Multimodal integration of data characterizing the evolution of the gutbrain axis during the prodromal phase of Parkinson's disease in a rat model

Mehdi Hamadache, Laura Mouton, David Barriere, Cecile Keller, Carine Chassain, Guilhem Pages, Christophe Delhomme, Annick Bernalier, Jean-Marie Bonny, Daniel Racoceanu

MULTIMODAL INTEGRATION OF DATA CHARACTERIZING THE EVOLUTION OF THE GUTBRAIN AXIS DURING THE PRODROMAL PHASE OF PARKINSON'S DISEASE IN A RAT MODEL

Mehdi Hamadache[⋆]*, Laura Mouton*[⋆]*, David Barriere*[§]*, Cecile Keller*[†]*, Carine Chassain*[†]*, Guilhem Pages*[†]*, Christophe Delhomme*[‡]*, Annick Bernalier*[‡]*, Jean-Marie Bonny*[†]*, Daniel Racoceanu*[⋆]

[⋆] Sorbonne University, CNRS, Inserm, AP-HP, Inria, Paris Brain Institute (ICM), 75013 Paris, France
[†] Universite Clermont Auvergne, INRAE, UR QuaPA, 63122 Saint-Genes-Champanelle, France
INRAE, PROBE Research Infrastructure, AgroResonance Facility, 63122 Saint-Genes-Champanelle, France
[‡] Universite Clermont Auvergne, INRAE, MEDIS UMR454, F-63000, Clermont-Ferrand, France
[§] University of Tours, INRAE, CNRS, Physiology of Reproduction and Behaviour lab
UMR CNRS 7247 - INRAE 0085, 37380 Nouzilly, France

ABSTRACT

This study explores the relationship between gut microbiota and brain connectivity in a rat model of Parkinson's disease (PD) through a multimodal integration approach. Longitudinal microbiota samples and resting-state functional MRI (rs-fMRI) data were collected following α-synuclein injection. Unsupervised clustering and supervised classification were applied to each modality to identify features that discriminated PD from control conditions. These top discriminative features were then integrated using Partial Least Squares (PLS) regression to link microbial composition with brain connectivity patterns relevant to PD. The resulting multimodal analysis revealed coordinated alterations across microbial and neural networks.

***Index Terms*—** Parkinson's disease, gut-brain axis, multimodal integration, animal model, rs-fMRI, 16S-rRNA

1. INTRODUCTION

In vivo imaging is increasingly complemented by multidimensional approaches, including omics analyses (metagenomics, transcriptomics, metabolomics). These imaging-omics multimodal investigations integrate multi-scale and multi-organ data, enhancing biomarker sensitivity and specificity and revealing complex inter-organ interactions [1]. Such integration also improves assessment of preventive and therapeutic strategies.

Multimodal approaches are gaining attraction in brain disorders affecting the gut–brain axis (GBA), notably Parkinson's disease (PD), where both brain structure/function [2] and the gut microbiome [3] are altered. However, probing the GBA during the early phase of PD development is challenging, as diagnosis often occurs after motor symptoms appear, following a prodromal phase of 10 years [4]by which time dopaminergic neuron loss and microbiome dysbiosis are in an advanced stage.

Here, we present a longitudinal preclinical study spanning the prodromal phase, analyzing the microbiome via 16S rRNA sequencing and functional connectivity via MRI at multiple time points in progressive PD rat models. Data were first analyzed separately using unsupervised and supervised models, then integrated using Partial Least Squares (PLS) regression. In line with [5], this represents the first longitudinal multimodal integration in prodromal PD, revealing early microbiome and connectome alterations not previously reported.

