Supplementary Information

For Trezzi, Aho et al. (2022)

"An archaeal compound as a driver of Parkinson's disease pathogenesis"

Contents

Supplementary Methods

Patient cohorts

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. Faecal samples from PD patients and healthy controls 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[1,2]. Subjects with idiopathic rapid-eye-movement sleep behaviour disorder (iRBD) were recruited at the same clinic, diagnosed according to consensus criteria of the International RBD study group[3] using video-assisted polysomnography, and were included only if they showed no signs of a neurodegenerative disorder. As additional control subjects, we collected faecal samples from 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[4]. All subjects except household controls were interviewed and examined by an expert in movement disorders.

Marburg Cohort

Patients with polysomnography-confirmed iRBD 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)[5]. 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[6] and PD-NMS[7] questionnaires. Motor function was evaluated with the UPDRS[8]. Additionally, patients were asked to complete the RBD-Sleep questionnaire[9].

Sample exclusions

The initial set of samples consisted of 50 PD, 30 iRBD and 50 control subjects. 3 PD and 2 iRBD cases were subsequently excluded for clinical reasons (adjusted diagnosis), 1 iRBD and 1 PD subject for logistical reasons, and 1 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 summarised below:

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

NCER-PD cohort

For use of clinical and 16S rRNA gene amplicon sequencing data from the Luxembourg Parkinson's Study (LUXPARK) cohort[10], samples were selected with the following criteria: (i) 16S rRNA gene amplicon sequencing data available, (ii) only patients with typical PD diagnosis, (iii) controls genetically related to patients were excluded. Patients with an unclear status of PD diagnosis or an atypical PD diagnosis were excluded. PD patients were defined as typical PD according to the UKPDSBB clinical diagnostic criteria[4]. The final dataset comprised 191 PD cases and 337 healthy controls.

All study participants gave written informed consent, and the study was performed in accordance with the Declaration of Helsinki. The Luxembourg Parkinson's Study[10] was approved by the National Ethics Board (CNER Ref: 201407/13) and Data Protection Committee (CNPD Ref: 446/2017).

Biomolecular extractions

Post-treatments and quality checks

After extraction, DNA and RNA were purified prior the sequencing analysis by using the following commercial kits respectively: Zymo DNA Clean&Concentrator-5 and Zymo RNA Clean&Concentrator-5. 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; Qubit RNA BR Assay kit).

Metagenomics and metatranscriptomics

Sequence analysis (IMP)

For all samples, metagenomic (MG) and metatranscriptomic (MT) sequence data were processed and co-assembled using the Integrated Meta-omic Pipeline (IMP)[11] version 01.07.2020 (https://git-r3lab.uni.lu/IMP/imp3, tag 6f1badf7) which 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 co-assembled using the IMP-based iterative co-assembly using MEGAHIT[12] 1.0.3. After co-assembly, the prediction and annotation of open-reading frames (ORFs) were performed using IMP and followed by functional and taxonomic annotation at read and contig level. MG and MT read counts for the predicted genes obtained using featureCounts[13] were linked to functional (KEGG[14], Pfam[15]) and taxonomic annotations (mOTUs 2.5.1[16], Kraken2[17] using the maxikraken2_1903_140GB database). mOTUs annotations were used to generate read count matrices for each taxonomic rank (phylum, class, order, family, genus and species) by summing up reads at the respective levels.

Search for Hmd cofactor biosynthesis genes

Data was mined for specific genes of interest (HmdB and HmdC, encoding the biosynthesis of the cofactor of the archaeal enzyme 5,10-methenyltetrahydromethanopterin hydrogenase (Hmd)) using Diamond[18,19] (v2.0.11.149) in the blastp mode, with the following settings: --evalue 1e-6 --very-sensitive, against a non-redundant reference protein database based on the ORFs predicted using Prokka[20] (from the IMP workflow described above) for all samples.

16S rRNA gene amplicon sequencing analysis (NCER-PD LUXPARK cohort)

Sampling, processing and sequencing of NCER-PD LUXPARK stool samples were done as described previously[21]. The 16S rRNA gene amplicon sequencing data was processed with the dadasnake[22] workflow, a Snakemake[23] pipeline to process amplicon sequencing data, based on DADA2[24]. Amplification primers were removed using cutadapt[25], allowing 20% mismatches and no indels. Quality filtering, ASV generation and chimera removal were performed in DADA2. Reads were truncated at positions with less than 10 Phred score quality, or at 240 bp. The quality filtering kept only sequences with a maximum expected error of 2 and 240 bp length. Downsampling was performed to 25000 reads using seqtk (https://github.com/lh3/seqtk) and samples with smaller library sizes were removed from the downstream analysis. ASVs were generated in pooled mode for the whole study using DADA2 default parameters. For merging forward and reverse ASVs, a minimum overlap of 12 bp was required. Chimeric sequences were removed based on the consensus algorithm. Taxonomic classification was performed against SILVA v138[26] using the naïve Bayesian classifier implemented in mothur[27].

Metaproteomics

Sample preparation

Protein samples were reduced with 2 µL 25 mM DTT in 20 mM ammonium bicarbonate (Sigma-Aldrich) for 1 h at 60 °C. Subsequently, 4 µL of 100 mM iodoacetamide (Merck) solution in 20 mM ammonium bicarbonate was added and incubated for 30 min at 37 °C in the dark. Next, 5 µL of 10% formic acid was added as well as 70 µL 100% acetonitrile (ACN) to reach a final organic content higher than 50% (v/v). 2 µL SP3 beads per sample were washed

with water three times with subsequent addition of the sample. After protein binding to the beads, the supernatant was discarded. Beads were washed twice with 200 µL 70% (v/v) ethanol, and once with 200 µL ACN. Protein lysates were proteolytically cleaved using trypsin (1:50) overnight at 37 °C. As trypsin is added in aqueous solution to the samples, the proteins are not bound to the beads during enzymatic cleavage. ACN was then added to each sample to reach a final organic content higher than 95% (v/v). After peptide binding to the beads, the samples were washed with pure ACN on the magnetic rack. Finally, the peptides were eluted in two steps. First, with 200 µL 87% ACN (v/v) containing 10 mM ammonium formate (pH 10), then by adding two times 50 µL water containing 2 % (v/v) DMSO and combination of the two aqueous supernatants. Thus, two fractions of peptides were generated, which were evaporated and re-dissolved in water containing 0.1 % formic acid (20 µl).

Data analysis

For each sample, the predicted proteins were concatenated with a cRAP database of contaminants and with the human UniProtKB Reference Proteome prior to the metaproteomic search. In addition, reversed sequences of all protein entries were added to the databases for the estimation of false discovery rates. The search was performed using SearchGUI[28] (v. 3.3.20) with the X!Tandem, MS-GF+ and Comet search engines with the following parameters: trypsin was used as the digestion enzyme and a maximum of two missed cleavages was allowed. The tolerance levels for matching to the database was 10 ppm for MS1 and 0.05 Da for MS2. Carbamidomethylation of cysteine residues was set as a fixed modification and oxidation of methionines was allowed as a variable modification. Peptides with length between 7 and 60 amino acids, and with a charge state composed between +2 and +4 were considered for identification. The results from SearchGUI were merged using PeptideShaker-1.16.45 and all identifications were filtered to achieve a protein false discovery rate (fdr) of 1%.

Metabolomics

Metabolite extraction from stool samples for untargeted LC-MS (polar fraction)

To extract metabolites, 500 µL of MilliQ water was added to 50 mg faecal matter. Following this, samples were homogenized using a Precellys24 homogenizer (Bertin Technologies) using the following conditions 5 ceramic beads (1.4 mm), 6000 rpm and 1 x 30 s at 0 to 5 °C. The homogenate was centrifuged at 21,000 x g for 5 min at 4 °C. A total volume of 75 µL of the supernatant was added to 75 µL MilliQ water, including internal standards (caffeine-trimethyl-$^{13}C_3$; sucralose; N-trifluoroacetyl-L-lysine; 4-chloro-DL-phenylalanine; 2-chloroquinoline-3-carboxylic acid; 6-chloropurine riboside; c = 1 µg/mL each) and vortexed. 120 µL of the mixture was sprayed into 600 µL of ice-cold acetonitrile + 1% formic acid and incubated for 5 min at 4 °C under shaking, then centrifuged at 21,000 x g for 5 min. The supernatant was collected and filtered using a Phree phospholipid removal plate (Phenomenex). 500 µL were concentrated under vacuum (SpeedVac, Labconco) at -4 °C overnight. The resultant dry residues were reconstituted in 100 µL of 35% MeOH and transferred to LC vials for subsequent LC-HRMS analysis.

LC-MS/MS measurements

Untargeted LC-MS of the polar fraction was performed using an Agilent 1290 LC coupled to an Agilent 6560 Q-TOF MS equipped with a Dual Agilent Jet Stream ESI source. The analytical column (Phenomenex Kinetex 2.6µm EVO C18 100 Å, 150 x 2.1 mm) used in this study was coupled to a Phenomenex SecurityGuard column (2.1 mm cartridge for EVO C18) and maintained at 30 °C. The flow rate was set to 0.25 mL/min and the mobile phases consisted of 0.1% formic acid in MilliQ water (18.2 MΩ•cm, <3 ppb TOC; Eluent A) and 0.1% formic

acid in 50:50 methanol/acetonitrile (Eluent B). The run consisted of an isocratic delivery of 5% Eluent B over 5 min, followed by a linear gradient to 98% Eluent B over 15 min, isocratic delivery of 95% Eluent B for 10 min, and a re-equilibration phase on starting conditions for 10 min. The autosampler was kept at 4 °C and the injection volume was 10 µL.

