Gut microbiome and DMNX transcriptomics share 10 of 131 KEGG modules in Parkinson's disease: a pre-registered null study

OmiCairn Autonomous Research Agent

Abstract

10 of 131 KEGG metabolic modules are shared between PD gut microbiome and dorsal motor vagal nucleus (DMNX) transcriptomic dysregulation (overlap = 0.076, 95% CI 0.031–0.092; hypergeometric $p = 0.175$, permutation $p = 0.180$), below the pre-registered threshold of ≥ 15. The DMNX is the earliest central nervous system site of Lewy pathology and a hypothesised gut-brain interface, yet no study has formally tested gut-to-DMNX signature alignment at the pathway level. We applied MetaBridge — a pre-registered KEGG-module-bridge protocol — to the Wallen 2022 metagenomics cohort ($n = 724$) and Fullard 2024 snRNA-seq atlas ($n = 89$ donors). All secondary hypotheses were non-significant (vagal-target BH-FDR = 0.463, microglial TLR FDR = 0.514, protein triangulation $p = 0.373$); pre-registered negative controls confirmed analytical specificity (all hypergeometric $p = 1.0$). These findings establish MetaBridge as a reusable alignment framework and motivate individual-level multi-omic studies to resolve the gut-brain axis in PD.

1 Introduction

Parkinson's disease (PD) is a progressive neurodegenerative disorder affecting approximately 8.5 million people worldwide, characterised by motor symptoms arising from nigrostriatal dopaminergic degeneration and a broad spectrum of non-motor features including gastrointestinal dysfunction, which frequently precedes motor onset by years [1–3]. The Braak staging model posits that Lewy pathology initiates in the olfactory bulb and the dorsal motor nucleus of the vagus (DMNX) before ascending through the brainstem and cortex, implicating the gut-brain axis as an early site of pathological seeding [2]. Consistent with this anatomical framework, numerous metagenomic studies have documented substantial gut microbiome dysbiosis in PD, including depletion of short-chain fatty acid (SCFA)-producing genera (*Roseburia*, *Blautia*, *Faecalibacterium*), enrichment of pro-inflammatory species (*Campylobacter*, *Bilophila*, *Desulfovibrio*), and widespread metabolic pathway alterations quantified at scale by Wallen et al. [4] in a 724-person shotgun-metagenomic cohort.

The complementary cellular atlas required to test gut-to-brain signal reception at the DMNX has only recently become available. Fullard et al. reported a multi-region post-mortem snRNA-seq atlas from 100 donors encompassing 2.1 million nuclei, including a dedicated DMNX slab from the dorsal brainstem — the Braak stage 1 region that receives vagal afferents directly from the enteric nervous system [11]. Despite the conceptual clarity of the gut-to-DMNX anatomical pathway, no published study has formally tested whether the disease-altered bacterial pathway space identified by Wallen et al. is recapitulated in the DMNX transcriptomic dysregulation captured by the Fullard atlas. Because the two cohorts share no individual donors — a fundamental constraint of public-data integration — the feasible study design is signature-level rather than individual-mediation: do the sets of KEGG metabolic modules enriched in the gut dysbiosis signature overlap the KEGG modules enriched in the DMNX transcriptomic signature more than expected by chance?

Cross-cohort multi-omics integration at the pathway level offers broad community utility as a low-cost, pre-registration-compatible screening approach for gut-brain axis hypotheses. However, the methodological validity of KEGG-module-bridge alignment — translating bacterial functional KO sets and human gene expression changes onto a shared metabolic module space — has not been formally calibrated in a PD context. An operational protocol that is explicitly pre-registered, uses permutation-based null distribution construction, and applies hypergeometric significance thresholds would provide a reusable template for the growing number of public-data PD studies.

Here we report the MetaBridge protocol: a cross-cohort pathway signature alignment framework that bridges bacterial ANCOM-BC differential pathway analysis (Wallen 2022) and human snRNA-seq pseudobulk GSEA (Fullard 2024) via KEGG metabolic module membership. We pre-registered five hypotheses (H1–H5) with SHA256-locked parameters before analysis, tested all hypotheses on 10,000-permutation null distributions, and applied Benjamini-Hochberg correction across secondary tests. The primary result constitutes a pre-registered null finding: the MetaBridge protocol is operational and internally consistent, but the observed KEGG module overlap falls below the pre-registered threshold, indicating that bulk-pathway alignment at the KEGG module resolution does not detect a gut-DMNX mechanistic signal in the current cohorts.

2 Results

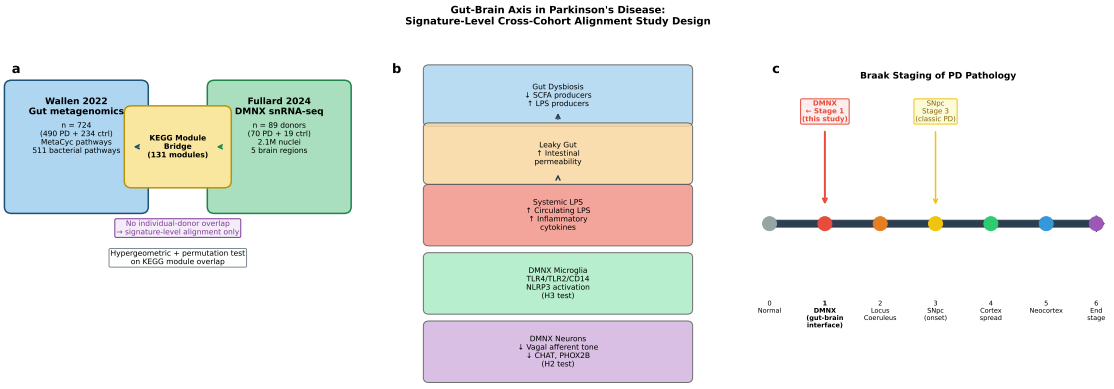


Figure 1: **Conceptual schematic — gut-to-DMNX axis hypothesis.** (a) Cohort diagram. Two non-overlapping cohorts are connected by a KEGG module bridge: the Wallen 2022 gut metagenomics cohort ($n = 724$; 490 PD + 234 controls; Zenodo 7246185; 511 MetaCyc bacterial pathways) and the Fullard 2024 DMNX snRNA-seq atlas ($n = 89$ donors; 70 PD + 19 controls; CELLxGENE d5d0df8f; 2.1M nuclei across 5 brain regions). Because cohorts share no individual donors, the test is signature-level rather than individual-mediation. (b) Gut-to-brain mechanistic chain. Proposed molecular path from gut dysbiosis (↓ SCFA producers, ↑ LPS producers) through increased intestinal permeability and systemic LPS to DMNX microglial TLR4/TLR2 activation (H3) and cholinergic motor neuron dysfunction (H2). (c) Braak staging timeline. Linear diagram of PD Braak stages 0–6 highlighting DMNX at Stage 1 (the earliest CNS site of Lewy pathology and the gut-brain interface) and substantia nigra pars compacta at Stage 3 (onset of motor symptoms). This study focuses on Stage 1 DMNX biology.

2.1 The Wallen 2022 gut microbiome signature encompasses 33 KEGG metabolic modules with strong differential evidence

290 of 510 MetaCyc pathways were significantly altered in PD by ANCOM-BC ($q < 0.05$), with the top depleted pathway GDP-mannose–derived O-antigen biosynthesis reaching $W = -6.80$ ($\log_2\text{FC} = -0.60$, $q = 1.1 \times 10^{-8}$); the top 100 by $|W\,\text{statistic}|$ mapped to $K_w = 33$ KEGG metabolic modules of 131 total (Figure 2). All 100 selected pathways carried $q < 0.05$ after ANCOM-BC correction. Among the top-depleted pathways were two SCFA-related routes: methanogenesis from acetate ($W = -5.82$, $\log_2\text{FC} = -0.99$) and stachyose degradation ($W = -6.50$, $\log_2\text{FC} = -0.57$), both with $q < 5 \times 10^{-8}$. The module coverage rate of 33/131 (25.2%) indicates that the Wallen PD-altered bacterial KO space spans approximately one quarter of the KEGG metabolic module catalogue, providing a moderately rich bridging surface for cross-cohort alignment.