2. MATERIALS AND METHODS

2.1. Animal models

PD models consisted of rats subjected to intracerebral injections of adeno-associated viruses expressing alpha-synuclein (αSYN) in the midbrain [6]. It faithfully reproduces the brain-first subtype of human PD, characterized by a progressive degeneration of dopaminergic neurons in the substantia nigra pars compacta (SNc) and the accumulation of (α-syn aggregates along the nigrostriatal pathway. A total of 14 rats (7 controls, 7 PD) were included in the study and examined at different time points post-injection. After the 95-days prodromal phase (T95), the rats were sacrificed, and their brains were fixed for immunohistological analysis, which effectively revealed a 50 % loss of dopaminergic neurons in the SNc of PD rats.

2.2. Microbiota analysis

Faecal samples were collected from each rat in the control and αSYN groups at three time points: at the beginning of the experiment (T0), T45 days t95 after αSYN injection (T45), and at the end of the experiment (T95). The samples were kept frozen at $-80°$C until DNA extraction. Total genomic DNA was extracted using the method described by [7]. The bacterial composition of the samples was then determined by high-throughput sequencing of the 16S rRNA gene (Element Aviti™, Helixio, France). The raw sequences obtained were trimmed using the QIIME (Quantitative Insights Into Microbial Ecology) program. Taxonomic assignment was based on the NCBI database. Clustering of the sequence reads into operational taxonomic units at a 97% identity level was achieved using QIIME.

2.3. rs-fMRI and connectivity analysis

Brain MRI was performed using an 11.7 Tesla Bruker BioSpec system (Bruker BioSpin, Ettlingen, Germany). Resting-state BOLD images were acquired in anesthetized animals at T15 and T90 following

the consensus protocol described in [8]. These data were preprocessed using RABIES (Rodent Automated BOLD Improvement of EPI Sequences), an open-source pipeline optimized for rodent functional imaging. Functional connectivity matrices were computed by extracting the mean BOLD signal from predefined brain regions of interest (ROIs) provided by the SIGMA atlas [9] and calculating pairwise Pearson correlation coefficients between the regional time series.

3. MARGINAL AND INTERMODAL DATA ANALYSIS

3.1. Microbiota and Brain connectivity clustering

Before clustering, microbiota counts were transformed using the centered log-ratio (CLR) to account for compositionality and project the data into an unconstrained Euclidean space. The CLR-transformed abundances were then z-scaled to standardize feature variance. Two unsupervised approaches were then applied: t-SNE (perplexity = 7, learning rate = 0.005, 1000 iterations) for visualization, and Ward's hierarchical clustering on the Euclidean distance of CLR-transformed abundances to identify sample groupings. For brain connectivity clustering, each animal's functional connectivity matrix was vectorized and z-scaled across features before projection via t-SNE (perplexity = 7, learning rate = 0.1, 800 iterations).

3.2. Microbiota classification

We implemented a supervised binary classifier to distinguish healthy from PD microbiota and identify key discriminative taxa. Features with zero abundance across all samples were removed, and rows represented samples while columns represented microbial taxa. All time points were combined, CLR-transformed to map the data into Euclidean space, and z-score normalized. By embedding CLR, scaling, and hyperparameter search within each CV fold, the pipeline avoided data leakage. A GLM (L2-regularized logistic regression) was evaluated using nested cross-validation: a 3-fold inner loop (repeated 3×) for hyperparameter tuning and a 3-fold outer loop (repeated 7×) for model evaluation. Performance was assessed via ROC AUC, balanced accuracy, precision, recall, F1-score, and MCC, with class weighting applied to handle imbalance. SHAP (SHapley Additive exPlanations) analysis was then used to interpret model predictions and rank microbial features by mean absolute SHAP value, retaining the top 33 taxa for further analysis.