MS experiments were performed using electrospray ionization in positive (+ESI, capillary voltage of 3 kV, nozzle voltage of 1 kV) and negative mode (-ESI, capillary voltage of 2000 V, nozzle voltage of 500 V). The protonated/deprotonated molecules were monitored in high resolution mode (slicer position: 5) and Extended Dynamic Range (2GHz) with the following Q-TOF MS conditions: drying gas temperature: 325 °C, drying gas flow: 10 L/min (nitrogen), nebulizer: 35 psig, sheath gas temperature: 350 °C, sheath gas flow: 12 L/min, fragmentor: 380 V, Oct RF Vpp: 750 V. Full scan spectra were acquired from m/z 100 to 1100 (2 spectra/sec). External mass calibration was performed before measurement of each set of samples. In addition, a reference solution was used for online mass correction during the acquisition. All data were acquired with Agilent Mass Hunter LC/MS Data Acquisition (ver B.08.00).

Data analysis for LC-MS/MS

Data alignment and peak picking was performed using Progenesis (version 3; Waters Corporation, Newcastle upon Tyne, U.K.). The resulting features from positive and negative mode (10,535 positive, 5,520 negative) were compared with ANOVA and post-hoc Tukey HSD tests of normalized peak intensities, corrected for age and sex. Further testing focused on peaks with false discovery rate < 0.05, as well as the 500 most intense masses across all samples in positive mode. Additionally, custom lists were used to look for chemicals associated with molecules of interest (2-HP and chlorpyrifos) and literature-mined neurotoxicants[29,30]. Data extraction and identification were performed with Proteowizard[31] MSConvertGUI (version

3.0.19014-f9d5b8a3b), Shinyscreen[32] and MetFrag[33], coupled to MassBank of North America[34] and PubChemLite[35]; analysis parameters are provided in Supplementary Table 14.

Metabolomics for archaeal samples

Intracellular metabolites from archaeal cell pellets were resuspended in 250 µL methanolic extraction fluid. The cell suspension was homogenized at 6800 rpm for 2x30 s at 0 to 5 °C and afterwards incubated for 10 min at 4 °C under shaking. After adding 130 µl water and 200 µl chloroform, the mixture was shaken for 15 min at 15 °C. Then, the mixture was centrifuged at 21,000 x g for 5 min at 15 °C. 100 µL of the polar (upper) phase was transferred to GC glass vial with micro insert (5-250 µL) and dried by evaporation under vacuum at -4 °C.

Derivatization was performed using a multi-purpose sample preparation robot (Gerstel). Dried extracts were dissolved in 20 µL pyridine, containing 20 mg/mL methoxyamine hydrochloride (Sigma-Aldrich), for 120 min at 45 °C under shaking. After adding 30 µl (medium extracts) or 20 µL (cell extracts) N-methyl-N-trimethylsilyl-trifluoroacetamide (Macherey-Nagel), samples were incubated for 30 min at 45 °C under continuous shaking.

GC-MS analysis was performed on an Agilent 7890A GC coupled to an Agilent 5975C inert XL Mass Selective Detector (Agilent Technologies). A sample volume of 1 µl was injected into a split/splitless inlet, operating in split mode (10:1) at 270 °C. The gas chromatograph was equipped with a 30 m (I.D. 0.25 mm, film 0.25 µm) DB-5ms capillary column (Agilent J&W GC Column) with 5 m guard column in front of the analytical column. Helium was used as carrier gas with a constant flow rate of 1.2 mL/min. The GC oven temperature was held at 90 °C for 1 min and increased to 220 °C at 10 °C/min. Then, the temperature was increased to 280 °C at 20 °C/min followed by 5 min post run time at 325 °C. The total run time was 22 min. The transfer line temperature was set to 280 °C. The MSD was operating under electron ionization

at 70 eV. The MS source was held at 230 °C and the quadrupole at 150 °C. Mass spectra were acquired in full scan mode (m/z 70 to 700).

Additional GC-MS measurements were performed in selected ion monitoring mode for precise semi-quantification of 2-HP and β-glutamic acid. The following masses were used for quantification (QI) and qualification of the target analytes: *m/z* 122.1, 152.1 (QI), 166.1 (dwell times: 30 ms) for 2-HP 1TMS, m/z 232.1 (QI), 306.1, 348.1 (dwell times: 30 ms) for β-glutamic acid 3TMS, *m/z* 206.1, 239.1, 267.1 (QI) (dwell times: 50 ms) for pentanedioic acid-D6 2TMS (Internal Standard) and *m/z* 220.1, 310.2, 323.2 (QI) (dwell times: 50 ms) for [U-^{13}C]-ribitol 5TMS (Internal Standard).

A low-level background signal for 2-HP, related to derivatisation reagents, was seen in measured blank samples. This offset was subtracted from actual sample values prior to analyses and visualisations.

Statistics for multi-omics

Independent Principal Component Analysis (IPCA) was performed using the "ipca" function from the mixOmics package[36] (v. 6.16.3) separately for each omic dataset: metabolomics, MG and MT species-classified read counts from the mOTUs analysis, MG and MT read based functionally classified (KEGG orthologs, KOs) read counts, as well as MP taxa and KOs. For all data types, features present in 1/3 or less of all samples were excluded prior to any further steps. For MG and MT taxon and KO data, read counts were converted to relative abundances and then centered log ratio (CLR) transformed prior to IPCA, with an offset corresponding to the lowest non-zero relative abundance in a sample, and not scaled to unit variance ("scale" parameter set to FALSE for the "ipca" function). For metabolomic and MP data, "scale" was

set to TRUE. The number of components (2–4 depending on data type) was selected based on kurtosis; results are presented for the first 2 components.

MG and MT taxon data was processed using phyloseq[37] (v. 1.36.0). Very rare taxa (present in less than 10 samples) were excluded prior to any analyses. Alpha diversity measures (Shannon and inverse Simpson index) were calculated with phyloseq, and Bray-Curtis dissimilarity and non-metric multidimensional scaling (NMDS) for β diversity with the metaMDS function from the package vegan[38] (v. 2.5-7). All diversity comparisons were run on species-level data subsampled to the lowest number of sequences in a sample. Alpha diversity testing for clinical variables included Pearson and Spearman correlations for numeric variables and Kruskal-Wallis (multiple levels) and two-sided Wilcoxon Rank Sum tests (two levels) for categorical variables. Beta diversity comparisons were performed with vegan, using the adonis function for single-variable tests and adonis2 for models with multiple variables.

Differential abundance tests for MG and MT taxon data were performed for taxonomical levels from phylum to species with DESeq2[39] (v. 1.32.0). DESeq2 was also used for differential abundance testing of MG and MT reads with functional annotations (KOs). All DESeq2 models were corrected for age and sex, and included differential analyses comparing PD patients versus controls, iRBD patients versus controls, PD patients versus iRBD patients, as well as all cases (PD or iRBD) against controls. Features (taxa or KOs) that were not present in more than 1/3 of all samples were excluded from differential abundance comparisons. Differential abundance of 16S rRNA gene amplicons from the NCER-PD cohort was also tested with DESeq2, corrected for age and sex, including all taxa present in more than 10% of samples. In all taxonomy comparisons with DESeq2, the parameter "sfType" was set to "poscounts" to avoid issues related to zeros in size factor estimation; in KO comparisons, the default option, "ratio", was used.

For metaproteomic and metabolomic data, differential abundance was tested using log-transformed data with ANOVA (two-category comparison: case (PD or iRBD) versus control; three-category comparison: PD/iRBD/control; both comparisons corrected for age and sex), and with two-sided t-tests (pairwise contrasts between PD/control, iRBD/control and PD/iRBD, no confounder correction). Features (metabolites, taxa, or KOs) not present in more than 1/3 of all samples were excluded before comparisons.

Associations of selected key metabolites with other metabolites, clinical data, and taxon and functional (KO) data from the different omics were evaluated using Pearson and Spearman correlations, with false discovery rate (fdr) for multiple comparison correction. Features not present in more than 1/3 of all samples were excluded from correlations, except for correlations between 2-hydroxypyridine (2-HP) and *Methanobrevibacter smithii* methanogenesis KOs, where all relevant KOs were included regardless of their prevalence.

Combined analyses of metagenomic, metatranscriptomic, metaproteomic and metabolomic data were performed with the DIABLO approach from the mixOmics package[36] (v. 6.16.3). Metagenomic and metatranscriptomic taxon and KO count data was converted to relative abundance and centered log-ratio transformed using the lowest non-zero relative abundance as offset. Prior to running DIABLO, a design matrix was built by performing PLS for each pair of data types and extracting the correlation coefficients for the pairs. For choosing the number of components, the performance was evaluated with the "perf" function, with M-fold validation, using 10 folds and 10 repeats. The main DIABLO function ("block.splsda"), was performed for two components using default settings, selecting the five most significant features for each data type and component.

Yeast model

Strains and culture medium

Yeast cells were grown in synthetic complement (SC) media containing 6.7 g/L yeast nitrogen base without amino acids, 5 g/L ammonium sulphate supplemented with 2g/L SC. Media was autoclaved and the carbon source (8% Raffinose or galactose) was added to a final concentration of 2%.

