNS Neuroscience &*
649 *Therapeutics* **19**, 889–896 (2013).
650 29. Wang, W. *et al.* Interactions between gut microbiota and Parkinson's disease: The role of
651 microbiota-derived amino acid metabolism. *Front. Aging Neurosci.* **14**, 976316 (2022).
652 30. Hughes, A. J., Daniel, S. E., Kilford, L. & Lees, A. J. Accuracy of clinical diagnosis of idiopathic
653 Parkinson's disease: a clinico-pathological study of 100 cases. *Journal of Neurology, Neurosurgery*
654 *& Psychiatry* **55**, 181–184 (1992).
655 31. Beheshti, I., Booth, S. & Ko, J. H. Differences in brain aging between sexes in Parkinson's disease.
656 *npj Parkinsons Dis.* **10**, 1–16 (2024).
657 32. Li, X. *et al.* Sex differences in rapid eye movement sleep behavior disorder: A systematic review
658 and meta-analysis. *Sleep Med Rev* **71**, 101810 (2023).
659 33. Mertsalmi, T. H. *et al.* More than constipation – bowel symptoms in Parkinson's disease and their
660 connection to gut microbiota. *European Journal of Neurology* **24**, 1375–1383 (2017).
661 34. Roume, H. *et al.* Comparative integrated omics: identification of key functionalities in microbial
662 community-wide metabolic networks. *npj Biofilms Microbiomes* **1**, 1–11 (2015).
663 35. Villette, R. *et al.* Unraveling Host-Gut Microbiota Dialogue and Its Impact on Cholesterol Levels.
664 *Front. Pharmacol.* **11**, (2020).
665 36. Hurley, M. J., Bates, R., Macnaughtan, J. & Schapira, A. H. V. Bile acids and neurological disease.
666 *Pharmacology & Therapeutics* **240**, 108311 (2022).
667 37. Kalecký, K. & Bottiglieri, T. Targeted metabolomic analysis in Parkinson's disease brain frontal
668 cortex and putamen with relation to cognitive impairment. *npj Parkinsons Dis.* **9**, 1–16 (2023).

669 38. Chen, S.-J. *et al.* Association of Fecal and Plasma Levels of Short-Chain Fatty Acids With Gut
670 Microbiota and Clinical Severity in Patients With Parkinson Disease. *Neurology* **98**, e848–e858
671 (2022).
672 39. Baert, F. *et al.* Parkinson's disease patients' short chain fatty acids production capacity after in vitro
673 fecal fiber fermentation. *npj Parkinsons Dis.* **7**, 1–14 (2021).
674 40. Tan, S. *et al.* Interaction between the gut microbiota and colonic enteroendocrine cells regulates
675 host metabolism. *Nat Metab* **6**, 1076–1091 (2024).
676 41. Qian, F. *et al. Salmonella* flagellin is a potent carrier–adjuvant for peptide conjugate to induce
677 peptide-specific antibody response in mice. *Vaccine* **33**, 2038–2044 (2015).
678 42. Gram, A. M. *et al.* Salmonella Flagellin Activates NAIP/NLRC4 and Canonical NLRP3
679 Inflammasomes in Human Macrophages. *J Immunol* **206**, 631–640 (2021).
680 43. Tran, H. Q., Ley, R. E., Gewirtz, A. T. & Chassaing, B. Flagellin-elicited adaptive immunity
681 suppresses flagellated microbiota and vaccinates against chronic inflammatory diseases. *Nat*
682 *Commun* **10**, 5650 (2019).
683 44. Cullender, T. C. *et al.* Innate and Adaptive Immunity Interact to Quench Microbiome Flagellar
684 Motility in the Gut. *Cell Host & Microbe* **14**, 571–581 (2013).
685 45. Clasen, S. J. *et al.* Silent recognition of flagellins from human gut commensal bacteria by Toll-like
686 receptor 5. *Science Immunology* **8**, eabq7001 (2023).
687 46. Quan, Y. *et al. Roseburia intestinalis*-derived flagellin is a negative regulator of intestinal
688 inflammation. *Biochemical and Biophysical Research Communications* **501**, 791–799 (2018).
689 47. Shen, Z. *et al.* Roseburia intestinalis stimulates TLR5-dependent intestinal immunity against
690 Crohn's disease. *eBioMedicine* **85**, (2022).
691 48. Wu, X. *et al.* Roseburia intestinalis-derived flagellin ameliorates colitis by targeting
692 miR-223-3p-mediated activation of NLRP3 inflammasome and pyroptosis. *Molecular Medicine*
693 *Reports* **22**, 2695–2704 (2020).
694 49. Liu, J. *et al.* Microbiota-microglia crosstalk between *Blautia producta* and neuroinflammation of
695 Parkinson's disease: A bench-to-bedside translational approach. *Brain, Behavior, and Immunity*
696 **117**, 270–282 (2024).

50. Tian, J. *et al.* Specific immune status in Parkinson's disease at different ages of onset. *NPJ Parkinsons Dis* **8**, 5 (2022).
51. Tansey, M. G. *et al.* Inflammation and immune dysfunction in Parkinson disease. *Nat Rev Immunol* **22**, 657–673 (2022).
52. Ho, M. S. Microglia in Parkinson's Disease. *Adv Exp Med Biol* **1175**, 335–353 (2019).
53. Scheiblich, H. *et al.* Microglial NLRP3 Inflammasome Activation upon TLR2 and TLR5 Ligation by Distinct α-Synuclein Assemblies. *The Journal of Immunology* **207**, 2143–2154 (2021).
54. Mollenhauer, B. *et al.* Nonmotor and diagnostic findings in subjects with de novo Parkinson disease of the DeNoPa cohort. *Neurology* **81**, 1226–1234 (2013).
55. Mollenhauer, B. *et al.* Monitoring of 30 marker candidates in early Parkinson disease as progression markers. *Neurology* **87**, 168–177 (2016).
56. Schenck, C. H. *et al.* Rapid eye movement sleep behavior disorder: devising controlled active treatment studies for symptomatic and neuroprotective therapy—a consensus statement from the International Rapid Eye Movement Sleep Behavior Disorder Study Group. *Sleep Medicine* **14**, 795–806 (2013).
57. Sateia, M. J. International Classification of Sleep Disorders-Third Edition. *Chest* **146**, 1387–1394 (2014).
58. Visser, M., Marinus, J., Stiggelbout, A. M. & Van Hilten, J. J. Assessment of autonomic dysfunction in Parkinson's disease: The SCOPA-AUT. *Movement Disorders* **19**, 1306–1312 (2004).
59. Chaudhuri, K. R. *et al.* The metric properties of a novel non-motor symptoms scale for Parkinson's disease: Results from an international pilot study. *Movement Disorders* **22**, 1901–1911 (2007).
60. Fahn, S., Elton, R. & Members of the UPDRS Development Committee. Unified Parkinson's disease rating scale. *Recent developments in Parkinson's disease* **0**, 153–163 (1987).
61. Stiasny-Kolster, K. *et al.* The REM sleep behavior disorder screening questionnaire—A new diagnostic instrument. https://movementdisorders.onlinelibrary.wiley.com/doi/abs/10.1002/mds.21740 (2007).

62. Roume, H. *et al.* A biomolecular isolation framework for eco-systems biology. *ISME J* **7**, 110–121 (2013).
63. Hughes, C. S. *et al.* Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat Protoc* **14**, 68–85 (2019).
64. Bannuscher, A. *et al.* A multi-omics approach reveals mechanisms of nanomaterial toxicity and structure–activity relationships in alveolar macrophages. *Nanotoxicology* **14**, 181–195 (2020).
65. De Saedeleer, B. *et al.* Systematic characterization of human gut microbiome-secreted molecules by integrated multi-omics. *ISME COMMUN.* **1**, 1–6 (2021).
66. Hiller, K. *et al.* MetaboliteDetector: Comprehensive Analysis Tool for Targeted and Nontargeted GC/MS Based Metabolome Analysis. *Anal. Chem.* **81**, 3429–3439 (2009).
67. Narayanasamy, S. *et al.* IMP: a pipeline for reproducible reference-independent integrated metagenomic and metatranscriptomic analyses. *Genome Biology* **17**, 260 (2016).
68. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
69. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
70. Queirós, P., Delogu, F., Hickl, O., May, P. & Wilmes, P. Mantis: flexible and consensus-driven genome annotation. *GigaScience* **10**, giab042 (2021).
71. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
72. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biology* **20**, 257 (2019).
73. Milanese, A. *et al.* Microbial abundance, activity and population genomic profiling with mOTUs2. *Nat Commun* **10**, 1014 (2019).
74. Guo, X. *et al.* Sipros Ensemble improves database searching and filtering for complex metaproteomics. *Bioinformatics* **34**, 795–802 (2018).
75. Wirbel, J. *et al.* Microbiome meta-analysis and cross-disease comparison enabled by the SIAMCAT machine learning toolbox. *Genome Biol* **22**, 1–27 (2021).

76. Fernandes, A. D., Macklaim, J. M., Linn, T. G., Reid, G. & Gloor, G. B. ANOVA-Like Differential Expression (ALDEx) Analysis for Mixed Population RNA-Seq. *PLOS ONE* **8**, e67019 (2013).
77. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57**, 289–300 (1995).
78. Hagberg, A. A., Schult, D. A. & Swart, P. J. Exploring Network Structure, Dynamics, and Function using NetworkX. (2008).
79. Argelaguet, R. *et al.* Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets | Molecular Systems Biology. *EMBOpress* https://www.embopress.org/doi/full/10.15252/msb.20178124 (2018).
80. Varrette, S., Bouvry, P., Cartiaux, H. & Georgatos, F. Management of an academic HPC cluster: The UL experience. in *2014 International Conference on High Performance Computing & Simulation (HPCS)* 959–967 (2014). doi:10.1109/HPCSim.2014.6903792.
81. Haak, B. W. *et al.* Integrative Transkingdom Analysis of the Gut Microbiome in Antibiotic Perturbation and Critical Illness. *mSystems* **6**, 10.1128/msystems.01148-20 (2021).
82. Al Bataineh, M. T. *et al.* Uncovering the relationship between gut microbial dysbiosis, metabolomics, and dietary intake in type 2 diabetes mellitus and in healthy volunteers: a multi-omics analysis. *Sci Rep* **13**, 17943 (2023).
83. Arıkan, M. *et al.* Metaproteogenomic analysis of saliva samples from Parkinson's disease patients with cognitive impairment. *npj Biofilms Microbiomes* **9**, 1–10 (2023).
84. Wilmes, P., Muller, E., Heintz-Buschart, A., Roume, H. & Lebrun, L. The sequential isolation of metabolites, RNA, DNA, and proteins from a single, undivided mixed microbial community sample. Preprint at https://doi.org/10.1038/protex.2014.051 (2014).
85. Heintz-Buschart, A. *et al.* Integrated multi-omics of the human gut microbiome in a case study of familial type 1 diabetes. *Nat Microbiol* **2**, 1–13 (2016).

802 Among the MOFA factors, F1 showed an association with the disease status, whereas F4 and
803 F9 were associated with patients' sex and recruitment cohort, respectively (Fig. 4A).

Paper 2: Microbiome Expression Network is Dysregulated in Parkinson's Disease Individuals

Contribution

- As the shared first author of the paper, participated in the conceptualization of the study, data preparation, bioinformatical and statistical analysis, data interpretation, visualization of results, and manuscript writing. Specific contributions include gene co-expression network construction and analysis, as well as hub gene inference and analysis.
- Responsible for the creation of visual elements of the manuscript, including figures 1-4, 6, S1-2:
    - Fig. 1 – developed the data analysis workflow and visualized the sketch.
    - Fig. 2 – conducted gene co-expression network construction and module-trait association analysis, calculated network topology metrics for co-expressed gene modules.
    - Fig. 3 – prepared and curated data for gene set enrichment analysis for each co-expressed gene module.
    - Fig. 4 – conducted hub gene analysis and prepared data tables for corresponding visualizations (A, B, C); prepared and curated data for statistical analysis (D, E) and taxonomy association of hub genes involved in bacterial microcompartments (G, H).
    - Fig. 6 – conducted hub gene analysis and prepared data tables for corresponding visualizations.
    - Fig. S1-2 – calculated network topology metrics for co-expressed gene modules and prepared data tables for corresponding visualizations.

Overview of the study

In our study, titled "Microbiome Expression Network is Dysregulated in Parkinson's Disease Individuals," we aimed to further elucidate the connections between the gut microbiome and PD. Traditional bioinformatics and biostatistical methods often fall short in capturing the complexity of microbiome-host interactions. To address this limitation, we employed Weighted Gene Co-expression Network Analysis to investigate microbial co-expression patterns in individuals with PD. This approach revealed dysregulation in microbial networks, offering novel insights into how microbial factors contribute to the pathogenesis of this neurodegenerative disorder. By using co-expression networks, our study identifies key regulatory mechanisms in PD, uncovering altered biological processes beyond mere differential gene expression.

The human gut microbiome consists of a diverse array of microbial genes that play crucial roles in maintaining host health. In this study, we constructed a signed co-expression network of the gut microbiome from both HC and PD individuals using WGCNA. This network included 4,789 genes from an initial set of 11,876, based on the ratio between metatranscriptomic and metagenomic reads. We identified 17 modules, 4 of which were significantly associated with HC and 5 with PD. We hypothesize that the 8 modules not associated with disease status represent stable, core functions critical for preserving microbial network integrity in both health and disease.

Following the identification of gene modules, we conducted an in-depth analysis of the co-expression network's topological properties. We calculated key metrics such as centrality measures (degree, betweenness, and eigenvector centrality) and the clustering coefficient for each module to assess their structural roles within the network. Additionally, microbial diversity within each module was measured using the Shannon index to explore the relationship between gene co-expression and taxonomic diversity. Although no significant differences in overall network topology were observed between modules associated with HC, PD, or unassociated modules, some unassociated modules exhibited higher centrality and clustering values despite their smaller size. Notably, the diversity of taxa expression within a module was positively correlated with connectivity and size but negatively correlated with betweenness and eigenvector centrality. These findings reveal complex interactions between microbial diversity and network centrality, offering new insights into the gut microbiome's structural and functional organization in both health and disease.

We also performed gene set enrichment analysis to investigate the pathway composition of the modules. Although relatively few significantly enriched pathways were identified – since approximately 43% of gene orthologs within modules were not annotated to any KEGG pathway – some notable enrichments were observed. Specifically, modules associated with HC showed significant enrichment for flagellar assembly and secondary bile acid biosynthesis, while a module unassociated with either HC or disease exhibited significant enrichment for biofilm formation. Although not statistically significant after correction, we also observed enrichment in pathways related to glycerolipid metabolism, peptidoglycan biosynthesis, lipoic acid metabolism, and valine degradation in modules associated with PD.

We identified and analyzed hub genes within the network, defining them as the top 100 most connected genes. Most of these hub genes were found in HC and non-associated modules, with many involved in energy production (oxidative phosphorylation, glycolysis/gluconeogenesis) and transporter activity (ABC transporters). From the perspective of functional redundancy, we observed no differences between HC and PD individuals but, we observed greater overall diversity in gene diversity for genes within HC associated modules. Genes were categorized into those with increased expression linked to higher bacterial diversity and those associated with lower diversity. Notably, genes involved in flagellar assembly were significantly upregulated in HC and associated with higher microbial diversity. In contrast, 37% of genes significantly upregulated in PD exhibited reduced diversity, compared to 22% in HC. This trend was more pronounced among hub genes, with 75% of hub genes upregulated in PD showing decreased diversity. In this paper, we emphasize the role of flagellar assembly genes as hub genes within the microbial network, underscoring the significance of this pathway for the gut microbiome. Additionally, we discovered a notable association between genes involved in bacterial microcompartment (BMC) formation – an essential mechanism for detoxification and energy production in microbes, – modules associated with HC and flagellar assembly.

In conclusion, this study used WGCNA to investigate the gut microbiome's role in PD. Gene modules associated with HC showed greater functional diversity and significant enrichment in flagellar assembly and secondary bile acid biosynthesis, whereas biofilm formation was enriched in modules unassociated with disease status. In contrast, genes upregulated in PD exhibited reduced microbial diversity, particularly among hub genes, where 75% showed decreased diversity. The enrichment and upregulation of flagellar assembly genes, and their association with higher microbial diversity in HC and bacterial microcompartment-associated genes, align with findings from our previous study, "Integrated Multi-Omics

Highlights Alterations of Gut Microbiome Functions in Prodromal and Idiopathic PD.". This dysregulation of microbial co-expression patterns suggests altered biological processes in PD, which may contribute to disease pathogenesis. The upregulation of citrate lyase genes in PD indicates a possible link to intestinal inflammation. Although not statistically significant after correction, pathways related to glycerolipid metabolism, peptidoglycan biosynthesis, lipoic acid metabolism, and valine degradation were observed in PD modules, highlighting potential metabolic changes. These findings suggest that modulating microbial functions could be a promising therapeutic approach to address inflammation and slow disease progression in PD.

The gut microbiome gene expression network is dysregulated in individuals with Parkinson's disease

Polina V. Novikova[1]*, Rémy Villette[1]*, Cédric C. Laczny[1], Patrick May[1#], Paul Wilmes[1,2,3,#].

1. Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

2. Department of Life Sciences and Medicine, Faculty of Science, Technology and Medicine, University of Luxembourg, Belvaux, Luxembourg

3. Departments of Genetics and Medicine (Hematology, Blood and Marrow Transplantation), Stanford University, Stanford, USA

*,# equal contributions

Abstract

In this work, we employed WGCNA on data from an integrated multi-omics study involving a cross-sectional cohort of PD patients and healthy controls (HC). Our integrated multi-omics analysis, incorporating metagenomics (MG) and metatranscriptomics (MT), allowed us to identify significant shifts in gene co-expression patterns associated with PD. Key findings include the observation that the PD-linked gene network exhibits decreased gene diversity compared to HC. In contrast, HC gene modules were more central, highly connected, and enriched in functions such as flagellar assembly and secondary bile acid biosynthesis. Furthermore, hub gene analysis revealed that most hub genes, which play crucial roles in microbiome network stability, belonged to the HC-linked network, and particularly were involved in processes related to microcompartment assembly and flagellin. We found that genera including *Blautia* and *Anaerobutyricum* were the main contributors to microcompartment assembly genes significantly decreased in PD. Interestingly, PD-associated gene expression was linked to reduced alpha-diversity, suggesting that increased gene expression in PD corresponds to a less diverse microbial ecosystem. Conversely, in HC, higher expression was associated with greater diversity. These findings reinforce the concept of microbial dysbiosis in PD and reveal a disruption of gut metabolic function at both functional and taxonomic levels, potentially contributing to the progression of the disease. Crucially, our work highlights critical microbiome-wide taxonomic and functional gene expression network features which would need to be restored in future rewilding efforts directed at the gut microbiome in PD.

Introduction

Parkinson's disease (PD) is the second most prevalent neurodegenerative disorder, primarily caused by the loss of dopaminergic neurons and the formation of Lewy bodies in the brain. PD is characterized by both motor and non-motor symptoms, with non-motor symptoms such as dysphagia, constipation, and bloating being linked to the gastrointestinal tract. Notably, idiopathic constipation, a common symptom of PD, often precedes motor symptoms by over a decade (Fasano et al., 2015), supporting the hypothesis that the disease can originate in the gut (Braak et al., 2003). Intestinal dysbiosis has been documented in PD (Cryan et al., 2020; Keshavarzian et al., 2015), and evidence suggests that the gut might initiate or worsen the development of PD (Bhattarai et al., 2021; Hirayama & Ohno, 2021; Qian et al., 2018). These changes in gut microbiota composition can compromise gut permeability and the integrity of the intestinal barrier, affecting gastrointestinal epithelial cells, the immune system, and the enteric nervous system (Stolfi et al., 2022; Weiss & Hennet, 2017a). Moreover, a dysbiotic gut is characterized by decreased microbial richness and diversity (Weiss & Hennet, 2017b). In the last decade, multiple studies have found evidence for dysbiosis in the gut microbiome of individuals with PD characterized by shifts in bacterial and archaeal taxa including *Methanobrevibacter*, *Akkermansia* and *Roseburia* (Boertien et al., 2019; Romano et al., 2021; Toh et al., 2022). In addition, we recently showed a decrease in flagellar assembly and chemotaxis transcripts, along with differences in the metabolome in PD gut microbiome (Villette et al., 2024, submitted).

Traditional bioinformatic and biostatistics methods frequently fall short of capturing the full complexity of microbiome-host interactions. Indeed, differential expression analysis or multi-variate approaches doesn't capture the importance of functions within a complex ecosystem such as the gut microbiome. Disease-associated phenotypes are believed to result from disruptions across the entire network rather than from single, isolated gene mutations (Barabási et al., 2010). Gene co-expression networks have been widely used to discover

functional gene clusters and pathways associated with various disease phenotypes (Cai et al., 2023; Meng & Mei, 2019). Therefore, we explored the relationship between the microbiome gene abundance normalized gene expression and the host through a network-based approach. This effort aimed to inspect how molecular interactions within the gene network are dysregulated, providing a deeper understanding of the complex phenotypes associated with diseases such as PD. Weighted Gene Co-Expression Network Analysis (D module generally consists of genes that are involved in similar functional processes (Galán-Vásquez & Perez-Rueda, 2019). Centrality analysis is a powerful tool for identifying significant elements within a network, particularly in biological networks (Brandes & Erlebach, 2005; Koschützki & Schreiber, 2008). While various centrality metrics are available to characterize networks, it is advised to consider multiple measures to gain a comprehensive understanding of biological networks (Koschützki & Schreiber, 2004; Wuchty & Stadler, 2003). Such an approach ensures a more nuanced and accurate exploration of the network's key players.

In this work, we employ WGCNA based on the ratio of MT to MG reads for individuals with PD and HC. We found modules associated with both disease groups. Amongst the modules, we find significant enrichment in flagellar assembly and secondary bile acid biosynthesis pathways. In addition, we find that most of the identified hub genes belong to HC associated modules and especially in module M2. Interestingly, within this module we identify genes involved in bacterial microcompartment formation and catabolism, most of these genes had decreased expression in individuals with PD especially in the context of commensals such as *Blautia obeum* and *Anaerobutyricum hallii*. Finally, we show that the majority of genes with increased expression in individuals with PD is associated with a decrease in gene diversity, especially for hub genes. This highlights another side of microbiome dysbiosis.

Results

Microbial co-expression network is linked with disease status.

We constructed a network representation of the gut microbiome including samples from both HC (n=49) and PD individuals (n=46) using WGCNA (Langfelder and Horvath, 2008). From an original set of 11 876 microbial genes, we inferred a signed co-expression network of 4 789 genes after WGCNA trimming and processing. For the multi-omics co-expression analysis, we used abundance-normalized gene expression using the ratio of MT and MG transcripts per million (Figure 1) (Roume et al., 2015). The co-expression network revealed 17 modules, four significantly associated with HC (M13, M2, M11 and M17, $p < 0.05$), five significantly associated with PD (M3, M6, M7, M8 and M15, $p \leq 0.05$) and eight neither significantly associated with HC nor PD (Figure 2A).

We next looked at topological features of the network and calculated the diversity of genes found within each module using the Shannon index, coined module diversity. Modules M2 and M4 showed the highest mean connectivity, highest intramodular connectivity and highest module diversity despite their large size (Figure 2B and Supplemental fig. 1). M1 had the lowest clustering coefficient, degree centrality, eigenvector centrality and closeness centrality, which led us to believe that M1 was a module with only genes not clustering with the rest of the modules (Figure 2B and Supplemental fig. 1). M6 exhibited high values of betweenness and eigenvector centrality but low connectivity (Figure 2B and Supplemental fig. 1A). When analysing the modules based on the trait association, we found no statistically significant differences in topological measures but important variation between modules, irrespective of trait association (Kruskal and Wallis test, $p > 0.05$, Figure 2B and Supplemental fig. 1A). However, M1 being a clear outlier for clustering coefficient degree and closeness centrality, we removed it to re-compare modules based on trait association and

found a significantly higher degree and closeness centrality (Kruskal and Wallis, $p = 0.023$ and $p = 0.013$, data not shown). Finally, we noticed the correlation between module diversity and the different topology metrics such as sum of connectivity and intramodular connectivity ($R^2 > 0.8$, $p < 0.001$, Figure 2C) while eigenvector and betweenness centrality were anticorrelated with module diversity ($R < -0.8$, $p < 0.001$, Figure 2C). Overall, we found interesting anti-correlation between connectivity measures and closeness/betweenness centrality while clustering coefficient was correlated to degree/eigenvector centrality (Supplemental fig. 1C).

Modules associated with HC showed enrichment in flagellar assembly and secondary bile acid biosynthesis.

We next proceeded with gene set enrichment analysis (GSEA) to obtain insights into the pathway composition of the modules. Based on KEGG KO annotations, we found that on average 43.7% (min: 26%, max: 59%, Figure 3A) of genes within modules did not belong to any KEGG pathway. Modules comprised on average 28.5 different pathways ranging from 10 (M6) to 42 pathways (M1) (Supplemental fig. 2). We identified an enrichment in flagellar assembly in M13 and secondary bile acid biosynthesis in M11 for modules associated with HC, an enrichment in biofilm formation for M16 (no association) and no statistically significant enrichment in PD-associated modules (Figure 3, $q < 0.05$, $q < 0.05$ and $q > 0.05$, respectively, GSEA). Although not statistically significant after correction, we noticed enrichments in the following transformation within modules associated with PD: glycerolipid metabolism (M3), peptidoglycan biosynthesis (M15), lipoic acid metabolism and valine degradation (M7) (Figure 3B, $p < 0.01$). We also noticed the presence of beta-lactam resistance genes (*oppA, oppB, oppC, oppD* and *oppF*) and quorum sensing genes in PD-associated module M6 (Supplemental fig. 1). In addition, methane metabolism genes were present in the PD-associated modules M6 and M8 (Supplemental fig. 1).

Hub genes are restricted to HC-associated modules

We next defined hub genes to appreciate the key functions of the microbial co-expression network. We first selected the top 100 genes with highest connectivity from all modules associated with HC and PD, excluding genes from modules not associated with either of the groups. Out of the 100 genes, 95 were from HC-associated modules and 5 with PD associated modules (Figure 4A). To access more genes from PD-associated modules, we selected 10% of the top connected genes from each modules retrieving 125 for HC and 108 for PD (Figure 4A). For the top 100 genes most genes belonged to module M2 (85%) the rest belonging to M13 and M3 (Figure 4B). Hub genes associated with HC were mainly involved in energy production (oxydative phosphorylation, glycolysis/gluconeogenesis), transporter activity (ABC transporters), nucleotide metabolism (pyrimidine and purine metabolism), saccharide metabolism (pentose and glucuronate interconversions, starch and sucrose metabolism) and microbial motility (flagellar assembly and two-component system) (Figure 4C). With the 10% per module approach, we noticed the presence of two glutamate synthases (*GLT1* and *gltB*) in the module M11 labelled as members of the alanine, aspartate and glutamate metabolism (Figure 4C).

Amongst the hub genes, we found a significant increase in their expression in HC including flagellar assembly (*flgB, fliQ, fliS, flgE, fliK, flgL, fliD, fliP,* $p < 0.05$, Figure 4D). We also found a significant increased expression for citrate lyase genes in individuals with PD (*citD, citC* and *citF,* $p < 0.05$, $q > 0.05$, Figure 4D). We noticed that members of bacterial microcompartments (BMCs) catabolism and metabolism (eut and pdu operons) genes were significantly increased in HC (*eutM, pduE, eutK,* $q < 0.05$, DeSeq2, Figure 4D and E). These genes are known to form BMCs in the cytoplasm, and we further investigated the taxa responsible for these differential expressions. We firstly noticed the disparity in orthologs annotation between the pdu and eut operons, these annotations were dependent on the taxonomy, with most of shell proteins annotated as members of the eut operon, while catabolism/anabolism genes were annotated as members of the pdu operon. We manually

grouped the genes according to their described functions in the literature (Supplemental table 1). Interestingly, the two genera expressing the most ethanolamine and propanediol/propionaldehyde genes were *Blautia* and *Anaerobutyricum.* Resolving the gene expression at the species level revealed a significant decrease in expression for these genes in PD, including in *Blautia obeum*, *Blautia massiliensis, CAG-41 spp.* and *Anaerobutyricum hallii* ($p < 0.05$, Wilcoxon test, Figure 4G and H). Of note, this decrease was also observed at the genus level (data not shown). Interestingly, we found an increase in Flavonifractor plautii expression of BMCs genes encoding for BMC-H and pduQ in PD (Figure 4G and H, $p < 0.05$, Wilcoxon test).

Bacterial microcompartment genes are correlated with flagellar expression.

We next wanted to investigate the links between bacterial microcompartments (BMCs) genes and the flagellar expression. In this manner, we tested the correlation between these genes using both normalized expression and MT TPM. In addition, we also tested the correlations for all the genes from all taxa or selected taxa based that were of particular interest (see material and methods for details). We found strong correlations between levels of BMCs gene expression and flagella assembly genes (Figure 5A and Supplemental fig. 3A-B). Indeed, we noticed significant correlations for 223 and 1 tests when using normalized expression (Figure 5A, $p < 0.05$, $q < 0.05$, respectively); 635 and 370 tests when using MT tpms (Figure 5A, $p < 0.05$, $q < 0.05$, respectively). In addition, correlations using only expression from selected taxa resulted in even more significant correlations both for normalized expression (Figure 5A, 500 genes with $p < 0.05$, 94 genes with $q < 0.05$, respectively) (Figure 5A, 569 genes with $p < 0.05$, 496 genes $q < 0.05$, respectively). Next, we checked specifically hub genes correlations, we found significant correlations between all the hub genes from BMCs and flagellar assembly genes both for normalized expression and MT tpms (Figure 5 B, $q < 0.05$, Spearman correlation test).

Genes enriched in Parkinson's disease show lower gene diversity

After noticing interesting modifications in the co-expression network and the lack of hub genes in PD associated modules, we decided to investigate functional redundancy and the gene expression diversity (GED) to appreciate the taxonomic differences of gene expression. In this context, GED refers to the diversity of taxa expressing a specific gene, while functional redundancy denotes the measure of taxonomic and functional diversity present within a sample (Tian et al., 2020). The following analyses will differentiate between genes found in module association to either one of the conditions (trait association, Figure 5A-C) and genes with differential expression showing an increase or decrease in PD (Figure 5D-F). We did not observe a difference in functional redundancy between HC and PD individuals (Figure 5A, $p > 0.05$, Wilcoxon test). Interestingly, we found no significant differences for overall GED for non-Hub genes but a significantly lower GED for hub genes in PD (Figure 5B, Wilcoxon test, $p > 0.05$ and $p < 0.01$, respectively). We also noticed hub genes had higher GED than non-hub genes (Figure 5C, Wilcoxon test, $p < 0.001$). We then compared the differential expression and functional diversity of a given gene. We found that gene expression was linked to either an increase of GED (more microbes expressing a given gene) or a loss of GED (less microbes expressing a given gene) (Figure 5D and E).

We were able to categorize genes into two groups: those with increased expression linked to a higher GED, and those with increased expression linked to a lower GED. We finally looked at the proportion of genes within the two above-mentioned categories. We found that a significantly higher proportion of genes with decrease GED in PD compared to HC for both hub genes and non-hub genes that had significantly different expression ($p = 0.016$ and $p = 0.002$, Fisher exact test and Chi square test, respectively, Figure 5F). Noteworthy, we noticed the same link for genes (hub and non-hub) that had were not significantly different in expression (Figure 5F). Interestingly, hub genes with significant increased expression in PD

had a higher ratio of decreased GED than non-hub genes (57% and 43%, respectively, Figure 5F).

Discussion

Previously, we demonstrated significant differences in the gut microbiome of individuals with PD (Villette et al., 2024, submitted). In the present study, we employed a network approach to assess the microbiome-wide impacts of these changes. Our goal was to uncover key differences in microbial metabolism between PD individuals and HC. We identified strong associations between co-expression network modules and disease status, with four modules linked to HC and five to PD. Additionally, we found eight modules that were not associated with either HC or PD. Except for module M1, that was mainly a module with genes not clustering with other genes, we found higher values of degree and closeness centrality for these modules compared to the trait associated modules. Given their overall high centrality and the lack of trait association, these modules may represent stable, core functions that uphold the integrity of microbial networks in both health and disease.

Hub genes are important to highlight key functions associated with PD (Calabrese et al., 2012; Farber, 2010; Horvath et al., 2006; Langfelder et al., 2013; Torkamani & Schork, 2009; Zhang & Horvath, 2005). Therefore, identifying hub genes is a highly effective strategy for uncovering genes that contribute to complex diseases such as PD. Using this approach we first decided to select the top 100 connected genes from modules associated with either HC or PD, which led to 95% of the genes being associated with HC. To get more insight on the PD most connected genes we selected the top 10% most connected genes per module. With this approach we uncovered citrate lyase genes being increased in PD, however, we did not manage to link these genes to other functions or to resolve the taxonomic expression of these genes. To our knowledge, there is no record of citrate lyase genes being associated with disease in humans. Interestingly, 13 genes from the flagellar assembly pathway are hub genes when using the top 10% approach and 9 when using the top 100 connected approach, showing once more the importance of this pathway in the microbial network, especially in the context of PD. Finally, the high number of hub genes from HC-associated modules is

noteworthy, as it indicates that there is an imbalance in the expression of key functions in the gut microbiome of PD. Hub genes attributed to the HC associated modules are sought to be key regulators of the microbial network.

Focusing on M2, a module associated with HC, comprises the most connected genes and especially genes from pdu and eut operons, two operons forming bacterial microcompartments. These operons are responsible for ethanolamine and 1,2-propanediol-utilization, an important energy source for bacteria and are typically associated with the survival of specific pathogenic bacteria (Ravcheev & Thiele, 2014; Tsoy et al., 2009), as they confer a growth advantage by utilizing abundantly present 1,2-propanediol and ethanolamine (EA) (Dank et al., 2021; Vance, 2018). However, it has been recently described that a wide range of commensals are also expressing these genes (Asija et al., 2021; Jallet et al., 2024; Q. Li et al., 2024; Reichardt et al., 2014). We find that the expression of these genes is decreased in PD compared to HC, especially in genera such as *Blautia* and *Anaerobutyricum* but increased in *Flavonifractor plautii*, a bacterium that we previously highlighted as associated to PD (Villette et al., 2024, submitted). Indeed, *F. plautii* and *Flavonifractor* genus, showed increased expression of flagellar assembly in PD. Interestingly, we found strong correlations between the expression of BMCs genes and flagellar assembly genes. Additionally, exposure to EA enhances *L. brevis*'s cellular aggregation and adhesion, potentially improving its probiotic efficacy via the prevention of pathogen attachment. Also, non-pathogenic *E. coli* isolates were also described to use EA as energy source to enhance growth, modulate gene expression, and outgrow pathogenic *E.coli* strains (Moreira de Gouveia et al., 2023; Rowley et al., 2018). So, our finding of loss of eut and pdu expression in genera such *Blautia* and *Anaerobutyricum,* suggest that these commensals might be losing access to the nutrients necessary to express these genes and therefore decrease their expression of flagella.

We demonstrate a decreased gene expression diversity in genes associated to PD but no differences in overall functional redundancy. Nonetheless, we show here that most of genes

overexpressed in PD are linked with a decrease in the diversity of taxa expressing these functions. Interestingly, this was even more present in genes that we defined as hub genes (using the second approach), headlining the loss of keystone genes expression in PD. In HC associated genes, we noticed an opposite relationship, increased expression was linked with increased gene expression diversity. We here show that the gut microbiome dysbiosis is indeed linked with a disruption of keystone functions in PD but also a disruption of diversity of functionality. We couldn't measure resilience and stability of PD gut microbiome, these measures being time dependent, but we strongly suppose that a general loss of gene expression diversity will result or is secondary to a decreased resilience and stability as hypothesized before (Ives & Carpenter, 2007; Loreau & Behera, 1999).

Material and methods

Patient cohorts and sampling

Kassel Cohort

The DeNoPa cohort represents a prospective, biannual follow-up study of (initially *de novo)* Parkinson's disease (PD) patients at the Paracelsus-Elena Klinik, Kassel, Germany. Fecal samples from PD patients (46) and healthy controls (29) were collected during the 4-year follow-up visit for the cohort. Details on inclusion and exclusion criteria and ancillary investigations have been published previously (Mollenhauer et al., 2013, 2016). Subjects with idiopathic rapid-eye-movement sleep behavior disorder (iRBD, 13) were recruited at the same clinic, diagnosed according to consensus criteria of the International RBD study group (Schenck et al., 2013) using video-assisted polysomnography, and were included only if they showed no signs of a neurodegenerative disorder. DeNoPa subjects were required to have a 4-week antibiotic free interval before fecal sample collection. As additional control subjects, we collected fecal samples from (20) neurologically healthy subjects living in the same household as the DeNoPa participants. Samples of *de novo* PD patients from a cross-sectional cohort at the same clinic were included if subjects were recently diagnosed, drug-naïve and met United Kingdom Parkinson's Disease Society Brain Bank (UKPDSBB) clinical diagnostic criteria (Hughes et al., 1992). All subjects except household HC were interviewed and examined by an expert in movement disorders. The study conformed to the Declaration of Helsinki and was approved by the ethics committee of the Physician's Board Hessen, Germany (FF 89/2008). The DeNoPa trial is registered at the German Register for Clinical trials (DRKS00000540).

Marburg Cohort

We also added samples from 14 patients with polysomnography-confirmed iRBD which were recruited from the outpatient clinic of the Department of Neurology, Philipps-University,

Marburg, Germany, between November 2015 and November 2016. iRBD was diagnosed according to the guidelines of the American Academy of Sleep Medicine (AASM ICSD-3) (Sateia, 2014). A detailed medical history was recorded, and a complete neurological examination performed to verify the subjects' suitability. Inclusion criteria were age above 18 years, no dopamimetic therapy, and no diagnosis of PD, MSA, DLB or PSP. Exclusion criteria were smoking, antibiotic therapy in the last 24 months, history of other neurological diseases or disorders of the gastrointestinal tract. Non-motor and autonomic symptoms were evaluated with the SCOPA-AUT (Visser et al., 2004) and PD-NMS (Chaudhuri et al., 2007) questionnaires. Motor function was evaluated with the UPDRS (Fahn et al., 1987). Additionally, patients were asked to complete the RBD-Sleep questionnaire (Stiasny-Kolster et al., 2007). The study conformed to the Declaration of Helsinki and was approved by the ethics committee of the Medical Faculty of the Philipps-University, Marburg, Germany (46/14).