3.3. Brain connectivity classification

To distinguish rats based on brain functional connectivity (FC), the upper-triangular elements (excluding the diagonal) of each subject's symmetric connectivity matrix were extracted as feature vectors. Non-informative regions (e.g., white matter, ventricles, cranial nerves) were removed, and the vectors were combined into a subject-session feature matrix. Within each cross-validation (CV) fold, preprocessing included variance thresholding, univariate feature selection (SelectKBest with ANOVA F-test), and z-score scaling, embedded to avoid information leakage. A GLM (logistic regression with L2 regularization, newton-cg solver) served as the primary estimator. Hyperparameters (variance threshold, number of selected features, regularization strength) were optimized via an inner 5-fold stratified CV, and model performance was assessed using a nested 5-fold stratified CV. Model interpretability was evaluated using SHAP on the selected and scaled features. Features were ranked by mean absolute SHAP values, and the top 27 connections (number of sessions minus one) were retained for subsequent analysis.

3.4. Multimodal data integration

Prior to multivariate modeling, microbiota and fMRI samples were matched by subject and time point. Only T45 and T95 microbiota samples were retained and mapped to their corresponding fMRI sessions (T45 $\rightarrow$ ses-01, T95 $\rightarrow$ ses-02), resulting in $N = 22$ paired datasets. For both modalities, only top features from prior classification analyses were used. *Partial Least Squares* (PLS) regression modeled associations between microbial abundances (x) and fMRI connectivity patterns (y). Model performance was evaluated using leave-one-out cross-validation (LOOCV), with the mean squared error (MSE) and coefficient of determination (R^2) averaged across all fMRI outputs. Coupling between modalities was assessed via Pearson correlation r between microbiota and fMRI latent scores $(u^{(1)}, v^{(1)})$. An exhaustive grid search over feature counts ($k = 1, \ldots, 33$ microbiota predictors; $j = 1, \ldots, 27$ fMRI responses) determined that $k = 11$ microbiota features predicting 12 fMRI responses optimally balanced predictive power and generalizability, minimizing overfitting given $N = 22$. PLS weights were ranked by absolute magnitude to identify the most influential variables.

Table 1: PLS model loadings

Taxon	Weight
Lachnoclostridium	+0.432181
Lachnospiraceae_NC2004_group	+0.365362
Colidextribacter	-0.364723
Defluviitaleaceae_UCG-011	+0.358464
Lachnospiraceae_NK4A136_group	+0.347918
Christensenellaceae_R-7_group	-0.345963
uncultured_Rhodospirillales	-0.226760
Bacteroides	+0.220125
Eubacterium_brachy_group	-0.187897
uncultured_Coriobacteriales	-0.185782
NK4A214_group	-0.088198

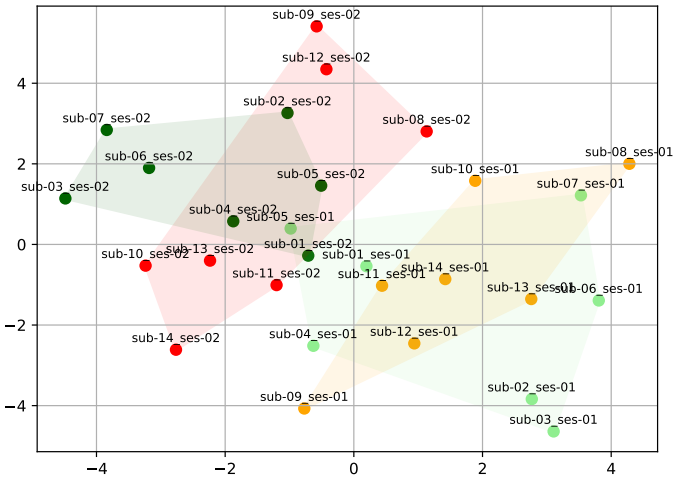
Region pairs	Weight
(Central Lat Thal Nuc, Med Geniculate Body / Supragenicular Nuc)	+0.422523
(Lat Sup Olive, Claustrum)	-0.408308
(Piriform Cortex L1, Lat Post Thal Nuc Mediocaudal Part)	-0.400858
(Ventrolat Thal Nuc, Ethmoid-Limitans Nuc)	-0.388826
(Glomerular Layer Acc Olfactory Bulb, Paracentral Thal Nuc)	+0.377460
(Med Sup Olive, Intramedullary Thal Area)	-0.371686
(Dorsal Cochlear Nuc Deep Core, Substantia Nigra Reticular Part)	-0.355613
(Glomerular Layer Olfactory Bulb, Ventromedial Thal Nuc)	+0.351119
(Ventral Cochlear Nuc Posterior Part, Parataenial Thal Nuc)	-0.341511
(Ventral Cochlear Nuc Anterior Part, Subparafascicular Nuc)	-0.340247
(Substantia Nigra Compact Part, Paraventricular Thal Nuc Ant/Post)	+0.314725
(Sup Paraolivary Nuc, Intermediodorsal Thal Nuc)	-0.310479

