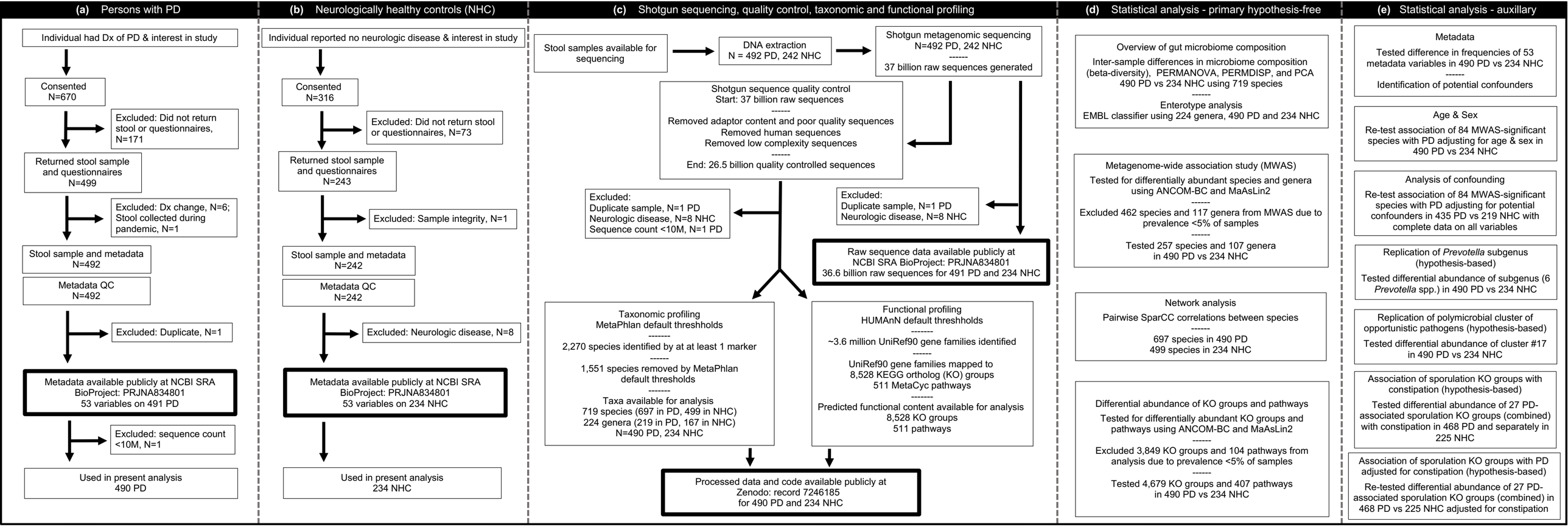
(a) Persons with PD
Individual had Dx of PD & interest in study
Consented N=670
Excluded: Did not return stool or questionnaires, N=171
Returned stool sample and questionnaires N=499
Excluded: Dx change, N=6; Stool collected during pandemic, N=1
Stool sample and metadata N=492
Metadata QC N=492
Excluded: Duplicate, N=1
Metadata available publicly at NCBI SRA BioProject: PRJNA834801 53 variables on 491 PD
Excluded: sequence count <10M, N=1
Used in present analysis 490 PD
(b) Neurologically healthy controls (NHC)
Individual reported no neurologic disease & interest in study
Consented N=316
Excluded: Did not return stool or questionnaires, N=73
Returned stool sample and questionnaires N=243
Excluded: Sample integrity, N=1
Stool sample and metadata N=242
Metadata QC N=242
Excluded: Neurologic disease, N=8
Metadata available publicly at NCBI SRA BioProject: PRJNA834801 53 variables on 234 NHC
Used in present analysis 234 NHC
(c) Shotgun sequencing, quality control, taxonomic and functional profiling
Stool samples available for sequencing
DNA extraction N = 492 PD, 242 NHC
Shotgun metagenomic sequencing N=492 PD, 242 NHC
37 billion raw sequences generated
Shotgun sequence quality control
Start: 37 billion raw sequences
Removed adaptor content and poor quality sequences
Removed human sequences
Removed low complexity sequences
End: 26.5 billion quality controlled sequences
Excluded: Duplicate sample, N=1 PD Neurologic disease, N=8 NHC Sequence count <10M, N=1 PD
Excluded: Duplicate sample, N=1 PD Neurologic disease, N=8 NHC
Raw sequence data available publicly at NCBI SRA BioProject: PRJNA834801 36.6 billion raw sequences for 491 PD and 234 NHC
Taxonomic profiling MetaPhlan default threshholds
2,270 species identified by at at least 1 marker
1,551 species removed by MetaPhlan default thresholds
Taxa available for analysis 719 species (697 in PD, 499 in NHC) 224 genera (219 in PD, 167 in NHC) N=490 PD, 234 NHC
Functional profiling HUMAnN default threshholds
~3.6 million UniRef90 gene families identified
UniRef90 gene families mapped to 8,528 KEGG ortholog (KO) groups 511 MetaCyc pathways
Predicted functional content available for analysis 8,528 KO groups 511 pathways
Processed data and code available publicly at Zenodo: record 7246185 for 490 PD and 234 NHC
(d) Statistical analysis - primary hypothesis-free
Overview of gut microbiome composition
Inter-sample differences in microbiome composition (beta-diversity), PERMANOVA, PERMDISP, and PCA 490 PD vs 234 NHC using 719 species
Enterotype analysis
EMBL classifier using 224 genera, 490 PD and 234 NHC
Metagenome-wide association study (MWAS)
Tested for differentially abundant species and genera using ANCOM-BC and MaAsLin2
Excluded 462 species and 117 genera from MWAS due to prevalence <5% of samples
Tested 257 species and 107 genera in 490 PD vs 234 NHC
Network analysis
Pairwise SparCC correlations between species
697 species in 490 PD 499 species in 234 NHC
Differential abundance of KO groups and pathways
Tested for differentially abundant KO groups and pathways using ANCOM-BC and MaAsLin2
Excluded 3,849 KO groups and 104 pathways from analysis due to prevalence <5% of samples
Tested 4,679 KO groups and 407 pathways in 490 PD vs 234 NHC
(e) Statistical analysis - auxillary
Metadata
Tested difference in frequencies of 53 metadata variables in 490 PD vs 234 NHC
Identification of potential confounders
Age & Sex
Re-test association of 84 MWAS-significant species with PD adjusting for age & sex in 490 PD vs 234 NHC
Analysis of confounding
Re-test association of 84 MWAS-significant species with PD adjusting for potential confounders in 435 PD vs 219 NHC with complete data on all variables
Replication of Prevotella subgenus (hypothesis-based)
Tested differential abundance of subgenus (6 Prevotella spp.) in 490 PD vs 234 NHC
Replication of polymicrobial cluster of opportunistic pathogens (hypothesis-based)
Tested differential abundance of cluster #17 in 490 PD vs 234 NHC
Association of sporulation KO groups with constipation (hypothesis-based)
Tested differential abundance of 27 PD-associated sporulation KO groups (combined) with constipation in 468 PD and separately in 225 NHC
Association of sporulation KO groups with PD adjusted for constipation (hypothesis-based)
Re-tested differential abundance of 27 PD-associated sporulation KO groups (combined) in 468 PD vs 225 NHC adjusted for constipation