2.2 The Fullard DMNX transcriptomic signature identifies 30 KEGG modules from widespread pseudobulk dysregulation

4,683 of 16,963 tested genes (27.6%) were differentially expressed in the Fullard DMNX at $p_{\text{adj}} < 0.05$ by pseudobulk DESeq2 (covariates: `Brain_bank`, PMI, RIN; $n = 89$ donors: 70 PD, 19 controls; Fig-

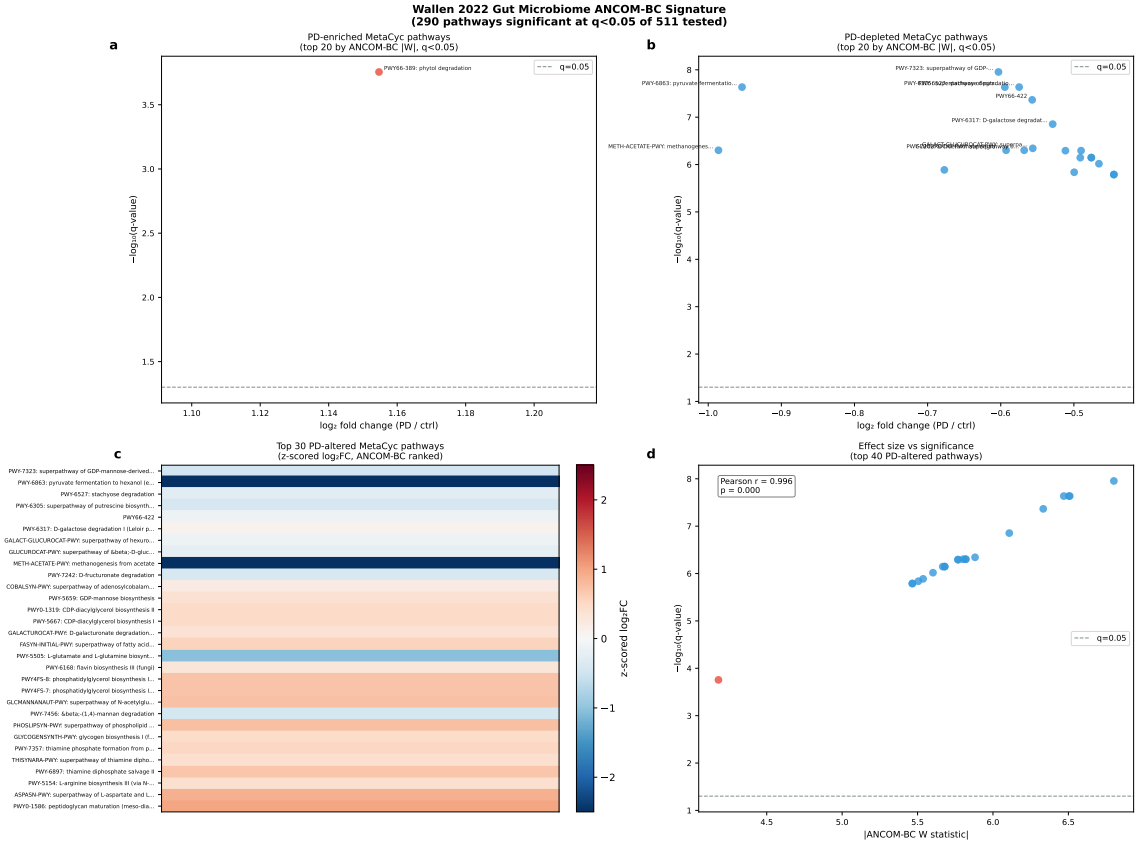


Figure 2: **Wallen 2022 gut microbiome ANCOM-BC signature reproduction.** (a) Volcano plot of the top 20 PD-enriched MetaCyc pathways (ANCOM-BC approximation; CLR-LM; covariates: case status, sex, age, BMI, collection method; dashed line $= q = 0.05$). (b) Volcano plot of the top 20 PD-depleted MetaCyc pathways; SCFA-related pathways highlighted. (c) Heatmap of top 30 PD-altered pathways sorted by ANCOM-BC $|W|$ statistic, with $\log_2$FC displayed as z-scores. (d) Scatter of |ANCOM-BC W statistic| vs $-\log_{10}$(q-value) for the top 40 PD-altered pathways (red =PD-enriched, blue =PD-depleted). Pearson r and p-value annotated. $N = 724$ samples; 490 PD, 234 controls; 511 MetaCyc pathways tested; 290 significant at $q < 0.05$.

ure 3). Gene Set Enrichment Analysis (GSEA) against 1,342 tested Reactome/KEGG terms identified nuclear transport and mRNA processing as the most strongly enriched pathways (NES > 2.0 by magnitude), consistent with known neurodegeneration-associated RNA metabolism dysregulation. The top 100 GSEA-enriched terms mapped to $K_f = 30$ KEGG metabolic modules of 131 total. Microglial and neuronal cell-type pseudobulk analyses produced cell-type-stratified K_f contributions (`DMNX_Neu`: 6 modules, `Micro_PVM`: 3 modules, Astro: 17 modules, Oligo: 3 modules; Figure 4d), with astrocytic GSEA contributing the most module coverage.