Consent

All subjects from both cohorts provided informed written consent, and the sample analysis was approved by the Comité National d'Ethique de Recherche of Luxembourg (reference no.: 140174_ND).

Fecal sample collection

Fecal samples were collected at the clinics via a stool specimen collector (MedAuxil) and collection tubes (Sarstedt), as previously described (Heintz-Buschart et al., 2018). Samples were immediately flash-frozen on dry ice after collection. Samples were subsequently stored at –80 °C and shipped on dry ice.

Sample exclusions

The initial set of samples consisted of 50 PD and 50 healthy control subjects (HC). Three PD and two iRBD cases were subsequently excluded for clinical reasons (adjusted diagnosis),

one iRBD and one PD subject for logistical reasons, and one control due to a combination of microbiome-altering medications (metformin, antidepressants, statins, and proton pump inhibitors). Additional samples were excluded due to missing values (metabolomics) or a low amount of identified analytes (metaproteomics), leading to the final numbers of samples 46 for individuals with PD and 49 for HC.

Metagenomic and metatranscriptomic sequencing

Extractions from fecal samples were performed according to a previously published protocol (Roume et al., 2012) conducted on a customized robotic system (Tecan Freedom EVO 200). After extraction, DNA and RNA were purified prior the sequencing analysis by using the following commercial kits respectively: Zymo DNA Clean&Concentrator-5 (D4014) and Zymo RNA Clean&Concentrator-5 (R1014). RNA quality was assessed and quantified with an Agilent 2100 Bioanalyser (Agilent Technologies) and the Agilent RNA 6000 Nano kit, and genomic DNA and RNA fractions with a NanoDrop Spectrophotometer 1000 (Thermo Scientific) as well as commercial kits from Qubit (Qubit ds DNA BR Assay kit, Q32850; Qubit RNA BR Assay kit, Q10210). All DNA samples were subjected to random shotgun sequencing. Following DNA isolation, 200-300 ng of DNA was sheared using a Bioruptor NGS (Diagenode) with 30s ON and 30s OFF for 20 cycles. Sequencing libraries were prepared using the TruSeq Nano DNA library preparation kit (Illumina) following the manufacturer's protocol, with 350 bp average insert size. For MT, 1 µg of isolated RNA was rRNA-depleted using the RiboZero kit (Illumina, MRZB12424). Library preparation was performed using the TruSeq Stranded mRNA library preparation kit (Illumina) following the manufacturer's protocol, apart from omitting the initial steps for mRNA pull down. MG and MT analyses, the qualities of the libraries were checked using a Bioanalyzer (Agilent) and quantified using Qubit (Invitrogen). Libraries were sequenced on an Illumina NextSeq500 instrument with 2x150 bp read length

Bioinformatics and statistical analyses

Sequencing data processing and analysis

For all samples, MG and MT sequencing data were processed and hybrid-assembled using the Integrated Meta-omic Pipeline (IMP) (Narayanasamy et al., 2016) (https://git-r3lab.uni.lu/IMP/imp3, commit 8c1bd6fa443d064511909c9eede20703f45e6c69). Data was quality trimmed, adapter sequences were removed, MT rRNA reads were removed by mapping against SILVA 138.1 (Quast et al., 2013) and human reads were removed from MT and MG after mapping against the human genome (hg38) and transcriptome (RefSeq 212). Pre-processed MG and MT reads were co-assembled using the IMP-based iterative hybrid-assembly pipeline using MEGAHIT 1.0.3 (D. Li et al., 2015). After assembly, the prediction and annotation of genomic features such as open-reading frames (ORFs) and non-coding genes was performed using a modified version of Prokka (Seemann, 2014) and followed by functional annotation of those using Mantis (Queirós et al., 2021). Genomic features were quantified on MG and MT level using featureCounts (Liao et al., 2014) from the final gff file. Taxonomic annotation of reads and contigs was performed using Kraken2 (Wood et al., 2019) with a GTDB release207 database (http://ftp.tue.mpg.de/ebio/projects/struo2/GTDB_release207/kraken2) and a 0.5 confidence threshold. Additionally, taxon abundances were estimated using mOTUs 2.5.1 (Milanese et al., 2019). The mOTU abundances were used to generate abundance matrices for each taxonomic rank (phylum, class, order, family, genus and species) by summing up taxon marker read counts at the respective levels.

Co-expression network construction

The Python package WGCNA (PyWGCNA, version 2.0.4) was used to construct a co-expression network of genes expressed in the microbiome of PD patients and HC according to the WGCNA procedure (Rezaie et al., 2023). As input to PyWGCNA, we used MG-normalized MT expression of KEGG orthologs (KO) (MT/MG ratio) in PD and HC samples

(Roume et al., 2015). MG-normalized expression of KOs was power transformed using PowerTransformer from *sklearn.preprocessing* (https://scikit-learn.org/stable/api/sklearn.preprocessing.html). The *WGCNA* function was run with the following parameters: minimum module size minModuleSize=20, dissimilarity threshold MEDissThres=0.18, networkType='signed'. Gene modules were identified using hierarchical clustering and the dynamic tree-cut function.

Diversity measures

To describe module diversity, functional redundancy and gene expression diversity, we used the Shannon index with different settings. We defined module diversity by the number and evenness of gene expression within a module, we summed normalized MT expression for each gene and used Shannon index from the vegan R package (2.6.6.1) (Oksanen et al., 2016). We then defined gene diversity by the richness and evenness of taxa expressing a given gene, also using the Shannon index. Functional redundancy was calculated using the R package SYNCSA (1.3.4) using the function *rao.diversity()* (Debastiani & Pillar, 2012).

Figure legends

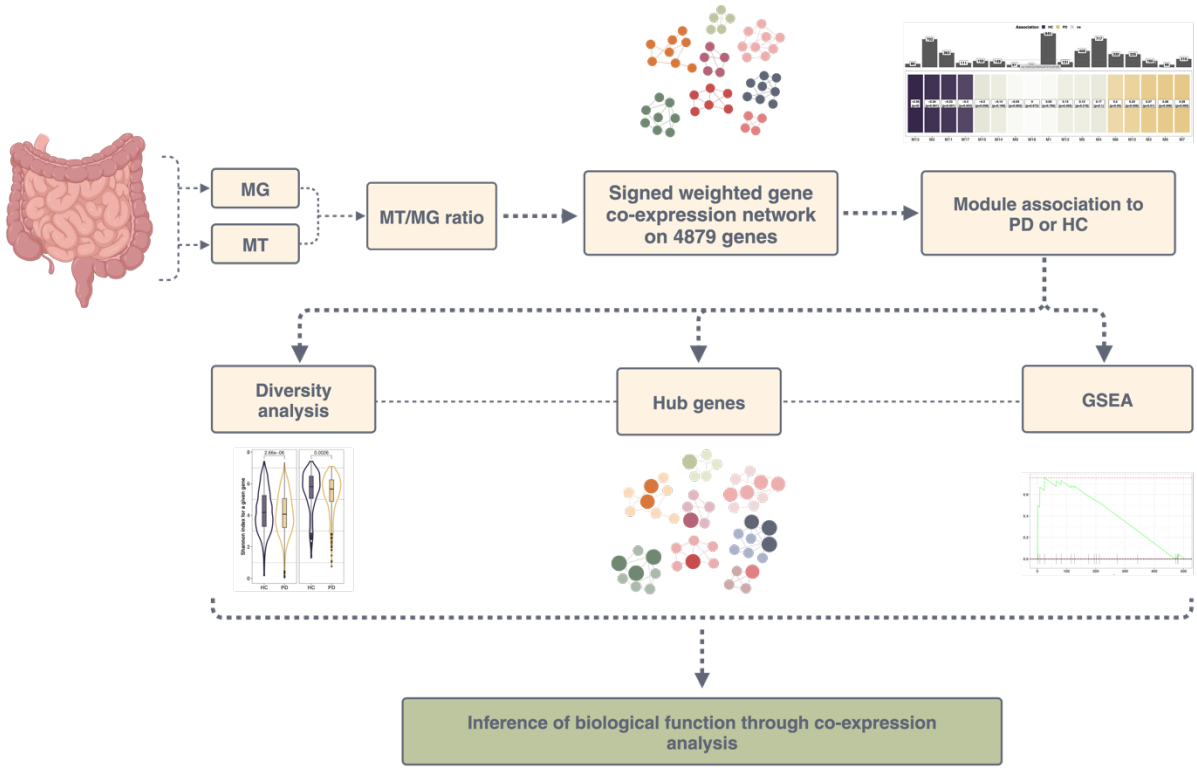


Figure 1. Schematic overview the data analysis workflow for identifying and analyzing co-expressed gene modules using WGCNA.

Initially, metagenomic (MG) and metatranscriptomic (MT) counts per gene were converted into values representing the normalized gene expression MT/MG ratio. A co-expression gene network was then constructed based on a dataset of 4879 genes derived from both PD and HC individuals. This network revealed 17 distinct gene modules. Among these, we selected modules significantly associated with either PD or HC. Further analyses focused on correlations between module genes and specific metabolites, diversity analysis, hub genes within these key modules, gene set enrichment analysis, aiming to uncover the biological relevance of these modules in relation to the disease.

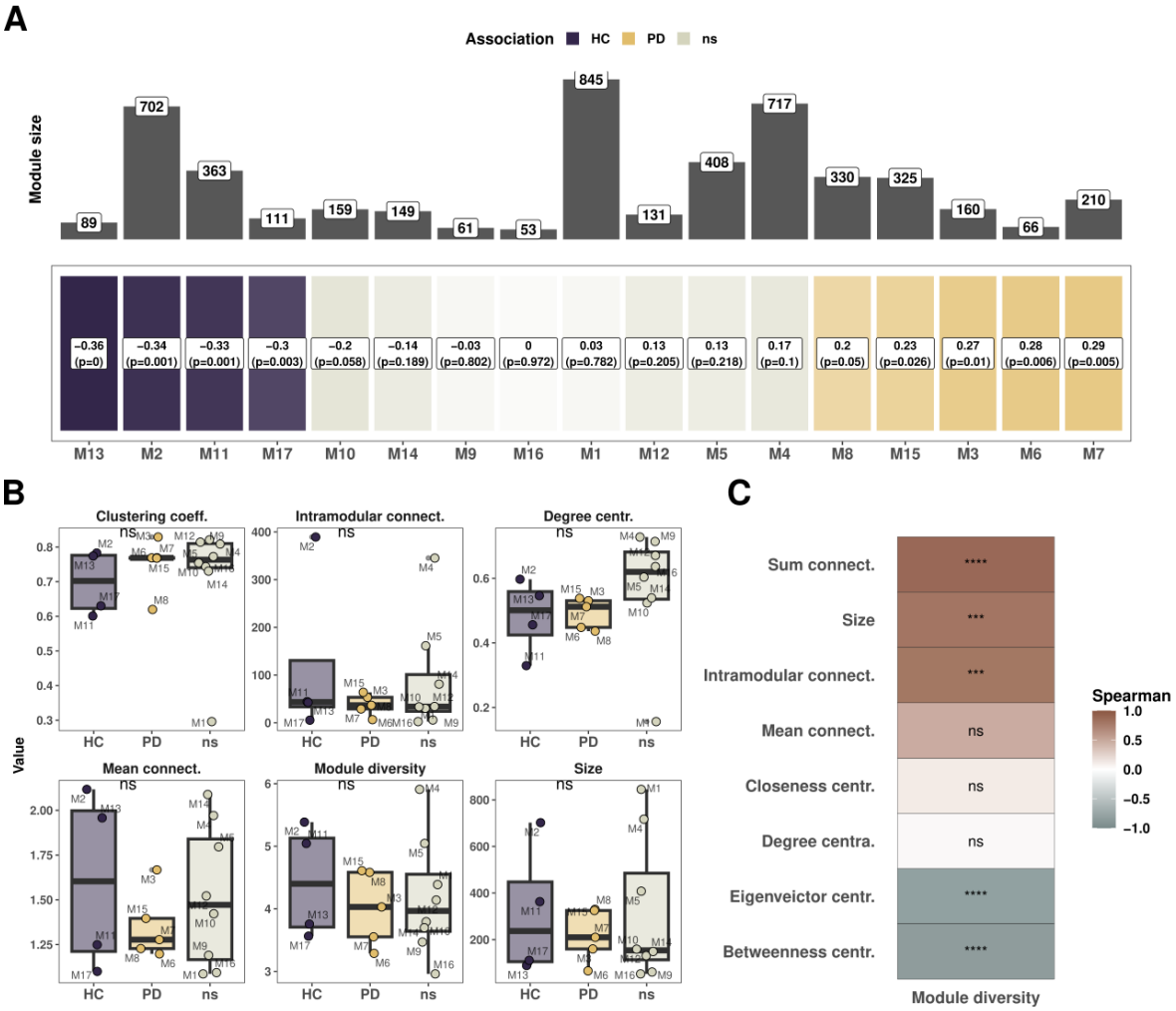


Figure 2. WGCNA reveals module association with disease.

A. Module trait relationship heatmap with correlation and p-values for each module. Modules are sorted based on the correlation value. The top panel represents the number of genes belonging to each module. **B.** Network topology analysis for the modules grouped by trait association. A Kruskal and Wallis test was conducted according to the trait association to compare modules based on their associations or not to one of the two groups. **C.** Correlation between module diversity and other topology features.

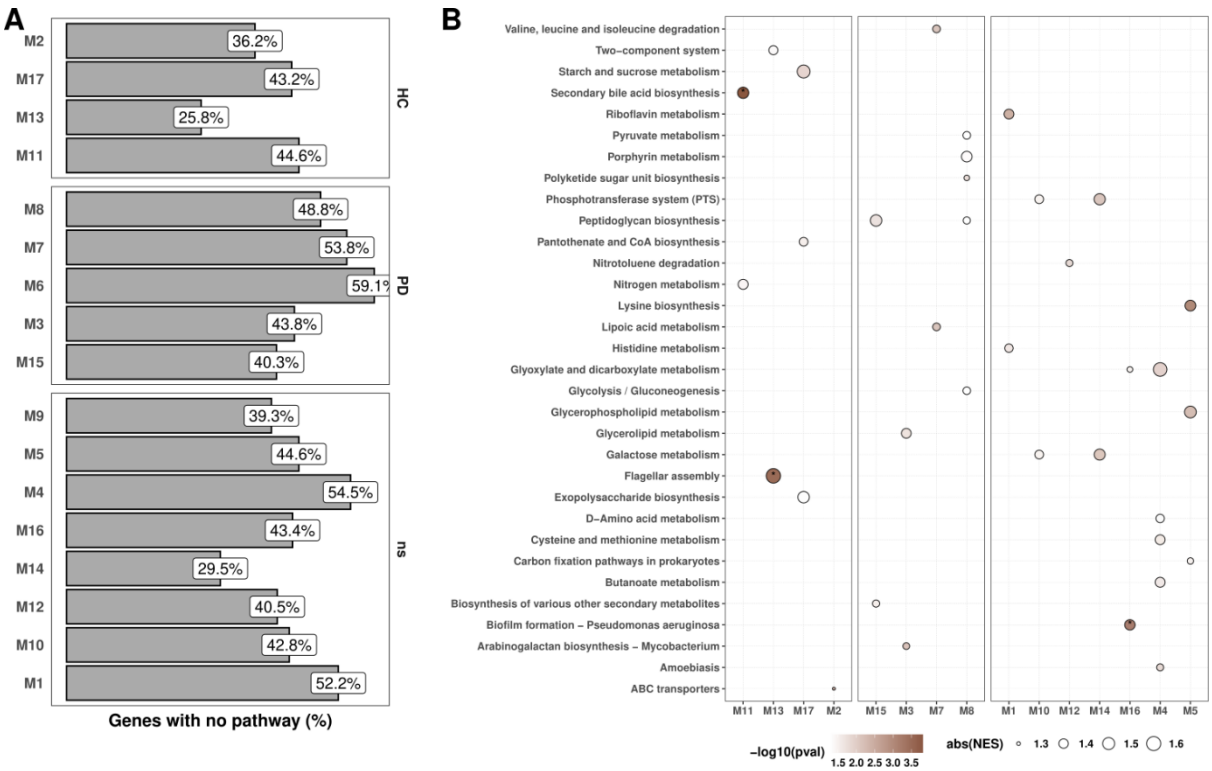


Figure 3. KEGG pathway analysis shows enrichment in HC modules but not in PD modules.

A. Count of genes with undescribed pathways or not belonging to any KEGG pathway. **B.** Gene Set Enrichment Analysis. Dots are only significant enrichment ($p < 0.05$), colored by -log10(p-value). Asterix represent significant enrichment after FDR correction ($q < 0.05$).

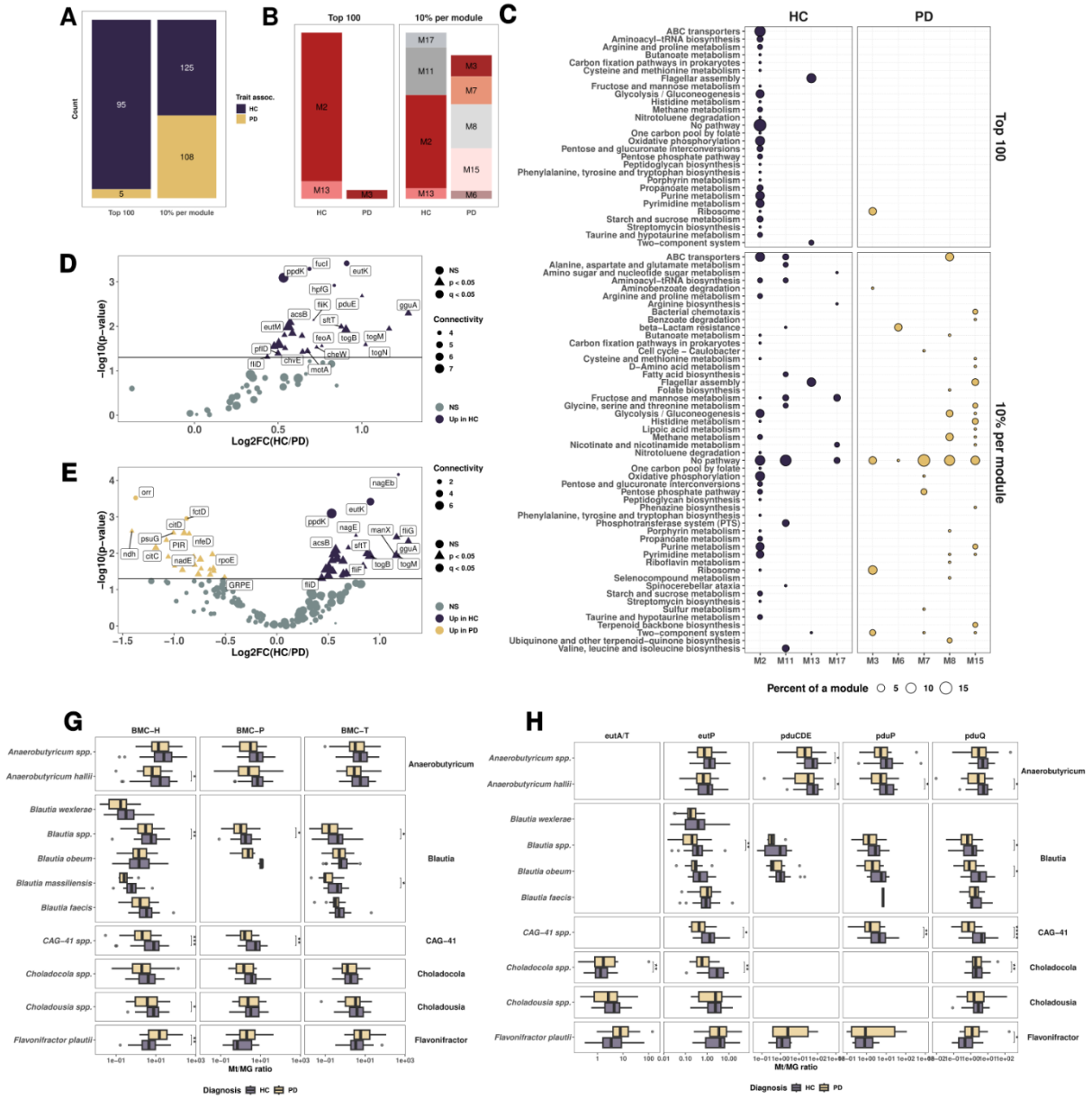


Figure 4. Hub genes are mainly associated with HC individuals.

A. Bar plot showing the count of hub genes selected based on top 100 connected genes (left panel) and 10 % top connected genes per module (right panel). **B.** Dot plot representing counts of pathways per module for the hub genes. The size of the dots represents the proportion of a given pathway within a module. **C and D.** Volcano plots of differential expression of genes for the hub genes selected with the top 100 connected genes in the network (**C.**) and top 10% connected per modules (**D.**), considering only modules significantly associated to one of the groups. Dots are colorized by disease category and shaped on the level of significance, triangular shape for p < 0.05 and round colored shape for

436 q < 0.05. **G and H.** Boxplots representing gene normalized expression resolved at the
437 species level for the BMC shell proteins (G.) and BMCs catabolism/anabolism (F.). All tests
438 are Wilcoxon tests with p values before correction.

439

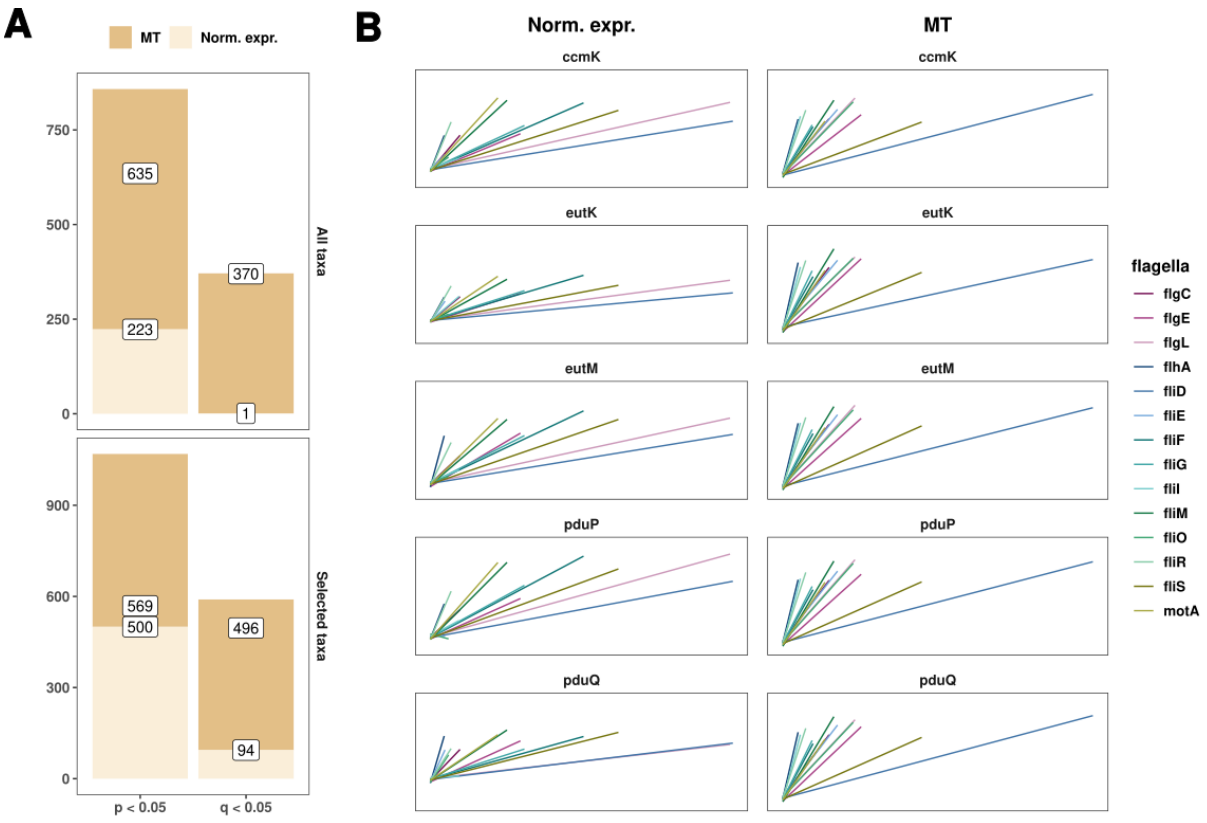


Figure 5. Bacterial microcompartments are correlated with flagella expression.

A. Bar plot counting the number of positive correlations before and after FDR correction for normalized expression and MT tpms when taking all taxa expressing the BMCs genes and flagella genes (upper panel) and relevant taxa from figure 4 (lower panel). All tests are spearman correlation. **B.** Correlation plots including selected taxa, both for normalized expression and MT tpms, considering only hub genes from the 10% per modules approach. Tests are spearman correlation and all correlation are significant after FDR correction ($q < 0.05$).

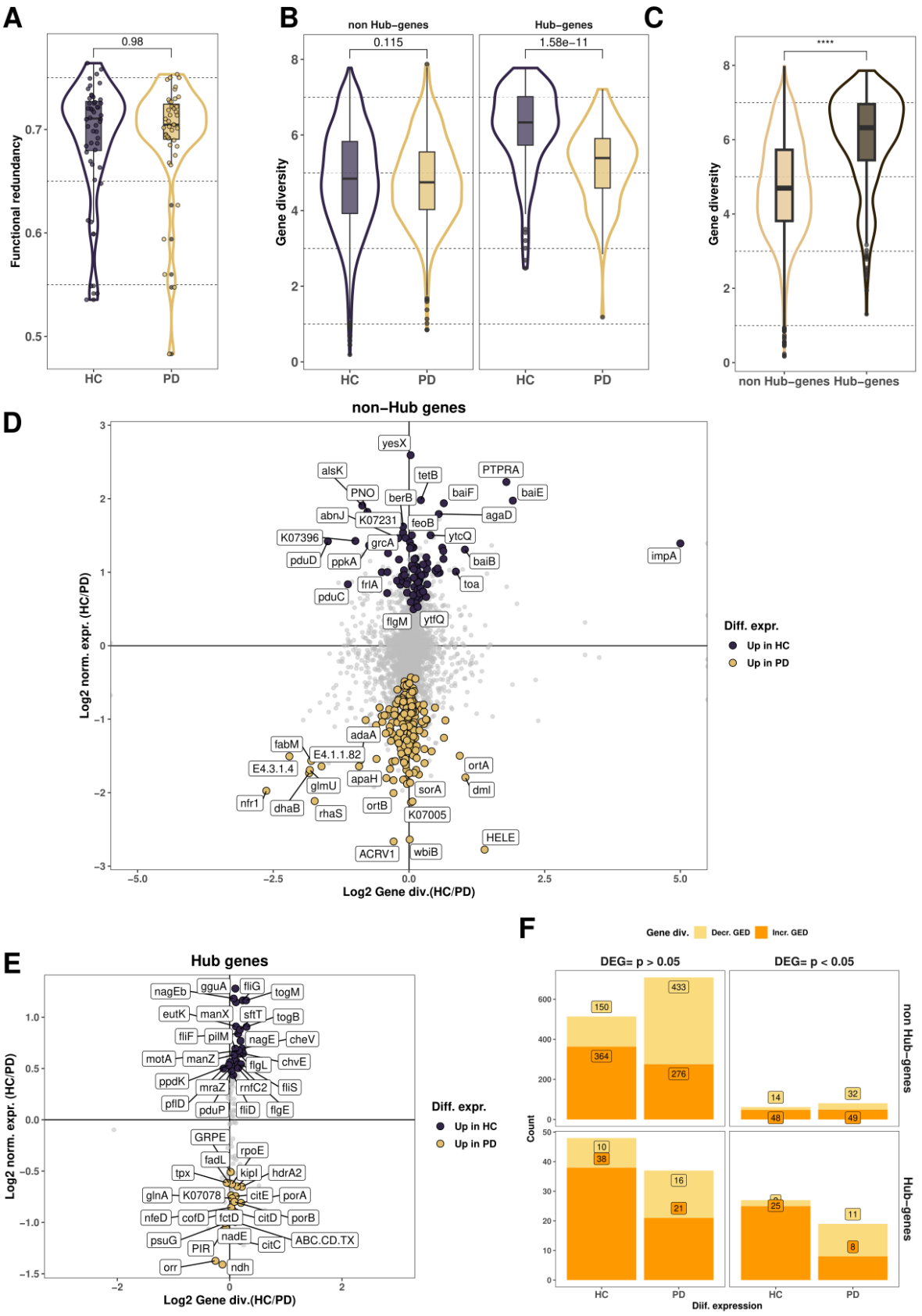
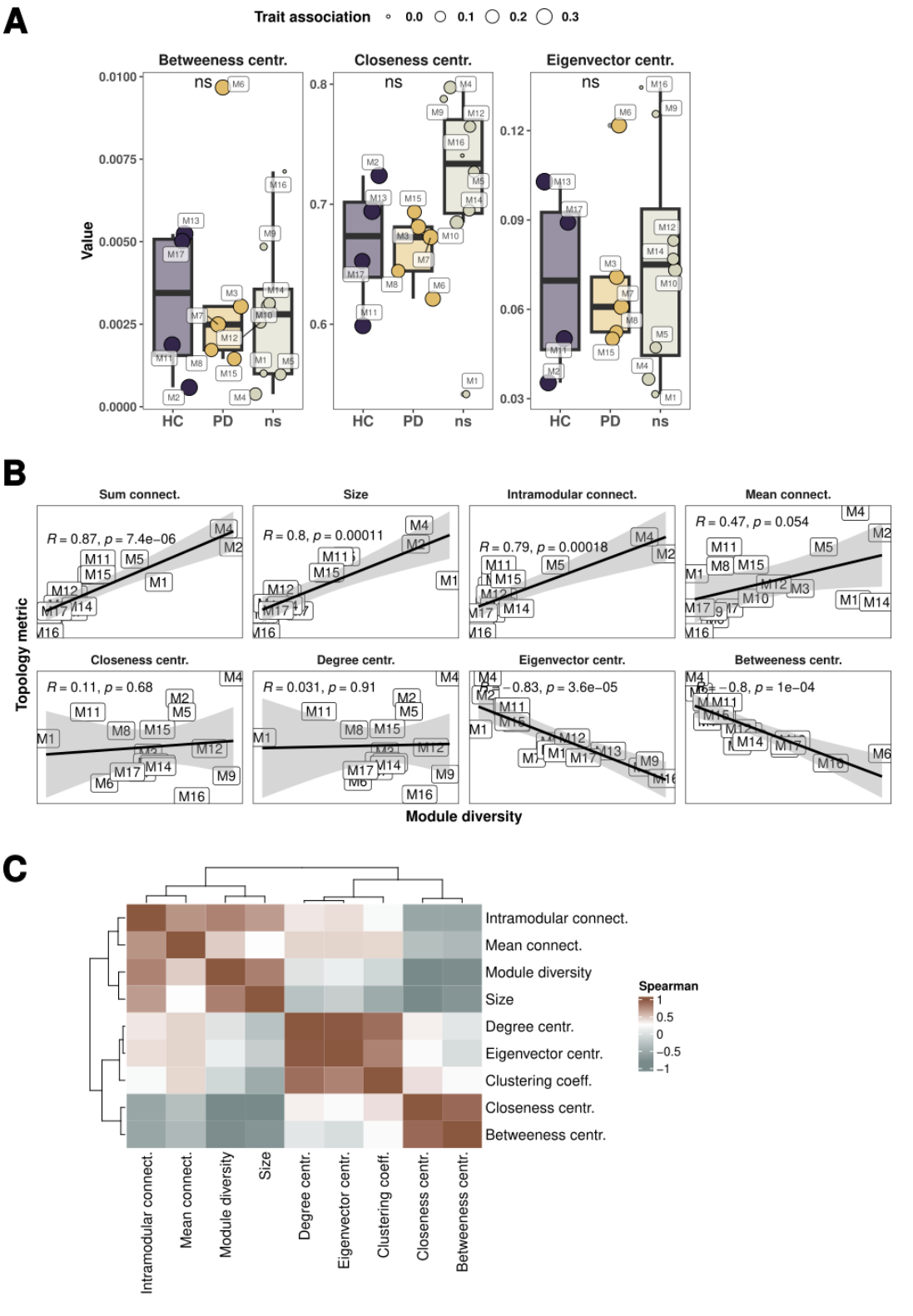

A
0.98
Functional redundancy
HC
PD
B
non Hub−genes
Hub−genes
0.115
1.58e−11
Gene diversity
HC
PD
HC
PD
C
****
Gene diversity
non Hub−genes
Hub−genes
D
non−Hub genes
Log2 norm. expr. (HC/PD)
Log2 Gene div.(HC/PD)
yesX
alsK
tetB
PTPRA
PNO
berB
baiF
baiE
abnJ
K07231
feoB
agaD
K07396
grcA
ytcQ
impA
pduD
ppkA
baiB
frlA
toa
pduC
flgM
ytfQ
adaA
fabM
E4.1.1.82
E4.3.1.4
apaH
ortA
glmU
sorA
dml
nfr1
dhaB
rhaS
ortB
K07005
HELE
ACRV1
wbiB
Diff. expr.
Up in HC
Up in PD
E
Hub genes
Log2 norm. expr. (HC/PD)
Log2 Gene div.(HC/PD)
nagEb
gguA
fliG
togM
eutK
manX
sftT
togB
fliF
pilM
nagE
cheV
motA
manZ
flgL
chvE
ppdK
mraZ
rnfC2
fliS
pflD
pduP
fliD
flgE
GRPE
fadL
rpoE
tpx
kipI
hdrA2
glnA
K07078
citE
porA
nfeD
cofD
fctD
citD
porB
psuG
PIR
nadE
citC
ABC.CD.TX
orr
ndh
Diff. expr.
Up in HC
Up in PD
F
Gene div.
Decr. GED
Incr. GED
DEG= p > 0.05
DEG= p < 0.05
Count
non Hub−genes
Hub−genes
HC
PD
HC
PD
Diif. expression

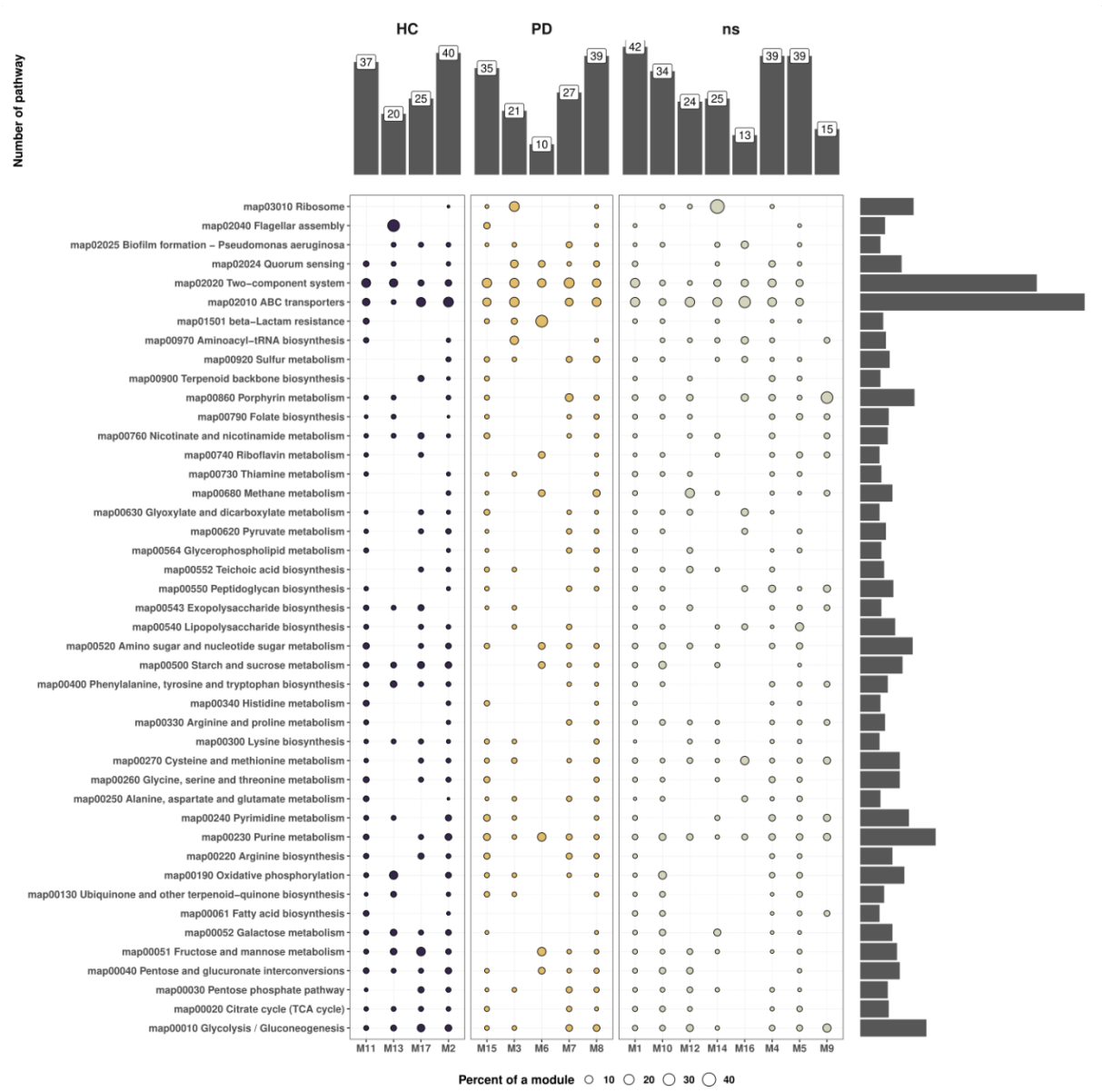
Figure 6. Gene diversity is decreased in PD individuals.

A. Boxplot representing functional redundancy for each sample according to disease status. **B.** Boxplot representing gene expression diversity according to disease status. Gene diversity is here defined by Shannon index of species expressing a given gene, the TPM are summed at the disease status level. **C.** Boxplot representing gene expression diversity grouped by hub genes belonging or not. Differential abundance versus differential diversity for a given gene for non-Hub genes (**D.**) and Hub genes (**E.**). Y axis represents log2-fold change of normalized expression and X axis the log 2-fold change of Gene diversity. Dots are labelled and colored for genes with p-value < 0.05. **F.** Stacked bar plot representing the count of genes with increase or decrease gene diversity for Hub and non-Hub genes. Genes are segregated into PD or HC group according to the sign of DEG and faceted according to DEG significance.