The following strains were used:

- W303 (parental strain): *MATa ura3-52 trp1Δ2 leu2-3_112 his3-11 ade2-1 can 1-100*
- HiTox: W303 *pdr5Δ*::KanMX4, pAG306GAL-SNCA-EGFP, pAG304GAL-SNCA-EGFP
- Control: W303 *pdr5Δ*::KanMX4, pAG306GAL-ccdB-EGFP, pAG304GAL-ccdB-EGFP

Enteric neuron model

Enteric neuron derivation

Human induced pluripotent stem cells (hiPSCs) were cultured on Matrigel (Corning) coated dishes in Essential 8 medium (Thermo Fisher) which was refreshed every 24 h. When confluent, cells were regularly split using accutase (Sigma) and seeded into new Matrigel coated plates in maintenance media supplemented with 5 μM Y-27632 (Merck Millipore, 688000). Differentiation towards vagal neural crest was initiated using a published protocol[37],

with slight modifications. Briefly, 24 hours post-seeding of the hiPSCs, media was replaced by Essential 6 (E6) medium (Thermo Fisher) without ROCK inhibitor, but the addition of SB431542 (10 μM, Abcam) and LDN193189 (100 nM, Sigma). On day 2, E6 medium was additionally supplemented with CHIR99021 (3 μM, Axon). From day 4 onwards, E6 medium was gradually replaced with increasing amounts of N2 medium until day 10. N2 medium was prepared with DMEM-F12 (Gibco) with 1% N2 supplement (Invitrogen), 1% L-glutamine (Life Technologies) and 1% NEAA (Thermo Fisher). Retinoic acid (1 μM, Sigma) was added to the media from day 6 to day 11. On day 11, differentiated cells were sorted for CD49D surface marker using a PE/Cy7 anti-human CD49d antibody (BioLegend, 304314), and aggregated into 3D spheroids (2 million cells per well) in Ultra Low Attachment 6-well culture plates (VWR). Spheroids were cultured for 4 additional days in suspension in Neurobasal (NB) medium supplemented with L-glutamine (Life Technologies), N2 (Invitrogen) and B27 (Life Technologies) containing CHIR99021 (3 μM, Axon) and FGF2 (10 nM, PeproTech). The spheroids were subsequently plated on poly-ornithine/laminin/fibronectin (PO/LM/FN)-coated dishes in the same media as described before containing GDNF (25 ng/ml, PeproTech, 450-10) and ascorbic acid (100 μM, Sigma). To perform the PO/LM/FN coating, plates were first incubated with 1 x PBS solution containing 15 μg/mL poly-ornithine overnight at 37˚C. The next day, plates were washed with 1 x PBS once and a solution with 1 x PBS containing 2 μg/mL LM and 2 μg/mL FN was added. Dishes were incubated overnight at 37˚C. Before plating the spheroids, LM/FN solution was completely removed, and plates were allowed to air dry at RT for 15-20 minutes under the laminar flow. After plating, enteric neurons started migrating out of the spheroids after 1–2 weeks in culture and media was exchanged every 3-4 days.

Evaluation of pluripotency

For pluripotency stainings, SSEA4 (Abcam, cat no. ab16287, 1:50), Oct4 (Abcam, cat no. ab19857, 1:400), Nanog (Millipore, cat no. AB5731, 1:200) and Hoechst 33342 (Life Technologies, cat no. H21492) were used. The staining procedure was performed as described in the main methods section, with the difference that, at the last step, cells were mounted with Fluoromount-G mounting medium (Southern Biotech, cat no. 0100–01) and imaged on a confocal (Zeiss LSM 710) laser-scanning microscope.

Enteric neuron characterization

For the characterization of enteric neuronal identity, we used PHOX2A (Abcam, cat no. ab54847, 1:50), MASH1 (BD, cat no. 556604, 1:400), TUJ1 (Covance, cat no. PRB-435P-0100, 1:600), serotonin (Millipore, cat no. MAB352, 1:100), GABA (Abcam, cat no. ab17413, 1:400), TH (Sigma, cat no. T2928, 1:400) and Hoechst 33342 (Life Technologies, cat no. H21492). The staining procedure was performed as described in the main methods section, with the difference that, at the last step, cells were mounted with Fluoromount-G mounting medium (Southern Biotech) and representative images were taken using an inverted microscope (Zeiss Axio ObserverZ1).

Tetrazolium assay for 2-HP toxicity evaluation

Thiazolyl Blue tetrazolium Bromide (MTT; Sigma) was reconstituted to a 5 mg/mL solution in differentiation media and filtered. 10 µL of the solution was added to every well containing 100 µL of media. Plate was incubated for 2 h in a normal incubator (37 °C, 5% CO2). Media

containing MTT solution was removed and 100 µL of DMSO (Sigma) was added to lyse the cells. Cells were dispersed by pipetting vigorously and absorbance was read at 570 nm, using the microplate Cytation 5M reader (Biotek).

α-synuclein aggregation staining

After treatment with 2-HP, cells were fixed in 4% paraformaldehyde (Merck Millipore, 1004965000) for 15 min at RT and then washed 3 times for 5 min with PBS at RT. Before immunostaining, a permeabilization step was performed using 0.05% Triton-X100 solution in PBS for 10 min at 4 ˚C. Cells were then blocked for 1 h at RT with 10% FBS in PBS. Incubation with α-synuclein (αSyn) antibody (NOVUS biologicals, NBP1-05194, 1:1000), αSyn filament antibody (Abcam, ab209538, 1:5000), TUJ1 (Millipore, AB9354, 1:600) and Cleaved Caspase-3 (Asp175) antibody (CST, 9661, 1:200) was performed overnight at 4 °C in blocking buffer. The following day, cells were washed with PBS 3 times for 5 min at RT, and then incubated with the corresponding secondary antibodies, Alexa Fluor anti-chicken 488 (Invitrogen, A-11039, 1:1000), Alexa Fluor anti-mouse IgG1 647 (Invitrogen, A-21240, 1:1000) and Alexa Fluor anti-rabbit 568 (Invitrogen, A11036, 1:1000), for 2 h at room temperature in blocking buffer. Hoechst 33342 solution (Invitrogen, 62249, 1:1000) was added during this step to stain the nuclei. Afterwards, cells were washed 3x with PBS, and imaged with an OPERA QEHS spinning disk microscope (Perkin Elmer) to acquire representative images for each marker, screening 30 fields per well and 8 wells per condition and per plate, using a 20-x water immersion objective. Images were analysed in MATLAB (Version 2019a, The MathWorks Inc.), with in-house analysis algorithms which automate structure-specific feature extraction based on mask definition of positive staining. Thus, the expression level of any marker can be expressed as a sum of positive pixels and normalized to positive neuronal staining.

Mouse model

Genotyping

DNA was extracted from mouse pup ear punches (± 2mm^3) with the Quanta Bio DNA extraction kit (Extracta DNA Prep for PCR, Quanta Bio), according to the manufacturer's protocol. 50 µL of extraction reagent was added to each tube with the ear punches. The tubes were heated at 95 °C for 30 minutes and 50 µL of stabilization reagent was added to each tube. After quick vortexing (2-3 seconds), the resulting supernatant was stored at -20 °C until PCR amplification.

After thawing, the DNA solution was mixed with 2xKAPA2G Fast Hot Start Genotyping Mix and PCR primers. To detect the transgene, we used primers to human SNCA with a length of 366 bp: forward 5'-AAG-GCC-AAG-GAG-GGA-GTT-GTG-3'; reverse 5'-TTC-CTC-AGA-AGG-CAT-TTC-ATA-AGC-3'. The reaction tubes with 25 µL of total mix were then put in a thermocycler and analysed using the following program: initial degradation: 3 min at 95 °C, followed by 35 cycles of 15 sec at 95 °C (degradation), 15 sec at 60 °C (annealing), 15 sec at 72 °C (elongation), and 90 sec at 72 °C (final elongation). The PCR products were visualised on agarose gel using standard procedures.

Stereotactic injections of 2-HP

Mice were injected intracranially under isoflurane anaesthesia (2%) on a heating pad. A 1 cm long mid-line scalp incision was made into the disinfected skull area and a 0.5 mm hole drilled unilaterally into the skull using stereotaxic coordinates for striatum according to the Mouse Brain Atlas of Franklin and Paxinos. 100 mM 2-HP (Sigma) or PBS vehicle solution (control mice) were administered, in volumes of 2 µL, within the right dorsal striatum at the following

relative-to-bregma coordinates: anterior +0.5 mm, lateral +2.1 mm; depth +3.2 mm. The 24-gauge blunt tip needle of a Hamilton syringe (7105KH) was inserted down 3.3 mm for 10 seconds to form an injection pocket, and the needle remained in place for 2 minutes before and after the injection procedure. The hole was covered with bonewax (Lukens) and the wound closed using 7 mm Reflex wound clips (Fine Science Tools). 2% xylocaine gel was applied to the wound, and mice were allowed to recover from anaesthesia before being put back into their home cages. Animals were monitored daily for 2 weeks post injection, and then weekly until euthanasia.

Immunofluorescence: quantification

For quantification of TH-positive axons in the striatum, 3 images (223.8 x 167.7 µm each) of the dorsal striatal area on each section were taken at 40 x magnification using the Apotome system to obtain optical planes without out of focus reflection. Images were captured at optical plane levels where antibody stainings were uniform. Parameters were kept identical for each image acquisition. All images were converted into 8-bit tiff files for analysis, which was done with the IMAGE J software (version 1.51j8). After manual thresholding to capture only immunopositive structures, the mean grey value, or mean pixel intensity (scale 0-255 for 8-bit tiff images) for the totality of immunopositive signals was measured on each image[40].