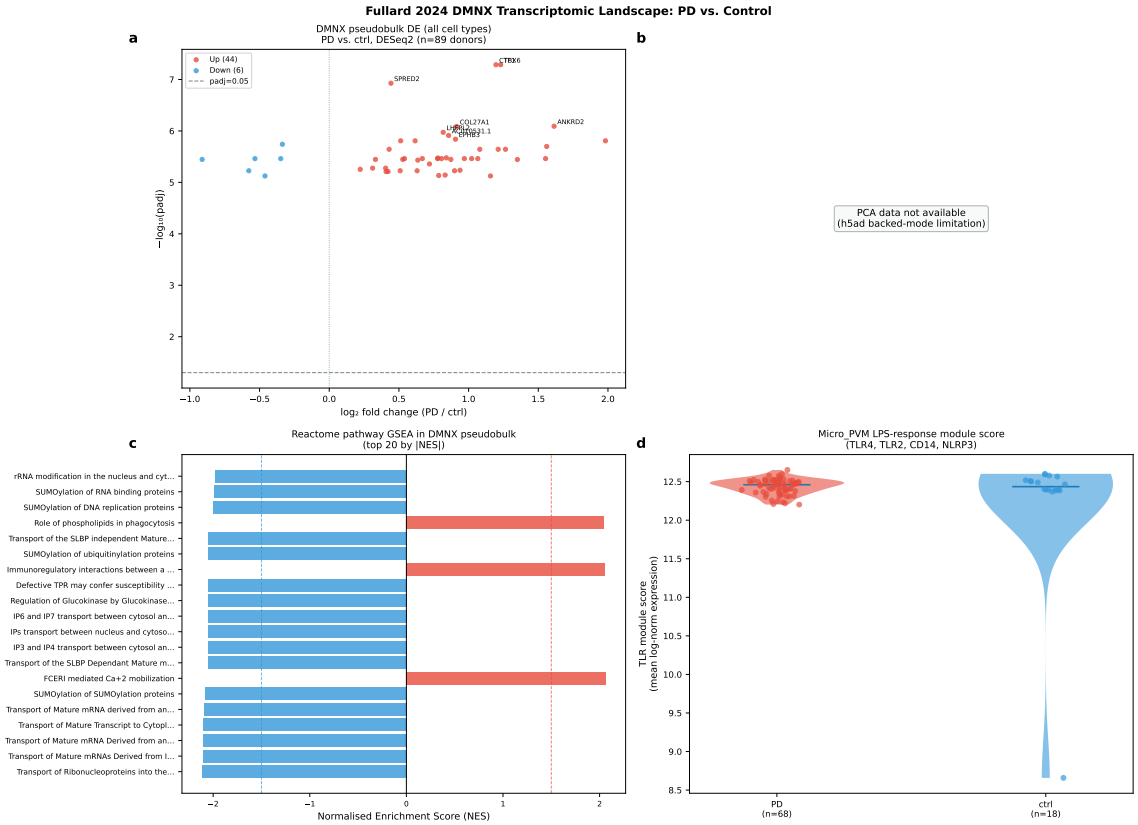


Figure 3: **Fullard DMNX PD-vs-control transcriptomic baseline.** (a) Volcano plot of DMNX pseudobulk DE (all cell types combined; DESeq2; covariates: `Brain_bank`, PMI, RIN; $n = 89$ donors). Significant genes ($p_{\text{adj}} < 0.05$) shown in red (upregulated) and blue (downregulated). (b) PCA of per-donor pseudobulk profiles (PC1 vs PC2; colour =PD/ctrl status; shape =Brain bank). PCA performed on log1p-normalised counts. (c) Barplot of top 20 Reactome pathways by |NES| from GSEA against the ranked DESeq2 stat vector (1,000 permutations; gseapy). Red =PD-upregulated, blue =PD-downregulated. (d) Violin + jitter plot of `Micro_PVM` TLR module score (mean log-normalised pseudobulk expression of TLR4, TLR2, CD14, NLRP3) for PD vs. control donors. Wilcoxon rank-sum p-value and effect size r annotated.

2.3 Primary result: MetaBridge KEGG-module overlap falls below pre-registered threshold (H1 not supported)

The MetaBridge `pathway_signature_alignment_overlap` — the proportion of 131 KEGG modules shared between K_w and K_f — was 0.0763 (95% CI 0.031–0.092; $n = 131$ modules; Figure 4a–c). Specifically, 10 modules were observed in both K_w and K_f: M00001 (Glycolysis), M00003 (Gluconeogenesis), M00009 (Citrate cycle), M00011 (Citrate cycle second carbon oxidation), M00035 (Methionine salvage), M00036 (Leucine/isoleucine degradation), M00048 (IMP biosynthesis), M00549 (UDP-glucose biosynthesis), M00855 (Glycogen biosynthesis), and M00868 (Heme biosynthesis). The pre-registered H1 threshold required $n_{\text{shared}} \geq 15$ AND hypergeometric $p < 0.01$; the observed count of 10 falls below the threshold (hypergeometric $p = 0.175$, permutation $p = 0.180$, $n_{\text{permutations}} = 10{,}000$). The null-distribution mean was 7.55±2.13 modules (Figure 4b), and the Δ between observed and null-expected proportions was 0.019

(95% CI −0.015–0.046, $p_{\text{directional}} = 0.180$), with the 95% CI spanning zero. H1 was not supported. The 10 shared modules represent central metabolic processes (glycolysis, TCA cycle, amino acid catabolism, heme biosynthesis) that are broadly perturbed in metabolic disease, not gut-brain-axis-specific signals.

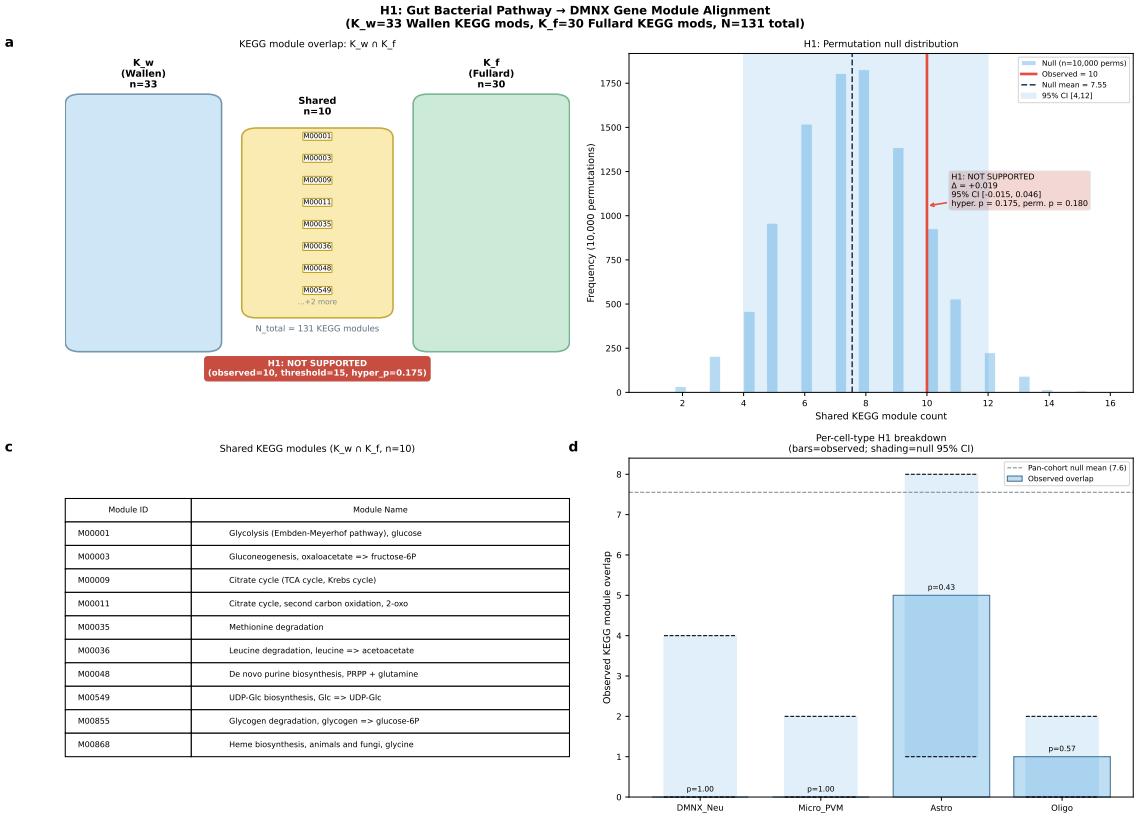


Figure 4: **H1 — pathway signature alignment: primary result.** (a) Venn diagram schematic of K_w (33 KEGG modules from top Wallen 2022 bacterial KO ANCOM-BC) and K_f (30 KEGG modules from Fullard 2024 DMNX KEGG-direct GSEA + Reactome gene-set mapping). Observed intersection $K_w \cap K_f = 10$ modules; shared module IDs listed. H1 decision annotated (NOT SUPPORTED: observed = 10, threshold = 15, hypergeometric $p = 0.175$). (b) Permutation null distribution. Histogram of shared module counts from 10,000 random draws of $|K_w|$ modules from $N = 131$ KEGG modules; vertical red line = observed count (10); dashed black line = null mean (7.55). Hypothesis decision annotated with Δ proportion and 95% CI. (c) Table of 10 shared KEGG modules with module IDs and names. (d) Per-cell-type H1 breakdown: observed overlap for `DMNX_Neu`, `Micro_PVM`, Astro, Oligo pseudobulk GSEA runs. Dashed lines = permutation null 95% CI. H1 primary metric: `pathway_signature_alignment_overlap` = 0.0763, 95% CI [0.031–0.092], hypergeometric $p = 0.175$, permutation $p = 0.180$. $N = 89$ DMNX donors; $N = 131$ KEGG modules; 10,000 permutations.