A
Trait association 0.0 0.1 0.2 0.3
Betweeness centr.
Closeness centr.
Eigenvector centr.
Value
ns
HC
PD
ns
B
Sum connect.
Size
Intramodular connect.
Mean connect.
Closeness centr.
Degree centr.
Eigenvector centr.
Betweeness centr.
Topology metric
Module diversity
R = 0.87, p = 7.4e−06
R = 0.8, p = 0.00011
R = 0.79, p = 0.00018
R = 0.47, p = 0.054
R = 0.11, p = 0.68
R = 0.031, p = 0.91
R = −0.83, p = 3.6e−05
R = −0.8, p = 1e−04
C
Intramodular connect.
Mean connect.
Module diversity
Size
Degree centr.
Eigenvector centr.
Clustering coeff.
Closeness centr.
Betweeness centr.
Spearman
1
0.5
0
−0.5
−1

Supplemental figure 1. Additional topology metrics from WGCNA.

A. Boxplots representing betweenness, closeness and eigenvector centrality. Kruskal and Wallis test. **B.** Correlation between topology metrics and module diversity. All tests are spearman correlation.

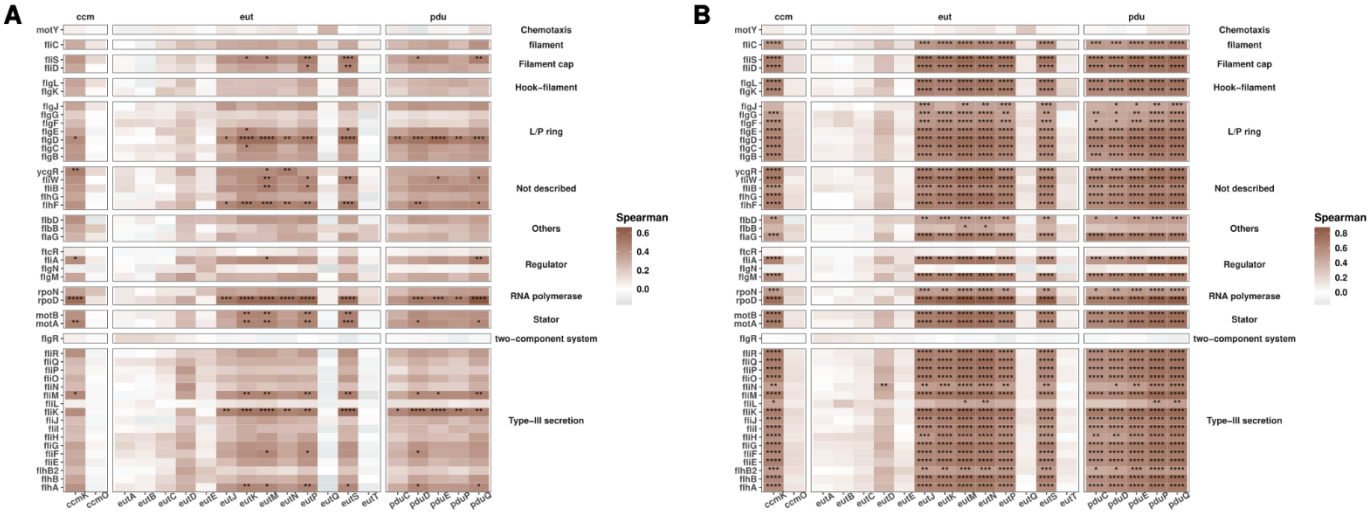


Supplemental figure 2. KEGG pathway for each module.

Count of pathway per module for the hub genes. The size of the dots represents the proportion of a given pathway within a module.

Supplemental figure 3. Bacterial microcompartments genes correlate with flagella assembly genes.

Heatmaps representing spearman correlation coefficients between genes involved in BMCs formation, catabolism or anabolism and genes involved in flagellar assembly. A. Heatmap correlation tests for selected bacteria using normalized expression. B. Heatmap correlation tests for selected bacteria using MT tpms.



Data availability

The datasets generated by this study are available in the following repositories: metagenomic and metatranscriptomic data at the NCBI BioProject collection with the ID PRJNA782492 (http://www.ncbi.nlm.nih.gov/bioproject/782492), metaproteomic data at the Proteomics Identifications (PRIDE) database with accession number PXD031457 (https://www.ebi.ac.uk/pride/archive/projects/PXD031457), and metabolomic data at MetaboLights with ID MTBLS5092 (https://www.ebi.ac.uk/metabolights/MTBLS5092). Due to privacy restrictions, clinical and demographic data are available on request from the corresponding author.

Code availability

The IMP pipeline, which was used for analysis of metagenomic and metatranscriptomic data, is available at https://gitlab.lcsb.uni.lu/IMP/imp3. The R and python code used for statistical analyses and visualizations is available at https://gitlab.lcsb.uni.lu/ESB/[TBA].

Acknowledgments

We thank the staff of the Luxembourg Centre for Systems Biomedicine (LCSB), particularly the sequencing platform and the metabolomics platform, for running the sequencing and metabolomic analyses. The bioinformatics presented in this paper were carried out using the HPC facilities of the University of Luxembourg (Varrette et al., 2022). In addition, we thank the clinician and hospital personnel that recruited the participants.

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 863664), and was further supported by the Luxembourg National Research Fund (FNR) CORE/16/BM/11333923 (MiBiPa), CORE/15/BM/10404093 (microCancer/MUST),

PRIDE/11823097 (MICROH DTU), and CORE/19/BM/13684739 (metaPUF), the Michael J. Fox Foundation under grant IDs 14701 (MiBiPa-PLUS) and MJFF-019228 (PARKdiet), the Parkinson's Foundation (MiBiPa Saliva), the Institute for Advanced Studies of the University of Luxembourg through an AUDACITY grant (ref. no. MCI-BIOME_2019), as well as an "Espoîr en tête" grant from the Rotary Club Luxembourg to P.W. This work was also supported by a Fulbright Research Scholarship from the Commission for Educational Exchange between the United States, Belgium and Luxembourg to P.W. Additional funding was provided by the FNR under INTERMOBILITY/23/17856242.

Contributions

Conceptualization P.M, P.W. Multi-omic data generation: P.M., R.V. Bioinformatics, statistics and data visualization: R.V, P.N. Initial manuscript draft: R.V, P.N. Review and editing: C.C.L., P.M. and P.W. Funding acquisition: C.C.L, P.W. All authors read and approved of the submitted version.

Bibliography

Asija, K., Sutter, M., & Kerfeld, C. A. (2021). A Survey of Bacterial Microcompartment Distribution in the Human Microbiome. *Frontiers in Microbiology*, *12*, 669024. https://doi.org/10.3389/FMICB.2021.669024/BIBTEX

Barabási, A. L., Gulbahce, N., & Loscalzo, J. (2010). Network medicine: a network-based approach to human disease. *Nature Reviews Genetics 2011 12:1*, *12*(1), 56–68. https://doi.org/10.1038/nrg2918

Bhattarai, Y., Si, J., Pu, M., Ross, O. A., McLean, P. J., Till, L., Moor, W., Grover, M., Kandimalla, K. K., Margolis, K. G., Farrugia, G., & Kashyap, P. C. (2021). Role of gut microbiota in regulating gastrointestinal dysfunction and motor symptoms in a mouse model of Parkinson's disease. *Gut Microbes*, *13*(1). https://doi.org/10.1080/19490976.2020.1866974

Boertien, J. M., Pereira, P. A. B., Aho, V. T. E., & Scheperjans, F. (2019). Increasing Comparability and Utility of Gut Microbiome Studies in Parkinson's Disease: A Systematic Review. *Journal of Parkinson's Disease*, *9*(s2), S297–S312. https://doi.org/10.3233/JPD-191711

Braak, H., Rüb, U., Gai, W. P., & Del Tredici, K. (2003). Idiopathic Parkinson's disease: possible routes by which vulnerable neuronal types may be subject to neuroinvasion by an unknown pathogen. *Journal of Neural Transmission (Vienna, Austria : 1996)*, *110*(5), 517–536. https://doi.org/10.1007/S00702-002-0808-2

Brandes, U., & Erlebach, T. (Eds.). (2005). *Network Analysis*. *3418*. https://doi.org/10.1007/B106453

Cai, L., Tang, S., Liu, Y., Zhang, Y., & Yang, Q. (2023). The application of weighted gene co-expression network analysis and support vector machine learning in the screening of Parkinson's disease biomarkers and construction of diagnostic models. *Frontiers in*

Molecular Neuroscience, *16*, 1274268. https://doi.org/10.3389/FNMOL.2023.1274268/BIBTEX

Calabrese, G., Bennett, B. J., Orozco, L., Kang, H. M., Eskin, E., Dombret, C., De Backer, O., Lusis, A. J., & Farber, C. R. (2012). Systems Genetic Analysis of Osteoblast-Lineage Cells. *PLoS Genetics*, *8*(12), 1003150. https://doi.org/10.1371/JOURNAL.PGEN.1003150

Chaudhuri, K. R., Martinez-Martin, P., Brown, R. G., Sethi, K., Stocchi, F., Odin, P., Ondo, W., Abe, K., MacPhee, G., MacMahon, D., Barone, P., Rabey, M., Forbes, A., Breen, K., Tluk, S., Naidu, Y., Olanow, W., Williams, A. J., Thomas, S., … Schapira, A. H. V. (2007). The metric properties of a novel non-motor symptoms scale for Parkinson's disease: Results from an international pilot study. *Movement Disorders : Official Journal of the Movement Disorder Society*, *22*(13), 1901–1911. https://doi.org/10.1002/MDS.21596

Cryan, J. F., O'Riordan, K. J., Sandhu, K., Peterson, V., & Dinan, T. G. (2020). The gut microbiome in neurological disorders. *The Lancet. Neurology*, *19*(2), 179–194. https://doi.org/10.1016/S1474-4422(19)30356-4

Dank, A., Zeng, Z., Boeren, S., Notebaart, R. A., Smid, E. J., & Abee, T. (2021). Bacterial Microcompartment-Dependent 1,2-Propanediol Utilization of Propionibacterium freudenreichii. *Frontiers in Microbiology*, *12*. https://doi.org/10.3389/FMICB.2021.679827/FULL

Debastiani, V. J., & Pillar, V. D. (2012). SYNCSA—R tool for analysis of metacommunities based on functional traits and phylogeny of the community components. *Bioinformatics*, *28*(15), 2067–2068. https://doi.org/10.1093/BIOINFORMATICS/BTS325

Fahn, S. , E. R. and M. of the U. D. C. (1987). (n.d.). *The Unified Parkinson's Disease Rating Scale. In Fahn, S., Marsden, C.D., Calne, D.B. and Goldstein, M., Eds., Recent Developments in Parkinson's Disease, Vol. 2, McMellam Health Care Information,*

Florham Park, 153-163. - References - Scientific Research Publishing. Retrieved October 9, 2024, from https://www.scirp.org/reference/referencespapers?referenceid=1365499

Farber, C. R. (2010). Identification of a gene module associated with BMD through the integration of network analysis and genome-wide association data. *Journal of Bone and Mineral Research : The Official Journal of the American Society for Bone and Mineral Research*, *25*(11), 2359–2367. https://doi.org/10.1002/JBMR.138

Fasano, A., Visanji, N. P., Liu, L. W. C., Lang, A. E., & Pfeiffer, R. F. (2015). Gastrointestinal dysfunction in Parkinson's disease. *The Lancet. Neurology*, *14*(6), 625–639. https://doi.org/10.1016/S1474-4422(15)00007-1

Galán-Vásquez, E., & Perez-Rueda, E. (2019). Identification of Modules With Similar Gene Regulation and Metabolic Functions Based on Co-expression Data. *Frontiers in Molecular Biosciences*, *6*. https://doi.org/10.3389/FMOLB.2019.00139

Heintz-Buschart, A., Pandey, U., Wicke, T., Sixel-Döring, F., Janzen, A., Sittig-Wiegand, E., Trenkwalder, C., Oertel, W. H., Mollenhauer, B., & Wilmes, P. (2018). The nasal and gut microbiome in Parkinson's disease and idiopathic rapid eye movement sleep behavior disorder. *Movement Disorders : Official Journal of the Movement Disorder Society*, *33*(1), 88–98. https://doi.org/10.1002/MDS.27105

Hirayama, M., & Ohno, K. (2021). Parkinson's Disease and Gut Microbiota. *Annals of Nutrition & Metabolism*, *77 Suppl 2*, 28–35. https://doi.org/10.1159/000518147

Horvath, S., Zhang, B., Carlson, M., Lu, K. V., Zhu, S., Felciano, R. M., Laurence, M. F., Zhao, W., Qi, S., Chen, Z., Lee, Y., Schecks, A. C., Liau, L. M., Wu, H., Geschwind, D. H., Febbo, P. G., Kornblum, H. I., Cloughesy, T. F., Nelson, S. F., & Mischel, P. S. (2006). Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. *Proceedings of the National Academy of Sciences of the United States*

of America, 103(46), 17402–17407. https://doi.org/10.1073/PNAS.0608396103/SUPPL_FILE/IMAGE418.GIF

Hughes, A. J., Daniel, S. E., Kilford, L., & Lees, A. J. (1992). Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *Journal of Neurology, Neurosurgery, and Psychiatry*, *55*(3), 181–184. https://doi.org/10.1136/JNNP.55.3.181

Ives, A. R., & Carpenter, S. R. (2007). Stability and diversity of ecosystems. *Science (New York, N.Y.)*, *317*(5834), 58–62. https://doi.org/10.1126/SCIENCE.1133258

Jallet, D., Soldan, V., Shayan, R., Stella, A., Ismail, N., Zenati, R., Cahoreau, E., Burlet-Schiltz, O., Balor, S., Millard, P., & Heux, S. (2024). Integrative in vivo analysis of the ethanolamine utilization bacterial microcompartment in Escherichia coli. *MSystems*, *9*(8). https://doi.org/10.1128/MSYSTEMS.00750-24

Keshavarzian, A., Green, S. J., Engen, P. A., Voigt, R. M., Naqib, A., Forsyth, C. B., Mutlu, E., & Shannon, K. M. (2015). Colonic bacterial composition in Parkinson's disease. *Movement Disorders*, *30*(10), 1351–1360. https://doi.org/10.1002/MDS.26307

Koschützki, D., & Schreiber, F. (2004). *Comparison of centralities for biological networks*. 199–206. https://dl.gi.de/handle/20.500.12116/28667

Koschützki, D., & Schreiber, F. (2008). Centrality Analysis Methods for Biological Networks and Their Application to Gene Regulatory Networks. *Gene Regulation and Systems Biology*, *2*(2), 193. https://doi.org/10.4137/GRSB.S702

Langfelder, P., Mischel, P. S., & Horvath, S. (2013). When Is Hub Gene Selection Better than Standard Meta-Analysis? *PLOS ONE*, *8*(4), e61505. https://doi.org/10.1371/JOURNAL.PONE.0061505

Li, D., Liu, C. M., Luo, R., Sadakane, K., & Lam, T. W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de

Bruijn graph. *Bioinformatics (Oxford, England)*, *31*(10), 1674–1676. https://doi.org/10.1093/BIOINFORMATICS/BTV033

Li, Q., Ruscheweyh, H. J., Østergaard, L. H., Libertella, M., Simonsen, K. S., Sunagawa, S., Scoma, A., & Schwab, C. (2024). Trait-based study predicts glycerol/diol dehydratases as a key function of the gut microbiota of hindgut-fermenting carnivores. *Microbiome*, *12*(1), 178. https://doi.org/10.1186/S40168-024-01863-4/FIGURES/7

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)*, *30*(7), 923–930. https://doi.org/10.1093/BIOINFORMATICS/BTT656

Loreau, M., & Behera, N. (1999). Phenotypic diversity and stability of ecosystem processes. *Theoretical Population Biology*, *56*(1), 29–47. https://doi.org/10.1006/TPBI.1998.1408

Meng, G., & Mei, H. (2019). Transcriptional Dysregulation Study Reveals a Core Network Involving the Progression of Alzheimer's Disease. *Frontiers in Aging Neuroscience*, *11*(MAY). https://doi.org/10.3389/FNAGI.2019.00101

Milanese, A., Mende, D. R., Paoli, L., Salazar, G., Ruscheweyh, H. J., Cuenca, M., Hingamp, P., Alves, R., Costea, P. I., Coelho, L. P., Schmidt, T. S. B., Almeida, A., Mitchell, A. L., Finn, R. D., Huerta-Cepas, J., Bork, P., Zeller, G., & Sunagawa, S. (2019). Microbial abundance, activity and population genomic profiling with mOTUs2. *Nature Communications*, *10*(1), 1014–1014. https://doi.org/10.1038/S41467-019-08844-4

Mollenhauer, B., Trautmann, E., Sixel-Döring, F., Wicke, T., Ebentheuer, J., Schaumburg, M., Lang, E., Focke, N. K., Kumar, K. R., Lohmann, K., Klein, C., Schlossmacher, M. G., Kohnen, R., Friede, T., & Trenkwalder, C. (2013). Nonmotor and diagnostic findings in subjects with de novo Parkinson disease of the DeNoPa cohort. *Neurology*, *81*(14), 1226–1234. https://doi.org/10.1212/WNL.0B013E3182A6CBD5

Mollenhauer, B., Zimmermann, J., Sixel-Döring, F., Focke, N. K., Wicke, T., Ebentheuer, J., Schaumburg, M., Lang, E., Trautmann, E., Zetterberg, H., Taylor, P., Friede, T., &

Trenkwalder, C. (2016). Monitoring of 30 marker candidates in early Parkinson disease as progression markers. *Neurology*, *87*(2), 168. https://doi.org/10.1212/WNL.0000000000002651

Moreira de Gouveia, M. I., Daniel, J., Garrivier, A., Bernalier-Donadille, A., & Jubelin, G. (2023). Diversity of ethanolamine utilization by human commensal Escherichia coli. *Research in Microbiology*, *174*(1–2), 103989. https://doi.org/10.1016/J.RESMIC.2022.103989

Narayanasamy, S., Jarosz, Y., Muller, E. E. L., Heintz-Buschart, A., Herold, M., Kaysen, A., Laczny, C. C., Pinel, N., May, P., & Wilmes, P. (2016). IMP: A pipeline for reproducible reference-independent integrated metagenomic and metatranscriptomic analyses. *Genome Biology*, *17*(1), 1–21. https://doi.org/10.1186/S13059-016-1116-8/FIGURES/5

Oksanen, O., Blanchet, F. G., & Kindt, R. (2016). *Community Ecology Package [R package vegan version 2.6-8]*. https://doi.org/10.32614/CRAN.PACKAGE.VEGAN

Qian, Y., Yang, X., Xu, S., Wu, C., Song, Y., Qin, N., Chen, S. Di, & Xiao, Q. (2018). Alteration of the fecal microbiota in Chinese patients with Parkinson's disease. *Brain, Behavior, and Immunity*, *70*, 194–202. https://doi.org/10.1016/J.BBI.2018.02.016

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, *41*(Database issue), D590. https://doi.org/10.1093/NAR/GKS1219

Queirós, P., Delogu, F., Hickl, O., May, P., & Wilmes, P. (2021). Mantis: flexible and consensus-driven genome annotation. *GigaScience*, *10*(6), 1–14. https://doi.org/10.1093/GIGASCIENCE/GIAB042

Ravcheev, D. A., & Thiele, I. (2014). Systematic genomic analysis reveals the complementary aerobic and anaerobic respiration capacities of the human gut

673 microbiota. *Frontiers in Microbiology*, *5*(DEC).
674 https://doi.org/10.3389/FMICB.2014.00674

675 Reichardt, N., Duncan, S. H., Young, P., Belenguer, A., McWilliam Leitch, C., Scott, K. P.,
676 Flint, H. J., & Louis, P. (2014). Phylogenetic distribution of three pathways for
677 propionate production within the human gut microbiota. *The ISME Journal 2014 8:6*,
678 *8*(6), 1323–1335. https://doi.org/10.1038/ismej.2014.14

679 Rezaie, N., Reese, F., & Mortazavi, A. (2023). PyWGCNA: a Python package for weighted
680 gene co-expression network analysis. *Bioinformatics*, *39*(7).
681 https://doi.org/10.1093/BIOINFORMATICS/BTAD415

682 Romano, S., Savva, G. M., Bedarf, J. R., Charles, I. G., Hildebrand, F., & Narbad, A. (2021).
683 Meta-analysis of the Parkinson's disease gut microbiome suggests alterations linked to
684 intestinal inflammation. *NPJ Parkinson's Disease*, *7*(1). https://doi.org/10.1038/S41531-
685 021-00156-Z

686 Roume, H., El Muller, E., Cordes, T., Renaut, J., Hiller, K., & Wilmes, P. (2012). A
687 biomolecular isolation framework for eco-systems biology. *The ISME Journal 2013 7:1*,
688 *7*(1), 110–121. https://doi.org/10.1038/ismej.2012.72

689 Roume, H., Heintz-Buschart, A., Muller, E. E. L., May, P., Satagopam, V. P., Laczny, C. C.,
690 Narayanasamy, S., Lebrun, L. A., Hoopmann, M. R., Schupp, J. M., Gillece, J. D.,
691 Hicks, N. D., Engelthaler, D. M., Sauter, T., Keim, P. S., Moritz, R. L., & Wilmes, P.
692 (2015). Comparative integrated omics: identification of key functionalities in microbial
693 community-wide metabolic networks. *NPJ Biofilms and Microbiomes*, *1*, 15007.
694 https://doi.org/10.1038/NPJBIOFILMS.2015.7

695 Rowley, C. A., Anderson, C. J., & Kendall, M. M. (2018). Ethanolamine Influences Human
696 Commensal Escherichia coli Growth, Gene Expression, and Competition with
697 Enterohemorrhagic E. coli O157:H7. *MBio*, *9*(5). https://doi.org/10.1128/MBIO.01429-18

Sateia, M. J. (2014). International classification of sleep disorders-third edition: highlights and modifications. *Chest*, *146*(5), 1387–1394. https://doi.org/10.1378/CHEST.14-0970

Schenck, C. H., Montplaisir, J. Y., Frauscher, B., Hogl, B., Gagnon, J. F., Postuma, R., Sonka, K., Jennum, P., Partinen, M., Arnulf, I., Cochen de Cock, V., Dauvilliers, Y., Luppi, P. H., Heidbreder, A., Mayer, G., Sixel-Döring, F., Trenkwalder, C., Unger, M., Young, P., … Oertel, W. (2013). Rapid eye movement sleep behavior disorder: devising controlled active treatment studies for symptomatic and neuroprotective therapy--a consensus statement from the International Rapid Eye Movement Sleep Behavior Disorder Study Group. *Sleep Medicine*, *14*(8), 795–806. https://doi.org/10.1016/J.SLEEP.2013.02.016

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, *30*(14), 2068–2069. https://doi.org/10.1093/BIOINFORMATICS/BTU153

Stiasny-Kolster, K., Mayer, G., Schäfer, S., Möller, J. C., Heinzel-Gutenbrunner, M., & Oertel, W. H. (2007). The REM sleep behavior disorder screening questionnaire--a new diagnostic instrument. *Movement Disorders : Official Journal of the Movement Disorder Society*, *22*(16), 2386–2393. https://doi.org/10.1002/MDS.21740

Stolfi, C., Maresca, C., Monteleone, G., & Laudisi, F. (2022). Implication of Intestinal Barrier Dysfunction in Gut Dysbiosis and Diseases. *Biomedicines*, *10*(2). https://doi.org/10.3390/BIOMEDICINES10020289

Toh, T. S., Chong, C. W., Lim, S. Y., Bowman, J., Cirstea, M., Lin, C. H., Chen, C. C., Appel-Cresswell, S., Finlay, B. B., & Tan, A. H. (2022). Gut microbiome in Parkinson's disease: New insights from meta-analysis. *Parkinsonism & Related Disorders*, *94*, 1–9. https://doi.org/10.1016/J.PARKRELDIS.2021.11.017

Torkamani, A., & Schork, N. J. (2009). Identification of rare cancer driver mutations by network reconstruction. *Genome Research*, *19*(9), 1570–1578. https://doi.org/10.1101/GR.092833.109

Tsoy, O., Ravcheev, D., & Mushegian, A. (2009). Comparative genomics of ethanolamine utilization. *Journal of Bacteriology*, *191*(23), 7157–7164. https://doi.org/10.1128/JB.00838-09

Vance, J. E. (2018). Historical perspective: phosphatidylserine and phosphatidylethanolamine from the 1800s to the present. *Journal of Lipid Research*, *59*(6), 923–944. https://doi.org/10.1194/JLR.R084004

Varrette, S., Cartiaux, H., Peter, S., Kieffer, E., Valette, T., & Olloh, A. (2022). Management of an Academic HPC & Research Computing Facility: The ULHPC Experience 2.0. *ACM International Conference Proceeding Series*, 14–24. https://doi.org/10.1145/3560442.3560445

Visser, M., Marinus, J., Stiggelbout, A. M., & van Hilten, J. J. (2004). Assessment of autonomic dysfunction in Parkinson's disease: the SCOPA-AUT. *Movement Disorders : Official Journal of the Movement Disorder Society*, *19*(11), 1306–1312. https://doi.org/10.1002/MDS.20153

Weiss, G. A., & Hennet, T. (2017a). Mechanisms and consequences of intestinal dysbiosis. *Cellular and Molecular Life Sciences 2017 74:16*, *74*(16), 2959–2977. https://doi.org/10.1007/S00018-017-2509-X

Weiss, G. A., & Hennet, T. (2017b). Mechanisms and consequences of intestinal dysbiosis. *Cellular and Molecular Life Sciences 2017 74:16*, *74*(16), 2959–2977. https://doi.org/10.1007/S00018-017-2509-X

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology*, *20*(1), 1–13. https://doi.org/10.1186/S13059-019-1891-0/FIGURES/2

Wuchty, S., & Stadler, P. F. (2003). Centers of complex networks. *Journal of Theoretical Biology*, *223*(1), 45–53. https://doi.org/10.1016/S0022-5193(03)00071-7

748 Zhang, B., & Horvath, S. (2005). A general framework for weighted gene co-expression
749 network analysis. *Statistical Applications in Genetics and Molecular Biology*, *4*(1).
750 https://doi.org/10.2202/1544-6115.1128/MACHINEREADABLECITATION/RIS

751

Paper 3: Functional Prediction of Proteins from the Human Gut Archaeome

Contribution

- As the first author of the paper, I participated in conceptualization of the study, data extraction, data preparation, bioinformatical analysis, data interpretation, visualization of results, and manuscript writing.
- Responsible for the creation of all visual elements of the manuscript.

Overview of the study

In our study titled “Functional Prediction of Proteins from the Human Gut Archaeome”, we undertake a comprehensive exploration of the underrepresented domain of archaea within the human gut microbiome. By employing advanced computational and structural methodologies, this research delineates the functional capacities of archaeal proteins, thereby providing a nuanced understanding of their roles within the gut ecosystem. This study addresses a critical domain in microbiome research, specifically the functional annotation of archaeal proteins, which has remained elusive due to the limited availability of homologous sequences in extant databases.

The human gastrointestinal tract harbors a diverse array of microbial communities, including a significant presence of archaea. Among these, *Methanobrevibacter smithii* emerges as a highly active and clinically relevant methanogenic archaeon, implicated in various gastrointestinal disorders such as inflammatory bowel disease and obesity. Our study presents an integrated approach to enhance the annotation of *M. smithii* proteins by leveraging advanced protein structure prediction and annotation tools, including AlphaFold2, trRosetta, ProFunc, and DeepFri.

In this research, we utilized an extensive dataset of archaeal proteins, from which a subset was identified as exclusive to the human gut. This dataset was further analyzed alongside bacterial proteins to discern unique archaeal proteins and archaeal-bacterial homologs. The study’s methodology involved predicting and characterizing the functional domains and structures of unique and homologous archaeal protein clusters associated with the human gut and *M. smithii*. This approach facilitated the refinement of existing sequence similarity-based annotations through the integration of predicted structural data.

One of the notable findings of this study was the identification of gut-specific archaeal proteins potentially involved in defense mechanisms, virulence, adhesion, and the degradation of toxic substances. The study also uncovered potential glycosyltransferases that could be linked to N-linked and O-glycosylation processes. Additionally, preliminary evidence suggested interdomain horizontal gene transfer between *Clostridia* species and *M. smithii*, including sporulation *Stage V proteins AE* and *AD*.

The implications of these findings may broaden the understanding of archaeal biology, particularly concerning *M. smithii*. The study underscores the importance of considering both sequence and structure for accurate protein function prediction. By integrating advanced computational tools, the research aims to provide a more comprehensive

annotation of archaeal proteins, which is crucial for elucidating their roles in the human gut microbiome.

Furthermore, the study highlights the evolutionary significance of archaea within the human gut. Historically, archaea were primarily associated with extreme environments; however, their presence and functional roles in more moderate environments, such as the human gut, have garnered increasing attention. The ability of archaea to thrive in diverse environments and resist various chemical stresses is partly attributed to their unique cell envelope structures. In natural ecosystems, archaea perform distinctive biogeochemical functions, such as methanogenesis, anaerobic methane oxidation, and ammonia oxidation.

In conclusion, the study's integrated approach to protein annotation, combining sequence and structure information, represents a significant contribution to the field of microbial genomics. The findings aim to enhance the understanding of the role of *M. smithii* in the human gut and pave the way for future research into the functional dynamics of archaeal proteins. This comprehensive annotation framework can be applied to other microbial communities, thereby contributing to a broader understanding of microbial ecology and evolution.

ISME Communications, 2024, 4(1), ycad014
https://doi.org/10.1093/ismeco/ycad014
Advance access publication: 10 January 2024
Original Article

Functional prediction of proteins from the human gut archaeome

Polina V. Novikova[1,*], Susheel Bhanu Busi[1,2], Alexander J. Probst[3], Patrick May[4], Paul Wilmes[1,5,*]

[1]Systems Ecology, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette L-4362, Luxembourg
[2]UK Centre for Ecology and Hydrology, Wallingford, OX10 8 BB, United Kingdom
[3]Environmental Metagenomics, Department of Chemistry, Research Center One Health Ruhr of the University Alliance Ruhr, for Environmental Microbiology and Biotechnology, University Duisburg-Essen, Duisburg 47057, Germany
[4]Bioinformatics Core, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette L-4362, Luxembourg
[5]Department of Life Sciences and Medicine, Faculty of Science, Technology and Medicine, University of Luxembourg, Esch-sur-Alzette L-4362, Luxembourg

*Corresponding authors: Paul Wilmes and Polina V. Novikova, Systems Ecology, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette L-4362, Luxembourg. Email: paul.wilmes@uni.lu, polina.novikova@uni.lu

Abstract

The human gastrointestinal tract contains diverse microbial communities, including archaea. Among them, *Methanobrevibacter smithii* represents a highly active and clinically relevant methanogenic archaeon, being involved in gastrointestinal disorders, such as inflammatory bowel disease and obesity. Herein, we present an integrated approach using sequence and structure information to improve the annotation of *M. smithii* proteins using advanced protein structure prediction and annotation tools, such as AlphaFold2, trRosetta, ProFunc, and DeepFri. Of an initial set of 873 481 archaeal proteins, we found 707 754 proteins exclusively present in the human gut. Having analysed archaeal proteins together with 87 282 994 bacterial proteins, we identified unique archaeal proteins and archaeal–bacterial homologs. We then predicted and characterized functional domains and structures of 73 unique and homologous archaeal protein clusters linked the human gut and *M. smithii*. We refined annotations based on the predicted structures, extending existing sequence similarity-based annotations. We identified gut-specific archaeal proteins that may be involved in defense mechanisms, virulence, adhesion, and the degradation of toxic substances. Interestingly, we identified potential glycosyltransferases that could be associated with *N*-linked and *O*-glycosylation. Additionally, we found preliminary evidence for interdomain horizontal gene transfer between *Clostridia* species and *M. smithii*, which includes *sporulation Stage V proteins AE and AD*. Our study broadens the understanding of archaeal biology, particularly *M. smithii*, and highlights the importance of considering both sequence and structure for the prediction of protein function.

Keywords: protein structure, archaea, methanogens, gut microbiome

Introduction

In 1977, Woese and Fox, and colleagues discovered the kingdom of *Archaebacteria*, later renamed Archaea, revealing a new branch in the tree of life [1-4]. The discovery of the *Asgard* superphylum and its close relationship with the eukaryotic branch supports the notion of an archaeal origin for eukaryotes, yet ongoing debates continue regarding whether the archaeal ancestor of eukaryotes belongs within the *Asgard* superphylum or represents a sister group to all other archaea [5, 6]. Historically, archaea were associated with extreme environments but have since been recognized for their general importance and prevalence [7, 8]. Their ability to thrive in extreme environments and to resist chemicals is attributed, in part, to their unique cell envelope structures. In nature, archaea perform distinctive biogeochemical functions, such as methanogenesis, anaerobic methane oxidation, and ammonia oxidation [9, 10]. By employing diverse ecological strategies for energy production, archaea can inhabit a wide variety of environments [11]. Archaea are also host-associated, such as on plants, in human and animal gastrointestinal tracts [12, 13], on human skin [14, 15], in respiratory airways [16], and in the oral cavity [17]. Based on recent estimates, archaea comprise up to 10% of the human gut microbiota [18].

Methanobrevibacter smithii, a ubiquitous and active methanogen in the human gut microbiome, has remarkable clinical relevance and is relatively well annotated [19]. It plays an important role in the degradation of complex carbohydrates, leading to the production of methane, which has significant physiological effects on human physiology. Imbalances in the population of *M. smithii* have been implicated as factors contributing to gastrointestinal disorders such as inflammatory bowel disease (IBD) [20, 21] and obesity [22-24]. Given the prevalence of *M. smithii* in the gut, further research aimed at *M. smithii* is key to understanding their role in disease. Archaeal proteins, including those of *M. smithii*, play a crucial role in adapting to diverse environments and showcase their unique biology. The knowledge about diverse archaea, including novel species, in the human gut microbiome has expanded, underscoring their significance [25]. Some host-associated taxa, like *Methanomassiliicoccales*, have potential beneficial effects on human health [26], while others like *Methanosphaera stadtmanae* have been linked to proinflammatory immune processes [27]. Given the current interest in the role of archaea in

Received 12 December 2023. **Revised:** 16 December 2023. **Accepted:** 19 December 2023

human health and disease, understanding the archaeal proteome is crucial for understanding the functional potential of archaea.

Studying archaeal proteins presents challenges both in experimental and computational aspects. Previous research has highlighted the potential for biotechnological applications in various archaeal genera [28]. However, genetic toolboxes for targeted genomic modifications are currently limited to mesophilic *Methanococcus* and *Methanosarcina* genera [29]. Although alternative methods like mass spectrometry-based searches exist, difficulties arise from inaccurate predictions of protein coding sequences (CDSs) due to limited knowledge of ribosomal binding sites and promoter consensus sequences [30]. Another unresolved challenge lies in the isolation and cultivation of archaea under laboratory conditions, although recent progress has been made [31, 32]. To overcome these challenges, metagenomic sequencing has emerged as a promising approach to study archaea and their ecological relationships. Metagenomics has enhanced our understanding of the archaeal branches within the tree of life [31-33], whereby assembled sequences allow prediction of protein CDSs and their functional characterization *in silico*. However, metagenome-assembled genomes (MAGs) face challenges in functional assignment due to incomplete sequences and difficulties in predicting and annotating open-reading frames (ORFs) [34, 35]. Sequence-based protein function annotation, commonly used but limited in cases of distant protein homologies, proves to be not particularly effective [36]. Moreover, the databases containing information about archaeal proteins and functions are not consistently updated, creating a 2-fold challenge in the sequence-based annotation of archaeal proteins. On one hand, Makarova *et al.* [37] report that *archaeal ribosomal proteins L45 and L47*, experimentally identified in 2011 [38], and pre-rRNA processing and ribosome biogenesis proteins of the *NOL1/NOP2/fmu* family, characterized in 1998 [39], were not added to annotation pipelines by 2019 and were labelled as "hypothetical." On the other hand, sequence similarity–based approaches fail to capture relationships between highly divergent proteins when aligned with a known database protein [40-42]. Archaea, the least characterized domain of life, suffer from incorrect protein annotations due to insufficient experimental data and outdated databases [43]. Furthermore, the study by Makarova *et al.* indicates that a substantial proportion of genes within archaeal genomes (30%–80%) have not been thoroughly characterized, leading to their classification as archaeal "dark matter" [37]. Poorly annotated proteins limit our study of microbial functionality and their roles in biological processes. However, protein structure prediction represents an alternative strategy addressing the gap in sequence–function annotation [44]. It complements sequence-based approaches, particularly when annotations are limited or conflicting across databases, by utilizing the conservation of tertiary structure to infer functional roles [45, 46]. Advanced computational techniques, such as AlphaFold2 (AF) [47] and trRosetta (TR) [48], offer accurate predictions of 3D structures, providing valuable functional insights.

Here, we present an integrated *in silico* approach to enhance protein functional characterization and improve accuracy of protein annotations in archaeon *M. smithii*. Having compared archaeal gut–specific proteins to bacterial gut proteins, we found 73 unique and homologous archaeal protein clusters. Our approach incorporates advanced protein structure prediction and annotation tools, such as AlphaFold2 (AF), trRosetta (TR), ProFunc (PF), and DeepFri (DF), into a comprehensive workflow. We predict and characterize the functional domains and structures of 73 gut-specific archaeal protein clusters. The predicted functions are linked to the adaptation to changing environments, survival, and nutritional capabilities of *M. smithii* within the human gut microbiome. We additionally identified sporulation-related archaeal proteins, presumably horizontally transferred to archaea from *Clostridium* species.

Materials and methods

Selection of gut-specific archaeal proteins

To select specific proteins of gut-associated archaea, we utilized archaeal MAGs obtained from the Genomes from Earth's Microbiomes (GEM) catalog [49] and the Unified Human Gastrointestinal Genome (UHGG) collection [50], along with bacterial MAGs from the UHGG collection (accessed in November 2020). Genomes were extracted based on available metadata and filtered by taxonomy to specifically target archaea.

Gene prediction was performed using Prodigal (V2.6.3) [51] on the archaeal and bacterial MAGs from the UHGG collection, while CDSs from the GEM catalog were downloaded from the provided source (https://portal.nersc.gov/GEM). Archaeal and bacterial proteins were further separately clustered using MMseqs2 (MM2) (v12.113e3-2) [52, 53] (Fig. 1) with the following parameters: –cov-mode 0 –min-seq-id 0.9 -c 0.9.