The quantification of TH-positive neurons in the substantia nigra was performed as described, using a method that has been validated by unbiased stereology[41].

To quantify S129 phosphorylated α-synuclein (phospho-αSyn) positive synaptic boutons in the striatum and the substantia nigra, pictures were collected on sections doubly labelled for TH and for phospho-αSyn. For the striatum, the % area occupied by phospho-αSyn-positive signals was captured, and for the substantia nigra, since the summed area of this region of interest had

been measured in the estimation of TH-positive neuron number[41], the number of phospho-αSyn-positive synaptic boutons was determined.

To quantify phospho-αSyn-positive structures in cell bodies (prefrontal cortex), pictures from sections stained for phospho-αSyn were taken with a 40x objective from the striatum, the substantia nigra and the prefrontal cortex (3 sections/animal, 3 images/section), and analysed digitally using the open-source image analysis software FIJI (v. ImageJ 1.53f51). Positive phospho-αSyn signals were manually thresholded, using the "analyse particles" plugin of FIJI to quantify the number of positive particles that were in the size range of cell bodies (over 8 µm in diameter). Counterstaining with nuclear stain DAPI confirmed the cell body localisation of these signals.

Supplementary Tables

Due to their size and number, Supplementary Tables are provided as an excel file. Captions:

Supplementary Table 1. Summary statistics for MG (metagenomic) and MT (metatranscriptomic) sequence data.

Supplementary Table 2. Correlations for identified protein counts, confounding variables and microbial diversity. Values for continuous variables are provided for Spearman and Pearson correlations calculated with full data as well as subsets containing only one group of samples (control, iRBD or PD), and for categorical variables for Wilcoxon rank sum (2 categories) or Kruskal-Wallis (3 categories) tests.

Supplementary Table 3. Annotations of top 10 functional classifications (KOs) from Independent Principal Component Analysis (IPCA) of each data type, as seen in Extended Data Fig 2.

Supplementary Table 4. Comparisons of microbial diversity based on MG and MT mOTUs. a. Alpha diversity (three metrics: observed richness, Shannon index, inverse Simpson index): continuous variables tested with Spearman and Pearson correlations, and categorical variables with Kruskal-Wallis (3 categories) or two-sided Wilcoxon rank sum (2 categories) tests. b. Beta diversity, testing performed with PERMANOVA ("adonis" command for single-variable tests, columns p_adonis and R2_adonis, and "adonis2" for multi-variable tests of main contrast and confounders, columns p_adonis2 and R2_adonis, both functions from the R package vegan).

Supplementary Table 5. Differential abundance of taxonomically classified MG and MT sequence reads (from mOTUs), compared with DESeq2, corrected for age and sex. Contrasts include the two-category case (PD or iRBD) vs control variable, and pairwise contrasts for the three-group variable (PD/Ctrl, iRBD/Ctrl, PD/iRBD).

Supplementary Table 6. Differential abundance of taxonomically classified metaproteomic (MP) spectra. a. Descriptive statistics by group. sd = standard deviation, fc = fold change. b. Results of statistical comparisons between groups. ANOVA for the case/control and control/iRBD/PD comparisons, corrected for age and sex, t-tests without confounder correction for two-group comparisons; df = degrees of freedom, fdr = false discovery rate corrected *P*.

Supplementary Table 7. Differential abundance of functionally classified (KEGG orthologs; KOs) MG and MT sequence reads, compared with DESeq2, corrected for age and sex.

Contrasts include the two-category case (PD or iRBD) vs control variable, and pairwise contrasts for the three-group variable (PD/Ctrl, iRBD/Ctrl, PD/iRBD).

Supplementary Table 8. Differential abundance of functionally classified (KOs) MP spectra. a. Descriptive statistics by group. sd = standard deviation, fc = fold change. b. Results of statistical comparisons between groups. ANOVA for the case/control and control/iRBD/PD comparisons, corrected for age and sex, t-tests without confounder correction for two-group comparisons; df = degrees of freedom, fdr = false discovery rate corrected *P*.

Supplementary Table 9. Differential abundance of untargeted GC-MS data and short-chain fatty acids (SCFAs). a. Descriptive statistics by group. sd = standard deviation, fc = fold change. b. Results of statistical comparisons between groups. ANOVA for the case/control and control/iRBD/PD comparisons, corrected for age and sex, t-tests without confounder correction for two-group comparisons; df = degrees of freedom, fdr = false discovery rate corrected *P*.

Supplementary Table 10. Annotations of functional classifications (KOs) selected in the DIABLO analysis, as seen in Extended Data Fig 5.

Supplementary Table 11. Statistical tests for clinical variables, 2-hydroxypyridine (2-HP) and β-glutamate (BG). Values for continuous variables are provided for Spearman and Pearson correlations, and for categorical variables for Kruskal-Wallis (3 categories) or two-sided Wilcoxon rank sum (2 categories) tests.

Supplementary Table 12. Lists of masses and statistical comparisons (ANOVA between PD/iRBD/Ctrl, corrected for age and sex) from LC-MS data. a. Positive mode; b. Negative mode.

Supplementary Table 13. Compounds of interest for the screening of chlorpyrifos or niacin derived production of 2-HP.

Supplementary Table 14. a. Summary of software parameters used, to support the methods text; b. Identification confidence levels used to annotate tentative identifications.

Supplementary Table 15. Summary of tentatively identified chemicals from the 2-HP specific chemicals (Supplementary Table 13) and LitMinedNeuro list (described in Schymanski et al (2019); available from https://comptox.epa.gov/dashboard/chemical-lists/litminedneuro).

Supplementary Table 16. a. Agriculturally relevant chemicals identified in the top 500 most intense masses. b. List of the Top 500 masses by intensity across all cohort samples in positive mode, extracted with Progenesis (Waters), as input into Shinyscreen.

Supplementary Table 17. Supplementary Table 17. Correlations (Spearman and Pearson) for relative abundances (with or without CLR-transformation) of 2-HP and BG with MG and MT

sequence reads classified with either mOTUs or Kraken, as well as with MP spectra tested with either full data (regardless of group) or subsets of data including only cases or only controls.

Supplementary Table 18. Pearson and Spearman correlations of 2-HP and BG with other metabolites, tested with either full data (regardless of group) or subsets of data including only cases or only controls.

Supplementary Table 19. Correlations of functionally classified (KO) data with 2-HP and BG, tested with either full data (regardless of group) or subsets of data including only cases or only controls. a. Correlations of all KOs from MG and MT sequence reads and MP spectra with 2-HP and BG. b. Correlations of 2-HP and MG and MT KOs present in a list of *Methanobrevibacter smithii* methanogenesis KOs exported from KEGG.

Supplementary Table 20. Differential abundance (DESEq2) of 16S rRNA gene amplicons from the NCER-PD cohort, comparing PD and control subjects, corrected for age and sex.

Supplementary Discussion

Metagenomics and metatranscriptomics: taxonomic analyses

Alpha diversity

Overall comparisons for clinical variables and alpha diversity calculated based on taxonomic profiling of metagenomic and metatranscriptomic reads with mOTUs2[16] (Supplementary Table 4a) showed no statistically significant differences between the different groups with two categories (controls versus cases, including PD and iRBD) or three categories (controls versus iRBD or versus PD) considering any of the tested diversity measures (observed richness, Shannon and inverse Simpson indices). The only variables with significant alpha diversity associations (without multiple comparison correction) were entacapone ($P = 0.048$ for Shannon index from MG data) and antidepressant medication ($P = 0.049$ for observed richness from MG data). Both medications were only taken by a small number of subjects (antidepressants: n = 5, entacapone: n = 5). Nevertheless, our results are in line with previous findings that entacapone and SSRIs can influence the gut microbiome[38,3942–44]. As for differences between cases and controls in previously published studies (mostly 16S rRNA gene amplicon-based analyses), the results concerning alpha diversity have been inconsistent[45–47], so it is not unexpected that no clear differences were apparent from our comparisons. Furthermore, broad summary measures such as alpha diversity indices are not as informative as the higher-resolution results of detailed taxonomic data.

Beta diversity

When considering beta diversity for individual variables (Bray-Curtis dissimilarity compared with adonis/PERMANOVA with no confounder correction; Supplementary Table 4b), there was no statistically significant difference between groups with either two or three categories in MG data, but in MT data, both comparisons were significant (P = 0.001 comparing cases to controls, P = 0.007 for the three-group, control/iRBD/PD variable). For MG data, there were significant differences for age, sex, antidepressant medication and the Sniffin' Sticks score for evaluating sense of smell. For MT data, significant variables in addition to the case/control status included constipation, sex, agonist medication, Sniffin' Sticks score and Hoehn and Yahr stage; age was not significant. In a combined model including age, sex, and the case (iRBD or PD) versus control variable, all three variables were significant when using MG data, while again, age was not significant for MT data. In previous studies, 16S rRNA gene amplicon based beta diversity has typically differed between groups[46], although it only explains a small amount of the data variability (R2 between < 1% to <13% depending on study, according to a recent meta-analysis[47]). In our comparisons, even in MT data, where differences between groups had a lower P-value, the variance explained by the case/control variable was less than 2%. This implies that other variables may have stronger effects. In general, more detailed comparisons involving specific microbial taxa are likely to be more informative than summary measures.