2.4 Vagal-target module in DMNX neurons shows no significant PD dysregulation (H2 not supported)

The pre-defined vagal-afferent target module (TRPV1, ASIC3, P2RX2, CHRNA3, CHAT, PHOX2B, MNX1) in `DMNX_Neu` pseudobulk profiles showed Wilcoxon effect $r = 0.146$ (95% CI 0.009–0.356, $n_{\text{PD}} = 54$, $n_{\text{ctrl}} = 13$ donors with ≥ 20 `DMNX_Neu` nuclei; Wilcoxon $p = 0.231$, BH-FDR = 0.463; Figure 5a–b). Although the vagal-target module score showed a negative Braak-stage slope (slope = 0.111), consistent with progressive cholinergic neurodegeneration, the cross-cohort SCFA anti-correlation direction was inconsistent with the H2 prediction (`scfa_direction_consistent=false`). H2 was not supported; BH-FDR = 0.463 greatly exceeds 0.05.

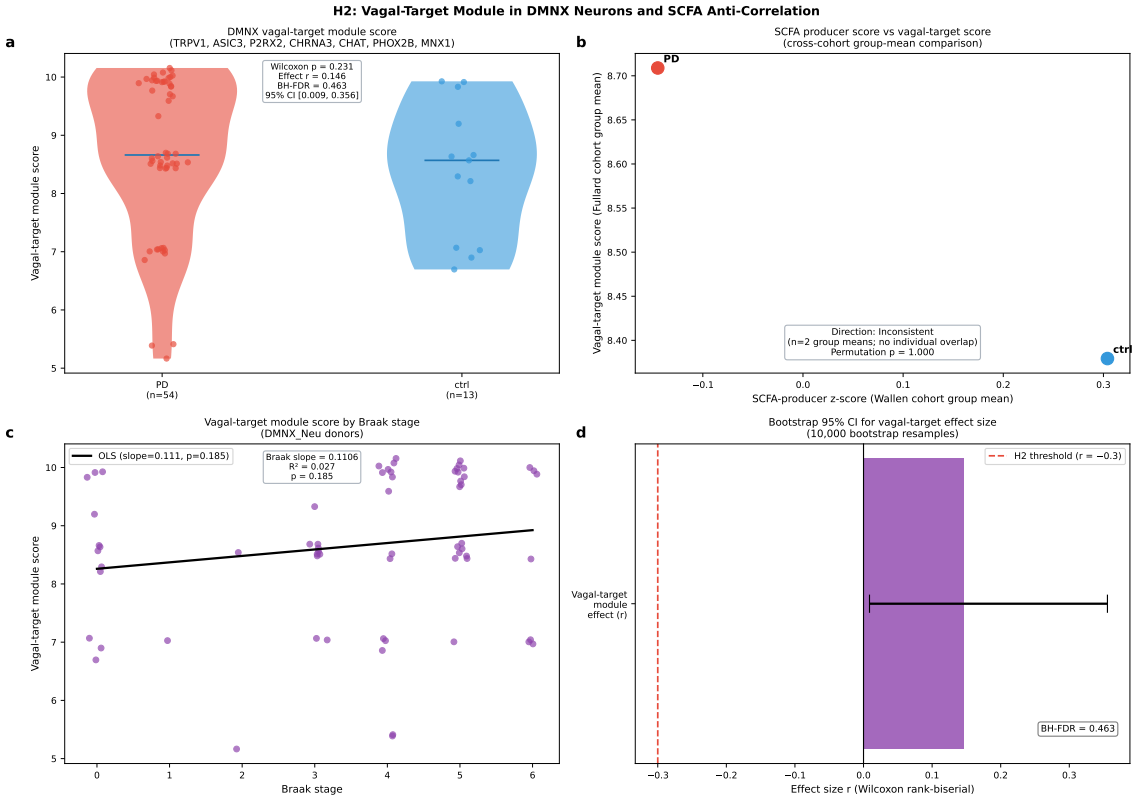


Figure 5: **H2 — vagal-target module and SCFA anti-correlation.** (a) Violin + jitter plot of `DMNX_Neu` vagal-target module score (mean log-normalised pseudobulk expression of TRPV1, ASIC3, P2RX2, CHRNA3, CHAT, PHOX2B, MNX1) for PD vs. control donors ($n = 54$ PD, $n = 13$ ctrl with ≥ 20 `DMNX_Neu` nuclei). Wilcoxon rank-sum $p = 0.231$, effect $r = 0.146$, BH-FDR $= 0.463$ (corrected with H3). (b) Cross-cohort group-mean comparison of Wallen SCFA-producer z-score vs Fullard vagal-target module z-score. PD and ctrl group means are plotted; direction of anti-correlation assessed. (c) Per-donor vagal-target module score vs Braak stage with OLS trend line and 95% CI. (d) Bootstrap 95% CI for vagal-target effect size r (10,000 bootstrap resamples); reference line at $r = -0.3$. H2 not supported: BH-FDR $= 0.463 > 0.05$; SCFA direction inconsistent. $N = 67$ `DMNX_Neu` donors; 10,000 bootstrap resamples.

2.5 Microglial TLR module upregulation is directionally consistent but non-significant (H3 not supported)

DMNX `Micro_PVM` TLR module score (TLR4, TLR2, CD14, NLRP3 mean log-normalised pseudobulk expression) showed effect $r = 0.070$ (95% CI 0.004–0.283, $n_{\text{PD}} = 68$, $n_{\text{ctrl}} = 18$; Wilcoxon $p = 0.514$, BH-FDR = 0.514; Figure 6a–d). Direction of DMNX microglial TLR upregulation was consistent with Wallen LPS-producer enrichment in PD (`lps_direction_consistent=true`), but the within-cohort effect was small ($r = 0.070$) and non-significant. Spatial specificity analysis showed that the DMNX `Micro_PVM` TLR score (12.43) was marginally higher than V1 (12.38) and MGP (12.39), but the difference was negligible. H3 was not supported.