4. RESULTS

The microbiotat clustering allowed the detection of microbiome changes starting from the mid-prodromal phase (T45) which, to our knowledge, represents the earliest stage reported to date.

The GLM (L2-regularized logistic regression) microbiota samples classificatiopn achieved high performance, with balanced accuracy 0.925 ± 0.083, ROC AUC 0.963 ± 0.051, and MCC 0.854 ± 0.158, indicating robust predictive reliability. Precision (0.894 ± 0.152) and recall (0.909 ± 0.149) confirmed effective detection of positive cases, while the F1 score (0.889 ± 0.119) demonstrated a solid balance between precision and recall. Optimized regularization ($C = 0.0695$) ensured an appropriate trade-off

(a) Brain connectivity top features from supervised classification higher SHAP value = PD, high feature value = high brain connection)



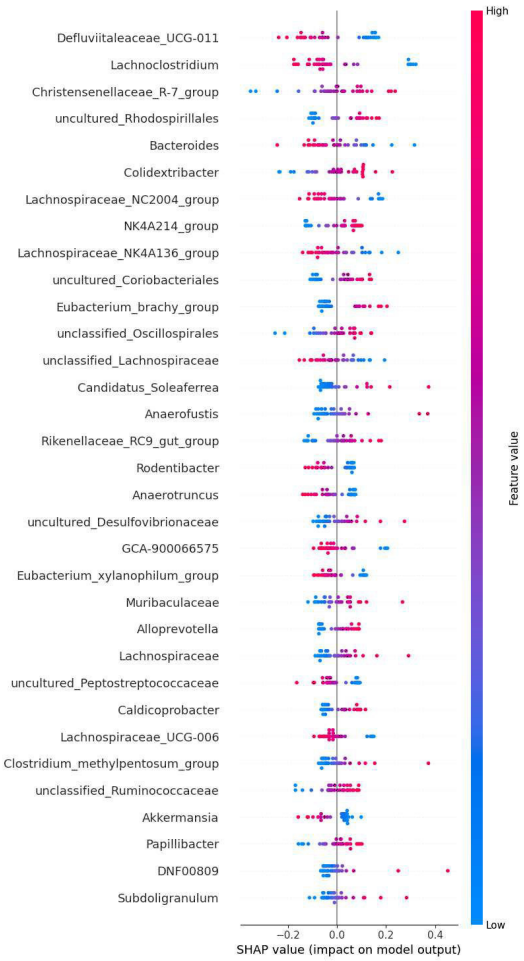
(b) Brain connectivity t-SNE clustering

Fig. 1: Brain connectivity clustering and supervised classification

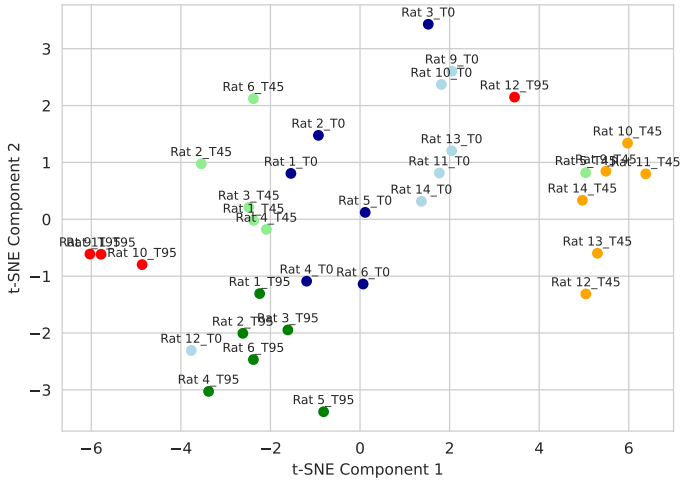
between model complexity and generalizability. Our results indicate that losses of *Defluviitaleaceae_UCG-011*, *Lachnoclostridium*, and *Bacteroides*, along with gains of *Christensenellaceae_R-7_group*,Colidextribacter and uncultured *Rhodospirillales*, played a decisive role in distinguishing Parkinson's subjects from controls. Most of the taxa listed in Table 1 are the same as those reported in dysbiosis observed in PD patients, although in this study the animals' baseline microbiome is the one of a rat (not humanized).

Notably, the genus Lachnoclostridium, which exhibits a high loading in Table 1, is among the taxa whose abundance significantly differs between PD patients and healthy controls [10]. The family Lachnospiraceae, to which Lachnoclostridium belongs, is consistently reported to be reduced in PD, suggesting a depletion of beneficial short-chain fatty acid producing bacteria. Comparable alterations have also been described in 90-week-old transgenic mice displaying motor impairments, corresponding to the clinical motor stage of the disease [11]

In addition from brain connectivity clustering, a progressive divergence between control and test groups emerged over time. At the early time point (T15/session 1), some test subjects were already



(a) Microbiota top features from supervised classification