To identify unique functions of gut-associated archaea, we selected proteins specific to the human gut and encoded by gut-associated archaea. MAGs were selected based on available metadata indicating their sampling location. First, we included protein clusters containing at least one protein from a MAG sampled in the human gut. We then excluded protein clusters that had proteins from MAGs sampled in other environments. The final selection included protein clusters where all proteins were encoded by MAGs sampled exclusively from the human gut.

From the selected gut-specific protein clusters, only those with complete KEGG annotations were included. Fully annotated archaeal and bacterial MM2 clusters were additionally clustered together with Sourmash (v4.0.0) [54, 55]. Archaeal protein clusters were categorized into two groups: those sharing KEGG Orthology identifiers (KOs) with bacterial proteins (prefix *h*) and those with unique KOs (prefix *u*) (Fig. 1).

Protein function annotation

Archaeal and bacterial proteins were annotated with KEGG orthologs (KOs) using Mantis (1.5.4) [56] (Fig. 1). AF [47, 57] and TR [48] were used as structure prediction tools. For each tool, the predicted protein structure was then annotated separately. The TR-based model was annotated using templates with the highest *identity* and *coverage* features. TR used a template for prediction if it met the criteria of confidence >0.6, *E*-value <0.001, and coverage >0.3. The protein model generated by AF was submitted to the PF [58] web server for structure-based annotation. "Sequence search vs existing PDB entries" and 3D functional template searches sections from the PF report were used for structure-based protein annotation. Structure matches were selected according to the reported highest possible likelihood of being correct as follows: *certain matches* (*E*-value < 10–6), *probable matches* (10–6 < *E*-value < 0.01), *possible matches* (0.01 < *E*-value < 0.1), and *long shots* (0.1 < *E*-value < 10.0). Only certain matches were used for the functional assignment. DeepFri [59] was used as an auxiliary tool, providing broad and general descriptions to verify or refute suggestions from AF and/or TR. DeepFri predictions with a certainty score > 0.7 were considered. Our combined approach integrates multiple methods to enhance the resolution of functional

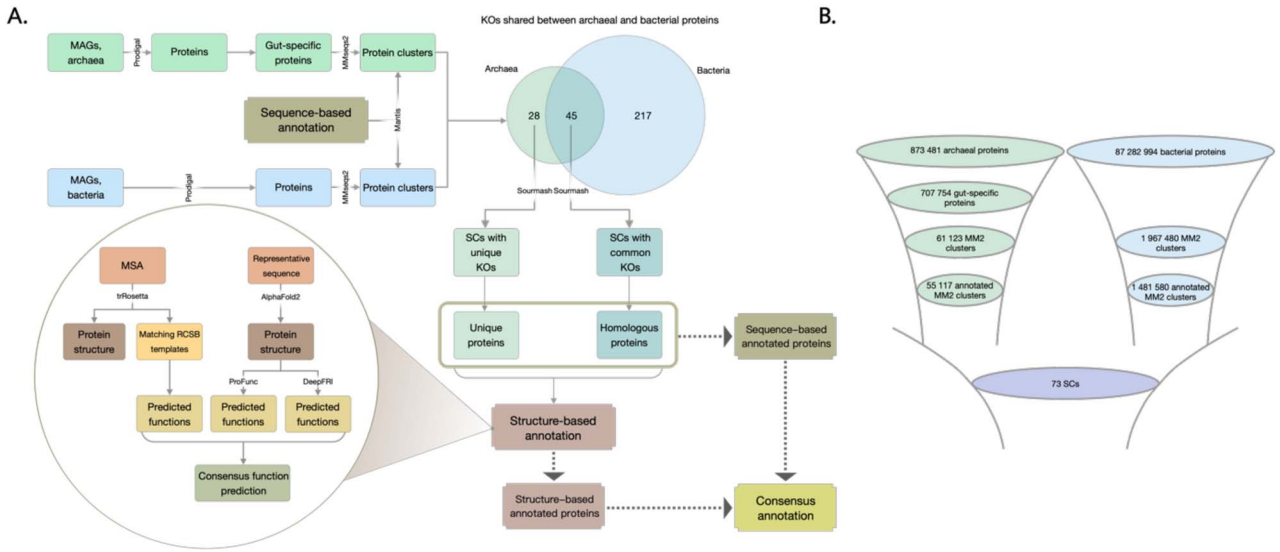


Figure 1. (A) Flowchart demonstrating major steps of the analysis; the Venn diagram demonstrates the number of shared KOs assigned to archaeal and bacterial sourmash clusters; (B) funnels illustrating the protein count at each stage of protein selection; MM2, MMseqs2 clusters; SCs, sourmash clusters.

Table 1. Relationships between PF likelihood and TR TM-scores.

PF likelihood	PF *E*-value	TR significance score	TR TM-score
Certain match	$<10^{-6}$	Very high	>0.7
Probable match	<0.01	High	>0.5
Possible match	<0.1	Medium	>0.4
Long shot	<10	Low	>0.3

annotation, particularly for challenges faced by traditional methods.

When TR- and AF-based annotations provided consistent results, the consensus was used as the final annotation of the protein function. However, when the reports gave different results, we prioritized the result with highest confidence. For instance, when the confidence of the model predicted by TR was *very high* and template matches were provided, and AF-based PF reported a match with a lower confidence (anything but *certain match*), the template hit by TR was used as the primary source for the annotation. The relationship between PF likelihood and TR template modeling scores (TM-scores) generated in our analysis is shown in Table 1. Similarly, any protein with a TR template match was considered as more reliable than an annotation with the "long shot" likelihood. In cases where there were no 3D functional hits, TR annotation was given priority. In cases when PF and TR provided annotations with the same level of significance/likelihood, the protein structure with highest *coverage* and *identity* was chosen. Here, we define coverage as *coverage* feature in TR and the ratio $\frac{\text{longest fitted segment}}{\text{query sequence length}}$ as in PF, and for identity, we take *identity* as in TR and *percentage sequence identity* as in PF.

The appropriateness of an annotation was determined based on the extent to which the assigned function of a protein was found to be directly relevant to archaea and supported by relevant literature. Any other annotations were classified as incorrect. Following this initial step, *sensitivity* was calculated as $\textbf{sensitivity} = \frac{N_{str}}{N_{str}+N_{seq}}$, *specificity* as $\textbf{specificity} = \frac{N_{seq}}{N_{seq}+N_{str}}$, *positive likelihood ratio* as $PLR = \frac{sensitivity}{1-specificity}$, negative likelihood ratio as $NLR = \frac{1-sensitivity}{specificity}$, where N_{seq} and N_{str} are the numbers of correct sequence- and structure-based annotations, respectively.

Protein relative occurrence calculation

Relative occurrence or frequency of protein functions in the groups of unique and homologous proteins was calculated. The measure was calculated as the ratio of the number of proteins with a specific KO to the total number of proteins of bacterial or archaeal proteins. For example, the relative occurrence of unique archaeal proteins annotated as *K20411* (sourmash Cluster 1) is $\frac{N_{select}}{N_{total}} * 10^6$, where N_{select} is the amount of proteins annotated with K20411 and N_{total} is the total number of archaeal proteins. The reason for using a constant factor of 10^6 in the equation is to scale the values and generate numbers better suited for graphical representation.

Gene expression analysis

To comprehensively assess the expression of archaeal proteins in the context of human health and disease, gene expression was verified using a dataset, which we previously published, by mapping metatranscriptomic reads of fecal samples of healthy individuals and patients with Type 1 diabetes mellitus (T1DM) [60] to nucleotide sequences of genes of interest using bwa mem [61]. Mapping files were processed with SAMtools (v1.6) [62]. Mosdepth (v0.3.3) [63] was used to calculate mean read coverage per gene of interest.

Horizontal gene transfer analysis

To assess the stability of gene structures in *M. smithii* genomes, we conducted a horizontal gene transfer (HGT) analysis using metaCHIP (v1.10.12) [64] on all *M. smithii* MAGs available in the included datasets. One *Methanobrevibacter_A oralis* MAG derived from UHGG were also included for the comparison of the number of HGT events.

Gene synteny analysis

pyGenomeViz (v0.3.2) [65] was used to build gene synteny for all archaeal genes of interest. Gene coordinates predicted with Prodigal were used as an input. An interval of 10 kb up- and downstream of the gene of interest was selected from the protein predictions. KEGG KOs were allocated based on the sequence-based annotations generated using Mantis [56]. Here, we exclusively focused on *M. smithii*, as our analysis revealed that all the gut-specific proteins encoded by gut-associated archaea were encoded by *M. smithii*, and thus, this taxon was considered representative for our analyses. The *M. smithii*–type strain DSM 861 was used to assess the presence of genes from flanking regions of specific genes in an archaeal culture.

Phylogenetic analysis

To build phylogenetic trees for selective sourmash clusters, additional similar sequences were added from Uniprot [66] using BLAST (v2.0.15.153) [67] with default parameters on the consensus sequences representing sourmash clusters of interest, namely *h9* and *h20*. Furthermore, Uniprot sequences and sourmash cluster sequences were used to build trees. Multiple sequence alignments were built using MAFFT (v7) [68] and trimmed with BMGE (v1.12) [69] using BLOSUM95 similarity matrix and the default cut-off 0.5. Maximum likelihood phylogenetic trees were built with IQ-TREE (v1.6.12) [70] and visualized using the R library *ggtree* (v3.6.2) [71].

Results and discussion

Our study aimed to analyze the gut-specific proteins encoded by *M. smithii* in the human gastrointestinal tract. As we focused on identifying archaeal unique proteins and archaeal–bacterial homologs, we analysed gut-specific archaeal and gut bacterial proteins together. Having compared the two subsets based on their sequence-based annotation, we categorized archaeal gut–specific proteins into two groups: unique and homologous proteins. To annotate them, we used KEGG KOs due to their consistent functional annotations across organisms and widespread usage. For structure-based functional assignment, we utilized a combination of structure prediction and annotation tools (Fig. 1), leveraging the higher prediction accuracy of AlphaFold2 and the rapid and accurate *de novo* predictions obtained via TR. Our central goal is to enhance the accuracy and reliability of protein structure predictions through the integration of these two approaches. Utilizing representative sequences of unique and homologous proteins, AF produced protein structures, and subsequent functional annotations were accomplished by integrating PF and DeepFRI. TR was employed to predict structures of unique and homologous proteins showing detectable homologous matches in the Protein Data Bank, which were subsequently used for further structure annotation.

It is important to note that our methodology includes semi-manual tools, making it most suitable for a limited number of select proteins. The primary design intent of our workflow was to facilitate the further refinement of functions for specific proteins of interest. Although alternative tools such as ESMFold [72] or EMBER3D [73] are available and hold promise for augmenting the potential of the described pipeline, our approach remains specialized and well-suited for in-depth protein analysis.

Enhancing annotations of proteins encoded by *M. smithii*

To explore the uncharted functional space of *M. smithii*, we first selected gut-specific proteins of gut-associated archaea. We collected the encoded proteins of a total of 1190 archaeal and 285 835 bacterial MAGs, resulting in 873 481 archaeal proteins and 87 282 994 bacterial proteins (Fig. 1). We focused on proteins associated with archaea of the human gut microbiome, which represented 37% (707 754 proteins) of all predicted archaeal proteins. These proteins were grouped into 61 123 MM2 clusters for archaea (≥2 proteins per cluster) and 1 967 480 MM2 clusters for bacteria (≥10 proteins per cluster). By retaining fully annotated protein clusters, we obtained 55 117 archaeal MM2 clusters and 1 481 580 bacterial MM2 clusters. Using our proposed functional prediction strategy (Fig. 1A), we analyzed the gut-associated archaeal proteins alongside bacterial proteins, resulting in 45 homologous sourmash clusters, *i.e.* shared between archaea and bacteria, and 28 unique sourmash clusters, *i.e.* composed exclusively of archaeal proteins. The bacterial data served as a reference to distinguish unique proteins encoded and transcribed by archaea, as well as archaeal proteins with homologs to bacterial ORFs. A summary of the annotations as well as comparison of annotations by structure-based tools is provided in Supplementary Tables 1–3.

All archaeal proteins from the abovementioned sourmash clusters were classified as *M. smithii*. We thus sought to extend our knowledge of *M. smithii* by exploring functions that could have implications for human health and disease. The investigation of the relative occurrence of identified proteins and their associated processes revealed distinct types of functions in unique and homologous protein clusters (Fig. 2). The most frequently identified functions in the unique sourmash clusters were related to adaptation to changing environments and protection mechanisms, e.g. defense against foreign DNA and oxidative stress, while processes such as RNA and DNA regulation, energy metabolism, and cell wall integrity and maintenance were less represented (Supplementary Table 4). Homologous sourmash clusters showed frequent functions related to adaptation, various protection mechanisms, energy metabolism, and cell structural integrity (Supplementary Table 5). Analysis of fecal metatranscriptomic data confirmed the transcription of the majority of encoded genes, with some unique and homologous genes exhibiting higher expression levels (Fig. 2). Two unique and 19 homologous sourmash clusters with relatively high expression levels were identified, including genes associated with adaptation to changing environments, defense against foreign DNA and oxidative stress, DNA/RNA regulation, and energy metabolism, while the rest were unannotated (Fig. 2).

Our analysis demonstrated disparity in annotations between sequence- and structure-based approaches. Notably, 46% (13 out of 28) and 31% (14 out of 45) of the unique and homologous sourmash clusters, respectively, lacked structure-based annotations, suggesting a reliance on sequence information for their functional annotation thus far. Literature searches suggest that the KEGG annotations may not provide reasonable or meaningful functional assignments for most of these unannotated proteins. For instance, a protein annotated as *mitochondrial import receptor subunit TOM40* by KEGG is predicted to be a *putative intimin/invasin-like protein* based on its structure, which is more relevant in the context of archaeal biology than being a eukaryotic protein involved in mitochondrial protein import. Similarly, a protein annotated as *Endophilin-A*, a eukaryotic protein involved in membrane curvature, shows structural similarity to *PilC, a Type IVa pilus subunit* of a prokaryotic adhesion filament. Although the presence of eukaryotic proteins in archaea is not surprising from an evolutionary perspective, the assignment of a protein to its evolutionary homolog from a different kingdom may not provide precise functional assignment of protein function. Moreover, examining the

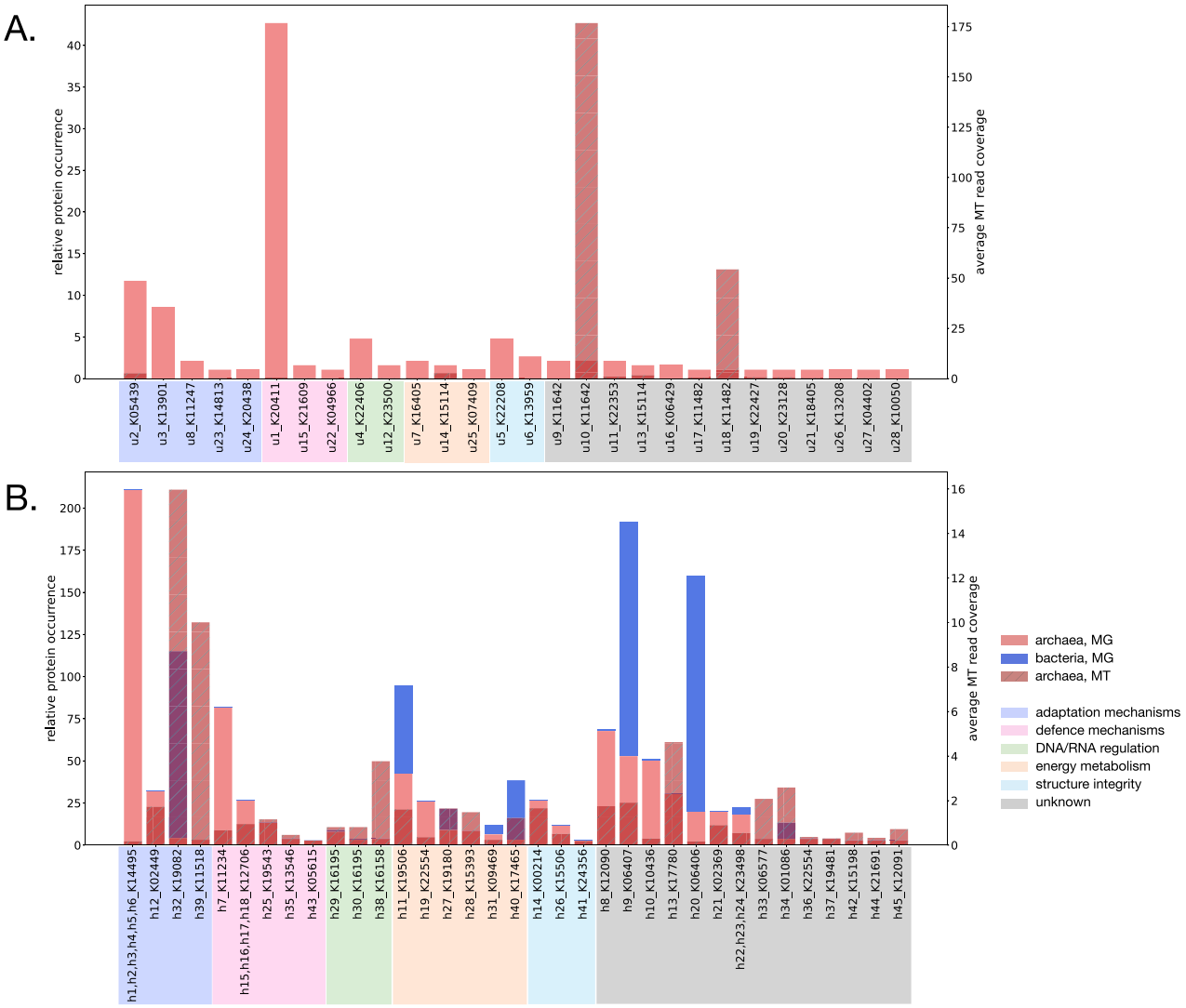


Figure 2. Relative metagenomic occurrence and average metatranscriptomic read coverage of proteins in the (A) unique and (B) homologous groups of clusters with archaeal proteins; MG, metagenomics; MT, metatranscriptomics.

sequence identities between protein clusters annotated through sequence-based methods and the corresponding sequences in UniProt, it is evident that the majority of proteins lack any discernible similarity with those in UniProt. Furthermore, for those instances where some degree of sequence identity is observed, they do not surpass 70% for archaea-specific, unique and 49% for homologous protein clusters (Supplementary Tables 6 and 7).

In general, the agreement between the sequence- and structure-based methods was limited, with 4% (1 out of 28) and 25% (11 out of 45) of the unique and homologous proteins showing consistent annotations, respectively (Supplementary Tables 4–5 and 8). The rest of the proteins exhibited disparity between sequence- and structure-based annotations, which was assessed by comparing their reported functions. For example, unique sourmash cluster *u24* yielded different annotations using EGGNOG, KEGG, and Pfam databases, which we used to potentially resolve disparities in the annotations (Supplementary Table 4). However, a consensus structure-based annotation identified it as *polypeptide N-acetylgalactosaminyltransferase*, providing additional annotation beyond sequence analysis. Similarly, the homologous protein clusters *h15*–*h18* had the same functional assignments as *novobiocin biosynthesis protein NovC* using KEGG, but structure-based annotation revealed further distinctions: *h16* and *h18* were classified as members of the *LytR-Cps2A-Psr protein family*, *h15* was annotated as *78 kDa glucose-regulated protein*, and *h17* remained unannotated (Supplementary Table 5). The incorporation of structural information in protein annotation enables the distinction between closely related sequences, offering additional insights into protein function, which highlights the crucial role of structural data in understanding protein functionality. In addition, the observed disparity between sequence and structure-based annotations, coupled with low sequence identities between sequence-based annotations and corresponding UniProt sequences, underscores the complementarity of structure-based methods to the abovementioned approach for protein function annotation.

We further identified glycosyltransferases responsible for N- and O-linked glycosylation from clusters *h1*–*h6* as prevalent archaeal gut-specific proteins. These proteins may contribute to the viability and adaptability of archaeal cells in the gut. For instance, the most prevalent unique archaeal glycosyltransferase is *4-amino-4-deoxy-L-arabinose (L-Ara4N) transferase*, which is essential for the protection from environmental stress, symbiosis, virulence, and resistance against antimicrobial activity [74, 75]. Moreover, one of the six glycosyltransferases is a

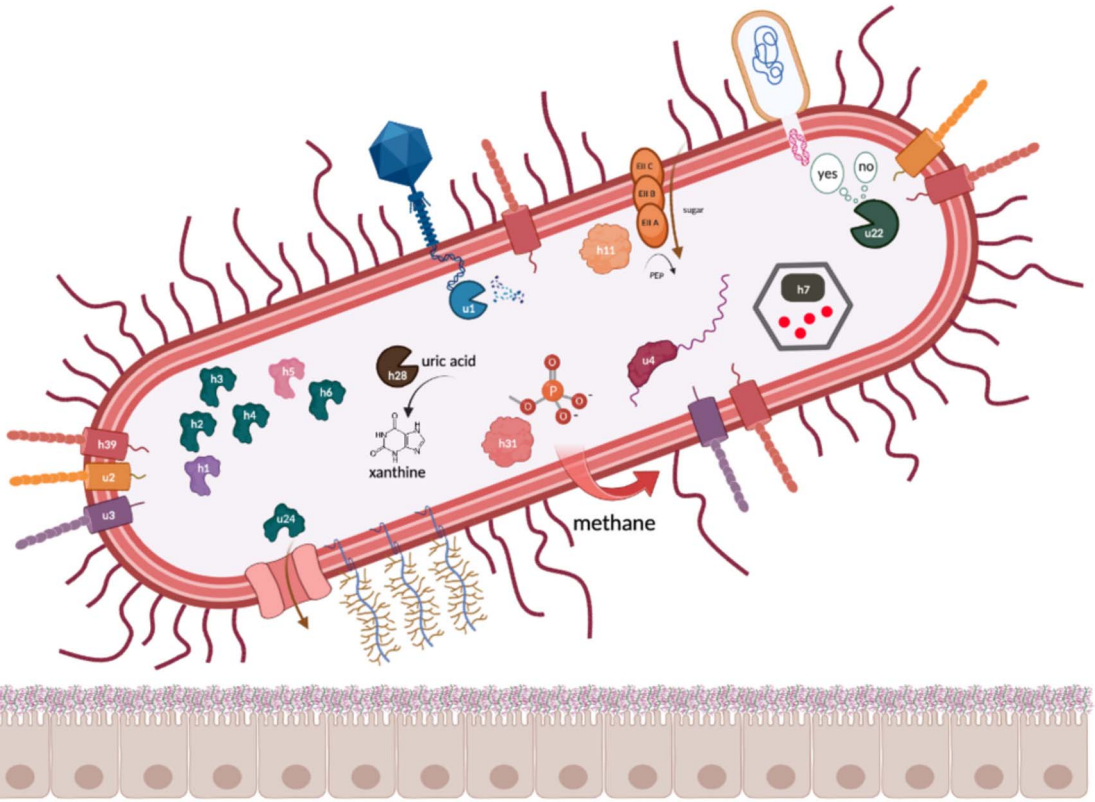


Figure 3. Schematic proposal highlighting proteins specific to gut-associated archaea with described functions: *u1, Type II restriction endonuclease BglII; u2, intimin/invasin-like protein with a Ig-like domain; u3, intimin/invasin-like protein; u4, Unr protein; u22, Type I restriction–modification EcoKI enzyme, specificity subunit; u24, polypeptide N-acetylgalactosaminyltransferase; h1, 4-amino-4-deoxy-L-arabinose transferase or related glycosyltransferases of PMT family; h2,3,4,6, dolichyl-phosphate-mannose–protein mannosyltransferase 1; h5, dolichyl-diphosphooligosaccharide–protein glycosyltransferase subunit STT3B; h7, Propanediol utilization protein pduA; h11, phosphoenolpyruvate-dependent PTS system, IIA component; h28, transthyretin-like protein; h31, 2-AEP aminotransferase.*

dolichyl-diphosphooligosaccharide–protein glycosyltransferase subunit STT3B (h5), which functions as an accessory protein in *N*-glycosylation and provides its maximal efficiency [76]. Archaeal *N*-glycosylation is known to play an important role in the viability and adaptivity of archaeal cells to external conditions such as high salinity [77], elevated temperatures [78], and an acidic environment [79] while also maintaining the structural integrity of cells [80, 81]. Four out of the six identified glycosyltransferases are *dolichyl-phosphate-mannose–protein mannosyltransferases 1 (POMT1)*, which are responsible for O-linked glycosylation of proteins in eukaryotes. Another *O*-glycosylation–associated protein, *polypeptide N-acetylgalactosaminyltransferase*, was found in the subset of unique archaeal proteins (u24). *M. smithii* has been found to decorate its cellular surface with sugar residues mimicking those present in the glycan landscape of the intestinal environment [82]. The presence of human mucus– and epithelial cell surface–associated glycans in *M. smithii*, along with the coding potential for enzymes involved in O-linked glycosylation in archaeal gut species, suggests that *M. smithii* cells might have the capability to emulate the surfaces of eukaryotic cells in the intestinal mucus. Beyond their structural role in proteins, *O*-glycans can also act as regulators of protein interactions, influencing both interprotein and cell-to-cell communication processes involved in cell trafficking and environmental recognition [83].

Further findings suggest that *2-aminoethylphosphonate-pyruvate (2-AEP) aminotransferase, transthyretin-like protein* and *phosphoenolpyruvate-dependent sugar phosphotransferase system system* encoded by *M. smithii* contribute to energy metabolism. 2-AEP is an enzyme commonly found in bacteria and is known to play a critical role in phosphonate degradation, which serves as an important source and production pathway for methane [84]. Additionally, cold-shock domains of *Unr protein* potentially provide *M. smithii* with adaptation strategies through stress-induced control of gene expression [85]. Furthermore, the predicted involvement of proteins such as the *specificity subunit of Type I restriction–modification EcoKI enzyme* [86] and *Type II restriction endonuclease BglII* [87] suggests their potential role in host defense strategies employed by *M. smithii* to protect themselves in the gut environment. Additionally, it is conceivable that archaeal proteins may play a role in protecting against toxicity from other organisms in the gut using *propanediol utilization protein pduA* [88-90], as well as acquiring genes of bacterial origin through HGT. If this is the case, the presence of adhesin-like proteins in archaea could potentially enable them to form symbiotic relationships with bacterial neighbors with diverse metabolic potentials [91]. Figure 3 provides a schematic representation emphasizing specific proteins identified in this study, which could potentially play a significant role in the functional dynamics of archaea within the human intestine. A more detailed description of all identified *M. smithii* proteins is provided in Supplementary Materials.

Characterization of select proteins and gene structures in *M. smithii* genomes

To elucidate the level of conservation among the identified genes recovered in our analyses, we assessed the level of genomic conservation within genomes of two strains of *M. smithii*, two strains *of Ca. Methanobrevibacter intestini* and the related species *Methanobrevibacter_A oralis* as a reference. *Ca. M. intestini* has been recently classified as an independent species within the *M. smithii* clade. We analysed HGT events and evaluated gene

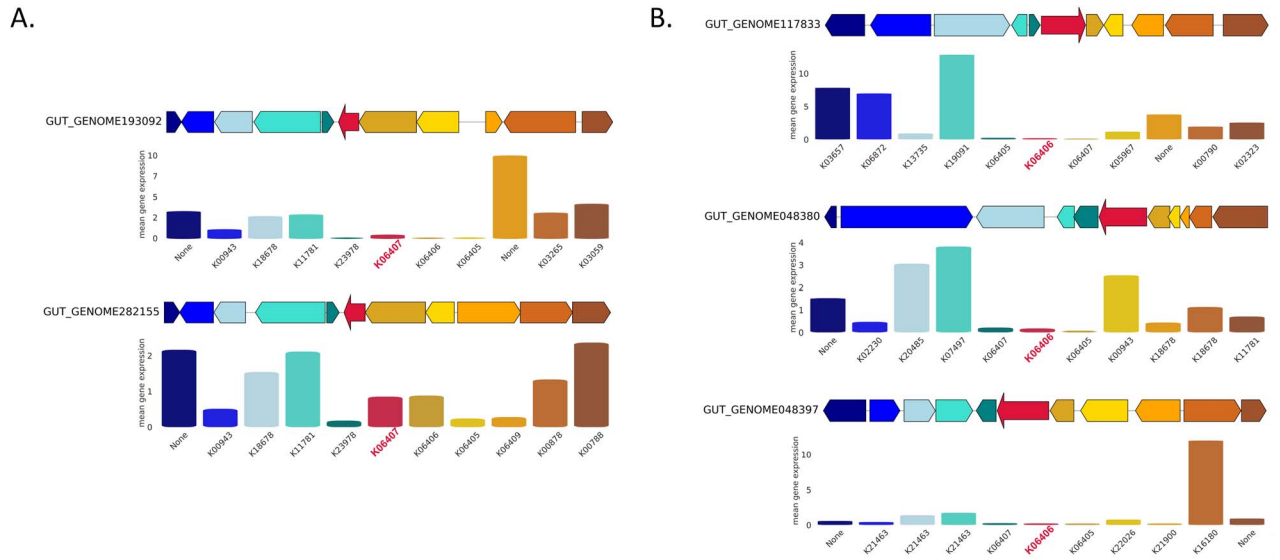


Figure 4. Gene synteny for *sporulation stage V genes AE* and *AD* from their respective sourmash clusters (A) *h9* and (**B**) *h20*; gene expression of target genes (*spoVAE* and *spoVAD*) as well as genes from flanking regions are demonstrated below each sequence and are colored correspondingly. Genes with key archaeal functions: (A) pyrimidine metabolism (K18678, *phytol kinase*), methane metabolism (K11781, *5-amino-6-(D-ribitylamino)uracil–L-tyrosine 4-hydroxyphenyl transferase*), and thiamine metabolism (K00878, *hydroxyethylthiazole kinase*; K00788, *thiamine-phosphate pyrophosphorylase*); (B) pyrimidine metabolism (K22026, *nucleoside kinase*; K18678, *phytol kinase*) and methane metabolism (K11781, *5-amino-6-(D-ribitylamino)uracil–L-tyrosine 4-hydroxyphenyl transferase*).

structure stability. Using 1022 available MAGs, we noted an increase in HGT events between 319 genomes of two *M. smithii* strains: *Methanobrevibacter_A smithii* and *GCF_000016525.1* (based on GTDB classification) (Supplementary Fig. 1). Specifically, 2.6% of the MAGs ($n = 27$) exhibited HGT events involving the transfer of $\sim 10 \pm 3$ genes to other MAGs. Intriguingly, MAGs exhibiting HGT events were sampled in diverse geographical locations such as Austria, France, the UK, and the USA. Our results suggest that the propensity of these MAGs to exchange genomic segments may be attributed to similarities in their respective local environments [92], including dietary and lifestyle factors of the individuals. Thus, it is plausible that exposure to similar diets or stresses may have influenced the evolution of these MAGs via HGT along comparable trajectories. Conversely, the low occurrence of HGT events among the majority (97.4%) of available *M. smithii* genomes indicates their overall genomic conservation and stability. This could be explained by the fact that these MAGs were sampled from individuals living under similar dietary and lifestyle conditions. Importantly, our findings support the concept of genomic stability in *M. smithii*, as we observed a high degree of conservation in the flanking regions of the genes of interest across various *M. smithii* genomes. Through synteny analyses, we found compelling evidence of conserved synteny for genes encoded in *M. smithii* genomes (https://doi.org/10.5281/zenodo.8024791).

Among the proteins specific for gut-associated archaea, we identified *Stage V sporulation proteins AE* (*spoVAE*) and *AD* (*spoVAD*) (*h9* and *h20*). Using BLAST searches, we extracted 250 bacterial protein sequences for SpoVAE and SpoVAD from Uniprot, including 12 *spoVAE* and 38 *spoVAD* proteins from environmental samples and the rest from isolate bacterial genomes belonging to the *Firmicutes* phylum. Phylogenetic trees demonstrated that proteins from *h9* and *h20* are phylogenetically and compositionally distinct from other sequences and form separate branches (Supplementary Figs 2 and 3). Gene synteny analyses revealed that sporulation genes are grouped in operons (K06405, K06406, and K06407; Fig. 4). Moreover, the flanking regions around sporulation genes include genes with key archaeal as well as methanogenic functions. In addition, the flanking regions of both *spoVAE* and *spoVAD* genes are also encoded in the *M. smithii* isolate DSM 861 genome (Fig. 5). This particular isolate served as the representative strain for our research. Furthermore, to further validate the representativeness of DSM 861, we also computed the average nucleotide identity (ANI) between the type strain DSM 861 and two other available strains, DSM 2374 and DSM 2375. The ANI calculations yielded estimates of 98.3 between *M. smithii* strains DSM 861 and DSM 2374, and 98.2 between DSM 861 and DSM 2375, respectively. However, in contrast to our MAGs, the isolate's genome did not encode the *spoVAE* and *spoVAD* genes. To assess whether *spoVAE* and *spoVAD* genes were acquired by *M. smithii* via HGT, we performed synteny analysis of bacterial sequences obtained from our human gut dataset that shared similarities with the archaeal sequences in clusters *h9* and *h20*. This analysis revealed that in the bacterial genomes found in the human intestine, the flanking regions of *spoVAE* and *spoVAD* genes include genes mediating and facilitating HGT, such as a site-specific DNA recombinase (K06400) encoded upstream from *spoVAE* and *Type IV pilus assembly proteins* (K02662, K02664) encoded downstream from *spoVAD* (Supplementary Figs 4 and 5). Genes originating from clusters *h9* and *h20* are found within bacterial genomes of *Firmicutes* phylum members, specifically *Clostridium* sp. CAG-302 and CAG-269, which highlights their association with known bacterial taxa in the gut and indicates HGT between these distantly related taxa.

Although sporulation has been primarily observed in spore-forming bacteria and not in archaea, it is known that non-sporulating bacterial species also encode sporulation genes. In these bacterial taxa, the genes likely encode regulatory proteins involved in peptidoglycan (PPG) turnover, thereby playing a role in cell division and/or development [93, 94]. Archaea lack PPG but methanogenic archaea, including *Methanobrevibacter* species, use pseudopeptidoglycan (pseudo-PPG) instead, which functions similarly to PPG in a bacterial cell and results in Gram-positive staining certain structural similarities between methanogens and bacteria described above leave open the question of whether

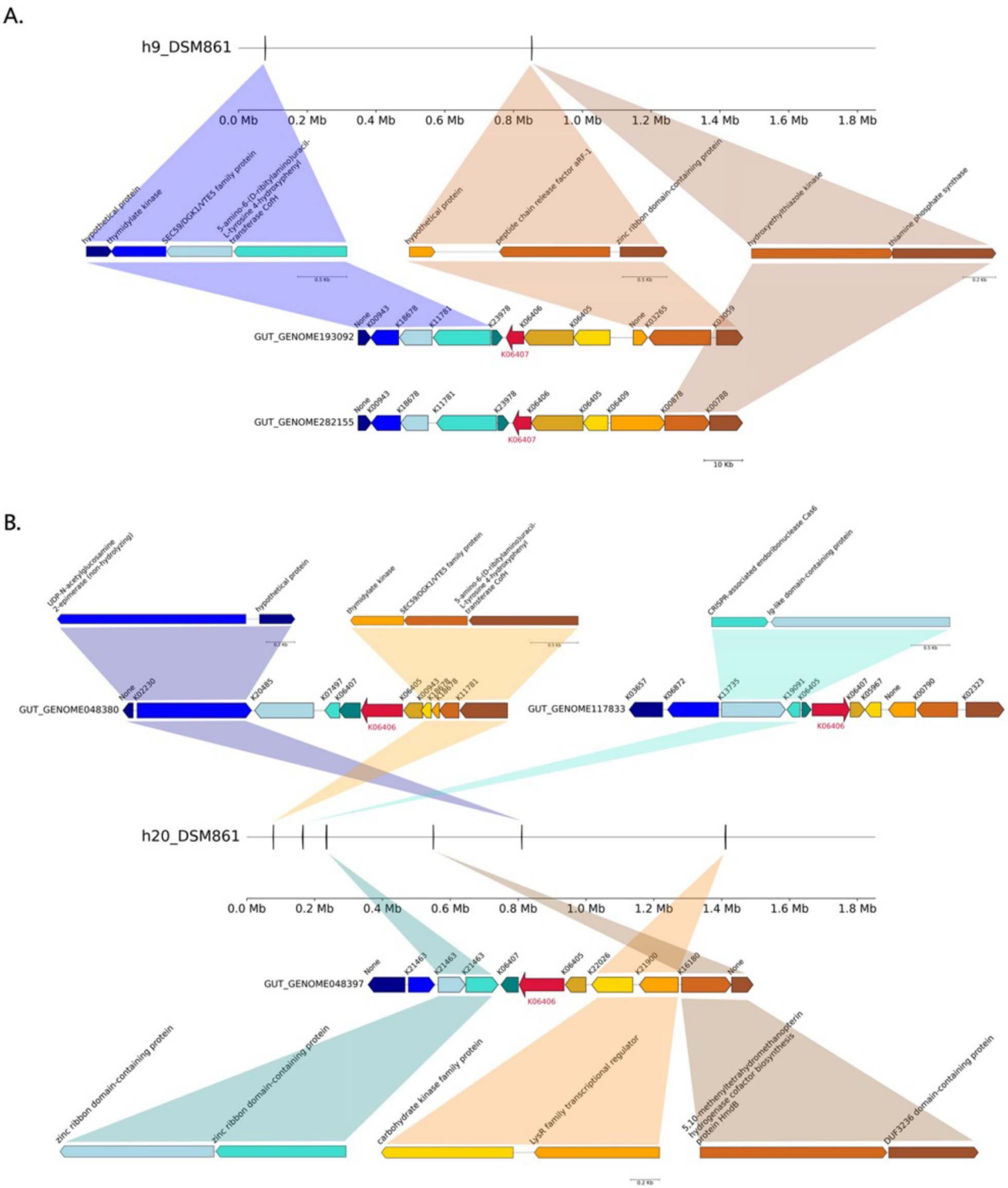


Figure 5. Genomic context of the archaeal flanking regions up- and downstream of the (A) *spoVAE* and (B) *spoVAD* gene clusters in the *M. smithii* strain DSM 861.

sporulation proteins could play a similar role in pseudo-PPG turnover in methanogenic archaea, analogous to their function in non-sporulating bacteria. The identification of these genes holds significant interest, especially in light of the work by Nelson Sathi *et al.*, suggesting that methanogens frequently acquire functionally active genes through horizontal transfer from bacteria. Comprehensive experimental analysis is required to determine their specific functions, but these findings present an exciting opportunity for further exploration. Phylogenetic analysis of *spoVAE* and *spoVAD* has demonstrated that sequences from the abovementioned clusters are compositionally homogeneous but phylogenetically distant from other known similar sequences in Uniprot and therefore might be unique to the human gut environment. Moreover, archaeal and bacterial sequences from sourmash clusters *h9* and *h20* branch out together, which suggests that sporulation genes encoded in archaea might be

the result of HGT from bacteria to archaea. This study provides evidence that archaeal genomes exhibit clustered sporulation genes surrounded by genes linked to archaea-specific functions like pyrimidine, thiamine, and methane metabolism. Moreover, genes in flanking regions up- and downstream of *spoVAE* and *spoVAD* genes are indeed encoded in the representative *M. smithii* isolate DSM 861. The study's intended scope did not include experimental investigations in the wet-lab, such as the application of a protocol using antibiotics, to confirm *M. smithii*'s sporulation capability [95, 96]. Such work represents a logical extension of our reported *in silico* results but goes beyond the scope of the present study. As bacteria encoding similar spoVAE and spoVAD proteins and bacterial sequences from clusters *h9* and *h20* belong to various species of the *Clostridium* genus, HGT probably occurred in the direction from the abovementioned species to *M. smithii*. Moreover, Ruaud, Esquivel-Elizondo, de la Cuesta-Zuluaga *et al.* have provided evidence of a syntrophic relationship between *Firmicutes* bacteria and *M. smithii*. The co-occurrence of these microorganisms is likely facilitated by physical and metabolic interactions. In addition to this, genes *h9* and *h20* as well as their surrounding genes are expressed by the archaeal genomes sampled from human fecal samples.