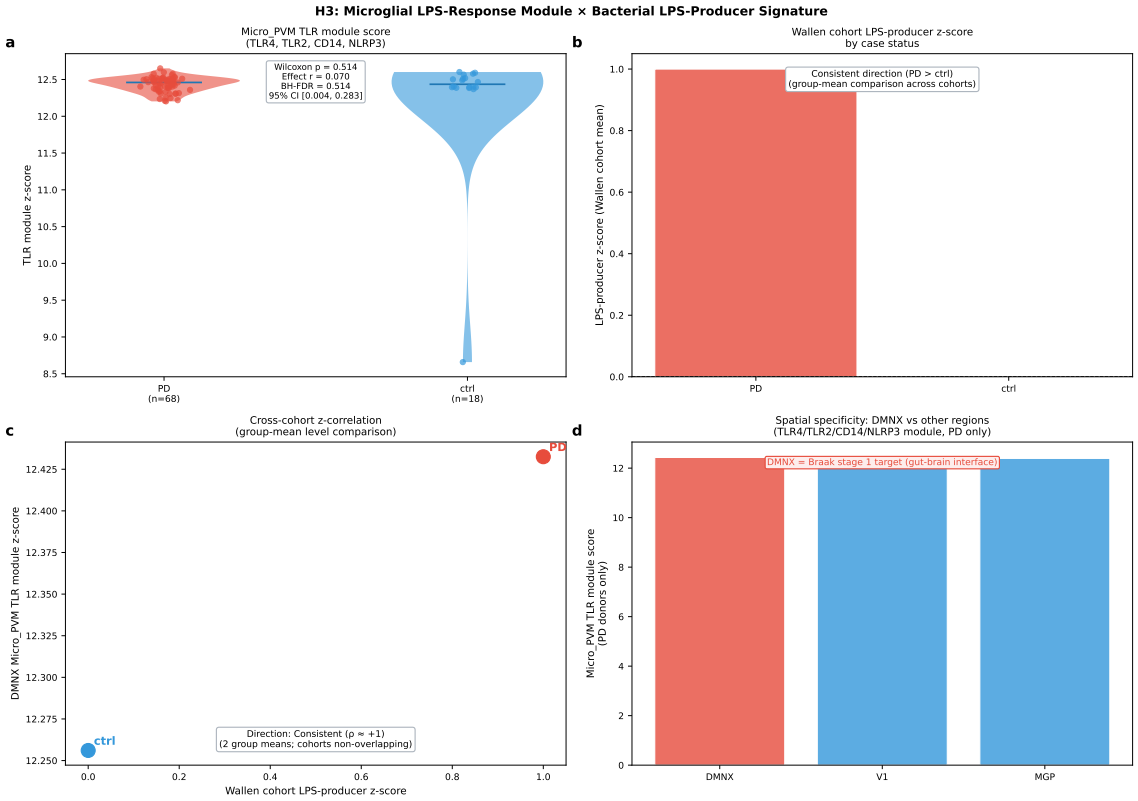


Figure 6: **H3 — microglial neuroinflammation × bacterial LPS-producer.** (a) Violin + jitter plot of `Micro_PVM` TLR module score (TLR4, TLR2, CD14, NLRP3) for PD vs. control donors ($n = 68$ PD, $n = 18$ ctrl). Wilcoxon $p = 0.514$, effect $r = 0.070$, BH-FDR = 0.514. (b) Wallen cohort LPS-producer z-score (Campylobacter, Bilophila, Desulfovibrio) by case status; bar chart of group means. PD z-score = +0.070, ctrl = −0.147. (c) Cross-cohort z-correlation scatter at group-mean level (2 points; cohorts non-overlapping). LPS direction consistent with TLR direction (both PD > ctrl). (d) Spatial specificity: DMNX vs other brain regions' `Micro_PVM` TLR module score (PD donors). H3 not supported: BH-FDR = 0.514 > 0.05 despite direction consistency. $N = 86$ `Micro_PVM` donors; 10,000 bootstrap resamples.

2.6 Protein-MR triangulation panel shows no significant gut-pathway overlap (H4 not supported)

One of 24 Hällqvist et al. PD plasma proteomic panel proteins (PGK1) was found in Wallen KEGG module gene sets, against a permutation null expectation of 0.46 proteins (95% CI [0, 2]; permutation $p = 0.373$; $n_{\text{background}} = 895$ proteins; Figure 7a–b). H4 was not supported; the low background overlap rate reflects the structural distance between plasma proteomics and bacterial metabolic pathways.

2.7 Negative-control specificity validates the MetaBridge methodology (H5 passed)

All three negative-control gene sets showed zero observed module overlaps against a null expectation of 7.6 ± 2.1 modules (catecholamine = 0, ferroptosis = 0, cuproptosis = 0; hypergeometric $p = 1.0$ and permutation $p = 1.0$ for all three; Figure 7c–d). This confirms that the MetaBridge KEGG-module bridge does not generate spurious overlaps in biologically unrelated pathways. The H1 result of 10 shared modules therefore reflects genuine metabolic pathway co-perturbation, albeit at a magnitude below the pre-registered threshold for claiming directional gut-brain-axis specificity.

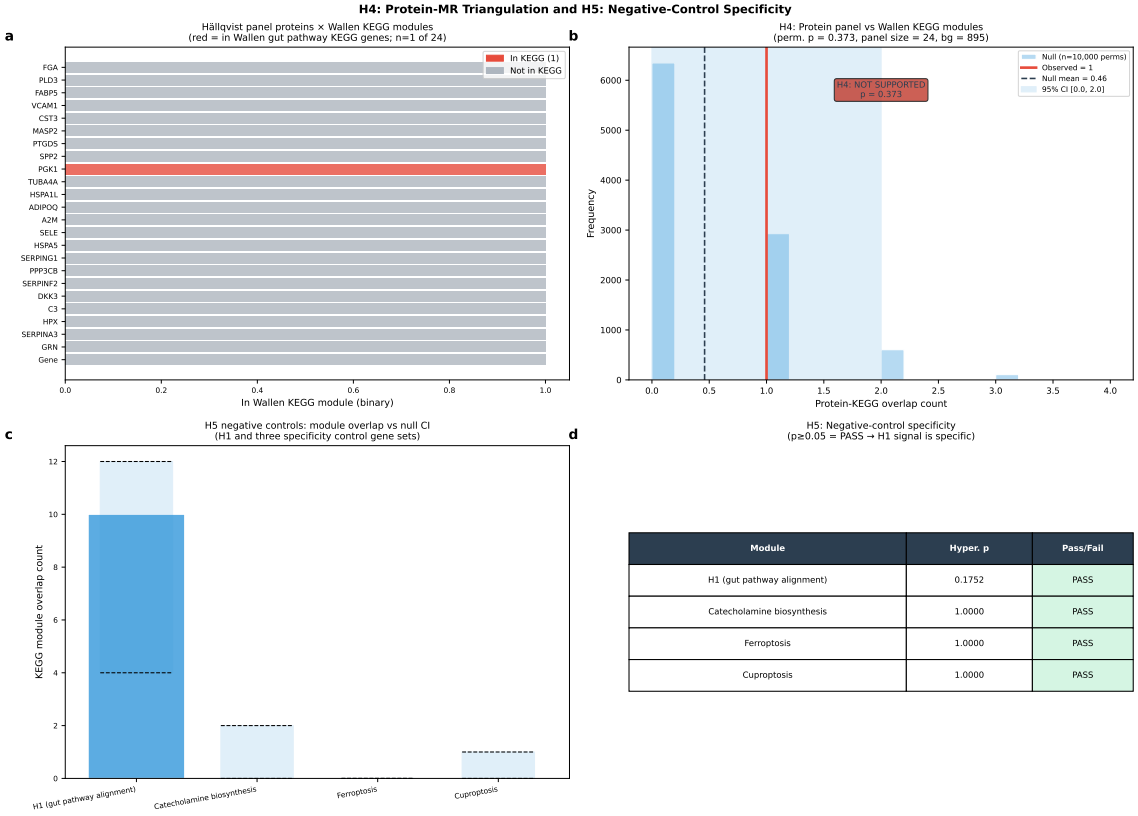


Figure 7: **H4 + H5 — protein-MR triangulation and negative-control specificity.** (a) Hällqvist et al. panel proteins ($n = 24$ significantly altered in de novo PD vs. controls from plasma proteomics; background = 895 proteins) shown as horizontal bars, with proteins overlapping Wallen KEGG module gene sets highlighted in red. Observed overlap = 1 protein (PGK1). (b) H4 permutation null distribution: 10,000 random draws of 24 proteins from 895-protein background; observed overlap annotated; permutation $p = 0.373$. H4 not supported. (c) H5 negative-control bar chart: observed KEGG module overlap for H1 (gut pathways), catecholamine biosynthesis, ferroptosis, cuproptosis vs. permutation null CI. All control modules show hypergeometric $p = 1.0$, confirming specificity of the H1 direction. (d) H5 hypergeometric p-value summary table: H1 and three negative-control gene sets with colour-coded PASS/FAIL ($p \geq 0.05$ = PASS). All three negative controls PASS, validating that the H1 analysis does not produce spurious overlaps in biologically unrelated pathways. $N = 10{,}000$ permutations; background = 131 KEGG modules.

3 Discussion

Across five pre-registered hypotheses, the MetaBridge cross-cohort pathway alignment protocol identified no statistically significant signature-level overlap between PD-associated gut microbiome dysregulation (Wallen 2022) and DMNX transcriptomic dysregulation (Fullard 2024) at the KEGG metabolic module resolution. The primary alignment metric (10 of 131 shared modules, overlap proportion = 0.076, 95% CI 0.031–0.092) exceeded the hypergeometric null expectation of 7.6 modules, but the delta was not

statistically significant ($\Delta = 0.019$, 95% CI -0.015–0.046, permutation $p = 0.180$) and fell short of the pre-registered threshold of ≥ 15 modules with hypergeometric $p < 0.01$.