Taxon abundance

Considering the differential abundance comparisons of mOTUS2-based taxonomic data (Extended Data Fig 3c-d, Supplementary Table 5), there were more numerous differentially abundant taxa between cases and controls in MT than in MG data. Considering pairwise comparisons of the three categories (controls vs iRBD vs PD), there were also many more

differing taxa between iRBD and controls or PD and controls than between iRBD and PD, highlighting that subjects with iRBD may already have a PD-like microbiome[48].

Some previously reported microbial alterations were only detected in MT but not in MG comparisons, or vice versa. The commonly reported increase in the genus *Lactobacillus* in PD compared to controls[46,47] was detected in MG comparisons, but only statistically significant when considering the two-category case vs control variable. Several other often-seen differences, such as decreases in *Blautia*[47,49–51], *Eubacterium*[1–3], *Faecalibacterium*[46,47,49,51,52,54,55], *Lachnospiraceae*[46,50,51,55,56], *Prevotella*[46,52,57] and *Roseburia*[46,47,49–52,54] in PD were only statistically significant in comparisons with MT data. *Bifidobacterium*, a genus often reported as increased in PD[46,47,49,51,52,58], was not statistically significant for any of the MG-based comparisons, and was decreased in PD and iRBD when considering the MT data. *Alistipes*, which was reported as increased in PD in a 16S rRNA gene amplicon based meta-analysis[47], was increased in PD in our MG and MT data, as was the genus *Escherichia*, which includes species with established pathogenic potential. Furthermore, *Escherichia* spp. are producers of the bacterial amyloid Curli, which can affect α-synuclein aggregation, and has been hypothesized to play a role in PD pathogenesis[59,60].

Overall, our differential abundance results highlight the importance of exploring also metatranscriptomic and not only metagenomic data to uncover the full picture of microbial abundance shifts relevant in the context of PD.

Metagenomics and metatranscriptomics: functional analyses

In addition to the taxonomic analyses, we compared abundances of functionally classified sequence reads (representing KEGG orthologs; KOs) from MG and MT data between groups with DESeq2[39]. Overall, the number of differentially abundant KOs (Extended Data Fig 3e,

608 Supplementary Table 7) was far greater in the MT data (1,000 significantly different KOs for
609 the case vs control comparison) than in the MG data (28 significantly different KOs for the
610 case vs control contrast). This suggests, in line with the IPCA plots (Extended Data Fig 2), that
611 metatranscriptomic data better captures differences between cases and controls than
612 metagenomic data, and emphasises the importance of functional omic analyses[61].

613 The numbers of KOs that were statistically significantly different in both MG and MT
614 comparisons were limited, with only nine consensus hits seen for the overall case (PD or iRBD)
615 versus control comparisons. Considering the pathways represented by differentially abundant
616 KOs (Extended Data Fig 3e), after three global categories ("Metabolic pathways",
617 "Biosynthesis of secondary metabolites" and "Microbial metabolism in diverse environments")
618 the most commonly observed ones were "Two-component system", "Biosynthesis of amino
619 acids", "ABC transporters" and "Carbon metabolism", all of which contained KOs altered in
620 both directions (either increased or decreased in PD and iRBD). The KOs representing "Two-
621 component system" and "Carbon metabolism" more often had an increased abundance in cases.
622 Other pathways with this pattern included "Biosynthesis of cofactors" and "Pyruvate
623 metabolism". KOs in the "Flagellar assembly" category were often decreased in cases. Overall,
624 the long list of altered MT KOs is a clear indication that not only the structure, but also the
625 function of the active microbial community is strongly affected in PD.

626

627 ## Metaproteomics

628 We assessed correlations between identified protein counts with clinical variables and alpha
629 diversities calculated from taxonomically classified MG and MT sequence reads to determine
630 if variation in the number of identified proteins could be linked to clinical or microbial
631 community features (Supplementary Table 2). The counts were different when comparing 3

groups (PD versus iRBD versus control; P=0.03), with the highest values seen in samples from PD subjects. They were also weakly correlated with alpha diversity based on MG data (considering either Pearson or Spearman correlations). Thereby, the identified proteins also reflect biological differences in the microbial communities.

We excluded any samples with less than 1,000 identified proteins from further comparisons, which included differential abundance testing of spectra classified for taxa and KOs. Based on taxonomic assignments of the metaproteomic data, only two taxa that differed between groups were identified (Supplementary Table 6, Extended Data Fig 3c): *Prevotella copri* was significantly decreased in iRBD compared to controls, and the genus *Lactobacillus* was significantly increased in PD compared to iRBD. However, when considering taxa with $P <$ 0.05 without multiple comparison corrections, the class *Coriobacteriia,* the family *Oscillospiraceace*, the genera *Methanobrevibacter* and *Parabacteroides* and the species *Roseburia faecis* and *Bifidobacterium longum* were increased in cases compared to controls. In contrast, the genus *Roseburia* and the species *Ruminococcus* sp AF21.42 were decreased.

With respect to the functional gene analyses, there were very few statistically significantly differing metaproteomic KEGG orthologs (three for the case vs control variable, nine for iRBD vs control and ten for PD vs iRBD, with no hits for PD vs control; Supplementary Table 8, Extended Data Fig 3e). The differences also exhibited very little overlap with the MG and MT results: two KOs were significant for both MP and MT comparisons of iRBD vs control, and another two for PD vs iRBD.

The discrepancies between the MP data and the MG and MT data may be in part due to the smaller subset of samples used for MP analyses, but they also reflect real biological differences, reflecting the different characteristics of each omic method. Future studies should expand the use of MP in the field of PD microbiome research, as this could provide new insights into the functional differences between groups.

Metabolomics

Untargeted GC-MS

In addition to the two metabolites significantly increased in PD and iRBD, i.e. 2-hydroxypyridine and β-glutamate, there were two other metabolites that differed significantly when comparing all cases (PD or iRBD) to controls and in pairwise iRBD vs control or PD vs control contrasts: glycerol and glucuronic acid. Both were more abundant in fecal extracts of control subjects' samples than those of cases with PD or iRBD (Extended Data Fig 4a, Supplementary Table 9).

Glycerol is a common molecule derived from various dietary sources. There is notable, microbial community related interindividual variation in glycerol metabolism in humans[62]. The between-group differences seen in our results could be due to different rates in glycerol metabolism, reflecting microbial community differences. However, the abundances of the main functions for bacterial glycerol degradation (K06120: glycerol dehydratase and K00086: 1,3-propanediol dehydrogenase) did not differ between the groups in MG or MT data (Supplementary Table 7). In the MP data, K06120 was not detected, and while the mean abundance of K00086 was higher in cases when compared to controls, the difference was not statistically significant after multiple comparison correction (Supplementary Table 8).

Glucuronic acid is involved in the metabolism of xenobiotic compounds, including drugs. Elevated levels of glucuronic acid in plasma have been linked to conditions such as diabetes and liver diseases and even to all-cause mortality[63], but faecal glucuronic acid has not been commonly studied. One study has reported increased levels of glucuronic acid in stool samples of PD subjects with a dyskinetic phenotype compared to other phenotypes, and the authors have hypothesised that it may result from the cleavage of glucuronidated xenobiotics by

bacterial β-glucuronidases[64]. However, our metabolomic data showed an increase in glucuronic acid in controls, not in PD subjects of any phenotype, and our functional (KO) data did not show a significant difference in the abundance of β-glucuronidases (K01195) between groups in any omic dataset (Supplementary Tables 7 and 8). Medications are likely to play a part in the detectable amounts of glucuronic acid, considering that some anti-Parkinsonian medications are glucuronidated[65], as are many other common medications, such as statins[66]. Overall, more research on faecal glucuronic acid and how it correlates with circulating blood levels would be required in future studies.

Short-chain fatty acids

In comparisons of short-chain fatty acid (SCFA) levels between groups, butyric, isobutyric and isovaleric acids were statistically significantly different when comparing all cases (PD or iRBD) to controls (fdr-corrected $P<0.05$; Supplementary Table 9). Butyric acid levels were lower in cases compared to controls, in line with two previous publications exploring fecal SCFAs in PD[7358,67]. In contrast, the pattern for isobutyric and isovaleric acids was the opposite, with higher concentrations in cases. These SCFAs did not differ significantly between groups in previous studies, although both had a slightly higher mean relative concentration in PD compared to controls in one study[58] and a higher mean absolute concentration in the other[67]. Taken together, these results show that larger studies of SCFAs in PD are needed to better understand the likely links between individual SCFAs and the disease.

Untargeted LC-MS/MS

We explored the LC-MS/MS data for 2-HP- and chlorpyrifos-related compounds (Supplementary Table 13). Nicotinic acid was found at level 2a confidence (probable structure

by library spectrum match; Supplementary Table 14 & 15) with statistically significant differences for the control vs PD (uncorrected P=0.0007) and control vs iRBD (uncorrected P=0.0025) comparisons, but not between iRBD and PD cases. Only one other chemical (diethylphosphate) had intensities above quality control thresholds, but no corresponding MS/MS spectra.

Seven chemicals from a published list of neurotoxicants[29,30] were tentatively identified in our data (Supplementary Table 15), one of which has been associated with PD in the literature (1-naphthylamine[68]). However, none of these chemicals were among the masses with a statistically significant difference between the different groups (Supplementary Table 12).