Conclusion

Our study aimed to uncover the potential functions of archaeal proteins, particularly those encoded by *M. smithii*, in the human gut. Sequence similarity–based methods, while effective for highly similar proteins (>70%–80% identity), may not accurately represent the functions of archaeal proteins due to the lack of experimental validation. More specifically, publicly available databases have limited experimentally validated archaeal sequences compared to bacterial and eukaryotic proteins (~7 000 000 archaeal, ~166 000 000 bacterial, and ~70 000 000 eukaryotic proteins, UniProtKB Jun 2023) making sequence-based protein annotations applicable to only a subset of archaeal proteins. In contrast, recent deep learning-based methods enable protein structure prediction and annotation without relying on high sequence similarity, allowing for functional similarity beyond close sequence matches. We used structural methods to improve the annotation of archaeal proteins, gaining better insights into their functions compared to traditional sequence-based methods. This approach allowed us to refine some existing annotations and discover new functions for others, giving us valuable insights into the roles of archaeal genes in the human gut. Our findings focus on the characterization of human-associated and gut-specific proteins identified in *M. smithii*, a metabolically proficient and clinically relevant methanogenic archaeon known to be linked to gastrointestinal disorders, including IBD and obesity. In upcoming research, the primary focus should be on improving the accuracy of determining translation initiation and termination sites through the integration of additional specialized tools [97, 98], as this holds significant promise for enhancing structural predictions. Furthermore, the refinement of our computational efforts with experimental approaches holds the key to elucidating the predicted protein structures and their corresponding functions.

Acknowledgements

All authors proof-read and approved of the content in this research paper. We are especially grateful for the critical feedback and suggestions provided by Dr Cedric Laczny. Experiments presented in this work were carried out using the HPC facility of the University of Luxembourg.

Supplementary material

Supplementary material is available at *The ISME Journal* online.

Conflicts of interest

None declared.

Funding

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 863664).

Data availability

Microbial MAGs from UHGG collection are available from the MGnify FTP site at http://ftp.ebi.ac.uk/pub/databases/metagenomics/mgnify_genomes/; MAGs from the GEM catalog are accessible at https://portal.nersc.gov/GEM/. Metatranscriptomic sequencing reads are available from NCBI BioProject PRJNA289586 and assembled contigs can be assessed at MG-RAST (submission IDs are indicated in MT_assembly_RAST_ids.xlsx). A description of the analyses including pre-processing steps along with the scripts for the main analysis, archaeal gut–specific unique and homologous sourmash clusters, and synteny plots can be found at GitLab: https://gitlab.lcsb.uni.lu/polina.novikova/archaea-in-gut.

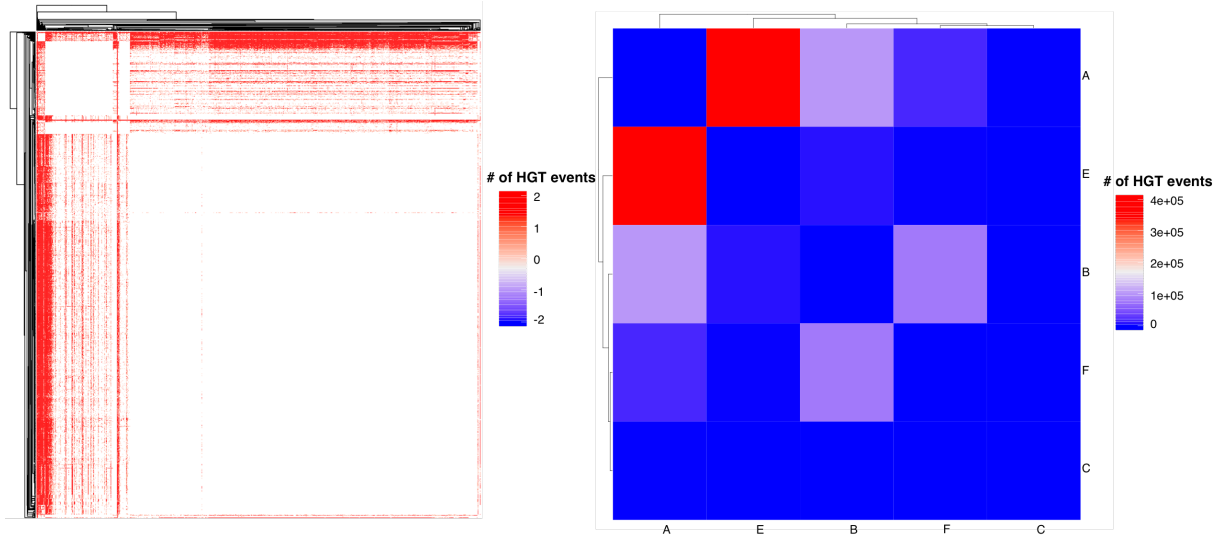
References

1. Woese CR, Fox GE. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* 1977;**74**:5088–90. https://doi.org/10.1073/pnas.74.11.5088
2. Balch WE, Magrum LJ, Fox GE *et al.* An ancient divergence among the bacteria. *J Mol Evol* 1977;**9**:305–11. https://doi.org/10.1007/BF01796092
3. Fox GE, Magrum LJ, Balch WE *et al.* Classification of methanogenic bacteria by 16S ribosomal RNA characterization. *Proc Natl Acad Sci U S A* 1977;**74**:4537–41. https://doi.org/10.1073/pnas.74.10.4537
4. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains archaea, bacteria, and eucarya. *Proc Natl Acad Sci U S A* 1990;**87**:4576–9. https://doi.org/10.1073/pnas.87.12.4576
5. Liu Y, Makarova KS, Huang WC *et al.* Expanded diversity of Asgard archaea and their relationships with eukaryotes. *Nature* 2021;**593**:553–7. https://doi.org/10.1038/s41586-021-03494-3
6. Williams TA, Cox CJ, Foster PG *et al.* Phylogenomics provides robust support for a two-domains tree of life. *Nat Ecol Evol* 2020;**4**:138–47. https://doi.org/10.1038/s41559-019-1040-x
7. Könneke M, Bernhard AE, de la Torre *et al.* Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* 2005;**437**:543–6. https://doi.org/10.1038/nature03911
8. Pester M, Schleper C, Wagner M. The Thaumarchaeota: an emerging view of their phylogeny and ecophysiology. *Curr Opin Microbiol* 2011;**14**:300–6. https://doi.org/10.1016/j.mib.2011.04.007

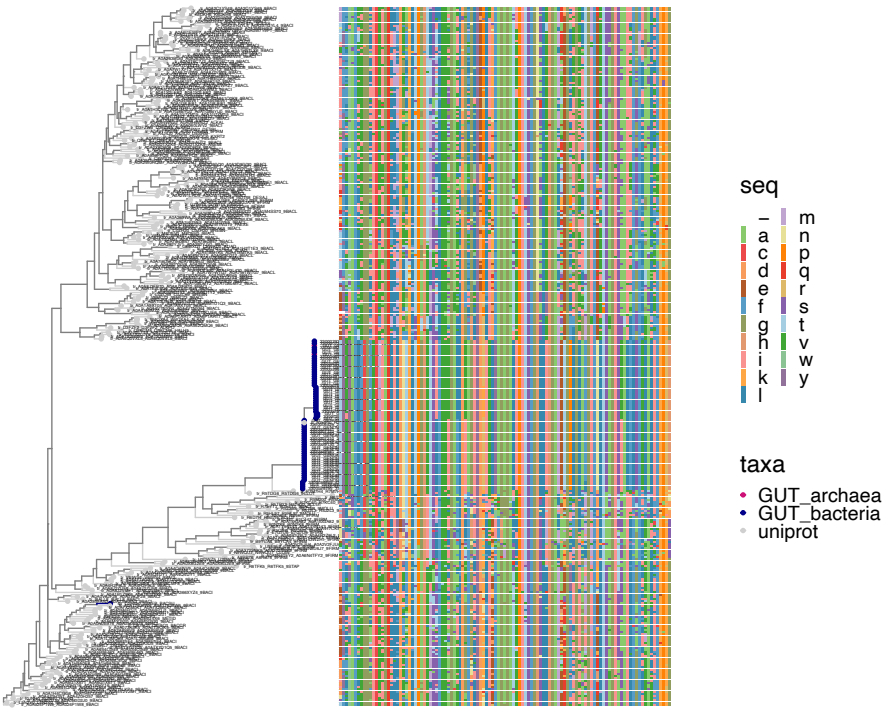
9. DeLong EF. Everything in moderation: archaea as "non-extremophiles". *Curr Opin Genet Dev* 1998;**8**:649–54. https://doi.org/10.1016/s0959-437x(98)80032-4
10. Schleper C. Ammonia oxidation: different niches for bacteria and archaea? *ISME J* 2010;**4**:1092–4. https://doi.org/10.1038/ismej.2010.111
11. Valentine DL. Adaptations to energy stress dictate the ecology and evolution of the archaea. *Nat Rev Microbiol* 2007;**5**:316–23. https://doi.org/10.1038/nrmicro1619
12. Hoegenauer C, Hammer HF, Mahnert A *et al.* Methanogenic archaea in the human gastrointestinal tract. *Nat Rev Gastroenterol Hepatol* 2022;**19**:805–13. https://doi.org/10.1038/s41575-022-00673-z
13. Thomas CM, Desmond-Le Quéméner E, Gribaldo S *et al.* Factors shaping the abundance and diversity of the gut archaeome across the animal kingdom. *Nat Commun* 2022;**13**:3358. https://doi.org/10.1038/s41467-022-31038-4
14. Moissl-Eichinger C, Probst AJ, Birarda G *et al.* Human age and skin physiology shape diversity and abundance of archaea on skin. *Sci Rep* 2017;**7**:4039. https://doi.org/10.1038/s41598-017-04197-4
15. Probst AJ, Auerbach AK, Moissl-Eichinger C. Archaea on human skin. *PLoS One* 2013;**8**:e65388. https://doi.org/10.1371/journal.pone.0065388
16. Kumpitsch C, Koskinen K, Schöpf V *et al.* The microbiome of the upper respiratory tract in health and disease. *BMC Biol* 2019;**17**:87. https://doi.org/10.1186/s12915-019-0703-z
17. Sogodogo E, Doumbo O, Aboudharam G *et al.* First characterization of methanogens in oral cavity in Malian patients with oral cavity pathologies. *BMC Oral Health* 2019;**19**:232. https://doi.org/10.1186/s12903-019-0929-8
18. Kim JY, Whon TW, Lim MY *et al.* The human gut archaeome: identification of diverse haloarchaea in Korean subjects. *Microbiome* 2020;**8**:114. https://doi.org/10.1186/s40168-020-00894-x
19. Eckburg PB, Bik EM, Bernstein CN *et al.* Diversity of the human intestinal microbial flora. *Science* 2005;**308**:1635–8. https://doi.org/10.1126/science.1110591
20. Ghavami SB, Rostami E, Sephay AA *et al.* Alterations of the human gut Methanobrevibacter smithii as a biomarker for inflammatory bowel diseases. *Microb Pathog* 2018;**117**:285–9. https://doi.org/10.1016/j.micpath.2018.01.029
21. Houshyar Y, Massimino L, Lamparelli LA *et al.* Going beyond bacteria: uncovering the role of archaeome and mycobiome in inflammatory bowel disease. *Front Physiol* 2021;**12**:783295. https://doi.org/10.3389/fphys.2021.783295
22. Basseri RJ, Basseri B, Pimentel M *et al.* Intestinal methane production in obese individuals is associated with a higher body mass index. *Gastroenterol Hepatol* 2012;**8**:22–8
23. Samuel BS, Gordon JI. A humanized gnotobiotic mouse model of host–archaeal–bacterial mutualism. *Proc Natl Acad Sci U S A* 2006;**103**:10011–6. https://doi.org/10.1073/pnas.0602187103
24. Mathur R, Amichai M, Chua KS *et al.* Methane and hydrogen positivity on breath test is associated with greater body mass index and body fat. *J Clin Endocrinol Metab* 2013;**98**:E698–702. https://doi.org/10.1210/jc.2012-3144
25. Borrel G, Brugère J-F, Gribaldo S *et al.* The host-associated archaeome. *Nat Rev Microbiol* 2020;**18**:622–36. https://doi.org/10.1038/s41579-020-0407-y
26. Borrel G, McCann A, Deane J *et al.* Genomics and metagenomics of trimethylamine-utilizing archaea in the human gut microbiome. *ISME J* 2017;**11**:2059–74. https://doi.org/10.1038/ismej.2017.72
27. Bang C, Weidenbach K, Gutsmann T *et al.* The intestinal archaea *Methanosphaera stadtmanae* and *Methanobrevibacter smithii* activate human dendritic cells. *PLoS One* 2014;**9**:e99411. https://doi.org/10.1371/journal.pone.0099411
28. Lyu Z, Whitman WB. Transplanting the pathway engineering toolbox to methanogens. *Curr Opin Biotechnol* 2019;**59**:46–54. https://doi.org/10.1016/j.copbio.2019.02.009
29. Thomsen J, Weidenbach K, Metcalf WW *et al.* Genetic methods and construction of chromosomal mutations in methanogenic archaea. *Methods Mol Biol* 2022;**2522**:105–17. https://doi.org/10.1007/978-1-0716-2445-6_6
30. Tebbe A, Klein C, Bisle B *et al.* Analysis of the cytosolic proteome of *Halobacterium salinarum* and its implication for genome annotation. *Proteomics* 2005;**5**:168–79. https://doi.org/10.1002/pmic.200400910
31. Zaremba-Niedzwiedzka K, Caceres EF, Saw JH *et al.* Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 2017;**541**:353–8. https://doi.org/10.1038/nature21031
32. Spang A, Saw JH, Jørgensen SL *et al.* Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* 2015;**521**:173–9. https://doi.org/10.1038/nature14447
33. Castelle CJ, Wrighton KC, Thomas BC *et al.* Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. *Curr Biol* 2015;**25**:690–701. https://doi.org/10.1016/j.cub.2015.01.014
34. Bork P. Powers and pitfalls in sequence analysis: the 70% hurdle. *Genome Res* 2000;**10**:398–400. https://doi.org/10.1101/gr.10.4.398
35. Wooley JC, Godzik A, Friedberg I. A primer on metagenomics. *PLoS Comput Biol* 2010;**6**:e1000667. https://doi.org/10.1371/journal.pcbi.1000667
36. Ellens KW, Christian N, Singh C *et al.* Confronting the catalytic dark matter encoded by sequenced genomes. *Nucleic Acids Res* 2017;**45**:11495–514. https://doi.org/10.1093/nar/gkx937
37. Makarova KS, Wolf YI, Koonin EV. Towards functional characterization of archaeal genomic dark matter. *Biochem Soc Trans* 2019;**47**:389–98. https://doi.org/10.1042/BST20180560
38. Márquez V, Fröhlich T, Armache JP *et al.* Proteomic characterization of archaeal ribosomes reveals the presence of novel archaeal-specific ribosomal proteins. *J Mol Biol* 2011;**405**:1215–32. https://doi.org/10.1016/j.jmb.2010.11.055
39. Wu P, Brockenbrough JS, Paddy MR *et al.* NCL1, a novel gene for a non-essential nuclear protein in *Saccharomyces cerevisiae*. *Gene* 1998;**220**:109–17. https://doi.org/10.1016/s0378-1119(98)00330-8
40. Vakirlis N, Carvunis A-R, McLysaght A, 'Synteny-based analyses indicate that sequence divergence is not the main source of orphan genes', *eLife* 2020;**9**:e53500. https://doi.org/10.7554/eLife.53500.
41. Weisman CM, Murray AW, Eddy SR. Many, but not all, lineage-specific genes can be explained by homology detection failure. *PLoS Biol* 2020;**18**:e3000862. https://doi.org/10.1371/journal.pbio.3000862
42. Sinha S, Lynn AM, Desai DK. Implementation of homology based and non-homology based computational methods for the identification and annotation of orphan enzymes: using *Mycobacterium tuberculosis* H37Rv as a case study. *BMC Bioinformatics* 2020;**21**:466. https://doi.org/10.1186/s12859-020-03794-x
43. Mahnert A, Blohs M, Pausan M-R *et al.* The human archaeome: methodological pitfalls and knowledge gaps. *Emerg Top Life Sci* 2018;**2**:469–82. https://doi.org/10.1042/ETLS20180037
44. Watson JD, Sanderson S, Ezersky A *et al.* Towards fully automated structure-based function prediction in structural genomics: a case study. *J Mol Biol* 2007;**367**:1511–22. https://doi.org/10.1016/j.jmb.2007.01.063

45. Chothia C, Lesk AM. The relation between the divergence of sequence and structure in proteins. *EMBO J* 1986;**5**:823–6. https://doi.org/10.1002/j.1460-2075.1986.tb04288.x
46. Skolnick J, Gao M, Zhou H *et al.* AlphaFold 2: why it works and its implications for understanding the relationships of protein sequence, structure, and function. *J Chem Inf Model* 2021;**61**:4827–31. https://doi.org/10.1021/acs.jcim.1c01114
47. Jumper J, Evans R, Pritzel A *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;**596**:583–9. https://doi.org/10.1038/s41586-021-03819-2
48. Du Z, Su H, Wang W *et al.* The trRosetta server for fast and accurate protein structure prediction. *Nat Protoc* 2021;**16**:5634–51. https://doi.org/10.1038/s41596-021-00628-9
49. Nayfach S, Roux S, Seshadri R *et al.* A genomic catalog of Earth's microbiomes. *Nat Biotechnol* 2021;**39**:499–509. https://doi.org/10.1038/s41587-020-0718-6
50. Almeida A, Nayfach S, Boland M *et al.* A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nat Biotechnol* 2021;**39**:105–14. https://doi.org/10.1038/s41587-020-0603-3
51. Hyatt D, Chen GL, LoCascio PF *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010;**11**:119. https://doi.org/10.1186/1471-2105-11-119.
52. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol* 2017;**35**:1026–8. https://doi.org/10.1038/nbt.3988
53. Steinegger M, Söding J. Clustering huge protein sequence sets in linear time. *Nat Commun* 2018;**9**:2542. https://doi.org/10.1038/s41467-018-04964-5
54. Brown CT, Irber L. Sourmash: a library for MinHash sketching of DNA. *J Open Source Softw* 2016;**1**:27. https://doi.org/10.21105/joss.00027
55. Pierce NT, Irber L, Reiter T *et al.* Large-scale sequence comparisons with sourmash. *F1000Res* 2019;**8**:1006. https://doi.org/10.12688/f1000research.19675.1
56. Queirós P, Delogu F, Hickl O *et al.* Mantis: flexible and consensus-driven genome annotation. *GigaScience* 2021;**10**:giab042. https://doi.org/10.1093/gigascience/giab042
57. Kryshtafovych A, Schwede T, Topf M, Fidelis K, Moult, J. Critical assessment of methods of protein structure prediction (CASP)-Round XIV. *Proteins* 2021;**89**(12):1607–17. https://doi.org/10.1002/prot.26237.
58. Laskowski RA, Watson JD, Thornton JM. ProFunc: a server for predicting protein function from 3D structure. *Nucleic Acids Res* 2005;**33**:W89–93. https://doi.org/10.1093/nar/gki414
59. Gligorijević V, Renfrew PD, Kosciolek T *et al.* Structure-based protein function prediction using graph convolutional networks. *Nat Commun* 2021;**12**:3168. https://doi.org/10.1038/s41467-021-23303-9
60. Heintz-Buschart A, May P, Laczny C et al. Integrated multi-omics of the human gut microbiome in a case study of familial type 1 diabetes. *Nat Microbiol* 2017;**2**:16180. https://doi.org/10.1038/nmicrobiol.2016.180.
61. Li H. *Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM*. 2013, doi: https://doi.org/10.6084/M9.FIGSHARE.963153.V1.
62. Li H, Handsaker B, Wysoker A *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* 2009;**25**:2078–9. https://doi.org/10.1093/bioinformatics/btp352
63. Pedersen BS, Quinlan AR. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* 2018;**34**:867–8. https://doi.org/10.1093/bioinformatics/btx699
64. Song W, Wemheuer B, Zhang S *et al.* MetaCHIP: community-level horizontal gene transfer identification through the combination of best-match and phylogenetic approaches. *Microbiome* 2019;**7**:36. https://doi.org/10.1186/s40168-019-0649-y
65. Shimoyama Y. *pyGenomeViz: A Genome Visualization Python Package for Comparative Genomics*. Jun. 2022. 23 July 2022, date last accessed. Available: https://github.com/moshi4/pyGenomeViz
66. Apweiler R, Bairoch A, Wu CH *et al.* UniProt: the universal protein knowledgebase. *Nucleic Acids Res* 2004;**32**:115D–19. https://doi.org/10.1093/nar/gkh131
67. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *Journal of molecular biology* 1990;**215**(3):403–10. https://doi.org/10.1016/S0022-2836(05)80360-2.
68. Katoh K, Misawa K, Kuma K *et al.* MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 2002;**30**:3059–66. https://doi.org/10.1093/nar/gkf436
69. Criscuolo A, Gribaldo S. BMGE (block mapping and gathering with entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol* 2010;**10**:210. https://doi.org/10.1186/1471-2148-10-210
70. Nguyen L-T, Schmidt HA, von Haeseler A *et al.* IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015;**32**:268–74. https://doi.org/10.1093/molbev/msu300
71. Yu G. Using ggtree to visualize data on tree-like structures. *Curr Protoc Bioinform* 2020;**69**:e96. https://doi.org/10.1002/cpbi.96
72. Lin Z, Akin H, Rao R *et al.* Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* 2023;**379**:1123–30. https://doi.org/10.1126/science.ade2574
73. Weissenow K, Heinzinger M, Steinegger M *et al.* Ultra-fast protein structure prediction to capture effects of sequence variation in mutation movies. bioRxiv. p. 2022.11.14.516473, 18 November 2022. https://doi.org/10.1101/2022.11.14.516473
74. Anandan A, Vrielink A. Structure and function of lipid A-modifying enzymes. *Ann N Y Acad Sci* 2020;**1459**:19–37. https://doi.org/10.1111/nyas.14244
75. Breazeale SD, Ribeiro AA, Raetz CRH. Origin of lipid a species modified with 4-Amino-4-deoxy-l-arabinose in polymyxin-resistant mutants of *Escherichia coli*: an aminotransferase (ArnB) that generates UDP-4-amino-4-deoxy-l-arabinose. *J Biol Chem* 2003;**278**:24731–9. https://doi.org/10.1074/jbc.M304043200
76. Dell A, Galadari A, Sastre F *et al.* Similarities and differences in the glycosylation mechanisms in prokaryotes and eukaryotes. *Int J Microbiol* 2011;**2010**:e148178. https://doi.org/10.1155/2010/148178
77. Abu-Qarn M, Eichler J. Protein N-glycosylation in archaea: defining *Haloferax volcanii* genes involved in S-layer glycoprotein glycosylation. *Mol Microbiol* 2006;**61**:511–25. https://doi.org/10.1111/j.1365-2958.2006.05252.x
78. Kärcher U, Schröder H, Haslinger E *et al.* Primary structure of the heterosaccharide of the surface glycoprotein of *Methanothermus fervidus*. *J Biol Chem* 1993;**268**:26821–6. https://doi.org/10.1016/S0021-9258(19)74185-4
79. Zähringer U, Moll H, Hettmann T *et al.* Cytochrome b558/566 from the archaeon *Sulfolobus acidocaldarius* has a unique Asn-linked highly branched hexasaccharide chain containing 6-sulfoquinovose. *Eur J Biochem* 2000;**267**:4144–9. https://doi.org/10.1046/j.1432-1327.2000.01446.x
80. Mescher MF, Strominger JL. Purification and characterization of a prokaryotic glycoprotein from the cell envelope of

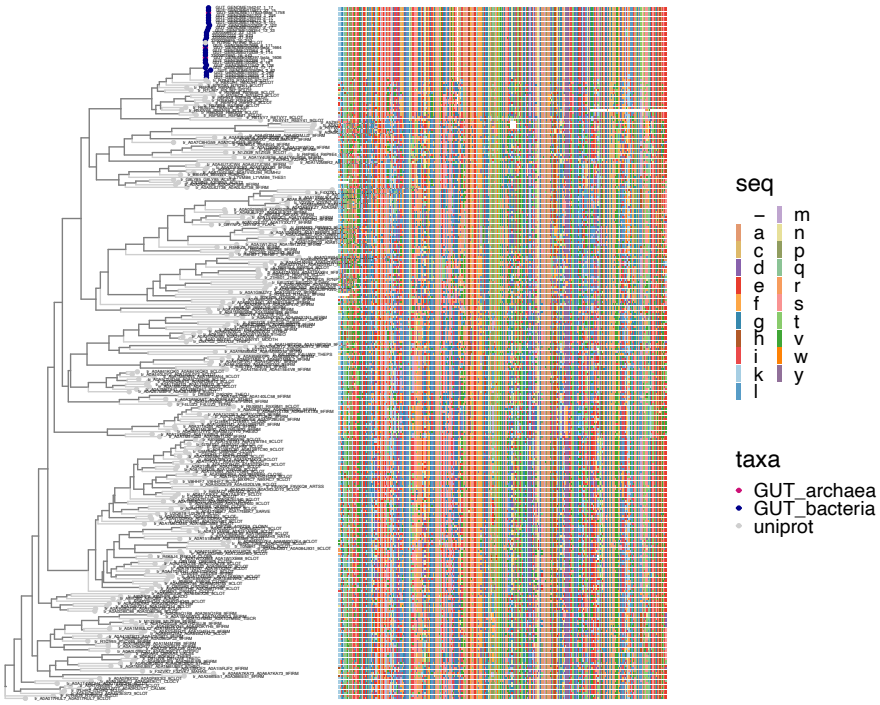
Halobacterium salinarium. *J Biol Chem* 1976;**251**:2005–14. https://doi.org/10.1016/S0021-9258(17)33647-5
81. Tamir A, Eichler J. N-glycosylation is important for proper *Haloferax volcanii* S-layer stability and function. *Appl Environ Microbiol* 2017;**83**:e03152–16. https://doi.org/10.1128/AEM.03152-16
82. Samuel BS, Hansen EE, Manchester JK *et al.* Genomic and metabolic adaptations of *Methanobrevibacter smithii* to the human gut. *Proc Natl Acad Sci U S A* 2007;**104**:10643–8. https://doi.org/10.1073/pnas.0704189104
83. Wandall HH, Nielsen MAI, King-Smith S *et al.* Global functions of O-glycosylation: promises and challenges in O-glycobiology. *FEBS J* 2021;**288**:7183–212. https://doi.org/10.1111/febs.16148
84. Metcalf WW, Griffin BM, Cicchillo RM *et al.* Synthesis of methylphosphonic acid by marine microbes: a source for methane in the aerobic ocean. *Science* 2012;**337**:1104–7. https://doi.org/10.1126/science.1219875
85. Dormoy-Raclet V, Markovits J, Malato Y *et al.* Unr, a cytoplasmic RNA-binding protein with cold-shock domains, is involved in control of apoptosis in ES and HuH7 cells. *Oncogene* 2007;**26**:2595–605. https://doi.org/10.1038/sj.onc.1210068
86. Roer L, Aarestrup FM, Hasman H. The EcoKI type I restriction-modification system in *Escherichia coli* affects but is not an absolute barrier for conjugation. *J Bacteriol* 2015;**197**:337–42. https://doi.org/10.1128/JB.02418-14
87. Pingoud A, Fuxreiter M, Pingoud V *et al.* Type II restriction endonucleases: structure and mechanism. *Cell Mol Life Sci* 2005;**62**:685–707. https://doi.org/10.1007/s00018-004-4513-1
88. Havemann GD, Sampson EM, Bobik TA. PduA is a shell protein of polyhedral organelles involved in coenzyme B(12)-dependent degradation of 1,2-propanediol in *Salmonella enterica* serovar typhimurium LT2. *J Bacteriol* 2002;**184**:1253–61. https://doi.org/10.1128/JB.184.5.1253-1261.2002
89. Kennedy NW, Ikonomova SP, Slininger Lee M *et al.* Self-assembling shell proteins PduA and PduJ have essential and redundant roles in bacterial microcompartment assembly. *J Mol Biol* 2021;**433**:166721. https://doi.org/10.1016/j.jmb.2020.11.020
90. Sampson EM, Bobik TA. Microcompartments for B12-dependent 1,2-propanediol degradation provide protection from DNA and cellular damage by a reactive metabolic intermediate. *J Bacteriol* 2008;**190**:2966–71. https://doi.org/10.1128/JB.01925-07
91. Hansen EE, Lozupone CA, Rey FE *et al.* Pan-genome of the dominant human gut-associated archaeon, *Methanobrevibacter smithii*, studied in twins. *Proc Natl Acad Sci U S A* 2011;**108**:4599–606. https://doi.org/10.1073/pnas.1000071108
92. Acar Kirit H, Bollback JP, Lagator M. The role of the environment in horizontal gene transfer. *Mol Biol Evol* 2022;**39**:msac220. https://doi.org/10.1093/molbev/msac220
93. Rigden DJ, Galperin MY. Sequence analysis of GerM and SpoVS, uncharacterized bacterial "sporulation" proteins with widespread phylogenetic distribution. *Bioinformatics* 2008;**24**:1793–7. https://doi.org/10.1093/bioinformatics/btn314
94. Onyenwoke RU, Brill JA, Farahi K *et al.* Sporulation genes in members of the low G+C Gram-type-positive phylogenetic branch (Firmicutes). *Arch Microbiol* 2004;**182**:182–92. https://doi.org/10.1007/s00203-004-0696-y
95. Pschorn W, Paulus H, Hansen J *et al.* Induction of sporulation in *Bacillus brevis*. *Eur J Biochem* 1982;**129**:403–7. https://doi.org/10.1111/j.1432-1033.1982.tb07064.x
96. Suárez JM, Edwards AN, McBride SM. The *Clostridium difficile* cpr locus is regulated by a noncontiguous two-component system in response to type a and B lantibiotics. *J Bacteriol* 2013;**195**:2621–31. https://doi.org/10.1128/JB.00166-13
97. Gleason AC, Ghadge G, Chen J *et al.* Machine learning predicts translation initiation sites in neurologic diseases with nucleotide repeat expansions. *PLoS One* 2022;**17**:e0256411. https://doi.org/10.1371/journal.pone.0256411
98. Zhang S, Hu H, Jiang T *et al.* TITER: predicting translation initiation sites by deep learning. *Bioinformatics* 2017;**33**:i234–42. https://doi.org/10.1093/bioinformatics/btx247



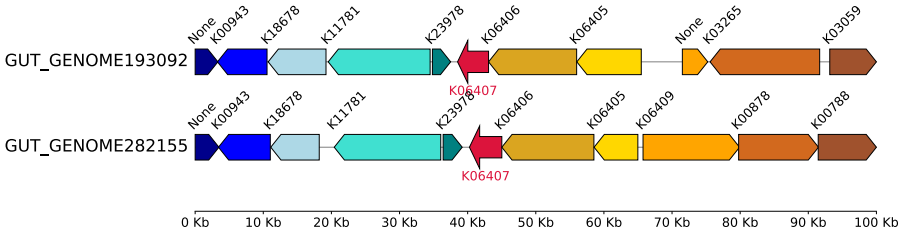
Supp. Figure 1. *Heatmaps demonstrating the intensity of HGT events between M. smithii genomes. A, HGT between taxonomic groups named as follows: A - Methanobrevibacter_A smithii, B - Methanobrevibacter_A smithii_A (Ca. Methanobreviabcter intestini), C - Methanobrevibacter_A oralis, E - GCF_000016525.1 (M. smithii), F - GCF_002252585.1 (Ca. Methanobreviabcter intestini); B, HGT events between individual genomes of same groups. The legend depicts the frequency of HGT events among the genomes of A, taxonomic groups and B, individual genomes.*



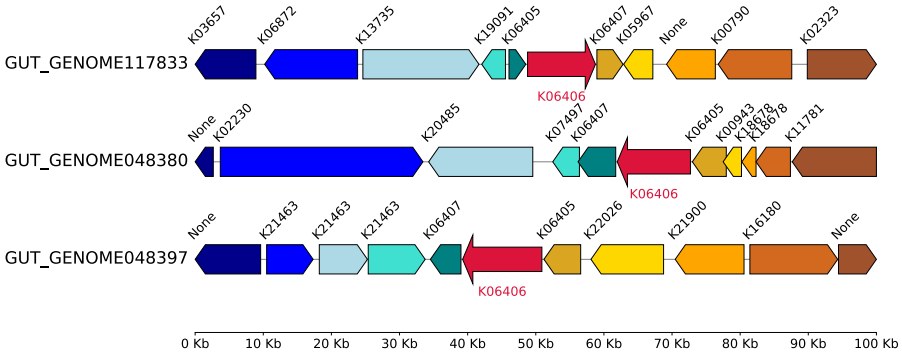
Supp. Figure 2. *Phylogenetic tree of stage V sporulation proteins AE from identified SC h9 and Uniprot. Bacterial and archaeal proteins from cluster h9 are depicted as GUT_bacteria and GUT_archaea in dark blue and pink, respectively.*



Supp. Figure 3. *Phylogenetic tree of stage V sporulation proteins AD from identified SC h20 and Uniprot. Bacterial and archaeal proteins from cluster h9 are depicted as GUT_bacteria and GUT_archaea in dark blue and pink, respectively.*



Supp. Figure 4. *Gene synteny of homologous bacterial sequences obtained from the human gut dataset that share similarities with the archaeal sequences from cluster h9 encoding stage V sporulation protein AE (spoVAE).*



Supp. Figure 5. *Gene synteny of homologous bacterial sequences obtained from the human gut dataset that share similarities with the archaeal sequences from cluster h20 encoding stage V sporulation protein AD (spoVAD).*

Conclusions and Perspectives

This chapter presents the overall conclusions drawn from the extensive research reached throughout this doctoral thesis. The work presented herein offers a comprehensive exploration of the intricate role played by the human gut microbiome in the pathophysiology of PD. By employing a robust combination of meta-omics approaches, network-based analyses, and advanced computational tools, this thesis has elucidated the multifaceted interactions between the gut microbiome and PD. The findings emphasize the critical importance of gut microbial imbalances in the context of PD, particularly through disruptions in microbial metabolism and reduced functional diversity. This research not only deepens our understanding of the gut-brain axis but also opens avenues for potential therapeutic interventions aimed at restoring microbial balance as a strategy for mitigating the effects of PD.

Our study has uncovered key alterations in microbial functions and metabolic pathways associated with PD and iRBD, including notable increases in β-glutamate levels and significant changes in glutamate metabolism, as described in the section “Paper 1: Integrated Multi-omics Highlights Alterations of Gut Microbiome Functions in Prodromal and Idiopathic Parkinson’s Disease.” While the neurotoxic effects of L-glutamate on neurons are well-documented (Iovino et al., 2020), β-glutamate remains poorly understood, with only one known enzymatic interaction described in current databases. We identified a central role for glutamate derivatives within the gut microbiome of PD individuals (Paper 1, Fig. 3A, C, D, Fig. 5). β-glutamate and related glutamate genes are particularly relevant due to L-glutamate's reported neurotoxicity and its association with microbial activity. However, the precise links between β-glutamate, L-, and D-glutamate are not yet clear. We have not identified any enzymatic interactions with β-glutamate in the gut microbiome, highlighting a critical knowledge gap in its microbial and host interactions. Furthermore, data on the kinetics and physiological effects of β-glutamate in the host are lacking, emphasizing the need for more research to clarify its functional role in disease. Experimentally, further investigation is necessary to determine β-glutamate's function and potential as a substrate for other microbial enzymes. To date, β-glutamate has only been described as an osmolyte in the *Methanogenium* and *Methanococcus* archaeal genera (Robertson et al., 1990) and *M. portucalensis* (Robinson et al., 2001), and wasn’t detected as a product of *M.smithii*. Additionally, there is a general lack of information about the activity of β-glutamate in *M. smithii*, the most prevalent methanogenic archaeon in the human gut. Given our observation of increased *M. smithii* expression correlating with elevated PD-related metabolites,

understanding its role is particularly significant. To advance this research, both experimental and *in silico* strategies could be instrumental. Computational approaches, such as using AlphaFold 3 (Abramson et al., 2024) to predict protein interactions involving β-glutamate, may offer valuable insights for experimental studies. Further work exploring interaction patterns between β-glutamate and microbiome-derived metabolites could shed light on its specific role in PD pathophysiology.