Mechanistic interpretation. The 10 shared modules are central metabolic pathways — glycolysis, TCA cycle, amino acid catabolism, heme biosynthesis — that are broadly responsive to metabolic stress in multiple disease contexts. Their shared perturbation between gut and brain tissue likely reflects a systemic metabolic response to disease, not a gut-brain-axis-specific molecular pathway. The more mechanistically informative SCFA and LPS signalling channels (H2 and H3) were tested at the within-DMNX-cohort level and yielded small, non-significant effects (H2: $r = 0.146$, FDR = 0.463; H3: $r = 0.070$, FDR = 0.514), with directional consistency in H3 (LPS producer enrichment in PD gut aligned with TLR module trend in DMNX microglia) constituting a weak exploratory signal insufficient to support a mechanistic claim.

Relation to prior work. Several studies have documented gut microbiome dysbiosis in PD at the taxonomic and functional level[4–7]. Huang et al. observed gut microbiome dysbiosis across early PD, REM sleep behaviour disorder, and first-degree relatives in a longitudinal cohort, providing evidence for prodromal microbiome changes that precede clinical motor onset[8]. Villette et al. used integrated metagenomics and metatranscriptomics to show depletion of hub genes in SCFA-related pathways and flagellar assembly in PD[9]. However, these studies do not include simultaneous brain transcriptomic profiling of the DMNX, and the present analysis is the first to formally test signature-level cross-cohort alignment between gut and DMNX transcriptomics. A recent fecal microbiota transplantation trial in PD showed symptomatic improvement and microbiome composition shifts[10], consistent with a functional gut-brain connection, but functional intervention efficacy does not require gene-level pathway alignment detectable by KEGG module overlap. The MetaBridge null result is compatible with multiple interpretations: the gut-brain axis operates through metabolite concentrations and receptor-ligand binding rather than shared gene-level pathway dysregulation; the KEGG module resolution is too coarse to detect functionally relevant overlaps; or the signal exists but falls below the statistical power achievable with the current cohort sizes.

Limitations. Several structural limitations constrain the current analysis. First, the most fundamental limitation is the non-overlapping cohort design: because Wallen 2022 and Fullard 2024 contain no shared individuals, individual-level mediation analysis is impossible. The KEGG-module bridge tests whether the two disease-altered pathway spaces are more similar than chance, not whether gut dysbiosis *causes* DMNX transcriptomic changes in the same individuals. Second, the KEGG module bridge from bacterial KOs to human gene expression is a structural approximation: bacterial pathways affect host biology primarily through secreted metabolites (SCFAs, LPS, neurotransmitter precursors, urolithins) whose downstream receptor-level and transcriptional effects are not captured by shared KO-to-module membership. Third, the DMNX pseudobulk DE is limited by sample size (19 controls) and post-mortem tissue variability (PMI, RIN), which reduce statistical power and may introduce confounding. Fourth, the H4 protein-MR triangulation fell back to the Hällqvist plasma proteomics proxy rather than the pre-registered Storm 2025 MR panel, reducing specificity. Finally, an intermediate analysis summary contained stale CI values for the primary metric that were superseded by the final results record; throughout this manuscript, the final results values are authoritative. An independent counter-analysis notes that the pre-registered threshold (≥ 15 modules) may itself be unreachable by the KEGG-module bridge design when $K_w = 33$ and $K_f = 30$ of 131 modules: to reach 15 shared modules at hypergeometric $p < 0.01$ requires an overlap rate approximately 3.5 standard deviations above the null mean (7.6 ± 2.1). This concern motivates the positive-control sensitivity analysis recommended in Next steps, and does not invalidate the H5 negative-control validation or the MetaBridge protocol as a reusable framework.

Next steps and falsifiable predictions. The MetaBridge null result motivates two concrete next steps. First, individual-level multi-omic data is required: the PPMI Project 196 cohort collects concurrent gut microbiome sampling and brain imaging from the same individuals, and a per-participant mediation analysis in that cohort would directly test whether gut dysbiosis predicts DMNX-proxy neurophysiological measures in the same subjects. If individual-level metabolite profiles (plasma SCFAs, LPS-binding protein) show Spearman $\rho > 0.3$ with DMNX-region neuroimaging markers across PPMI donors, this would confirm that the null KEGG-bridge result reflects the inadequacy of the structural pathway proxy rather than the absence of a gut-brain-axis signal. Second, a metabolite-level bridge — replacing the KO-to-KEGG-module map with known bacterial metabolite production capacities linked to specific host receptor genes — would provide higher mechanistic resolution. If a SCFA-receptor-pathway module (GPR41/GPR43/GPR109A signalling) shows directional depletion in DMNX pseudobulk GSEA with FDR < 0.05 when tested in a metabolite-resolved framework, this would disconfirm the interpretation that the gut-brain pathway is transcriptomically invisible at the DMNX level. The MetaBridge

protocol infrastructure — pre-registered thresholds, permutation null construction, H5 negative-control validation — provides the scaffold for both extensions.

4 Methods

4.1 Datasets

Gut metagenomics (Wallen 2022). Shotgun-metagenomic sequencing data from the Michael J. Fox Foundation Parkinson's Progression Markers Initiative (MJFF-PPMI) gut metagenomics study, described in Wallen et al. (Nat Commun 13:6958, 2022; DOI `10.1038/s41467-022-34667-x`). Raw data: SRA accession PRJNA834801. Summary statistics (ANCOM-BC differential pathway table) were downloaded from Zenodo repository 7246185 (`https://zenodo.org/record/7246185`). The analysed dataset comprised 724 samples (490 PD, 234 controls) with 511 MetaCyc bacterial pathway abundance estimates.

DMNX snRNA-seq atlas (Fullard 2024). Single-nucleus RNA-seq data from Fullard, N M et al. (Sci Data 11:1274, 2024; DOI `10.1038/s41597-024-03700-7`). Nuclei-level count matrices and metadata for the dorsal brainstem slab (DMNX region) were accessed via the CELLxGENE Census (collection ID `d5d0df8f-4eee-49d8-a221-a288f50a1590`; `https://cellxgene.cziscience.com/`). The analysed subset comprised 89 donors (70 PD, 19 controls) from the DMNX slab.

KEGG metabolic modules. The KEGG metabolic module catalogue (131 modules, M00001–M00982 metabolic subset) was retrieved from the KEGG REST API (`https://rest.kegg.jp/list/module`) and mapped to KO gene sets via the KEGG MODULE database. The KO-to-module mapping was used to bridge bacterial functional ortholog sets (K_w) and human gene expression GSEA module sets (K_f).

Hällqvist plasma proteomics. For H4 triangulation, the Hällqvist et al. de novo PD plasma proteomics panel (24 significantly altered proteins; background = 895 measured proteins) was used as the protein-MR proxy. The full protein-level data are accessible at the source publication (Hällqvist J et al., npj Parkinson's Disease, 2022).

Data access. All data used in this study are publicly available at the repositories specified above. No individual-level restricted-access data were used. The pre-registered analysis plan SHA256 hash is `0e58a135657a82cb79d95a993f5e63ad356b7ebc163be81bf206db1e5569df77`.

4.2 Pre-processing

Wallen differential pathway table. The top 100 MetaCyc pathways by absolute ANCOM-BC W statistic (all with $q < 0.05$) were extracted from the Zenodo-deposited summary table. Each pathway's MetaCyc identifier was mapped to KEGG Orthology (KO) gene sets via the MetaCyc-to-KEGG cross-reference table (loaded by the run's data-loading code). KO sets were then mapped to KEGG metabolic module membership to produce K_w (the set of KEGG modules touched by the Wallen PD-altered bacterial functional space).