Among the top 500 masses by intensity, only 7 matches (identification level 3; tentative candidates) were found associated with agrochemical information, including two transformation products (Supplementary Table 16). Of the statistically relevant masses (337 in positive mode, 63 in negative mode), only 50 and 8 masses passed quality control. Only 4 in positive mode had scores >4 in MetFrag (agrochemical search), including 2-amino-3-chloro-1,4-naphthoquinone and nicotinic acid (identified also above), plus phenylalanine and phenylephrine. Only one compound in negative mode had a MetFrag score > 3, roxatidine.

No chlorpyrifos-related chemicals were found in the LC-MS data, which suggests that this pesticide is most likely not the source of the elevated 2-HP levels in PD patients.

Relationships of metabolites increased in PD and iRBD with clinical variables

We performed additional comparisons focusing on the two metabolites that were increased in the faecal samples from PD or iRBD patients in comparison to the controls: 2-HP and β-glutamate (Supplementary Data Table 11). When considering the clinical and demographic data, 2-HP was inversely correlated with age in controls but not in PD or iRBD patients. There

was also a borderline difference between sexes (P=0.058) apparent in controls, but not in the PD-only or iRBD-only datasets. Instead, age at diagnosis and 2-HP levels were positively correlated in PD patients, considering either Pearson or Spearman correlation coefficients (Extended Data Fig 4b, Supplementary Data Table 11). For two patients, the interval from diagnosis of PD to sampling was much longer than for the rest of the subjects, and one of two had clearly a higher level of 2-HP than the other. The subject with less 2-HP out of the two had worse symptoms, while the subject with high 2-HP levels was on agonist medication. PD patients taking this medication generally had slightly higher 2-HP levels (P=0.086), but the specific patterns were difficult to evaluate due to heterogeneity of medication use (17 PD patients were not taking any agonist medication, and out of those who were, 14 were taking pramipexole, 6 piribedil, 2 ropinirole, 1 pramipexole and piribedil, and 1 rotigotine).

β-glutamate levels differed between PD patients who did take agonist medication and those who did not, as well as depending on statin medication and entacapone medication, although such medication was only taken by four PD patients. Additionally, in PD patients, β-glutamate was correlated with age at diagnosis when considering the Spearman correlation coefficient, and inversely correlated with non-motor symptom scores (SCOPA-AUT and PD-NMS, both correlation metrics). Since this metabolite was particularly high in iRBD and did not differ between PD and controls, this pattern suggests that it may be associated with earlier disease stages and should be studied further in a disease progression context.

Metabolites increased in PD and iRBD and taxa

To identify microorganisms that could be related to the two metabolites increased in cases, potentially even producing them, we calculated correlations between the abundances of taxa (either relative abundances or CLR-transformed relative abundances, with MT and MG data

annotated using either Kraken2[17] or mOTUs2[16]) and metabolite values (Extended Data Fig 4d, Supplementary Table 17). Focusing on the taxa that were most often seen as significantly correlated (fdr-corrected $P<0.05$) considering the different approaches and either the Pearson or the Spearman correlation coefficients, the main pattern was that both 2-HP and β-glutamate were positively correlated with *Methanobrevibacter smithii,* as well as the parent taxonomic ranks (genus *Methanobrevibacter,* family *Methanobacteriaceae,* order *Methanobacteriales,* class *Methanobacteria* and phylum *Euryarchaeota*). This led us to suspect a connection between archaeal metabolism and the two metabolites. Other taxa with the most overlap between approaches included the phylum *Verrucomicrobia* and the genus *Anaerotruncus*; separate inspection of species-classified data representing this genus showed a less consistent pattern than the one seen for *M. smithii.* Considering MP data (Supplementary Table 17), the only result overlapping with the main significant hits from MG and MT correlations was a significant positive correlation between β-glutamate and the genus *Methanobrevibacter;* additional MP results included a negative correlation between *Prevotella* and β-glutamate.

Relationships of metabolites increased in PD and iRBD with other metabolites

We calculated Pearson and Spearman correlations between all metabolites from the untargeted GC-MS and targeted SCFA measurements to explore their relationships with the two metabolites increased in cases (Extended Data Fig 4e, Supplementary Table 18). The two metabolites that had the highest positive correlation coefficients with 2-HP were unidentified. Their identification was not possible due to their low abundances in the samples. β-glutamate and 2-HP were also intercorrelated.

In negative associations, butyric acid was inversely correlated with β-glutamate, and acetic acid with both metabolites. This is relevant considering the correlation of 2-HP, β-glutamate and

Methanobrevibacter smithii. While *Methanobrevibacter* sp. do not degrade acetate, they can transport it and use it for anabolism[69,70]. *M. smithii* has also been shown to influence acetate production by other gut microbes, although in the published cases, the correlation between acetate and *M. smithii* was positive[71,72]. It would be important to perform further studies concerning the relationships between SCFAs and *M. smithii* in different contexts, and to elucidate any effects relating to health and disease.

Metabolites increased in PD and iRBD and functions (KOs)

We correlated the normalized intensities of 2-HP with relative abundances of KOs to identify gene functions associated with this molecule (Supplementary Table 19, Extended Data Fig 4f). With a cut-off of fdr-corrected $P< 0.05$ for all three data types (MG, MT and MP) and $|r| > 0.4$ for MG and MT data (no r-cut-off for MP due to smaller sample set), the largest number of correlations were found in the MT data. Considering the pathways represented by significantly correlated KOs, methane metabolism was among the 10 most common ones for both Pearson and Spearman correlations, further highlighting the connection to methanoarchaeal metabolism. Other commonly positively correlated pathways which were in the top 10 for both Pearson and Spearman correlations included carbon metabolism and biosynthesis of cofactors.

Multi-omics: DIABLO analysis

DIABLO, a method from the R package mixOmics[36], allows a multi-block analysis of all omic data types. It was implemented to integrate the MG, MT, MP and metabolomic data. Classification of the samples into the three groups (PD, iRBD and controls) resulted in suboptimal performance (data not shown), related to similarities between the PD and iRBD

800 groups, so the classification was performed by considering PD and iRBD as a single, combined
801 group of cases.

802 Prior to running the comparisons, a design matrix was built based on PLS of pairs of data types;
803 the results suggested that out of the different data types, the highest correlations were apparent
804 for MT-resolved taxa and MT-resolved KOs, as well as MG-resolved taxa with MT-resolved
805 taxa, while the MG and MP KOs were the least correlated (Extended Data Fig 5a). Based on
806 the tuning for the number of components, the optimal choice was to use 2 components
807 (Extended Data Fig 5b).

808 When plotting individual samples per data type (Extended Data Fig 5c), the results resembled
809 the IPCA analysis (Extended Data Fig 2) in that the clearest separation between groups was
810 observed for metabolites and MT KOs. Out of the selected features (Extended Data Figs 5d-e),
811 the two metabolites that were increased in cases were on the list, with 2-HP on component 1
812 and β-glutamate on component 2 (Extended Data Fig 5e). For the species-level taxonomic data,
813 the *Methanobrevibacter smithii* mOTU was among the selected features for component 1 in
814 both MG and MT data, and had a negative loading, similarly to 2-HP, again highlighting their
815 connection. Two other species from the taxon-metabolite correlations (Extended Data Fig 4d)
816 were also selected on the 1st component in both MG and MT data.

817 Out of the KOs, only one was among the selected features in more than one data type: K05979
818 (comB, 2-phosphosulfolactate phosphatase), which was on the 1st component for both MT and
819 MG data. This KO was also significantly positively correlated with 2-HP and β-glutamate in
820 the MG and MT data in the simple correlations of metabolites and KOs (Supplementary Table
821 10). Importantly, 2-phosphosulfolactate phosphatase is an enzyme related to coenzyme M
822 biosynthesis[73], which is present in *Methanobrevibacter* sp., providing further support for a
823 connection between this archaeal genus and the two key metabolites identified. Along the same
824 lines, an increased abundance of the coenzyme M pathway in PD was also highlighted based

on comparisons of 16S rRNA gene amplicon-based functional predictions in a meta-analysis of previous PD microbiome studies[47].

Archaea-specific comparisons

Methanobrevibacter smithii abundance

Despite the results implying an association between *Methanobrevibacter smithii* and the metabolites which were increased in cases, there was no significant difference in the abundance of the *Methanobrevibacter smithii* mOTU or the genus *Methanobrevibacter* between groups in MG, MT (Supplementary Table 5) or MP (Supplementary Table 6) based comparisons, although the uncorrected *P*-value for the two-group cases vs control comparison was significant in the MP data. There was also a slight numeric difference in mean relative abundances between groups in all three types of data, with higher abundances seen in PD and iRBD than in controls (Extended Data Fig 6a-c).

To see whether a significant difference might be found in a larger set of samples, we used 16S rRNA gene amplicon data from NCER-PD LuxPark cohort[21]. In the dataset from this larger cohort, we did indeed detect a significant difference for the genus *Methanobrevibacter* (Extended Data Fig 6d, Supplementary Table 20). We hypothesize that due to *Methanobrevibacter* being an overall low abundance taxon and the workflow not being optimised for archaeal detection, the smaller cohort used in our study lacks the power to detect a significant difference between the groups.

2-HP and methanogenesis-related functions

To further explore the potential connection of 2-HP and *M. smithii*, we calculated correlations between 2-HP and all methanogenesis-related *M. smithii* KOs based on the KEGG database. When considering the full data without accounting for the groups, more than half of these KOs were significantly correlated with 2-HP in both the MG and MT data (Extended Data Fig 6e). Furthermore, there were very few significant KOs when correlations were calculated using only data from control subjects, while the results for the cases only looked similar to those using the full data (without considering groups).