Besides altered levels of β-glutamate, we observed substantial differences in chemotaxis and flagellar assembly pathways (Paper 1, Fig. 3D-F, Fig. 4; Paper 2 Fig. 3B, Fig. 4C-E, Supp. Fig. 3), processes typically linked to pro-inflammatory responses (Gram et al., 2021; F. Qian et al., 2015; Tran et al., 2019). These changes likely reflect the indirect influence of the gut microbiome on the gut-brain axis through interactions with the immune system. Notably, we demonstrated that these alterations occur not at the genomic level (Boktor et al., 2023), as previously reported, but at the transcriptomic level, emphasizing the potential regulatory impact of these pathways in PD. We additionally highlighted strong correlation between flagellar assembly genes and genes involved in bacterial microcompartments (Paper 2, Fig. 5). Based on the findings presented in the studies “Paper 1: Integrated Multi-omics Highlights Alterations of Gut Microbiome Functions in Prodromal and Idiopathic Parkinson’s Disease” and “Paper 2: Microbiome Expression Network is Dysregulated in Parkinson’s Disease Individuals,” we hypothesize that certain bacterial taxa, such as *Anaerobutyricum* and *Blautia*, exhibit a decreased expression of chemotaxis and flagella genes, which is likely due to their inability to utilize ethanolamine and 1,2-propanediol as energy sources. These compounds are crucial for the metabolic activities of these bacteria (Engels et al., 2016; Trischler et al., 2023), and their absence may impair the bacteria’s motility and chemotactic responses. The observed dysregulation in gene expression suggests a significant shift in the metabolic capabilities of these bacterial communities, potentially contributing to the pathophysiology of PD. However, it is important to note that experimental validation of this hypothesis was beyond the scope of my current work. Therefore, this intriguing finding warrants further experimental investigation to confirm and elucidate the underlying mechanisms.

Interestingly, taxa encoding and expressing chemotaxis and flagellar assembly-related genes, such as *Roseburia*, are commensals typically associated with either ‘silently recognized’ (Clasen et al., 2023) or anti-inflammatory properties (Quan et al., 2018; Shen et al., 2022; X. Wu et al., 2020). For example, we observed a decrease in *Roseburia*'s transcription of flagellin in the gut microbiome of PD patients (Paper 1, Fig. 4, Ext. Fig. 5), further suggesting a complex interplay between microbial community shifts and host

immune responses. Flagellins, in particular, could serve as immune-modulating compounds affecting both microbiome composition and immune responses in PD. Additionally, our analysis revealed significant alterations in BMCs, particularly in *Blautia* and *Anaerobutyricum* genera, as described earlier. These microcompartments, besides being crucial for energy catabolism, are commonly associated with the survival of specific pathogenic bacteria (Dank et al., 2021; Vance, 2018), seem to confer a competitive advantage to certain commensal bacteria within the PD microbiome, as demonstrated in non-pathogenic species (Akouris et al., 2024; Moreira de Gouveia et al., 2023). This mechanism may play a role in protecting against pathogenic invasions. The downregulation of BMC-associated functions in PD, coupled with elevated glycerol levels in healthy controls (Paper 1, Fig. 3A), further suggests that BMCs may have a protective role in gut homeostasis. To our knowledge, our work is the first to demonstrate the involvement of BMCs and flagellar assembly in the context of PD. Our findings highlight core microbiome functions that are disrupted in disease, including chemotaxis, flagellar assembly and BMC activity. These observations provide valuable insights into the microbial mechanisms potentially contributing to PD pathogenesis and offer a foundation for further experimental validation. *In vitro* experiments, such as gene knockouts or the use of advanced models like HuMiX (Shah et al., 2016), which simulate human-microbe interactions, are essential to confirming these microbiome-host interactions. Such approaches will help elucidate the mechanistic links between microbial changes and PD symptoms, and ultimately, these findings could inform the development of microbiome-targeted therapies for PD.

The findings on genes involved in chemotaxis and flagellar assembly in the gut microbiome of PD patients suggest several promising translational applications that could contribute to future therapeutic developments. First, the identification of anti-inflammatory flagellin expression in commensal bacteria such as *Roseburia* presents an opportunity to explore the therapeutic introduction of these bacteria or similar strains to manage PD-related inflammation. Prior research has shown that anti-inflammatory bacterial components, such as specific flagellins, can modulate immune responses (Quan et al., 2018; Shen et al., 2022), suggesting that supplementation with these bacteria might reduce pro-inflammatory signals in the gut-brain axis of PD patients. Such interventions could potentially take the form of live biotherapeutics or engineered probiotic formulations designed to express anti-inflammatory flagellin, potentially moderating gut and systemic inflammation, which has been associated with neurodegenerative progression in PD. Probiotics can enhance the gut microbiome by altering the intestinal environment and suppressing the growth of harmful

bacteria, and this approach has been shown in application to the treatment of other gut-associated disorders such as IBD (Praveschotinunt et al., 2019).

Additionally, since anti-inflammatory flagellin expression appears downregulated in PD, therapeutic strategies could aim to reactivate these flagellins within the existing microbiome. For instance, studies have indicated that dietary components or prebiotics can modulate bacterial gene expression, including genes related to immune modulation (Burr et al., 2020; Pérez-Cano, 2022). Future research could investigate whether specific dietary interventions or small molecules could selectively enhance the expression of anti-inflammatory flagellins in *Roseburia* and similar taxa. Such an approach would focus on functionally boosting the patient's own microbiota to reduce inflammation, thus supporting microbiome resilience and restoring immune modulation within the gut-brain axis.

Furthermore, the role of BMCs in gut homeostasis, particularly in relation to genera such as *Blautia* and *Anaerobutyricum*, points toward another potential avenue for translational research. The observed downregulation of BMC-associated functions in PD, with higher levels of BMC activity in the healthy group, suggests that enhancing BMC expression might protect the gut microbiome from inflammatory and pathogenic shifts. Targeted approaches to upregulate BMC-associated pathways in PD could involve dietary compounds or prebiotics promoting BMC-expressing bacteria, aiming to protect the microbial ecosystem and improve the host's gut barrier function. Additionally, the use of bacterial strains with modified BMC expression represents a promising approach to restoring gut homeostasis in PD. Such engineered strains could potentially restore or enhance the activity of BMCs in PD-affected gut ecosystems, providing a competitive advantage to beneficial bacteria and supporting the gut barrier against inflammatory and pathogenic shifts. This aligns with studies that explore how bacterial modifications can regulate immune responses and gut environment stabilization, supporting the therapeutic potential of engineered probiotics for gut-brain-related disorders (Paudel et al., 2023). Synthetic strains could be created to express both BMC-related resilience and anti-inflammatory flagellin, potentially offering a targeted live therapeutic intervention that reinforces gut health without altering the microbiome's broader composition. However, it is important to note that the use of genetically modified organisms in human treatments remains controversial and faces significant technical, regulatory, and ethical challenges, making human application currently unlikely.

For verification of these approaches, advanced *in vitro* models, such as HuMiX, which simulate human-microbe interactions, could provide valuable insights. Using HuMiX to

model PD-specific conditions would allow detailed examination of microbiome-host interactions, specifically focusing on flagellin or BMC modulation effects on inflammatory pathways. Such models provide a preclinical validation framework, supporting the development of targeted microbiome-based therapies that leverage anti-inflammatory and structural-support functions of the microbiome. Overall, harnessing these microbial functions therapeutically presents a promising translational opportunity to bridge findings from this thesis into functional therapies that may influence PD management by targeting inflammation and enhancing gut-brain communication.

Our findings regarding *Methanobrevibacter smithii*, a methanogenic archaeon, underscore several key insights. Although *M. smithii* is generally low in abundance within the gut microbiome, studies report that its levels are higher in PD patients (Cem Duru et al., 2024; Rosario et al., 2021). In this work, we demonstrated that *M. smithii* exhibits notable transcriptional activity, correlating with significantly elevated metabolite levels in PD (Paper 1, Fig. 3C). In the section “Paper 3: Functional Prediction of Proteins from the Human Gut Archaeome” we explored the functional capabilities of *M. smithii* within the gut through a structure-based strategy, advancing our understanding of its roles in this environment. We enhanced functional annotation and highlighted potentially impactful functions for *M. smithii* in the gut (Paper 3, Fig. 3), with some proteins even corresponding to those described in Paper 2. This study reinforces the considerable but underexplored potential of archaea, which remain challenging to investigate experimentally and computationally due to their unique biological characteristics. Developing experimental methods and workflows specifically tailored to archaea is a promising avenue for future research initiated by this thesis. Our findings, alongside studies reporting an increased abundance of *M. smithii* in PD, reveal an intriguing signal at the transcriptomic level. Its correlation with metabolites elevated in PD suggests that *M. smithii* may have a more significant role in gut-microbiome-related pathologies than previously understood. Further investigation into *M. smithii* activity in relation to PD could offer valuable insights.

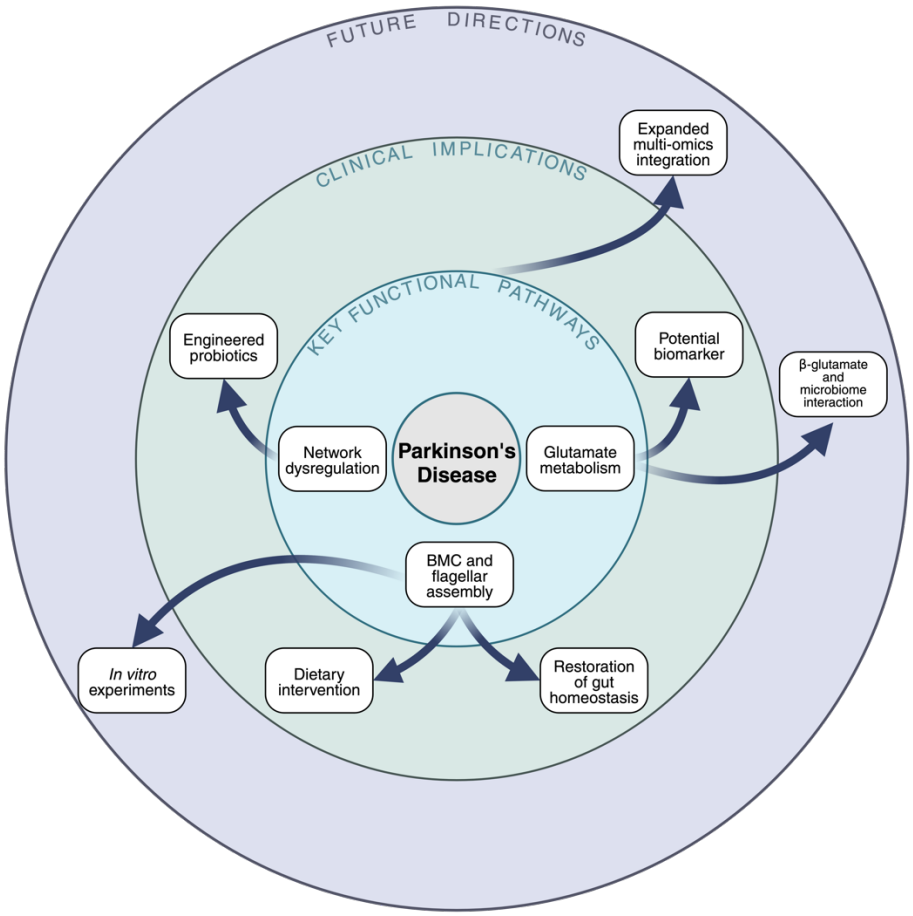


Figure 4. Diagram demonstrating the summary on key findings, clinical implications and future directions of research.

This study has several limitations, primarily its reliance on MG and MT data as the main sources of information. While MG captures microbial gene abundance and provides a static overview, it only indicates the presence of organisms and their potential functional roles. MT offers insights into gene activity of microbial communities under specific environmental conditions. However, since transcription is an intermediate stage in gene expression, incorporating synthesized protein data would enhance understanding of microbial functionality. Proteomics enables identification and quantification of proteins, while MP characterizes proteins actively expressed by a microbial community at a given time (Wilmes & Bond, 2004). Including MP would deepen the analysis, complementing MG and MT data, and provide insights into functionally active microbial members and their protein products, which may serve as potential biomarkers linked to specific clinical conditions. Thus, MP is emerging as a vital complement to MG and MT approaches, successfully applied in studies

of Crohn's disease (Erickson et al., 2012; Henry et al., 2022), inflammatory bowel disease (Zhang et al., 2018) and cystic fibrosis (Hardouin et al., 2021).

Despite its potential, the application of MP is less widespread than that of MG and MT, primarily due to a lack of standardized protocols for sample preparation, limited bioinformatics tools, and challenges in detecting low-abundance proteins in complex biological samples (Heyer et al., 2017; Zhang et al., 2018). Additionally, MP relies heavily on reference databases derived from genomic and metagenomic data for accurate peptide identification and pathway analysis (Lai et al., 2019). While advancements have been made, MP is still in its developmental phase, hindered by the complexity of the gut microbiome (Wilmes et al., 2015). Alongside computational challenges posed by vast data, experimental obstacles exist (Glatter et al., 2015; Zhang et al., 2018). Identifying peptides from homologous proteins can lead to redundant protein identifications, potentially skewing analyses since homologous proteins may perform distinct functional roles across species. Nonetheless, MP could identify proteins specifically dysregulated in PD patients, revealing altered pathways and gene expression patterns. For instance, disruptions in flagellar assembly pathways or microcompartment-associated genes could be validated by measuring the abundance or activity of corresponding proteins, elucidating the mechanistic link between microbial changes and PD symptoms, and providing stronger evidence for targeted therapeutic interventions.

MP is valuable for identifying post-translational modifications (PTMs), which are crucial for various bacterial processes, including protein synthesis, cell cycle regulation, biofilm formation, virulence, and nitrogen metabolism (Christensen et al., 2019; Macek et al., 2019). The enzymes responsible for PTMs vary significantly across bacterial species, and environmental conditions heavily influence the extent of these modifications (Bastos et al., 2017; Q. Ma et al., 2021). Consequently, modified proteins may display different functional behaviors depending on their environment, highlighting MP's relevance in identifying molecular agents linked to disease. Given that the human microbiome is significantly affected by factors such as geography, diet, and medication, studying modified proteins can improve our understanding of how protein activity adapts to these changes.

However, it is important to recognize that proteins encoded by orthologous genes do not always retain equivalent functions, as some may have been repurposed for novel roles (Kuraku & Ukena, 2021). Additionally, bacterial orthologs may exhibit distinct functional roles and regulatory mechanisms depending on species and environmental contexts (Price et al., 2007). Consequently, *in silico* predictions of gene or protein functions, while

informative, are not fully reliable without experimental validation. The validation process, however, presents challenges. For instance, proteins may behave differently in isolated systems compared to their natural environments, and microbial communities often undergo compositional changes when removed from their natural habitat (Dantas et al., 2013). This introduces variability into experimental outcomes and complicates efforts to draw accurate conclusions about *in vivo* microbial activity. To overcome these limitations, engineered microenvironments, such as "gut-on-a-chip" devices, offer a promising solution. These platforms simulate both mono- and multi-environment conditions, enabling real-time and continuous monitoring of gut microbiota interactions with human cells, tissues, and even other organs (Kim et al., 2012; Lucchetti et al., 2021). By providing a closer approximation to the in vivo conditions of the human gut, these systems help address the challenges associated with studying microbial communities in simplified or artificial models. As a result, they hold considerable potential for improving our understanding of microbial dynamics and their contributions to health and disease.

Although bioinformatics tools have advanced significantly, they continue to have limitations in fully elucidating the complexities of the human gut microbiome and its connections to related diseases. Bioinformatics challenges in microbiome research stem from the complexity and scale of data generated in omics studies. The vast diversity of microbial communities and their dynamic interactions with host environments complicate data processing and analysis, leading to issues with reproducibility and interpretation. This is not unexpected, given that the human gut microbiota consists of over one thousand microbial species, collectively containing approximately 150 times more genes than the entire human genome (Lagier et al., 2016). Furthermore, it is estimated that around one hundred trillion microbes inhabit the human body, contributing significantly to various biological processes associated with health and disease (B. Wang et al., 2017). Additionally, integrating multi-omics data poses significant computational challenges, as existing workflows often struggle to manage the volume and variety of data. The absence of standardized protocols for data collection and analysis further contributes to inconsistencies across studies, limiting the ability to derive generalizable conclusions.

Nevertheless, integration of multi-omics techniques including MG, MT and MP offers enhanced insights into the functional cellular and metabolic pathways that characterize microbial ecosystems (X. Zhang et al., 2019). The research presented in this thesis further underscores the invaluable contribution of multi-omics approaches in understanding human health. This trend is not surprising, as the increasing accessibility of these techniques has led to their more frequent application across a range of diseases, beyond just PD (Ali et al.,

2023; Mills et al., 2022; Worby et al., 2022). Combining these techniques enhances functional predictability and facilitates a more comprehensive conceptualization of the roles within microbial ecosystems (Ferrocino et al., 2023).

On this topic, the results presented in section “Paper 1: Integrated Multi-omics Highlights Alterations of Gut Microbiome Functions in Prodromal and Idiopathic Parkinson’s Disease” reveal a notable discrepancy between microbial composition and activity, evidenced by the absence of differences at the MG level alongside substantial variations at the MT level. This indicates that while the structure of the gut microbiome remains stable between health and disease states, the functional activity of these microbial communities is significantly altered. And therefore, microbial functionality could be a more sensitive marker of disease status than compositional changes alone. To substantiate this hypothesis, further research incorporating dynamic scenarios is essential. Longitudinal analyses, such as time-series studies assessing health and disease trajectories under various interventions – like dietary changes and pharmacological treatments – could provide critical insights into how microbial activity shifts in response to these influences. While acknowledging that such studies are often resource-intensive and complex, their significance in enhancing our understanding of disease pathophysiology, including PD, cannot be overstated. Additionally, findings from our work “Paper 2: Microbiome Expression Network is Dysregulated in Parkinson’s Disease Individuals” highlight that PD is characterized by decreased gene expression diversity, also emphasizing the need for a comprehensive approach to observe systemic changes over time. Conducting such experiments will deepen our understanding of the multifaceted interactions within the gut microbiome and their implications for health and disease. By continuously collecting data over time and integrating multi-omics approaches, the field can advance our understanding of the intricate interplay between microbial functions, host genetics, and environmental factors. This progress may pave the way for targeted therapeutic strategies designed to restore microbial eubiosis in PD.

In the end, microbiome science is an exciting and rapidly growing field that is still finding its footing. The technical advancements and bioinformatics developments showcased in this work have fueled the remarkable growth of this emerging area of research. One of the most intriguing strengths of microbiome research is its interdisciplinary nature. It’s unrealistic for anyone to be an expert in every method related to microbiome studies, so our greatest strength comes from bringing together talented individuals from various disciplines. By combining our knowledge and skills, we can work together to unlock the mysteries of the microbiome. The future of microbiome science is filled with opportunities for preventing, diagnosing, and treating complex human diseases. By striking a balance between optimism

and realism, we can pave the way for meaningful advancements in how we approach health and disease research.

Bibliography

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C. C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., … Jumper, J. M. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature 2024 630:8016*, *630*(8016), 493–500. https://doi.org/10.1038/s41586-024-07487-w

Adams-Carr, K. L., Bestwick, J. P., Shribman, S., Lees, A., Schrag, A., & Noyce, A. J. (2016). Constipation preceding Parkinson's disease: a systematic review and meta-analysis. *Journal of Neurology, Neurosurgery & Psychiatry*, *87*(7), 710–716. https://doi.org/10.1136/JNNP-2015-311680

Aho, V. T. E., Houser, M. C., Pereira, P. A. B., Chang, J., Rudi, K., Paulin, L., Hertzberg, V., Auvinen, P., Tansey, M. G., & Scheperjans, F. (2021). Relationships of gut microbiota, short-chain fatty acids, inflammation, and the gut barrier in Parkinson's disease. *Molecular Neurodegeneration*, *16*(1), 1–14. https://doi.org/10.1186/S13024-021-00427-6/FIGURES/7

Aho, V. T. E., Pereira, P. A. B., Voutilainen, S., Paulin, L., Pekkonen, E., Auvinen, P., & Scheperjans, F. (2019). Gut microbiota in Parkinson's disease: Temporal stability and relations to disease progression. *EBioMedicine*, *44*, 691–707. https://doi.org/10.1016/J.EBIOM.2019.05.064

Akahoshi, D. T., & Bevins, C. L. (2022). Flagella at the Host-Microbe Interface: Key Functions Intersect With Redundant Responses. *Frontiers in Immunology*, *13*. https://doi.org/10.3389/FIMMU.2022.828758

Akouris, P. P., Stuivenberg, G. A., Chmiel, J. A., Kiattiburut, W., Poon, A., Reid, G., & Burton, J. P. (2024). Ethanolamine enhances adhesion, promotes microcompartment formation, and modulates gene expression in Levilactobacillus brevis ATCC 14869. *Gut Microbes*, *16*(1). https://doi.org/10.1080/19490976.2024.2350778

Ali, R. O., Quinn, G. M., Umarova, R., Haddad, J. A., Zhang, G. Y., Townsend, E. C., Scheuing, L., Hill, K. L., Gewirtz, M., Rampertaap, S., Rosenzweig, S. D., Remaley, A. T., Han, J. M., Periwal, V., Cai, H., Walter, P. J., Koh, C., Levy, E. B., Kleiner, D. E., … Heller, T. (2023). Longitudinal multi-omics analyses of the gut-liver axis reveals

metabolic dysregulation in hepatitis C infection and cirrhosis. *Nature Microbiology*, *8*(1), 12–27. https://doi.org/10.1038/S41564-022-01273-Y

Almeida, A., Nayfach, S., Boland, M., Strozzi, F., Beracochea, M., Shi, Z. J., Pollard, K. S., Sakharova, E., Parks, D. H., Hugenholtz, P., Segata, N., Kyrpides, N. C., & Finn, R. D. (2020). A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nature Biotechnology 2020 39:1*, *39*(1), 105–114. https://doi.org/10.1038/s41587-020-0603-3

Ashtiani, M., Salehzadeh-Yazdi, A., Razaghi-Moghadam, Z., Hennig, H., Wolkenhauer, O., Mirzaie, M., & Jafari, M. (2018). A systematic survey of centrality measures for protein-protein interaction networks. *BMC Systems Biology*, *12*(1), 1–17. https://doi.org/10.1186/S12918-018-0598-2/FIGURES/8

Azeloglu, E. U., & Iyengar, R. (2015). Signaling Networks: Information Flow, Computation, and Decision Making. *Cold Spring Harbor Perspectives in Biology*, *7*(4). https://doi.org/10.1101/CSHPERSPECT.A005934

Badia-i-Mompel, P., Wessels, L., Müller-Dott, S., Trimbour, R., Ramirez Flores, R. O., Argelaguet, R., & Saez-Rodriguez, J. (2023). Gene regulatory network inference in the era of single-cell multi-omics. *Nature Reviews Genetics 2023 24:11*, *24*(11), 739–754. https://doi.org/10.1038/s41576-023-00618-5

Baj, A., Moro, E., Bistoletti, M., Orlandi, V., Crema, F., & Giaroni, C. (2019). Glutamatergic Signaling Along The Microbiota-Gut-Brain Axis. *International Journal of Molecular Sciences*, *20*(6). https://doi.org/10.3390/IJMS20061482

Bao, G., Wang, M., Doak, T. G., & Ye, Y. (2015). Strand-specific community RNA-seq reveals prevalent and dynamic antisense transcription in human gut microbiota. *Frontiers in Microbiology*, *6*(SEP), 149568. https://doi.org/10.3389/FMICB.2015.00896/BIBTEX

Barabási, A. L., & Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nature Reviews Genetics 2004 5:2*, *5*(2), 101–113. https://doi.org/10.1038/NRG1272

Barichella, M., Severgnini, M., Cilia, R., Cassani, E., Bolliri, C., Caronni, S., Ferri, V., Cancello, R., Ceccarani, C., Faierman, S., Pinelli, G., De Bellis, G., Zecca, L., Cereda, E., Consolandi, C., & Pezzoli, G. (2019). Unraveling gut microbiota in Parkinson's

disease and atypical parkinsonism. *Movement Disorders : Official Journal of the Movement Disorder Society*, *34*(3), 396–405. https://doi.org/10.1002/MDS.27581

Barthélemy, M. (2004). Betweenness centrality in large complex networks. *The European Physical Journal B 2004 38:2*, *38*(2), 163–168. https://doi.org/10.1140/EPJB/E2004-00111-4

Bascompte, J. (2010). Structure and dynamics of ecological networks. *Science*, *329*(5993), 765–766. https://doi.org/10.1126/SCIENCE.1194255/ASSET/B03418CD-0EDA-47BF-B4C7-87BC3D87FE61/ASSETS/GRAPHIC/329_765_F1.JPEG

Bashiardes, S., Zilberman-Schapira, G., & Elinav, E. (2016). Use of Metatranscriptomics in Microbiome Research. *Https://Doi.Org/10.4137/BBI.S34610*, *10*, 19–25. https://doi.org/10.4137/BBI.S34610

Bastos, P. A. D., da Costa, J. P., & Vitorino, R. (2017). A glimpse into the modulation of post-translational modifications of human-colonizing bacteria. *Journal of Proteomics*, *152*, 254–275. https://doi.org/10.1016/J.JPROT.2016.11.005

Batah, J., Kobeissy, H., Pham, P. T. B., Denève-Larrazet, C., Kuehne, S., Collignon, A., Janoir-Jouveshomme, C., Marvaud, J. C., & Kansau, I. (2017). Clostridium difficile flagella induce a pro-inflammatory response in intestinal epithelium of mice in cooperation with toxins. *Scientific Reports 2017 7:1*, *7*(1), 1–10. https://doi.org/10.1038/s41598-017-03621-z

Beach, T. G., Corbillé, A. G., Letournel, F., Kordower, J. H., Kremer, T., Munoz, D. G., Intorcia, A., Hentz, J., Adler, C. H., Sue, L. I., Walker, J., Serrano, G., & Derkinderen, P. (2016). Multicenter Assessment of Immunohistochemical Methods for Pathological Alpha-Synuclein in Sigmoid Colon of Autopsied Parkinson's Disease and Control Subjects. *Journal of Parkinson's Disease*, *6*(4), 761. https://doi.org/10.3233/JPD-160888

Bedarf, J. R., Hildebrand, F., Coelho, L. P., Sunagawa, S., Bahram, M., Goeser, F., Bork, P., & Wüllner, U. (2017). Functional implications of microbial and viral gut metagenome changes in early stage L-DOPA-naïve Parkinson's disease patients. *Genome Medicine*, *9*(1). https://doi.org/10.1186/S13073-017-0428-Y

Bergman, H., Wichmann, T., Karmon, B., & DeLong, M. R. (1994). The primate subthalamic nucleus. II. Neuronal activity in the MPTP model of parkinsonism.

Https://Doi.Org/10.1152/Jn.1994.72.2.507, *72*(2), 507–520. https://doi.org/10.1152/JN.1994.72.2.507

Bikel, S., Valdez-Lara, A., Cornejo-Granados, F., Rico, K., Canizales-Quinteros, S., Soberón, X., Del Pozo-Yauner, L., & Ochoa-Leyva, A. (2015). Combining metagenomics, metatranscriptomics and viromics to explore novel microbial interactions: towards a systems-level understanding of human microbiome. *Computational and Structural Biotechnology Journal*, *13*, 390–401. https://doi.org/10.1016/J.CSBJ.2015.06.001

Bima, A. I. H., Elsamanoudy, A. Z., Albaqami, W. F., Khan, Z., Parambath, S. V., Al-Rayes, N., Kaipa, P. R., Elango, R., Banaganapalli, B., Shaik, N. A., Bima, A. I. H., Elsamanoudy, A. Z., Albaqami, W. F., Khan, Z., Parambath, S. V., Al-Rayes, N., Kaipa, P. R., Elango, R., Banaganapalli, B., & Shaik, N. A. (2022). Integrative system biology and mathematical modeling of genetic networks identifies shared biomarkers for obesity and diabetes. *Mathematical Biosciences and Engineering 2022 3:2310*, *19*(3), 2310–2329. https://doi.org/10.3934/MBE.2022107

Bisaglia, M., Filograna, R., Beltramini, M., & Bubacco, L. (2014). Are dopamine derivatives implicated in the pathogenesis of Parkinson's disease? *Ageing Research Reviews*, *13*(1), 107–114. https://doi.org/10.1016/J.ARR.2013.12.009

Blander, J. M., Longman, R. S., Iliev, I. D., Sonnenberg, G. F., & Artis, D. (2017). Regulation of inflammation by microbiota interactions with the host. *Nature Immunology*, *18*(8), 851–860. https://doi.org/10.1038/NI.3780

Blandini, F. (2010). An update on the potential role of excitotoxicity in the pathogenesis of Parkinson's disease. *Functional Neurology*, *25*(2), 65–71. https://europepmc.org/article/med/20923603

Blandini, F., Porter, R. H. P., & Greenamyre, J. T. (1996). Glutamate and Parkinson's disease. *Molecular Neurobiology 1996 12:1*, *12*(1), 73–94. https://doi.org/10.1007/BF02740748

Blesa, J., Foffani, G., Dehay, B., Bezard, E., & Obeso, J. A. (2021). Motor and non-motor circuit disturbances in early Parkinson disease: which happens first? *Nature Reviews Neuroscience 2021 23:2*, *23*(2), 115–128. https://doi.org/10.1038/s41583-021-00542-9

Bloch, F., Jackson, M. O., & Tebaldi, P. (2023). Centrality measures in networks. *Social Choice and Welfare*, *61*(2), 413–453. https://doi.org/10.1007/S00355-023-01456-4/TABLES/5

Boertien, J. M., Murtomäki, K., Pereira, P. A. B., van der Zee, S., Mertsalmi, T. H., Levo, R., Nojonen, T., Mäkinen, E., Jaakkola, E., Laine, P., Paulin, L., Pekkonen, E., Kaasinen, V., Auvinen, P., Scheperjans, F., van Laar, T., Verwey, N. A., van Harten, B., Portman, A. T., … Hoogendijk, J. E. (2022). Fecal microbiome alterations in treatment-naive de novo Parkinson's disease. *NPJ Parkinson's Disease*, *8*(1). https://doi.org/10.1038/S41531-022-00395-8

Boktor, J. C., Sharon, G., Verhagen Metman, L. A., Hall, D. A., Engen, P. A., Zreloff, Z., Hakim, D. J., Bostick, J. W., Ousey, J., Lange, D., Humphrey, G., Ackermann, G., Carlin, M., Knight, R., Keshavarzian, A., & Mazmanian, S. K. (2023). Integrated Multi-Cohort Analysis of the Parkinson's Disease Gut Metagenome. *Movement Disorders : Official Journal of the Movement Disorder Society*, *38*(3), 399–409. https://doi.org/10.1002/MDS.29300

Bonacich, P. (1987a). Power and Centrality: A Family of Measures. *Https://Doi.Org/10.1086/228631*, *92*(5), 1170–1182. https://doi.org/10.1086/228631

Bonacich, P. (1987b). Power and Centrality: A Family of Measures'. *American Journal of Sociology*, *92*(5), 1170–1182.

Boucher, B., & Jenna, S. (2013). Genetic interaction networks: Better understand to better predict. *Frontiers in Genetics*, *4*(DEC), 68624. https://doi.org/10.3389/FGENE.2013.00290/BIBTEX

Braak, H., Del Tredici, K., Rüb, U., De Vos, R. A. I., Jansen Steur, E. N. H., & Braak, E. (2003). Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiology of Aging*, *24*(2), 197–211. https://doi.org/10.1016/S0197-4580(02)00065-9

Bray, D. (2003). Molecular networks: The top-down view. *Science*, *301*(5641), 1864–1865. https://doi.org/10.1126/SCIENCE.1089118/ASSET/CC56BB59-4D35-4724-8BA4-95A240B1F3E4/ASSETS/GRAPHIC/SE3831882001.JPEG

Burr, A. H. P., Bhattacharjee, A., & Hand, T. W. (2020). Nutritional modulation of the microbiome and immune response. *Journal of Immunology (Baltimore, Md. : 1950)*, *205*(6), 1479. https://doi.org/10.4049/JIMMUNOL.2000419

Carlson, M. R. J., Zhang, B., Fang, Z., Mischel, P. S., Horvath, S., & Nelson, S. F. (2006). Gene connectivity, function, and sequence conservation: Predictions from modular yeast co-expression networks. *BMC Genomics*, *7*(1), 1–15. https://doi.org/10.1186/1471-2164-7-40/TABLES/2

Cava, F., Lam, H., De Pedro, M. A., & Waldor, M. K. (2011). Emerging knowledge of regulatory roles of d-amino acids in bacteria. *Cellular and Molecular Life Sciences: CMLS*, *68*(5), 817. https://doi.org/10.1007/S00018-010-0571-8

Cem Duru, I., Lecomte, A., Shishido, T. K., Laine, P., Suppula, J., Paulin, L., Scheperjans, F., Pereira, P. A. B., Auvinen, P., & Pereira, A. B. (2024). Metagenome-assembled microbial genomes from Parkinson's disease fecal samples. *Scientific Reports 2024 14:1*, *14*(1), 1–13. https://doi.org/10.1038/s41598-024-69742-4

Cerri, S., Mus, L., & Blandini, F. (2019). Parkinson's Disease in Women and Men: What's the Difference? *Journal of Parkinson's Disease*, *9*(3), 501. https://doi.org/10.3233/JPD-191683

Chalancon, G., Kruse, K., & Babu, M. M. (2013). Metabolic Networks, Structure and Dynamics. *Encyclopedia of Systems Biology*, 1263–1267. https://doi.org/10.1007/978-1-4419-9863-7_561

Chang, H. T., Tseng, L. J., Hung, T. J., Kao, B. T., Lin, W. Y., Fan, T. chi, Chang, M. D., & Pai, T. W. (2010). A general co-expression network-based approach to gene expression analysis: comparison and applications. *BMC Systems Biology*, *4*(SUPPL. 2). https://doi.org/10.1186/1752-0509-4-8

Charitou, T., Bryan, K., & Lynn, D. J. (2016). Using biological networks to integrate, visualize and analyze genomics data. *Genetics Selection Evolution*, *48*(1), 1–12. https://doi.org/10.1186/S12711-016-0205-1/FIGURES/4

Charitou, T., Kontou, P. I., Tamposis, I. A., Pavlopoulos, G. A., Braliou, G. G., & Bagos, P. G. (2022). Drug genetic associations with COVID-19 manifestations: a data mining and network biology approach. *The Pharmacogenomics Journal 2022 22:5*, *22*(5), 294–302. https://doi.org/10.1038/s41397-022-00289-1

Checco, P., Biey, M., & Kocarev, L. (n.d.). *Synchronization in random networks with given expected degree sequences*. https://doi.org/10.1016/j.chaos.2006.05.063

Chen, S. G., Stribinskis, V., Rane, M. J., Demuth, D. R., Gozal, E., Roberts, A. M., Jagadapillai, R., Liu, R., Choe, K., Shivakumar, B., Son, F., Jin, S., Kerber, R., Adame, A., Masliah, E., & Friedland, R. P. (2016). *Exposure to the Functional Bacterial Amyloid Protein Curli Enhances Alpha-Synuclein Aggregation in Aged Fischer 344 Rats and Caenorhabditis elegans*. https://doi.org/10.1038/srep34477

Cheng, Y., Tong, Q., Yuan, Y., Song, X., Jiang, W., Wang, Y., Li, W., Li, Y., & Zhang, K. (2023). α-Synuclein induces prodromal symptoms of Parkinson's disease via activating TLR2/MyD88/NF-κB pathway in Schwann cells of vagus nerve in a rat model. *Journal of Neuroinflammation*, *20*(1), 1–18. https://doi.org/10.1186/S12974-023-02720-1/FIGURES/6

Christensen, D. G., Xie, X., Basisty, N., Byrnes, J., McSweeney, S., Schilling, B., & Wolfe, A. J. (2019). Post-translational Protein Acetylation: An Elegant Mechanism for Bacteria to Dynamically Regulate Metabolic Functions. *Frontiers in Microbiology*, *10*(JULY). https://doi.org/10.3389/FMICB.2019.01604

Cirstea, M. S., Yu, A. C., Golz, E., Sundvick, K., Kliger, D., Radisavljevic, N., Foulger, L. H., Mackenzie, M., Huan, T., Finlay, B. B., & Appel-Cresswell, S. (2020). Microbiota Composition and Metabolism Are Associated With Gut Function in Parkinson's Disease. *Movement Disorders : Official Journal of the Movement Disorder Society*, *35*(7), 1208–1217. https://doi.org/10.1002/MDS.28052

Clairembault, T., Leclair-Visonneau, L., Coron, E., Bourreille, A., Le Dily, S., Vavasseur, F., Heymann, M. F., Neunlist, M., & Derkinderen, P. (2015). Structural alterations of the intestinal epithelial barrier in Parkinson's disease. *Acta Neuropathologica Communications*, *3*, 12. https://doi.org/10.1186/S40478-015-0196-0

Clasen, S. J., Bell, M. E. W., Borbón, A., Lee, D. H., Henseler, Z. M., de la Cuesta-Zuluaga, J., Parys, K., Zou, J., Wang, Y., Altmannova, V., Youngblut, N. D., Weir, J. R., Gewirtz, A. T., Belkhadir, Y., & Ley, R. E. (2023). Silent recognition of flagellins from human gut commensal bacteria by Toll-like receptor 5. *Science Immunology*, *8*(79). https://doi.org/10.1126/SCIIMMUNOL.ABQ7001

Cline, M. S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Campilo, I., Creech, M., Gross, B., Hanspers, K., Isserlin, R., Kelley, R.,

Killcoyne, S., Lotia, S., Maere, S., Morris, J., Ono, K., Pavlovic, V., … Bader, G. D. (2007). Integration of biological networks and gene expression data using Cytoscape. *Nature Protocols 2007 2:10*, *2*(10), 2366–2382. https://doi.org/10.1038/nprot.2007.324

Colin, R., Ni, B., Laganenka, L., & Sourjik, V. (2021). Multiple functions of flagellar motility and chemotaxis in bacterial physiology. *FEMS Microbiology Reviews*, *45*(6), 1–19. https://doi.org/10.1093/FEMSRE/FUAB038

Corsini, G. U., Pintus, S., Chiueh, C. C., Weiss, J. F., & Kopin, I. J. (1985). 1-Methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) neurotoxicity in mice is enhanced by pretreatment with diethyldithiocarbamate. *European Journal of Pharmacology*, *119*(1–2), 127–128. https://doi.org/10.1016/0014-2999(85)90331-0

Costa, A., Martín González, A. M., Guizien, K., Doglioli, A. M., Gómez, J. M., Petrenko, A. A., & Allesina, S. (2019). Ecological networks: Pursuing the shortest path, however narrow and crooked. *Scientific Reports 2019 9:1*, *9*(1), 1–13. https://doi.org/10.1038/s41598-019-54206-x