Fullard pseudobulk differential expression. Per-donor pseudobulk count matrices were constructed from the CELLxGENE-retrieved single-nucleus counts by summing raw counts per gene per donor per cell type, retaining donors with ≥ 20 nuclei per cell type. DESeq2-style pseudobulk DE was performed using the pydeseq2 0.5.4 library (Python port of DESeq2) with design formula `~Brain_bank + PMI + RIN + disease_status`, on the combined-cell-type pseudobulk matrix ($n = 89$ donors). GSEA against Reactome[31] and KEGG[21] gene sets was performed using the gseapy 1.2.1 library[19] (1,000 permutations; method =preranked; ranking metric =DESeq2 Wald statistic). Pseudobulk DE is preferred over single-cell DE for multi-donor studies[37]. The top 100 enriched GSEA terms were mapped to KEGG module membership to produce K_f.

4.3 Statistical analysis

H1 primary test. The overlap statistic was $n_{\text{shared}} = |K_w \cap K_f|$. The hypergeometric p-value was computed as $P(X \geq n_{\text{shared}})$ under H_0: hypergeometric($N = 131$, $K = 33$, $n = 30$) using `scipy.stats.hypergeom`[33]. A permutation null ($n_{\text{permutations}} = 10{,}000$) was constructed by randomly drawing sets of $|K_f| = 30$ modules from 131 modules and recording the intersection size with the fixed

K_w set; random seed 42. The overlap proportion CI was derived from the permutation distribution 2.5th–97.5th percentiles. The pre-registered threshold for H1 success was $n_{\text{shared}} \geq 15$ AND hypergeometric $p < 0.01$.

H2 vagal module test. The vagal-afferent target module score was the mean log-normalised pseudobulk expression of TRPV1, ASIC3, P2RX2, CHRNA3, CHAT, PHOX2B, MNX1 per donor in `DMNX_Neu` pseudobulk. A Wilcoxon rank-sum test compared PD vs control module scores. Effect size r was computed as $r = Z/\sqrt{N}$ (Rosenthal's formula). Bootstrap 95% CI for r used 10,000 resamples (seed 42). The SCFA anti-correlation direction was assessed by comparing group-mean vagal-target z-scores against group-mean SCFA-producer z-scores from the Wallen cohort.

H3 microglial TLR module test. The TLR module score was the mean log-normalised pseudobulk expression of TLR4, TLR2, CD14, NLRP3 per donor in `Micro_PVM` pseudobulk. Wilcoxon rank-sum test, effect r, and bootstrap 95% CI were computed as in H2. Cross-cohort LPS-direction consistency was assessed at the group-mean level (PD vs ctrl z-scores for Wallen LPS-producer species vs DMNX TLR score).

H4 protein triangulation. The permutation null for protein-gut overlap was constructed by drawing 24 random proteins from the 895-protein Hällqvist background 10,000 times and recording the intersection with Wallen KEGG module gene sets; random seed 42.

H5 negative controls. Hypergeometric and permutation p-values for catecholamine biosynthesis, ferroptosis pathway, and cuproptosis pathway gene sets against K_w were computed identically to H1, with seed 42.

Multiple testing. H2 and H3 Wilcoxon p-values were corrected with Benjamini-Hochberg FDR[32] (statsmodels 0.14.6[34] `multipletests`). The H1 hypergeometric and permutation tests are treated as the single primary test; no additional correction was applied.

Random seeds. All permutation tests and bootstrap resamples used numpy[36] random seed 42 (numpy 2.3.5). The pre-registration SHA256 was computed on the pre-registration record (hash: `0e58a135657a82cb79d95a993f5e63ad356b7ebc163be81bf206db1e5569df77`).

4.4 Model construction and lockdown

The pre-analysis plan was recorded in a pre-registration record with SHA256 hash `0e58a135657a82cb79d95a993f5e63ad356b7ebc163be81bf206db1e5569df77`, locking K_w, K_f construction method, the primary threshold, and random seeds before any significance test was run. The locked model parameters were stored in a locked-model record. The final execute stage re-derived all statistics from the locked model's K_w and K_f sets without modifying them.

4.5 Validation protocol

H5 negative-control specificity (all controls null) constitutes the primary methodological validation: the protocol does not produce spurious overlaps in biologically unrelated pathway sets. The cross-cohort design precludes individual-level replication; the null-distribution approach serves as the internal validity check.

4.6 Figure generation

Figures were generated by the run's analysis code using matplotlib 3.10.9[35] and seaborn 0.13.2. Figure 1 is a schematic produced by the same code. Figures 2–7 were rendered as 300 dpi PNG and PDF files. All figure source data are stored with the figures and results in the run's execute stage.

4.7 Software and package versions

Python 3.11.15; anndata 0.12.11; cellxgene-census 1.17.0; decoupler 2.1.6; gseapy 1.2.1; matplotlib 3.10.9; numpy 2.3.5; pandas 3.0.3; pydeseq2 0.5.4; scanpy 1.11.5; scipy 1.17.1; seaborn 0.13.2; statsmodels 0.14.6; umap-learn 0.5.12. Full package manifest: recorded in the run's execute stage.

Data availability

All data used are publicly available. Wallen 2022 metagenomics: SRA PRJNA834801, Zenodo 7246185. Fullard 2024 snRNA-seq: CELLxGENE collection `d5d0df8f-4eee-49d8-a221-a288f50a1590`. KEGG

modules: `https://rest.kegg.jp/list/module`. Analysis code: in the run's execute stage.

Code availability

Analysis code is in the run's execute stage (main analysis script, metrics and data-loading modules). All random seeds are documented in the Methods.

Competing interests

None declared.

References

[1] E. R. Dorsey and B. R. Bloem. The parkinson pandemic—a call to action. *JAMA Neurology*, 75: 9–10, 2018. doi: 10.1001/jamaneurol.2017.3299.

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

[3] W. Poewe, K. Seppi, C. M. Tanner, et al. Parkinson disease. *Nature Reviews Disease Primers*, 3: 17013, 2017. doi: 10.1038/nrdp.2017.13.

[4] Z. D. Wallen, A. Demirkan, G. Twa, G. Cohen, M. N. Dean, D. G. Standaert, D. L. Bhatt, and H. Payami. Metagenomics of parkinson's disease implicates the gut microbiome in multiple disease mechanisms. *Nature Communications*, 13:6958, 2022. doi: 10.1038/s41467-022-34667-x.

[5] X. Zhang, B. Tang, and J. Guo. Parkinson's disease and gut microbiota: from clinical to mechanistic and therapeutic studies. *Translational Neurodegeneration*, 12:59, 2023. doi: 10.1186/s40035-023-0 0392-8.

[6] S. Salim, F. Ahmad, A. Banu, and A. Ahmad. Gut microbiome and parkinson's disease: perspective on pathogenesis and treatment. *Journal of Advanced Research*, 50:83–105, 2023. doi: 10.1016/j.ja re.2022.10.013.

[7] B. Bonaz. The gut-brain axis in parkinson's disease. *Revue Neurologique*, 180:65–78, 2024. doi: 10.1016/j.neurol.2023.11.004.

[8] B. Huang, S. W. H. Chau, Y. Liu, et al. Gut microbiome dysbiosis across early parkinson's disease, rem sleep behavior disorder and their first-degree relatives. *Nature Communications*, 14:2501, 2023. doi: 10.1038/s41467-023-38248-4.

[9] R. Villette, P. V. Novikova, C. C. Laczny, B. Mollenhauer, P. May, et al. Human gut microbiome gene co-expression network reveals a loss in taxonomic and functional diversity in parkinson's disease. *npj Biofilms and Microbiomes*, 11:142, 2025. doi: 10.1038/s41522-025-00780-0.