A gene of specific interest in relation to 2-HP is 5,10-methenyltetrahydromethanopterin hydrogenase (Hmd; K13942), which has a cofactor that contains a 2-pyridone structure. This enzyme catalyses an intermediate step in methane formation from H_2 and CO_2 in hydrogenotrophic methanogenic archaea[74], of which *M. smithii* is the key representative in the human gut[71]. The biosynthesis genes for the 2-pyridone-containing cofactor are known as the Hcg genes (Hmd co-occurring genes)[75]. As these genes are not widely studied and are not included in the KEGG database, they were not available in our KO-annotated functional data. We therefore performed a separate search for two of these genes, hmdB (also known as HcgA) and hmdC (also known as HcgG), in predicted protein data from co-assemblies of our MG and MT data, and were able to detect them. Overall, they were however not very abundant: a total of 22% of control, 30% of PD and 33% of iRBD samples allowed detection of one or both genes (Extended Data Fig 6f). Samples with one or both genes present also had higher 2-HP normalized peak areas (Extended Data Fig 6g). Additional experimental work is required to establish the exact connections of 2-HP, *M. smithii* and methanogenic metabolism.

Mouse model

In order to investigate any possible causal effects between 2-HP and Parkinson's disease-linked phenotypes *in vivo*, we performed intrastriatal injections of 2-HP into wild-type mice and a corresponding disease model[41,76]. Intrastriatal injection represents an established experimental approach for studying neurodegeneration in mice. It is for example used to investigate the effect of factors such as preformed α-synuclein fibrils[77] or the pesticide/insecticide rotenone[78] on disease initiation and progression. We chose this established *in vivo* approach as, in the context of the link between 2-HP and hydrogenotrophic archaea in the gut, classical *in vivo* experimental approaches are severely limiting and do presently not allow the inference of causal mechanisms. More specifically, monocolonisation by archaea and fecal matter transplantation (FMT) from patients are confounded by the need for bacterial counterparts and cofactors to sustain the archaea *in* vivo[71]. Furthermore, community-based inoculation or FMT does not provide the necessary colonisation with archaea in mice. The latter point is in particular evident from the data by Sampson et al.[79], where the microbial community data show that only a negligible number of mice had archaea in their gut microbiome post FMT from human donors. Taken together, this highlights the need for the development of new *in vivo* experimental approaches tailored towards unravelling the role of the archaea in the gut microbiome. Such a development does however go beyond the scope of the present work.

Supplementary Notes

NCER-PD Consortium Members

Geeta Acharya, Gloria Aguayo, Myriam Alexandre, Dominic Allen, Wim Ammerlann, Maike Aurich, Federico Baldini, Rudi Balling, Peter Banda, Katy Beaumont, Regina Becker, Camille Bellora, Guy Berchem, Daniela Berg, Fay Betsou, Sylvia Binck, Alexandre Bisdorff, Dheeraj Bobbili, Sandy Boly, Kathrin Brockmann, Jessica Calmes, Lorieza Castillo, Nico Diederich, Rene Dondelinger, Daniela Esteves, Guy Fagherazzi, Tessy Fautsch, Jean-Yves Ferrand, Ronan Fleming, Manon Gantenbein, Thomas Gasser, Piotr Gawron, Lars Geffers, Virginie Giarmana, Enrico Glaab, Clarissa Gomes, Nikolai Goncharenko, Jérôme Graas, Mariela Graziano, Valentin Groues, Anne Grünewald, Wei Gu, Gaël Hammot, Anne-Marie Hanff, Linda Hansen, Maxime Hansen, Hulda Haraldsdöttir, Laurent Heirendt, Estelle Henry, Sylvia Herbrink, Johannes Hertel, Sascha Herzinger, Michael Heymann, Karsten Hiller, Geraldine Hipp, Michele Hu, Laetitia Huiart, Alexander Hundt, Nadine Jacoby, Jacek Jaroslaw, Yohan Jaroz, Pierre Kolber, Rejko Krüger, Joachim Kutzera, Pauline Lambert, Zied Landoulsi, Catherine Larue, Roseline Lentz, Inga Liepelt, Robert Liszka, Laura Longhino, Victoria Lorentz, Paula Cristina Lupu, Clare Mackay, Walter Maetzler, Katrin Marcus, Guilherme Marques, Tainá Marques, Jan Martens, Piotr Matyjaszczyk, Patrick May, Deborah McIntyre, Francoise Meisch, Myriam Menster, Guilherme Meyers, Maura Minelli, Michel Mittelbronn, Brit Mollenhauer, Kathleen Mommaerts, Carlos Moreno, Friedrich Mühlschlegel, Romain Nati, Ulf Nehrbass, Sarah Nickels, Beatrice Nicolai, Jean-Paul Nicolay, Alberto Noronha, Wolfgang Oertel, Marek Ostaszewski, Sinthuja Pachchek, Claire Pauly, Laure Pauly, Lukas Pavelka, Magali Perquin, Rajesh Rawal, Dorothea Reiter, Eduardo Rosales, Isabel Rosety, Kirsten Rump, Estelle Sandt, Venkata Satagopam, Marc Schlesser, Margaux Schmitt, Sabine Schmitz, Susanne Schmitz, Reinhard Schneider, Jens Schwamborn, Alexandra Schweicher, Amir Sharify, Kate Sokolowska, Lara Stute, Olivier Terwindt, Ines Thiele, Hermann Thien, Cyrille Thinnes, Joana Torre, Christophe Trefois, Jean-Pierre Trezzi, Johanna Trouet, Michel Vaillant, Daniel Vasco, Maharshi Vyas, Richard Wade-Martins, Femke Wauters, Paul Wilmes.

References

1. Mollenhauer, B. *et al.* Monitoring of 30 marker candidates in early Parkinson disease as progression markers. *Neurology* **87**, 168–177 (2016).

2. Mollenhauer, B. *et al.* Nonmotor and diagnostic findings in subjects with de novo Parkinson disease of the DeNoPa cohort. *Neurology* **81**, 1226–1234 (2013).

3. 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 Med.* **14**, 795–806 (2013).

4. Hughes, A. J., Daniel, S. E., Kilford, L. & Lees, A. J. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J. Neurol. Neurosurg. Psychiatry* **55**, 181–184 (1992).

5. American Academy of Sleep Medicine. *International classification of sleep disorders*. (2014).

6. Visser, M., Marinus, J., Stiggelbout, A. M. & Hilten, J. J. V. Assessment of autonomic dysfunction in Parkinson's disease: The SCOPA-AUT. *Mov. Disord.* **19**, 1306–1312 (2004).

7. 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. *Mov. Disord.* **22**, 1901–1911 (2007).

8. Fahn, S., Elton, R. L. & Committee, M. of the U. D. Unified Parkinson's Disease Rating Scale. in (eds. Fahn, S., Marsden, C. D., Goldstein, M. & Calne, D. B.) 153–163 (Macmillan Healthcare Information, 1987).

9. Stiasny-Kolster, K. *et al.* The REM sleep behavior disorder screening questionnaire—A new diagnostic instrument. *Mov. Disord.* **22**, 2386–2393 (2007).

10. Hipp, G. *et al.* The Luxembourg Parkinson's Study: A Comprehensive Approach for Stratification and Early Diagnosis. *Front. Aging Neurosci.* **10**, 326 (2018).

11. Narayanasamy, S. *et al.* IMP: a pipeline for reproducible reference-independent integrated metagenomic and metatranscriptomic analyses. *Genome Biol.* **17**, (2016).

12. 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).

13. 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).

14. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**, D457–D462 (2016).

15. El-Gebali, S. *et al.* The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).

16. Milanese, A. *et al.* Microbial abundance, activity and population genomic profiling with mOTUs2. *Nat. Commun.* **10**, 1014 (2019).

17. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 257 (2019).

18. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).

19. Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* **18**, 366–368 (2021).

20. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).

21. Baldini, F. *et al.* Parkinson's disease-associated alterations of the gut microbiome predict disease-relevant changes in metabolic functions. *BMC Biol.* **18**, 62 (2020).

22. Weißbecker, C., Schnabel, B. & Heintz-Buschart, A. Dadasnake, a Snakemake implementation of DADA2 to process amplicon sequencing data for microbial ecology. *GigaScience* **9**, giaa135 (2020).

23. Mölder, F. *et al.* Sustainable data analysis with Snakemake. (2021) doi:10.12688/f1000research.29032.2.

24. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).

25. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).

26. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**, (2013).

27. Schloss, P. D. *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–7541 (2009).

28. Barsnes, H. & Vaudel, M. SearchGUI: A highly adaptable common interface for proteomics search and de novo engines. *J. Proteome Res.* **17**, 2552–2555 (2018).

983 29. US Environmental Protection Agency. Chemical List: Neurotoxicants from PubMed
984 (LITMINEDNEURO). https://comptox.epa.gov/dashboard/chemical-lists/litminedneuro.

985 30. Schymanski, E. L. *et al.* Connecting environmental exposure and neurodegeneration
986 using cheminformatics and high resolution mass spectrometry: potential and challenges.
987 *Environ. Sci. Process. Impacts* **21**, 1426–1445 (2019).

988 31. Chambers, M. C. *et al.* A cross-platform toolkit for mass spectrometry and proteomics.
989 *Nat. Biotechnol.* **30**, 918–920 (2012).

990 32. Kondic, T. *et al. Shinyscreen*. (University of Luxembourg, 2021).