Cruz-Córdova, A., Rocha-Ramírez, L. M., Ochoa, S. A., Gónzalez-Pedrajo, B., Espinosa, N., Eslava, C., Hernández-Chiñas, U., Mendoza-Hernández, G., Rodríguez-Leviz, A., Valencia-Mayoral, P., Sadowinski-Pine, S., Hernández-Castro, R., Estrada-García, I., Muñoz-Hernández, O., Rosas, I., & Xicohtencatl-Cortes, J. (2012). Flagella from Five Cronobacter Species Induce Pro-Inflammatory Cytokines in Macrophage Derivatives from Human Monocytes. *PLOS ONE*, *7*(12), e52091. https://doi.org/10.1371/JOURNAL.PONE.0052091

Cryan, J. F., & Dinan, T. G. (2012). Mind-altering microorganisms: the impact of the gut microbiota on brain and behaviour. *Nature Reviews Neuroscience 2012 13:10*, *13*(10), 701–712. https://doi.org/10.1038/nrn3346

Dalile, B., Van Oudenhove, L., Vervliet, B., & Verbeke, K. (2019). The role of short-chain fatty acids in microbiota–gut–brain communication. *Nature Reviews Gastroenterology & Hepatology 2019 16:8*, *16*(8), 461–478. https://doi.org/10.1038/s41575-019-0157-3

Dank, A., Zeng, Z., Boeren, S., Notebaart, R. A., Smid, E. J., & Abee, T. (2021). Bacterial Microcompartment-Dependent 1,2-Propanediol Utilization of Propionibacterium freudenreichii. *Frontiers in Microbiology*, *12*. https://doi.org/10.3389/FMICB.2021.679827/FULL

Dantas, G., Sommer, M. O. A., Degnan, P. H., & Goodman, A. L. (2013). Experimental Approaches for Defining Functional Roles of Microbes in the Human Gut. *Annual Review of Microbiology*, *67*, 459. https://doi.org/10.1146/ANNUREV-MICRO-092412-155642

David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., Ling, A. V., Devlin, A. S., Varma, Y., Fischbach, M. A., Biddinger, S. B., Dutton, R. J., & Turnbaugh, P. J. (2014). Diet rapidly and reproducibly alters the human gut microbiome. *Nature*, *505*(7484), 559. https://doi.org/10.1038/NATURE12820

De Saedeleer, B., Malabirade, A., Ramiro-Garcia, J., Habier, J., Trezzi, J.-P., Peters, S. L., Daujeumont, A., Halder, R., Jäger, C., Busi, S. B., May, P., Oertel, W., Mollenhauer, B., Laczny, C. C., Hettich, R. L., & Wilmes, P. (2021). Systematic characterization of human gut microbiome-secreted molecules by integrated multi-omics. *ISME Communications*, *1*(1). https://doi.org/10.1038/S43705-021-00078-0

Deng, M., Zhang, K., Mehta, S., Chen, T., & Sun, F. (2004). Prediction of Protein Function Using Protein–Protein Interaction Data. *Https://Home.Liebertpub.Com/Cmb*, *10*(6), 947–960. https://doi.org/10.1089/106652703322756168

Dodet, P., Houot, M., Leu-Semenescu, S., Corvol, J. C., Lehéricy, S., Mangone, G., Vidailhet, M., Roze, E., & Arnulf, I. (2024). Sleep disorders in Parkinson's disease, an early and multiple problem. *Npj Parkinson's Disease 2024 10:1*, *10*(1), 1–11. https://doi.org/10.1038/s41531-024-00642-0

dos-Santos-Pereira, M., Acuña, L., Hamadat, S., Rocca, J., González-Lizárraga, F., Chehín, R., Sepulveda-Diaz, J., Del-Bel, E., Raisman-Vozari, R., & Michel, P. P. (2018). Microglial glutamate release evoked by α-synuclein aggregates is prevented by dopamine. *Glia*, *66*(11), 2353–2365. https://doi.org/10.1002/GLIA.23472

Du, Z., Su, H., Wang, W., Ye, L., Wei, H., Peng, Z., Anishchenko, I., Baker, D., & Yang, J. (2021). The trRosetta server for fast and accurate protein structure prediction. *Nature Protocols 2021 16:12*, *16*(12), 5634–5651. https://doi.org/10.1038/s41596-021-00628-9

Engels, C., Ruscheweyh, H. J., Beerenwinkel, N., Lacroix, C., & Schwab, C. (2016). The common gut microbe Eubacterium hallii also contributes to intestinal propionate formation. *Frontiers in Microbiology*, *7*(MAY), 184615. https://doi.org/10.3389/FMICB.2016.00713/BIBTEX

Erickson, A. R., Cantarel, B. L., Lamendella, R., Darzi, Y., Mongodin, E. F., Pan, C., Shah, M., Halfvarson, J., Tysk, C., Henrissat, B., Raes, J., Verberkmoes, N. C., Fraser, C. M., Hettich, R. L., & Jansson, J. K. (2012). Integrated Metagenomics/Metaproteomics Reveals Human Host-Microbiota Signatures of Crohn's Disease. *PLOS ONE*, *7*(11), e49138. https://doi.org/10.1371/JOURNAL.PONE.0049138

Evans, T. S., & Chen, B. (2022). Linking the network centrality measures closeness and degree. *Communications Physics 2022 5:1*, *5*(1), 1–11. https://doi.org/10.1038/s42005-022-00949-5

Fang, X., Lloyd, C. J., & Palsson, B. O. (2020). Reconstructing organisms in silico: genome-scale models and their emerging applications. *Nature Reviews Microbiology 2020 18:12*, *18*(12), 731–743. https://doi.org/10.1038/S41579-020-00440-4

Ferrocino, I., Rantsiou, K., McClure, R., Kostic, T., de Souza, R. S. C., Lange, L., FitzGerald, J., Kriaa, A., Cotter, P., Maguin, E., Schelkle, B., Schloter, M., Berg, G., Sessitsch, A., & Cocolin, L. (2023). The need for an integrated multi-OMICs approach in microbiome science in the food system. *Comprehensive Reviews in Food Science and Food Safety*, *22*(2), 1082–1103. https://doi.org/10.1111/1541-4337.13103

Figorilli, M., Meloni, M., Lanza, G., Casaglia, E., Lecca, R., Saibene, F. L., Congiu, P., & Puligheddu, M. (2023). Considering REM Sleep Behavior Disorder in the Management of Parkinson's Disease. *Nature and Science of Sleep*, *15*, 333. https://doi.org/10.2147/NSS.S266071

Figura, M., Kuśmierska, K., Bucior, E., Szlufik, S., Koziorowski, D., Jamrozik, Z., & Janik, P. (2018). Serum amino acid profile in patients with Parkinson's disease. *PloS One*, *13*(1). https://doi.org/10.1371/JOURNAL.PONE.0191670

Fricova, D., Harsanyiova, J., & Trancikova, A. K. (2020). Alpha-Synuclein in the Gastrointestinal Tract as a Potential Biomarker for Early Detection of Parkinson's Disease. *International Journal of Molecular Sciences*, *21*(22), 1–21. https://doi.org/10.3390/IJMS21228666

Fuller, T., Langfelder, P., Presson, A., & Horvath, S. (2011). Review of Weighted Gene Coexpression Network Analysis. *Handbook of Statistical Bioinformatics*, 369–388. https://doi.org/10.1007/978-3-642-16345-6_18

Funayama, M., Nishioka, K., Li, Y., & Hattori, N. (2022). Molecular genetics of Parkinson's disease: Contributions and global trends. *Journal of Human Genetics 2022 68:3*, *68*(3), 125–130. https://doi.org/10.1038/s10038-022-01058-5

Galbiati, A., Verga, L., Giora, E., Zucconi, M., & Ferini-Strambi, L. (2019). The risk of neurodegeneration in REM sleep behavior disorder: A systematic review and meta-analysis of longitudinal studies. *Sleep Medicine Reviews*, *43*, 37–46. https://doi.org/10.1016/J.SMRV.2018.09.008

Gargalovic, P. S., Imura, M., Zhang, B., Gharavi, N. M., Clark, M. J., Pagnon, J., Yang, W. P., He, A., Truong, A., Patel, S., Nelson, S. F., Horvath, S., Berliner, J. A., Kirchgessner, T. G., & Lusis, A. J. (2006). Identification of inflammatory gene modules based on variations of human endothelial cell responses to oxidized lipids. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(34), 12741–12746. https://doi.org/10.1073/PNAS.0605457103/SUPPL_FILE/05457TABLE4.XLS

Ghoshal, U., Shukla, R., Srivastava, D., & Ghoshal, U. C. (2016). Irritable Bowel Syndrome, Particularly the Constipation-Predominant Form, Involves an Increase in Methanobrevibacter smithii, Which Is Associated with Higher Methane Production. *Gut and Liver*, *10*(6), 932–938. https://doi.org/10.5009/GNL15588

Glaab, E., Baudot, A., Krasnogor, N., Schneider, R., & Valencia, A. (2012). EnrichNet: network-based gene set enrichment analysis. *Bioinformatics*, *28*(18), i451–i457. https://doi.org/10.1093/BIOINFORMATICS/BTS389

Glatter, T., Ahrné, E., & Schmidt, A. (2015). Comparison of Different Sample Preparation Protocols Reveals Lysis Buffer-Specific Extraction Biases in Gram-Negative Bacteria and Human Cells. *Journal of Proteome Research*, *14*(11), 4472–4485. https://doi.org/10.1021/ACS.JPROTEOME.5B00654

Gligorijević, V., Renfrew, P. D., Kosciolek, T., Leman, J. K., Berenberg, D., Vatanen, T., Chandler, C., Taylor, B. C., Fisk, I. M., Vlamakis, H., Xavier, R. J., Knight, R., Cho, K., & Bonneau, R. (2021). Structure-based protein function prediction using graph convolutional networks. *Nature Communications 2021 12:1*, *12*(1), 1–14. https://doi.org/10.1038/s41467-021-23303-9

Golbeck, J. (2013). Network Structure and Measures. *Analyzing the Social Web*, 25–44. https://doi.org/10.1016/B978-0-12-405531-5.00003-1

Gram, A. M., Wright, J. A., Pickering, R. J., Lam, N. L., Booty, L. M., Webster, S. J., & Bryant, C. E. (2021). Salmonella Flagellin Activates NAIP/NLRC4 and Canonical NLRP3 Inflammasomes in Human Macrophages. *Journal of Immunology (Baltimore, Md. : 1950)*, *206*(3), 631–640. https://doi.org/10.4049/JIMMUNOL.2000382

Gröger, A., Kolb, R., Schäfer, R., & Klose, U. (2014). Dopamine Reduction in the Substantia Nigra of Parkinson's Disease Patients Confirmed by In Vivo Magnetic Resonance Spectroscopic Imaging. *PLOS ONE*, *9*(1), e84081. https://doi.org/10.1371/JOURNAL.PONE.0084081

Gu, Y., Zhang, J., Zhao, X., Nie, W., Xu, X., Liu, M., & Zhang, X. (2024). Olfactory dysfunction and its related molecular mechanisms in Parkinson's disease. *Neural Regeneration Research*, *19*(3), 583. https://doi.org/10.4103/1673-5374.380875

Hagberg, A., Swart, P. J., & Schult, D. A. (2008). *Exploring network structure, dynamics, and function using NetworkX*.

Hahn, M. W., & Kern, A. D. (2005). Comparative Genomics of Centrality and Essentiality in Three Eukaryotic Protein-Interaction Networks. *Molecular Biology and Evolution*, *22*(4), 803–806. https://doi.org/10.1093/MOLBEV/MSI072

Hamnett, R., Bendrick, J. L., Robertson, K., Zhao, E. T., & Kaltschmidt, J. A. (2024). Enteric glutamatergic interneurons regulate intestinal motility. *BioRxiv*, 2024.03.24.586153. https://doi.org/10.1101/2024.03.24.586153

Hanash, S. M., Baik, C. S., & Kallioniemi, O. (2011). Emerging molecular biomarkers—blood-based strategies to detect and monitor cancer. *Nature Reviews Clinical Oncology 2011 8:3*, *8*(3), 142–150. https://doi.org/10.1038/nrclinonc.2010.220

Hardouin, P., Chiron, R., Marchandin, H., Armengaud, J., & Grenga, L. (2021). Metaproteomics to Decipher CF Host-Microbiota Interactions: Overview, Challenges and Future Perspectives. *Genes 2021, Vol. 12, Page 892*, *12*(6), 892. https://doi.org/10.3390/GENES12060892

Hashish, S., & Salama, M. (2023). The Role of an Altered Gut Microbiome in Parkinson's Disease: A Narrative Review. *Applied Microbiology 2023, Vol. 3, Pages 429-447*, *3*(2), 429–447. https://doi.org/10.3390/APPLMICROBIOL3020030

Heintz-Buschart, A., May, P., Laczny, C. C., Lebrun, L. A., Bellora, C., Krishna, A., Wampach, L., Schneider, J. G., Hogan, A., De Beaufort, C., & Wilmes, P. (2016). Integrated multi-omics of the human gut microbiome in a case study of familial type 1 diabetes. *Nature Microbiology 2016 2:1*, *2*(1), 1–13. https://doi.org/10.1038/nmicrobiol.2016.180

Heintz-Buschart, A., Pandey, U., Wicke, T., Sixel-Döring, F., Janzen, A., Sittig-Wiegand, E., Trenkwalder, C., Oertel, W. H., Mollenhauer, B., & Wilmes, P. (2018). The nasal and gut microbiome in Parkinson's disease and idiopathic rapid eye movement sleep behavior disorder. *Movement Disorders : Official Journal of the Movement Disorder Society*, *33*(1), 88–98. https://doi.org/10.1002/MDS.27105

Heintz-Buschart, A., & Wilmes, P. (2018). Human Gut Microbiome: Function Matters. *Trends in Microbiology*, *26*(7), 563–574. https://doi.org/10.1016/J.TIM.2017.11.002

Henry, C., Bassignani, A., Berland, M., Langella, O., Sokol, H., & Juste, C. (2022). Modern Metaproteomics: A Unique Tool to Characterize the Active Microbiome in Health and Diseases, and Pave the Road towards New Biomarkers—Example of Crohn's Disease and Ulcerative Colitis Flare-Ups. *Cells*, *11*(8). https://doi.org/10.3390/CELLS11081340/S1

Hertel, J., Harms, A. C., Heinken, A., Baldini, F., Thinnes, C. C., Glaab, E., Vasco, D. A., Pietzner, M., Stewart, I. D., Wareham, N. J., Langenberg, C., Trenkwalder, C., Krüger, R., Hankemeier, T., Fleming, R. M. T., Mollenhauer, B., & Thiele, I. (2019). Integrated Analyses of Microbiome and Longitudinal Metabolome Data Reveal Microbial-Host Interactions on Sulfur Metabolism in Parkinson's Disease. *Cell Reports*, *29*(7), 1767-1777.e8. https://doi.org/10.1016/J.CELREP.2019.10.035

Heyer, R., Schallert, K., Zoun, R., Becher, B., Saake, G., & Benndorf, D. (2017). Challenges and perspectives of metaproteomic data analysis. *Journal of Biotechnology*, *261*, 24–36. https://doi.org/10.1016/J.JBIOTEC.2017.06.1201

Hirayama, M., Tsunoda, M., Yamamoto, M., Tsuda, T., & Ohno, K. (2016). Serum Tyrosine-to-Phenylalanine Ratio is Low in Parkinson's Disease. *Journal of Parkinson's Disease*, *6*(2), 423–431. https://doi.org/10.3233/JPD-150736

Horvath, S., Zhang, B., Carlson, M., Lu, K. V., Zhu, S., Felciano, R. M., Laurence, M. F., Zhao, W., Qi, S., Chen, Z., Lee, Y., Schecks, A. C., Liau, L. M., Wu, H., Geschwind, D. H., Febbo, P. G., Kornblum, H. I., Cloughesy, T. F., Nelson, S. F., & Mischel, P. S.

(2006). Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(46), 17402–17407. https://doi.org/10.1073/PNAS.0608396103/SUPPL_FILE/IMAGE418.GIF

Hou, Y. fang, Shan, C., Zhuang, S. yue, Zhuang, Q. qian, Ghosh, A., Zhu, K. cheng, Kong, X. ke, Wang, S. min, Gong, Y. ling, Yang, Y. ying, Tao, B., Sun, L. hao, Zhao, H. Y., Guo, X. zhi, Wang, W. qing, Ning, G., Gu, Y. yun, Li, S. tian, & Liu, J. min. (2021). Gut microbiota-derived propionate mediates the neuroprotective effect of osteocalcin in a mouse model of Parkinson's disease. *Microbiome*, *9*(1). https://doi.org/10.1186/S40168-020-00988-6

Hurley, M. J., Bates, R., Macnaughtan, J., & Schapira, A. H. V. (2022). Bile acids and neurological disease. *Pharmacology & Therapeutics*, *240*, 108311. https://doi.org/10.1016/J.PHARMTHERA.2022.108311

Iovino, L., Tremblay, M. E., & Civiero, L. (2020). Glutamate-induced excitotoxicity in Parkinson's disease: The role of glial cells. *Journal of Pharmacological Sciences*, *144*(3), 151–164. https://doi.org/10.1016/J.JPHS.2020.07.011

Iranzo, A., Fernández-Arcos, A., Tolosa, E., Serradell, M., Molinuevo, J. L., Valldeoriola, F., Gelpi, E., Vilaseca, I., Sánchez-Valle, R., Lladó, A., Gaig, C., & Santamaría, J. (2014). Neurodegenerative Disorder Risk in Idiopathic REM Sleep Behavior Disorder: Study in 174 Patients. *PLoS ONE*, *9*(2), 89741. https://doi.org/10.1371/JOURNAL.PONE.0089741

Jameson, B. D., Murdock, S. A., Ji, Q., Stevens, C. J., Grundle, D. S., & Kim Juniper, S. (2023). Network analysis of 16S rRNA sequences suggests microbial keystone taxa contribute to marine N2O cycling. *Communications Biology 2023 6:1*, *6*(1), 1–14. https://doi.org/10.1038/s42003-023-04597-5

Jankovic, J. (2008). Parkinson's disease: clinical features and diagnosis. *Journal of Neurology, Neurosurgery & Psychiatry*, *79*(4), 368–376. https://doi.org/10.1136/JNNP.2007.131045

Jayaraj, R. L., Beiram, R., Azimullah, S., Nagoor Meeran, M. F., Ojha, S. K., Adem, A., & Jalal, F. Y. (2020). Valeric Acid Protects Dopaminergic Neurons by Suppressing Oxidative Stress, Neuroinflammation and Modulating Autophagy Pathways.

International Journal of Molecular Sciences, *21*(20), 1–18. https://doi.org/10.3390/IJMS21207670

Jeong, H., Mason, S. P., Barabási, A. L., & Oltvai, Z. N. (2001a). Lethality and centrality in protein networks. *Nature*, *411*(6833), 41–42. https://doi.org/10.1038/35075138

Jeong, H., Mason, S. P., Barabási, A. L., & Oltvai, Z. N. (2001b). Lethality and centrality in protein networks. *Nature 2001 411:6833*, *411*(6833), 41–42. https://doi.org/10.1038/35075138

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature 2021 596:7873*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kennedy, S. A., Jarboui, M. A., Srihari, S., Raso, C., Bryan, K., Dernayka, L., Charitou, T., Bernal-Llinares, M., Herrera-Montavez, C., Krstic, A., Matallanas, D., Kotlyar, M., Jurisica, I., Curak, J., Wong, V., Stagljar, I., LeBihan, T., Imrie, L., Pillai, P., … Kolch, W. (2020). Extensive rewiring of the EGFR network in colorectal cancer cells expressing transforming levels of KRASG13D. *Nature Communications 2020 11:1*, *11*(1), 1–14. https://doi.org/10.1038/s41467-019-14224-9

Keshavarzian, A., Green, S. J., Engen, P. A., Voigt, R. M., Naqib, A., Forsyth, C. B., Mutlu, E., & Shannon, K. M. (2015). Colonic bacterial composition in Parkinson's disease. *Movement Disorders*, *30*(10), 1351–1360. https://doi.org/10.1002/MDS.26307

Kim, G., Deepinder, F., Morales, W., Hwang, L., Weitsman, S., Chang, C., Gunsalus, R., & Pimentel, M. (2012). Methanobrevibacter smithii is the predominant methanogen in patients with constipation-predominant IBS and methane on breath. *Digestive Diseases and Sciences*, *57*(12), 3213–3218. https://doi.org/10.1007/S10620-012-2197-1/FIGURES/8

Kim, H. J., Huh, D., Hamilton, G., & Ingber, D. E. (2012). Human gut-on-a-chip inhabited by microbial flora that experiences intestinal peristalsis-like motions and flow. *Lab on a Chip*, *12*(12), 2165–2174. https://doi.org/10.1039/C2LC40074J

Kline, E. M., Houser, M. C., Herrick, M. K., Seibler, P., Klein, C., West, A., & Tansey, M. G. (2021). Genetic and Environmental Factors in Parkinson's Disease Converge on Immune Function and Inflammation. *Movement Disorders*, *36*(1), 25–36. https://doi.org/10.1002/MDS.28411

Kolicheski, A., Turcano, P., Tamvaka, N., McLean, P. J., Springer, W., Savica, R., & Ross, O. A. (2022). Early-Onset Parkinson's Disease: Creating the Right Environment for a Genetic Disorder. *Journal of Parkinson's Disease*, *12*(8), 2353. https://doi.org/10.3233/JPD-223380

Kolli, U., Jalodia, R., Moidunny, S., Singh, P. K., Ban, Y., Tao, J., Cantu, G. N., Valdes, E., Ramakrishnan, S., & Roy, S. (2023). Multi-omics analysis revealing the interplay between gut microbiome and the host following opioid use. *Gut Microbes*, *15*(2). https://doi.org/10.1080/19490976.2023.2246184

Kondoh, T., Mallick, H. N., & Torii, K. (2009). Activation of the gut-brain axis by dietary glutamate and physiologic significance in energy homeostasis. *The American Journal of Clinical Nutrition*, *90*(3), 832S-837S. https://doi.org/10.3945/AJCN.2009.27462V

Kondoh, T., & Torii, K. (2008). Brain Activation by Umami Substances via Gustatory and Visceral Signaling Pathways, and Physiological Significance. *Biological and Pharmaceutical Bulletin*, *31*(10), 1827–1832. https://doi.org/10.1248/BPB.31.1827

Koonin, E. V., Wolf, Y. I., & Karev, G. P. (2002). The structure of the protein universe and genome evolution. *Nature 2002 420:6912*, *420*(6912), 218–223. https://doi.org/10.1038/NATURE01256

Koschützki, D., & Schreiber, F. (2008). Centrality analysis methods for biological networks and their application to gene regulatory networks. *Gene Regulation and Systems Biology*, *2008*(2), 193–201. https://doi.org/10.4137/GRSB.S702/ASSET/IMAGES/LARGE/10.4137_GRSB.S702-FIG4.JPEG

Kuraku, S., & Ukena, K. (2021). Molecular evolution of peptide and protein hormones in vertebrates. *Handbook of Hormones: Comparative Endocrinology for Basic and Clinical Research*, 7–9. https://doi.org/10.1016/B978-0-12-820649-2.00002-4

Lagier, J. C., Khelaifia, S., Alou, M. T., Ndongo, S., Dione, N., Hugon, P., Caputo, A., Cadoret, F., Traore, S. I., Seck, E. H., Dubourg, G., Durand, G., Mourembou, G.,

Guilhot, E., Togo, A., Bellali, S., Bachar, D., Cassir, N., Bittar, F., … Raoult, D. (2016). Culture of previously uncultured members of the human gut microbiota by culturomics. *Nature Microbiology*, *1*. https://doi.org/10.1038/NMICROBIOL.2016.203

Lai, L. A., Tong, Z., Chen, R., & Pan, S. (2019). Metaproteomics Study of the Gut Microbiome. *Methods in Molecular Biology (Clifton, N.J.)*, *1871*, 123–132. https://doi.org/10.1007/978-1-4939-8814-3_8

Langfelder, P., & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*(1), 1–13. https://doi.org/10.1186/1471-2105-9-559/FIGURES/4

Lanpher, B., Brunetti-Pierri, N., & Lee, B. (2006). Inborn errors of metabolism: the flux from Mendelian to complex diseases. *Nature Reviews. Genetics*, *7*(6), 449–460. https://doi.org/10.1038/NRG1880

Laskowski, R. A., Watson, J. D., & Thornton, J. M. (2005). ProFunc: a server for predicting protein function from 3D structure. *Nucleic Acids Research*, *33*(suppl_2), W89–W93. https://doi.org/10.1093/NAR/GKI414

Lau, M. K., Borrett, S. R., Baiser, B., Gotelli, N. J., & Ellison, A. M. (2017). Ecological network metrics: opportunities for synthesis. *Ecosphere*, *8*(8), e01900. https://doi.org/10.1002/ECS2.1900

Lema, N. K., Gemeda, M. T., & Woldesemayat, A. A. (2023). Recent Advances in Metagenomic Approaches, Applications, and Challenges. *Current Microbiology 2023 80:11*, *80*(11), 1–14. https://doi.org/10.1007/S00284-023-03451-5

Li, D., Liu, C. M., Luo, R., Sadakane, K., & Lam, T. W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics (Oxford, England)*, *31*(10), 1674–1676. https://doi.org/10.1093/BIOINFORMATICS/BTV033

Li, P., Killinger, B. A., Ensink, E., Beddows, I., Yilmaz, A., Lubben, N., Lamp, J., Schilthuis, M., Vega, I. E., Woltjer, R., Andrew Pospisilik, J., Brundin, P., Brundin, L., Graham, S. F., & Labrie, V. (2021a). Gut Microbiota Dysbiosis Is Associated with Elevated Bile Acids in Parkinson's Disease. *Metabolites*, *11*(1), 1–15. https://doi.org/10.3390/METABO11010029

Li, P., Killinger, B. A., Ensink, E., Beddows, I., Yilmaz, A., Lubben, N., Lamp, J., Schilthuis, M., Vega, I. E., Woltjer, R., Andrew Pospisilik, J., Brundin, P., Brundin, L., Graham, S. F., & Labrie, V. (2021b). Gut Microbiota Dysbiosis Is Associated with Elevated Bile Acids in Parkinson's Disease. *Metabolites*, *11*(1), 1–15. https://doi.org/10.3390/METABO11010029

Li, Q., Meng, L. B., Chen, L. J., Shi, X., Tu, L., Zhou, Q., Yu, J. L., Liao, X., Zeng, Y., & Yuan, Q. Y. (2023). The role of the microbiota-gut-brain axis and intestinal microbiome dysregulation in Parkinson's disease. *Frontiers in Neurology*, *14*, 1185375. https://doi.org/10.3389/FNEUR.2023.1185375/BIBTEX

Li, W., Wu, X., Hu, X., Wang, T., Liang, S., Duan, Y., Jin, F., & Qin, B. (2017). Structural changes of gut microbiota in Parkinson's disease and its correlation with clinical features. *Science China. Life Sciences*, *60*(11), 1223–1233. https://doi.org/10.1007/S11427-016-9001-4

Li, Y., Dong, J., Xiao, H., Zhang, S., Wang, B., Cui, M., & Fan, S. (2020). Gut commensal derived-valeric acid protects against radiation injuries. *Gut Microbes*, *11*(4), 789–806. https://doi.org/10.1080/19490976.2019.1709387

Li, Z., Liang, H., Hu, Y., Lu, L., Zheng, C., Fan, Y., Wu, B., Zou, T., Luo, X., Zhang, X., Zeng, Y., Liu, Z., Zhou, Z., Yue, Z., Ren, Y., Li, Z., Su, Q., & Xu, P. (2023). Gut bacterial profiles in Parkinson's disease: A systematic review. *CNS Neuroscience & Therapeutics*, *29*(1), 140. https://doi.org/10.1111/CNS.13990

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)*, *30*(7), 923–930. https://doi.org/10.1093/BIOINFORMATICS/BTT656

Liddle, R. A. (2018). Parkinson's disease from the gut. *Brain Research*, *1693*, 201–206. https://doi.org/10.1016/J.BRAINRES.2018.01.010

Lin, C. L. G., Kong, Q., Cuny, G. D., & Glicksman, M. A. (2012). Glutamate Transporter EAAT2: A New Target For The Treatment Of Neurodegenerative Diseases. *Future Medicinal Chemistry*, *4*(13), 1689–1700. https://doi.org/10.4155/FMC.12.122

Liu, J., Chen, Q., & Su, R. (2024). Interplay of human gastrointestinal microbiota metabolites: Short-chain fatty acids and their correlation with Parkinson's disease. *Medicine*, *103*(17), E37960. https://doi.org/10.1097/MD.0000000000037960

Liu, L., Wei, J., & Ruan, J. (2017). Pathway Enrichment Analysis with Networks. *Genes*, *8*(10). https://doi.org/10.3390/GENES8100246

Liu, M. T., Rothstein, J. D., Gershon, M. D., & Kirchgessner, A. L. (1997). Glutamatergic Enteric Neurons. *Journal of Neuroscience*, *17*(12), 4764–4784. https://doi.org/10.1523/JNEUROSCI.17-12-04764.1997

Liu, W., Lim, K. L., & Tan, E. K. (2022). Intestine-derived α-synuclein initiates and aggravates pathogenesis of Parkinson's disease in Drosophila. *Translational Neurodegeneration*, *11*(1), 1–18. https://doi.org/10.1186/S40035-022-00318-W/FIGURES/7

Lucchetti, M., Kaminska, M., Oluwasegun, A. K., Mosig, A. S., & Wilmes, P. (2021). Emulating the gut-liver axis: Dissecting the microbiome's effect on drug metabolism using multiorgan-on-chip models. *Current Opinion in Endocrine and Metabolic Research*, *18*, 94–101. https://doi.org/10.1016/J.COEMR.2021.03.003

Luck, K., Kim, D. K., Lambourne, L., Spirohn, K., Begg, B. E., Bian, W., Brignall, R., Cafarelli, T., Campos-Laborie, F. J., Charloteaux, B., Choi, D., Coté, A. G., Daley, M., Deimling, S., Desbuleux, A., Dricot, A., Gebbia, M., Hardy, M. F., Kishore, N., … Calderwood, M. A. (2020). A reference map of the human binary protein interactome. *Nature 2020 580:7803*, *580*(7803), 402–408. https://doi.org/10.1038/S41586-020-2188-X

Ma, Q., Zhang, Q., Chen, Y., Yu, S., Huang, J., Liu, Y., Gong, T., Li, Y., & Zou, J. (2021). Post-translational Modifications in Oral Bacteria and Their Functional Impact. *Frontiers in Microbiology*, *12*, 784923. https://doi.org/10.3389/FMICB.2021.784923/BIBTEX

Ma, T., Huo, Z., Kuo, A., Zhu, L., Fang, Z., Zeng, X., Lin, C. W., Liu, S., Wang, L., Liu, P., Rahman, T., Chang, L. C., Kim, S., Li, J., Park, Y., Song, C., Oesterreich, S., Sibille, E., & Tseng, G. C. (2019). MetaOmics: analysis pipeline and browser-based software suite for transcriptomic meta-analysis. *Bioinformatics*, *35*(9), 1597–1599. https://doi.org/10.1093/BIOINFORMATICS/BTY825

Macek, B., Forchhammer, K., Hardouin, J., Weber-Ban, E., Grangeasse, C., & Mijakovic, I. (2019). Protein post-translational modifications in bacteria. *Nature Reviews Microbiology 2019 17:11*, *17*(11), 651–664. https://doi.org/10.1038/s41579-019-0243-0

Macfarlane, G. T., Steed, H., & Macfarlane, S. (2008). Bacterial metabolism and health-related effects of galacto-oligosaccharides and other prebiotics. *Journal of Applied Microbiology*, *104*(2), 305–344. https://doi.org/10.1111/J.1365-2672.2007.03520.X

Mahbub, N. U., Islam, M. M., Hong, S. T., & Chung, H. J. (2024). Dysbiosis of the gut microbiota and its effect on α-synuclein and prion protein misfolding: consequences for neurodegeneration. *Frontiers in Cellular and Infection Microbiology*, *14*, 1348279. https://doi.org/10.3389/FCIMB.2024.1348279/BIBTEX

Majumdar, A., Siva Venkatesh, I. P., & Basu, A. (2023). Short-Chain Fatty Acids in the Microbiota–Gut–Brain Axis: Role in Neurodegenerative Disorders and Viral Infections. *ACS Chemical Neuroscience*, *14*(6), 1045–1062. https://doi.org/10.1021/ACSCHEMNEURO.2C00803

Mamelak, M. (2018). Parkinson's Disease, the Dopaminergic Neuron and Gammahydroxybutyrate. *Neurology and Therapy*, *7*(1), 5. https://doi.org/10.1007/S40120-018-0091-2

Marcone, G. L., Rosini, E., Crespi, E., & Pollegioni, L. (2019). D-amino acids in foods. *Applied Microbiology and Biotechnology 2019 104:2*, *104*(2), 555–574. https://doi.org/10.1007/S00253-019-10264-9

Masato, A., Plotegher, N., Boassa, D., & Bubacco, L. (2019). Impaired dopamine metabolism in Parkinson's disease pathogenesis. *Molecular Neurodegeneration 2019 14:1*, *14*(1), 1–21. https://doi.org/10.1186/S13024-019-0332-6

Mattammal, M. B., Haring, J. H., Chung, H. D., Raghu, G., & Strong, R. (1995). An endogenous dopaminergic neurotoxin: Implication for Parkinson's disease. *Neurodegeneration*, *4*(3), 271–281. https://doi.org/10.1016/1055-8330(95)90016-0

McCarville, J. L., Chen, G. Y., Cuevas, V. D., Troha, K., & Ayres, J. S. (2020). Microbiota Metabolites in Health and Disease. *Annual Review of Immunology*, *38*(Volume 38, 2020), 147–170. https://doi.org/10.1146/ANNUREV-IMMUNOL-071219-125715/CITE/REFWORKS

Meldrum, B. S. (2000). Glutamate as a Neurotransmitter in the Brain: Review of Physiology and Pathology. *The Journal of Nutrition*, *130*(4), 1007S-1015S. https://doi.org/10.1093/JN/130.4.1007S

Meredith, G. E., Totterdell, S., Beales, M., & Meshul, C. K. (2009). Impaired glutamate homeostasis and programmed cell death in a chronic MPTP mouse model of Parkinson's disease. *Experimental Neurology*, *219*(1), 334–340. https://doi.org/10.1016/J.EXPNEUROL.2009.06.005

Milanese, A., Mende, D. R., Paoli, L., Salazar, G., Ruscheweyh, H. J., Cuenca, M., Hingamp, P., Alves, R., Costea, P. I., Coelho, L. P., Schmidt, T. S. B., Almeida, A., Mitchell, A. L., Finn, R. D., Huerta-Cepas, J., Bork, P., Zeller, G., & Sunagawa, S. (2019). Microbial abundance, activity and population genomic profiling with mOTUs2. *Nature Communications 2019 10:1*, *10*(1), 1–11. https://doi.org/10.1038/s41467-019-08844-4

Mills, R. H., Dulai, P. S., Vázquez-Baeza, Y., Sauceda, C., Daniel, N., Gerner, R. R., Batachari, L. E., Malfavon, M., Zhu, Q., Weldon, K., Humphrey, G., Carrillo-Terrazas, M., Goldasich, L. D. R., Bryant, M. K., Raffatellu, M., Quinn, R. A., Gewirtz, A. T., Chassaing, B., Chu, H., … Gonzalez, D. J. (2022). Multi-omics analyses of the ulcerative colitis gut microbiome link Bacteroides vulgatus proteases with disease severity. *Nature Microbiology*, *7*(2), 262–276. https://doi.org/10.1038/S41564-021-01050-3

Miyauchi, E., Shimokawa, C., Steimle, A., Desai, M. S., & Ohno, H. (2022). The impact of the gut microbiome on extra-intestinal autoimmune diseases. *Nature Reviews Immunology 2022 23:1*, *23*(1), 9–23. https://doi.org/10.1038/s41577-022-00727-y

Moniruzzaman, M., Wurch, L. L., Alexander, H., Dyhrman, S. T., Gobler, C. J., & Wilhelm, S. W. (2017). Virus-host relationships of marine single-celled eukaryotes resolved from metatranscriptomics. *Nature Communications 2017 8:1*, *8*(1), 1–10. https://doi.org/10.1038/ncomms16054

Mor, D. E., Daniels, M. J., & Ischiropoulos, H. (2019). The usual suspects, dopamine and alpha synuclein, conspire to cause neurodegeneration. *Movement Disorders : Official Journal of the Movement Disorder Society*, *34*(2), 167. https://doi.org/10.1002/MDS.27607

Mor, D. E., Tsika, E., Mazzulli, J. R., Gould, N. S., Kim, H., Daniels, M. J., Doshi, S., Gupta, P., Grossman, J. L., Tan, V. X., Kalb, R. G., Caldwell, K. A., Caldwell, G. A., Wolfe, J. H., & Ischiropoulos, H. (2017). Dopamine induces soluble α-synuclein oligomers and

nigrostriatal degeneration. *Nature Neuroscience 2017 20:11*, *20*(11), 1560–1568. https://doi.org/10.1038/nn.4641

Morais, L. H., Schreiber, H. L., & Mazmanian, S. K. (2020). The gut microbiota–brain axis in behaviour and brain disorders. *Nature Reviews Microbiology 2020 19:4*, *19*(4), 241–255. https://doi.org/10.1038/s41579-020-00460-0

Moreira de Gouveia, M. I., Daniel, J., Garrivier, A., Bernalier-Donadille, A., & Jubelin, G. (2023). Diversity of ethanolamine utilization by human commensal Escherichia coli. *Research in Microbiology*, *174*(1–2), 103989. https://doi.org/10.1016/J.RESMIC.2022.103989

Morrison, D. J., & Preston, T. (2016). Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. *Gut Microbes*, *7*(3), 189–200. https://doi.org/10.1080/19490976.2015.1134082

Mu, L., Chen, J., Sobotka, S., Nyirenda, T., Benson, B., Gupta, F., Sanders, I., Adler, C. H., Caviness, J. N., Shill, H. A., Sabbagh, M., Samanta, J. E., Sue, L. I., & Beach, T. G. (2015). Alpha-Synuclein Pathology in Sensory Nerve Terminals of the Upper Aerodigestive Tract of Parkinson's Disease Patients. *Dysphagia*, *30*(4), 404–417. https://doi.org/10.1007/S00455-015-9612-7