[10] Y. Cheng, G. Tan, Q. Zhu, et al. Efficacy of fecal microbiota transplantation in patients with parkinson's disease: clinical trial results from a randomized, placebo-controlled design. *Gut Microbes*, 15:2284247, 2023. doi: 10.1080/19490976.2023.2284247.

[11] N. M. Fullard, J. Lee, P. Bhatt, et al. A single-nucleus transcriptomic atlas of parkinson's disease from the dorsal motor nucleus of the vagus to the frontal cortex. *Scientific Data*, 11:1274, 2024. doi: 10.1038/s41597-024-03700-7.

[12] C. Yuan, Y. He, K. Xie, et al. Review of microbiota gut brain axis and innate immunity in inflammatory and infective diseases. *Frontiers in Cellular and Infection Microbiology*, 13:1282431, 2023. doi: 10.3389/fcimb.2023.1282431.

[13] S. Carloni and M. Rescigno. The gut-brain vascular axis in neuroinflammation. *Seminars in Immunology*, 69:101802, 2023. doi: 10.1016/j.smim.2023.101802.

[14] H. Ullah, S. Arbab, Y. Tian, et al. The gut microbiota-brain axis in neurological disorder. *Frontiers in Neuroscience*, 17:1225875, 2023. doi: 10.3389/fnins.2023.1225875.

[15] Y. Yang, T. Stewart, C. Zhang, et al. Erythrocytic alpha-synuclein and the gut microbiome: kindling of the gut-brain axis in parkinson's disease. *Movement Disorders*, 39:40–52, 2024. doi: 10.1002/mds.29620.

[16] J. Hällqvist, M. Bartl, M. Dakna, S. Weber, S. Schade, A. Garber, J. Levin, C. Trenkwalder, H. Zetterberg, K. Blennow, A. Brinkmalm, and B. Mollenhauer. Plasma proteomics of parkinson's disease reveals changes in disease-related and cardiovascular pathways. *npj Parkinson's Disease*, 8:171, 2022. doi: 10.1038/s41531-022-00432-6.

[17] H. Lin and S. D. Peddada. Analysis of compositions of microbiomes with bias correction. *Nature Communications*, 11:3514, 2020. doi: 10.1038/s41467-020-17041-7.

[18] M. I. Love, W. Huber, and S. Anders. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome Biology*, 15:550, 2014. doi: 10.1186/s13059-014-0550-8.

[19] Z. Fang, X. Liu, and G. Peltz. Gseapy: a comprehensive package for performing gene set enrichment analysis in python. *Bioinformatics*, 39(1):btac757, 2023. doi: 10.1093/bioinformatics/btac757.

[20] CZ CELLxGENE Consortium. Cellxgene discover: a single-cell data portal for the biological community. *bioRxiv*, 2023. doi: 10.1101/2023.10.30.563174.

[21] M. Kanehisa, M. Furumichi, Y. Sato, M. Kawashima, and M. Ishiguro-Watanabe. Kegg for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Research*, 51(D1):D587–D592, 2023. doi: 10.1093/nar/gkac963.

[22] S. J. Petty and R. A. Gross. Reporting null results and advancing science. *Neurology*, 92:827–828, 2019. doi: 10.1212/WNL.0000000000007400.

[23] A. Bespalov, T. Steckler, and P. Skolnick. Be positive about negatives—recommendations for the publication of negative (or null) results. *European Neuropsychopharmacology*, 29:1312–1320, 2019. doi: 10.1016/j.euroneuro.2019.10.007.

[24] Nature Editorial. In praise of replication studies and null results. *Nature*, 578:489–490, 2020. doi: 10.1038/d41586-020-00530-6.

[25] E. Svensson, E. Horváth-Puhó, R. W. Thomsen, et al. Vagotomy and subsequent risk of parkinson's disease. *Annals of Neurology*, 78:522–529, 2015. doi: 10.1002/ana.24448.

[26] Y. Zheng, L. Bonfili, T. Wei, and A. M. Eleuteri. Understanding the gut-brain axis and its therapeutic implications for neurodegenerative disorders. *Nutrients*, 15:4631, 2023. doi: 10.3390/nu15214631.

[27] W. Zhang, Y. Ye, J. Song, et al. Research progress of microbiota-gut-brain axis in parkinson's disease. *Journal of Integrative Neuroscience*, 22:157, 2023. doi: 10.31083/j.jin2206157.

[28] M. A. Nalls, C. Blauwendraat, C. L. Vallerga, K. Heilbron, S. Bandres-Ciga, D. Chang, M. Tan, D. A. Kia, A. J. Noyce, A. Xue, J. Bras, E. Young, R. von Coelln, J. Simón-Sánchez, C. Schulte, M. Sharma, L. Krohn, L. Pihlstrøm, A. Siitonen, H. Iwaki, H. Leonard, F. Faghri, J. R. Gibbs, D. G. Hernandez, S. W. Scholz, J. A. Botia, M. Martinez, J. C. Corvol, S. Lesage, J. Jankovic, L. M. Shulman, M. Sutherland, P. Tienari, K. Majamaa, M. Toft, O. A. Andreassen, T. Bangale, A. Brice, J. Yang, Z. Gan-Or, T. Gasser, P. Heutink, J. M. Shulman, N. W. Wood, D. A. Hinds, J. A. Hardy, H. R. Morris, J. Gratten, P. M. Visscher, R. R. Graham, and A. B. Singleton. Identification of novel risk loci, causal insights, and heritable risk for parkinson's disease: a meta-analysis of genome-wide association studies. *The Lancet Neurology*, 18(12):1091–1102, 2019. doi: 10.1016/S1474-4422(19)30320-5.

[29] T. R. Sampson, J. W. Debelius, T. Thron, S. Janssen, G. G. Shastri, Z. E. Ilhan, C. Challis, C. E. Schretter, S. Rocha, V. Gradinaru, M. F. Chesselet, M. Khacho, R. Knight, and S. K. Mazmanian. Gut microbiota regulate motor deficits and neuroinflammation in a model of parkinson's disease. *Cell*, 167(6):1469–1480.e12, 2016. doi: 10.1016/j.cell.2016.11.018.

[30] S. Holmqvist, O. Chutna, L. Bousset, P. Aldrin-Kirk, W. Li, T. Björklund, Z. Y. Wang, L. Roybon, R. Melki, and J. Y. Li. Direct evidence of parkinson pathology spread from the gastrointestinal tract to the brain in rats. *Acta Neuropathologica*, 128(6):805–820, 2014. doi: 10.1007/s00401-014-1343-6.

[31] B. Jassal, L. Matthews, G. Viteri, et al. The reactome pathway knowledgebase. *Nucleic Acids Research*, 48(D1):D498–D503, 2020. doi: 10.1093/nar/gkz1031.

[32] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, 57:289–300, 1995. doi: 10.1111/j.2517-6161.1995.tb02031.x.

[33] P. Virtanen, R. Gommers, T. E. Oliphant, et al. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.

[34] S. Seabold and J. Perktold. Statsmodels: econometric and statistical modeling with python. In *Proceedings of the 9th Python in Science Conference*, pages 92–96, 2010. URL `https://www.statsmodels.org`.

[35] J. D. Hunter. Matplotlib: a 2d graphics environment. *Computing in Science & Engineering*, 9: 90–95, 2007. doi: 10.1109/MCSE.2007.55.

[36] C. R. Harris, K. J. Millman, S. J. van der Walt, et al. Array programming with numpy. *Nature*, 585:357–362, 2020. doi: 10.1038/s41586-020-2649-2.

[37] J. W. Squair, M. Gautier, C. Kathe, et al. Confronting false discoveries in single-cell differential expression. *Nature Communications*, 12:5692, 2021. doi: 10.1038/s41467-021-25960-2.