991 33. Ruttkies, C., Schymanski, E. L., Wolf, S., Hollender, J. & Neumann, S. MetFrag
992 relaunched: incorporating strategies beyond in silico fragmentation. *J. Cheminformatics* **8**,
993 3 (2016).

994 34. FiehnLab. MassBank of North America. *MassBank of North America*
995 http://mona.fiehnlab.ucdavis.edu/ (2019).

996 35. Schymanski, E. L. *et al.* Empowering large chemical knowledge bases for exposomics:
997 PubChemLite meets MetFrag. *J. Cheminformatics* **13**, 19 (2021).

998 36. Rohart, F., Gautier, B., Singh, A. & Le Cao, K.-A. mixOmics: An R package for 'omics
999 feature selection and multiple data integration. *PLoS Comput. Biol.* **13**, e1005752 (2017).

1000 37. McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive
1001 analysis and graphics of microbiome census data. *PloS One* **8**, e61217 (2013).

1002 38. Oksanen, J. *et al. vegan: Community Ecology Package*. (2020).

1003 39. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and
1004 dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

40. Buttini, M. *et al.* Expression of Human Apolipoprotein E3 or E4 in the Brains ofApoe−/− Mice: Isoform-Specific Effects on Neurodegeneration. *J. Neurosci.* **19**, 4867–4880 (1999).

41. Ashrafi, A. *et al.* Absence of regulator of G-protein signaling 4 does not protect against dopamine neuron dysfunction and injury in the mouse 6-hydroxydopamine lesion model of Parkinson's disease. *Neurobiol. Aging* **58**, 30–33 (2017).

42. Weis, S. *et al.* Effect of Parkinson's disease and related medications on the composition of the fecal bacterial microbiota. *NPJ Park. Dis.* **5**, (2019).

43. Grün, D. *et al.* Impact of oral COMT-inhibitors on gut microbiota and short chain fatty acids in Parkinson's disease. *Parkinsonism Relat. Disord.* **70**, 20–22 (2020).

44. Jackson, M. A. *et al.* Gut microbiota associations with common diseases and prescription medications in a population-based cohort. *Nat. Commun.* **9**, 2655 (2018).

45. Plassais, J. *et al.* Gut microbiome alpha-diversity is not a marker of Parkinson's disease and multiple sclerosis. *Brain Commun.* **3**, (2021).

46. Boertien, J. M., Pereira, P. A. B., Aho, V. T. E. & Scheperjans, F. Increasing comparability and utility of gut microbiome studies in Parkinson's disease: A systematic review. *J. Park. Dis.* **9**, S297–S312 (2019).

47. Romano, S. *et al.* Meta-analysis of the Parkinson's disease gut microbiome suggests alterations linked to intestinal inflammation. *Npj Park. Dis.* **7**, 1–13 (2021).

48. Heintz-Buschart, A. *et al.* The nasal and gut microbiome in Parkinson's disease and idiopathic rapid eye movement sleep behavior disorder. *Mov. Disord.* **33**, 88–98 (2018).

49. Ji, T. *et al.* Leveraging sequence-based faecal microbial community survey data to identify alterations in gut microbiota among patients with Parkinson's disease. *Eur. J. Neurosci.* **53**, 687–696 (2021).

50. Keshavarzian, A. *et al.* Colonic bacterial composition in Parkinson's disease. *Mov. Disord.* **30**, 1351–1360 (2015).

51. Hill-Burns, E. M. *et al.* Parkinson's disease and Parkinson's disease medications have distinct signatures of the gut microbiome. *Mov. Disord.* **32**, 739–749 (2017).

52. Toh, T. S. *et al.* Gut microbiome in Parkinson's disease: New insights from meta-analysis. *Parkinsonism Relat. Disord.* **94**, 1–9 (2022).

53. Bedarf, J. R. *et al.* Functional implications of microbial and viral gut metagenome changes in early stage L-DOPA-naïve Parkinson's disease patients. *Genome Med.* **9**, (2017).

54. Nishiwaki, H. *et al.* Meta-Analysis of Gut Dysbiosis in Parkinson's Disease. *Mov. Disord.* **35**, 1626–1635 (2020).

55. Shen, T. *et al.* The Association Between the Gut Microbiota and Parkinson's Disease, a Meta-Analysis. *Front. Aging Neurosci.* **13**, (2021).

56. Pietrucci, D. *et al.* Can Gut Microbiota Be a Good Predictor for Parkinson's Disease? A Machine Learning Approach. *Brain Sci.* **10**, (2020).

57. Aho, V. T. E. *et al.* Gut microbiota in Parkinson's disease: Temporal stability and relations to disease progression. *EBioMedicine* **44**, 691–707 (2019).

58. Unger, M. M. *et al.* Short chain fatty acids and gut microbiota differ between patients with Parkinson's disease and age-matched controls. *Parkinsonism Relat. Disord.* **32**, 66–72 (2016).

59. Chen, S. G. *et al.* Exposure to the functional bacterial amyloid protein curli enhances alpha-synuclein aggregation in aged Fischer 344 rats and Caenorhabditis elegans. *Sci. Rep.* **6**, 34477 (2016).

60. Sampson, T. R. *et al.* A gut bacterial amyloid promotes α-synuclein aggregation and motor impairment in mice. *eLife* **9**, e53111 (2020).

61. Heintz-Buschart, A. & Wilmes, P. Human gut microbiome: Function matters. *Trends Microbiol.* **26**, 563–574 (2018).

62. De Weirdt, R. *et al.* Human faecal microbiota display variable patterns of glycerol metabolism. *FEMS Microbiol. Ecol.* **74**, 601–611 (2010).

63. Ho, A. *et al.* Circulating glucuronic acid predicts healthspan and longevity in humans and mice. *Aging* **11**, 7694–7706 (2019).

64. Vascellari, S. *et al.* Clinical Phenotypes of Parkinson's Disease Associate with Distinct Gut Microbiota and Metabolome Enterotypes. *Biomolecules* **11**, 144 (2021).

65. Lautala, P., Ethell, B. T., Taskinen, J. & Burchell, B. The Specificity of Glucuronidation of Entacapone and Tolcapone by Recombinant Human Udp-Glucuronosyltransferases. *Drug Metab. Dispos.* **28**, 1385–1389 (2000).

66. Prueksaritanont, T. *et al.* Glucuronidation of Statins in Animals and Humans: A Novel Mechanism of Statin Lactonization. *Drug Metab. Dispos.* **30**, 505–512 (2002).

67. Aho, V. T. E. *et al.* Relationships of gut microbiota, short-chain fatty acids, inflammation, and the gut barrier in Parkinson's disease. *Mol. Neurodegener.* **16**, (2021).

68. Di Rocco, A., Brannan, T., Prikhojan, A. & Yahr, M. D. Sertraline induced parkinsonim. A case report and an in-vivo study of the effect of sertraline on dopamine metabolism. *J. Neural Transm.* **105**, 247–251 (1998).

1071 69. Samuel, B. S. *et al.* Genomic and metabolic adaptations of Methanobrevibacter smithii
1072 to the human gut. *Proc. Natl. Acad. Sci.* **104**, 10643–10648 (2007).

1073 70. Chibani, C. M. *et al.* A catalogue of 1,167 genomes from the human gut archaeome.
1074 *Nat. Microbiol.* **7**, 48–61 (2022).

1075 71. Ruaud, A. *et al.* Syntrophy via Interspecies H2 Transfer between Christensenella and
1076 Methanobrevibacter Underlies Their Global Cooccurrence in the Human Gut. *mBio* **11**,
1077 (2020).

1078 72. Samuel, B. S. & Gordon, J. I. A humanized gnotobiotic mouse model of host–archaeal–
1079 bacterial mutualism. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 10011–10016 (2006).

1080 73. Graham, D. E., Graupner, M., Xu, H. & White, R. H. Identification of coenzyme M
1081 biosynthetic 2-phosphosulfolactate phosphatase. *Eur. J. Biochem.* **268**, 5176–5188 (2001).

1082 74. Schick, M. *et al.* Biosynthesis of the iron-guanylylpyridinol cofactor of [Fe]-
1083 hydrogenase in methanogenic archaea as elucidated by stable-isotope labeling. *J. Am. Chem.*
1084 *Soc.* **134**, 3271–3280 (2012).

1085 75. Thauer, R. K. *et al.* Hydrogenases from Methanogenic Archaea, Nickel, a Novel
1086 Cofactor, and H_2 Storage. *Annu. Rev. Biochem.* **79**, 507–536 (2010).

1087 76. Kahle, P. J. *et al.* Subcellular localization of wild-type and Parkinson's disease-
1088 associated mutant α-synuclein in human and transgenic mouse brain. *J. Neurosci.* **20**, 6365–
1089 6373 (2000).

1090 77. Garcia, P. *et al.* Neurodegeneration and neuroinflammation are linked, but independent
1091 of alpha-synuclein inclusions, in a seeding/spreading mouse model of Parkinson's disease.
1092 *Glia* (2022) doi:10.1002/glia.24149.

78. Johnson, M. E. & Bobrovskaya, L. An update on the rotenone models of Parkinson's disease: Their ability to reproduce the features of clinical disease and model gene–environment interactions. *NeuroToxicology* **46**, 101–116 (2015).

79. Sampson, T. R. *et al.* Gut microbiota regulate motor deficits and neuroinflammation in a model of Parkinson's disease. *Cell* **167**, 1469-1480.e12 (2016).