Muzio, L., Viotti, A., & Martino, G. (2021). Microglia in Neuroinflammation and Neurodegeneration: From Understanding to Therapy. *Frontiers in Neuroscience*, *15*, 742065. https://doi.org/10.3389/FNINS.2021.742065/BIBTEX

Narayanasamy, S., Jarosz, Y., Muller, E. E. L., Heintz-Buschart, A., Herold, M., Kaysen, A., Laczny, C. C., Pinel, N., May, P., & Wilmes, P. (2016). IMP: A pipeline for reproducible reference-independent integrated metagenomic and metatranscriptomic analyses. *Genome Biology*, *17*(1), 1–21. https://doi.org/10.1186/S13059-016-1116-8/FIGURES/5

Nayfach, S., Roux, S., Seshadri, R., Udwary, D., Varghese, N., Schulz, F., Wu, D., Paez-Espino, D., Chen, I. M., Huntemann, M., Palaniappan, K., Ladau, J., Mukherjee, S., Reddy, T. B. K., Nielsen, T., Kirton, E., Faria, J. P., Edirisinghe, J. N., Henry, C. S., … Eloe-Fadrosh, E. A. (2020). A genomic catalog of Earth's microbiomes. *Nature Biotechnology 2020 39:4*, *39*(4), 499–509. https://doi.org/10.1038/s41587-020-0718-6

Ngo, K. J., Paul, K. C., Wong, D., Kusters, C. D. J., Bronstein, J. M., Ritz, B., & Fogel, B. L. (2024). Lysosomal genes contribute to Parkinson's disease near agriculture with high intensity pesticide use. *Npj Parkinson's Disease 2024 10:1*, *10*(1), 1–10. https://doi.org/10.1038/s41531-024-00703-4

Nie, S., & Ge, Y. (2023). The link between the gut microbiome, inflammation, and Parkinson's disease. *Applied Microbiology and Biotechnology 2023 107:22*, *107*(22), 6737–6749. https://doi.org/10.1007/S00253-023-12789-6

Nikoloski, Z., Grimbs, S., May, P., & Selbig, J. (2008). Metabolic networks are NP-hard to reconstruct. *Journal of Theoretical Biology*, *254*(4), 807–816. https://doi.org/10.1016/J.JTBI.2008.07.015

Nikoloski, Z., May, P., & Selbig, J. (2010). Algebraic connectivity may explain the evolution of gene regulatory networks. *Journal of Theoretical Biology*, *267*(1), 7–14. https://doi.org/10.1016/J.JTBI.2010.07.028

Novikova, P. V, Busi, S. B., Probst, A. J., May, P., & Wilmes, P. (2024). Functional prediction of proteins from the human gut archaeome. *ISME Communications*, *4*(1), 14. https://doi.org/10.1093/ISMECO/YCAD014

O'Gorman Tuura, R. L., Baumann, C. R., & Baumann-Vogel, H. (2018). Beyond dopamine: GABA, glutamate, and the axial symptoms of Parkinson disease. *Frontiers in Neurology*, *9*(SEP), 397277. https://doi.org/10.3389/FNEUR.2018.00806/BIBTEX

Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, *32*(3), 245–251. https://doi.org/10.1016/J.SOCNET.2010.03.006

Ortiz de Ora, L., Balsamo, J. M., Uyeda, K. S., & Bess, E. N. (2024). Discovery of a Gut Bacterial Metabolic Pathway that Drives α-Synuclein Aggregation. *ACS Chemical Biology*, *19*(4), 1011–1021. https://doi.org/10.1021/ACSCHEMBIO.4C00095/ASSET/IMAGES/LARGE/CB4C00095_0005.JPEG

Paudel, D., Nair, D. V. T., Joseph, G., Castro, R., Tiwari, A. K., & Singh, V. (2023). Gastrointestinal microbiota-directed nutritional and therapeutic interventions for inflammatory bowel disease: opportunities and challenges. *Gastroenterology Report*, *12*. https://doi.org/10.1093/GASTRO/GOAE033

Pavlopoulos, G. A., Secrier, M., Moschopoulos, C. N., Soldatos, T. G., Kossida, S., Aerts, J., Schneider, R., & Bagos, P. G. (2011). Using graph theory to analyze biological networks. *BioData Mining*, *4*(1), 1–27. https://doi.org/10.1186/1756-0381-4-10/FIGURES/11

Pei, G., Chen, L., & Zhang, W. (2017). WGCNA Application to Proteomic and Metabolomic Data Analysis. *Methods in Enzymology*, *585*, 135–158. https://doi.org/10.1016/BS.MIE.2016.09.016

Pérez-Cano, F. J. (2022). Dietary Modulation of the Immune Function: Direct and Microbiota-Dependent Effect. *Nutrients*, *14*(9), 1957. https://doi.org/10.3390/NU14091957

Perrin, L., Spinosi, J., Chaperon, L., Kab, S., Moisan, F., & Ebaz, A. (2021). Pesticides expenditures by farming type and incidence of Parkinson disease in farmers: A French nationwide study. *Environmental Research*, *197*, 111161. https://doi.org/10.1016/J.ENVRES.2021.111161

Peterson, C. T., Perez Santiago, J., Iablokov, S. N., Chopra, D., Rodionov, D. A., & Peterson, S. N. (2022). Short-Chain Fatty Acids Modulate Healthy Gut Microbiota Composition and Functional Potential. *Current Microbiology*, *79*(5), 1–13. https://doi.org/10.1007/S00284-022-02825-5/FIGURES/5

Pitz, V., Makarious, M. B., Bandres-Ciga, S., Iwaki, H., Aslibekyan, S., Auton, A., Babalola, E., Bell, R. K., Bielenberg, J., Bryc, K., Bullis, E., Coker, D., Partida, G. C., Dhamija, D., Das, S., Elson, S. L., Eriksson, N., Filshtein, T., Fitch, A., … Blauwendraat, C. (2024). Analysis of rare Parkinson's disease variants in millions of people. *Npj Parkinson's Disease 2024 10:1*, *10*(1), 1–10. https://doi.org/10.1038/s41531-023-00608-8

Plaimas, K., Eils, R., & König, R. (2010). Identifying essential genes in bacterial metabolic networks with machine learning methods. *BMC Systems Biology*, *4*(1), 1–16. https://doi.org/10.1186/1752-0509-4-56/FIGURES/4

Postuma, R. B., Gagnon, J. F., Bertrand, J. A., Génier Marchand, D., & Montplaisir, J. Y. (2015). Parkinson risk in idiopathic REM sleep behavior disorder: Preparing for neuroprotective trials. *Neurology*, *84*(11), 1104. https://doi.org/10.1212/WNL.0000000000001364

Postuma, R. B., Gagnon, J. F., Vendette, M., Fantini, M. L., Massicotte-Marquez, J., & Montplaisir, J. (2009). Quantifying the risk of neurodegenerative disease in idiopathic REM sleep behavior disorder. *Neurology*, *72*(15), 1296. https://doi.org/10.1212/01.WNL.0000340980.19702.6E

Pouchieu, C., Piel, C., Carles, C., Gruber, A., Helmer, C., Tual, S., Marcotullio, E., Lebailly, P., & Baldi, I. (2018). Pesticide use in agriculture and Parkinson's disease in the AGRICAN cohort study. *International Journal of Epidemiology*, *47*(1), 299–310. https://doi.org/10.1093/IJE/DYX225

Praveschotinunt, P., Duraj-Thatte, A. M., Gelfat, I., Bahl, F., Chou, D. B., & Joshi, N. S. (2019). Engineered E. coli Nissle 1917 for the delivery of matrix-tethered therapeutic domains to the gut. *Nature Communications 2019 10:1*, *10*(1), 1–14. https://doi.org/10.1038/s41467-019-13336-6

Price, M. N., Dehal, P. S., & Arkin, A. P. (2007). Orthologous Transcription Factors in Bacteria Have Different Functions and Regulate Different Genes. *PLOS Computational Biology*, *3*(9), e175. https://doi.org/10.1371/JOURNAL.PCBI.0030175

Qian, F., Guo, A., Li, M., Liu, W., Pan, Z., Jiang, L., Wu, X., & Xu, H. (2015). Salmonella flagellin is a potent carrier-adjuvant for peptide conjugate to induce peptide-specific antibody response in mice. *Vaccine*, *33*(17), 2038–2044. https://doi.org/10.1016/J.VACCINE.2015.03.006

Qian, Y., Yang, X., Xu, S., Huang, P., Li, B., Du, J., He, Y., Su, B., Xu, L. M., Wang, L., Huang, R., Chen, S., & Xiao, Q. (2020). Gut metagenomics-derived genes as potential biomarkers of Parkinson's disease. *Brain : A Journal of Neurology*, *143*(8), 2474–2489. https://doi.org/10.1093/BRAIN/AWAA201

Qian, Y., Yang, X., Xu, S., Wu, C., Song, Y., Qin, N., Chen, S. Di, & Xiao, Q. (2018). Alteration of the fecal microbiota in Chinese patients with Parkinson's disease. *Brain, Behavior, and Immunity*, *70*, 194–202. https://doi.org/10.1016/J.BBI.2018.02.016

Quan, Y., Song, K., Zhang, Y., Zhu, C., Shen, Z., Wu, S., Luo, W., Tan, B., Yang, Z., & Wang, X. (2018). Roseburia intestinalis-derived flagellin is a negative regulator of intestinal inflammation. *Biochemical and Biophysical Research Communications*, *501*(3), 791–799. https://doi.org/10.1016/J.BBRC.2018.05.075

Queirós, P., Delogu, F., Hickl, O., May, P., & Wilmes, P. (2021). Mantis: flexible and consensus-driven genome annotation. *GigaScience*, *10*(6), 1–14. https://doi.org/10.1093/GIGASCIENCE/GIAB042

Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., & Barabási, A. L. (2002). Hierarchical organization of modularity in metabolic networks. *Science*, *297*(5586), 1551–1555. https://doi.org/10.1126/SCIENCE.1073374/SUPPL_FILE/1073374S.PDF

Ravenhill, S. M., Evans, A. H., & Crewther, S. G. (2023). Escalating Bi-Directional Feedback Loops between Proinflammatory Microglia and Mitochondria in Ageing and Post-Diagnosis of Parkinson's Disease. *Antioxidants 2023, Vol. 12, Page 1117*, *12*(5), 1117. https://doi.org/10.3390/ANTIOX12051117

Reeds, P. J., Burrin, D. G., Stoll, B., & Jahoor, F. (2000). Intestinal glutamate metabolism. *The Journal of Nutrition*, *130*(4S Suppl). https://doi.org/10.1093/JN/130.4.978S

Rezaie, N., Reese, F., & Mortazavi, A. (2023). PyWGCNA: a Python package for weighted gene co-expression network analysis. *Bioinformatics*, *39*(7). https://doi.org/10.1093/BIOINFORMATICS/BTAD415

Ridlon, J. M., Kang, D. J., & Hylemon, P. B. (2006). Bile salt biotransformations by human intestinal bacteria. *Journal of Lipid Research*, *47*(2), 241–259. https://doi.org/10.1194/JLR.R500013-JLR200

Rios-Covian, D., González, S., Nogacka, A. M., Arboleya, S., Salazar, N., Gueimonde, M., & de los Reyes-Gavilán, C. G. (2020). An Overview on Fecal Branched Short-Chain Fatty Acids Along Human Life and as Related With Body Mass Index: Associated Dietary and Anthropometric Factors. *Frontiers in Microbiology*, *11*, 513909. https://doi.org/10.3389/FMICB.2020.00973/BIBTEX

Robertson, D. E., Roberts, M. F., Belay, N., Stetter, K. O., & Boone, D. R. (1990a). Occurrence of beta-glutamate, a novel osmolyte, in marine methanogenic bacteria. *Applied and Environmental Microbiology*, *56*(5), 1504–1508. https://doi.org/10.1128/AEM.56.5.1504-1508.1990

Robertson, D. E., Roberts, M. F., Belay, N., Stetter, K. O., & Boone, D. R. (1990b). Occurrence of beta-glutamate, a novel osmolyte, in marine methanogenic bacteria. *Applied and Environmental Microbiology*, *56*(5), 1504–1508. https://doi.org/10.1128/AEM.56.5.1504-1508.1990

Robinson, P., Neelon, K., Schreier, H. J., & Roberts, M. F. (2001a). β-Glutamate as a Substrate for Glutamine Synthetase. *Applied and Environmental Microbiology*, *67*(10), 4458. https://doi.org/10.1128/AEM.67.10.4458-4463.2001

Robinson, P., Neelon, K., Schreier, H. J., & Roberts, M. F. (2001b). β-Glutamate as a Substrate for Glutamine Synthetase. *Applied and Environmental Microbiology*, *67*(10), 4458. https://doi.org/10.1128/AEM.67.10.4458-4463.2001

Romano, S., Savva, G. M., Bedarf, J. R., Charles, I. G., Hildebrand, F., & Narbad, A. (2021). Meta-analysis of the Parkinson's disease gut microbiome suggests alterations linked to intestinal inflammation. *Npj Parkinson's Disease 2021 7:1*, *7*(1), 1–13. https://doi.org/10.1038/s41531-021-00156-z

Roos, D. S., Klein, M., Deeg, D. J. H., Doty, R. L., & Berendse, H. W. (2022). Prevalence of Prodromal Symptoms of Parkinson's Disease in the Late Middle-Aged Population. *Journal of Parkinson's Disease*, *12*(3), 967. https://doi.org/10.3233/JPD-213007

Rosario, D., Bidkhori, G., Lee, S., Bedarf, J., Hildebrand, F., Le Chatelier, E., Uhlen, M., Ehrlich, S. D., Proctor, G., Wüllner, U., Mardinoglu, A., & Shoaie, S. (2021). Systematic analysis of gut microbiome reveals the role of bacterial folate and homocysteine metabolism in Parkinson's disease. *Cell Reports*, *34*(9). https://doi.org/10.1016/J.CELREP.2021.108807

Roume, H., Heintz-Buschart, A., Muller, E. E. L., May, P., Satagopam, V. P., Laczny, C. C., Narayanasamy, S., Lebrun, L. A., Hoopmann, M. R., Schupp, J. M., Gillece, J. D., Hicks, N. D., Engelthaler, D. M., Sauter, T., Keim, P. S., Moritz, R. L., & Wilmes, P. (2015). Comparative integrated omics: identification of key functionalities in microbial community-wide metabolic networks. *NPJ Biofilms and Microbiomes*, *1*, 15007. https://doi.org/10.1038/NPJBIOFILMS.2015.7

Sampson, T. R., Debelius, J. W., Thron, T., Janssen, S., Shastri, G. G., Ilhan, Z. E., Challis, C., Schretter, C. E., Rocha, S., Gradinaru, V., Chesselet, M. F., Keshavarzian, A., Shannon, K. M., Krajmalnik-Brown, R., Wittung-Stafshede, P., Knight, R., & Mazmanian, S. K. (2016). Gut Microbiota Regulate Motor Deficits and Neuroinflammation in a Model of Parkinson's Disease. *Cell*, *167*(6), 1469-1480.e12. https://doi.org/10.1016/J.CELL.2016.11.018/ATTACHMENT/D4F49F21-9693-4F57-9D34-5E243AB4FCFB/MMC1.PDF

Savica, R., Grossardt, B. R., Bower, J. H., Eric Ahlskog, J., & Rocca, W. A. (2016). Time Trends in the Incidence of Parkinson Disease. *JAMA Neurology*, *73*(8), 981–989. https://doi.org/10.1001/JAMANEUROL.2016.0947

Scheffer, D. D. L., Freitas, F. C., Aguiar, A. S., Ward, C., Guglielmo, L. G. A., Prediger, R. D., Cronin, S. J. F., Walz, R., Andrews, N. A., & Latini, A. (2021). Impaired dopamine metabolism is linked to fatigability in mice and fatigue in Parkinson's disease patients. *Brain Communications*, *3*(3). https://doi.org/10.1093/BRAINCOMMS/FCAB116

Scheperjans, F., Aho, V., Pereira, P. A. B., Koskinen, K., Paulin, L., Pekkonen, E., Haapaniemi, E., Kaakkola, S., Eerola-Rautio, J., Pohja, M., Kinnunen, E., Murros, K., & Auvinen, P. (2015). Gut microbiota are related to Parkinson's disease and clinical phenotype. *Movement Disorders : Official Journal of the Movement Disorder Society*, *30*(3), 350–358. https://doi.org/10.1002/MDS.26069

Schmit, K. J., Garcia, P., Sciortino, A., Aho, V. T. E., Pardo Rodriguez, B., Thomas, M. H., Gérardy, J. J., Bastero Acha, I., Halder, R., Cialini, C., Heurtaux, T., Ostahi, I., Busi, S. B., Grandmougin, L., Lowndes, T., Singh, Y., Martens, E. C., Mittelbronn, M., Buttini, M., & Wilmes, P. (2023). Fiber deprivation and microbiome-borne curli shift gut bacterial populations and accelerate disease in a mouse model of Parkinson's disease. *Cell Reports*, *42*(9). https://doi.org/10.1016/J.CELREP.2023.113071

Schwiertz, A., Spiegel, J., Dillmann, U., Grundmann, D., Bürmann, J., Faßbender, K., Schäfer, K. H., & Unger, M. M. (2018). Fecal markers of intestinal inflammation and intestinal permeability are elevated in Parkinson's disease. *Parkinsonism & Related Disorders*, *50*, 104–107. https://doi.org/10.1016/J.PARKRELDIS.2018.02.022

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, *30*(14), 2068–2069. https://doi.org/10.1093/BIOINFORMATICS/BTU153

Seidler, A., Hellenbrand, W., Robra, B. P., Vieregge, P., Nischan, P., Joerg, J., Oertel, W. H., Ulm, G., & Schneider, E. (1996). Possible environmental, occupational, and other etiologic factors for Parkinson's disease: a case-control study in Germany. *Neurology*, *46*(5), 1275–1284. https://doi.org/10.1212/WNL.46.5.1275

Sfondrini, L., Rossini, A., Besusso, D., Merlo, A., Tagliabue, E., Mènard, S., & Balsari, A. (2006). Antitumor activity of the TLR-5 ligand flagellin in mouse models of cancer. *Journal of Immunology (Baltimore, Md.: 1950)*, *176*(11), 6624–6630. https://doi.org/10.4049/JIMMUNOL.176.11.6624

Shah, P., Fritz, J. V., Glaab, E., Desai, M. S., Greenhalgh, K., Frachet, A., Niegowska, M., Estes, M., Jäger, C., Seguin-Devaux, C., Zenhausern, F., & Wilmes, P. (2016). A microfluidics-based in vitro model of the gastrointestinal human–microbe interface. *Nature Communications 2016 7:1*, *7*(1), 1–15. https://doi.org/10.1038/ncomms11535

Sharma, A., Yan, Y., Karunaratne, T., Herekar, A. A., Kurek, J., Morgan, J., & Rao, S. S. (2020). S0500 Intestinal Methanogen Overgrowth (IMO) in Parkinson's Disease: High Prevalence and Correlation With Constipation. *American Journal of Gastroenterology*, *115*(1), S249–S250. https://doi.org/10.14309/01.AJG.0000704048.99013.18

Shen, Z., Luo, W., Tan, B., Nie, K., Deng, M., Wu, S., Xiao, M., Wu, X., Meng, X., Tong, T., Zhang, C., Ma, K., Liao, Y., Xu, J., & Wang, X. (2022). Roseburia intestinalis stimulates TLR5-dependent intestinal immunity against Crohn's disease. *EBioMedicine*, *85*. https://doi.org/10.1016/J.EBIOM.2022.104285

Shirlee Tan, B. S. P., David Schubert, B. S. P., & Pamela Maher, B. S. P. (2001). Oxytosis: A novel form of programmed cell death. *Current Topics in Medicinal Chemistry*, *1*(6), 497–506. https://doi.org/10.2174/1568026013394741

Shrestha, N., Abe, R. A. M., Masroor, A., Khorochkov, A., Prieto, J., Singh, K. B., Nnadozie, M. C., Abdal, M., & Mohammed, L. (2021). The Correlation Between Parkinson's Disease and Rapid Eye Movement Sleep Behavior Disorder: A Systematic Review. *Cureus*, *13*(8). https://doi.org/10.7759/CUREUS.17026

Silva, Y. P., Bernardi, A., & Frozza, R. L. (2020). The Role of Short-Chain Fatty Acids From Gut Microbiota in Gut-Brain Communication. *Frontiers in Endocrinology*, *11*. https://doi.org/10.3389/FENDO.2020.00025

Song, J., & Singh, M. (2009). How and when should interactome-derived clusters be used to predict functional modules and protein function? *Bioinformatics*, *25*(23), 3143–3150. https://doi.org/10.1093/BIOINFORMATICS/BTP551

Spillantini, M. G., & Goedert, M. (2018). Neurodegeneration and the ordered assembly of α-synuclein. *Cell and Tissue Research*, *373*(1), 137–148. https://doi.org/10.1007/S00441-017-2706-9

Strandwitz, P. (2018). Neurotransmitter modulation by the gut microbiota. *Brain Research*, *1693*(Pt B), 128. https://doi.org/10.1016/J.BRAINRES.2018.03.015

Sveinbjornsdottir, S. (2016). The clinical symptoms of Parkinson's disease. *Journal of Neurochemistry*, *139 Suppl 1*, 318–324. https://doi.org/10.1111/JNC.13691

Tan, A. H., Chong, C. W., Lim, S. Y., Yap, I. K. S., Teh, C. S. J., Loke, M. F., Song, S. L., Tan, J. Y., Ang, B. H., Tan, Y. Q., Kho, M. T., Bowman, J., Mahadeva, S., Yong, H. Sen, & Lang, A. E. (2021). Gut Microbial Ecosystem in Parkinson Disease: New Clinicobiological Insights from Multi-Omics. *Annals of Neurology*, *89*(3), 546–559. https://doi.org/10.1002/ANA.25982

Tanei, Z. ichi, Saito, Y., Ito, S., Matsubara, T., Motoda, A., Yamazaki, M., Sakashita, Y., Kawakami, I., Ikemura, M., Tanaka, S., Sengoku, R., Arai, T., & Murayama, S. (2021). Lewy pathology of the esophagus correlates with the progression of Lewy body disease: a Japanese cohort study of autopsy cases. *Acta Neuropathologica*, *141*(1), 25. https://doi.org/10.1007/S00401-020-02233-8

Tran, H. Q., Ley, R. E., Gewirtz, A. T., & Chassaing, B. (2019). Flagellin-elicited adaptive immunity suppresses flagellated microbiota and vaccinates against chronic inflammatory diseases. *Nature Communications 2019 10:1*, *10*(1), 1–15. https://doi.org/10.1038/s41467-019-13538-y

Tredici, K. Del, & Braak, H. (2013). *Idiopathic Parkinson's Disease: Staging an α-Synucleinopathy with a Predictable Pathoanatomy*. https://www.ncbi.nlm.nih.gov/books/NBK6077/

Trischler, R., Poehlein, A., Daniel, R., & Müller, V. (2023). Ethanologenesis from glycerol by the gut acetogen Blautia schinkii. *Environmental Microbiology*, *25*(12), 3577–3591. https://doi.org/10.1111/1462-2920.16517

Tsalenchuk, M., Gentleman, S. M., & Marzi, S. J. (2023). Linking environmental risk factors with epigenetic mechanisms in Parkinson's disease. *Npj Parkinson's Disease 2023 9:1*, *9*(1), 1–12. https://doi.org/10.1038/s41531-023-00568-z

Tsurugizawa, T., Kondoh, T., & Torii, K. (2008). Forebrain activation induced by postoral nutritive substances in rats. *NeuroReport*, *19*(11), 1111–1115. https://doi.org/10.1097/WNR.0B013E328307C414

Van Den Bossche, T., Arntzen, M., Becher, D., Benndorf, D., Eijsink, V. G. H., Henry, C., Jagtap, P. D., Jehmlich, N., Juste, C., Kunath, B. J., Mesuere, B., Muth, T., Pope, P. B., Seifert, J., Tanca, A., Uzzau, S., Wilmes, P., Hettich, R. L., & Armengaud, J. (2021).

The Metaproteomics Initiative: a coordinated approach for propelling the functional characterization of microbiomes. *Microbiome*, *9*(1), 1–4. https://doi.org/10.1186/S40168-021-01176-W/METRICS

Vance, J. E. (2018). Historical perspective: phosphatidylserine and phosphatidylethanolamine from the 1800s to the present. *Journal of Lipid Research*, *59*(6), 923–944. https://doi.org/10.1194/JLR.R084004

Vascellari, S., Palmas, V., Melis, M., Pisanu, S., Cusano, R., Uva, P., Perra, D., Madau, V., Sarchioto, M., Oppo, V., Simola, N., Morelli, M., Santoru, M. L., Atzori, L., Melis, M., Cossu, G., & Manzin, A. (2020). Gut Microbiota and Metabolome Alterations Associated with Parkinson's Disease. *MSystems*, *5*(5). https://doi.org/10.1128/MSYSTEMS.00561-20

Vazquez, A. (2010). Protein Interaction Networks. *Neuroproteomics*, 135–145. https://doi.org/10.1201/b17409-14

Vijay-Kumar, M., Aitken, J. D., Sanders, C. J., Frias, A., Sloane, V. M., Xu, J., Neish, A. S., Rojas, M., & Gewirtz, A. T. (2008). Flagellin treatment protects against chemicals, bacteria, viruses, and radiation. *Journal of Immunology (Baltimore, Md.: 1950)*, *180*(12), 8280–8285. https://doi.org/10.4049/JIMMUNOL.180.12.8280

Wahlström, A., Sayin, S. I., Marschall, H. U., & Bäckhed, F. (2016). Intestinal Crosstalk between Bile Acids and Microbiota and Its Impact on Host Metabolism. *Cell Metabolism*, *24*(1), 41–50. https://doi.org/10.1016/J.CMET.2016.05.005

Wakabayashi, K., Takahashi, H., Ohama, E., & Ikuta, F. (1990). Parkinson's disease: an immunohistochemical study of Lewy body-containing neurons in the enteric nervous system. *Acta Neuropathologica*, *79*(6), 581–583. https://doi.org/10.1007/BF00294234

Wall, M. E., Hlavacek, W. S., & Savageau, M. A. (2004). Design of gene circuits: lessons from bacteria. *Nature Reviews Genetics 2004 5:1*, *5*(1), 34–42. https://doi.org/10.1038/NRG1244

Wallen, Z. D., Demirkan, A., Twa, G., Cohen, G., Dean, M. N., Standaert, D. G., Sampson, T. R., & Payami, H. (2022a). Metagenomics of Parkinson's disease implicates the gut microbiome in multiple disease mechanisms. *Nature Communications*, *13*(1). https://doi.org/10.1038/S41467-022-34667-X

Wallen, Z. D., Demirkan, A., Twa, G., Cohen, G., Dean, M. N., Standaert, D. G., Sampson, T. R., & Payami, H. (2022b). Metagenomics of Parkinson's disease implicates the gut microbiome in multiple disease mechanisms. *Nature Communications 2022 13:1*, *13*(1), 1–20. https://doi.org/10.1038/s41467-022-34667-x

Wan, Q., Tang, J., Han, Y., & Wang, D. (2018). Co-expression modules construction by WGCNA and identify potential prognostic markers of uveal melanoma. *Experimental Eye Research*, *166*, 13–20. https://doi.org/10.1016/J.EXER.2017.10.007

Wang, B., Yao, M., Lv, L., Ling, Z., & Li, L. (2017). The Human Microbiota in Health and Disease. *Engineering*, *3*(1), 71–82. https://doi.org/10.1016/J.ENG.2017.01.008

Wang, C., Lau, C. Y., Ma, F., & Zheng, C. (n.d.). *Genome-wide screen identifies curli amyloid fibril as a bacterial component promoting host neurodegeneration*. https://doi.org/10.1073/pnas.2106504118/-/DCSupplemental

Wang, J., Wang, F., Mai, D., & Qu, S. (2020). Molecular Mechanisms of Glutamate Toxicity in Parkinson's Disease. *Frontiers in Neuroscience*, *14*. https://doi.org/10.3389/FNINS.2020.585584

Wang, L., Li, F., Gu, B., Qu, P., Liu, Q., Wang, J., Tang, J., Cai, S., Zhao, Q., & Ming, Z. (2022). Metaomics in Clinical Laboratory: Potential Driving Force for Innovative Disease Diagnosis. *Frontiers in Microbiology*, *13*, 883734. https://doi.org/10.3389/FMICB.2022.883734/BIBTEX

Warnecke, T., Schäfer, K. H., Claus, I., Del Tredici, K., & Jost, W. H. (2022). Gastrointestinal involvement in Parkinson's disease: pathophysiology, diagnosis, and management. *Npj Parkinson's Disease 2022 8:1*, *8*(1), 1–13. https://doi.org/10.1038/s41531-022-00295-x

Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. https://doi.org/10.1017/CBO9780511815478

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature 1998 393:6684*, *393*(6684), 440–442. https://doi.org/10.1038/30918

Weingarten, C. P., Sundman, M. H., Hickey, P., & Chen, N. kuei. (2015). Neuroimaging of Parkinson's disease: Expanding views. *Neuroscience & Biobehavioral Reviews*, *59*, 16–52. https://doi.org/10.1016/J.NEUBIOREV.2015.09.007

Wetherington, J., Serrano, G., & Dingledine, R. (2008). Astrocytes in the Epileptic Brain. *Neuron*, *58*(2), 168–178. https://doi.org/10.1016/J.NEURON.2008.04.002

Wilmes, P., & Bond, P. L. (2004). The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms. *Environmental Microbiology*, *6*(9), 911–920. https://doi.org/10.1111/J.1462-2920.2004.00687.X

Wilmes, P., Heintz-Buschart, A., & Bond, P. L. (2015). A decade of metaproteomics: Where we stand and what the future holds. *Proteomics*, *15*(20), 3409. https://doi.org/10.1002/PMIC.201500183

Wilmes, P., Martin-Gallausiaux, C., Ostaszewski, M., Aho, V. T. E., Novikova, P. V., Laczny, C. C., & Schneider, J. G. (2022). The gut microbiome molecular complex in human health and disease. *Cell Host and Microbe*, *30*(9), 1201–1206. https://doi.org/10.1016/j.chom.2022.08.016

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology*, *20*(1), 1–13. https://doi.org/10.1186/S13059-019-1891-0/FIGURES/2

Woodburn, S. C., Bollinger, J. L., & Wohleb, E. S. (2021). The semantics of microglia activation: neuroinflammation, homeostasis, and stress. *Journal of Neuroinflammation 2021 18:1*, *18*(1), 1–16. https://doi.org/10.1186/S12974-021-02309-6

Worby, C. J., Schreiber, H. L., Straub, T. J., van Dijk, L. R., Bronson, R. A., Olson, B. S., Pinkner, J. S., Obernuefemann, C. L. P., Muñoz, V. L., Paharik, A. E., Azimzadeh, P. N., Walker, B. J., Desjardins, C. A., Chou, W. C., Bergeron, K., Chapman, S. B., Klim, A., Manson, A. L., Hannan, T. J., … Earl, A. M. (2022). Longitudinal multi-omics analyses link gut microbiome dysbiosis with recurrent urinary tract infections in women. *Nature Microbiology 2022 7:5*, *7*(5), 630–639. https://doi.org/10.1038/s41564-022-01107-x

Wu, J. X., Pascovici, D., Wu, Y., Walker, A. K., & Mirzaei, M. (2023). Application of WGCNA and PloGO2 in the Analysis of Complex Proteomic Data. *Methods in Molecular Biology (Clifton, N.J.)*, *2426*, 375–390. https://doi.org/10.1007/978-1-0716-1967-4_17

Wu, S., Li, C., Huang, W., Li, W., & Li, R. W. (2012). Alternative Splicing Regulated by Butyrate in Bovine Epithelial Cells. *PLoS ONE*, *7*(6), 39182. https://doi.org/10.1371/JOURNAL.PONE.0039182

Wu, X., Pan, S., Luo, W., Shen, Z., Meng, X., Xiao, M., Tan, B., Nie, K., Tong, T., & Wang, X. (2020). Roseburia intestinalis-derived flagellin ameliorates colitis by targeting miR-223-3p-mediated activation of NLRP3 inflammasome and pyroptosis. *Molecular Medicine Reports*, *22*(4), 2695–2704. https://doi.org/10.3892/MMR.2020.11351

Wuyts, S., Alves, R., Zimmermann-Kogadeeva, M., Nishijima, S., Blasche, S., Driessen, M., Geyer, P. E., Hercog, R., Kartal, E., Maier, L., Müller, J. B., Garcia Santamarina, S., Schmidt, T. S. B., Sevin, D. C., Telzerow, A., Treit, P. V, Wenzel, T., Typas, A., Patil, K. R., … Bork, P. (2023). Consistency across multi-omics layers in a drug-perturbed gut microbial community. *Molecular Systems Biology*, *19*(9), 11525. https://doi.org/10.15252/MSB.202311525/SUPPL_FILE/MSB202311525-SUP-0002-EVFIGS.PDF

Xie, A., Ensink, E., Li, P., Gordevičius, J., Marshall, L. L., George, S., Pospisilik, J. A., Aho, V. T. E., Houser, M. C., Pereira, P. A. B., Rudi, K., Paulin, L., Tansey, M. G., Auvinen, P., Brundin, P., Brundin, L., Labrie, V., & Scheperjans, F. (2022). Bacterial Butyrate in Parkinson's Disease Is Linked to Epigenetic Changes and Depressive Symptoms. *Movement Disorders*, *37*(8), 1644–1653. https://doi.org/10.1002/MDS.29128

Xu, H., & Yang, F. (2022). The interplay of dopamine metabolism abnormalities and mitochondrial defects in the pathogenesis of schizophrenia. *Translational Psychiatry 2022 12:1*, *12*(1), 1–13. https://doi.org/10.1038/s41398-022-02233-0

Yamaguchi, S. (1991). Basic properties of umami and effects on humans. *Physiology & Behavior*, *49*(5), 833–841. https://doi.org/10.1016/0031-9384(91)90192-Q

Yuan, X., Tian, Y., Liu, C., & Zhang, Z. (2022). Environmental factors in Parkinson's disease: New insights into the molecular mechanisms. *Toxicology Letters*, *356*, 1–10. https://doi.org/10.1016/J.TOXLET.2021.12.003

Zacharias, H. U., Kaleta, C., Cossais, F., Schaeffer, E., Berndt, H., Best, L., Dost, T., Glüsing, S., Groussin, M., Poyet, M., Heinzel, S., Bang, C., Siebert, L., Demetrowitsch, T., Leypoldt, F., Adelung, R., Bartsch, T., Bosy-Westphal, A., Schwarz, K., & Berg, D. (2022). Microbiome and Metabolome Insights into the Role of the Gastrointestinal–

Brain Axis in Parkinson's and Alzheimer's Disease: Unveiling Potential Therapeutic Targets. *Metabolites*, *12*(12). https://doi.org/10.3390/METABO12121222

Zhang, B., & Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology*, *4*(1). https://doi.org/10.2202/1544-6115.1128/MACHINEREADABLECITATION/RIS

Zhang, F., Yue, L., Fang, X., Wang, G., Li, C., Sun, X., Jia, X., Yang, J., Song, J., Zhang, Y., Guo, C., Ma, G., Sang, M., Chen, F., & Wang, P. (2020). Altered gut microbiota in Parkinson's disease patients/healthy spouses and its association with clinical features. *Parkinsonism & Related Disorders*, *81*, 84–88. https://doi.org/10.1016/J.PARKRELDIS.2020.10.034

Zhang, K., Paul, K. C., Jacobs, J. P., Chou, H. C. (Lori), Duarte Folle, A., Del Rosario, I., Yu, Y., Bronstein, J. M., Keener, A. M., & Ritz, B. (2022). Parkinson's Disease and the Gut Microbiome in Rural California. *Journal of Parkinson's Disease*, *12*(8), 2441–2452. https://doi.org/10.3233/JPD-223500

Zhang, L., Chen, F. X., Zeng, Z., Xu, M., Sun, F., Yang, L., Bi, X., Lin, Y., Gao, Y. J., Hao, H. X., Yi, W., Li, M., & Xie, Y. (2021). Advances in Metagenomics and Its Application in Environmental Microorganisms. *Frontiers in Microbiology*, *12*, 766364. https://doi.org/10.3389/FMICB.2021.766364/BIBTEX

Zhang, X., Deeke, S. A., Ning, Z., Starr, A. E., Butcher, J., Li, J., Mayne, J., Cheng, K., Liao, B., Li, L., Singleton, R., Mack, D., Stintzi, A., & Figeys, D. (2018). Metaproteomics reveals associations between microbiome and intestinal extracellular vesicle proteins in pediatric inflammatory bowel disease. *Nature Communications 2018 9:1*, *9*(1), 1–14. https://doi.org/10.1038/s41467-018-05357-4

Zhang, X., Li, L., Butcher, J., Stintzi, A., & Figeys, D. (2019). Advancing functional and translational microbiome research using meta-omics approaches. *Microbiome 2019 7:1*, *7*(1), 1–12. https://doi.org/10.1186/S40168-019-0767-6

Zhang, X., Li, L., Mayne, J., Ning, Z., Stintzi, A., & Figeys, D. (2018). Assessing the impact of protein extraction methods for human gut metaproteomics. *Journal of Proteomics*, *180*, 120–127. https://doi.org/10.1016/J.JPROT.2017.07.001

Zhang, X., Tang, B., & Guo, J. (2023). Parkinson's disease and gut microbiota: from clinical to mechanistic and therapeutic studies. *Translational Neurodegeneration 2023 12:1*, *12*(1), 1–30. https://doi.org/10.1186/S40035-023-00392-8

Zhao, Q., Chen, Y., Huang, W., Zhou, H., & Zhang, W. (2023). Drug-microbiota interactions: an emerging priority for precision medicine. *Signal Transduction and Targeted Therapy 2023 8:1*, *8*(1), 1–27. https://doi.org/10.1038/s41392-023-01619-w

Zhao, Y., Walker, D. I., Lill, C. M., Bloem, B. R., Darweesh, S. K. L., Pinto-Pacheco, B., McNeil, B., Miller, G. W., Heath, A. K., Frissen, M., Petrova, D., Sánchez, M. J., Chirlaque, M. D., Guevara, M., Zibetti, M., Panico, S., Middleton, L., Katzke, V., Kaaks, R., … Vermeulen, R. (2023). Lipopolysaccharide-binding protein and future Parkinson's disease risk: a European prospective cohort. *Journal of Neuroinflammation*, *20*(1), 1–8. https://doi.org/10.1186/S12974-023-02846-2/FIGURES/1