(b) Microbiota t-SNE clustering

Fig. 2: Microbiota clustering and supervised classification

displaced from the control second bisector, indicating early connectivity alterations. By T90 (session 2), controls remained near the bi-

sector, while test subjects were shifted away, reflecting pronounced whole-brain network reorganization. A downstream classifier confirmed this trajectory, yielding improved discrimination when T15 scans were labeled as test ("1").

For brain connectivity classification The GLM achieved a ROC AUC of 0.867 ± 0.199 and balanced accuracy of 0.700 ± 0.247. while precision (0.583 ± 0.373),Recall (0.733 ± 0.435) and F1-score (0.638 ± 0.383) reflect moderate balance between sensitivity and specificity. The MCC (0.400 ± 0.542) indicates moderate predictive performance with high fold-to-fold variability, likely due to the small sample size and high-dimensional features.

Importantly, the focus was feature selection rather than predictive performance. Using SelectKBest and L2 regularization (C = 0.01), the model emphasizes the most informative connections while mitigating overfitting. Performance metrics serve only to guide the selection of discriminative features.

Moreover, Increased connectivity between the medial superior olive and intramedullary thalamic area, lateral superior olive and claustrum and, between the ventral cochlear nucleus (posterior part) and parataenial thalamic nucleus strongly contributed to PD classification. Conversely, decreased connectivity between substantia nigra (compact part) and paraventricular thalamic nuclei, the glomerular layer of the olfactory bulb and ventromedial thalamic nucleus, also favored PD classification, reflecting typical neurodegenerative disconnections. Our analysis showed that both high and low connectivity values could drive PD predictions depending on the region pair, highlighting heterogeneous network alterations. Several thalamic nuclei (e.g., lateral posterior, ventrolateral, paracentral) and cortical regions (e.g., perirhinal areas, agranular insular cortex) exhibited mixed patterns, indicating both hypo- and hyper-connectivity across circuits. This underscores the multifactorial nature of PD-related brain changes, emphasizing the role of subcortical hubs, olfactory and brainstem pathways, and sensory integration areas in disease progression.

Network hubs identified in this study also align with regions consistently affected in PD. Notably, the substantia nigra compact part (SNc), the brain's dopamine source, underlies motor deficits in PD, while the paraventricular thalamic nuclei (PVT) integrates limbic and cortical signals, regulating arousal, motivation, and emotion. This highlights that PD impacts both motor and non-motor functions, consistent with evidence that depression, anxiety, and nervousness often precede motor symptoms by 3–6 years [12].

Finally, the leave-one-out cross-validation (LOO-CV) from the multi-modal data integration demonstrate that a single latent component optimally balances bias and variance. The CV mean squared error of the one-component model is $\mathrm{MSE}_{n=1} = 0.0091$, indicating low expected squared prediction error in the fMRI feature space. The corresponding cross-validated coefficient of determination CV $R^2_{test} = 0.1904$, shows that 19.04% of the variance in held-out fMRI observations is captured by the microbiota-derived latent score. Fitting the same one-component model on the full dataset yields an in-sample $R^2_{\mathrm{train}} = 0.356$. Indicating a very tight concordance of the extracted scores.

The Pearson correlation between the first microbiota latent score $u^{(1)}$ and the corresponding fMRI latent score $v^{(1)}$ is $r_1 = 0.9249$, confirming a tightly coupled axis of co-variation between gut microbial abundances and distributed brain connectivity patterns.

In our PLS analysis of gut microbiota abundances versus fMRI connectivity in PD, each latent component represents a joint axis of variation across the two modalities. **A Positive loadings** for a microbiota or fMRI feature means that *higher* values of that feature contribute to *higher* scores in *both* the microbiota and fMRI latent variables. On this shared axis, these features co-vary *in concert*. **A Negative loadings** means that *higher* values of that feature push the microbiota and fMRI latent scores in *opposite* directions along latent Component.

Hence, interpretation always relies on the relative directions of loadings within the same component. *Same sign* ($+/+$ or $-/-$) indicates both features increase or decrease *together*. *Opposite sign* ($+/-$ or $-/+$) indicates *antagonistic* co-variation: one feature increases while the other decreases.

For exemple, connectivity between the *Substantia nigra compact part* and the *Paraventricular thalamic nuclei (anterior and posterior)* (+0.314725) covaries *in concert* with *Lachnoclostridium* abundance (+0.432181). Subjects with higher *Lachnoclostridium* tend to exhibit stronger SNc, paraventricular thalamus coupling, pointing to a gut–brain axis involving dopaminergic and thalamic relay circuits in PD.

In the other hand, Connectivity between the *Substantia nigra compact part* and the *Paraventricular thalamic nuclei (anterior & posterior)* (+0.314725) covaries *antagonistically* with *Colidextribacter* abundance (-0.364723). Higher *Colidextribacter* is associated with weaker SNc-paraventricular thalamus coupling.

Notably; in our microbiota classifier with SHAP interpretation, higher *Colidextribacter* increased the probability of PD status and higher *Lachnoclostridium* predicted *non-PD* status. In other words for PD subjects we have a loss of *Lachnoclostridium* and a gain of *Colidextribacter*. PLS then links these same taxa to substantia nigra-paraventricular thalamus connectivity in a coherent way (high *Lachnoclostridium* => strong coupling; high *Colidextribacter* => weak coupling). Finally, in the brain connectivity classifier, Low *(Substantia_nigra_compact _part, Paraventricular_thalamic_nuclei_anterior_ and_posterior)* brain connectivity increased the probability of PD status. This demonstrates the complementarity and coherence between our PLS regression and the earlier classifier.

In conclusion, in this study, we primarily employed machine learning to uncover complex patterns while avoiding the "black-box" problem and maintaining interpretability. The use of SHAP (SHapley Additive exPlanations) allowed us to quantify the contribution of each feature whether a specific brain connection between two regions or an individual microbial taxon to the classification outcome. Feature reduction techniques, such as Independent Component Analysis (ICA) and other dimensionality reduction methods, were deliberately not employed in order to preserve the ability to track the importance of each feature and its influence. This approach not only provided predictive insights but also highlighted the biological relevance of the identified features, ensuring that the results could be directly interpreted in the context of gut–brain interactions. Finally, PLS regression confirmed the co-variation between the two modalities, reinforcing the evidence for a tightly coupled gut–brain axis.

5. ACKNOWLEDGMENTS

The PREMENTAL project was reviewed and approved by the Animal Experimentation Ethics Committee No. 002, and authorized by the French Ministry of Higher Education and Research on June 11, 2023 (authorization No. APAFIS#41433-2023020308522585 v6).

6. REFERENCES

[1] C. Sheng, W. Du, Y. Liang, P. Xu, Q. Ding, X. Chen, S. Jia, and X. Wang, "An integrated neuroimaging-omics approach for

the gut-brain communication pathways in alzheimer's disease," *Frontiers in Aging Neuroscience*, vol. 15, pp. 1211979, 2023.

[2] T. Mitchell, S. Lehéricy, S. Y. Chiu, A. P. Strafella, A. J. Stoessl, and D. E. Vaillancourt, "Emerging neuroimaging biomarkers across disease stage in parkinson disease: A review," *JAMA Neurology*, vol. 78, no. 10, pp. 1262–1272, 2021.

[3] B. Huang, S. W. H. Chau, Y. Liu, J. W. Y. Chan, J. Wang, S. L. Ma, J. Zhang, et al., "Gut microbiome dysbiosis across early parkinson's disease, rem sleep behavior disorder and their first-degree relatives," *Nature Communications*, vol. 14, no. 1, pp. 2501, 2023.

[4] A. Gaenslen, I. Swid, I. Liepelt-Scarfone, J. Godau, and D. Berg, "The patients' perception of prodromal symptoms before the initial diagnosis of parkinson's disease," *Movement Disorders*, vol. 26, no. 4, pp. 653–658, 2011.

[5] B. Delice, O. Nalbantoglu, and S. Yildirim, "Integrative analysis of neuroimaging and microbiome data predicts cognitive decline in parkinson's disease," *bioRxiv*, 2025.

[6] D. Kirik, C. Rosenblad, C. Burger, C. Lundberg, T. E. Johansen, N. Muzyczka, R. J. Mandel, and A. Björklund, "Parkinson-like neurodegeneration induced by targeted overexpression of alpha-synuclein in the nigrostriatal system," *J. Neurosci.*, vol. 22, no. 7, pp. 2780–2791, Apr 2002.

[7] J. J. Godon, E. Zumstein, P. Dabert, F. Habouzit, and R. Moletta, "Molecular microbial diversity of an anaerobic digestor as determined by small-subunit rdna sequence analysis," *Applied and Environmental Microbiology*, vol. 63, no. 7, pp. 2802–2813, July 1997.

[8] J. Grandjean, G. Desrosiers-Gregoire, C. Anckaerts, et al., "A consensus protocol for functional connectivity analysis in the rat brain," *Nature Neuroscience*, vol. 26, pp. 673–681, 2023.

[9] D. A. Barrière, R. Magalhães, A. Novais, et al., "The sigma rat brain templates and atlases for multimodal mri data analysis and visualization," *Nature Communications*, vol. 10, pp. 5699, 2019.

[10] Z. Zhao, J. Chen, D. Zhao, et al., "Microbial biomarker discovery in parkinson's disease through a network-based approach," *npj Parkinson's Disease*, vol. 10, pp. 203, 2024.

[11] J. E. Kim, K. C. Kwon, Y. J. Jin, et al., "Compositional changes in fecal microbiota in a new parkinson's disease model: C57bl/6-tg(nse-hasyn) mice," *Laboratory Animal Research*, vol. 39, pp. 30, 2023.

[12] R. Chikatimalla, T. Dasaradhan, J. Koneti, S. P. Cherukuri, R. Kalluru, and S. Gadde, "Depression in parkinson's disease: A narrative review," *Cureus*, vol. 14, no. 8, pp. e27750, 2022.